Diffraction corrections to the equilibrium properties of the classical electron gas. Pair correlation function

C. Deutsch and M. M. Gombert

Laboratoire de Physique des Plasmas, Université Paris XI, Centre d'Orsay, 91405-Orsay, France
(Received 1 December 1975)

We perform a systematic study of the diffraction corrections ($\hbar \neq 0$) in the high-temperature range ($k_B T > 1$ Ry) for the pair correlation function of the one-component classical electron gas with a neutralizing background, up to the third-order in the plasma parameter $\Lambda = e^2/k_B T\lambda_D$. This program is achieved through the effective interaction $V_{ee}(r) = (e^2/r)(1 - e^{-cr})$ with $c \sim$ (thermal De Broglie wavelength)$^{-1}$, allowing for a straightforward and tractable generalization of the one-component classical plasma model. The nodal expansion of the potential of average force is performed order-by-order with finite Mayer-Salpeter diagrams. The classical results of De Witt (second-order) and Cohen–Murphy (third-order) are recovered in the $\hbar \to 0$ limit. The resummation to all orders of the bubble diagrams gives access to the short-range behavior of the pair correlation function, which is found similar with the Monte-Carlo results.

I. INTRODUCTION

It is a well-known fact that the diffraction effects due to the uncertainty principle are not negligible in the high-temperature electron gas, while the symmetry effects (Fermi statistics) may be safely ignored.

More especially, the diffraction effects are the only quantum corrections of some importance to the equilibrium properties of the classical one-component plasma with a neutralizing background when the Landau length $e^2/k_B T$ becomes smaller than the electron thermal wavelength $\lambda\!\!\!^{-} = \hbar/(2\pi m_e k_B T)^{1/2}$ with $k_B T > 1$ Ry ($T \geqslant 135000\,°$K). As shown be De Witt,[1] it is possible to perform a direct approach to this problem through an exact quantum many-body calculation. However, as is often the case, this rigorous approach is not amenable to a complete treatment, in view of formidable computational difficulties. Fortunately, it is possible to go far enough along this way to obtain the first-order corrections in the quantum parameter $\gamma = \lambda\!\!\!^{-}/\lambda_D = h\omega_p/k_B T$ (ω_p = electron plasma frequency) to the expansion of the canonical free energy with respect to the classical plasma parameter $\Lambda = e^2/k_B T\lambda_D$. Therefore, taking these exact results as a starting point we are allowed to pay attention to another approach to the same problem. This amounts to replacing the classical Coulomb interaction r^{-1} with the effective temperature-dependent interaction[2]

$$V_{ee}(r) = \frac{e^2}{r}(1 - e^{-cr}), \quad c = \frac{1}{\sqrt{2}\lambda\!\!\!^{-}}, \quad \cdots. \tag{I.1}$$

This latter may be introduced in the nodal expansion[3,4] of the potential of average force

$$W_2(r) = k_B T \ln g_2(r), \quad g_2(r) = \text{pair correlation function}, \tag{I.2}$$

with respect to Λ. Obviously, such an expansion is meaningful only for temperature and densities fulfilling the double inequality

$$\frac{e^2}{k_B T} \leqslant \lambda\!\!\!^{-} < \lambda_D, \quad \Lambda \leqslant 1. \tag{I.3}$$

Although known for a long time[5] and despite its appealing simplicity, Eq. (I.1) has not been much considered as a valuable tool for the computation of the diffraction corrections to the thermodynamic properties of the electron gas taken as a collection of charged Boltzmann particles. It must be mentioned that it was used[6] to investigate the first-order corrections in γ to the free energy[2,3,6] taken in the ring approximation. Moreover, the functional form (I.1) has been mostly considered[6] as an appropriate description for soft extended charges used in electrolyte theory in the framework of the soft-sphere model [$c \sim$ (average ion diameter)$^{-1}$]. Apparently, this status has something in common with a lack of confidence in the capability of such a simple expression to quantitatively reproduce the diffraction effects. However, several completely[4] independent derivations reproduce the same effective interaction, while recent numerical studies have given additional support to it. Eq. (I.1) provides an excellent approximation to the effective potential in the medium and high-temperature regime where the quantum effects arising from symmetry and diffraction may be safely decoupled from each other, so that the wavefunction of the interacting electron gas may be given the symmetry of the ideal fermion gas. The latter is negligible for $k_B T \gtrsim 1$ Ry, and we shall restrict ourselves to Boltzmann statistics in the sequel. In Sec. II, a direct derivation of Eq. (I.1) without spurious approximation due to Kelbg[5] makes the above statements clear to the reader. With Sec. III, we develop the Λ-expansion of the pair correlation function and obtain the diffraction corrections for the classical Mayer—Salpeter graphs.[4] The corresponding Debye interaction (first-order) retains the finite $r = 0$ behavior of Eq. (I.1), so that the Fourier transform of any power of it is also finite. As a consequence, the nodal expansion may be pursued order-by-order to high-order without further short-range resummation of the Meeron graphs.[7] This feature appears as a very important one, for it allows the effective interaction (I.1) to provide a straightforward and tractable generalization of the well-known one-component classical plasma model reached in the $\hbar \to 0$ limit. As a by-product, we obtain the possibility to completely visualize the structure of the nodal expansion. This explains why we give peculiar attention to the third-order (Sec. IV) where new qualitative features such as nonconvolution (Bridge) graphs appear for the first time. Such a study paves the way to a systematic investigation of the asymptotic behavior[8] of $g_2(r)$, through a resummation to all orders of the appropriate diagrams. The canonical thermodynamic functions will be considered at length in another work.[9]

II. THE EFFECTIVE POTENTIAL

The purpose of this section is to clarify the kind of assumptions underlying the derivation of the effective interaction (I.1). We are interested in a two-body interaction retaining the diffraction corrections arising from the uncertainty principle, i.e., we consider a gas of Boltzmann wavepackets and neglect Fermi statistics. Therefore, we restrict ourselves to a high-temperature plasma ($k_BT \geq 1$ Ry), although as shown below, the effective interaction (I.1) remains a good approximation at lower temperatures. Among the various derivations[2,10] of Eq. (I.1), we choose the simpler one due to Kelbg.[5] It is based upon the familiar idea, going back to Wigner and Kirkwood and also considered by Dunn and Broyles, of approximating the two-body high-temperature Slater sum with a Gibbs expression through the ansatz

$$\exp[-\beta(H_0+H')] = \exp(-\beta H')\exp(-\beta H_0)G, \tag{II.1}$$

where

$$H_0 = \frac{1}{2m_e}\sum_{k=1}^{N} P_K^2, \quad H' = \sum_{1\leq k<l\leq N} \frac{e^2}{|\bar{r}_k - \bar{r}_l|}, \tag{II.2}$$

and G is a measure of the noncommutativity of H and H' in the small $\beta=(k_BT)^{-1}$ range, given as a solution of the Bloch-like equation

$$\frac{dG}{d\beta} = \exp(\beta H_0)[H_0 - \exp(\beta H')H_0\exp(-\beta H')]\exp(-\beta H_0)G. \tag{II.3}$$

Expanding the bracket with respect to β, one gets a series stopping exactly after the second-order, i.e.,

$$\frac{dG}{d\beta} = -\exp(\beta H_0)\left[\beta[H', H_0] + \frac{\beta^2}{2}[H',[H',H_0]]\right] \times \exp(-\beta H_0)G. \tag{II.4}$$

Restricting ourselves to the term linear in H', we have

$$G = 1 + \int_0^{\beta}\frac{d}{d\beta_1}\exp(\beta_1 H_0)H'\exp(-\beta_1 H_0)\,d\beta_1, \tag{II.5}$$

which allows the density operator $\rho=\exp[\beta(F-H)]$ (F=free energy, $H=H_0+H'$), to be given the form

$$\begin{aligned}\rho\exp(-\beta F) &= \exp(-\beta H)\\ &= \exp(-\beta H')\exp(-\beta H_0) + \exp(-\beta H')\\ &\quad\times\int_0^{\beta}\beta_1\frac{d}{d\beta_1}[\exp(\beta_1-\beta)H'\exp(-(\beta_1-\beta)H_0)]\\ &\quad\times\exp(-\beta H_0)\,d\beta_1.\end{aligned} \tag{II.6}$$

The Slater sum is realized in the coordinate representation $r^N = \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ by

$$\begin{aligned}\langle r^N|\exp(-\beta H)|r^N\rangle &= \left(\frac{2\pi m_e}{h^2\beta}\right)^{3N/2}\exp\left(-\frac{\beta}{8\pi^3}\right)\sum_{1\leq i<j\leq N} e_ie_j\\ &\quad\times\int V_q(k)\exp[-i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)]\,d\mathbf{k},\end{aligned} \tag{II.7}$$

with

$$\begin{aligned}V_q(k) &= \frac{4\pi}{k^2}\int_0^1\exp\left[-\alpha(1-\alpha)\frac{\hbar^2\beta}{m_e}k^2\right]d\alpha\\ &= \frac{4\pi}{k^2}\exp\left(-\frac{\hbar^2\beta k^2}{2m_e}\right){}_1F_1\left(\frac{1}{2},\frac{3}{2};\frac{\hbar^2\beta}{4m_e}k^2\right)\end{aligned} \tag{II.8}$$

in terms of the confluent hypergeometric function. The corresponding effective potential taking into account the diffraction effects is therefore given by the Fourier transform

$$\begin{aligned}u_q(r) &= \frac{1}{2\pi^2 r}\int_0^{\infty} dk\,k\sin kr\,V_q(k)\\ &= r^{-1}\left[1-\exp\left(-\frac{r^2}{2\pi\lambda^2}\right)\right] + \left(\frac{1}{\sqrt{2}\lambda}\right)\left[1-\phi\left(\frac{r}{2\pi\lambda}\right)\right]\end{aligned} \tag{II.9}$$

where $\phi(x) = (2/\sqrt{\pi})\int_0^x \exp(-t^2)\,dt$. The interest of the present derivation lies in the absence of any ad hoc assumptions to reach Eq. (II.9). However, this expression is much too involved to be useful in a nodal expression,[4,11] The diffraction corrections are mostly concentrated in the $r\to 0$ limit where

$$\lim_{r\to 0} u_q(r) = 1/\sqrt{2}\lambda, \tag{II.10}$$

while at large r, $u_q(r)$ reduces to r^{-1}, as it should. So, one is naturally lead to approximate numerically Eq. (II.9) with the simpler one-parameter function

$$u(r) = [1-\exp(-Cr)/r], \quad C = 1/\sqrt{2}\lambda, \tag{II.11}$$

which deviates at most by two per cent from the exact expression (II.9). Table I provides an accurate idea of the status of the effective interaction (I.1). It must be kept in mind that a very small variation of c is able to produce important modifications in the thermodynamic functions.[1,9] Dunn and Broyles[2] have also obtained an expression analogous to (II.11) with $c = (\sqrt{\pi}\lambda)^{-1}$, which does not give as good an agreement with V_{e+e} as does (II.11). Moreover, their expression has been derived as the $T\to\infty$ limit of an involved quadrature through a Bloch-like equation but with additional assumptions such as the validity of the random phase approximation and the neglect of triple correlations which are difficult to assert. Isihara and Wadati[10] also obtained Eq. (II.9) as a first-order approximation in an expression with respect to the interaction in a more general formalism based upon the Montroll—Ward analysis of the partition function in the grand canonical ensemble. The interest of this approach lies in the possibility of including the symmetry effects in a systematic way at lower temperatures, and also of improving these corrections through the inclusion of high order diffraction contributions. The present effort must be considered as a first step in a systematic perturbative program devoted to the calculation of quantum corrections for the high-temperature and nondegenerate classical electron gas. Worthy of note is that the first-order diffractive corrections keep the form (II.9) for all T, while the symmetry zero and first-order contributions

$$V_{ee}^0(r) = \frac{e^2\exp(-r^2/2\pi\lambda^2)}{2\beta}$$

TABLE I. Two particles effective potentials (in rydbergs) at $T=5\times10^4$, 10^6, and 10^4 °K. The distance r is in numbers of the Bohr radius a_0. The last two columns give the numerical values for electrons with antiparallel and parallel spins respectively obtained by Barker (Ref. 12) through a numerical evaluation of the Slater sum.

r	Eq. (II. 9)	Eq. (II. 11)	$2/r$	$V_{e\uparrow e\downarrow}$	$V_{e\uparrow e\uparrow}$
		$T=5\times10^4$ °K			
0	1.4079	1.4079	∞	1.3086	∞
0.5	1.2512	1.1868	4.0000	1.1317	1.8631
1	1.1005	1.0108	2.0000	0.9892	1.3749
15	0.960 82	0.8695	1.3333	0.8642	1.0863
2	0.8358	0.7553	1.0000	0.7572	0.8874
4	0.4946	0.470 07	0.500 00	0.4721	0.4840
6	0.3332	0.3284	0.3333	0.3278	0.3283
		$T=10^6$ °K			
0	6.2964	6.2964	∞	6.3083	∞
0.5	3.4991	3.1712	4.0000	3.4160	6.6855
1	1.9903	1.9141	2.0000	1.9724	2.1961
1.5	1.3333	1.3215	1.3333	1.3300	1.3342
2	1.0000	0.9981	1.0000	0.9995	0.9995
		$T=10^7$ °K			
0	19.9112	19.9112	∞	19.9486	∞
0.5	3.9999	3.9724	4.0000	3.9974	4.0200
1	2.0000	1.9999	2.0000	2.0000	2.0000

$$V'_{ee}(r)=\frac{\sqrt{\pi}}{4}\frac{e^2}{r}-\exp\left[-\frac{1}{2\pi}\left(\frac{r}{\chi}\right)^2\right]\times\int_0^\infty\frac{d\xi\phi(r\xi^{1/2}/\sqrt{2}\beta)}{\xi(1+\xi)}\qquad\text{(II. 12)}$$

are no longer negligible at lower temperatures. Although we do not intend to discuss the thermodynamic functions, it is yet still of interest to comment a little on the reliability of the effective interaction (II. 11) to reproduce the first exact diffraction corrections to the free energy[1] displayed in

$$\frac{\beta(F-F_0)}{N}=-\frac{\Lambda}{3}\left(\Lambda-3\frac{(2\pi)^{1/2}}{2^4}+\frac{\gamma^2}{4}+\cdots\right)+\frac{\Lambda^2}{12}\left[\ln\frac{\lambda}{\lambda_D}+\left(\ln3+\frac{C}{2}-\frac{1}{2}\right)\right].\qquad\text{(II. 13)}$$

It is an easy matter to check out that both the ring sum and the virial methods yield the above dominant terms in the $\hbar\to0$ limit, thus supporting the relevance of expression (II. 11) in the near classical region, where the present analysis is mostly confined.

Nevertheless, it appears possible to obtain a more accurate, effective interaction (especially in the $r\sim\chi$ range) through a Padé interpolation of the short-range results given in Ref. 10 with the asymptotic bare Coulomb interaction. This point will be examined in a forthcoming work. Finally, it must be appreciated that the best support provided to the one-parameter effective interaction (I. 1) arises from the comparison shown in Table I with the two-body Slater sum for electrons with antiparallel spins.[12]

III. PAIR CORRELATION FUNCTION

A. Introduction and first order

The effective interaction (I. 1) is ideally suited to analyze the Λ-expansion of the potential of average force $W_2(r)$. Its finite behavior at $r=0$ allows for an order-by-order evaluation of the Mayer–Salpeter diagrams,[4,7,11] taking into account with the first approximation the long-range resummation of the Coulomb tail, without worrying about the resummation of the short-range behavior at higher order ($n\geq3$). We are thus allowed to proceed step-by-step along the lines of the perturbative expansion detailed in a recent work,[11] where a similar situation was provided by the locally summable two-dimensional Coulomb potential itself. This explains why we skip the fundamentals of the formalism, and proceed directly with the calculations. The first-order approximation in Λ produces the long-range resummation of the bare interaction $r^{-1}(1-e^{-Cr})$ in the form

$$W_2^1(r)=(2\pi)^{-1}\int d\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{r})\frac{V(k)}{1+\beta\rho V(k)}=\frac{e^2e^2}{\alpha_2^2-\alpha_1^2}\,\frac{\exp(-\alpha_1r)-\exp(-\alpha_2r)}{r}\qquad\text{(III. 1)}$$

with

$$\alpha_1^2=\frac{c^2}{2}\left[1-\left(1-\frac{4}{c^2\lambda_D^2}\right)^{1/2}\right],$$

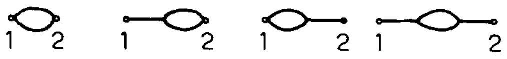
$$\alpha_2^2=\frac{c^2}{2}\left[1+\left(1-\frac{4}{c^2\lambda_D^2}\right)^{1/2}\right],\quad c\lambda_D>2,$$

and ρ, the electron number density. The corresponding high-temperature quantity is

$$W_2^1(r)=\frac{e^2}{r(1-4/c^2\lambda_D^2)^{1/2}}\left[\exp(-\alpha_1r)-\exp(-\alpha_2r)\right],\qquad\text{(III. 2)}$$

$$\lim_{T\to\infty}W_2^1(r)=\frac{e^2}{r}\left[\exp(-r/\lambda_D)-\exp(-cr)\right]\approx\frac{e^2\exp(-r/\lambda_D)}{r}.\qquad\text{(III. 3)}$$

Equation (III. 1) exhibits a nontrivial mixing of $\lambda_D=(k_BT/4\pi\rho e^2)^{1/2}$ and λ at moderate temperature, already obtained by previous authors,[1,6] with the limit behavior $\lim_{r\to0}W_2^1(r)=e^2c/(\alpha_1^2+\alpha_2^2)$.



(2a) (2b) (2c) (2d)

FIG. 1. Second-order Mayer—Salpeter diagrams.

It may also be of interest to consider a formal extension of the expression (III. 1) valid for any $c\lambda_D$ value, in connection with the so-called soft-sphere model. In this case, λ is given the meaning of an average diameter[6] for penetrable charged spheres with[13] $(c\lambda_D < 2)$

$$W_2^1(r) = \frac{2e^2}{r(4/c^2\lambda_D^2 - 1)^{1/2}} \exp - \frac{cr}{\sqrt{2}} \times \sin\left(\frac{cr}{2\sqrt{2}} \; \frac{4}{c^2\lambda_D^2} - 1\right)^{1/2} \qquad \text{(III. 4)}$$

showing periodic oscillations associated with the appearance of the long-range order. In the sequel, we shall restrict ourselves to $c\lambda_D > 2$.

B. Second order[14] (general results)

The second-order contribution $W_2^2(r)$ may again be worked out with the aid of the convolution diagrams, shown in Fig. 1, built upon the Debye chain (III. 1) with $[W_2^b(r) = W_2^c(r) \equiv W_2^{bc}(r)]$

$$g_2(r) = [-\beta W_2^1(r)]^2/2! + 2\beta W_2^{bc}(r) + \beta W_2^d(r). \qquad \text{(III. 5)}$$

The bubble contribution (2a)

$$g_2^a(r) = \frac{\Lambda^2}{2r^2(1 - 4/c^2\lambda_D^2)} [\exp(-\alpha_1 r) - \exp(-\alpha_2 r)]^2 \qquad \text{(III. 6)}$$

is trivial, while the following graphs have to be computed through the Fourier transforms

$$g_2^a(k) = \frac{4\pi\beta^2 e^4}{2k(1 - 4/c^2\lambda_D^2)} \cdot \left(\tan^{-1}\frac{k}{2\alpha_1} + \tan^{-1}\frac{k}{2\alpha_2} - 2\tan^{-1}\frac{k}{\alpha_1 + \alpha_2}\right), \qquad \text{(III. 7)}$$

$$g_2^{bc}(k) = -\beta W_2^{bc}(k) = \rho g_2^a(k)(-\beta W_2^1(k)), \qquad \text{(III. 8)}$$

with the expressions

$$g_2^{bc}(r) = \frac{-\Lambda^2 \,\mathrm{Im}}{2r(1 - 4/c^2\lambda_D^2)^{3/2}} \int_{-\infty}^{\infty} dk \exp(ikr) \times\left(\frac{1}{k^2 + \alpha_1^2} - \frac{1}{k^2 + \alpha_2^2}\right)\left(\tan^{-1}\frac{k}{2\alpha_1} + \tan^{-1}\frac{k}{2\alpha_2} - 2\tan^{-1}\frac{k}{\alpha_1 + \alpha_2}\right) \qquad \text{(III. 9)}$$

and (see Appendix A)

$$I \equiv \int_{-\infty}^{\infty} dk \frac{\exp(ikr)}{k^2 + A^2} \tan^{-1}\frac{k}{B} = \frac{i\pi}{2A}\left[\exp(-Ar)\ln\left(\frac{A+B}{|A-B|}\right) + \exp(-Ar)E_i(-Br + Ar) - \exp(-Ar)E_i(-Br - Ar)\right], \qquad \text{(III. 10)}$$

where $E_i(-x) = -\int_x^\infty dt e^{-t}/t$, $x > 0$. The final result is a very lengthy and cumbersome expression detailed in Ref. 13, which is not illuminating in its own right. This explains why we restrict ourselves to the quasiclassical result with $\alpha_2 \approx c \gg \alpha_1 \approx \lambda_D^{-1}$ and $c\lambda_D \gg 1$, i. e.,

$$\begin{aligned} g_2^{bc}(r) = -\frac{\Lambda^2}{4(1 - 4/c^2\lambda_D^2)^{3/2}} \cdot \Bigg(&\frac{\exp(-\alpha_1 r)}{\alpha_1 r} \ln\frac{3(2\alpha_2 + \alpha_1)\alpha_2^2}{(2\alpha_2 - \alpha_1)(2\alpha_1 + \alpha_2)^2} - \frac{\exp(-\alpha_2 r)}{\alpha_2 r}\ln\frac{3(2\alpha_1 + \alpha_2)\alpha_1^2}{(\alpha_2 - 2\alpha_1)(2\alpha_2 + \alpha_1)^2} \\ &+ \frac{\exp(-\alpha_1 r)}{\alpha_1 r}[E_i(-\alpha_1 r) + E_i(-2\alpha_2 + \alpha_1 r) - 2E_i(-\alpha_2 r)] - \frac{\exp(-\alpha_2 r)}{\alpha_2 r}[E_i(-\alpha_2 r) + E_i(-\alpha_2 r) - 2E_i(-\alpha_2 r)] \\ &- \frac{\exp(\alpha_1 r)}{\alpha_1 r}[E_i(-3\alpha_1 r) + E_i(-2\alpha_2 r - \alpha_1 r) - 2E_i(-2\alpha_1 r - \alpha_2 r)] \\ &+ \frac{\exp(\alpha_2 r)}{\alpha_2 r}[E_i(-3\alpha_2 r) + E_i(-2\alpha_1 r - \alpha_2 r) - 2E_i(-2\alpha_2 r - \alpha_1 r)]\Bigg). \end{aligned} \qquad \text{(III. 11)}$$

In the same way

$$\begin{aligned} g_2^d(r) &= +\beta W_2^d(r) \\ &= \frac{1}{\lambda_D^2 2\pi r(1 - 4/c^2\lambda_D^2)^2} \cdot \mathrm{Im} \int_{-\infty}^{\infty} dk \exp(ikr)\left[\frac{1}{(k^2 + \alpha_1^2)^2} + \frac{1}{(k^2 + \alpha_2^2)^2} + \frac{2}{\alpha_2^2 - \alpha_1^2}\left(\frac{1}{k^2 + \alpha_2^2} - \frac{1}{k^2 + \alpha_1^2}\right)\right] \\ &\quad \cdot \left(\tan^{-1}\frac{k}{2\alpha_1} + \tan^{-1}\frac{k}{2\alpha_2} - 2\tan^{-1}\frac{k}{\alpha_1 + \alpha_2}\right) \end{aligned} \qquad \text{(III. 12)}$$

is evaluated with the aid of the derivative of Eq. (III. 10) with respect to A.

$$\begin{aligned} &\int_{-\infty}^{\infty} dk \frac{\exp(ikr)}{(k^2 + A^2)} \tan^{-1}\frac{k}{B} \\ &= \frac{i\pi}{4A^2} \cdot \left(\frac{2B}{B^2 - A^2}[\exp(-Br) - \exp(-Ar)] + (r + A^{-1})\exp(-Ar)E_i(-Br + Ar) \right. \\ &\quad \left. + (r - A)^{-1}\exp(Ar)E_i(-Br - Ar) + (r + A^{-1})\exp(-Ar)\ln\frac{A+B}{|A-B|}\right) \end{aligned} \qquad \text{(III. 13)}$$

in the quasiclassical approximation

$$g_2^d(r) = \frac{\Lambda^2}{2\lambda_D^2(1-4/C^2\lambda_D^2)^2}\Bigg\{\frac{\exp(-\alpha_1 r)}{\alpha_1 r}\Bigg[\left(\frac{\alpha_1 r+1}{4\alpha_1^2} - \frac{1}{\alpha_2^2-\alpha_1^2}\right)\Big(E_i(-2\alpha_2 r+\alpha_1 r)+E_i(-\alpha_1 r)-2E_i(-\alpha_2 r)$$

$$+\ln\frac{3(\alpha_1+2\alpha_2)\alpha_2^2}{(2\alpha_1-\alpha_1)(2\alpha_1+\alpha_2)^2}\Big) - \frac{(\alpha_2-\alpha_1)^2}{2\alpha_1^2}\left(\frac{1}{\alpha_2(2\alpha_1+\alpha_2)}+\frac{1}{3(2\alpha_1+\alpha_2)(2\alpha_2+\alpha_1)}\right)\Bigg] + \frac{\exp(-\alpha_2 r)}{\alpha_2 r}\Bigg[\left(\frac{\alpha_2 r+1}{4\alpha_2^2}+\frac{1}{\alpha_2^2-\alpha_1^2}\right)$$

$$\cdot\left(E_i(-2\alpha_1 r+\alpha_2 r)+E_i(-\alpha_2 r)-2E_i(-\alpha_1 r)+\ln\frac{3(\alpha_2+2\alpha_1)\alpha_1^2}{(\alpha_2-2\alpha_1)(2\alpha_2+\alpha_1)^2}\right) - \frac{(\alpha_2-\alpha_1)^2}{2\alpha_2^2}$$

$$\cdot\left(\frac{1}{\alpha_1(2\alpha_1-\alpha_2)}+\frac{1}{3(2\alpha_1+\alpha_2)(2\alpha_2+\alpha_1)}\right)\Bigg] + \frac{\exp(+\alpha_1 r)}{\alpha_1 r}\left(\frac{\alpha_1 r-1}{4\alpha_1^2}+\frac{1}{\alpha_2^2-\alpha_1^2}\right)$$

$$\times[E_i(-3\alpha_1 r)+E_i(-2\alpha_2 r-\alpha_1 r)-2E_i(-2\alpha_1 r-\alpha_2 r)] + \frac{\exp(\alpha_2 r)}{\alpha_2 r}\left(\frac{\alpha_2 r-1}{4\alpha_2^2}-\frac{1}{\alpha_2^2-\alpha_1^2}\right)$$

$$\times[E_i(-3\alpha_2 r)+E_i(-2\alpha_1 r-\alpha_2 r)-2E_i(-2\alpha_2 r-\alpha_1 r)] + \frac{\exp(-2\alpha_1 r)}{\alpha_1 r}\left(\frac{1}{3\alpha_1^2}+\frac{\alpha_1^2}{\alpha_2^2(4\alpha_1^2-\alpha_2^2)}\right)$$

$$+\frac{\exp(-2\alpha_2 r)}{\alpha_2 r}\cdot\left(\frac{1}{3\alpha_2^2}+\frac{\alpha_2^2}{\alpha_1^2(4\alpha_2^2-\alpha_1^2)}\right) - \frac{\exp[-(\alpha_1+\alpha_2)r]}{(\alpha_1+\alpha_2)r}\left(\frac{1}{2\alpha_1}+\frac{1}{2\alpha_2}+\frac{\alpha_2}{2\alpha_1(2\alpha_1+\alpha_2)}+\frac{\alpha_1}{2\alpha_2(2\alpha_2+\alpha_1)}\right)\Bigg\}. \quad \text{(III. 14)}$$

C. Second order (classical limits)

In order to recover the familiar one-component plasma model ($\hbar=0$) results,[4] and also to see how the first diffraction corrections appear, we explicit the formulas (III. 6), (III. 11), (III. 14) with

$$\alpha_1 \approx \lambda_D^{-1}\left(1+\frac{1}{2c^2\lambda_D^2}\right), \quad \alpha_2 \approx c\left(1-\frac{1}{2c^2\lambda_D^2}\right). \quad \text{(III. 15)}$$

So, we first obtain

$$\lim_{(c\lambda_D)^{-1}\to 0} g_2^d(r) \sim \frac{\Lambda^2}{2r^2}[\exp(-2r) - 2\exp(-r-c\lambda_D r),$$

$$r \text{ is a number of } \lambda_D, \quad \text{(III. 16)}$$

as a sum of the well-known classical term and a quantum correction vanishing with $(c\lambda_D)^{-1}$. The relation

$$E_i(-x+\epsilon) = E_i(-x) - \frac{\epsilon e^{-x}}{x}, \quad \epsilon \ll x$$

gives

$$g_2^{bc}(r) \underset{(c\lambda_D)^{-1}\to 0}{\sim} g_{cl}^{bc}(r) + g_{qu}^{bc}(r), \quad r \text{ is a number of } \lambda_D,$$

$$g_{cl}^{bc}(r) = -\frac{\Lambda^2}{4}\left[\frac{e^{-r}}{r}\ln 3 + \frac{e^{-r}}{r}E_i(-r) - \frac{e^{-r}}{r}E_i(-3r)\right], \quad \text{(III. 17)}$$

$$g_{cl}^{qu}(r) = \frac{\Lambda^2}{2}\left[\frac{2e^{-r}}{c\lambda_D r} + \frac{\exp(-rc\lambda_D)}{c\lambda_D r}\times\left(\ln\frac{\sqrt{3}}{2c\lambda_D} - \frac{e^{-r}}{r} - 2e^r - E_i(-r)\right)\right],$$

and

$$g_2^d(r) \underset{(c\lambda_D)^{-1}\to 0}{\sim} g_{cl}^d(r) + g_{qu}^d(r), \quad r \text{ is a number of } \lambda_D,$$

$$g_{cl}^d(r) = \frac{\Lambda^2}{8r}[(1+r)e^{-r}\ln 3 - \tfrac{4}{3}(e^{-r}-e^{-2r}) + (1+r)e^{-r}E_i(-r) - (1-r)e^r E_i(-3r)], \quad \text{(III. 18)}$$

$$g_{qu}^d(r) = \frac{\Lambda^2}{8c\lambda_D r}\left[3e^{-r} + 2\exp(-c\Lambda_D r)\times\left((r-1)\frac{e^r}{r} - 2e^r - 2e^{-r}\right)\right].$$

$g_{cl}^{hc}(r)$ and $g_{cl}^d(r)$ have already been obtained by DeWitt.[4] The quantum corrections vanish with $\hbar$, thus allowing for a straightforward generalization of the classical one-component plasma model.

D. Second order (limits behavior)

The finite behavior of Eq. (I. 1) at $r=0$ is expected to survive in the $W_2(r)$ nodal expansion. We already checked this point in the first-order approximation. It will appear in the second-order with the introduction of the well-known relations

$$-E_i(-x) = -\gamma - \ln x - \sum_{n=1}^{\infty}\frac{(-x)^n}{nn!}, \quad x>0$$

$$E_i(x) = \gamma + \ln x + \sum_{n=1}^{\infty}\frac{(-x)^n}{nn!}$$

in $g_2^{bc}(r)$ and $g_2^d(r)$, so that

$$g_2^{bc}(r) \underset{r\to 0}{\sim} -\frac{\Lambda^2}{2(1-4/c^2\lambda_D^2)^{3/2}}\cdot\left[\ln\frac{2(2\alpha_1+\alpha_2)^3}{\alpha_1(\alpha_1+2\alpha_2)^3} + \lambda_D^2 r^2\left(\alpha_1^2\ln\frac{(2\alpha_1+\alpha_2)^2}{3\alpha_1(2\alpha_2+\alpha_1)} - \alpha_2^2\ln\frac{(\alpha_1+2\alpha_2)^2}{3\alpha_2(2\alpha_1+\alpha_2)}\right)\right], \quad \text{(III. 19)}$$

$$g_2^d(r) \underset{r\to 0}{\sim} \frac{\Lambda^2}{\lambda^2(1-4/c^2\lambda_D^2)^2}\cdot\left(\frac{1}{\alpha_2^2-\alpha_1^2}\cdot\ln\frac{\alpha_1(2\alpha_2+\alpha_1)^3}{\alpha_2(\alpha_2+2\alpha_1)^3} + \frac{\alpha_1\alpha_2[9(\alpha_2^4+\alpha_1^4)-16\alpha_1\alpha_2(\alpha_1^2+\alpha_2^2)+18\alpha_1^2\alpha_2^2]-2(\alpha_1^6+\alpha_2^6)}{6\alpha_1^2\alpha_2^2(4\alpha_2^2-\alpha_1^2)(4\alpha_1^2-\alpha_2^2)}\right),$$

while

$$g_2^d(r) \underset{r\to 0}{\sim} \frac{\Lambda^2(\alpha_2-\alpha_1)^2}{2(1-4/c^2\lambda_D^2)}\lambda_D^2. \quad \text{(III. 20)}$$

On the other hand, the asymptotic expressions ($x \gg 1$)

$$-E_i(-x) = x^{-1}e^{-x}\left(\sum_{n=0}^{N}\frac{n!}{(-x)^n} + O(|x|^{-N-1})\right),$$

$$E_i(x) = x^{-1}e^{-x}\left(\sum_{n=0}^{N}\frac{n!}{x^n} = O(|x|^{-N-1})\right),$$

monitor the $r\to\infty$ behavior with (r is number of λ_D)

$$g_2^d(r) \underset{r\to\infty}{\sim} \frac{\Lambda^2\exp(-2\alpha_1 r)}{2r^2(1-4/c^2\lambda_D^2)},$$

$$g_2^{bc}(r) \underset{r\to+\infty}{\sim} \frac{-\Lambda^2}{2(1-4/c^2\lambda_D^2)^{3/2}}\cdot\frac{\exp(-\lambda_D\alpha_1 r)}{\lambda_D\alpha_1 r}\times\ln\frac{3(2\alpha_2+\alpha_1)\alpha_2^2}{(2\alpha_2-\alpha_1)(2\alpha_1+\alpha_2)^2}, \quad \text{(III. 21)}$$

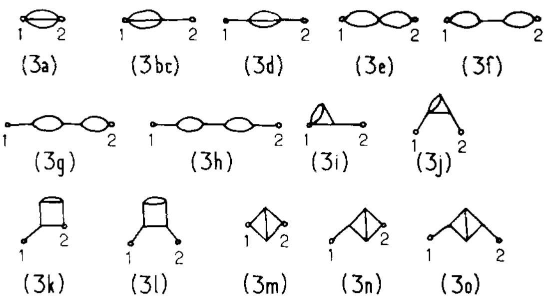


FIG. 2. Numerically different third-order Mayer—Salpeter diagrams.

$$g_2^a(r) \underset{r\to\infty}{\sim} \frac{\Lambda^2 \exp(-\lambda_D \alpha_1 r)}{8\alpha_1^2\lambda_D^2(1-4/c^2\lambda_D^2)} \cdot \ln\frac{3(2\alpha_2+\alpha_1)\alpha_2^2}{(2\alpha_2-\alpha_1)(2\alpha_1+\alpha_2)^2},$$

extending to the quantum situation the well-known asymptotic preeminence[15] of the longest convolution chain.

IV. THIRD-ORDER CONTRIBUTIONS TO $W_2(r)$

A. General

Until now the diffraction contributions appear as corrections to the already well-known classical results and vanish with $\hbar$. From the third-order, the interaction (I. 1) becomes instrumental in allowing for a very simple although significant extension of the classical plasma model. Paralleling the two-dimensional Coulomb gas[11] situation, the Fourier transform of any power of the Debye interaction

$$\int d\mathbf{r} \exp(i\mathbf{k}\cdot\mathbf{r}) C_D^n = \frac{4\pi}{k}\int_0^\infty dr\, r \sin kr \left(\frac{\exp(-\alpha_1 r) - \exp(-\alpha_2 r)}{r}\right)^n < +\infty, \quad \text{all } n, \tag{IV. 1}$$

with C_D = Debye chain, remains finite, while its classical analog

$$\frac{4\pi}{k}\int_0^\infty dr\, r \sin kr \frac{\exp(-nr)}{r^n} < +\infty, \quad n \leq 2, \tag{IV. 2}$$

becomes meaningless for $n \geq 3$. This fortunate behavior allows us to consider the "simple 12-irreducible cluster diagrams" as constructed from Debye chains and nodal points to every order n, without further short-range resummation of the n-bubbles[7] (Meeron) sum,

$$\int_{-1}^{+1} d\mu \int_0^\infty dr\, r^2 \exp(ipr\mu)[\exp(-\Lambda e^{-r}/r) - 1 + \Lambda e^{-r}/r]. \tag{IV. 3}$$

We are left with finite diagrams, in contrast to the purely classical situation[7] ($\hbar = 0$) where some third-order graphs become infinite.

B. Convolution diagrams

With the third-order appears the new qualitative feature that some Mayer-Salpeter graphs are not evaluatable with the standard Fourier-transform convolution techniques. Hopefully, most of them may be computed with the methods already used for the second-order. Keeping in mind a possible extrapolation to high orders,[8] we develop our calculations analytically as far as possible and postpone the numerical evaluation to the very end. As a by-product, we shall elucidate the asymptotic behavior of convolution chains built from second-order graphs taken as basic bubbles, otherwise hidden in a brute force numerical approach.[7] The third-order graphs listed in Fig. 2 are now evaluated as follows:

$$g_3^a(r) = \frac{\Lambda^3}{3!\,(1-4/c^2\lambda_D^2)^{3/2}} \quad \frac{\exp(-\alpha_1 r) - \exp(-\alpha_2 r)^3}{r} \tag{IV. 4}$$

$$g_3^{bc}(r) = \frac{\rho}{2\pi^2 r}\int_0^\infty dk\, k \sin kr\, g_3^a(k)(-\beta W_2^1(k)) \tag{IV. 5}$$

where

$$\begin{aligned} g_3^a(k) = \frac{4\pi\beta^3 e^6}{6(1-4/c^2\lambda_D^2)^{3/2}k}\Bigg\{&\Bigg[3\alpha_1 \tan^{-1}\left(\frac{k}{3\alpha_1}\right) \\ &- 3(2\alpha_1+\alpha_2)\tan^{-1}\frac{k}{2\alpha_1+\alpha_2} + 3(\alpha_1+2\alpha_2) \\ &\times \tan^{-1}\frac{k}{\alpha_1+2\alpha_2} - 3\,\alpha_2\tan^{-1}\frac{k}{3\alpha_2}\Bigg] \\ &+ \frac{k}{2}[\ln(k^2+9\alpha_1^2) - 3\ln(k^2+(2\alpha_1+\alpha_2)^2) \\ &+ 3\ln(k^2+(\alpha_1+2\alpha_2)^2) - \ln(k^2+9\alpha_2^2)]\Bigg\}. \end{aligned} \tag{IV. 6}$$

As before, Eq. (IV. 5) could be given a tractable form with the aid of the quadratures given in the previous section, and with

$$\begin{aligned} J \equiv \operatorname{Im}\int_{-\infty}^{+\infty} \frac{dk\, k \exp(ikr)}{k^2+A^2}\ln\left(\frac{B^2+k^2}{C^2+k^2}\right) &= \exp(-rA)\Bigg[E_i(-Cr+Ar) - E_i(-Br+Ar) \\ &- \ln\left(\frac{A-C}{A-B}\right) - \ln\left(\frac{C+A}{B+A}\right) - \exp(rA)(E_i(-Br-Ar) \\ &- E_i(-Cr-Ar)) \end{aligned} \tag{IV. 7}$$

explicated in Appendix B with $A = \alpha_1, \alpha_2$, $B = 3\alpha_2, \alpha_1 + 2\alpha_2$, $C = 3\alpha_1, \alpha_2 + 2\alpha_1$. The limit values

$$\lim_{r\to 0} J = 0,$$

$$\lim_{r\to\infty} J = \pi\exp(-rA)\ln\frac{B^2-A^2}{C^2-A^2}, \quad A < C$$

$$\lim_{r\to\infty} J = \frac{\pi\exp(-Cr)}{(A+C)r}, \quad C < A,$$

together with the corresponding I values control $\lim_{r\to 0} W_2^{bc}(r)$ and $\lim_{r\to\infty} W_3^{bc}(r)$.

More precisely, $W_3^{bc}(r)$ decreases at infinity as $\exp(-\alpha_1 r/r)$. A convolution chain vanishes faster at infinity when the number of lines within its bubbles increases.

Again, $W_3^a(r)$ may explicated through the derivative with respect to A of both sides of Eq. (IV. 7):

$$\operatorname{Im}\int_{-\infty}^{\infty}\frac{dk\, k\exp(ikr)}{(k^2+A^2)^2}\ln\frac{(k^2+B^2)}{k^2+C^2} = \frac{\pi}{2A}\,\alpha\Big[r\exp(-rA)\Big(E_i(-Cr+Ar) - E_i(-Br+Ar)$$

TABLE II. Limit behaviors of some diagrams (r in number of λ_D).

a. $r \to 0$ with $\hbar \neq 0$ Finite results

b. $r \to 0$ and $h \to 0$

$$g_2^a(r) \sim \frac{\Lambda^2}{2!} c^2 \lambda_D^2, \quad g_3^a(r) \sim -\frac{\Lambda^3}{3!} c^3 \lambda_D^3$$

$$g_2^{bc}(r) \sim -\frac{\Lambda^2}{2!} \ln c\lambda_D, \quad g_3^{bc} \sim \frac{\Lambda^3}{3!} c\lambda_D \ln \frac{32}{27}$$

$$g_2^d(r) \sim \frac{\Lambda^2}{12}, \quad g_3^d(r) \sim -\frac{\Lambda^3}{12} \ln c\lambda_D$$

c. $\lim r \to \infty$

$g_2^a(r) \sim \exp(-2\alpha_1 r)/r^2, \quad g_3^a(r) \sim \exp(-3\alpha_1 r)/r^3$

$g_2^{bc}(r)$ and $g_r^{bc}(r)$ are $O(\exp(-\alpha_1 r)/r)$

$g_2^d(r), \quad g_3^d(r) \sim O(\exp(-\alpha_1 r))$

d. $\lim r \to \infty$ and $\hbar \to 0$

$$g_2^a(r) \sim \frac{\Lambda^2 \exp(-2\alpha_1 r)}{2! r^2}, \quad g_3^a(r) \sim -\frac{\Lambda^3 \exp(-3\alpha_1 r)}{3! r^3}$$

$$g_2^{bc}(r) \sim -\frac{\Lambda^2}{4} \ln 3 \frac{e^{-r}}{r}, \quad g_3^{bc}(r) \simeq \frac{\Lambda^3 e^{-r}}{6r}\left(\ln c\lambda_D - \ln\frac{64}{3} + 1\right)$$

$$g_2^d(r) \sim \frac{\Lambda^2}{8} e^{-r} \ln 3, \quad g_3^d(r) \simeq -\frac{\Lambda^2}{12} e^{-r}\left(\ln c\lambda_D - \ln\frac{64}{3} + 1\right)$$

$$\frac{g_2^d(r)}{g_2^{bc}(r)} = \frac{g_3^d(r)}{g_3^{bc}(r)} = -\frac{r}{2}$$

$$+ \ln\frac{B^2 - A^2}{|C^2 - A^2|}\Bigg) + r\exp(rA)[E_i(-Br - Ar) - E_i(-Cr - Ar)] - [\exp(-Cr) - \exp(-Ar)]\frac{2A}{A^2 - C^2} + [\exp(-Br) - \exp(-Ar)]\frac{2A}{A^2 - B^2}\Bigg]. \qquad \text{(IV. 8)}$$

Equations (IV. 8) and (III. 13) show that $W_3^d(r) \underset{r\to\infty}{\sim} \exp(-\alpha_1 r)$. The limit behaviors of $g_3^{bc}(r)$ and $g_3^d(r)$ are displayed in Table II, together with their second-order counterparts. Their variations with $(c\lambda_D)^{-1}$ are much more important than the corresponding $W_2^{bc}(r)$ and $W_2^d(r)$ ones, as expected from $\lim_{\hbar\to 0} W_3^a(k) \to \infty$. They diverge as $\ln c\lambda_D$ with $r \to 0$, while $W_3^a(r) \underset{r\to 0}{\sim} (\lambda_D c)^3$. Analogous behavior of more compact graphs built upon n-bubbles ($l = n$, $k = 0$) may be extrapolated to $n \geq 4$. The chain diagrams (3e), (3f), (3g), and (3h) are easily estimated from ($n' = 4 + m$, $m = 0, 1, 2, 3$)

$$(-1)^m \frac{\Lambda^3}{(r/\lambda_D) 2\pi (1 - 4/c^2\lambda_D^2)^{n'/2} \lambda_D^{2m}} \int_0^\infty \frac{dk}{k} \sin kr \times \left(\frac{1}{k^2 + \alpha_1^2} - \frac{1}{k^2 + \alpha_2^2}\right)^m \cdot \left(\tan^{-1}\frac{k}{2\alpha_1} + \tan^{-1}\frac{k}{2\alpha_2} - 2\tan^{-1}\frac{k}{\alpha_1 + \alpha_2}\right)^2. \qquad \text{(IV. 9)}$$

Again the graphs decrease in absolute value with $(c\lambda_D)^{-1} \to 0$, until they reach the classical Cohen–Murphy results.[7] The usual asymptotic behavior is recovered. The longest chain remains the more important graph when $r \to \infty$, while its $(c\lambda_D)^{-1} = 0$ limit is obtained (see Table III). Usually, it is reached monotonically from the lower side. However, in its immediate vicinity, the diffraction corrections could produce a larger contribution arising from the $1 + 4n'/c^2\lambda_D^2$ term incompletely cancelled by other factors. For instance, the longest chain (3h) displays the asymptotic behavior

$$\lim_{r\to\infty}(3h) \sim \frac{\Lambda^3}{2\pi(1 - 4/c^2\lambda_D^2)^{7/2}\lambda_D^5} \int_0^\infty \frac{dk \sin kr}{k(k^2 + \alpha_1^2)^3} \tan^{-1}\frac{k}{2\alpha_1}$$
$$\approx \frac{-\Lambda^3(1 + 14/c^2\lambda_D^2)}{2\pi r (2\alpha_1)^2 \lambda_D^2} \int_0^\infty \frac{dk\, k \sin kr}{(k^2 + \alpha_1^2)^3} \qquad \text{(IV. 10)}$$
$$= \frac{-\Lambda^3}{128}\left(1 + \frac{4}{c^2\lambda_D^2}\right)\frac{r}{\alpha_1^4} \cdot \frac{\exp(-r\alpha_1)}{\lambda_D^5}\left(\frac{1}{\alpha_1} + r\right)$$
$$\underset{r\to\infty}{\sim} \frac{-\Lambda^3 r \exp(-\alpha_1 r)}{128\alpha_1^4\lambda_D^5}\left(1 + \frac{14}{c^2\lambda_D^2}\right)$$

TABLE III. Numerically evaluated third-order convolution graphs as a function of reduced distance. The comparison with the classical Cohen–Murphy results[7] is obtained by multiplying these latter with an overall $\pi/2$ factor.

r	2π(3e)	2π(3f)	2π(3g)	2π(3h)
	$(c\lambda_D)^{-1} = 10^{-1}$			
0.2	0.2138	−0.186 09	0.046 88	−0.020 59
0.4	1.1061	−0.156 70	0.044 41	−0.020 09
0.6	0.5340	−0.124 28	0.040 84	−0.019 30
0.8	0.264 98	−0.095 69	0.036 68	−0.018 28
1.0	0.136 29	−0.062 67	0.032 96	−0.017 10
1.2	0.073 36	−0.054 89	0.028 15	−0.015 81
1.4	0.039 44	−0.041 42	0.024 22	−0.014 48
1.6	0.021 96	−0.031 27	0.020 67	−0.013 14
1.8	0.012 44	−0.023 66	0.017 57	−0.011 81
2.0	0.007 15	−0.017 95	0.014 77	−0.010 58
2.4	0.002 44	−0.010 42	0.010 37	−0.008 31
3.0	0.000 513	−0.004 71	0.007 20	−0.005 59
4.0	0.000 04	−0.001 32	0.002 31	−0.002 70
5.0		−0.000 39	0.000 87	−0.001 23
	$(c\lambda_D)^{-1} = 10^{-2}$			
0.2	7.5468	−0.4778	0.091 24	−0.036 53
0.4	2.1349	−0.335 43	0.083 97	−0.035 46
0.6	0.829 88	−0.238 23	0.074 99	−0.033 83
0.8	0.370 31	−0.171 25	0.065 65	−0.031 79
1.0	0.178 76	−0.124 40	0.056 67	−0.029 49
1.2	0.091 33	−0.091 18	0.048 43	−0.027 06
1.4	0.048 93	−0.067 35	0.041 07	−0.024 59
1.6	0.027 10	−0.050 10	0.034 63	−0.022 16
1.8	0.015 16	−0.037 50	0.029 06	−0.019 83
2.0	0.008 4	−0.028 23	0.024 30	−0.017 63
2.4	0.002 77	−0.016 23	0.016 85	−0.013 72
3.0	0.000 63	−0.007 29	0.009 59	−0.009 12
4.0	0.000 10	−0.002 04	0.003 66	−0.004 35
5.0		−0.000 60	0.001 38	−0.001 97
	$(c\lambda_D)^{-1} = 10^{-3}$			
0.2	8.107	−0.524 53	0.097 83	−0.038 93
0.4	2.3134	−0.361 48	0.089 79	−0.037 77
0.6	0.904 83	−0.254 64	0.080 01	−0.036 017
0.8	0.378 779	−0.182 20	0.069 921	−0.033 831
1.0	0.172 61	−0.131 97	0.060 281	−0.031 371
1.2	0.085 256	−0.096 54	0.051 46	−0.028 77
1.4	0.050 94	−0.071 22	0.043 605	−0.026 136
1.6	0.033 23	−0.052 93	0.036 74	−0.023 55
1.8	0.018 80	−0.039 59	0.030 818	−0.021 066
2.0	0.007 41	−0.029 79	0.025 76	−0.018 73
2.4	0.001 52	−0.017 11	0.017 85	−0.014 567
3.0	0.0011	−0.0073	0.010 15	−0.0097
4.0	0.00083	−0.0021	0.0039	−0.0046
5.0		−0.000 64	0.001 46	−0.002 09

with the classical limit reached from above, thus extending to the present situation the well-known asymptotic behavior of the longest chain, i.e.,

$$\lim_{r\to\infty}\int_0^\infty \frac{dk\,k\sin kr}{(k^2+\alpha_1^2)^n}\approx\frac{\pi}{2(n-1)!}\left(\frac{r}{2\alpha_1}\right)^{n-1}\exp(-\alpha_1 r). \tag{IV.11}$$

Now, we should pay attention to the four remaining convolution diagrams (3i) and (3j) built from (2bc). We write them in the from [see Eq. (III.17)]

$$g_3^i(r)=\frac{2\Lambda}{\pi\lambda_D(1-4/c^2\lambda_D^2)r}\int_0^\infty dk\sin kr\left(\frac{1}{k^2+\alpha_1^2}-\frac{1}{k^2+\alpha_2^2}\right)$$

$$\cdot\int_0^\infty dr'\sin kr'[\exp(-\alpha_1 r')-\exp(-\alpha_2 r')]g_2^{bc}(r')$$

$$\equiv I_1+I_2, \tag{IV.12}$$

evaluated with the aid of

$$\int_0^\infty dr'\sin kr'\frac{[\exp(-\alpha_1 r')-\exp(-\alpha_2 r')]}{r'}\exp(-ar')=\tan^{-1}\frac{k}{\alpha_1+a}-\tan^{-1}\frac{k}{\alpha_2+a}, \tag{IV.13}$$

$$\int_a^b dk\tan^{-1}\frac{k}{u}=b\tan^{-1}\frac{k}{b}-a\tan^{-1}\frac{k}{a}-\frac{u}{2}\ln\frac{u^2+b^2}{u^2+a^2},$$

$$E_i(-r')=-r'\int_1^\infty dt\ln t\exp(-r't),$$

as

$$g_3^i(r)=\frac{\Lambda^3}{\pi r}\int_0^\infty dk\sin kr\left(\frac{1}{k^2+\alpha_1^2}-\frac{1}{k^2+\alpha_2^2}\right)G(k)+\frac{\Lambda^3}{2\lambda_D r}\cdot\int_1^\infty dt\ln tH(t), \tag{IV.14}$$

where

$$\begin{aligned}G(k)=&-\left(\frac{\ln 3}{2}+\frac{2}{c\lambda_D}\right)\left(\tan^{-1}\frac{k}{\alpha_1+\lambda_D^{-1}}-\tan^{-1}\frac{k}{\alpha_2+\lambda_D^{-1}}\right)+\ln\left(\frac{\sqrt{3}}{2c\lambda_D}\right)\cdot\frac{1}{c\lambda_D}\cdot\tan^{-1}\left(\frac{k}{\alpha_1+c}-\tan^{-1}\frac{k}{\alpha_2+c}\right)\\&-\frac{2}{c\lambda_D}\left(\tan^{-1}\frac{k}{\alpha_1+c-\lambda_D^{-1}}-\tan^{-1}\frac{k}{\alpha_2+c-\lambda_D^{-1}}\right)+\frac{1}{c\lambda_D}\cdot\left[(c\lambda_D-1+\alpha_2\lambda_D)\tan^{-1}\frac{k}{c-\lambda_D^{-1}+\alpha_2}\right.\\&\left.-(c\lambda_D-1+\alpha_1\lambda_D)\tan^{-1}\frac{k}{c-\lambda_D^{-1}+\alpha_1}\right]+\frac{k\lambda_D}{2}\ln\frac{k^2+(c-\lambda_D^{-1}+\alpha_2)^2}{k^2+(c-\lambda_D^{-1}+\alpha_1)^2}-(c\lambda_D+1+\alpha_2\lambda_D)\tan^{-1}\frac{k}{c+1/\lambda_D+\alpha_2}\\&+(c\lambda_D+1+\alpha_1\lambda_D)\tan^{-1}\frac{k}{c+1/\lambda_D+\alpha_1}-\frac{k\lambda_D}{2}\ln\frac{k^2+(c+\lambda_D^{-1}+\alpha_2)^2}{k^2+(c+\lambda_D^{-1}+\alpha_2)^2},\end{aligned} \tag{IV.15}$$

and

$$\begin{aligned}2H(t)=&\left(1+\frac{2}{c\lambda_D}\right)\left(\frac{\exp(-\alpha_1 r)-\exp[-(\alpha_1+r/\lambda_D)r]}{(\alpha_1+t/\lambda_D)^2-\alpha_1^2}-\frac{\exp(-\alpha_1 r)-\exp[-(\alpha_2+t/\lambda_D)r]}{(\alpha_2+t/\lambda_D)^2-\alpha_1^2}\right.\\&\left.-\left(\frac{\exp(-\alpha_2 r)-\exp[-r(\alpha_1+t/\lambda_D)]}{(\alpha_2+t/\lambda_D)^2-\alpha_2^2}\right)+\frac{\exp(-\alpha_2 r)-\exp[-(\alpha_2+t/\lambda_D)r]}{(\alpha_2+t/\lambda_D)^2-\alpha_2^2}\right)\\&-3\left[\text{same terms with }\alpha_1+\frac{t}{\lambda_D}\to\alpha_1+\frac{3t}{\lambda_D}\text{ and }\alpha_2+\frac{t}{\lambda_D}\to\alpha_2+\frac{3t}{\lambda_D}\right],\end{aligned} \tag{IV.16}$$

with the asymptotic limits

$$\begin{aligned}&I_1\underset{r\to\infty}{\sim}\frac{A}{r/\lambda_D}[\exp(-\alpha_1 r)-\exp(-\alpha_2 r)],\\&A=\frac{\Lambda^3}{2\lambda_D k}G(0)\\&\approx\frac{\Lambda^3}{2\lambda_D}\left[\left(-\frac{\ln 3}{2}+\frac{2}{c\lambda_D}\quad\frac{1}{\alpha_1+\lambda_D^{-1}}-\frac{1}{\alpha_1+\lambda_D^{-1}}\right)+\ln\left(\frac{\sqrt{3}}{2c\lambda_D}\right)\cdot\frac{1}{c\lambda_D}\circ\left(\frac{1}{c+\alpha_1}-\frac{1}{c+\alpha_1}\quad+\frac{2}{c\lambda_D}\quad\frac{1}{\alpha_2+c-\lambda_D^{-1}}-\frac{1}{\alpha_1+c-\lambda_D^{-1}}\right)\right.\\&\left.+c^{-1}\ln\frac{(\alpha_1+c+1/\lambda_D)(\alpha_2+c-1/\lambda_D)}{(\alpha_2+c+1/\lambda_D)(\alpha_1+c-1/\lambda_D)}\right],\\&I_2\underset{r\to\infty}{\sim}\frac{\mathbf{A}\exp(-\alpha_1 r)}{r}+\frac{\mathbf{B}\exp(-\alpha_2 r)}{r},\end{aligned} \tag{IV.17}$$

and the corresponding classical expressions ($\hbar=0$)

$$I_1\underset{r\to\infty}{\sim}-\frac{\Lambda^3}{8}\ln 3\frac{e^{-r}}{r},\quad \mathbf{A}\sim 0.1\Lambda^3,\quad \mathbf{B}\sim O\left(\frac{1}{c^2\lambda_D^2}\right)=0. \tag{IV.17a}$$

In the same way,

$$g_3^j(r)=-\frac{\Lambda^3}{\lambda_D^2\pi r}\cdot\int_0^\infty dk\sin kr\left(\frac{1}{k^2+\alpha_1^2}-\frac{1}{k^2+\alpha_2^2}\right)^2G(k)-\frac{\Lambda^3}{2\lambda_D^2 r}\int_0^\infty dt\ln tH'(t)\equiv I_1'+I_2', \tag{IV.18}$$

$$
\begin{aligned}
H'(t)=&\frac{2}{\alpha_1^2-\alpha_2^2}\,H(t)+\frac{3\{-\exp(-\alpha_1 r)+\exp[-r(\alpha_1+3+\lambda_D^{-1})]\}}{[\alpha_1^2-(\alpha_1+3t/\lambda_D)^2]^2}-\frac{3\exp[-r(\alpha_2+3t/\lambda_D)]-\exp(-\alpha_1 r)}{[\alpha_1^2-(\alpha_1+3t/\lambda_D)^2]^2}\\
&-\left(1+\frac{2}{c\lambda_D}\right)\frac{\exp[-r(\alpha_1+t/\lambda_D)]-\exp(-\alpha_1 r)}{[\alpha_1^2-(\alpha_1+t/\lambda_D)^2]^2}+\left(1+\frac{1}{c\lambda_D}\right)\frac{\exp[-r(\alpha_2+t/\lambda_D)]-\exp(-\alpha_1 r)}{[\alpha_1^2-(\alpha_2+t/\lambda_D)^2]^2}\\
&+\frac{3\{\exp[-r(\alpha_1+3t/\lambda_D)]-\exp(-r\alpha_2)\}}{[\alpha_2^2-(\alpha_1+3t/\lambda_D)]^2}-\frac{3\{\exp[-r(\alpha_2+3t\lambda_D^{-1})]-\exp(-\alpha_2 r)\}}{[\alpha_2^2-(\alpha_2+3t/\lambda_D)^2]^2}\\
&-\left(1+\frac{2}{c\lambda_D}\right)\frac{\exp[-r(\alpha_1+t/\lambda_D)]-\exp(-r\alpha_2)}{[\alpha_2^2-(\alpha_1+t/\lambda_D)^2]^2}+\left(1+\frac{2}{c\lambda_D}\right)\frac{\exp[-r(\alpha_2+t/\lambda_D)]-\exp(-\alpha_2 r)}{[\alpha_2^2-(\alpha_2+t/\lambda_D)^2]^2}\\
&-\frac{r}{2\alpha_1}\exp(-\alpha_1 r)\left(\frac{3}{\alpha-(\alpha_1-3t/\lambda_D)^2}-\frac{3}{\alpha_1^2-(\alpha_2+3t/\lambda_D)^2}-\frac{(1+2/c\lambda_D)}{\alpha_1^2-(\alpha_1+t/\lambda_D)^2}+\frac{(1+2/c\lambda_D)}{\alpha_1^2-(\alpha_2+t/\lambda_D)^2}\right)\\
&-\frac{r}{2\alpha_2}\exp(-\alpha_2 r)\left[\frac{3}{\alpha_2^2-(\alpha_1+3t/\lambda_D)^2}-\frac{3}{\alpha_2^2-(\alpha_2+3t/\lambda_D)^2}-\frac{(1+2/c\lambda_D)}{\alpha_2^2-(\alpha_1+t/\lambda_D)^2}+\frac{(1+2/c\lambda_D)}{\alpha_2^2-(\alpha_2+t/\lambda_D)^2}\right],
\end{aligned}
\tag{IV. 19}
$$

with

$$
I_1' \underset{r\to\infty}{\simeq} \frac{A\exp(-\alpha_1 r)}{2\alpha_1},\quad I_2' \underset{r\to\infty}{\sim} \frac{\mathbf{A}\exp(-\alpha_1 r)}{\alpha_1}-\frac{\mathbf{B}\exp(-\alpha_2 r)}{\alpha_2}. \tag{IV. 20}
$$

The diagrams (3k) and (3l) may be given a similar treatment based upon $W_2^d(r)$ taken in the form (III. 18), with

$$
g_3^k(r)=\frac{\Lambda^3}{4\pi r(1-4/c^2\lambda_D^2)}\int_0^\infty dk\,\sin kr\left(\frac{1}{k^2+\alpha_1^2}-\frac{1}{k^2+\alpha_2^2}\right)G'(k)+\frac{\Lambda^3}{8\lambda_D r}\int_1^\infty dt\,\ln t H_1(t)+\Lambda^3 F(r), \tag{IV. 21}
$$

$$
\begin{aligned}
g_3^k(r)\underset{r\to\infty}{\sim}&\frac{\Lambda^3}{8r}[\exp(-\alpha_1 r)-\exp(-\alpha_2 r)]\left(\lim_{k\to\infty}\frac{G'(k)}{k}\right)+\frac{\Lambda^3}{8r\lambda_D}[\mathbf{A}'\exp(-\alpha_1 r)+\mathbf{B}'\exp(-\alpha_2 r)]\\
&+\frac{\Lambda^3}{8\lambda_D r}[C\exp(-\alpha_1 r)+D\exp(-\alpha_2 r)],
\end{aligned}
\tag{IV. 22}
$$

$$
g_3^k(r)\underset{r\to\infty,\,\hbar\to 0}{\sim}\frac{0.042\Lambda^3\exp(-\alpha_1 r)}{r/\lambda_D}, \tag{IV. 23}
$$

and also

$$
g_3^l(r)\underset{r\to\infty}{\simeq}\frac{\Lambda^3}{8\lambda_D^2}[E_1\exp(-\alpha_1 r)+E_2\exp(-\alpha_2 r)], \tag{IV. 24}
$$

$$
g_3^l(r)\underset{r\to\infty,\,\hbar\to 0}{\sim}-0.021\Lambda^3\exp(-\alpha_1 r). \tag{IV. 25}
$$

A, B, $\lim_{k\to\infty}[G'(k)/k]$, A′, B′, C, D, E_1, and E_2 are detailed in Appendix C.

C. Bridge diagrams (classical limits)

We consider together the diagrams (3m), (3n), and (3o) which have to be computed with other techniques. (3n) and (3o) could be worked out through the standard convolution techniques, once the first one written as

$$
g_3^m(r)=\left(\frac{-\beta e^2}{(1-4/c^2\lambda_D^2)^{1/2}}\right)^5\rho^2\int d\mathbf{r}_3\,d\mathbf{r}_4\frac{\exp(-\alpha_1 r_{13})}{r_{13}}\cdot\frac{\exp(-\alpha_1 r_{32})}{r_{32}}\cdot\frac{\exp(-\alpha_1 r_{34})}{r_{34}}\cdot\frac{\exp(-\alpha_1 r_{14})}{r_{14}}\cdot\frac{\exp(-\alpha_1 r_{34})}{r_{34}} \tag{IV. 26}
$$

in the classical limit is known. 1 and 2 label the root points while 3 and 4 index the nodal points. In view of the complexity of this expression and of the fact shown below that this graph does not provide the most important limit contributions as $r\to 0$ and $r\to\infty$, we restrict ourselves to Eq. (IV. 26). This is the first and more important term ($\alpha_2\gg\alpha_1\sim 1/\lambda_D$) of a sum of 32 analogous contributions. Such a simplification does not imply any loss of physical information, because every term remains finite. A numerical study[7] has already shown that Eq. (IV. 26) remains finite for $\alpha_1=1$ and $r/\lambda_D\geq 0.2$. The very rapid variation in the vicinity of $r=0$ of the bubble graphs makes it useful to investigate more throughly the present $r\to 0$ limit. Therefore, we invoke again the Nijboer—Van Hove procedure already considered for the two-dimensional plasma.[11] Eq. (IV. 26) then becomes

$$
\begin{aligned}
g_3^m(r)&=\frac{\rho^2(-\beta e^2)^5}{(2\pi)^3}\int F(k)\,|G(\mathbf{k})|^2\,d\mathbf{k},\\
F(k)&=\frac{4\pi}{k^2+\alpha_1^2},\\
G(k)&=\int d\mathbf{r}_3\exp(i\mathbf{k}\cdot\mathbf{r}_3')\frac{\exp(-\alpha_1 r_{13}')}{r_{13}'}\cdot\frac{\exp(-\alpha_2 r_{32}')}{r_{32}'},
\end{aligned}
\tag{IV. 27}
$$

with $\mathbf{r}_i'=\mathbf{r}_i-(\mathbf{r}_1+\mathbf{r}_2)/2$. The reference system explicated in Fig. 3 ($r_{12}\equiv r$) gives

$$
\begin{aligned}
g(\mathbf{k})=2\pi&\int_{-1}^{+1}d\cos\theta\int_0^\infty J_0(kW\sin\theta\sin\theta')\\
&\times\exp(ikW\cos\theta\cos\theta')\\
&\cdot\frac{\exp(-\alpha_1|W^2+r_{12}^2/4-r_{12}W\cos\theta|^{1/2})}{|W^2+r_{12}^2/4-r_{12}W\cos\theta|^{1/2}}\\
&\cdot\frac{\exp(-\alpha_1|W^2+r_{12}^2/4+r_{12}W\cos\theta|^{1/2})}{|W^2+r_{12}^2/4+r_{12}W\cos\theta)|^{1/2}},
\end{aligned}
\tag{IV. 28}
$$

with

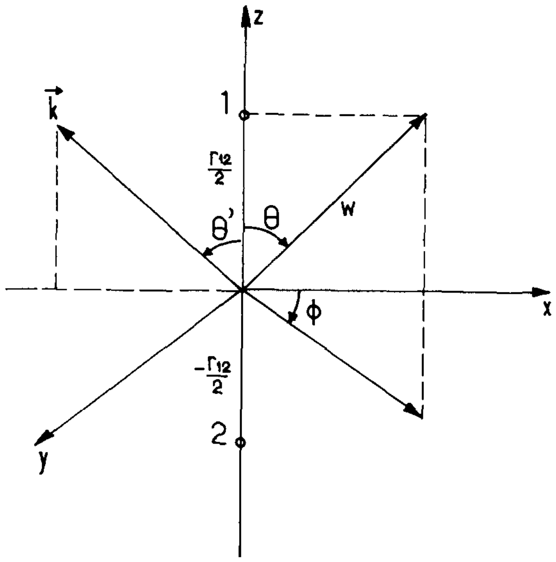


FIG. 3. Reference system for (3m).

$$G(\mathbf{k}) \underset{r_{12}\to\infty}{\sim} 2\pi \int_{-1}^{+1} d\cos\theta \int_0^\infty dW J_0(kW\sin\theta\sin\theta') \times \exp(ikW\cos\theta\cos\theta' - 2\alpha_1 W). \tag{IV.29}$$

The inequality

$$\int_0^\infty dW \exp(ikWA) J_0(BkW) \exp(-2W\alpha_1) \leq (\alpha_1^{-1}B^2k^2/\alpha_1^2 + 4)^{-1/2} \tag{IV.30}$$

yields

$$\lim_{r_{12}\to\infty} |G(\mathbf{k})| \leq \frac{4\pi}{k\sin\theta'} \arcsin\left(\frac{k^2\sin^2\theta'}{k^2\sin^2\theta' + 4\alpha_1^2}\right)^{1/2} \tag{IV.31}$$

where

$$\lim_{r_{12}\to 0} |g_3^m(r_{12})| < \frac{\Lambda^3}{\lambda_D\pi} \int_0^\infty dk \int_{-1}^{+1} \frac{d\cos\theta'}{k^2+\alpha_1^2} \cdot \frac{1}{\sin^2\theta'} \cdot \left[\arcsin\left(\frac{k^2\sin^2\theta'}{k^2\sin^2\theta' + 4\alpha_1^2}\right)^{1/2}\right]^2 \approx \frac{1.783\Lambda^3}{\alpha_1\lambda_D\pi} \tag{IV.32}$$

excluding any diverging short-range behavior. On the other hand, the asymptotic behavior of Eq. (IV.26) is obtained[16] by expressing each e^{-r}/r factor in terms of $(k^2+1)^{-1}$, and considering the resulting Fourier transform

$$W_3^m\left(\frac{k}{\alpha_1}\right) = -\frac{4\pi\Lambda^3}{\alpha_1\lambda_D} \cdot \int_0^1 dx \int_0^1 dy \frac{[1+x(1-x)k^2/\alpha_1^2]^{-1/2}[1+y(1-y)k^2/\alpha_1^2]^{-1/2}}{(x-y)^2k^2/\alpha_1^2 + \{1+[1+x(1-x)k^2/\alpha_1^2]^{1/2} + [1+y(1-y)k^2/\alpha_1^2]^{1/2}\}} \cdot \tag{IV.33}$$

Its $k\to 0$ limit,

$$-\frac{4\pi\Lambda^3}{\alpha_1\lambda_D} \int_0^1 dx \int_0^1 dy \frac{[1-x(1-x)k^2/2\alpha_1^2][1-y(1-y)k^2/2\alpha_1^2]}{(k^2/\alpha_1^2)(x-y)^2 + 9[1+x(1-x)(k^2/3\alpha_1^2) + y(1-y)(k^2/3\alpha_1^2)]} = -9^{-1}\left(1-\frac{8k^2}{27\alpha_1^2}\right)\frac{4\pi\Lambda^3}{\alpha_1\lambda_D} \approx -\frac{3\Lambda^3}{2\lambda_D\alpha_1} \cdot \left(\frac{k^2}{\alpha_1}+\frac{27}{8}\right)^{-1} \tag{IV.34}$$

yields

$$g_3^m(r) \underset{r\to\infty}{\sim} -\frac{3\Lambda^3}{8r/\lambda_D} \cdot \exp(-\sqrt{27/8}\,\alpha_1 r), \qquad g_3^m(r) \underset{r\to\infty,\hbar\to 0}{\sim} -\frac{3\Lambda^3}{8r/\lambda_D} \cdot \exp\left(-\sqrt{27/8} \cdot \frac{r}{\lambda_D}\right). \tag{IV.35}$$

Equations (IV.32), (IV.35) confirm the intermediate behavior of the Bridge graphs. They do not diverge at $r=0$ and decrease at infinity faster than the first-order Debye term $\exp(-r/b)$. A systematic study[8] of these graphs with $n>3$ is given elsewhere.[8] The limit behaviors (IV.32), (IV.35) could be easily transferred to the graphs (3n) and (30) with

$$g_3^n(r) \underset{r\to\infty}{\sim} \frac{-\Lambda^3}{2\pi r[1-4/c^2\lambda_D^2)]^{3/2}} \int_0^\infty \frac{dk\,k\sin kr}{k^2+\alpha_1^2} \cdot \frac{3}{8} \cdot \frac{k}{\frac{27}{8}+k^2} = \frac{3\Lambda^3}{32}\left(1+\frac{1}{c^2\lambda_D^2}\right) \cdot \frac{1}{\frac{27}{8}-\alpha_1^2} \cdot \frac{\exp(-\alpha_1 r)}{r}, \tag{IV.36}$$

$$g_3^0(r) \underset{r\to\infty}{\sim} \frac{-\Lambda^3}{2\pi r(1-4/c^2\lambda_D^2)^{7/2}} \int_0^\infty \frac{dk\,k\sin kr}{(k^2+\alpha_1^2)^2} \cdot \frac{3}{8} \cdot \frac{1}{\frac{27}{8}+k^2} = -\frac{3\Lambda^3}{64}\left(1+\frac{14}{c^2\lambda_D^2}\right)\frac{\exp(-\alpha_1 r)}{\alpha_1(\frac{27}{8}-\alpha_1^2)} \cdot \tag{IV.37}$$

Equations (IV.35)–(IV.77) are explicated numerically in Table IV. It is rewarding that these asymptotic expressions fall very close to the complete and much more involved Cohen–Murphy results,[7] on the whole r-range.

V. SHORT-RANGE BEHAVIOR

Although we have given considerable attention to $\lim_{r\to\infty} g_2(r)$ in the previous analysis, it must be emphasized that the present model is equally well suited to investigate the $g_2(r)$ short-range behavior through a resummation of all orders of the most important diagrams in the $r\to 0$ limit, which obviously are the n-bubbles ($l=n$, $k=0$) followed by the n-bubbles decorated with one ($l=n+1$, $k=1$) and two ($l=n+2$, $k=2$) Debye lines. For instance, we have

$$\frac{(n\text{-bubble with one Debye line})_{r=0}}{(n\text{-bubble})_{r=0}} \leq \frac{\ln\alpha_2}{\alpha_2^2} \underset{\hbar\to 0}{\to} 0, \tag{V.1}$$

exhibiting the relative divergence of the first categories of diagrams in the classical limit. Then, the first contribution to $\lim_{r\to 0} g_2(r)$ is

$$\sum_{n=1}^\infty \frac{(-\Lambda)^n}{n!}\left(\frac{\exp(-\alpha_1 r)-\exp(-\alpha_2 r)}{r}\right)^n = \exp\{-\Lambda[\exp(-\alpha_1 r)-\exp(-\alpha_2 r)]r^{-1}\} - 1 \tag{V.2}$$

giving back the classical expression ($\alpha_2 \gg \alpha_2 \sim 1$)

$$g_2(r) \underset{r\to\infty}{\sim} \exp[-(\Lambda/r)+H(0)], \quad r \text{ is a number of } \lambda_D, \tag{V.3}$$

already obtained by Cooper and DeWitt[17] through an estimate for the short-range part of the chain diagrams

TABLE IV. Nonconvolution (Bridge) diagrams evaluated with Eqs. (IV. 35), (IV. 36), and IV. 37) as a function of reduced distance $[(c\lambda_D)^{-1} = 10^{-3}]$.

r	(3m)	(3n)	(3o)
0.2	−0.2066	0.024 91	−0.005 67
0.4	−0.0715	0.018 82	−0.005 30
0.6	−0.03303	0.014 25	−0.004 83
0.8	−0.0171	0.010 82	−0.004 31
1.0	−0.0095	0.008 23	−0.003 79
1.2	−0.005 48	0.006 28	−0.0033
1.4	−0.003 25	0.004 80	−0.002 84
1.6	−0.001 97	0.003 67	−0.002 43
1.8	−0.001 21	0.002 82	−0.002 07
2.0	−0.000 75	0.002 17	−0.001 75
2.2	−0.000 47	0.001 67	−0.001 48
2.4	−0.000 30	0.001 29	−0.001 24
3.0	−0.000 08	0.000 60	−0.000 73
4.0		0.000 17	−0.000 29

building up the hyper-netted chain approximation. This classical limit is obtained through the limits $\alpha_2 \to \infty$ and $\alpha_1 \to 1$ taken first, followed by $r \to 0$. Had we retained some nonnegligible diffraction effects, we should have done $r \to 0$ first, and thus obtained

$$g_2(r) \sim \exp[H'(0)] = \exp[-\beta e^2/\lambda\!\!\!{}^- + H(0)], \quad r \to 0, \qquad (\text{V. }3')$$

in accord with the Davies–Storer analysis,[18] and a conjecture[19] made sometime ago by DeWitt.[19] It must also be appreciated that the $\hbar \to 0$ limit of Eq. (V. 3′) does not reproduce the classical Eq. (V. 3). $H(0)$ may be given a good approximation with the $r \to 0$ limit of the two series of decorated bubbles previously considered, i.e.,

$$H(0) = S(1) + S(2) + \alpha_1 \lambda_D \Lambda, \qquad (\text{V. }4)$$

where

$$S(n) = \lim_{r\to\infty} \frac{2}{\pi r} \int_0^\infty dk \sin kr \left(\frac{1}{k^2+\alpha_1^2} - \frac{1}{k^2+\alpha_2^2}\right)^n \times \int_0^\infty du\, u \sin ku \cdot \exp\left[-\Lambda[\exp(-\alpha_1 u) - \exp(-\alpha_2 u)]u^{-1} - 1 + \Lambda\left(\frac{\exp(-\alpha_1 u) - \exp(-\alpha_2 u)}{u}\right)\right]. \qquad (\text{V. }5)$$

The corresponding data are available in Table V, altogether with the expressions

TABLE V. Comparison of Eqs. (V.4), (V.6), and (V.7) for $H(0)$ with $1/c\lambda_D = 10^{-6}$.

Λ	Eq. (5)	Eq. (7)	Eq. (8)
0.01	0.0098	0.0096	0.0179
0.05	0.0471	0.0446	0.0470
0.09	0.0825	0.0773	0.0780
0.13	0.1167	0.109	0.107
0.17	0.150	0.143	0.135
0.21	0.182	0.178	0.162
0.25	0.214	0.215	0.188
0.29	0.246	0.256	0.213
0.37	0.307	0.348	0.263
0.49	0.397		0.335
0.65	0.5136		0.427
0.85	0.655		0.533

$$H(0) = \Lambda + \Lambda^2(\ln\Lambda + 0.834), \qquad (\text{V. }6)$$

$$H(0) = 0.619\Lambda^{0.86}, \qquad (\text{V. }7)$$

proposed in Ref. 20 on the basis of a careful examination of the Monte-Carlo data. It turns out that for $\Lambda < 0.37$, Eqs. (V. 6), (V. 7) are in reasonable agreement with the present resummations, although Eq. (V. 6) was expected to be reliable only for $\Lambda < 0.2$, and Eq. (V. 7) for $\Lambda > 0.3$, respectively.[20] As expected, Eq. (V. 7) is the better approximation for larger Λ. However, further resummations are needed in order to reach a closer agreement. It must also be kept in mind that the analytical inversion of the Monte-Carlo data is a very tricky procedure[20] at $r \to 0$, in view of the vanishing Coulomb term in the rhs of Eq. (V. 3). Great care must be exercised in deriving analytical expressions such as Eqs. (V. 6), (V. 7). We think that our approach could provide a firm basis to check such attempts. Actually, we have also to take into account a series of graphs starting from higher order (Fig. 4). The corresponding contribution may be computed through the Fourier transform of $W_2^A(k)^n$,

$$I_n \equiv \frac{4\pi}{k} \int_0^\infty dr\, r \sin kr \left(\frac{\exp(-\alpha_1 r) - \exp(-\alpha_2 r)}{r}\right)^n = \frac{4\pi}{k} \int_0^\infty dr\, r \sin kr \int_{\alpha_1}^{\alpha_2} dX_n \int_{\alpha_1+X_n}^{\alpha_2+X_n} dX_{n-1} \times \int \cdots \int_{\alpha_1+X_3}^{\alpha_2+X_3} dX_2 \int_{\alpha_1+X_2}^{\alpha_2+X_2} dX_1 \cdot \exp(-rX_1). \qquad (\text{V. }8)$$

The first orders $n = 1, 2, 3$ give back the quantities used previously, while

$$I_4 = I_1' + I_2', \qquad (\text{V. }9)$$

with

$$\frac{I_1'}{2\pi} = k^{-1}\Big[[k^2 + (4\alpha_1)^2]\tan^{-1}\frac{k}{4\alpha_1} - 4[k^2 + (3\alpha_1+\alpha_2)^2] \times \tan^{-1}\frac{k}{3\alpha_1+\alpha_2} + 6[k^2 + (2\alpha_1+2\alpha_2)^2] \times \tan^{-1}\frac{k}{2\alpha_1+2\alpha_2} - 4[k^2(\alpha_1+3\alpha_2)^2]\tan^{-1}\frac{k}{\alpha_1+3\alpha_2} + [k^2 + (4\alpha_2)^2]\tan^{-1}\frac{k}{4\alpha_2}\Big] \qquad (\text{V. }10)$$

$$\frac{I_2'}{\pi} = 8\{4\alpha_1 \ln[k^2 + (4\alpha_1)^2] - 4(3\alpha_1+\alpha_2)\ln[k^2 + (3\alpha_1+\alpha_2)^2] + 6(2\alpha_1+2\alpha_2)\ln[k^2 + (2\alpha_1+2\alpha_2)^2] - 4(\alpha_1+3\alpha_2) \times \ln[k^2 + (\alpha_1+3\alpha_2)^2] + 4\alpha_2 \ln[k^2 + (4\alpha_2)^2]\}$$

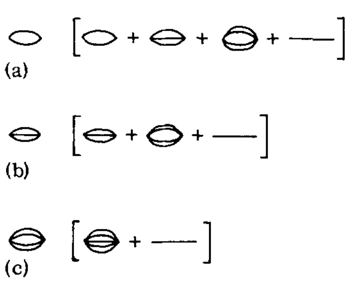


FIG. 4. High order series contributing to $\lim_{r\to 0} g_2(r)$.

$$+k\left[\tan^{-1}\left(\frac{4\alpha_1}{k}\right)-4\tan^{-1}\left(\frac{3\alpha_1+\alpha_2}{k}\right)\right.$$
$$+6\tan^{-1}\left(\frac{2\alpha_1+2\alpha_2}{k}\right)-4\tan^{-1}\left(\frac{\alpha_1+3\alpha_2}{k}\right)$$
$$\left.+\tan^{-1}\left(\frac{4\alpha_2}{k}\right)\right] \tag{V.11}$$

enable us to confirm that the higher order series pictured in Fig. 4 provide a contribution to $H(0)$ smaller than Eq. (V.5). This point may be checked out diagram by diagram. For instance, the four-bubble once decorated is more important in the vicinity of $r=0$ than its homolog (with the same number of lines) given in the second term of the series (a) in Fig. 4.

APPENDIX A

In order to prove the relation (III.10), let us introduce the representation

$$\tan^{-1}\frac{k}{B}=\int_0^\infty dx\exp(-Bx)\frac{\sin kx}{x} \tag{A1}$$

is its l.h.s. So, we get

$$\int_{-\infty}^{\infty}dk\frac{\exp(ikr)}{k^2+A^2}\tan^{-1}\frac{k}{B}$$
$$=\int_0^\infty\frac{dx}{x}\exp(-Bx)\cdot\mathrm{Im}\int_{-\infty}^{\infty}dk\frac{(\cos kr+i\sin kr)}{k^2+A^2}\sin kx$$
$$=\frac{\pi}{2A}\left[\int_r^\infty\frac{dx}{x}\exp(-Bx)[\exp(+rA)\exp(-xA)\right.$$
$$\left.-\exp(-rA)\exp(-xA)\right]+\int_0^r\frac{dx}{x}\exp(-Bx)$$
$$\left.\times[\exp(-rA)\exp(xA)-\exp(-rA)\exp(-xA)]\right]$$
$$=\frac{\pi}{2A}\cdot\left[-\exp(+rA)E_i(-A+B)r)+\exp(-rA)\right.$$
$$\times E_i(-(A+B)r)+\int_0^\infty\frac{dx}{x}\exp(-Bx)$$
$$\times[\exp(-rA)\exp(xA)-\exp(-rA)\exp(-xA)]$$
$$-\int_r^\infty\frac{dx}{x}\exp(-Bx)[\exp(-rA)\exp(xA)$$
$$\left.-\exp(-rA)\exp(-xA)\right]. \tag{A2}$$

The relations

$$\int_0^\infty dx\exp(-Bx)\,\frac{\exp(Ax)-\exp(-Ax)}{x}=\ln\left(\frac{B+A}{|B-A|}\right) \tag{A3}$$

and

$$-\int_r^\infty\frac{dx}{x}\exp(-Bx)[\exp(-rA)\exp(xA)-\exp(-rA)\exp(-xA)]$$
$$=\exp(-rA)E_i(-(B-A)r)-\exp(-rA)E_i(-(A+B)r), \tag{A4}$$

introduced in the last line of Eq. (A2), gives back Eq. (III.10).

APPENDIX B

Equation (IV.7) is evaluated with the aid of the representation

$$\ln(k^2+B^2)-\ln(k^2+C^2)$$
$$=2\int_0^\infty du\frac{\exp(-Bu)-\exp(-Cu)}{u}\cos ku \tag{B1}$$

so that the k-quadrature becomes

$$\mathrm{Im}\int_{-\infty}^{\infty}\frac{dk\,k}{k^2+A^2}(\cos kr+i\sin kr)\cos ku$$
$$=2\int_0^\infty\frac{dk\,k\sin kr\cos ku}{k^2+A^2}$$
$$=\pi\begin{cases}\exp(-rA)\cosh uA, & r>u,\\ -\exp(-uA)\sinh Ar, & r<u,\end{cases} \tag{B2}$$

while the u-quadrature may be written as

$$\int_0^\infty|\cdots|\,du$$
$$=\pi\int_0^r du\exp(-rA)\cosh uA\left(\frac{\exp(-Bu)-\exp(-Cu)}{u}\right)$$
$$-\pi\int_r^\infty du\exp(-uA)\sinh Ar\left(\frac{\exp(-Bu)-\exp(-Cu)}{u}\right)$$
$$=\pi\exp(-rA)\int_0^r du\cosh uA\int_{-C}^{-B}dt\exp(ut)$$
$$-\pi\sinh(Ar)(E_i(-Cr-Ar)-E_i(-Br-Ar))$$
$$=\pi\exp(-rA)\int_{-C}^{-B}dt\left(\frac{\exp[(A+t)r]}{2(A+t)}+\frac{\exp[(t-A)r]}{2(t-A)}\right.$$
$$\left.-\frac{1}{2(A+t)}-\frac{1}{2(t-A)}\right)-\pi\sinh Ar(E_i(-Cr-Ar)$$
$$-E_i(-Br-Ar))$$
$$=\pi\exp(-rA\left[\int_{-C+A}^{-B-A}\frac{dV\exp(Vr)}{2V}+\int_{-C-A}^{-B-A}\frac{dV\exp(Vr)}{2V}\right.$$
$$\left.-\frac{1}{2}\ln\left|\frac{A-B}{A-C}\right|-\frac{1}{2}\ln\left(\frac{B+A}{C+A}\right)\right]$$
$$-\pi\sinh Ar(E_i(-Cr-Ar)-E_i(-Br-Ar))$$
$$=\frac{\exp(-rA)\pi}{2}\left[E_i(-Br+Ar)-E_i(-Cr+Ar)\right.$$
$$\left.-\ln\left|\frac{A-B}{C+A}\right|-\ln\left(\frac{B+A}{C+A}\right)\right]-\frac{\exp(rA)\pi}{2}(E_i(-Cr-Ar)$$
$$-E_i(-Br-Ar))\cdots. \tag{B3}$$

The last line of (B3) becomes the rhs of Eq. (IV.7).

APPENDIX C

Here we detail the quantities monitoring the asymptotic behavior of the graphs (3i), (3j), (3k), and (3l):

$$\mathbf{A}=-\frac{\Lambda^3}{4\lambda_D^2}\int_1^\infty dt\,\ln t\left(\frac{3}{(\alpha_1+3t/\lambda_D)^2-\alpha_1^2}-\frac{3}{(\alpha_2+3t/\lambda_D)^2-\alpha_1^2}-\frac{1+2c\lambda_D}{(\alpha_1+t/\lambda_D)^2-\alpha_1^2}+\frac{1+2/c\lambda_D}{(\alpha_2+t/\lambda_D)^2-\alpha_1^2}\right), \tag{C1}$$

$$\mathbf{B}=-\frac{\Lambda^3}{4\lambda_D^2}\int_1^\infty dt\,\ln t\left(-\frac{3}{(\alpha_1+3t/\lambda_D)^2-\alpha_2^2}+\frac{3}{(\alpha_2+3t/\lambda_D)-\alpha_2^2}+\frac{1+2c\lambda_D}{(\alpha_1+t/\lambda_D)^2-\alpha_2^2}-\frac{1+2/c\lambda_D}{(\alpha_2+t/\lambda_D)^2-\alpha_2^2}\right), \tag{C2}$$

$$\lim_{k\to\infty}\frac{G'(k)}{k}=\left(\frac{3}{c\lambda_D}+\ln 3-\frac{4}{3}\right)\left(\frac{1}{\alpha_1+\lambda_D^{-1}}-\frac{1}{\alpha_2+\lambda_D^{-1}}\right)+\frac{4}{3}\left(\frac{1}{\alpha_1+2/\lambda_D}-\frac{1}{\alpha_2+2/\lambda_D}\right)+\frac{2(\alpha_1/c+1+2/c\lambda_D)}{\alpha_1+c-\lambda_D^{-1}}$$
$$-\frac{2(\alpha_2/c+1+2/c\lambda_D)}{\alpha_1+c-\lambda_D^{-1}}-\frac{4}{c\lambda_D}\left(\frac{1}{\alpha_1+c+\lambda_D^{-1}}-\frac{1}{\alpha_2+c+\lambda_D^{-1}}\right)+\frac{2}{c}\ln\frac{\alpha_1+c-\lambda_D^{-1}}{\alpha_2+c-\lambda_D^{-1}}, \tag{C3}$$

$$\mathbf{A}'=\frac{\ln 3}{(\alpha_1+\lambda_D^{-1})^2-\alpha_1^2}-\frac{\ln 3}{(\alpha_2+\lambda_D^{-1})^2-\alpha_1^2}-\frac{4}{c\lambda_D}\left(\frac{1}{(\alpha_1+c-\lambda_D^{-1})^2-\alpha_1^2}-\frac{1}{(\alpha_2+c-\lambda_D^{-1})^2-\alpha_1^2}\right), \tag{C4}$$

$$\mathbf{B}'=\frac{-\ln 3}{(\alpha_1+\lambda_D^{-1})^2-\alpha_2^2}+\frac{\ln 3}{(\alpha_2+\lambda_D^{-1})^2-\alpha_2^2}+\frac{4}{c\lambda_D}\left(\frac{1}{(\alpha_1+c-\lambda_D^{-1})^2-\alpha_1^2}-\frac{1}{(\alpha_2+c-\lambda_D^{-1})^2-\alpha_1^2}\right), \tag{C5}$$

$$C=\int_1^\infty dt\,\ln t\left(\frac{-1}{(\alpha_1+\lambda_D^{-1}+t/\lambda_D)^2-\alpha_1^2}+\frac{1}{(\alpha_2+1/\lambda_D+t/\lambda_D)}+\frac{3}{(\alpha_1+3t/\lambda_D-1/\lambda_D)^2-\alpha_1^2}-\frac{3}{(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_1^2}\right.$$
$$\left.-\frac{(2/\lambda_D)(\alpha_1+1/\lambda_D+t/\lambda_D)}{[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2]^2}+\frac{(2/\lambda_D)(\alpha_2+1/\lambda_D+t/\lambda_D)}{[\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2]^2}-\frac{(6/\lambda_D)(\alpha_1+3t/\lambda_D-1/\lambda_D)}{[(\alpha_2+3t/D-\lambda_D^{-1})^2-\alpha_1^2]^2}+\frac{(6/\lambda_D)(\alpha_2+3t/\lambda_D-1/\lambda_D)}{[(\alpha_2+3t/\lambda_D-1/\lambda_D)^2-\alpha_1^2]^2}\right) \tag{C6}$$

$$D=\int_1^\infty dt\,\ln t\left(\frac{1}{(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2}-\frac{1}{(\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2}-\frac{3}{(\alpha_1+3t/\lambda_D-\lambda_D^{-1})^2-\alpha_2^2}\right.$$
$$+\frac{3}{(\alpha_2+3t/\lambda_D-1/\lambda_D)^2-\alpha_2^2}+\frac{(2/\lambda_D)(\alpha_1+1/\lambda_D+t/\lambda_D)}{[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2]^2}-\frac{(2/\lambda_D)(\alpha_2+1/\lambda_D+t/\lambda_D)}{[(\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2]^2}$$
$$\left.+\frac{(6/\lambda_D)(\alpha_1+3t/\lambda_D-1/\lambda_D)}{[(\alpha_1+3t/\lambda_D-1/\lambda_D)^2-\alpha_2^2]^2}\;\frac{(6/\lambda_D)(\alpha_2+3t/\lambda_D-1/\lambda_D)}{[(\alpha_2+3t/\lambda_D-1/\lambda_D)^2-\lambda_2^2]^2}\right), \tag{C7}$$

$$\mathbf{A}''=\frac{\ln 3}{2\alpha_1[(\alpha_1+\lambda_D^{-1})^2-\alpha_1^2]}-\frac{\ln 3}{2\alpha_1[(\alpha_2+\lambda_D^{-1})^2-\alpha_1^2]}-\frac{2}{\alpha_1 c\lambda_D[(\alpha_1-1/\lambda_D+c)^2-\alpha_1^2]}+\frac{2}{\alpha_1 c\lambda_D[(\alpha_2-1/\lambda_D+c)^2-\alpha_1^2]}, \tag{C8}$$

$$\mathbf{B}''=\frac{\ln 3}{2\alpha_2[(\alpha_1+\lambda_D^{-1})^2-\alpha_2^2]}-\frac{\ln 3}{2\alpha_2[(\alpha_2-\lambda_D^{-1})^2-\alpha_2^2]}-\frac{2}{\alpha_2 c\lambda_D[(\alpha_1-\lambda_D^{-1}+c)^2-\alpha_2^2]}+\frac{2}{\alpha_2 c\lambda_D[(\alpha_2-\lambda_D^{-1}+c)^2-\alpha_1^2]}, \tag{C9}$$

$$C'=\int_1^\infty dt\,\ln t\left(\frac{-1}{2\alpha_1[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2)}+\frac{1}{2\alpha_1[(\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2]}+\frac{3}{2\alpha_1[(\alpha_1-1/\lambda_D+3t/\lambda_D)^2-\alpha_1^2)}\right.$$
$$-\frac{3}{2\alpha_1[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_1^2]}-\frac{\alpha_1+1/\lambda_D+t/\lambda_D}{\alpha_1\lambda_D[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2]^2}+\frac{\alpha_2+1/\lambda_D+t/\lambda_D}{\alpha_1\lambda_D[(\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_1^2]^2}$$
$$\left.-\frac{3(\alpha_1-\lambda_D^{-1}+3t/\lambda_D)}{\lambda_D\alpha_1[(\alpha_1-1/\lambda_D+3t/\lambda_D)^2-\alpha_1^2]^2}+\frac{3(\alpha_2-1/\lambda_D+3t/\lambda_D)}{\lambda_D\alpha_1[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_1^2]^2}\right), \tag{C10}$$

$$D'=\int_1^\infty dt\,\ln t\left(-\frac{1}{2\alpha_2[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2]}+\frac{1}{2\alpha_2[(\alpha_2+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2]}+\frac{3}{2\alpha_2[(\alpha_1-1/\lambda_D+3t/\lambda_D)^2-\alpha_2^2]}\right.$$
$$-\frac{3}{2\alpha_2[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_2^2]}-\frac{\alpha_1+1/\lambda_D+t/\lambda_D}{\alpha_2\lambda_D[(\alpha_1+1/\lambda_D+t/\lambda_D)^2-\alpha_2^2]^2}+\frac{\alpha_2+1/\lambda_D+t/\lambda_D}{\alpha_2\lambda_D[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_2^2]^2}$$
$$\left.-\frac{3(\alpha_1-\lambda_D^{-1}+3t\lambda_D^{-1})}{\alpha_2\lambda_D[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_2^2]^2}+\frac{3(\alpha_2-\lambda_D^{-1}+3t/\lambda_D)}{\alpha_2\lambda_D[(\alpha_2-1/\lambda_D+3t/\lambda_D)^2-\alpha_2^2]^2}\right), \tag{C11}$$

$$E_1=\frac{1}{2\alpha_1}\lim_{k\to 0}\frac{G'(k)}{k}+\frac{\mathbf{A}''}{\lambda_D}+\frac{C'}{\lambda_D}, \tag{C12}$$

$$E_2=\frac{1}{2\alpha_2}\lim_{k\to 0}\frac{G'(k)}{k}+\frac{\mathbf{B}''}{\lambda_D}+\frac{C'}{\lambda_D}. \tag{C13}$$

ACKNOWLEDGMENTS

We have to thank A.A. Broyles, H.E. De Witt, Y. Furutani, B. Jancovici, and M. Pokrant for stimulating discussions and useful advice. One of us (C.D.) acknowledges the hospitality of the Research Laboratory of Electronics, Massachusetts Institute of Technology, where part of this work was completed.

[1]H.E. DeWitt, J. Math. Phys. 7, 616 (1966).
[2]T. Dunn and A.A. Broyles, Phys. Rev. **157**, 156 (1967).
[3]J. Coste, Nucl. Fusion 5, 284 (1965).
[4]H.E. DeWitt, Phys. Rev. **140**, A 466 (1965).
[5]G. Kelbg, Ann. Phys. (Leipz.) 12, 219 (1963); **12**, 354 (1964); A.E. Glauberman and I.R. Yuknovski, Dokl. Akad. Nauk SSSR 93, 999 (1953).
[6]S.G. Brush, H.E. DeWitt, and J.G. Trulio, Nucl. Fusion **3**, 5 (1963).
[7]E.G.D. Cohen and T.J. Murphy, Phys. Fluids **12**, 1404 (1969).
[8]C. Deutsch, Y. Furutani, and M.M. Gombert, Phys. Rev. A 13, 2244 (1976).
[9]M.M. Gombert and C. Deutsch, to be published.
[10]A. Isihara and M. Wadati, Physica 57, 237 (1972).
[11]C. Deutsch and M. Lavaud, Phys. Rev. A 9, 2598 (1974).
[12]A.A. Barker, J. Chem. Phys. 55, 1751 (1971).
[13]M.M. Gombert, Thèse 3^e cycle, Université Paris XI, Orsay, June 1974 (unpublished).
[14]A preliminary account of these results has been given in M.M. Gombert and C. Deutsch, Phys. Lett. A 47, 473 (1974).
[15]F. Del Rio and H.E. DeWitt, Phys. Fluids **12**, 791 (1969).
[16]D.J. Mitchell and B.W. Ninham, Phys. Rev. **174**, 280 (1968).
[17]M.S. Cooper and H.E. DeWitt, Phys. Lett. A **40**, 391 (1972).
[18]B. Davis and R.G. Storer, Phys. Rev. **171**, 150 (1968).
[19]H.E. DeWitt, private communication (1968).

New representation of the Tomimatsu–Sato solution

Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 14 August 1975)

We devise a new representation of the simplest Tomimatsu–Sato solution of Einstein's vacuum field equations. This permits us to dispose of the previously troublesome "directional singularities" through the introduction of an advanced (or retarded) time coordinate. In the neighborhood of the locations in question the T–S space is shown to possess a Killing tensor of valence two, which allows us to solve the geodesic problem in this neighborhood completely. Finally, we present for future analysis a plausible toroidal model of the material source for the T–S solution.

I. INTRODUCTION

Several years ago Tomimatsu and Sato (T–S) constructed[1] a spinning mass solution of the Einstein vacuum field equations. The nonvanishing components of the metric tensor were given by

$$g_{zz}=g_{\rho\rho}=B/[p^4(x^2-y^2)^4],\quad g_{\Phi\Phi}=(1-y^2)D/(p^2B),$$
$$g_{\Phi T}=4(q/p)(1-y^2)C/B,\quad g_{TT}=-A/B, \tag{1.1}$$

where $z=xy$, and $\rho^2=(x^2-1)\,(1-y^2)$, and A, B, C, and D denote the following polynomials:

$$A=p^4(x^2-1)^4+q^4(1-y^2)^4-2p^2q^2(x^2-1)(1-y^2)$$
$$\times[2(x^2-1)^2+2(1-y^2)^2+3(x^2-1)(1-y^2)],$$
$$B=[p^2(x^2+1)(x^2-1)-q^2(y^2+1)(1-y^2)+2px(x^2-1)]^2$$
$$+4q^2y^2[px(x^2-1)+(px+1)(1-y^2)]^2, \tag{1.2}$$
$$C=-p^3x(x^2-1)[2(x^2+1)(x^2-1)+(x^2+3)(1-y^2)]-p^2(x^2-1)$$
$$\times[4x^2(x^2-1)+(3x^2+1)(1-y^2)]+q^2(px+1)(1-y^2)^3,$$
$$D=p^6(x^2-1)(x^8+28x^6+70x^4+28x^2+1)-16q^6(1-y^2)^3$$
$$+p^4q^2\{(x^2-1)[32x^2(x^4+4x^2+1)-4(1-y^2)(x^2-1)^3$$
$$+(-6x^4+12x^2+10)(1-y^2)^3]-4(1-y^2)^3(x^4+6x^2+1)\}$$
$$+p^2q^4\{(x^2-1)[64x^4+(1-y^2)^2(y^4+14y^2+1)]$$
$$-16(1-y^2)^3(x^2+2)\}+8p^5x(x^4-1)(x^4+6x^2+1)$$
$$-32pq^4x(1-y^2)^3+8p^3q^2x\{(x^2-1)[8x^2(x^2+1)$$
$$+(1-y^2)^2(2y^2-x^2+1)]-4(1-y^2)^3\}.$$

Several even more complicated solutions corresponding to spinning masses were constructed by the same authors, but none of these solutions have been thoroughly studied.[2] We have recently succeeded, however, in casting the simplest T–S solution into an alternative representation which is much more amenable to serious investigations. Preliminary indications are that the structure of the space is extremely interesting, and a complete study is currently being made.[3] Furthermore, we anticipate that similar procedures may be applied to the more complicated T–S solutions.

II. NEW REPRESENTATION

It is well known that Boyer–Lindquist coordinates are inappropriate for studies of the structure of Kerr space, and that the transition to the more advantageous Kerr coordinates is facilitated by writing the Kerr line element in the form

$$ds^2=\Sigma[(dr^2/\Delta)+d\theta^2]+\Sigma^{-1}\{\sin^2\theta[(r^2+a^2)\,d\Phi-a\,dT]^2$$
$$-\Delta(dT-a\sin^2\theta\,d\Phi)^2\}. \tag{2.1}$$

We, therefore, sought and found an analogous representation of the line element of T–S space.

The essential step in our treatment consists of observing that the polynomials A, B, C, and D can be expressed in the following manner:

$$A=(p^2V^2+q^2W^2)^2-4p^2q^2VW(V+W)^2,$$
$$B=(p^2V^2+q^2W^2)\{p^2V^2+q^2W^2+4[V+(2+V)(1+px)]\}$$
$$-4q^2W(V+W)[p^2V(V+W)+2(1+px)W],$$
$$C=\tfrac{1}{2}(p^2V^2+q^2W^2)[p^2V(V+W)+2(1+px)W] \tag{2.2}$$
$$-\tfrac{1}{2}p^2V(V+W)\{p^2V^2+q^2W^2+4[V+(2+V)(1+px)]\},$$
$$D=p^2V\{p^2V^2+q^2W^2+4[V+(2+V)(1+px)]\}^2$$
$$-4q^2W[p^2V(V+W)+2(1+px)W]^2,$$

where $V=x^2-1$ and $W=1-y^2$. From Eq. (1.1) it follows immediately that

$$ds^2=[B/p^4(V+W)^4](dz^2+d\rho^2)+(VW/B)$$
$$\times(\{p^2V^2+q^2W^2+4[V+(2+V)(1+px)]\}\,d\Phi$$
$$-2pq(V+W)\,dT)^2-(1/B)\{(p^2V^2+q^2W^2)\,dT$$
$$-2(q/p)W[p^2V(V+W)+2(1+px)W]\,d\Phi\}^2. \tag{2.3}$$

For our present purposes, however, we find it convenient to express this result in the form

$$ds^2=\Lambda^2Q(dz^2+d\rho^2)+[VW/(V+W)^2]\Lambda^{-2}$$
$$\times Q^{-1}[F\,d\Phi-2(q/p)\,dT]^2-\Lambda^2Q^{-1}[dT-2(q/p)G\,d\Phi]^2, \tag{2.4}$$

where

$$F=\frac{p^2V^2+q^2W^2+4[V+(2+V)(1+px)]}{p^2(V+W)}, \tag{2.5}$$

$$G=W\frac{p^2V(V+W)+2(1+px)W}{p^2V^2+q^2W^2}, \tag{2.6}$$

$$Q=[F-(2q/p)^2G]/(V+W), \tag{2.7}$$

$$\Lambda^2=\frac{p^2V^2+q^2W^2}{p^2(V+W)^2}. \tag{2.8}$$

It should be noted that as $x\to\infty$, $F\to V$, $G\to W$, $Q\to 1$ and $\Lambda^2\to 1$, while $Q\to 0$ at the equatorial ring singularities.

We now introduce new coordinates ϕ and u such that

$$d\Phi = d\phi + \frac{q}{p}\alpha(F^{-1})\frac{dF}{F},\quad dT = du + \tfrac{1}{2}\alpha(F^{-1})\,dF, \tag{2.9}$$

where $\alpha(F^{-1})$ is a function of F^{-1} to be specified later. This transformation results in

$$\begin{aligned} &F\,d\Phi - 2\frac{q}{p}\,dT = F\,d\phi - 2\frac{q}{p}\,du,\\ &dT - 2\frac{q}{p}G\,d\Phi = du - 2\frac{q}{p}G\,d\phi + \tfrac{1}{2}\alpha Q(V+W)\frac{dF}{F}. \end{aligned} \tag{2.10}$$

Therefore, Eq. (2.4) assumes the form

$$\begin{aligned} ds^2 = {}&\Lambda^2 Q\left(dz^2 + d\rho^2 - \tfrac{1}{4}\alpha^2(V+W)^2\frac{dF^2}{F^2}\right) + \left(\frac{VW}{(V+W)^2}\right)\\ &\times\Lambda^{-2}Q^{-1}\left(F\,d\phi - 2\frac{q}{p}\,du\right)^2 - \Lambda^2 Q^{-1}\left(du - 2\frac{q}{p}G\,d\phi\right)^2\\ &-\Lambda^2\alpha(V+W)\frac{dF}{F}\left(du - 2\frac{q}{p}G\,d\phi\right). \end{aligned} \tag{2.11}$$

In a recent paper by Economou and Ernst[4] it was suggested that the so-called directional singularities which plague the T—S solutions at the poles, $x=1$, $y=\pm 1$, may simply be a coordinate effect, and these "points" may in fact be surfaces. In the next section we shall employ the line element (2.11) to show that this is indeed the case.

III. GEOMETRY NEAR THE NORTH POLE

In the neighborhood of the point $x=1$, $y=+1$, to which we refer as the "north pole," spherical polar coordinates provide a better chart than do the symmetrical x—y coordinates. The specification

$$z = 1 + [2(1+p)/p^2]r\cos\theta,\quad \rho = [2(1+p)/p^2]r\sin\theta, \tag{3.1}$$

when expressed in terms of x and y, assumes the form

$$r = [p^2/2(1+p)](x-y),\quad \cos\theta = (xy-1)/(x-y), \tag{3.2}$$

where initially we contemplate the range of θ to be $0\le\theta\le\pi$, but it should be noted that $\theta=\pi$ corresponds to the surface $x=1$, which cannot be regarded as part of the axis.

If one is just interested in the geometry in the neighborhood of the north pole, it is convenient to replace the coordinates defined in Eqs. (3.1) and (3.2) by coordinates

$$r = [p^2/4(1+p)](V+W),\quad \sin^2\theta = 4VW/(V+W)^2, \tag{3.3}$$

which closely approximate the spherical polar coordinates. All our former equations can be expressed in terms of this r—θ coordinate system by substituting the expressions

$$V = [4(1+p)/p^2]r\cos^2(\theta/2),\quad W = [4(1+p)/p^2]r\sin^2(\theta/2). \tag{3.4}$$

In particular, Λ^2 is observed to be a function of θ alone; namely,

$$\Lambda^2 = [p^2\cos^4(\theta/2) + q^2\sin^4(\theta/2)]/p^2, \tag{3.5}$$

which for $q\ne 0$ has finite positive upper and lower bounds. On the other hand, for small values of r,

$$\begin{aligned} &F\approx 2/r,\quad G\approx[2(1+p)/p^2]\Lambda^{-2}\sin^4(\theta/2),\\ &Q\approx[p^2/2(1+p)r^2]. \end{aligned} \tag{3.6}$$

It is immediately apparent from Eq. (2.11) that if a singularity is to be avoided at $r=0$ it will be necessary for

$$dz^2 + d\rho^2 - \tfrac{1}{4}\alpha^2(V+W)^2\frac{dF^2}{F^2}$$

to contain a factor r^2. Using the relations

$$dz^2 + d\rho^2 = \tfrac{1}{4}(V+W)\left(\frac{dV^2}{(1+V)V} + \frac{dW^2}{(1-W)W}\right) \tag{3.7}$$

and

$$\frac{dF}{F}\approx -\frac{dV+dW}{V+W} + dV, \tag{3.8}$$

we find that the necessary factor r^2 can be obtained by attributing to the function $\alpha(F^{-1})$ the power series expansion

$$\alpha(F^{-1}) = 1 + [4(1+p)/p^2]F^{-1} + \cdots. \tag{3.9}$$

If α is chosen in the manner which we have indicated, then the line element induced upon the surface $r=0$ is simply

$$[p^2/2(1+p)]\,ds^2 = \Lambda^2\,d\theta^2 + \Lambda^{-2}\sin^2\theta\,d\phi^2. \tag{3.10}$$

Wild has evaluated[5] the Gaussian curvature of this 2-surface, which is everywhere negative and finite, exhibiting no singularities whatsoever.

IV. APPROXIMATE METRIC

If one replaces the components of the metric tensor by power series in the coordinate r, the question naturally arises of how one should properly truncate the series in order to obtain a good approximation to the actual geometry near $r=0$. We propose replacing r by λr and u by $\lambda^{-1}u$, after which the limit $\lambda\to 0$ will be taken. The resulting approximate line element,

$$\begin{aligned} [p^2/2(1+p)]\,ds^2 = {}&\Lambda^2(2\,dr\,du - r^2\,du^2 + d\theta^2)\\ &+\Lambda^{-2}\sin^2\theta\,[d\phi - (q/p)r\,du]^2, \end{aligned} \tag{4.1}$$

has been the subject of an extensive study[6] by Economou, who not only identified the space in question as a known type-D vacuum solution of Einstein's field equations, but also verified that the Weyl tensor approximates the Weyl tensor invariant evaluated earlier[4] by Economou and Ernst for the full T—S solution. For this reason we are confident that Eq. (4.1) does adequately represent the geometry near the north pole of T—S space.

The further study of the approximate line element of Eq. (4.1) is facilitated by introducing a null tetrad (t, m, t, t^*) such that

$$\begin{aligned} &k = du,\quad m = \Lambda^2(dr - \tfrac{1}{2}r^2\,du),\\ &t = (1/\sqrt{2})\{\Lambda\,d\theta + i\Lambda^{-1}\sin\theta[d\phi - (q/p)r\,du]\}. \end{aligned} \tag{4.2}$$

The corresponding tangent vectors are given by the expressions

$$\mathbf{k}=\Lambda^{-2}\mathbf{a}_r, \quad \mathbf{m}=\mathbf{a}_u+(q/p)r\mathbf{a}_\phi+\tfrac{1}{2}r^2\mathbf{a}_r,$$
$$\mathbf{t}=(1/\sqrt{2})\{\Lambda^{-1}\mathbf{a}_\theta+i\Lambda\csc\theta\,\mathbf{a}_\phi\}, \tag{4.3}$$

so the Hamiltonian assumes the form

$$H=\Lambda^{-2}p_r[p_u+(q/p)rp_\phi+\tfrac{1}{2}r^2p_r]$$
$$+\tfrac{1}{2}\Lambda^{-2}p_\theta^2+\tfrac{1}{2}\Lambda^2\csc^2\theta\,p_\phi^2. \tag{4.4}$$

p_ϕ, p_u, and H are obviously constants of the motion, which we shall denote, respectively, by l, $-\epsilon$, and $-\frac{1}{2}\mu^2$. Recalling that Λ^2 depends only upon θ, it is obvious also that this space possesses a Killing tensor of the type which has been studied so extensively[7] by Hauser and Malhiot. The associated constant of the motion will be denoted by K. As in the case of the Kerr metric[8] the existence of an extra constant of the motion will enable us to solve completely the geodesic problem in the neighborhood of the north pole of T—S space.

Multiplying Eq. (4.4) by Λ^2 and regarding p_r as a function of r alone and p_θ as a function of θ alone, we obtain the pair of equations

$$p_\theta^2+\Lambda^4\csc^2\theta\, l^2+\Lambda^2\mu^2=K, \tag{4.5}$$

$$p_r[-\epsilon+(q/p)lr+\tfrac{1}{2}r^2p_r]=-\tfrac{1}{2}K, \tag{4.6}$$

which determines the two functions $p_r(r)$ and $p_\theta(\theta)$. On the other hand, Hamilton's equations of motion yield

$$p_r=\Lambda^2\dot{u}, \quad p_\theta=\Lambda^2\dot{\theta}, \tag{4.7}$$

where dots denote differentiation with respect to proper time in the case of timelike geodesics, or an appropriate parameter in the case of other types of geodesics.

Because the positive function Λ^2 is bounded and never vanishes (for $q\neq 0$), there is a monotonic relation between the proper time (affine parameter, or proper distance) and an auxilliary parameter τ defined by

$$\Lambda^2\dot{\tau}=1. \tag{4.8}$$

It is convenient to discuss the geodesics in terms of the auxilliary parameter τ, for it turns out that the equations of motion can be solved *explicitly* in terms of τ.

Hamilton's equations also give the constants of motion,

$$l=\Lambda^{-4}\sin^2\theta\left(\frac{d\phi}{d\tau}-\frac{q}{p}r\frac{du}{d\tau}\right), \tag{4.9}$$

$$\epsilon=r^2\frac{du}{d\tau}-\frac{dr}{d\tau}+\frac{q}{p}lr, \tag{4.10}$$

while Eqs. (4.5) and (4.6) can be written as follows:

$$\left(\frac{d\theta}{d\tau}\right)^2+\Lambda^4\csc^2\theta\, l^2+\Lambda^2\mu^2=K, \tag{4.11}$$

$$\frac{du}{d\tau}\left[-\epsilon+\frac{q}{p}lr+\frac{r^2}{2}\frac{du}{d\tau}\right]=-\tfrac{1}{2}K. \tag{4.12}$$

Using Eq. (4.10) to eliminate $du/d\tau$ from Eq. (4.12), we obtain the key equation of motion

$$\left(\epsilon-\frac{q}{p}lr\right)^2-\left(\frac{dr}{d\tau}\right)^2=Kr^2, \tag{4.13}$$

which is readily solved for r in terms of the auxilliary parameter τ.

After $r(\tau)$ is determined, the function $u(\tau)$ can be determined using Eq. (4.10). Equation (4.11) yields the function $\theta(\tau)$, and Eq. (4.9) yields the function $\phi(\tau)$. Finally, Eq. (4.8) can be employed in order to determine the relation between the auxilliary parameter τ and the proper time (affine parameter, or proper distance).

V. TIMELIKE GEODESICS WITH $l=0$

Equation (4.11) implies that when $l\neq 0$ there is a repulsive barrier which keeps θ away from 0 or π. In the case $l=0$ there is no such barrier, for Λ^2 is a bounded function of θ. Consequently, if K is larger than μ^2, θ will undergo excursions to the axis, $\theta=0$, while if K is larger than μ^2q^2/p^2, θ will undergo excursions to $\theta=\pi$, to which we shall refer as the "cut." If one recalls that $\theta=\pi$ corresponds to the surface $x=1$ in the original T—S space, the need to extend θ beyond the value π will not seem so surprising. Furthermore, it has been shown by Wild[5] that if θ runs from 0 to 2π, then the Euler characteristic of the 2-surface $r=0$ is zero, which corresponds to a toroidal or Klein bottle topology for the surface. On the other hand, if θ runs from 0 to π, then in general there is an extrinsic curvature cusp, which is only avoided for $q=p=1/\sqrt{2}$. In the latter special case, the topology is that of a 2-sphere (Euler characteristic 2).[9]

All of the timelike geodesics with $l=0$ pass twice through the surface $r=0$, for since $K>0$, the solution of Eq. (4.13) is simply

$$r=(\epsilon/\sqrt{K})\cos(\sqrt{K}\,\tau). \tag{5.1}$$

The proper time between successive passages through $r=0$ is obviously finite. The greater the energy ϵ is, the further the particle can get from $r=0$. We anticipate that in the exact T—S solution particles with energy exceeding a certain amount could escape from the north polar region.

When K is sufficiently large, the term $\Lambda^2\mu^2$ can be neglected in Eq. (4.11), so one readily obtains an approximate orbital equation,

$$r\approx(\epsilon/\sqrt{K})\cos(\theta-\theta_0). \tag{5.2}$$

All the exact expressions for the geodesics can be worked out as necessary, for the problem has been reduced to quadratures.

Clearly, it is necessary to extend our space across the $r=0$ surface. We intend to discuss this question at length in a future paper. At the present time let it suffice to mention that one of the more interesting possibilities involves the identification of the north and south polar $r=0$ surfaces.

Since a distinct $r=0$ surface arises when one uses a retarded time coordinate instead of an advanced time coordinate, one can identify the *future* north polar $r=0$ surface with the *past* south polar $r=0$ surface, and the *past* north polar $r=0$ surface with the *future* south polar $r=0$ surface. The resulting "wormhole" structure is not terribly far-fetched if one contemplates that the natural source for the T—S solution is a rotating torus of matter. In the Weyl $z-\rho$ coordinates the surface of the material torus would resemble a spheroid, large enough at the equator to contain the outer ring singularity, and extending to the north pole ($z=+1$, $\rho=0$), and to the south pole ($z=-1$, $\rho=0$), where the wormhole mouths are located.

VI. NULL GEODESICS WITH $l = 0$

The positive-K null geodesics resemble the large-K timelike geodesics. The $K=0$ null geodesics constitute the shear-free, twist-free, and expansion-free principal null geodesic congruence upon which our null tetrad was based. For these null rays θ is constant, while r varies linearly with τ. It would seem reasonable to expect the corresponding null congruence in the exact T—S space to consist of light rays coming in from all directions to impinge upon the future north polar $r=0$ surface. By symmetry one would expect another congruence of incoming light rays which impinge upon the future south polar $r=0$ surface. In the specific model which we have been considering these null geodesics could be extended through the wormhole and generate a pair of outgoing null congruences. This picture would be consistent with the known algebraically general Petrov type of the T—S space, for there must be four different principal null directions at each space—time point.

VII. SPACELIKE GEODESICS WITH $l = 0$

In the case of spacelike geodesics K must be no less than the smaller of the two quantities μ^2 and $\mu^2 q^2/p^2$. Spacelike geodesics with $K \geq 0$ reach $r=0$ with a finite proper distance, and these geodesics can be extended past that surface. On the other hand, spacelike geodesics with $K<0$ correspond to

$$r = r_0 \exp(-\sqrt{-K}\,\tau), \tag{7.1}$$

and $\epsilon = 0$. These geodesics approach $r=0$ infinitely slowly.

If one restricts attention to the symmetry axis, where $\Lambda^2 = 1$ and $K=\mu^2$, our findings based upon the approximate metric (4.1) are completely consistent with an earlier study[10] of axial geodesics in T—S space by Gibbons and Russell-Clark.

VIII. CONCLUSIONS

We believe that the discovery of the simplified representation (2.4) of the $\delta=2$ T—S solution of Einstein's vacuum field equations has transformed this solution from one spurned for its complexity into one which can now be studied seriously. It would not be surprising to us if a coordinate system were discovered soon in which a principal null congruence played a dominant role. Our speculations will then be easier to prove or disprove.

ACKNOWLEDGMENTS

The author has profited from thrice weekly discussions with J. E. Economou and W. J. Wild at the IIT Summer Relativity Seminar.

[1]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972); Progr. Theor. Phys. **50**, 95 (1973).
[2]There was, of course, the hotly debated question concerning whether or not $x=1$ is an event horizon, which was finally resolved in the negative. (See Ref. 10.)
[3]We expect to submit an illustrated article to another journal.
[4]J. E. Economou and F. J. Ernst, J. Math. Phys. **17**, 52 (1976).
[5]Private communication.
[6]J. E. Economou, J. Math. Phys. **17**, 1095 (1976), following paper.
[7]I. Hauser and R. J. Malhiot, J. Math. Phys. **16**, 150 (1975); **16**, 1625 (1975).
[8]B. Carter, Phys. Rev. **174**, 1559 (1968).
[9]R. Geroch has suggested, however, that considered as a 2-surface, $r=0$ does not constitute a manifold if θ runs from 0 to 2π. Private communication.
[10]G. W. Gibbons and R. A. Russell-Clark, Phys. Rev. Lett. **30**, 398 (1973).

Approximate form of the Tomimatsu–Sato $\delta=2$ solution near the poles $x=1$, $y=\pm 1$

John E. Economou

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 14 August 1975)

A consideration of the line element for the Tomimatsu–Sato $\delta=2$ solution near the poles $x=1$, $y=\pm 1$ reveals the existence of a metric which is nonsingular there. The further study of this metric indicates that it corresponds to a vacuum, type D gravitational field, and, as such, it is among those type D vacuum solutions specified by W. Kinnersley. Reasons are given in support of the belief that the derived metric is a valid approximation to the exact T–S solution close to the poles.

I. INTRODUCTION

The recent determination[1] of the curvature invariants for the Tomimatsu–Sato $\delta=2$ spinning mass field has done much to clarify the singularity structure of this space–time. The study of those invariants has shown that the directional singularity of Tomimatsu and Sato[2] is not actually a singularity; the values which the invariants attain depend on the manner in which one approaches the poles, $x=1$, $y=\pm 1$, but these values are never infinite. This immediately suggests the possibility of casting the metric tensor into a form which remains nonsingular at the poles. Such a transformation of coordinates could possibly reveal that the points $x=1$, $y=\pm 1$ are actually surfaces; the directional dependence of the values of the curvature invariants would then arise as a result of approaching different points on these surfaces.

Some recent work[3] by Ernst has shown that it is possible to express the exact T–S $\delta=2$ solution in a more tractable form. With the T–S solution in this newly discovered form, Ernst has been able to define a system of coordinates in terms of which the metric tensor remains finite at the poles. The present paper considers that metric which results from a specific lower order approximation to the exact T–S solution. The space–time defined by this metric is found to correspond to a vacuum gravitational field which is Petrov type D. This fact is not too surprising, because the exact T–S solution is type D along the symmetry axis $y^2=1$ and along the surfaces $x=\pm 1$.[1] Furthermore, the calculation of the Weyl tensor component C_0 for the approximate metric gives a result which agrees with the exact expression for C_0, when this expression is approximated to the lowest nonvanishing order. These two results seem to indicate that the "type D" metric is a valid approximation to the T–S solution near the poles, in the sense that the geometry of the T–S solution can be represented by the type D geometry there. Finally, the approximate metric is shown to correspond to one of the type D vacuum solutions studied by Kinnersley.[4]

II. APPROXIMATE FORM OF THE T-S $\delta=2$ SOLUTION

In the complex potential formulation of axially symmetric, stationary gravitational fields, which is due to Ernst,[5] the determination of the geometry of space–time is reduced to finding the solution of a nonlinear partial differential equation of the form:

$$(\xi\xi^*-1)\nabla^2\xi=2\xi^*\nabla\xi\cdot\nabla\xi. \tag{1}$$

The T–S $\delta=2$ solution corresponds to the rational function $\xi=N/D$,[6] where (in prolate spheroidal coordinates)

$$N=p^2(x^4-1)+q^2(y^4-1)-2ipqxy(x^2-y^2), \tag{2a}$$

$$D=2px(x^2-1)-2iqy(1-y^2). \tag{2b}$$

With this result, the metric can be written in the Weyl form as

$$ds^2=f^{-1}\left[P^{-2}\left(\frac{dx^2}{x^2-1}+\frac{dy^2}{1-y^2}\right)+\rho^2\,d\phi^2\right]-f(dT-\omega\,d\phi)^2, \tag{3}$$

where $f=A/B$,

$$P^2=p^4(x^2-y^2)^3A^{-1},\qquad \omega=4qp^{-1}CA^{-1}(1-y^2),\ \text{and} \tag{4a}$$

$$A=[p^2(x^2-1)^2+q^2(1-y^2)^2]^2-4p^2q^2(x^2-1)(1-y^2)(x^2-y^2)^2,$$

$$B=[p^2(x^4-1)+q^2(y^4-1)+2px(x^2-1)]^2+4q^2y^2[px(x^2-y^2)+(1-y^2)]^2, \tag{4b}$$

$$C=p^2(x^2-1)[(x^2-1)(1-y^2)-4x^2(x^2-y^2)]-p^3x(x^2-1)\times[2(x^4-1)+(x^2+3)(1-y^2)]+q^2(1+px)(1-y^2)^3. \tag{4c}$$

The specific method of approximation which was utilized in our work consisted of keeping only the dominant term in each of the expressions (4a)–(4c) when close to the pole region $x=1$, $y=1$ (we restrict attention to $y=1$, although the entire analysis is also valid for $y=-1$, with some minor modifications). This manner of approximating the metric must, of course, be justified at some point. If one assumes its validity for the present and denotes x^2-1 and $1-y^2$ by V and W respectively, then the metric assumes the following approximate form:

$$ds^2=\frac{a_1}{V+W}\left(\frac{dV^2}{V}+\frac{dW^2}{W}\right)+a_2\,d\phi^2-2a_3(V+W)\,d\phi\,dT-a_4(V+W)^2\,dT^2, \tag{5}$$

where the functions a_1, a_2, a_3, and a_4,

$$a_1=2(1+p)(p^2V^2+q^2W^2)/[p^4(V+W)^2], \tag{6a}$$

$$a_2=8(1+p)VW/(p^2V^2+q^2W^2), \tag{6b}$$

$$a_3 = 2pqVW/(p^2V^2 + q^2W^2), \tag{6c}$$

$$a_4 = (p^2V^2 + q^2W^2)/[8(1+p)(V+W)^2] - p^2q^2VW/[2(1+p)(p^2V^2+q^2W^2)], \tag{6d}$$

essentially depend only on the ratio of V and W.

It was found by Ernst[3] that by introducing the coordinates

$$r = \tfrac{1}{4}p^2(1+p)^{-1}(V+W), \quad \theta = 2\cot^{-1}\sqrt{V/W}, \tag{7}$$

and by defining

$$du = dT + dr/r^2, \tag{8a}$$

$$d\phi' = d\phi + q\,dr/pr, \tag{8b}$$

$$\Lambda^2 = (p^2V^2 + q^2W^2)/[p^2(V+W)^2], \tag{8c}$$

the metric can be put in the form:

$$\tfrac{1}{2}(1+p)^{-1}p^2\,ds^2 = \Lambda^2[(2\,dr - r^2\,du)\,du + d\theta^2] + \Lambda^{-2}\sin^2\theta[d\phi' - (q/p)r\,du]^2, \tag{9}$$

which remains nonsingular at the pole, where $r=0$. (It should be noted that for the T–S metrics x is interpreted as a radial coordinate, with the range $-\infty < x < \infty$, while y is interpreted as an angular coordinate, with the range $-1 \leq y \leq 1$. In this paper, the region $x \geq 1$ is being studied. Therefore, $r=0$ corresponds to the pole $x=y=1$.) With the line element in the above form, it is evident that the pole $r=0$ corresponds to a surface whose metrical properties are described by the following induced metric:

$$\tfrac{1}{2}(1+p)^{-1}p^2\,ds^2 = \Lambda^2\,d\theta^2 + \Lambda^{-2}\sin^2\theta\,d\phi'^2. \tag{10}$$

III. ANALYSIS OF THE APPROXIMATE METRIC

At this point it is appropriate to consider the approximate metric as a given exact solution to the Einstein field equations, in order to determine certain characteristics of the space–time defined by the metric. Such an analysis will reveal whether or not the derived metric can be considered as a valid approximation to the T–S metric close to the pole. One cannot know initially, for example, what the Petrov type of the metric is, or if the metric corresponds to a vacuum solution.

During the course of our work, it was found advantageous to introduce the following null tetrad (for the present, Λ is being considered as an arbitrary function of θ):

$$k = du, \tag{11a}$$

$$m = \Lambda^2(dr - \tfrac{1}{2}r^2\,du), \tag{11b}$$

$$t = \Lambda^{-1}\sin\theta[d\zeta - (1/\sqrt{2})\,qp^{-1}r\,du], \tag{11c}$$

with

$$d\zeta = (1/\sqrt{2})(d\phi' + i\Lambda^2\csc\theta\,d\theta), \tag{12}$$

in terms of which the metric tensor is given by the expression $g_{\mu\nu} = 2k_{(\mu}m_{\nu)} + 2t_{(\mu}t^*_{\nu)}$. Alternatively, one has

$$du = k, \tag{13a}$$

$$dr = \Lambda^{-2}m + \tfrac{1}{2}r^2k, \tag{13b}$$

$$d\zeta = \Lambda\csc\theta\,t + (1/\sqrt{2})qp^{-1}rk, \tag{13c}$$

$$d\theta = (i/\sqrt{2})\Lambda^{-1}(t^* - t). \tag{13d}$$

In null tetrad form, the equations involving the affine connection can be written as[7]

$$dk = Pk + v^*t + vt^*, \tag{14a}$$

$$dm = -Pm + wt + w^*t^*, \tag{14b}$$

$$dt = -w^*k - vm + iQt, \tag{14c}$$

where v, $u = P + iQ$, and w are the connection 1-forms. One can express v, as well as u and w, in terms of its null tetrad components as

$$v = v_km + v_mk + v_tt^* + v_{t^*}t. \tag{15}$$

The null tetrad components of the connection forms are also referred to as the spin coefficients. In particular, v_k, v_t, $\mathrm{Re}\,v_{t^*}$, and $\mathrm{Im}\,v_{t^*}$ correspond to the geodesy, shear, expansion, and twist, respectively, of a congruence of rays. The use of expressions (11a)–(11c), (13a)–(13d), and (14a)–(14c) results in the following expressions for the spin coefficients:

$$v_k = v_t = v_{t^*} = 0, \tag{16a}$$

$$v_m = -(1/2\sqrt{2})qp^{-1}\Lambda^{-3}\sin\theta - (i/2\sqrt{2})\Lambda^{-1}\frac{\partial\ln\Lambda^2}{\partial\theta}, \tag{16b}$$

$$w_m = w_t = w_{t^*} = 0, \quad w_k = -v_m, \tag{16c}$$

$$P_k = 0, \quad P_m = -r, \quad P_t = v_m, \quad Q_k = Q_m = 0, \tag{16d}$$

$$Q_t = (1/\sqrt{2})\Lambda^{-1}\frac{\partial}{\partial\theta}\ln(\Lambda^{-1}\sin\theta). \tag{16e}$$

Therefore, $v = v_mk$, $w = w_km$, and $u = P_mk + u_tt^* + u_{t^*}t$. Proceeding further, the field equations are given by the following expressions:

$$dv + vu = C_2B_{-1} + C_1B_0 + (C_0 + R/12)B_1 + \tfrac{1}{2}S_{kk}B^*_{-1} + \tfrac{1}{2}S_{kt}B^*_0 + \tfrac{1}{2}S_{tt}B^*_1, \tag{17a}$$

$$du - 2wv = -2C_1B_{-1} - 2(C_0 - R/24)B_0 - 2C_{-1}B_1 - S_{kt^*}B^*_{-1} - S_{tt^*}B^*_0 + S_{mt}B^*_1, \tag{17b}$$

$$dw - wu = (C_0 + R/12)B_{-1} + C_{-1}B_0 + C_{-2}B_1 + \tfrac{1}{2}S_{t^*t^*}B^*_{-1} - \tfrac{1}{2}S_{mt^*}B^*_0 + \tfrac{1}{2}S_{mm}B^*_1, \tag{17c}$$

where $B_{+1} = kt$, $B_0 = km + tt^*$, and $B_{-1} = mt^*$, and where the numerical subscript denotes spin-weight. Using the field equations, in conjunction with the connection one-forms, one can obtain expressions for the Weyl tensor components and the stress-energy tensor components. These are:

$$C_2 = C_1 = C_{-1} = C_{-2} = 0, \tag{18a}$$

$$S_{kk} = S_{kt} = S_{kt^*} = S_{mm} = S_{mt} = S_{mt^*} = 0, \tag{18b}$$

$$C_0 + \frac{R}{12} = \frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial v_m}{\partial\theta} + v_mu_{t^*}, \tag{18c}$$

$$C_0 + \frac{R}{12} = -\frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial w_k}{\partial\theta} - \frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial\ln\Lambda^2}{\partial\theta}\circ w_k - w_ku_t, \tag{18d}$$

$$-2\left(C_0 - \frac{R}{24}\right) - S_{tt^*} = -\Lambda^{-2}\frac{\partial P_m}{\partial r} + 2w_kv_m + \frac{1}{\sqrt{2}}qp^{-1}\Lambda^{-3}\sin\theta(u_t + u_{t^*}), \tag{18e}$$

$$-2\left(C_0-\frac{R}{24}\right)+S_{tt^*}=-\frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial}{\partial\theta}(u_t+u_{t^*})$$

$$-\frac{i}{\sqrt{2}}\Lambda^{-1}(u_t+u_{t^*})\frac{\partial}{\partial\theta}\ln(\Lambda^{-1}\sin\theta), \qquad (18f)$$

$$\tfrac{1}{2}S_{tt}=-\frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial v_m}{\partial\theta}+v_m u_t, \qquad (18g)$$

$$\tfrac{1}{2}S_{t^*t^*}=\frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial w_k}{\partial\theta}+\frac{i}{\sqrt{2}}\Lambda^{-1}\frac{\partial\ln\Lambda^2}{\partial\theta}\circ w_k-w_k u_{t^*}. \qquad (18h)$$

From the spin coefficients (16a)–(16e), it is found that Eqs. (18c) and (18d) are identical. Furthermore, Eqs. (18g) and (18h) imply that $S_{tt}=S_{t^*t^*}$; that is, S_{tt} is real. Restricting our attention to the particular expression for Λ given by (8c), Eqs. (18a)–(18h) give

$$R=S_{tt}=S_{t^*t^*}=S_{tt^*}=0, \qquad (19)$$

and

$$C_0=-\tfrac{1}{2}p^2(p-iq)^2[p^2-\sin^2(\theta/2)-ipq]^{-3}. \qquad (20)$$

The above results show that the metric corresponds to a vacuum, type D ($C_0\neq 0$; $C_2=C_1=C_{-1}=C_{-2}=0$) solution of Einstein's gravitational field equations. The vanishing of the Weyl component C_2 and the spin coefficients v_k, v_t, and v_{t^*} indicate that k is a principal null direction and that the space–time admits a nontwisting, nondiverging, shear-free null geodesic congruence of rays. Since all of the possible type D vacuum solutions have been enumerated by Kinnersley,[4] the above metric must be a known solution. This point will be considered in greater detail below.

IV. CONCERNING THE VALIDITY OF THE APPROXIMATION

The question which has remained unanswered, thus far, is concerned with whether, or not, the space–time geometry of the exact T–S solution can be accurately portrayed by the geometry of the "type D" metric when one is close to the pole. We believe that there are several good reasons which justify the use of the approximate metric. To begin with, the exact T–S $\delta=2$ solution has been shown to be algebraically general everywhere except along the symmetry axis $y^2=1$ and along the surfaces $x=\pm 1$, where it is Petrov type D.[1] Thus, it seems plausible that the region of space–time near the point $x=1$, $y=1$ could be approximated by a type D solution of the gravitational field equations. Secondly, if one considers the exact expression for the Weyl tensor component C_0[1]:

$$C_0=p^4Z^{-3}T, \qquad (21)$$

where $T=2-px(x^2-3)-iqy(3-y^2)$ and $Z=(N+D)/(x^2-y^2)$, then, in the neighborhood of $x=1$, $y=1$, it can be shown to be approximately equal to

$$C_0=\tfrac{1}{8}p^4(1+p)^{-1}(p-iq)^2[p^2-\sin^2(\theta/2)-ipq]^{-3}, \qquad (22)$$

which, except for a constant factor of $-\frac{1}{4}(1+p)^{-1}p^2$, agrees exactly with Eq. (20), the expression for C_0 obtained through use of the type D metric.[8] Finally, one would expect to be able to find a *vacuum* solution which approximates the vacuum T–S solution near the pole.

V. COMPARISON WITH KNOWN TYPE D METRICS

As was mentioned above, the approximate metric, Eq. (9), must correspond to a known type D vacuum solution. Our tetrad can, in fact, be shown to be related to the tetrad found by Kinnersley[4] for those type D vacuum metrics which admit nontwisting, nondiverging rays. In terms of the covectors **k**, **m**, and **t**, our tetrad can be written as

$$\mathbf{k}=\Lambda^{-2}\mathbf{a}_r, \qquad (23a)$$

$$\mathbf{m}=\mathbf{a}_u+qp^{-1}r\mathbf{a}_\phi+\tfrac{1}{2}r^2\mathbf{a}_r, \qquad (23b)$$

$$\mathbf{t}=(1/\sqrt{2})(\Lambda\csc\theta\,\mathbf{a}_\phi+i\Lambda^{-1}\mathbf{a}_\theta), \qquad (23c)$$

where, for example, $a_r\Gamma d=\partial/\partial r$ in a more familiar notation. To facilitate comparison with Kinnersley's work, we give here the expression which he obtains for the only nonvanishing component of the Weyl conform tensor:

$$\psi=\psi_2=(m+il)(x+ia)^{-3}, \qquad (24)$$

which is related to expression (20) for C_0. By setting $C_0=-\psi$, one can identify

$$x=4[p^2-\sin^2(\theta/2)], \qquad (25a)$$

$$a=-4pq, \qquad (25b)$$

$$m=32p^2(p^2-q^2), \qquad (25c)$$

$$l=-64p^3q, \qquad (25d)$$

and, with a little algebra, it becomes evident that

$$x^2+a^2=16p^2\Lambda^2, \qquad (26a)$$

$$\xi^2=2\Lambda^{-2}\sin^2\theta, \qquad (26b)$$

where $\xi^2=[2amx+l(a^2-x^2)]/[2a(x^2+a^2)]$. The further identification

$$r'=\Lambda^2 r,\quad y=\phi,\quad u'=-u \qquad (27)$$

enables one to express the tetrad vectors, Eqs. (23), as:

$$\mathbf{k}=\mathbf{a}_r', \qquad (28a)$$

$$\mathbf{m}=-\mathbf{a}_u'+lr'^2[2a(x^2+a^2)]^{-1}\mathbf{a}_r'-4ar'(x^2+a^2)^{-1}\mathbf{a}_y, \qquad (28b)$$

$$\mathbf{t}=-2ir'x\xi(x^2+a^2)^{-1}\mathbf{a}_r'-i\xi\mathbf{a}_x+\xi^{-1}\mathbf{a}_y, \qquad (28c)$$

which is consistent with the tetrad found by Kinnersley.

VI. CONCLUSION

It is extremely interesting that the algebraically general T–S solution can be approximated by a type D solution near the points $x=1$, $y=\pm 1$. The type D metric has several desirable properties which make it a natural tool for studying the geometry of the T–S solution in the regions of the poles. To begin with, the metric reveals the nonsingular nature of the points $x=1$, $y=\pm 1$ and the fact that these points are actually surfaces. This latter result verifies what was previously suspected from the Weyl tensor calculation.[1] Furthermore, the approximate metric is of importance when one is considering the geodesic problem in the neighborhood of either pole. It has been shown by Ernst[3] that this space–time admits an additional constant of the motion which is associated with a Killing tensor; it is possible, therefore, to analyze the geodesic problem completely. In his work, Ernst discusses the nature of timelike, spacelike, and null geodesics close to the poles and the relevance of these geodesics for the exact T—S solution.

A preliminary analysis by Ernst and Wild has indicated that the study of the topology of the T—S solution near the poles merits further attention. Hopefully, future work on this problem will contribute to an even deeper understanding of the T—S solution. One aspect of this work, for example, would be concerned with possible sources of the T—S field. Another interesting area of research involves the T—S $\delta = 3$ solution and the possibility of analyzing this space—time in terms of an approximate metric. The question naturally arises as to whether the geometry of the pole regions can be approximated by a type D metric. A determination of the Weyl tensor invariants would indicate whether a type D geometry is feasible. If it happens that a similar analysis can, in fact, be carried to completion for the $\delta = 3$ solution, then one would be in a good position to speculate on the entire class of T—S solutions.

ACKNOWLEDGMENTS

The author would like to acknowledge gratefully the help and encouragement of Dr. Frederick J. Ernst. He has also benefited from several discussions with Mr. Walter J. Wild.

[1]J. E. Economou and F. J. Ernst, J. Math. Phys. 17, 52 (1976).
[2]A. Tomimatsu and H. Sato, Progr. Theor. Phys. 50, 95 (1973).
[3]F. J. Ernst, J. Math. Phys. 17, 1091 (1976), preceding paper.
[4]W. Kinnersley, J. Math. Phys. 10, 1195 (1969).
[5]For a description of the complex potential formulation of axially symmetric, stationary gravitational fields, the reader is referred to F. J. Ernst, Phys. Rev. 167, 1175 (1968); 168, 1415 (1968); J. Math. Phys. 15, 1409 (1974).
[6]A. Tomimatsu and H. Sato, Phys. Rev. Lett. 29, 1344 (1972).
[7]The null tetrad formalism which is used throughout this paper is due mainly to Dr. Isidore Hauser. It is described in detail in a number of unpublished articles of the IIT relativity seminar. The reader may also refer to the previously mentioned article by F. J. Ernst, J. Math. Phys. 15, 1409 (1974).
[8]The appearance of the additional constant factor can be attributed to using an entirely different tetrad system.

The solution to an inverse problem in stratified dielectric media

B. Gopinath

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 15 December 1975)

We solve the problem of determining the average dielectric constant and thickness of the layers that constitute a stratified dielectric medium from measurements of transmitted power at a single frequency. Each sample of the medium that is available for measurement is modeled as a stack of "n" layers of dielectrics of thicknesses l_i and dielectric constants K_i (see Fig. 1). We assume that n, l_i, K_i, $i=1,2,\ldots$, are all independent random variables and their values, of course, depend on the particular sample and the layer indexed by "i". Furthermore, it is assumed that the l_i are identically distributed with some exponential distribution and that the K_i are identically distributed. There are no other constraints or assumptions about these distributions except the following which are made more precise in the text: (1) The various averages of interest are finite. (2) Distributions which precipitate certain singular conditions (which is not a problem in "almost all cases") are excluded. Then the method described in this paper determines uniquely: (1) the average of the thicknesses of the layers El_i; (2) the average of the dielectric constants EK_i from measurements of transmitted power and without any further knowledge of the distributions of l_i, K_i, or n. We remark that the theory presented here applies also to acoustic or mechanical systems with the appropriate interpretation of the physical parameters.

The direct problem of determining the properties of wave propagation in a nonuniform medium has been the subject of many recent papers.[1-4] The medium is generally modelled as an assemblage of layers of varying thicknesses and dielectric constants. Related are the models of disordered linear chains (DLC) introduced by Dyson.[5] (A DLC is a chain of one-dimensional simple harmonic oscillators each coupled to its nearest neighbors where the moments of inertia and strength of coupling are random variables.) These models serve to characterize nonuniformities in the medium. Though considerable attention has been focused on the direct problems, e.g., determination of moments of solutions of the wave equation, transmission coefficients, spectral functions, etc., the author knows of no work which attempts to determine the averages of the parameters of a model of a medium from averages of solutions to the wave equation or transmission coefficient, etc. Of course for a linear medium when measurements can be made at all frequencies the problem for each sample of the medium reduces to the inverse problem[6,7] in either the discrete or continuous version depending on the model. However when measurements are made at only a single frequency, the solution in the nonrandom case is of no apparent help.

In this paper we solve the problem of determining the average dielectric constant and thickness of the layers that constitute the medium from power type measurements at one frequency. Each sample of the medium that is available for measurement is modelled here as a stack of n layers of dielectrics of thicknesses l_i and dielectric constants K_i (see Fig. 1). Here n, l_i, and K_i are random variables and their values depend on the particular sample of the medium and the layer indexed by i. We assume throughout that n, l_i, K_i, $i=1,\ldots,$ are all independent random variables. Furthermore, it is assumed that the l_i are identically distributed with some exponential distribution, and that the K_i are identically distributed. Such models have arisen in modeling plasmas as in Ref. 8 for example.

There are no other constraints or assumptions about these distributions except the following:

(1) The various averages of interest [made precise in Eq. (20)] are finite.

(2) Distributions which precipitate certain singular conditions (which is not a problem in "almost all cases") are excluded.

Then we show that the method described below determines uniquely:

(1) The average of the thicknesses of the layers El_i.

(2) The average of the dielectric constants EK_i.

Each sample of the medium is sandwiched between distinct dielectric systems L_i, R_j, $i=1,2,3,4$, $j=1,2,3,4$ as shown in Fig. (3). For each i, j the transmitted power from a plane monochromatic electromagnetic wave of known intensity, normally incident on the system is measured. The average reciprocal power transmission coefficient[9] of the samples of the medium is calculated for the sixteen possible choices of L_i and R_j. From these sixteen measurements the "average power transfer matrix" can be determined. The eigenvectors of this matrix are shown to contain all the information needed to determine the required averages.

The surprising nature of this result, namely that with no knowledge about the distributions of K_i and n it is possible to determine uniquely the averages of l_i and K_i, stems from the rich structure that the power transfer matrix of a single layer has [see Eqs. (22)–(28)]. Since this structure has primarily only mathematical interest the proofs of various propositions are relegated to the Appendix.

1. PRELIMINARIES

Let us consider a sample of the transmission medium. Suppose it consist of n dielectric layers of thicknesses

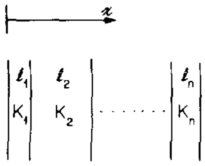


FIG. 1.

$l_1, l_2, \ldots, l_n$ stacked as shown in Fig. 1. The dielectric constant of the ith layer is K_i. We assume plane wave propagation throughout. Denote by x the distance along the direction of propagation (see Fig. 1). The space dependent part of the electric field $e(x)$ in the sample of the medium then satisfies Eq. (1)

$$\frac{d^2e(x)}{dx^2} + \beta^2(x, \omega)e(x) = 0, \quad 0 \le x < x_n, \tag{1}$$

where

$$\beta^2(x, \omega) = \omega^2\mu_0\epsilon_0 K_i(\omega) = \beta_i^2, \quad x_{i-1} \le x < x_i, \tag{2}$$

$$0 = x_0 \text{ and } x_i = \sum_{j=1}^{i} l_j, \quad i = 1, \ldots, n, \tag{3}$$

$\omega\sqrt{\mu_0\epsilon_0}$ is the free space wavenumber,[10] and $K_i(\omega)$ of course is independent of ω for a linear medium. Let $Z_i^t = [e(x_i), e'(x_i)]$. Then it is easily seen from (2) that

$$\begin{bmatrix} e(x_i) \\ e'(x_i) \end{bmatrix} = \begin{bmatrix} \cos\beta_i l_i & (1/\beta_i)\sin\beta_i l_i \\ -\beta_i \sin\beta_i l_i & \cos\beta_i l_i \end{bmatrix} \begin{bmatrix} e(x_{i-1}) \\ e'(x_{i-1}) \end{bmatrix}. \tag{4}$$

(Superscript "t" denotes transpose of a matrix or vector and prime denotes differentiation w.r.t. x. Whenever convenient, arguments of functions will be omitted.) Hence

$$Z_i = F_i Z_{i-1}, \tag{5}$$

where F_i stands for the matrix appearing on the right-hand side of (4).

Represent the two independent solutions of (1) for initial conditions $u(0) = 1$, $u'(0) = 0$ and $u(0) = 0$, $u'(0) = 1$ by $u_1(x)$ and $u_2(x)$ respectively. Then

$$\begin{bmatrix} u_1(x_n) & u_2(x_n) \\ u_1'(x_n) & u_2'(x_n) \end{bmatrix} = F_n F_{n-1} \cdots F_1 \triangleq \Phi_n. \tag{6}$$

Φ_n is called the fundamental matrix of the system of dielectrics consisting of layers $1, 2, \ldots, n$.

2. DESCRIPTION OF THE MEASUREMENTS

In order to understand the measurement contemplated in this paper let us sandwich this sample between two dielectric media in which measurements are made as shown in Fig. 2.

Let β_l and β_r be related in the obvious way[2] to K_l and K_r for the dielectric media at the left and right respectively. Let a plane monochromatic electromagnetic wave of radian frequency ω be normally incident on the sample of transmission media. Part of this wave will be reflected into the boundary dielectric on the left, and part will be transmitted into the boundary dielectric on the right. Normalizing the amplitude of the incident wave to unity we represent the amplitude of the reflected and transmitted waves by R and T. By power type measurements we mean measurement of $T\tilde{T}$ or $R\tilde{R}$, which

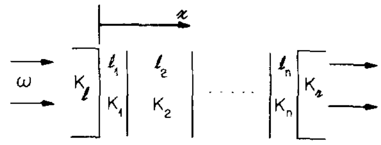


FIG. 2.

of course have no information in them of the phase of the transmitted or reflected wave. Here ~ denotes complex conjugation. We now calculate $T\tilde{T}$ by relating it to the fundamental solutions of (1). Then $e(x)$ can be expressed as follows:

$$e(x) = \exp(-i\beta_l x) + R\exp(i\beta_l x), \quad x \le 0, \tag{7}$$

$$e(x) = A_1 u_1(x) + A_2 u_2(x), \quad 0 \le x \le x_n, \tag{8}$$

$$e(x) = T\exp[-i\beta_r(x - x_n)], \quad x \ge x_n. \tag{9}$$

Making use of the continuity of $e(x)$ and $e'(x)$ at $x = 0$ and $x = x_n$, we can eliminate A_1 and A_2 and obtain

$$\begin{bmatrix} T \\ -i\beta_r T \end{bmatrix} = \begin{bmatrix} u_1(x_n) & u_2(x_n) \\ u_1'(x_n) & u_2'(x_n) \end{bmatrix} \begin{bmatrix} R+1 \\ i\beta_l(R-1) \end{bmatrix}, \tag{10}$$

or

$$\frac{1}{T}\begin{bmatrix} R+1 \\ i\beta_l(R-1) \end{bmatrix} = \begin{bmatrix} u_1(x_n) & u_2(x_n) \\ u_1'(x_n) & u_2'(x_n) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -i\beta_r \end{bmatrix}, \tag{11}$$

from which

$$\frac{2i\beta_l}{T} = [i\beta_l - 1]\begin{bmatrix} u_1(x_n) & u_2(x_n) \\ u_1'(x_n) & u_2'(x_n) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -i\beta_r \end{bmatrix}. \tag{12}$$

Let $[i\beta_l - 1] = b_l^t$ and $[1, -i\beta_r] = b_r^t$. Then from (6)

$$\frac{4\beta_l^2}{T\tilde{T}} = b_l^t \Phi_n^{-1} b_r \tilde{b}_l^t \Phi_n^{-1} \tilde{b}_r. \tag{13}$$

The reciprocal power transmission coefficient (RPTC) is $1/T\tilde{T}$. Hence

$$\frac{4\beta_l^2}{T\tilde{T}} = (b_l \otimes \tilde{b}_l)^t \Phi_n^{-1} \otimes \Phi_n^{-1}(b_r \otimes \tilde{b}_r), \tag{14}$$

$\otimes$ here denotes the standard tensor product.

It was shown in Ref. 9 that if $n = 1$ then just by changing the dielectric media on the left and right of the "slab" the average of interest can be determined. However when $n > 1$, this is not so. The problem still can be solved by placing known dielectric systems on the right and left as shown in Fig. 3.

Let L_i, $i = 1, 2, 3, 4$, R_j, $j = 1, 2, 3, 4$ be the fundamental matrices for two sets of dielectric systems L_i and R_j. By changing L_i and R_j we get information about the power transfer matrix $\Phi_n \otimes \Phi_n$. The fundamental matrix for the composite system then is $R_j \Phi_n L_i$. Then a total of sixteen measurements yields

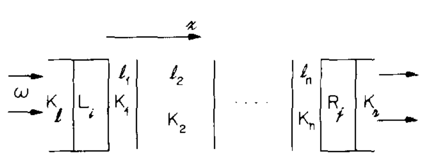


FIG. 3.

$$\frac{4\beta^2_l}{(T\tilde{T})_{ij}} = (b_l \otimes \tilde{b}_l)^t (L_i^{-1} \otimes L_i^{-1}) \Phi_n^{-1} \otimes \Phi_n^{-1} (R_j^{-1} \otimes R_j^{-1}) \times (b_r \otimes \tilde{b}_r), \quad i, j = 1, 2, 3, 4. \tag{15}$$

If $(b_l \otimes \tilde{b}_l)^t L_i^{-1} \otimes L_i^{-1}$ and $(R_j^{-1} \otimes R_j^{-1}) b_r \otimes \tilde{b}_r$ are linearly independent, then the sixteen measurements in (15) yield $\Phi_n^{-1} \otimes \Phi_n^{-1}$ uniquely. Since L_i and R_j have a special structure we have to prove that we can find dielectrics which allow these vectors to be linearly independent. That is the content of the following lemma.

Lemma 1: Let K_{-i}, l_{-i} be the dielectric constants and lengths for a set of dielectrics L_i, $i = 1, 2, 3, 4$, such that $2\beta_{-i} l_{-i} = \pi/2$ for $i = 1, 2, 3, 4$ [β_{-i} as in (2)]. If β_{-i} are all distinct, then the set of vectors $(b_l \otimes b_l)^t L_i^{-1} \otimes L_i^{-1}$ are linearly independent.

Proof: See Appendix.

It is clear that a similar result holds for $(R_j^{-1} \otimes R_j^{-1}) \times (b_r \otimes b_r)$, $j = 1, 2, 3, 4$.

Let G denote the fundamental matrix of a sample of the medium then the average value of $G \otimes G$ can be determined uniquely by measuring the RPTC for the 16 cases given by (15) as follows. From (15),

$$E \frac{4\beta_l^2}{(T\tilde{T})_{ij}} = (b_l \otimes \tilde{b}_l)^t (L_i^{-1} \otimes L_i^{-1}) E[(G^{-1} \otimes G^{-1})] \times (R_j^{-1} \otimes R_j^{-1})(b_r \otimes \tilde{b}_r), \quad i, j = 1, 2, 3, 4, \tag{16}$$

so that from the 16 measurements of RPTC we have $E[G^{-1} \otimes G^{-1}]$. But

$$E(G^{-1} \otimes G^{-1}) = \sum_{n=1}^{\infty} E_n(\Phi_n^{-1} \otimes \Phi_n^{-1}) p_n, \tag{17}$$

where p_n is the probability that the sample consists of n layers and E_n denotes conditioning with respect to n.

Since the determinant of $\Phi_n = 1$ for all n, the elements of Φ_n^{-1} are related simply to the elements of Φ_n and hence those of G and G^{-1},

$$\begin{aligned} (G^{-1})_{11} &= G_{22}, \quad (G^{-1})_{12} = -G_{21}, \\ (G^{-1})_{22} &= G_{11}, \quad (G^{-1})_{21} = -G_{12}. \end{aligned} \tag{18}$$

Since these relations are independent of the values of any of the random variables we can determine $EG \otimes G$ from $EG^{-1} \otimes G^{-1}$ using (18),

$$EG \otimes G = \sum_{n=1}^{\infty} p_n E_n \Phi_n \otimes \Phi_n.$$

Using F to denote the fundamental matrix of a single layer

$$E_n \Phi_n \otimes \Phi_n = [EF \otimes F]^n, \tag{19}$$

since the parameters of each layer are independent identically distributed, hence

$$EG \otimes G = \sum_{n=1}^{\infty} (EF \otimes F)^n p_n. \tag{20}$$

If τ is the eigenvalue of largest magnitude of $EF \otimes F$ then $\lim_{n\to\infty} p_n^{1/n} < 1/\tau$ is sufficient to guarantee the finiteness of $EG \otimes G$. We will assume throughout that this is so.

We will be determining all the averages from the eigenvectors of $EG \otimes G$, when $EG \otimes G$ has distinct eigenvalues.

As we will show in Sec. 3, the eigenvalues of $EF \otimes F$ are distinct in "almost all cases." However even if this is so, $EG \otimes G$ may have eigenvalues which are equal. No two eigenvalues of $EG \otimes G$ are equal when with λ_i, $i = 1, 2, 3, 4$, denoting the eigenvalues of $E(F \otimes F)$ we have

$$\sum_{n=0}^{\infty} p_n(\lambda_i^n - \lambda_j^n) \neq 0 \quad \text{if } i \neq j. \tag{21}$$

This is only a mild condition on the distribution of n. We need not know *a priori* whether this is so, because the measurements in (16) yield $EG \otimes G$ and hence we just have to check whether $EG \otimes G$ has distinct eigenvalues or not. We will assume in what follows that $EG \otimes G$ has distinct eigenvalues wherever needed.

3. POWER TRANSFER MATRIX OF A SINGLE LAYER

We derive here the power transfer matrix for a single layer. Since from (4), F_i is known in terms of β_i and l_i, for the ith layer, we can omit the subscript i and write for a single layer,

$$F = \begin{bmatrix} \cos\beta l & (1/\beta)\sin\beta l \\ -\beta \sin\beta l & \cos\beta l \end{bmatrix}.$$

From this it can tediously be checked that

$$EF \otimes F = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_2 & \alpha_4 \\ \alpha_3 & \alpha_1 & \alpha_5 & \alpha_2 \\ \alpha_3 & \alpha_5 & \alpha_1 & \alpha_2 \\ \alpha_6 & \alpha_3 & \alpha_3 & \alpha_1 \end{bmatrix} \triangleq A, \tag{22}$$

where

$$\alpha_1 = \tfrac{1}{2}E(1 + \cos 2\beta l), \tag{23}$$

$$\alpha_2 = \tfrac{1}{2}E(\sin 2\beta l/\beta), \tag{24}$$

$$\alpha_3 = -\tfrac{1}{2}E(\beta \sin 2\beta l), \tag{25}$$

$$\alpha_4 = \tfrac{1}{2}E(1/\beta^2)(1 - \cos 2\beta l), \tag{26}$$

$$\alpha_5 = -\tfrac{1}{2}E(1 - \cos 2\beta l), \tag{27}$$

$$\alpha_6 = \tfrac{1}{2}E\beta^2(1 - \cos 2\beta l). \tag{28}$$

From (23) and (27) it is clear that $\alpha_1 - \alpha_5 = 1$, $\alpha_5 < 0$. From (26) and (28) it is easily seen that α_4 and α_6 are positive.

From (22) we can show that one of the eigenvalues of A is 1 and $(0, -1, 1, 0)^t$ is the corresponding eigenvector since $\alpha_1 - \alpha_5 = 1$. If we denote the other eigenvalues of A by λ_1, λ_2, λ_3 we can by calculating the characteristic polynomial of A show that $\theta_i = \lambda_i - \alpha_i$ are the roots of the following polynomial:

$$\chi(\theta) = \theta^3 - \alpha_5\theta^2 - (\alpha_4\alpha_6 + 4\alpha_2\alpha_3)\theta + \alpha_4\alpha_5\alpha_6 - 2\alpha_3^2\alpha_4 - 2\alpha_2^2\alpha_6 \tag{29}$$

$$= (\theta - \alpha_5)(\theta^2 - \alpha_4\alpha_6) - (4\alpha_2\alpha_3\theta + 2\alpha_3^2\alpha_4 + 2\alpha_2^2\alpha_6). \tag{30}$$

It can easily be shown that $\chi(\sqrt{\alpha_4\alpha_6}) < 0$ which implies that $\chi(\theta)$ has a positive root greater than $\sqrt{\alpha_4\alpha_6}$. Denote this root by θ_1. Now from (26), (27), and (28) an application of Schwartz's inequality[11] shows that $\alpha_5^2 < \alpha_4\alpha_6$ whenever β and l can assume more than one value. Assuming this is so

$$\theta_1 + \alpha_1 = \lambda_1 > \sqrt{\alpha_4 \alpha_6} + \alpha_1 > \alpha_1 - \alpha_5 = 1. \tag{31}$$

Furthermore it can be seen from the representation of $\chi(\theta)$ in (30) that the other two roots θ_2, θ_3 are either complex or, if real they are negative. Since $\chi(\theta)$ has real coefficients the roots are distinct when they are complex. It is conceivable that these two roots coincide when they are real. However, it can easily be checked out that when β can assume only one value, these are complex, hence also when β has small variance. Therefore except at most for a set of values of α_i, $i \geq 2$ which are solutions of some nontrivial algebraic equation, the eigenvalues are distinct. Hence, for our purposes we will assume that the eigenvalues of A are distinct.

The structure of the eigenvectors of A is the subject of the lemmas that follow. The proofs are found in the Appendix.

Lemma 2: Let $(e_1, e_2, e_3, e_4)^t$ be an eigenvector of A corresponding to the eigenvalue e, then $(e_4, e_3, e_2, e_1)^t$ is the eigenvector of A^t corresponding to e.

Lemma 3: If l is distributed exponentially and β can assume more than one value, then an eigenvector $(e_1, e_2, e_3, e_4)^t$ corresponding to any eigenvalue of A not equal to 1 can be normalized such that $e_2 = e_3 = 1$.

Hence for our purposes we can assume that the eigenvectors of A can be normalized in such a way that the eigenvectors corresponding to the eigenvalues 1, λ_1, λ_2, λ_3 are respectively

$$\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_1 \\ 1 \\ 1 \\ b_1 \end{bmatrix}, \begin{bmatrix} a_2 \\ 1 \\ 1 \\ b_2 \end{bmatrix}, \begin{bmatrix} a_3 \\ 1 \\ 1 \\ b_3 \end{bmatrix}. \tag{32}$$

With a_i, b_i as above we have the following lemmas.

Lemma 4: $a_i b_j + b_i a_j = -2, \quad i \neq j.$

Lemma 5: $a_i b_j - b_i a_j \neq 0, \quad i \neq j.$

4. DETERMINATION OF THE AVERAGE OF l_i, β_i

We now show that knowledge of a_i, b_i give us a set of equations that the α_i's must satisfy. From these equations we can uniquely determine the averages of interest.

Since we assume that the eigenvalues of $EG \times G$ [Eq. (20)] are distinct, the eigenvectors of $EG \times G$ are the eigenvectors of $EF \otimes F$ hence of A. Thus, from the set of 16 measurements described previously we can determine uniquely the eigenvectors of A normalized as in (32).

Writing out the equations that the eigenvectors satisfy we see that

$$\lambda_i = \alpha_3 a_i + \alpha_1 + \alpha_5 + b_i \alpha_2. \tag{33}$$

Using these expressions for λ_i we can arrive at the following equations for α_i, $i = 2, 3, 4, 5, 6$ in terms of a_i, b_i, $i = 1, 2, 3$:

$$\begin{bmatrix} 2 - a_1 b_1 & -a_1^2 & b_1 & -a_1 & 0 \\ 2 - a_2 b_2 & -a_2^2 & b_2 & -a_2 & 0 \\ 2 - a_3 b_3 & -a_3^2 & b_3 & -a_3 & 0 \\ -b_1^2 & 2 - a_1 b_1 & 0 & -b_1 & a_1 \\ -b_2^2 & 2 - a_2 b_2 & 0 & -b_2 & a_2 \\ -b_3^2 & 2 - a_3 b_3 & 0 & -b_3 & a_3 \end{bmatrix} \times \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} = 0. \tag{34}$$

It can be shown that any four rows of this matrix are linearly dependent so that the ratios α_i / α_j are not uniquely determined by these equations.

However, we can use the fact that the l_i's are exponentially distributed to simplify the expressions for the α_i's in (23)–(28). Suppose the l_i's are exponentially distributed with parameter λ, that is

$$\text{Prob}\{l_i \leq x\} = 1 - \exp(-\lambda x). \tag{35}$$

Then the average of each of the l_i is

$$El_i = 1/\lambda. \tag{36}$$

In this case then we can further simplify the expressions for α_i, $i = 1, 2, \ldots, 6$. Take, for example, α_6:

$$\begin{aligned} \alpha_6 &= \tfrac{1}{2} E\beta^2 (1 - \cos 2\beta l) \\ &= \tfrac{1}{2} E\beta^2 - \tfrac{1}{2} E\beta^2 [E_\beta (\cos 2\beta l)], \end{aligned} \tag{37}$$

i.e.,

$$\begin{aligned} \alpha_6 &= \tfrac{1}{2} E\beta^2 - \tfrac{1}{2} E\left(\beta^2 \frac{\lambda^2}{\lambda^2 + 4\beta^2}\right) \\ &= \tfrac{1}{2} E\beta^2 - \frac{1}{2}\frac{\lambda^2 E}{4}\left(1 - \frac{\lambda^2}{\lambda^2 + 4\beta^2}\right) \\ \therefore \alpha_6 &= \tfrac{1}{2} E\beta^2 + \frac{\lambda^2}{4}\left(\frac{\lambda^2 a}{2} - \frac{1}{2}\right), \end{aligned} \tag{38}$$

where

$$a = E\frac{1}{\lambda^2 + 4\beta^2}, \tag{39}$$

and E_β denotes conditioning with respect to β. A similar treatment of α_i, $i = 1, 2, \ldots, 5$ yields the following:

$$\alpha_1 = \frac{\lambda^2 a}{2} + \frac{1}{2}, \tag{40}$$

$$\alpha_2 = \lambda a, \tag{41}$$

$$\alpha_3 = \frac{\lambda}{2}\left(\frac{\lambda^2 a}{2} - \frac{1}{2}\right), \tag{42}$$

$$\alpha_4 = 2a, \tag{43}$$

$$\alpha_5 = \frac{\lambda^2 a}{2} - \frac{1}{2}, \tag{44}$$

$$\alpha_6 = \frac{\lambda^2}{4}\left[\frac{\lambda^2 a}{2} - \frac{1}{2}\right] + \frac{1}{2} E\beta^2. \tag{45}$$

Now using (41) and (44) we have

$$\alpha_2 = (\lambda/2)\,\alpha_4, \tag{46}$$

and (42) and (44) give

$$\alpha_3 = (\lambda/2)\,\alpha_5. \tag{47}$$

Using these relations we can eliminate α_2, α_3 from the first two equations in (34) yielding

$$\frac{\lambda}{2}\begin{bmatrix} 2-a_1b_1 & -a_1^2 \\ 2-a_2b_2 & -a_2^2 \end{bmatrix}\begin{bmatrix}\alpha_4\\ \alpha_5\end{bmatrix}+\begin{bmatrix} b_1 & -a_1 \\ b_2 & -a_2\end{bmatrix}\begin{bmatrix}\alpha_4\\ \alpha_5\end{bmatrix}=0. \tag{48}$$

Since $\Delta_1 \triangleq a_1b_2 - b_1a_2 \neq 0$ (see Lemma 5), for $(\alpha_4, \alpha_5) \neq 0$ to be a solution of (48), λ must be a root of the equation (writing the appropriate determinant equals zero, and using Lemma 4),

$$\frac{4\Delta_1}{\lambda^2} = 2(a_1^2 - a_2^2) - a_1a_2\Delta_1. \tag{49}$$

Hence the value of $\lambda > 0$ satisfying (49) gives the average thickness! [From (36)]. Furthermore, using this value of λ, we can find $\alpha_5/\alpha_4 = r$ (say) from (48). Hence from (44)

$$\lambda^2 a/2 - \tfrac{1}{2} = 2ra,$$

$$a = 1/(\lambda^2 - 4r). \tag{50}$$

Now since $a_i \neq 0$ for some i from one of the last three equations of (34) α_6 can be determined and hence $E\beta^2$ from (45).

5. SUMMARY OF THE METHOD

We summarize now in 12 steps the method of determining El_i, EK_i.

Step 1: Select boundary dielectric systems such that $(b_l \otimes \tilde{b}_l)^t L_i^{-1} \otimes L_i^{-1}$ and $(R_j^{-1} \otimes R_j^{-1}) b_r \otimes \tilde{b}_r$ are for $i, j = 1, 2, 3, 4$ two sets of linearly independent vectors.

Step 2: Sandwich each sample between L_i and R_j as shown in Fig. 3.

Step 3: Let a monochromatic plane wave of known intensity and radian frequency ω be directed normally at the composite system as in Fig. 3. Then measure the amplitude of the transmitted wave for each of the sixteen pairs of boundary dielectrics.

Step 4: Calculate the average of the reciprocal power transmission coefficient (15) for each of the 16 pairs of boundary dielectrics.

Step 5: From the 16 quantities of Step 4 we have the left-hand side of (16). Therefore $E[G^{-1} \otimes G^{-1}]$ is uniquely determined.

Step 6: From the formulas (18) determine $EG \otimes G$.

Step 7: Check whether the eigenvalues of $EG \otimes G$ are distinct. If they are not distinct we cannot use the method. If they are distinct we determine the required averages by completing the following remaining steps.

Step 8: Find the 4 linearly independent eigenvectors of $E(G \otimes G)$ which are the same as the eigenvectors of $EF \otimes F$.

Step 9: Normalize the eigenvectors as in (32), and find a_i, b_i, $i = 1, 2, 3$.

Step 10: Find λ as in (49).

Step 11: Find "a" from (50) and α_6 from one of the last three equations of (40) for which $a_i \neq 0$.

Step 12: Find $E\beta^2$ from (45). Then $E\beta^2 = \omega^2\mu_0\epsilon_0 EK(\omega)$. Therefore $EK(\omega)$ is determined.

APPENDIX

Lemma 1: Let K_{-i}, l_{-i} be the dielectric constants for the set of dielectrics L_i, $i = 1, 2, 3, 4$ such that $2\beta_{-i}l_{-i} = \pi/2$ for $i = 1, 2, 3, 4$, β_{-i} as in (2). If β_{-i} are all distinct then the set of vectors

$$(b_l \otimes \tilde{b}_l)^t (L_i^{-1} \otimes L_i^{-1}), \quad i = 1, 2, 3, 4,$$

are linearly independent.

Proof: Using formulas (4) we can show that $2L_i^{-1} \otimes L_i^{-1}$ is represented by

$$\begin{bmatrix} 1 & -1/\beta_{-i} & -1/\beta_{-i} & 1/\beta_{-i}^2 \\ \beta_{-i} & 1 & -1 & -1/\beta_{-i} \\ \beta_{-i} & -1 & 1 & -1/\beta_{-i} \\ \beta_{-i}^2 & \beta_{-i} & \beta_{-i} & 1 \end{bmatrix}. \tag{A1}$$

Whence $2(b_l \otimes \tilde{b}_l)^t L_i^{-1}$ becomes

$$B_i \triangleq \left(\beta_l^2 + \beta_{-i}^2,\ -\frac{\beta_l^2}{\beta_{-i}} + \beta_{-i} - 2i\beta_l,\right.$$
$$\left. -\frac{\beta_l^2}{\beta_{-i}} + \beta_{-i} + 2i\beta_l,\ \frac{\beta_l^2}{\beta_{-i}^2} + 1\right).$$

We can show that vectors B_i are linearly dependent (by column manipulations on the matrix whose rows are B_i) *iff* the vectors C_i given by

$$C_i = [\beta_{-i}^4, \beta_{-i}^3 - \beta_l^2\beta_{-i}, \beta_{-i}^2, \beta_l^2] \tag{A2}$$

are also linearly dependent. That is iff there exist constants not all zero such that

$$d_0\beta_{-i}^4 + d_1(-\beta_l^2\beta_{-i} + \beta_{-i}^3) + d_2\beta_{-i}^2 + d_3 = 0, \tag{A3}$$

i.e., the β_{-i} which are all distinct and positive are the roots of

$$d_0x^4 + d_1\left(-\beta_l^2 x + x^3\right) + d_2x^2 + d_3 = 0. \tag{A4}$$

Note that $d_0 \neq 0$ if the β_i are to be distinct. Using the representation of the coefficients of a polynomial in terms of the roots, since $d_0 \neq 0$,

$$\beta_l^2 \sum_{i=1}^{4} \beta_{-i} = -\sum_{i_1<i_2<i_3} \beta_{-i_1}\beta_{-i_2}\beta_{-i_3}, \tag{A5}$$

$$i_1, i_2, i_3 \in \{1, 2, 3, 4\}.$$

However, this does not hold since β_{-i}, β_l are assumed positive.

Lemma 2: Let $(e_1, e_2, e_3, e_4)^t$ be an eigenvector of A corresponding to the eigenvalue e, then $(e_4, e_3, e_2, e_1)^t$ is the eigenvector of A^t corresponding to e.

Proof: Since $(e_1, e_2, e_3, e_4)^t$ is an eigenvector of A corresponding to e,

$$\sum_{j=1}^{4} A_{ij}e_j = ee_i, \quad i = 1, 2, 3, 4. \tag{A6}$$

However, since $A_{ij} = A_{5-j,5-i}$, $i, j = 1, 2, 3, 4$,

$$\sum_{j=1}^{4} A_{5-j,5-i}e_j = ee_i. \tag{A7}$$

If we put $5 - j = l$, $5 - i = m$,

$$\sum_{l=1}^{4} A_{lm} e_{5-l} = e e_{5-m}, \quad m = 1, 2, 3, 4. \tag{A8}$$

Hence (e_4, e_3, e_2, e_1) is an eigenvector of A^t corresponding to e.

Lemma 3: If an eigenvalue e of A is not equal to 1 the corresponding eigenvector has $e_2 = e_3 \neq 0$, (hence e_2 can be normalized to 1) when β can assume more than one value, and l is distributed exponentially.

Proof: Let $e \neq 1$ be an eigenvalue of A and $(e_1, e_2, e_3, e_4)^t$ an eigenvector corresponding to e. Since $e \neq 1$, $(e_1, e_2, e_3, e_4)^t$ is orthogonal to $(0, -1, 1, 0)^t$ which is an eigenvector of A^t corresponding to the eigenvalue 1. Therefore $e_2 = e_3$. Suppose $e_2 = e_3 = 0$, then it can easily be shown from the equation $A\underline{e} = e\underline{e}$ that $e_2 = e_3 = 0$ iff

$$\alpha_2^2 \alpha_6 - \alpha_3^2 \alpha_4 = 0. \tag{A9}$$

However,

$$\alpha_2^2 \alpha_6 - \alpha_3^2 \alpha_4 = (\lambda^2 a/2)(\alpha_4 \alpha_6 - \alpha_5^2) \tag{A10}$$

from Eqs. (40)—(45). But, $\alpha_4 \alpha_6 > \alpha_5^2$ whenever β can assume more than one value from Schwartz Inequality.

Hence $e_2 = e_3 \neq 0$ whenever the distribution of β is not singular. So we can normalize the eigenvector (e_1, e_2, e_3, e_4) by dividing by e_2, hence the following lemma.

Lemma 4: $a_i b_j + b_i a_j = -2, \quad i \neq j.$

From Lemma 2, the eigenvector of A^t corresponding to λ_j is $(b_j, 1, 1, a_j)^t$. Hence the orthogonality of the eigenvectors of A and A^t gives

$$a_i b_j + b_i a_j = -2, \quad i \neq j. \tag{A11}$$

Lemma 5: $a_i b_j - b_i a_j \neq 0, \quad i \neq j.$

Let

$$a_1 b_2 - a_2 b_1 = \Delta_1, \tag{A12}$$

$$a_2 b_3 - a_3 b_2 = \Delta_2, \tag{A13}$$

$$a_3 b_1 - a_1 b_3 = \Delta_3. \tag{A14}$$

Then using Lemma 4, and putting $\Delta_i/2 = \delta_i$ we have

$$a_1 b_2 = (\delta_1 - 1), \quad a_2 b_3 = (\delta_2 - 1), \tag{A15}$$

$$a_2 b_1 = -(1 + \delta_1), \quad a_3 b_2 = -(1 + \delta_2), \tag{A16}$$

$$a_3 b_1 = (\delta_3 - 1), \quad a_1 b_3 = -(1 + \delta_3). \tag{A17}$$

Now

$$\begin{aligned} a_1 b_2 a_2 b_3 a_3 b_1 &= (\delta_1 - 1)(\delta_2 - 1)(\delta_3 - 1) \\ &= a_2 b_1 a_3 b_2 a_1 b_3 = -(1 + \delta_1)(1 + \delta_2)(1 + \delta_3), \end{aligned} \tag{A18}$$

i.e.,

$$\begin{aligned} &+ \delta_1 + \delta_2 + \delta_3 + \delta_1\delta_2\delta_3 - \delta_1\delta_2 - \delta_2\delta_3 - \delta_3\delta_1 - 1 \\ &= -1 - \delta_1 - \delta_2 - \delta_3 - \delta_1\delta_2 - \delta_2\delta_3 - \delta_3\delta_1 - \delta_1\delta_2\delta_3, \end{aligned} \tag{A19}$$

i.e.,

$$\delta_1 + \delta_2 + \delta_3 + \delta_1\delta_2\delta_3 = 0. \tag{A20}$$

However since the eigenvectors corresponding to λ_i, $i = 1, 2, 3$, are linearly independent,

$$\det \begin{bmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \neq 0, \tag{A21}$$

i.e., $(a_1 b_2 - a_2 b_1) - (a_1 b_3 - a_3 b_1) + (a_2 b_3 - a_3 b_2) \neq 0$, i.e.,

$$\delta_1 + \delta_2 + \delta_3 \neq 0. \tag{A22}$$

Hence each of the $\delta_i \neq 0$ from (A20).

[1]J. McKenna and J.A. Morrison, J. Math. Phys. 11, 2348—60 (1970).

[2]H.E. Rowe, IEEE Trans. Microwave Theory Tech. 19, 73—80 (1971).

[3]V.N. Tutukalin, Theory Probabilities Appl. 14, 547—66 (1969).

[4]G.C. Papanicolaou, SIAM J. Appl. Math. 21, No. 1 (1971).

[5]F.J. Dyson, Phys. Rev. 192, 1331—38 (1953).

[6]M.G. Krein, Dokl. Akad. Nauk SSSR 76, 21—4 (1951).

[7]I.M. Gel'fand and B.M. Levitan, Izv. Akad. Nauk SSSR, Ser. Mat. 15, 309—60 (1951) [Am. Math. Soc. Transl., Ser. 2 1, 253—304 (1955)].

[8]A.R. Hochstim and C.P. Martens, "Radar scattering from a plane parallel turbulent plasma slab with step function fluctuations of electron density," Institute for Defense Analysis Science and Technology Division Research paper, p. 316 (1967).

[9]B. Gopinath, "Determination of Average Dielectric Constant and Length of Slabs from Power Type Measurements," preprint No. Bell Laboratories TM 1211—23 (1973).

[10]M. Born and E. Wolf, *Principles of Optics* (New York, Pergamon, 1965), 3rd ed., p. 52.

[11]W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1966).

On the equivalence of nonrelativistic quantum mechanics based upon sharp and fuzzy measurements

S. Twareque Ali* and H. D. Doebner†

International Centre for Theoretical Physics, Trieste, Italy
(Received 20 January 1976)

Starting from the idea that physical measurements may have residual imprecisions, the possibility of replacing the nonrelativistic, three-dimensional configuration space by a so-called fuzzy configuration space, having an isomorphic Borel structure, is discussed. A quantization procedure with respect to such a space is developed, and the invariance of nonrelativistic quantum mechanics under such Borel isomorphisms is exploited to prove the equivalence of this quantization procedure to the usual quantization procedure on a fuzzy-free configuration space. Further, any Galilean invariant dynamics is shown to be insensitive to such imprecisions in the measurements of position and momentum.

I. INTRODUCTION

1. A recent point of accumulating interest, in the study of the axiomatic foundations of quantum mechanics, has been a consideration of fuzzy sets and fuzzy observables.[1-3] Briefly, the motivation for such a study is the following: It is a recognized fact that physical measurements are subject to errors and that in most cases such errors are not arbitrarily controllable. For instance, the measurement of an observable, such as the position of a particle, can only be made as precise as the applied instrument will allow. The residual imprecision can perhaps be improved by using better (i. e., more sensitive) instruments, but it cannot be eliminated altogether. However, most physical theories are built upon the assumption that an ideal measurement process exists for any observable quantity—viz. one in which no errors of observation are involved. In other words, the predictions of any theory correspond to the limit of infinitely sharp measurements.[4] To check a theory against experiment, one has therefore to rely upon some sort of statistical inference—something that is mostly quite outside the original scope of the theory itself. On the other hand, one may take the attitude that the necessary experimental imprecision involved in the measurement process is in itself a fact of physical reality, and hence ought to be incorporated into the formulation of any physical theory. If this second attitude is taken, a number of interesting questions immediately arise, both physical and mathematical.

2. To begin with, it becomes interesting to study properties of the configuration space in such theories. For example, since sharp measurements of position are never possible, the mathematical concept of a three-dimensional space $\mathbb{R}^3$ consisting of individual points loses its significance. Instead, points in space must now be replaced by probability distributions. That is to say, when one speaks of a particle as being located at the point $\mathbf{x}_0$ in $\mathbb{R}^3$, one ought to specify, along with $\mathbf{x}_0$, a distribution function $\mathbf{x} \mapsto f_{\mathbf{x}_0}(\mathbf{x})$, which would provide a measure for our confidence of actually finding the particle[5] also at some other point $\mathbf{x} \neq \mathbf{x}_0$. To be a "good" measurement, the distribution function $f_{\mathbf{x}_0}$ would have to be sharply peaked at $\mathbf{x}_0$.

Once one agrees to consider the aggregate of points $\mathbf{x} \in \mathbb{R}^3$, together with the associated distribution functions $f_{\mathbf{x}}$ as some sort of a "fuzzy configuration space," one ought to study the possible topological or Borel structures on them. One should then go further and try to build physical theories on such fuzzy configuration spaces, and then compare these resulting theories against the standard "fuzzy-free" theories. Next one could try to see if these new theories lead to any genuine enrichment of the older theories, in the sense of whether effects exist which find justifications only on the basis of the new theories, but *not* on the basis of fuzzy-free theories.

On the mathematical side, it further becomes interesting to study probability theory on fuzzy sets and fuzzy spaces. Since all measurements in quantum mechanics must be related to certain probability measures defined on the observed spectra of some self-adjoint operators on a Hilbert space, it is also necessary to develop a spectral theory, of these operators, based upon fuzzy spaces. One such attempt has been made in Ref. 2.

3. In this paper we study nonrelativistic quantum mechanics based upon fuzzy configuration spaces which are Borel isomorphic to $\mathbb{R}^3$. As a quantization method we use Mackey's imprimitivity theory[6] which is well suited in this case, because in that theory the Borel structure of $\mathbb{R}^3$ is used explicity and it has there a direct and simple physical meaning. We choose the Euclidean group $\mathcal{E}^3$ as the kinematical group, so that the position operators Q_j may be defined and from there the momentum operators P_j derived. The P_j's and the Q_j's then satisfy the usual commutation relations. We generalize this method to fuzzy configuration spaces in a natural manner, according to which instead of the usual projection valued measures on the Borel sets there now arise the more general positive operator valued measures (see also Ref. 3). It is then possible to construct fuzzy position operators $\widetilde{Q}_j$ and the conjugate momentum operators $\widetilde{P}_j$ in a manner completely analogous to that of the construction of the Q_j's and the P_j's. The remarkable result which emerges from our analysis is that Q_j, P_j and $\widetilde{Q}_j, \widetilde{P}_j$ turn out to be unitarily equivalent, that is, that there exists a unitary transformation V for which $VQ_jV^{-1} = \widetilde{Q}_j$ and $VP_jV^{-1} = \widetilde{P}_j$. Independently of this we also prove that the information contained in the fuzzy localization operators, constituting the positive operator valued measures, is the same as that

contained in the sharp localization operators given by the projection valued measures. Our result also shows that for a physical system which is invariant under Galilean transformations, the dynamics (i.e., the Hamiltonian) is unaltered if the assumption of sharp position measurements is replaced by that of fuzzy position measurements. While our result does not imply the futility of studying quantum mechanics on fuzzy configuration spaces, it does however mean that the usual assumption of ideally sharp measurements does not lead to any unphysical result. To that extent, of course, the introduction of fuzzy configuration spaces isomorphic to $\mathbb{R}^3$ in nonrelativistic quantum mechanics is superfluous. However, for relativistic systems the situation is quite different. A prototype of a relativistic result, obtainable only in a quantum mechanics based upon fuzzy measurements was discussed in Ref. 3. It was shown there, for example, how a localization operator for a photon could be constructed on the basis of such a theory, but not on the basis of a fuzzy-free theory. In other words, there do exist physical effects which are intrinsically "fuzzy-dependent," so that a theory based upon a fuzzy configuration space does in fact lead to an enrichment of the older theories.

4. The rest of this paper is organized as follows. Mackey's quantization method is sketched in Sec. II.1 in a manner which lends itself to a generalization to fuzzy configuration spaces. The physical argument as to why fuzzy spaces should be introduced and how they should be defined is discussed in Secs. II.2 and II.3. The definition of fuzzy configuration spaces and of the Borel structure on them is contained in Sec. II.4, while the consistency of the physical arguments with the mathematical formulas is analyzed in Sec. II.5. Section III gives in III.1 the properties of fuzzy configuration spaces under the action of the group $\mathcal{E}^3$, and the smoothness of $\mathcal{E}^3$-covariant fuzzy localization operators is proved in III.2 (Theorem 1). The relative amount of information contained in fuzzy and sharp localization operators is discussed in III.3 and it is shown in III.4 (Theorem 2) that both give the same information. This result points to the main result in Sec. IV, where in IV.2 the fuzzy position operators $\tilde{Q}_j$ are constructed from the fuzzy localization operators, analogously to the construction of the operators Q_j from the sharp localization operators (IV.1). The corresponding momentum operators $\tilde{P}_j$ and P_j are defined via the Fourier transform. In Sec. IV.3 it is proved that the $\tilde{P}_j$'s and $\tilde{Q}_j$'s are unitarily equivalent to the P_j's and Q_j's (Theorem 3). The result and its implications are discussed in Sec. V.

II. THE BOREL STRUCTURE AND FUZZY CONFIGURATION SAPCES

1. We first sketch Mackey's quantization method and the essential role played by the Borel structure imposed upon the physical configuration space $\mathbb{R}^3$. Let us consider a massive, nonrelativistic, spinless particle moving in $\mathbb{R}^3$. To quantize its kinematics, we recall that its momentum and angular momentum are given at the geometrical level by the three-dimensional Euclidean group $\mathcal{E}^3 = T^3 \circledS SO(3)$ acting on $\mathbb{R}^3$ as $[\mathbf{x}]g, \mathbf{x} \in \mathbb{R}^3$, $g \in \mathcal{E}^3$. To describe the position of the system, we introduce the Borel sets E in $\mathbb{R}^3$ as localization volumes and denote the family of all such sets by $\mathcal{B}(\mathbb{R}^3)$. The group $\mathcal{E}^3$ acts as $[E]g$ on E and gives a Borel automorphism of $\mathcal{B}(\mathbb{R}^3)$. Quantization of the system now means constructing a (continuous) unitary representation of $\mathcal{E}^3$, i.e., $g \in \mathcal{E}^3 \mapsto U_g$, and a projection valued measure $E \mapsto P(E)$ defined on the Borel sets $E \in \mathcal{B}(\mathbb{R}^3)$, on some Hilbert space $\mathcal{H}$, such that the covariance condition (imprimitivity relation)

$$U_g^* P(E) U_g = P([E]g) \tag{2.1}$$

is satisfied. This covariance condition follows from considerations of homogeneity of the physical space. A pair $(U_g, P(E))$ or $g \mapsto U_g$, $E \mapsto P(E)$, fulfilling the condition (2.1) is called a system of imprimitivity (SI). In our case, for a spinless particle moving in $\mathbb{R}^3$, the SI is unique up to an isometric isomorphism of Hilbert spaces. That is a result of the imprimitivity theorem of Mackey.[7] Physically, the generators of U are the momenta and angular momenta and the $P(E)$'s are the operators of localization in the regions E (see below). It is reasonable to take $\mathcal{H} = L^2(\mathbb{R}^3, d^3x)$ with $\mathbb{R}^3$ as the base space which is identified with the physical position space. Then the action of U_g is

$$(U_g\psi)(\mathbf{x}) = \psi([\mathbf{x}]g), \tag{2.2}$$

and $P(E)$ acts simply as

$$(P(E)\psi)(\mathbf{x}) = \chi_E(\mathbf{x})\, \psi(\mathbf{x}), \tag{2.3}$$

for all $\psi \in \mathcal{H}$. (χ_E is the characteristic function of the set E.) Being a normalized projection valued (PV) measure, $E \mapsto P(E)$ satisfies the following conditions:

(I) $P(\phi) = 0$, ϕ denoting the null set,

(II) $P(\mathbb{R}^3) = I$, I being the identity operator on $\mathcal{H}$,

(III) $P(\cup_i E_i) = \sum_i P(E_i)$, where $\{E_i\}$ is any countable family of mutually disjoint sets in $\mathcal{B}(\mathbb{R}^3)$. The sum here is assumed to converge weakly (hence strongly).

2. It was mentioned above that geometrically the $P(E)$'s are operators of localization (cf. Ref. 3, for example) in the regions $E \subset \mathbb{R}^3$, so that, given that the system is in the state $\psi \in \mathcal{H}$, the probability $p_\psi(E)$ of finding it in the region E is

$$p_\psi(E) = (\psi, P(E)\,\psi) = \int_E |\psi(\mathbf{x})|^2 d^3x. \tag{2.4}$$

Clearly, $E \mapsto p_\psi(E)$ is a probability measure on $\mathbb{R}^3$. It was argued in Ref. 3, from physical considerations based, for example, upon the impossibility of accurately determining the boundaries of the localization volumes $E \subset \mathbb{R}^3$, that the use of the $P(E)$'s to define the $p_\psi(E)$'s in (2.4) implies that sharp measurements of position are possible. On the other hand, since no physical instrument has an infinite accuracy, it is necessary, from an operational point of view, to generalize this approach. This was done in Ref. 3, where the PV measure in (2.7) was replaced by a normalized positive operator valued (POV) measure, $E \mapsto a(E)$. That is, for each $E \in \mathcal{B}(\mathbb{R}^3)$ the corresponding localization operator is taken to be $a(E)$, which is a positive operator on $\mathcal{H}$ satisfying:

(Ia) $a(\phi)=0$,

(IIa) $a(\mathbb{R}^3)=1$,

(IIIa) $a(\cup_i E_i)=\sum_i a(E_i)$, weakly, hence strongly, for any countable family $\{E_i\}$ of mutually disjoint sets in $\mathcal{B}(\mathbb{R}^3)$.

We now give another physical justification for the introduction of the POV measures, which will stress the Borel structure aspect of $\mathcal{B}(\mathbb{R}^3)$ (cf. also Ref. 2).

3. Let us suppose that, as a result of the finite precision of the available measuring apparatus, to each point $\mathbf{x}\in\mathbb{R}^3$ we must associate the confidence function $f_{\mathbf{x}}$, that is, for each point $\mathbf{x}$, $f_{\mathbf{x}}(\mathbf{x}')$ gives the probability density that when the particle is observed to be at $\mathbf{x}$, it could actually be at some other point $\mathbf{x}'\in\mathbb{R}^3$. More generally, we may associate a probability measure $\nu_{\mathbf{x}}$, defined on $\mathcal{B}(\mathbb{R}^3)$, with each $\mathbf{x}\in\mathbb{R}^3$. [In the above case, for example, $\nu_{\mathbf{x}}$ can be chosen to be the measure which has the density $f_{\mathbf{x}}$ with respect to the Lebesgue measure on $\mathbb{R}^3$: $\nu_{\mathbf{x}}(dx')=f_{\mathbf{x}}(\mathbf{x}')\,d^3x'$.] We shall assume that for each $E\in\mathcal{B}(\mathbb{R}^3)$ the function $\mathbf{x}\mapsto\nu_{\mathbf{x}}(E)$ is measurable (with respect to $\mathbf{x}$).

4. To connect these ideas with a mathematical property of localization operators, on configuration space, we recall that sharp localization operators are given by the projectors $P(E)$, which act in the manner [cf. Eq. (2.3)]

$$(P(E)\psi)(\mathbf{x})=\delta_{\mathbf{x}}(E)\,\psi(\mathbf{x}), \tag{2.5}$$

with $\delta_{\mathbf{x}}$ being the Dirac measure at $\mathbf{x}$. In the case of imprecise measurements we should have to replace $\delta_{\mathbf{x}}$ in (2.5) by the probability measure $\nu_{\mathbf{x}}$, referred to above. To formalize this idea we introduce a notation which points to the physical situation described above, in which which to every $\mathbf{x}\in\mathbb{R}^3$, considered as a measurable quantity, is associated the probability measure $\nu_{\mathbf{x}}$ on $\mathcal{B}(\mathbb{R}^3)$. Let us introduce the pairs $\tilde{x}=(\mathbf{x},\nu_{\mathbf{x}})$ and call the set

$$\{\tilde{x}\,|\,\mathbf{x}\in\mathbb{R}^3\}\equiv(\mathbb{R}^3,\nu)$$

a fuzzy configuration space. Clearly $(\mathbb{R}^3,\delta)$ is a sharp configuration space which is essentially $\mathbb{R}^3$ itself. The spaces $\mathbb{R}^3$, $(\mathbb{R}^3,\nu)$, and $(\mathbb{R}^3,\delta)$ are isomorphic, with

$$i:\mathbf{x}\in\mathbb{R}\mapsto i(\mathbf{x})=\tilde{x}=(\mathbf{x},\nu_{\mathbf{x}}). \tag{2.6}$$

A Borel structure can be induced on $(\mathbb{R}^3,\nu)$ from that on $\mathbb{R}^3$ by which a set $\tilde{E}\subset(\mathbb{R}^3,\nu)$ will be called Borel if and only if $i^{-1}(\tilde{E})$ is a Borel set in $\mathbb{R}^3$. Then both $\mathcal{B}(\mathbb{R}^3,\nu)$ and $\mathcal{B}(\mathbb{R}^3)$ are Borel isomorphic. Because the measures ν are essential for the physical interpretation of sets in the base space $\mathbb{R}^3$ of $L^2(\mathbb{R}^3,d^3x)$, it is reasonable to add this measure also in the notation for the Hilbert space. We shall therefore write $L^2((\mathbb{R}^3,\delta),d^3x)$ for $L^2(\mathbb{R}^3,d^3x)$, and $L^2(\mathbb{R}^3,\nu),d^3x)$ when we mean that the base space $\mathbb{R}^3$ has been changed to the fuzzy space $(\mathbb{R}^3,\nu)$. The Hilbert spaces $L^2((\mathbb{R}^3,\nu),d^3x)$ and $L^2(\mathbb{R}^3,d^3x)$ are clearly the same space, by virtue of the isomorphism in Eq. (2.6).

We are now prepared to define fuzzy (i.e., imprecise) localization operators. We write them as positive operators $a(E)$, $E\in\mathcal{B}(\mathbb{R}^3)$ such that

$$(a(E)\,\psi)(\mathbf{x})=\nu_{\mathbf{x}}(E)\,\psi(\mathbf{x}), \tag{2.7}$$

for all $\psi\in\mathcal{H}$. Let $E=i^{-1}(\tilde{E})$, for $\tilde{E}\in\mathcal{B}(\mathbb{R}^3,\nu)$. Then clearly, $\tilde{E}\mapsto a(E)$ defines a POV measure on $\mathcal{B}(\mathbb{R}^3,\nu)$. In Ref. 3 this was identified as a fuzzy localization operator. The introduction of the rather complicated notation involving the fuzzy configuration space is intended to stress the fact that as far as physically realizable localization operators are concerned, the localization volumes $E\in\mathcal{B}(\mathbb{R}^3)$ always appear in conjunction with a set of measures $\{\nu_{\mathbf{x}}\,|\,\mathbf{x}\in E\}$. For the discussion presented here, the fuzzy configuration space will turn out to be a useful and self-explanatory concept, and its definition already indicates how the POV measure $E\mapsto a(E)$ could be embedded in a PV measure $E\mapsto P'(E)$ in an enlarged Hilbert space $\mathcal{H}'\supset\mathcal{H}$ $=L^2((\mathbb{R}^3,\nu),d^3x)$ (cf. Ref. 3).

5. To check the consistency of our arguments, we now ask the following question. If the particle is in the state $\psi\in L^2((\mathbb{R}^3,\nu),d^3x)$ and is observed to be localized in the set $\tilde{E}\in\mathcal{B}(\mathbb{R}^3,\nu)$ with a probability $\tilde{p}_\psi(\tilde{E})$, then how is the measure $\tilde{p}_\psi$ related to the sharp probability measure p_ψ of Eq. (2.4)? To answer this question, let $\lambda(\mathbf{x})$ be the Radon–Nikodym derivative of p_ψ with respect to the Lebesgue measure d^3x on $\mathbb{R}^3$, at $\mathbf{x}\in\mathbb{R}^3$. Clearly $\lambda(\mathbf{x})=|\psi(\mathbf{x})|^2$ and it is the probability density of finding the particle at the sharp point $\mathbf{x}\in\mathbb{R}^3$. Now let $\tilde{\lambda}(\tilde{x})$ be the probability density of finding the particle at the fuzzy point $\tilde{x}$. Then it is easily seen (cf. Ref. 2) that we ought to have

$$\tilde{\lambda}(\tilde{x})=\int_{\mathbb{R}^3}\lambda(\mathbf{x}')\,\nu_{i^{-1}(\tilde{x})}(dx'), \tag{2.8}$$

and

$$\tilde{p}_\psi(\tilde{E})=\int_{\tilde{E}}\tilde{\lambda}(\tilde{x})\,\mu(d\tilde{x}), \tag{2.9}$$

μ being the image of the measure d^3x on $\mathcal{B}(\mathbb{R}^3,\nu)$, obtained in the obvious way through the isomorphism (2.6). The justification for Eqs. (2.8) and (2.9) is that they provide us with a relationship between $\tilde{p}_\psi$ and p_ψ which is compatible with the condition[8] that $\tilde{p}_\psi$ reduce to p_ψ when $\nu_{\mathbf{x}}$ is replaced by the delta measure $\delta_{\mathbf{x}}$ at $\mathbf{x}$ (the limit of sharp position measurements). It is now straightforward to check that

$$\tilde{p}_\psi(\tilde{E})=(\psi,a(E)\psi), \tag{2.10}$$

where $E=i^{-1}(\tilde{E})$, and $\tilde{E}\mapsto a(E)$ is the same POV measure as defined in (2.7).

6. To sum up the discussion up to now, we have tailored certain Borel spaces to be isomorphic to $\mathbb{R}^3$ in such a manner that it is possible to introduce the notion of fuzzy localization operators. It turned out that they are POV measures on $\mathcal{B}(\mathbb{R}^3)$ with values in $L^2(\mathbb{R}^3,\nu),d^3x)$. The converse of this statement is also interesting to study, and we shall see in the next section to what extent an arbitrary normalized POV measure on $\mathcal{B}(\mathbb{R}^3)$, with values in $L^2(\mathbb{R}^3,d^3x)$ and under the influence of the symmetry group $\mathcal{E}^3$, determines a fuzzy configuration space and also how much information it gives, relative to the sharp localization operators $P(E)$.

III. THE INFORMATION CONTAINED IN FUZZY LOCALIZATION OPERATORS

1. To study the transformation properties of fuzzy localization operators $a(E)$ under $\mathcal{E}^3$, we transfer first the group action of $\mathcal{E}^3$ from $\mathbb{R}^3$ to $(\mathbb{R}^3, \nu)$ as $[\tilde{x}]g = ([\mathbf{x}]g, \nu_{[\mathbf{x}]g})$. Both spaces $\mathbb{R}^3$ and $(\mathbb{R}^3, \nu)$ are isomorphic as transitive Borel $\mathcal{E}^3$ spaces (cf. Ref. 9). The stability group of the origin $\mathbf{0}$ of $\mathbb{R}^3$ is SO(3). In analogy with the covariance condition (2.1) imposed upon $P(E)$, we demand that

$$U_g{}^* a(E) U_g = a([E]g). \tag{3.1}$$

We shall call $\tilde{E} \mapsto a(E)$, $g \mapsto U_g$, satisfying (3.1) a POV system of imprimitivity (POVSI). By contrast, $E \mapsto P(E)$, $g \mapsto U_g$, satisfying (2.1) will be called a projective system of imprimitivity PVSI. Further, from Eq. (2.7) it is clear that for almost all $\mathbf{x} \in \mathbb{R}^3$ (with respect to the Lebesgue measure),

$$\nu_{[\mathbf{x}]g^{-1}}(E) = \nu_{\mathbf{x}}([E]g), \tag{3.2}$$

and hence, that for all $h \in$ SO(3)

$$\nu_{\mathbf{0}}([E]h) = \nu_{\mathbf{0}}(E), \tag{3.3}$$

for all $E \in \mathcal{B}(\mathbb{R}^3)$, $\mathbf{0}$ being the origin of $\mathbb{R}^3$. (Thus, for example, when $\nu_{\mathbf{x}}(dx') = f_{\mathbf{x}}(\mathbf{x}')\, d^3x'$, $f_{\mathbf{x}}$ is only a function of the norm $|\mathbf{x} - \mathbf{x}'|$.) Also, since $\tilde{x}$ and $\mathbf{x}$ have the same stability subgroups, if (3.2) holds for any $h \in \mathcal{E}^3$ then $h \in$ SO(3).

2. We now state a few properties of POVSI's on $\mathbb{R}^3$, which we shall also try to interpret physically. They will further clarify the relationship between fuzzy localization operators and the Borel structure of $\mathbb{R}^3$. Proofs of mathematical results have been deferred to the Appendix.

Theorem 1: Let $E \mapsto a(E)$, $g \mapsto U_g$ be a POVSI on $\mathbb{R}^3$ for the group $\mathcal{E}^3$. Then $a(E) = 0$ if and only if $E \in \mathcal{B}(\mathbb{R}^3)$ has Lebesgue measure zero.

An observable which possesses the property just mentioned, i.e., $a(E) = 0$ if and only if $\mu(E) = 0$ for some Borel measure μ, is sometimes referred to as a smooth observable.[10] Theorem 1 thus says that all covariant [i.e., satisfying Eq. (3.1)] localization operators are smooth with respect to the Lebesgue measure, and it follows that localization probabilities are concentrated on the same sets for both sharp and fuzzy measurements.

3. To state our next result, we need some notion of comparison between different sets of fuzzy localization operators. Let $E \mapsto a_1(E)$ and $E \mapsto a_2(E)$ be two sets of fuzzy localization operators on $\mathcal{B}(\mathbb{R}^3)$. We shall say that a_1 *gives more information than* a_2, and write $a_1 \supset a_2$, if for any two vectors $\phi, \psi \in \mathcal{H}$ the equality $(\psi, a_1(E)\psi) = (\phi, a_1(E)\phi)$ implies $(\psi, a_2(E)\psi) = (\phi, a_2(E)\phi)$ (cf. also Ref. 11). If both $a_1 \supset a_2$ and $a_2 \supset a_1$ hold, we shall say that a_1 and a_2 *give the same information.* Physically, if $a_1 \supset a_2$, then using a_2 alone we cannot distinguish between states which are indistinguishable using a_1 alone. For any fuzzy localization a, let $\mathcal{A}(a)$ be the von Neumann algebra generated by the operators $a(E)$, $E \in \mathcal{B}(\mathbb{R}^3)$. If then for any two fuzzy localizations a_1 and a_2 we have $a_1 \supset a_2$, it follows that $\mathcal{A}(a_1) \supset \mathcal{A}(a_2)$.

Finally, let $C(\mathbb{R}^3)$ denote the set of all complex valued continuous functions on $\mathbb{R}^3$ and $K(\mathbb{R}^3)$ the subset of those functions in $C(\mathbb{R}^3)$ which have compact supports. Let $L^\infty(\mathbb{R}^3)$ be the *-algebra (with respect to the "essential sup norm") of all equivalence classes of bounded Lebesgue measurable functions on $\mathbb{R}^3$. Then, for the PVSI $E \mapsto P(E)$, $g \mapsto U_g$ of Eqs. (2.1) and (2.3) it is well known that $\mathcal{A}(P)$ is isometrically isomorphic to $L^\infty(\mathbb{R}^3)$. Let j denote this isometry, so that $j[\mathcal{A}(P)] = L^\infty(\mathbb{R}^3)$. For any $f \in K(\mathbb{R}^3)$ let $P(f) = \int_{\mathbb{R}^3} f(\mathbf{x})\, P(dx)$ and $a(f) = \int_{\mathbb{R}^3} f(\mathbf{x}) \quad (dx)$. Then quite trivially $j[P(f)] \in C(\mathbb{R}^3)$ (in fact it is the equivalence class of f itself). Also, if $P \supset a$ then $j[\mathcal{A}(a)] \subset L^\infty(\mathbb{R}^3.)$

4. The main result of this section is now stated.

Theorem 2: Let $E \mapsto P(E)$, $g \mapsto U_g$ be a PVSI for $\mathcal{E}^3$ on $L^2(\mathbb{R}^3, d^3x)$. Take the POVSI $E \mapsto a(E)$, $g \mapsto U_g$ determined by a fuzzy configuration space $(\mathbb{R}^3, \nu)$ which is Borel isomorphic to $\mathbb{R}^3$ as an $\mathcal{E}^3$ space. Then the fuzzy localization operators $a(E)$ satisfy

(i) a gives the same information as P,

(ii) $j[a(f)] \in C(\mathbb{R}^3)$, for all $f \in K(\mathbb{R}^3)$.

Conversely, any POVSI on $L^2(\mathbb{R}^3, d^3x)$ for the group $\mathcal{E}^3$, satisfying (i) and (ii) determines a fuzzy configuration space $(\mathbb{R}^3, \nu)$ which is Borel isomorphic to $\mathbb{R}^3$ as an $\mathcal{E}^3$ space. Further, given any POVSI $E \mapsto a(E)$, $g \mapsto U_g$, satisfying (i), there exists a sequence $\{E \mapsto a^{(n)}(E),\ g \mapsto U_g\}$ of POVSI's satisfying (i) and (ii) such that $a^{(n)}(E) \mapsto a(E)$ weakly, for all $E \in \mathcal{B}(\mathbb{R}^3)$.

This theorem tells us exactly how general a fuzzy localization $E \mapsto a(E)$ can be when it arises from a replacement of $\mathbb{R}^3$ by a Borel isomorphic fuzzy space $(\mathbb{R}^3, \nu)$. The result is that the generalization from $E \mapsto P(E)$ to $E \mapsto a(E)$ is restricted to the class of localizations $\{a\}$ which give the same information as P itself. In other words, the amount of physical information contained in the localization operators $a(E)$ is only a function of the Borel structure of the space $\mathbb{R}^3$ and not the possible individual realizations of this structure. This tends to indicate already that the result of the quantization procedure itself, i.e., the quantum mechanics of the system, should not change if P were to be replaced by a, since no information would really be gained or lost in the process. This we shall prove explicitly in the following section (cf. in particular, Theorem 3).

IV. FUZZY AND SHARP POSITION OPERATORS, MOMENTA, AND UNITARY EQUIVALENCE

1. The preceding discussion naturally raises the question as to whether canonical operators $\tilde{Q}_j$ and $\tilde{P}_j$, of position and momentum, respectively, can be defined with respect to the fuzzy configuration space $(\mathbb{R}^3, \nu)$ in a manner similar to that of the definitions of the usual position and momentum operators Q_j and P_j on the sharp configuration space $(\mathbb{R}^3, \delta)$.

For a construction of Q_j, P_j in $L^2((\mathbb{R}^3, \delta), d^3x)$, consider the PVSI $E \mapsto P(E)$, $g \mapsto U_g$ describing the system [Eqs. (2.1)–(2.3)].[12] Q_j is then given as an integral, with respect to the measure P of the component x_j of $\mathbf{x} \in \mathbb{R}^3$, i.e.,

$$Q_j = \int_{\mathbb{R}^3} x_j \, P(dx), \quad j=1,2,3, \tag{4.1}$$

on a dense domain $\mathcal{D}(Q_j) \subset \mathcal{H}$, on which it acts as

$$(Q_j\psi)(\mathbf{x}) = \delta_{\mathbf{x}}(x_j)\psi(\mathbf{x}). \tag{4.2}$$

Here $\delta_{\mathbf{x}}(x_j)$ is the average value of x_j with respect to the Dirac measure $\delta_{\mathbf{x}}$,

$$\delta_{\mathbf{x}}(x_j) = \int_{\mathbb{R}^3} x_j \, \delta_{\mathbf{x}}(dx'). \tag{4.3}$$

The matrix elements of Q_j are

$$(\psi, Q_j\psi) = \int_{\mathbb{R}^3} x_j \, p_\psi(dx), \tag{4.4}$$

with p_ψ being the measure defined in (2.4). The momenta P_j are the generators of the translation subgroup of T^3 of $\mathcal{E}^3$, derived as the differentials of the unitary representation $U(\mathcal{E}^3)$ of $\mathcal{E}^3$ on $\mathcal{H}$. In the special case at hand, where $\mathcal{E}^3$ acts transitively on $\mathbb{R}^3$ an alternative method can be used to construct P_j. For this let $\mathcal{F}$ denote the Fourier transform operator on $L^2(\mathbb{R}^3, d^3x)$,

$$(\mathcal{F}\psi)(\mathbf{x}) = (1/(2\pi)^{3/2}) \int_{\mathbb{R}^3} \exp(-i\mathbf{k}\cdot\mathbf{x}) \, \psi(\mathbf{k}) \, d^3k, \tag{4.5}$$

for all $\psi \in L^2(\mathbb{R}^3, d^3x)$. Then we may set[13] P_j to be the negative Fourier transform of Q_j, i.e.,

$$P_j = -\mathcal{F} Q_j \mathcal{F}^{-1}, \tag{4.6}$$

on a dense set $\mathcal{D}(P_j) \subset L^2(\mathbb{R}^3, d^3x)$. A closer inspection shows that the P_j's and the Q_j's are essentially self-adjoint operators on a common dense domain $\mathcal{D}(H) \subset \mathcal{H}$, fulfilling the (canonical) commutation relations of the three-dimensional Heisenberg algebra H^3: $[Q_j, P_k] = i\delta_{jk} I$, $[Q_j, Q_k] = 0 = [P_j, P_k]$. Further, the skew-adjoint representation of H^3 given by P_j, Q_j is integrable.

2. To extend the above method to $(\mathbb{R}^3, \nu)$ we ought to start with the POVSI $\tilde{E} \in \mathcal{B}(\mathbb{R}^3, \nu) \mapsto a(E)$, $g \mapsto U_g$. The discussion in Sec. III.4 then forces us to define a *fuzzy position operator* $\tilde{Q}_j$ on $L^2((\mathbb{R}^3, \nu), d^3x)$ as

$$\tilde{Q}_j = \int_{\mathbb{R}^3} x_j \, a(dx) \tag{4.7}$$

on a domain $\mathcal{D}(\tilde{Q}_j)$ spanned by those vectors $\psi \in \mathcal{H}$ for which the right-hand side of the equation

$$(\psi, \tilde{Q}_j\psi) = \int_{\mathbb{R}^3} x_j \, \tilde{p}_\psi(dx) \tag{4.8}$$

is finite, $\tilde{p}_\psi$ being the measure defined as $\tilde{p}_\psi(E) = (\psi, a(E)\psi)$, $E \in \mathcal{B}(\mathbb{R}^3)$. $\tilde{Q}_j$ acts on a vector $\psi \in \mathcal{D}(\tilde{Q}_j)$ in the manner

$$(\tilde{Q}_j\psi)(\mathbf{x}) = \nu_{\mathbf{x}}(x_j)\psi(\mathbf{x}), \tag{4.9}$$

with

$$\nu_{\mathbf{x}}(x_j) = \int_{\mathbb{R}^3} x_j' \nu_{\mathbf{x}}(dx') = x_j + \nu_0(x_j). \tag{4.10}$$

Since $\nu_0(x_j)$ is a fixed constant, whenever it is finite, we shall put

$$\nu_0(x_j) = c_j, \tag{4.11}$$

so that

$$(\tilde{Q}_j\psi)(\mathbf{x}) = ([Q_j + c_j I]\psi)(\mathbf{x}). \tag{4.12}$$

To ensure the existence of $\tilde{Q}_j$ on a dense domain $\mathcal{D}(\tilde{Q}_j)$, one must restrict the measures $\nu_{\mathbf{x}}$ such that they correspond to physical averaging procedures. We shall call those $\nu_{\mathbf{x}}$'s *physically admissible* for which c_j is finite and consequently $\tilde{Q}_j$ exists on a dense domain $\mathcal{D}(\tilde{Q}_j)$ for $j=1,2$ or 3. For such $\nu_{\mathbf{x}}$'s

$$\tilde{Q}_j = Q_j + c_j I \tag{4.13}$$

on $\mathcal{D}(Q_j)$.

Physically, the finiteness of c_j is easily justified. Consider the quantity

$$\Delta_j^2 = \int_{\mathbb{R}^3} x_j^2 \, \nu_0(dx).$$

Clearly, Δ_j is the dispersion of the error distribution for the observed j component of the localization point of the particle under study. To have a physically meaningful apparatus, Δ_j should be small, or at least finite. This implies that $\nu_0(x_j)$ should also be finite.

3. Now to define a momentum operator $\tilde{P}_j$ on $L^2((\mathbb{R}^3, \nu)\, d^3x)$ we use, as above, the Fourier transform of $\tilde{Q}_j$ and write, on a domain $\mathcal{D}(\tilde{P}_j)$,

$$\mathcal{F}\tilde{Q}_j\mathcal{F}^{-1} = -\tilde{P}_j. \tag{4.14}$$

From this it follows, using Eq. (4.13), that

$$\tilde{P}_j = P_j - c_j I, \tag{4.15}$$

which holds on $\mathcal{D}(P_j)$ for physically admissible $\nu_{\mathbf{x}}$'s. Because of the properties of $\tilde{P}_j$, $\tilde{Q}_j$, and I, we immediately find that for physically admissible $\nu_{\mathbf{x}}$'s the fuzzy position operators $\tilde{Q}_j$ and the corresponding momentum operators $\tilde{P}_j$ span a skew-adjoint integrable representation of the Heisenberg algebra H^3 on $L^2((\mathbb{R}^3\nu), d^3x)$, exactly as the Q_j's and P_j's do on $L^2((\mathbb{R}^3, \delta), d^3x)$. Both representations are irreducible. Then, according to the well-known theorem of von Neumann, $\tilde{P}_j, \tilde{Q}_j$ and P_j, Q_j are unitarily equivalent. In fact, there exists a unitary operator V in $L^2(\mathbb{R}^3, d^3x)$, viz.

$$V = \exp(-i\mathbf{c}\cdot\mathbf{P}\exp(i\mathbf{c}\cdot\mathbf{Q}), \tag{4.16}$$

where $\mathbf{c}$ is the 3-vector (c_1, c_2, c_3) such that the relations

$$VQ_jV^{-1} = \tilde{Q}_j, \quad VP_jV^{-1} = \tilde{P}_j \tag{4.17}$$

hold. Thus the two descriptions based upon sharp and fuzzy base spaces are physically equivalent.

We formulate this result as:

Theorem 3: Let $\nu_{\mathbf{x}}$ be physically admissible, let the position and momentum operators Q_j, P_j on $L^2((\mathbb{R}^3, \delta), d^3x)$ and $\tilde{Q}_j, \tilde{P}_j$ on $L^2((\mathbb{R}^3, \nu), d^3x)$ be such that Q_j and $\tilde{Q}_j$ are constructed via $E \in \mathcal{B}(\mathbb{R}^3, \delta) \mapsto P(E)$ and $\tilde{E} \in \mathcal{B}(\mathbb{R}^3, \nu) \mapsto a(E)$, respectively, and $P_j, \tilde{P}_j$ through their Fourier transforms. Then Q_j, P_j and $\tilde{Q}_j, \tilde{P}_j$ are unitarily equivalent.

Finally, since from the nature of Eqs. (4.13) and (4.15) it follows that the various systems of imprimitivity discussed above, which all arise from the transitive action of $\mathcal{E}^3$ on $\mathbb{R}^3$, can be derived canonically from unitary irreducible representations of the Galilei group, we have the other result that a Hamiltonian on $L^2((\mathbb{R}^3, \delta), d^3x)$, which is covariant with respect to Galilean transformations, leads to an equivalent dynamics on $L^2((\mathbb{R}^3, \nu), d^3x)$, where a fuzzy base space is used. In other words, the Hamiltonians formulated in terms of the Q_j, P_j and $\tilde{Q}_j, \tilde{P}_j$ are unitarily equivalent.

V. CONCLUSIONS

1. The results of the preceding sections may be summed up in the following way. Given a (spinless) free, massive particle, a quantization procedure consists in assigning to it a Hilbert space $\mathcal{H}$ of possible states, and the set of operators Q_j, P_j obeying the canonical commutation relations. To interpret these latter quantities as operators of position and momentum, respectively, it is necessary to find a realization of $\mathcal{H}$ as $L^2(\mathbb{R}^3, d^3x)$. In that case the desired representation of the Q_j's and the P_j's may be obtained in a natural manner by using the localization operators $P(E)$. However, we still have the freedom to replace $\mathbb{R}^3$ by any other space which is Borel isomorphic to it, and yet obtain a unitarily equivalent representation for the commutation relations. But each individual realization of these Borel isomorphisms need not have a physical interpretation. We have shown above that the class of isomorphisms which correspond to the replacement of $\mathbb{R}^3$ by fuzzy configuration spaces $(\mathbb{R}^3, \nu)$ can be interpreted physically in terms of imprecise measurements. Hence, conversely, the problem of imprecise measurements of position and momentum is subsumed naturally in the invariance of the quantization procedure under Borel isomorphisms. It would of course be interesting to find possible physical interpretations for other Borel isomorphisms of $\mathbb{R}^3$—i. e., those which do not lead to $(\mathbb{R}^3, \nu)$. As a further remark in this direction, we should note that it is really the manner in which the transformation property of $\mathbb{R}^3$ under $\mathcal{E}^3$ is transported, as it were, to be measures $\nu_{\mathbf{x}}$:

$$\nu_{[\mathbf{x}]g^{-1}}(E) = \nu_{\mathbf{x}}([E]g),$$

which is responsible for the rather strong result that the quantization procedure remains invariant.

2. It is worthwhile emphasizing here that we have proved the invariance of the dynamics on fuzzy configuration spaces only in a nonrelativistic context. Since the amount of fuzziness inherent in a position measurement is clearly not a Lorentz invariant quantity, we do not expect similar results in the relativistic case as well. On the contrary, as mentioned in the introduction, considerations of imprecise localizations do in fact lead to interesting and new results for relativistic systems. We hasten to add that our results on fuzzy-observables are only restricted to considerations of the position and momentum observables and all other observables which are functions of these two observables. One could perhaps also introduce the concept of fuzzy time in a related manner, but it is not our intention to suggest that here. Further it does not seem to us very useful at this point to try to apply the same considerations to observables having discrete spectra.

3. From our analysis a curious difference between a quantum system and its classical counterpart seems to show up. A classical dynamical system moves in general on a differentiable manifold, and its description depends very much upon the local structure of this manifold. By contrast, its quantization[14] depends only upon the Borel structure generated by the manifold. Since different manifolds could, in general, lead to the same Borel structure, in the quantization procedure some detailed structural information seems to be ignored. In other words, quantum mechanics depends in some sense only upon the global properties of the manifold. Perhaps this could be the deeper reason why the quantization of gravity is inherently so difficult. One tries there to quantize the very local structure of space—time to which the usual quantization procedure is insensitive.

4. As a final comment we mention an associated, interesting mathematical problem. Let $E \mapsto a(E)$, $g \mapsto U_g$ be a POVSI determined by a fuzzy configuration space $(\mathbb{R}^3, \nu)$. It is easily verified that one can write, for any two $\phi, \psi \in L^3(\mathbb{R}^3, d^3x)$, a relation of the sort

$$(\phi,\, a(E)\psi) = \int (\phi, P_{\mathbf{x}}(E)\psi)\, \nu_0(dx), \tag{5.1}$$

where $P_{\mathbf{x}}$ is the PV measure on $\mathcal{B}(\mathbb{R}^3)$:

$$P_{\mathbf{x}}(E) = U_{\mathbf{x}} P(E)\, U_{\mathbf{x}}^*, \tag{5.2}$$

and $E \mapsto P(E)$, $g \mapsto U_g$ is the canonical PVSI of Eqs. (2.1)–(2.3). We further have $U_g^* \mathcal{A}(P)\, U_g = \mathcal{A}(P)$, for all $g \in \mathcal{E}^3$ [$\mathcal{A}(P)$ is the von Neumann algebra generated by P] and $\mathcal{A}(a) \subset \mathcal{A}(P)$. Thus the measure ν_0, which now "represents" a may be thought of as being a measure defined on the compact convex set[15] of all regular normalized POV measures on $\mathcal{B}(\mathbb{R}^3)$ with ranges lying in $\mathcal{A}(P)$. In particular, since ν_0 has suport on the PV measures $P_{\mathbf{x}}$, it is "carried" by the extreme points σ_I of this convex set. We may therefore write $a(E)$ as

$$(\phi,\, a(E)\psi) = \int_{\sigma_I} (\phi, P(E)\psi)\, \nu(dP), \tag{5.3}$$

in terms of a probability measure defined on σ_I.

The question arising now is, starting from an arbitrary locally compact space X, a separable Hilbert space $\mathcal{H}$, and a normalized POV measure defined on $\mathcal{B}(X)$, whether it is possible to represent a as an integral over PV measures as in (5.3). Further, if a is a POVSI for the action of some group G, how does this affect ν? It has been shown in Ref. 15 that as far as the first question is concerned, it is always possible to find such a representation for a if $\mathcal{A}(a)$ is commutative. The representing measure ν is also unique, in this case (actually whenever it exists). Physically this result implies the possibility of writing any fuzzy localization as some sort of a probability average over sharp ones.

APPENDIX

1. Proof of Theorem 1 .

By Ref. 16, for any $\psi \in \mathcal{H}$, we have

$$\int_{\mathbb{R}^3} \chi_E(\mathbf{x})\, d^3x = \frac{1}{p_\psi(\mathbb{R}^3)} \int_{\mathbb{R}^3} p_\psi([E^{-1}]\mathbf{x})\, d^3x, \tag{A1}$$

where p_ψ is the postive measure $p_\psi(E) = (\psi, a(E)\psi)$ and E^{-1} is the inverse of the set E, $\mathbb{R}^3$ being considered as the subgroup T^3 of $\mathcal{E}^3$, and $[E^{-1}]\mathbf{x}$ is its translate through $\mathbf{x} \in \mathbb{R}^3$, considered as an element in $\mathcal{E}^3$. Thus, E has Lebesgue measure zero $\Rightarrow p_\psi([E^{-1}]\mathbf{x}) = 0$ for almost all $\mathbf{x}$. However, the null set on which $p_\psi([E^{-1}])\mathbf{x})$ may fail to be zero could presumably depend on ψ. To show that it may indeed be chosen to be independent of ψ, let $\{\phi_j\}$, $j=1,2,3,\ldots$, be a complete orthonormal set of vectors in $\mathcal{H}$ and let $\mathcal{N}_j$ be the null set on which $p_{\phi_j}([E^{-1}]\mathbf{x})$ is not zero. Then $\mathcal{N}=\cup_j\mathcal{N}_j$ being a countable union of null sets is itself a null set, and if follows that $p_\psi([E^{-1}]\mathbf{x})\neq 0$ if and only if $\mathbf{x}\in\mathcal{N}$. Next we note that E has Lebesgue measure zero iff E^{-1} has Lebesgue measure zero, so that

$$\int_{\mathbb{R}^3}\chi_E(\mathbf{x}')\,d^3x'=0\Rightarrow(U_\mathbf{x}\psi,\,a(E)U_\mathbf{x}\psi)=0,$$

for almost all $\mathbf{x}$ and all $\psi\in\mathcal{H}$. Hence E has Lebesgue measure zero $\Rightarrow a(E)=0$.

Conversely, by (A1), if for any $\psi\in\mathcal{H}$, $p_\psi([E^{-1}]\mathbf{x})=0$ for almost all $\mathbf{x}$, then E has Lebesgue measure zero. From this it follows that $p_\psi(E^{-1})=0$, for all $\psi\in\mathcal{H}\Rightarrow E$ has Lebesgue measure zero, i.e., $a(E)=0\Rightarrow E$ has Lebesgue measure zero. ∎

2. Proof of Theorem 2

Let $\phi,\psi\in L^2(\mathbb{R}^3,d^3x)$ be such that $(\phi,a(E)\phi)=(\psi,a(E)\psi)$ for all $E\in\mathcal{B}(\mathbb{R}^3)$. Using (2.7) and (3.2) it is easy to see that this implies the equality

$$|\phi(\hat{\mathbf{x}})|^2\,\hat{\nu}_0(\mathbf{x})=|\psi(\hat{\mathbf{x}})|^2\,\hat{\nu}_0(\mathbf{x}),\tag{A2}$$

for almost all $\mathbf{x}\in\mathbb{R}^3$, where $\mathbf{x}\mapsto|\phi(\hat{\mathbf{x}})|^2$, $\mathbf{x}\mapsto|\psi(\hat{\mathbf{x}})|^2$, and $\mathbf{x}\mapsto\hat{\nu}_0(\mathbf{x})$ are, respectively, the Fourier transforms of the functions $\mathbf{x}\mapsto|\phi(x)|^2$, $\mathbf{x}\mapsto|\psi(x)|^2$, and of the measure ν_0. Further, since $\nu_\mathbf{x}=\nu_{\mathbf{x}'}$ iff $\mathbf{x}=\mathbf{x}'$ (recall that $\mathbf{x}$ and $\nu_\mathbf{x}$ have the same stability subgroups, it follows that

$$\hat{\nu}_0(\mathbf{x}+\mathbf{x}')=\hat{\nu}_0(\mathbf{x})\quad\text{iff }\mathbf{x}=0.\tag{A3}$$

Thus $\hat{\nu}_0$ has support on the whole of $\mathbb{R}^3$, which fact together with (A2) implies that

$$|\phi(\mathbf{x})|^2=|\psi(\mathbf{x})|^2$$

for almost all $\mathbf{x}\in\mathbb{R}^3$. Thus $(\phi,a(E)\phi)=(\psi,a(E)\psi)\Rightarrow(\phi,P(E)\phi)=(\psi,P(E)\psi)$ for all $E\in\mathcal{B}(\mathbb{R}^3)$. The implication in the other direction is trivial. Hence (i) follows.

To prove (ii) let $f\in K(\mathbb{R}^3)$. Then it is straightforward to verify that

$$\begin{aligned}(j[a(f)])(\mathbf{x})&=\int_{\mathbb{R}^3}f(\mathbf{x}')\,\nu_\mathbf{x}(dx')\\&=\int_{\mathbb{R}^3}f(\mathbf{x}'-\mathbf{x})\,\nu_0(dx'),\end{aligned}\tag{A4}$$

in virtue of Eq. (3.2). But the quantity on the right-hand side of (A4) defines a continuous function of $\mathbf{x}$ (cf. Ref. 17, Chap. XIV, Sec. 9).

Conversely, if (i) and (ii) are satisfied then for any $\psi\in L^2(\mathbb{R}^3,d^3x)$ and $f\in K(\mathbb{R}^3)$ [note (i) $\Rightarrow\mathcal{A}(a)=\mathcal{A}(P)$],

$$(a(f)\psi)(\mathbf{x})=F_f(\mathbf{x})\,\psi(\mathbf{x}),\tag{A5}$$

for some bounded continuous function $\mathbf{x}\mapsto F_f(\mathbf{x})$. Further, from the linearity and positivity of $a(f)$ in f, it follows that for each $\mathbf{x}\in\mathbb{R}^3$, $f\mapsto F_f(\mathbf{x})$ is a bounded positive linear form on $K(X)$. Hence, there exists a measure $\nu_\mathbf{x}$ on $\mathcal{B}(\mathbb{R}^3)$ such that

$$\nu_\mathbf{x}(f)=F_f(\mathbf{x}).\tag{A6}$$

From the covariance of a under U_g it follows that $\nu_\mathbf{x}$ satisfies Eq. (3.2). The construction of the space $(\mathbb{R}^3,\nu)$ is now obvious.

Finally if $E\mapsto a(E)$, $g\mapsto U_g$ is a POVSI satisfying (i) only, the function $\mathbf{x}\mapsto F_f(\mathbf{x})$ in (A5) is only a bounded measurable function and is not necessarily defined for all $\mathbf{x}$ in $\mathbb{R}^3$. But by Ref. 17 (Chap. XIV, Sec. 11) there exists a sequence of bounded continuous functions $F_f^{(n)}$ of $\mathbf{x}$ which converges to F_f in the manner

$$F_f^{(n)}(\mathbf{x})\to F_f(\mathbf{x})$$

for almost all $\mathbf{x}$. Further $F_f^{(n)}(\mathbf{x})$ can be chosen to be linear in f. Thus there exists a sequence of measures $\nu_\mathbf{x}^{(n)}$ and hence a sequence of POVSI's satisfying (i) and (ii) which converges weakly to the given POVSI. ∎

ACKNOWLEDGMENTS

The authors would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. One of them (S. T. A.) would like to thank Professor E. Prugovečki for many useful discussions and critical comments.

*Present address: Department of Mathematics, University of Toronto, Toronto, Canada M5S, 1A1.

†Permanent address: Institut für Theoretische Physik, Technische Universität Clausthal, Clausthal-Zellerfeld, Fed. Rep. Germany.

[1]E. Prugovečki, Found. Phys. 6, 557 (1975), and references cited therein.

[2]E. Prugovečki, J. Math. Phys. 17, 517 (1976).

[3]S. Twareque Ali and Gérard G. Emch, J. Math. Phys. 15, 176 (1974).

[4]Note that even theories, like quantum mechanics, which predict that sharp and *simultaneous* measurements of certain observables are not possible, also assume that sharp measurements of these observables, treated *singly*, are possible.

[5]For simplicity, we assume the particle to be pointlike.

[6]See, for example, G.W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

[7]G.W. Mackey, Proc. Natl. Acad. Sci. USA 35, 537 (1949).

[8]In fact, Eqs. (2.8) and (2.9) provide us with the only possible linear functional relationship between $\tilde{p}_\psi$ and p_ψ subject to this stipulation (see also Ref. 2).

[9]V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand Reinhold, New York, 1970), Vol. II.

[10]S. Gudder and C. Piron, J. Math. Phys. 12, 1583 (1971).

[11]E. B. Davies, J. Funct. Anal. 6, 318 (1970).

[12]Note that for a system moving on a homogeneous space a definition of a (pointlike) position operator Q_j is not possible in general (cf. Ref. 14).

[13]See, for instance, N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1966), Vol. I, Sec. 37.

[14]H.D. Doebner and J. Tolar, J. Math. Phys. 16, 975 (1975).

[15]S. Twareque Ali, "On some geometrical properties of fuzzy observables", ICTP, Trieste, preprint IC/75/54.

[16]G.W. Mackey, Duke Math. J. 16, 313 (1949), Lemma 3.3.

[17]J. Dieudonné, *Treastise on Analysis* (Academic, New York and London, 1970), Vol. II.

Remarks on conformal space

F. Bayen and M. Flato

Physique Mathématique, Collège de France, Paris 5, France
(Received 26 September 1975)

Some remarks about a recent article dealing with Maxwell equations written on conformal space M_c^4 are presented. The relevance of the manifold M_c^4 in conformal physics, the cotransformation of fields under the conformal group, and the presence of external currents in Maxwell equations are discussed. A simple proof of the conformal covariance of a linear gauge condition for Maxwell equations is then exhibited.

In a recent paper,[1] Mayer has studied the conformal covariance of Maxwell equations with source terms written on the compactified Minkowski space M_c^4. This was achieved after an investigation of the transformation properties under $SO_0(4,2)$ of vector and tensor fields on the manifold M_c^4. Besides, a conformally covariant linear gauge condition, which is compatible with Maxwell equations, was given in this article. This condition has been found independently by us.[2]

In the present note we shall make some remarks on the previously quoted article. We discuss critically the relevance of the manifold M_c^4 in conformal physics, the cotransformation of fields under the conformal group and the presence of external currents in Maxwell equations. We then show that to prove the conformal covariance of the gauge condition, one does not really need to utilize fibre bundle theory (though this theory is a nice mathematical instrument).

A. The idea to utilize manifolds such as M_4^c, i.e., roughly speaking Minkowski space M plus points at infinity, though very fashionable is not very new.[3] The usual argument is based on the singularities of the conformal transformations in M. The conformal group C is in fact defined as a pseudogroup of analytic diffeomorphisms in M. This situation, which is not necessarily unphysical, may lead to some difficulties if one tries to apply global conformal transformations in quantum field theory. However a formulation of the covariance under the conformal group in second-quantized theory, in which the fields are considered as operator-valued distributions, exists.[4] In order to carry such a formulation in M_c^4 one has first to construct a second-quantized theory in M_c^4, a thing which mildly speaking is not straightforward.

Besides M_c^4 does not admit a global causal ordering: M_c^4 contains closed timelike curves.[5] This is in contrast with M which admits a chronogeometry covariant under the Weyl group which may be extended to the conformal group locally in the following sense[3]: to any bounded region $B \subset M$ containing the origin of M there exists a neighborhood V_B of the identity in C such that the action of V_B on B is causal. This fact is sufficient for physical applications.

Any particular reason to utilize M_c^4 (if at all) should exhibit cosmological arguments.

B. The cotransformation under C_0 (the connected component of the identity in C) of scalars, four-vectors, antisymmetric tensors of rank two, and more generally of a field Ψ cotransforming according to a finite-dimensional irreducible representation R of $SL(2,\mathbb{C})$ is clear and well-defined: Let $g, g' \in SU(2,2)$ and suppose that when $x \mapsto x' = g \cdot x$, Ψ cotransforms according to $\Psi(x) \mapsto \Psi'(x') = S(g,x)\,\Psi(x)$, where S is a matrix function depending on $g \in SU(2,2)$ and $x \in M$. Under the assumptions of

(1) consistency with the group structure

$$S(g'g,x) = S(g', g \circ x)S(g,x), \tag{1}$$

whenever both sides are defined, and of

(2) compatibility with Poincaré covariance

$$S((a,A),x) = R(A), \tag{2}$$

where $(a,A) \in SL(2,\mathbb{C}) \circ \mathbb{R}^4$, then according to a theorem[4] $S(g,x)$ is defined for any $g \in SU(2,2)$. In fact let us denote by $S(a,x)$ the value of S for the usual special conformal transformations,

$$x \mapsto x' = \omega(a,x)^{-1}(x + ax^2), \tag{3}$$

where $\omega(a,x) = 1 + 2ax + a^2x^2$ and $ax = a^\mu x_\mu$, $x^2 = x^\mu x_\mu$. Then it is clear that $S(g,x)$ will be known once $S(a,x)$ is given. [Use a factorization of g and formulas (1) and (2)]. As shown in Ref. 4, $S(a,x)$ is uniquely defined up to some possible arbitrariness in the power of ω, called the conformal degree, i.e., $S(a,x) = \omega(a,x)^n S_0(a,x)$. Here $S_0(a,x)$ is the unique unimodular solution of a matrix differential system with initial value $S_0(a,0) = I$ (the identity matrix).

The expression of $S_0(a,x)$ for any irreducible representation (j,j') of $SL(2,\mathbb{C})$ may be computed from symmetrized tensor products of its value for the particular representation $(\frac{1}{2},0)$[6]: $S_0(a,x) = \omega(a,x)^{-1/2}(I + a_\mu x_\nu \bar{\sigma}^\mu \sigma^\nu)$, where $\sigma^0 = I$, σ^j the Pauli matrices, and $\bar{\sigma}^0 = \sigma^0, \bar{\sigma}^j = -\sigma^j$. Accordingly an expression for S_0 was given in Ref. 4 for the particular representation $(j,0) \oplus (0,j)$. Moreover a compact expression for $S(g,x)$ for any $g \in SU(2,2)$ is easy to find with the help of the previously quoted theorem and will be given elsewhere.[7] An extension with appropriate modifications of this unicity theorem to any finite-dimensional representation of $SL(2,\mathbb{C})$ is straightforward.[7] This is not the case for unitary representations of $SL(2,\mathbb{C})$, though it is trivial to find operators satisfying (1) and (2).

In the case of a tensor (p times covariant and q times contravariant) cotransforming according to an irreducible representation of $SO_0(3,1)$ the theorem allows to write[8]

$$S_0(a,x)=\omega(a,x)^{q-p}\left(\frac{\partial x}{\partial x'}\right)^p\left(\frac{\partial x'}{\partial x}\right)^q. \qquad (4)$$

All these well-known results show that:

(1) Various global expressions of $S(g,x)$ have been known for many years. In particular, though not understood by the author of Ref. 1, some of them have been given in Ref. 4. These expressions are not expansions up to order two in the group parameters although such an expansion was also given there. These expressions have been used to prove the conformal covariance of a certain number of field equations.

(2) These cotransformations of the fields do not have to be postulated nor do they need to be derived from cotransformations of fields defined on M_c^4.

(3) In the three cases quoted at the beginning of Part B (and more generally for tensors) it seems more natural to use the compact formula (4) than the six lines formulas given in Ref. 1, the consistency of which with the group structure is by no means obvious.

C. The well-known conformal covariance of Maxwell equations in the vacuum

$$\partial_\mu F^{\mu\nu}=0$$

trivially implies the conformal covariance of Maxwell equations in the presence of an external divergenceless current j,

$$\partial_\mu F^{\mu\nu}=j^\nu, \qquad (5)$$

the conformal degree of j being three in accordance with the conformal covariance of the continuity equation[4] $\partial_\mu j^\mu=0$. However Eq. (5) implies an interaction of a certain type. Therefore the real problem is to investigate the conformal covariance of the whole set of equations which are involved. In particular one has to check that the conformal degree of j^μ is three.

Before we give a simple treatment of the conformal covariance of a linear gauge condition for Maxwell equations we make the following remark. The nonlinear covariant condition $\partial_\mu(A_\nu A^\nu A^\mu)=0$ introduced in Ref. 4 did by no means preclude the existence of such a linear equation. Besides, for a plane wave $A^\mu(x)=a^\mu\exp(ikx)$, this nonlinear equation (with Maxwell equations) gives the Lorentz condition $k_\mu a^\mu=0$. This fact holds also for the linear gauge condition $\Box\partial_\mu A^\mu=0$ to be introduced below.

Let us now consider the following system of differential equations:

$$\Box A_\mu-\partial_\mu\Lambda=0, \qquad (6)$$

$$\partial_\mu A^\mu-\Lambda=0, \qquad (7)$$

$$\Box\Lambda=0, \qquad (8)$$

where A_μ is a four-vector and Λ a scalar. This system implies the conformally covariant equations $\Box A_\mu-\partial_\mu\partial_\nu A^\nu=0$ for the electromagnetic potential and the gauge condition $\Box\partial_\mu A^\mu=0$. We now prove the conformal covariance of the system. The pair (A,Λ) cotransforms under $SL(2,\mathbb{C})$ according to the reducible representation $(0,0)\oplus(\frac12,\frac12)$. The extension of the unicity theorem of Ref. 4 (see Ref. 7 and also Ref. 9) proves us after some computation that under the transformation (3) one has for the cotransformation of the pair (A,Λ) the following three possibilities:

$$\begin{aligned}
&(1)\ A'^\mu(x')=\omega(a,x)^{n+1}\,\partial_\nu x'^\mu A^\nu(x),\\
&\quad\ \Lambda'(x')=\omega(a,x)^m\,\Lambda(x);\\
&(2)\ A'^\mu(x')=\omega(a,x)^{n+1}\,\partial_\nu x'^\mu A^\nu(x),\\
&\quad\ \Lambda'(x')=\omega(a,x)^{n+1}\,\Lambda(x)+\alpha\,\omega(a,x)^n\\
&\qquad\qquad\times\partial_\mu\,\omega(a,x)A^\mu(x);\\
&(3)\ A'^\mu(x\ )=\omega(a,x)^{n+1}\\
&\qquad\qquad\times\partial_\nu x'^\mu[A^\nu(x)+\beta\,\omega(a,x)^{-1}\partial^\nu\omega(a,x)\,\Lambda(x)],\\
&\quad\ \Lambda'(x')=\omega(a,x)^{n-1}\,\Lambda(x),
\end{aligned} \qquad (9)$$

where m,n, and α,β are fixed scalars. [In cases (2) and (3) the pair cotransforms under the conformal group in an indecomposable manner.] The first and the third possibilities are excluded if one considers Eqs. (7) and (8). If one considers the second possibility and looks again at Eqs. (7) and (8) an easy calculation shows that the system will be conformally covariant only if $\alpha=n-3$ and $n=1$, namely $\alpha=-2$. These conditions are sufficient. In fact a straightforward computation shows that if the pair (A,Λ) cotransforms according to (9) with $n=1$ and $\alpha=-2$ one has

$$\begin{aligned}
\Box' A'_\mu(x')-\partial'_\mu\Lambda'(x')&=\omega(a,x)^2\,\partial'_\mu x^\rho\\
&\quad\times[\Box A_\rho(x)-\partial_\rho\Lambda(x)]\\
&\quad+2\omega(a,x)\partial_\rho\omega\partial'_\mu x^\rho[\partial_\nu A^\nu(x)-\Lambda(x)]
\end{aligned}$$

$$\partial'_\mu A'^\mu(x')-\Lambda'(x')=\omega(a,x)^2\,[\partial_\mu A^\mu(x)-\Lambda(x)]$$

$$\begin{aligned}
\Box'\Lambda'(x')&=\omega(a,x)^4\,\Box\Lambda(x)-8a^2\,\omega(a,x)^3\,[\partial_\mu A^\mu(x)-\Lambda(x)]\\
&\quad-2\omega(a,x)^3\,\partial^\mu\omega(a,x)\,[\Box A_\mu(x)-\partial_\mu\Lambda(x)].
\end{aligned}$$

This treatment of the covariance of the gauge condition, which is slightly different from the one given in Ref. 2 gives the result obtained also by Mayer[1] in a straightforward manner.

It is interesting to note that the condition $\Box\partial_\mu A^\mu=0$ has been met in quantum electrodynamics.[10,11] Haller and Landowitz[12] have shown that the Gupta–Bleuler condition $\partial^\mu A_\mu^{(+)}|n\rangle=0$ is not covariant and fails to define state vectors that remain in the physical subspace. For this reason a generalized Lorentz gauge formulation was introduced.[10] This formulation, which is obtained by supposing the gauge condition $\Box\partial_\mu A^\mu=0$, ensures in the presence of interaction between charged particles and photons the existence of an invariant positive-frequency part of $\partial_\mu A^\mu$.

Note added in proof. In a private letter to the authors of the present article (for which he is cordially acknowledged) Mayer raises a new argument in favor of the M_c^4 space. This argument has to do with some claims that occurred in the literature concerning the obiligatory utilization of representations of the universal covering group of the conformal group in quantum field theory (arguments of which we have been perfectly aware).

Our reasons for ignoring these particular arguments are the following:

(1) For many decades most physicists have been aware of the fact, as far as covariance and symmetry principles are concerned, one can utilize projective representations instead of true representations.

(2) One of us (M. F.) stressed already a long time ago that for many reasons (and in particular local causality in the case of the conformal group) one is led to the possibility of utilization of Lie algebra representations of the conformal group in conformal field theory (representations of the universal covering group are certainly included in the class of representations of the Lie algebra of the conformal group). In such a case the covariance principle will only hold in its (weaker) commutator form.

(3) Every argument known to us tending to prove that one is *obliged* to utilize unitary representations of the universal covering instead of representations of the conformal group itself is either mathematically incorrect or in the best cases is based upon a much stronger hypothesis than one really needs in usual versions of conformal Wightman-type field theories.

(4) At last, even if one would have been obliged to utilize representations of the universal covering group of the conformal group (which we do not believe is the case) the theory would have had as a natural space $\tilde{M}$, which is the universal covering space of M_c^4 and *at the same time* a covering space of the Minkowski space M.

[1]D. H. Mayer, J. Math. Phys. 16, 884 (1975).
[2]F. Bayen, C.R. Acad. Sci. Paris 280, 1083 (1975).
[3]M. Flato and D. Sternheimer, C.R. Acad. Sci. Paris 263, 935 (1966).
[4]M. Flato, J. Simon, and D. Sternheimer, Ann. Phys. (N. Y.) 61, 78 (1970).
[5]T. H. Go, H. A. Kastrup, and D. H. Mayer, to be published in Rep. Math. Phys. (1974).
[6]J. Wess, Nuovo Cimento 18, 1086 (1960).
[7]F. Bayen, in the book devoted to A. Lichnerowicz' sixtieth birthday, edited by M. Cahen and M. Flato (D. Reidel, Dordrecht, 1977); C.R. Acad. Sci. Paris, to be published (1976).
[8]M. Flato, P. Hillion, and J. Simon, C.R. Acad. Sci. Paris 268, 1065 (1969).
[9]G. Mack and A. Salam, Ann. Phys. (N. Y.) 53, 174 (1969).
[10]S. P. Tomczak and K. Haller, Nuovo Cimento B 8, 1 (1972).
[11]K. Haller, Phys. Rev. D 10, 4299 (1974).
[12]K. Haller and L. F. Landowitz, Phys. Rev. 171, 1749 (1968).

Angular momentum of systems of electric and magnetic charges and of singular flux surfaces

A. O. Barut* and H. Schneider

Sektion Physik der Universität München, 8 München 2, Germany
(Received 28 July 1975)

The angular momentum of a system consisting of an electric charge e and a magnetic charge g is, as is well known, (eg). We derive general formulas for systems consisting of arbitrary electric and magnetic charges, dyons and of singular magnetic flux lines or surfaces of arbitrary integrable topological shapes. The total angular momentum is then quantized and related to the quantized flux.

I. INTRODUCTION

It is known[1,2] that the angular momentum residing in the field produced by an electric charge e, and a magnetic charge g has the magnitude (eg), and that produced by two dyons (particles endowed with both electric and magnetic charges) has the magnitude $(e_1g_2 - e_2g_1)$. The purpose of this paper is to derive general formulas for systems composed of arbitrary electric and magnetic charges, dyons, and of manifolds with magnetic flux lines. An example for the latter is a cylindrical quantized flux enclosed by a superconducting ring in which electric charges are moving (along the axis). The total angular momentum will then be quantized and related to the quantized flux.

We wish to evaluate the angular momentum

$$\mathbf{J} = \int_V dV\, \mathbf{r}\times(\mathbf{E}\times\mathbf{B}), \tag{1}$$

where $V \subset \mathbb{R}_3$ is an arbitrary integrable region, in general multiply connected, with boundary ∂V. We enclose point charges and singular magnetic lines along one- or more-dimensional surfaces by small spheres or mantels which we will then let shrink to points or to singular surfaces. The surfaces of these spheres and mantels are part of the boundary ∂V.

II. SYSTEMS OF ELECTRIC AND MAGNETIC CHARGES

We evaluate (1) for the charge e_1 at $\mathbf{r}_1$ and then can sum over all charges.

If the total linear momentum $\int_V (\mathbf{E}\times\mathbf{B})\,dV = 0$, we can replace (1) by

$$\mathbf{J}_1 = \int_V (\mathbf{r} - \mathbf{r}_1)\times(\mathbf{E}\times\mathbf{B})\,dV. \tag{1'}$$

Let $\mathbf{a}$ be an arbitrary constant vector and $\mathbf{J}_1$ the angular momentum about the position of the charge:

$$\begin{aligned}\mathbf{a}\cdot\mathbf{J}_1 &= \int_V dV\,\mathbf{a}\cdot((\mathbf{r}-\mathbf{r}_1)\times(\mathbf{E}\times\mathbf{B}))\\ &= \int_V dV\,(\mathbf{E}\times\mathbf{B})\cdot(\mathbf{a}\times(\mathbf{r}-\mathbf{r}_1))\\ &= -\int_V dV\,\mathbf{B}\cdot(\mathbf{E}\times(\mathbf{a}\times(\mathbf{r}-\mathbf{r}_1))).\end{aligned} \tag{2}$$

Taking $\mathbf{E} = (e/|r-r_1|^3)(\mathbf{r}-\mathbf{r}_1)$ we have with $\mathbf{s} \equiv \mathbf{r}-\mathbf{r}_1$,

$$\begin{aligned}\mathbf{E}\times(\mathbf{a}\times\mathbf{s}) &= (\mathbf{E}\cdot\mathbf{s})\mathbf{a} - (\mathbf{E}\cdot\mathbf{a})\mathbf{s}\\ &= e[(1/s)\mathbf{a} - (\mathbf{s}\cdot\mathbf{a}/s^3)\mathbf{s}] = e\nabla(\mathbf{s}\cdot\mathbf{a}/s),\end{aligned} \tag{3}$$

hence

$$\begin{aligned}\mathbf{a}\cdot J_1 &= -e\int_V dV\,\mathbf{B}\cdot\nabla\left(\frac{(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{a}}{|r-r_1|}\right)\\ &= e\int_V dV\frac{(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{a}}{|r-r_1|}\nabla\cdot\mathbf{B} - e\int_{\partial V} d\mathbf{f}\cdot\mathbf{B}\frac{(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{a}}{|r-r_1|}.\end{aligned}$$

Because $\mathbf{a}$ is arbitrary we have the result that

$$\mathbf{J}_1 = e\int_V dV\frac{\mathbf{r}-\mathbf{r}_1}{|r-r_1|}\nabla\cdot\mathbf{B} - e\int_{\partial V} d\mathbf{f}\cdot\mathbf{B}(\mathbf{r}-\mathbf{r}_1)/|r-r_1|. \tag{4}$$

The total angular momentum from all electric charges with the charge distribution $e(\mathbf{r})$ is then

$$\mathbf{J} = \int dV\, e(\mathbf{r})\mathbf{J}_1(\mathbf{r}). \tag{5}$$

Equations (4) and (5) allow us to evaluate the angular momentum simply and directly without first calculating $\mathbf{E}$ and $\mathbf{B}$ and then carrying out the integral (1), although earlier results have been obtained that way.

Results

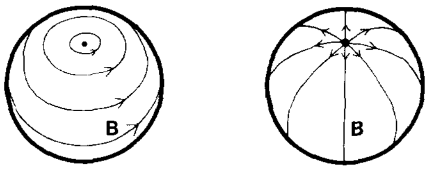
(1) For a standard Maxwell field $\nabla\cdot\mathbf{B} = 0$ everywhere; thus in (4) only the integral over the boundary around the point charges and around V remains and gives $\mathbf{J} = 0$.

(2) In the case of point magnetic charges we can evaluate (4) either as a volume integral over the whole space (first term), or as the integral over the boundary which is a small sphere around the magnetic charge (second term). Either way we obtain from (4),

$$\mathbf{J}_1 = e\sum_i g_i\hat{n}_i, \tag{6}$$

where g_i is the ith magnetic charge and $\hat{n}_i$ the unit vector connecting the electric charge to the ith magnetic charge. In particular, for a pair of magnetic charges and with an electric charge situated along the line connecting the two magnetic charges with unit vector $\hat{n}$ we find

$$\begin{aligned}\mathbf{J} &= 2eg\hat{n}, && \text{if } e \text{ is in between,}\\ &= 0, && \text{if } e \text{ is outside,}\\ &= eg\hat{n}, && \text{if } e \text{ is on one of the magnetic charges (dyon).}\end{aligned} \tag{7}$$



(a) (b)

FIG. 1.

Note that the passage from an e and g to a dyon (e,g) is discontinuous, as of course, coinciding charges are always discontinuous in Maxwell theory.

For a number of electric charges we have from (5)

$$\mathbf{J}=\sum_k e_k \sum_i g_i \hat{n}_i^k, \tag{8}$$

where $\hat{n}^k_i$ is the unit vector in the direction from the kth electric charge to the ith magnetic charge.

III. SINGULAR FLUX SURFACES

Let F be a submanifold of $\mathbb{R}_3$. By a suitable choice of coordinates we can diagonalize locally the induced metric g_F on F. Let the magnetic field $\mathbf{B}$ be concentrated on F and tangential to F. The identity (3) remains correct with ∇ as the covariant derivative on F. Hence

$$\begin{aligned}\mathbf{a}\cdot\mathbf{J}_1 &= -e\int_F df B^k(\mathbf{s}\cdot\mathbf{a}/s)_{;k}\\ &= -e\int_F df[(B^k\mathbf{s}\cdot\mathbf{a}/s)_{;k}-(\mathbf{s}\cdot\mathbf{a}/s)B^k{}_{;k}]\end{aligned} \tag{9}$$

or

$$\mathbf{J}_1 = e\int_F df(\mathbf{s}/s)\nabla^k B_k - e\int_{\partial F} d\sigma h_k B^k \mathbf{s}/s,$$

where σ is the line element on the boundary ∂F of F, and if $u(\sigma)$ is the tangent vector along ∂F,

$$h_i(\sigma)=\epsilon_{ki}(\sigma)\mathrm{u}^k(\sigma),\quad \epsilon_{ki}=\sqrt{g_F}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}. \tag{10}$$

Thus

$$\begin{aligned}\mathbf{J}_1 &= e\int_F df[(\mathbf{r}-\mathbf{r}_1)/|r-r_1|]\\ &\quad\times\nabla_k B_k - e\int_{\partial F} d\sigma h_k B^k(\mathbf{r}-\mathbf{r}_1)/|r-r_1|.\end{aligned} \tag{11}$$

Note that

$$\phi\equiv\int_{\partial F} d\sigma h_k B^k \tag{12}$$

is the *magnetic flux into* the surface F.

Closed surfaces: The boundary ∂F is zero. Hence in (11) only the singular points of the flow of magnetic lines on F contribute (Fig. 1). Denoting $\nabla^k B_k = g\delta_F$ at singular points, we again obtain the formula (8). It is clear that the result does not depend on the form of the closed surface, but only on the strength and number of the singular points of the flow on the submanifold.

For open surfaces (Fig. 2), $\nabla^k B_k=0$ inside F, hence only the second term in (11) survives. The boundary ∂F behaves like lines of magnetic charge distributions.

Singular flux lines

Let L be a line from the point A to the point B with the magnetic field concentrated on (and tangential to) L alone. Equation (11) now becomes

$$\mathbf{J}_1 = e\int_A^B ds(\mathbf{r}-\mathbf{r}_1)/|r-r_1|)\nabla^k B_k - e(g_B\hat{n}_B+g_A\hat{n}_A),\quad (n_B, n_A\neq 0). \tag{13}$$

But the first term is zero for a normal flux line and we obtain the negative of the expression (6), as it should be because the endpoints of a flux string behave like pair of positive and negative magnetic charges ($g_B = -g_A$). The flux string is *complementary* to the pair of magnetic charges in the sense that string plus the external field of the magnetic charges combined gives a Maxwellian system with $\nabla\cdot B=0$ everywhere.[3] The result (13) had been obtained previously for a straight-line geometry by a tedious direct integration of Eq. (1).[3] If the charge e coincides with B (*dyon*), for example, the corresponding term $g_B n_B$ is absent in (13).

Closed strings: The boundary $\partial L=0$, $\nabla\circ B=0$ on L hence $\mathbf{J}=0$.

The result (13) can also be obtained from (4) by surrounding the string by a tube of radius a and going to the limit $a\to 0$.

Because only the sources and sinks of the magnetic flux contribute, we have thus proved that *the flow of magnetic flux from a point A to another point B along any region of any shape and of any dimensionality (line, surface, or volume) gives rise to the same angular momentum for a charge e, namely* $e(g_B\hat{n}_B+g_A\hat{n}_A)$.

Remark: The considerations of Sec. III extend easily to singular surfaces of the electric field and point magnetic charges.

IV. QUANTIZATION OF THE ANGULAR MOMENTUM AND OF FLUX

The main results of this paper, Eqs. (6), (8), (11), and (13) connect the angular momentum to the magnetic flux. Consequently, the quantization of both of these quantities are intimately related.[4] In the case of a single magnetic charge g, the flux $\phi=4\pi g$ quantized according to London

$$\phi=(h/e)\nu,\quad \nu=0,1,2,3,\cdots \tag{14}$$

gives $J=eg=\frac{1}{2}\hbar\nu$, i.e., precisely half-integer spins *and* Dirac's quantization condition[1] for the product (eg): both this latter condition and (14) are in turn related to the requirement of continuity of the wavefunction around the singular potential lines, the phase of the wavefunction however being discontinuous by $2\pi\nu$.

For the more general case (8) we obtain weaker conditions for the product of electric and magnetic

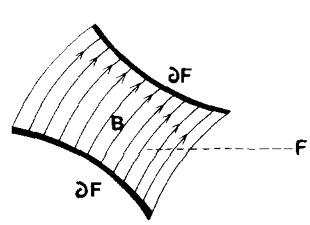


FIG. 2.

charges. The projection of **J** along some direction $\hat{n}$ is, from (8),

$$\mathbf{J}\cdot\hat{n}=\sum_{i,k} e_k g_i(\hat{n}_i^k\cdot\hat{n}). \tag{15}$$

If, however, the values of g_i are already determined from the Dirac rule $eg=\frac{1}{2}\hbar\nu$, then Eq. (15) is a restriction on the lowest possible total angular momentum of the system of electric and magnetic charges.

In the case of flows of magnetic lines from A to B along singular surfaces of any shape and dimensionality each quantum of flux correspond to an angular momentum

$$\mathbf{J}=\tfrac{1}{2}\hbar(\hat{n}_B-\hat{n}_A), \tag{16}$$

if e does not coincide with B or A, and

$$\mathbf{J}=\tfrac{1}{2}\hbar\hat{n}_B, \tag{16'}$$

if e does coincide with A (dyon). Thus we have a purely electromagnetic origin for the spin degree of freedom either from magnetic charges, or the complementary singular flux lines.[4]

*On leave of absence from the University of Colorado,

[1]P.A.M. Dirac, Proc. R. Soc. Lond. A 133, 60 (1931); Phys. Rev. 74, 817 (1948).

[2]M.N. Saha, Phys. Rev. 75, 1968 (1949); J. Schwinger, Science 165, 757 (1969); D. Zwanziger, Phys. Rev. 176, 1480, 1489 (1968); A.O. Barut and G. Bornzin, J. Math. Phys. 12, 841 (1971).

[3]A.O. Barut and G. Bornzin, Nucl. Phys. B 81, 477 (1974).

Boulder, Colorado. Recipient of an Alexander von Humboldt Senior Scientist Award.

[4]A.O. Barut, Acta. Phys. Austriaca, Suppl. 11, 565 (1973).

Cluster expansions for fermion fields by the time dependent Hamiltonian approach

David C. Brydges*

Department of Mathematics, The University of Michigan, Ann Arbor, Michigan 48104
(Received 3 November 1975)

A cluster expansion is given for a fermion field moving in an external field according to the interaction $\bar{\psi}\psi\phi$ in one space dimension.

The purpose of this paper is pedagogic: to demonstrate, in a problem relatively untrammelled by technical complications, how a cluster expansion may be developed for the Yukawa model in one space dimension. The cluster expansion is based on the time dependent Hamiltonian approach devised by Federbush.[1] It appears to the author that the method of integrating out all the fermions[2-4] being developed by McBryan, Seiler, and Simon should be capable of producing a more streamlined version of the cluster expansion than the one implied by this paper. However, since it may be slightly surprising that the cluster expansion is not tied to a manifestly covariant approach, and, furthermore, the same general idea has the applications[1] in statistical mechanics, perhaps this paper is not devoid of interest.

Federbush's idea is to replace the use of covariances with Dirichlet conditions in Ref. 5 by Hamiltonians for which regions are isolated by potential barriers. The extremes wherein two regions are completely isolated or in full communication are interpolated by varying the heights of these barriers. It will be seen that barriers parallel to the spatial axis may be incorporated by making the Hamiltonians suitably time dependent. The main disadvantage of this approach is that the heights of the barriers get involved in the estimates. The other technical complication which will occur in all approaches to a cluster expansion for the Yukawa model is the nonpositivity (in any sense) of the propagator for fermions. The positivity properties of the free Euclidean boson covariance were a useful aid in Ref. 5.

The formal structure of the expansion is the same (except that integrations over interpolating variables run from zero to infinity) as in Ref. 5, [see Eqs. (3.13)]. This will become evident in Sec. 1. Therefore, rather than repeating much of the material in Ref. 5, this paper is confined to establishing suitable analogs to the key ingredients of the convergence proof in Ref. 5. These are the formula for differentiating the measure (1.7) in Ref. 5 (this is discussed in Sec. 2), and Proposition (5.3) in Ref. 5, which is the subject of Secs. 3–5. The technical complications mentioned above are all buried in Sec. 5. The main point is the identity (1.11), which relates part of the free fermion propagator to Brownian motion. A "hand-wave" at the full Yukawa theory in one space dimension is given in the Appendix.

The external field model discussed here is almost as singular as the Yukawa model (the vacuum energies of both are logarithmically divergent), hence the cluster expansion is given for fields with a momentum cutoff and the convergence is established uniformly in this cutoff. The difference between the two models may be expressed in the following way: The external field $\phi=\phi(x)$ in this paper is assumed to belong to $L^3(\mathbb{R})$ locally uniformly. If ϕ were a boson field, this condition would be logarithmically divergent in a momentum cutoff. Needless to say, this extra divergence entails considerable complications for the Yukawa model. Nevertheless, the author feels that the interesting parts of a cluster expansion (by this method) for the Yukawa model are contained in this paper.

1. NOTATION

With the same notation and representation as in Ref. 6 the fermion field ψ at time zero is given by

$$\psi(x)=(4\pi)^{-1/2}\int \exp(ipx) \times \{v(p)\,b'^*(p)+u(-p)\,b(-p)\}\,\omega^{-1/2}(p)\,dp, \tag{1.1}$$

where the spinors are

$$u(p)=\begin{bmatrix}[\omega(p)-p]^{1/2}\\ -[\omega(p)+p]^{1/2}\end{bmatrix},\quad v(p)=\begin{bmatrix}[\omega(p)-p]^{1/2}\\ [\omega(p)+p]^{1/2}\end{bmatrix}. \tag{1.2}$$

The cluster expansion will be developed by devising a family of time dependent Hamiltonians which will interpolate between the usual one and one which does not propagate across or into any of the shaded regions in Fig. 1. Each shaded region is centered on a line of integral ordinate. The width ϵ of the "barriers" will be chosen below. To this end, the free Hamiltonian is modified as follows: The single particle kinetic energy is given by the operator $(M_0^2-\Delta)^{1/2}$, where $\Delta=d^2/dx^2$. Define the (τ,s) dependent operator,

$$\omega(\tau,s)=(M_0^2-\Delta+\sum_b s_b\chi_b)^{1/2}, \tag{1.3}$$

where $\chi_b=\chi_b(x,\tau)$ is a characteristic function of one of the shaded regions in Fig. 1, which particular one being specified by the subscript b, and $s=(s_b)$ is a multivariable (each $s_b\geqslant 0$) parametrizing the interpolation. By a well-known theorem, $\omega^2(s,\tau)$ converges pointwise in τ in the sense of strong convergence of the resolvent, as a given $s_b\to\infty$, to the analogous operator with Dirichlet conditions on b. (A proof can easily be constructed using the path space representation.) This implies, by functional analysis, that $\exp[-\omega(\tau,s)]$ converges strongly, pointwise in τ. Define the corresponding time dependent free-field Hamiltonian by

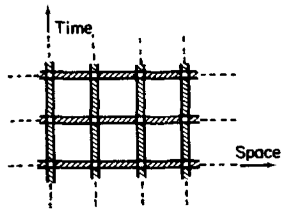


FIG. 1. Illustration of an infinite array of barriers partitioning $\mathbb{R}^2$.

second quantization of $\omega(\tau, s)$; i.e.,

$$H_{OF}(\tau, s) = \int dx\, dy\, \omega(\tau, s, x, y)[b^*(x)\, b(y) + b'^*(x)\, b'(y)], \tag{1.4}$$

where $\omega(\tau, s, x, y)$ is the distribution kernel of the operator (1.3).

The interpolation (1.4) is not sufficient to make different squares independent because ψ contains the nonlocal operators $\omega^{-1/2}$, u, v. Also it is necessary to use a cutoff on the momentum which likewise is not local. Choose a positive function $\kappa = \kappa(x)$ in $C^\infty(\mathbb{R})$ with support in the interval $(-\frac{1}{10}, \frac{1}{10})$, normalized so that $\int \kappa(x)\, dx = 1$. Define $\kappa_\epsilon(x) = (1/\epsilon)\kappa(x/\epsilon)$ where $\epsilon > 0$. At the outset of the cluster expansion, choose some $\epsilon > 0$ and cut off all fields by the replacement $\psi(x) \to (\kappa_\epsilon * \psi)\ (x)$. The nonlocal operators in $\kappa_\epsilon * \psi$ are the factors $\hat{\kappa}_\epsilon(p)\, \omega^{-1/2}(p)\, u(p)$ and $\hat{\kappa}_\epsilon(p)\omega^{-1/2}(p) v(p)$. Let $f(\tau, s, x, y) = \Pi_{b \in \Gamma(\tau, x, y)}[1/(1 + s_b)]$ where $\Gamma(\tau, x, y)$ is the set of barriers b that intersect the straight line joining (x, τ) and (y, τ). Now let

$$\begin{aligned} &w_+(x, y, \tau, s, \epsilon) \\ &\quad = (4\pi)^{-1/2}\, (\hat{\kappa}_\epsilon \omega^{-1/2} u)^\vee(x - y) f(\tau, s, x, y), \\ &w_-(x, y, \tau, s, \epsilon) \\ &\quad = (4\pi)^{-1/2}\, (\hat{\kappa}_\epsilon \omega^{-1/2} v)^\vee(x - y) f(\tau, s, x, y), \end{aligned} \tag{1.5}$$

and define

$$\begin{aligned} \psi(x, \tau, s, \epsilon) = \int dy\, [\, & w_+(x, y, \tau, s, \epsilon)\, b(y) \\ & + w_-(y, x, \tau, s, \epsilon)\, b'^*(y)]. \end{aligned} \tag{1.6}$$

Note that $\psi(x, \tau, 0, \epsilon) = (\kappa_\epsilon * \psi)\ (x)$; $s = 0$ means $s_b = 0$ for all b; at $s = \infty$, $w_\pm(x, y) = 0$, whenever x, y are separated by a barrier or whenever x or y is in a barrier. From now on, dependences on s, τ, ϵ will frequently be suppressed. Corresponding to these interpolated fields, define the propagator, letting $x = (x', t)$, $y = (y', u) \in \mathbb{R}^2$, by

$$S(x, y, s) = T\ \langle \bar{\psi}(x', t, s) \exp[-\textstyle\int_{-\infty}^{\infty} H_{OF}(\tau, s)\, d\tau]\, \psi(y', u, s)\rangle, \tag{1.7}$$

where T is the time ordering operator. All time dependences are piecewise constant, so (1.7) is easily defined. More explicitly, the propagator is given by

$$S(x, y, s) = \begin{cases} \int dx_1'\, dy_1'\, \bar{w}_-(x', x_1', t, s)\, p(x_1', t, y_1', u, s) \\ \quad \times w_-(y_1', y', u, s), \quad t > u, \\ -\int dx_1'\, dy_1'\, w_+(y', y_1', u, s)\, p(y_1', u, x_1', t, s) \\ \quad \times \bar{w}_+(x_1', x', t, s), \quad u > t, \end{cases} \tag{1.8}$$

where $p(x', t, y', u)$ is the kernel of the operator defined by the time ordered exponential

$$T \exp\left(-\int_{\min(t,u)}^{\max(t,u)} \omega(\tau, s)\, d\tau\right). \tag{1.9}$$

At $s = 0$, S reduces to the usual propagator. At $s = \infty$, all unshaded squares in Fig. 1 become independent. The proof is left to the reader.

The kernel $p(x', t, y', u)$ is a positive function and can be related to Brownian motion, which will be essential for estimates. The principle whereby this is obtained in probability theory is called subordination. Within the context of semigroups it can be obtained by: $t > 0$, $\beta > 0$,

$$\exp(-t\beta) = (1/i\pi) \int_{-\infty}^{\infty} dk_0\, k_0 (k_0^2 + \beta^2)^{-1} \exp(ik_0 t) \tag{1.10}$$

$$\begin{aligned} &= \int_0^\infty ds ((1/i\pi) \int_{-\infty}^{\infty} dk_0\, k_0 \exp(-k_0^2 s) \exp(ik_0 t)) \\ &\qquad \times \exp(-s\beta^2) \\ &= \int_0^\infty ds\, t^{-2} \mu(s/t^2) \exp(-s\beta^2), \end{aligned} \tag{1.11}$$

where $\mu(s) = (1/2\sqrt{\pi}) \exp(-1/4s)\, s^{-3/2}$. The essential point is that μ is positive. Hence the kernel of $\exp[-t\omega(\tau, s)]$ at fixed (τ, s) may be expressed in terms of the kernel of $\exp[-t\omega^2(\tau, s)]$ which is positive by virtue of its relation to Brownian motion. The operator (1.9) is a product of operators with the form $\exp[-t\omega(\tau, s)]$.

The family of interacting Hamiltonians will now be defined. They are time dependent, but at $s = 0$ they reduce to the normal time independent cutoff Hamiltonian for a fermion field interacting with an external field ϕ.

$$\begin{aligned} H(\tau, s, \Lambda) = H_{OF}(\tau, s) + \int_{\Lambda_\tau} dx\, \phi(x) \\ \times{:}\bar{\psi}(x, \tau, s, \epsilon)\, \psi(x, \tau, s, \epsilon){:}\, dx + E_\epsilon(\Lambda) \end{aligned} \tag{1.12}$$

where $\Lambda \subset \mathbb{R}^2$, is a bounded, measurable set and Λ_τ denotes its spatial cross-section at time τ. The external field ϕ is assumed to satisfy, for all $x_1 \in \mathbb{R}$,

$$[\int_{x_1}^{Hx_1} |\phi(x)|^3 dx]^{1/3} \leq \text{const}. \tag{1.13}$$

The vacuum energy renormalization is inspired by perturbation theory:

$$\begin{aligned} E_\epsilon(\Lambda) = (4\pi)^{-1} \int dp_1\, dp_2 |(\chi_\Lambda \phi)^\wedge(p_1 + p_2)^2\ \frac{\omega_1\omega_2 - p_1 p_2 - M_0^2}{\omega_1 \omega_2} \\ \times \frac{1}{\omega_1 + \omega_2} [\hat{\kappa}^2(\epsilon p_1) \hat{\kappa}^2(\epsilon p_2)\, (2\pi)^2], \end{aligned} \tag{1.14}$$

where $\chi_\Lambda = \chi_\Lambda(x, \tau)$ is the characteristic function of Λ, and $\wedge$ denotes the fourier transform with respect to x only. Hence $E_\epsilon(\Lambda)$ is a function of τ.

From Fig. 1 it will be seen that $\mathbb{R}^2$ has been partitioned into the following subsets: large squares (unshaded), small squares (unshaded), and rectangles (shaded). Let ∇ denote an element of this partition. The projection of this partition onto the spatial axis yields a partition of the latter into long and short intervals. Let I label an element in this partition. Let J label the elements of the similarly obtained partition of the time axis.

The barriers, i.e., shaded rectangles in Fig. 1, will be chosen of width ϵ. Those barriers whose short dimension is in the space direction are called "space barriers;" the others are "time barriers." The subset of $\mathbb{R}^2$, Λ in (1.12) is restricted to be a union of ∇s so that (1.12) has a stepwise time dependence and time orderings are easily defined.

The cluster expansion is used to analyze quantities of the form

$$\int dx\, W(x)\, T\left\langle \prod_{i=1}^{p} \psi\#(x_i) \exp[-\textstyle\int H(\tau,\Lambda)\,d\tau]\right\rangle, \quad (1.15)$$

where $x=x_1,x_2,\ldots,x_p$ with $x_i\in\mathbb{R}^2$, $x_i=(x_i^1,t_i)$, $W\in L^2(\mathbb{R}^{2p})$, $\psi\#=\psi$ or $\overline{\psi}$. Generally, at the beginning of the expansion, $s=0$ and ψ,H are the usual fields and Hamiltonian obtained by setting $s=0$ in (1.6) and (1.4). In this case the $\psi\#$s will have no time dependence; however, the t_is and τ must still be left in (1.15) as "dummy" variables because of their other role, time ordering the expression. The expansion is obtained by replacing $\psi\#$s and H by their s dependent analogs and using the fundamental theorem of calculus together with factorizations and resummations exactly as in Ref. 5. The proof that (1.15) factors in the appropriate manner along contours where $s=0$ can be carried out by first approximating the exponential by a polynomial, whereupon the factorization is an immediate consequence of Wick's theorem.

2. THE ANALOG TO THE FORMULA FOR DIFFERENTIATING THE MEASURE IN REF. 5

As essential ingredient in the convergence proof of the cluster expansion in Ref. 5 is its formula (1.7) for differentiating the measure. A similar formula also holds in this framework, namely

$$\frac{d}{ds_b}\, T\langle P(\overline{\psi}(s),\psi(s))\exp[-\textstyle\int H(\tau,s,\Lambda)\,d\tau]\rangle$$
$$=T\langle (S\,!\,\Delta_{\overline{\psi}\psi})\,P(\overline{\psi}(s),\ \psi(s))\exp[-\textstyle\int H(\tau,s,\Lambda)\,d\tau]\rangle, \quad (2.1)$$

where P is a polynomial and

$$(S\,!\,\Delta_{\overline{\psi}\psi})=-\int\left(\frac{d}{ds_b}\, S(x,y,s)\right)\frac{\delta}{\delta\overline{\psi}(x)}\,\frac{\delta}{\delta\psi(y)}\,dx\,dy, \quad (2.2)$$

with $x,y\in\mathbb{R}^2$ and $\delta/\delta\overline{\psi}$, $\delta/\delta\psi$ are formal anticommuting functional derivatives. The proof of (2.1) is very easy in the special case where $\phi=0$ so that $H(\tau,s,\Lambda)=H_{0F}(\tau,s)$, for then the expectation on the left of (2.1) may be evaluated by using the anticommutation rules and the identity

$$b(x)\exp[-\lambda H_{0F}(s,\tau)]$$
$$=\exp[-\lambda H_{0F}(s,\tau)]\,(\exp[-\lambda\omega(s,\tau)]b)(x), \quad (2.3)$$

together with similar ones for b^*,b',b'^*. This special case can be used to approximate the case when $\phi\neq 0$.

3. THE EXPONENT

In the next section, the following estimate will be used to control the exponential part of (1.15). For M_0 sufficiently large,

$$H(\tau,s,\Lambda)\geq -\text{const}\,|\Lambda_\tau| \quad (3.1)$$

uniformly in ϵ,τ, and s, where $|\Lambda_\tau|$ is the length of $\Lambda_\tau\subset\mathbb{R}$. The s dependence in the $H_{0F}(s)$ is bounded below by $H_{0F}(s)\geq H_{0F}(0)$. This is implied by the operator estimate $\omega(s)\geq\omega(0)$ which is implied by $\omega^2(s)\geq\omega^2(0)$ because operator estimates remain valid on taking square roots. The last bound is obvious. Thus (3.1) is implied by

$$H_{0F}+\int_{\Lambda_\tau}\phi(x){:}\,\overline{\psi}(x,\tau,s,\epsilon)\,\psi(x,\tau,s,\epsilon){:}\,dx+E_\epsilon(\Lambda)$$
$$\geq -\text{const}\,|\Lambda_\tau|. \quad (3.2)$$

The pure creation and annihilation parts, V_p, of the interaction have the form

$$\int dp_1\,dp_2[\upsilon(p_1,p_2)b^*(p_1)\,b'^*(p_2)$$
$$-\overline{\upsilon}(p_1,p_2)\,b(p_1)\,b'(p_2)] \quad (3.3)$$

for some ϵ,τ,s,ϕ dependent function υ. The operator $H_{0F}+V_p+E_\epsilon(\Lambda)$ will now be bounded below by a first-order dressing transformation originally due to Glimm.[7] In this situation, it is possible to use a particularly simple form of this transformation.[6,8] First construct an operator ΓV_p to satisfy

$$[H_{0F},\ \Gamma V_p]=V_p. \quad (3.4)$$

Note that (3.4) requires ΓV_p be antisymmetric. Thus let

$$\Gamma V_p=\left(\int dp_1\,dp_2\,\upsilon(p_1,p_2)\,\frac{1}{\omega(p_1)+\omega(p_2)}\right.$$
$$\left.\times b^*(p_1)\,b'^*(p_2)\right)-(\text{its adjoint}). \quad (3.5)$$

Define dressed operators by

$$\tilde{b}(p)=b(p)+[b(p),\ \Gamma V_p],$$
$$\tilde{b}'(p)=b'(p)+[b'(p),\ \Gamma V_p]. \quad (3.6)$$

Obtain an inequality by calculating the positive operator $\int dp\,\omega(p)[\tilde{b}^*(p)\,\tilde{b}(p)+\tilde{b}'^*(p)\,\tilde{b}'(p)]$. Thus

$$\tilde{b}^*(p)\,\tilde{b}(p)=b^*(p)\,b(p)+[b^*(p)\,b(p),\ \Gamma V_p]$$
$$+[b^*(p),\ \Gamma V_p]\,[b(p),\ \Gamma V_p]. \quad (3.7)$$

Normal order the operators in the last term of (3.7). The normal ordered form is a negative operator because of the anticommutation relations. Hence,

$$H_{0F}+V_p+\int dp_1\,dp_2\,|\upsilon(p_1,p_2)|^2\,(\omega_1+\omega_2)^{-1}\geq 0. \quad (3.8)$$

The third term corresponds to the contraction during the normal ordering. It diverges as $\epsilon\to 0$, however, up to a term which is greater than the right-hand side of (3.2), it cancels with $E_\epsilon(\Lambda)$. This calculation is not very difficult, using some of the devices developed in Sec. 5.

By repeating the above argument with $\int dp\,\tilde{\omega}[\tilde{b}^*\tilde{b}+\tilde{b}'^*\tilde{b}']$ where $\tilde{\omega}=\omega-c\omega^{\tau'}$ with $\tau'<1$ and c chosen so that $\tilde{\omega} > 0$, the result can be strengthened to

$$H_{0F} + V_p + E_\epsilon(\Lambda) \geq cN_{\tau'} - \text{const}\,|\Lambda_\tau|, \tag{3.9}$$

where $N_{\tau'}$ is the second quantization of $\omega^{\tau'}$. For M_0 sufficiently large the remaining part of the interaction, V_s, can be bounded by $cN_{\tau'}$ using local N_τ estimates.[9] Given all the machinery in Ref. 9, this is a fairly standard calculation, so it will be omitted.

4. CONVERGENCE

The result of "cluster expanding" (1.15), leads in analogy to (3.13) in Ref. 5 to an expansion whose coefficients have the general form (Γ is a set of barriers; $\partial^\Gamma = \Pi_{b\in\Gamma}\, d/ds_b$; $\int ds_\Gamma$ integrates the s_b for $b \in \Gamma$ from 0 to ∞), i.e.,

$$\int dx\, W(x) \int ds_\Gamma\, \partial^\Gamma T \times \left\langle \prod_{i=1}^{p} \psi\#(x_i, s) \exp[-\textstyle\int H(\tau, s, X_0)\, d\tau] \right\rangle. \tag{4.1}$$

where X_0 is a union of ∇s. It is now shown how to estimate (4.1) to obtain expressions very similar to those in Ref. 5 and thus obtain a convergence proof by the same kind of combinatorics as in Ref. 5.

First (2.1) is used in combination with Leibniz' rule to perform ∂^Γ. The result is the s integral of a sum of terms of the form, $y_j \in \mathbb{R}^2$, $y = (y_j)$, i.e.,

$$\int dy\, K_{\mathcal{G}}(y, s)\, T \left\langle \prod_{j=1}^{m} \psi\#(y_j, s) \exp[-\textstyle\int H(\tau, s, X_0)\, d\tau] \right\rangle, \tag{4.2}$$

where $\mathcal{G}$ indexes each term in the sum. Each field $\psi\#$ in (4.2) will be assumed to be localized by the kernel $K_{\mathcal{G}}$ in a space–time region ∇, so $\mathcal{G}$ not only indexes the ways in which the differentiations are applied (terms resulting from Leibniz' rule), but also the possible localizations of the fields.

(4.2) is estimated by taking the operator norm over the fermion Fock space in the following manner: Introduce an orthonormal product basis, f_α of $L^2(\mathbb{R}) \times \cdots \times L^2(\mathbb{R})$, with m factors. Write $K_{\mathcal{G}}(s)$ as a sum of f_α with coefficients that depend on s and the time variables in y. Estimate (4.2) by applying $\|\psi\#(f)\| \leq \|f\|_2$ to this sum together with the lower bound on the exponent, (3.1). Let t be the multivariable $t_1, \ldots, t_m$, the times in y. (4.2) is less than, in absolute value,

$$\int dt\, \|K_{\mathcal{G}}(t, s)\|_D \exp(\text{const}\,|X_0|), \tag{4.3}$$

where the (defermiation) norm $\|\ \|_D$ is defined on a function F of m space variables $x_1', \ldots, x_m'$ by the prescription (sum over norms of components if F is a matrix)

$$\|F\|_D = \inf \sum_\alpha |\langle F, f_\alpha\rangle|, \tag{4.4}$$

where the infimum is over product bases of $\Pi L^2(\mathbb{R})$ as above.

In the special case where W is constant and there is no $\int dx$ in (4.1), the quantity $K_{\mathcal{G}}$ factors, together with the integrals $\int dt$, into terms $K_{\mathcal{G}_c}$, corresponding to the way in which any Feynman graph for this theory factors, into its connected subgraphs which are either closed loops or lines with open ends [an open end corresponding to a field $\psi\#$ in (4.2)]. Because the f_α in (4.4) are product bases, the $\|\ \|_D$ norm in (4.3) also factors. A factor $K_{\mathcal{G}_c}$ corresponding to a graph consisting of a connected line with n vertices and both ends open has the form

$$K_{\mathcal{G}_c}(x_1, x_n, s) = \int dx_2 \cdots dx_{n-1} (\phi\chi_{\nabla_1})(x_1) S_{\gamma_1}(x_1, x_2)(\phi\chi_{\nabla_2})(x_2) \times S_{\gamma_2}(x_2, x_3)(\phi\chi_{\nabla_3})(x_3) \cdots S_{\gamma_{n-1}}(x_{n-1}, x_n)(\phi\chi_{\nabla_n})(x_n), \tag{4.5}$$

where $\gamma_1, \ldots, \gamma_n$ are disjoint subsets of Γ and $S_\gamma = \Pi_{b\in\Gamma}\, d/ds_b S$. $x_i \in \mathbb{R}^2$. By making a specific choice of product basis f_α in (4.4), it is easy to see that, letting $x_i = (x_i', t_i)$,

$$\|K_{\mathcal{G}_c}(t_1, t_n)\|_D \leq \text{const tr}\,|K_{\mathcal{G}_c}(t_1, t_n)|, \tag{4.6}$$

where $|K_{\mathcal{G}_c}(t_1, t_n)|$ means the operator absolute value, i.e., $[K_{\mathcal{G}_c}^*(t_1, t_n) K_{\mathcal{G}_c}(t_1, t_n)]^{1/2}$, where $K_{\mathcal{G}_c}(t_1, t_n)$ denotes the operator corresponding to the kernel $K_{\mathcal{G}_c}(x_1', t_1, x_n', t_n)$ for fixed t_1, t_n. Except when $n = 2$ the trace norm in (4.6) may be estimated by first taking the time integrals in (4.5) outside the trace norm and then majorizing by a product of Hilbert Schmidt norms of the form, $i = 1, 2, \ldots, n-1$,

$$\sum_{\alpha,\beta} [\textstyle\int dx_i'\, dx_{i+1}' |(|\phi|^{1/2}\chi_{\nabla_i})(x_i) \times S_{\gamma_i}^{(\alpha,\beta)}(x_i, x_{i+1})(|\phi|^{1/2}\chi_{\nabla_{i+1}})(x_{i+1})|^2]^{1/2}, \tag{4.7}$$

where α, β, are spinor indices. By the Hölder inequality (4.7) is less than

$$\|\phi\chi_{\nabla_i}(t_i)\|_3^{1/2} \|S_{\gamma_i}(t_i, t_{i+1})\|_{3,\nabla_i\times\nabla_{i+1}} \|\phi\chi_{\nabla_{i+1}}(t_{i+1})\|_3^{1/2}, \tag{4.8}$$

where the $\|\ \|_{3,\nabla\times\nabla'}$ norm is defined for a function $F = F(x, y)$ with $x = (x', t)$, $y = (y', u)$, to be the L^3 norm with respect to x', y' of $\chi_\nabla(x) F(x, y) \chi_{\nabla'}(y)$. When F is a matrix, the norm is defined to be the sum over the norms of the components. $\|\phi\chi_\nabla\|_3$ is the L^3 norm of $\phi\chi_\nabla$ with respect to the spatial variable. So by the hypothesis (1.13), for $n \geq 2$,

$$\int dt_1 dt_n \|K_{\mathcal{G}_c}(t_1, t_n)\|_D \leq (\text{const})^n \int dt_1 \cdots dt_n \times \prod_{i=1}^{n-1} \|S_{\gamma_i}(t_i, t_{i+1})\|_{3,\nabla_i\times\nabla_{i+1}}. \tag{4.9}$$

The time integral is estimated by the Cauchy Schwarz inequality to obtain

$$\int dt_1 dt_n \|K_{\mathcal{G}_c}\|_D \leq (\text{const})^n \prod_{i=1}^{n-1} \|S_{\gamma_i}\|_{3,\nabla_i\times\nabla_{i+1}}, \tag{4.11}$$

where

$$\|S_{\gamma_i}\|_{3,\nabla_i\times\nabla_{i+1},2} = (\textstyle\int dt_i\, dt_{i+1} \|S_{\gamma_i}(t_i, t_{i+1})\|_{3,\nabla_i\times\nabla_{i+1}}^2)^{1/2}. \tag{4.12}$$

Suppose now that $n = 2$ in (4.5). In the next section, identities and estimates for S_γ are developed and with their aid it is a simple matter to prove directly that for some $q>0$,

$$\int ds_{\gamma_1} \int dt_1 dt_2 \operatorname{tr}|K_{\mathcal{G}_c}(t_1,t_2)|$$
$$\leq \text{const} M_0^{-|\gamma_1|/q} \exp[-M_0 d(j,\gamma_1)/2], \qquad (4.13)$$

where $\int ds_{\gamma_1}$ integrates only over those s_b such that $b\in\gamma_1$. The estimate is uniform in the remaining s_b. $d(j,\gamma)$ is defined for a localization j and a line γ to be the distance between ∇_1 and ∇_2 measured by the shortest path which touches every barrier $b\in\gamma$ where ∇_1 and ∇_2 are the localizations specified by j for the variables in S_γ.

In general, $K_{\mathcal{G}}$ also contains numerical factors corresponding to closed loops. These already have the form of a trace of an operator, hence in absolute value are less than the trace norm of that operator. The operator has a form similar to (4.5) and can be bounded by the same methods, even when $n=2$ since there is an extra propagator in the operator. Also in the next section, it will be proved that, letting $j(\gamma)$ denote the localization specified for S_γ,

$$\int ds_\gamma \| S_\gamma\|_{3,j(\gamma),2}$$
$$\leq \text{const}\, M_0^{-|\gamma|/q} \exp[-M_0 d(j,\gamma)/2]. \qquad (4.14)$$

Putting all this together and using a simple argument to include W gives

$$\left| \int dx\, W(x) \int ds_\Gamma\, \partial^\Gamma T \left\langle \prod_{i=1}^{p} \psi^\#(x_i) \right.\right.$$
$$\left.\left. \times \exp\left[- \int H(\tau,s,X_0)\, d\tau\right]\right\rangle \right|$$
$$\leq \|W\|_2 \exp(\text{const}\, |X_0|)$$
$$\times \sum_{\mathcal{G}} (\text{const})^{n(\mathcal{G})} \prod_\gamma M_0^{-|\gamma|/q} \exp[-M_0 d(j,\gamma)/2].$$

The $\|W\|_2$ is with respect to space and time variables. This estimate is almost the same (a different localization) as estimates in Ref. 5 and leads to a proof of convergence for M_0 sufficiently large via the same combinatorics as in Ref. 3. In particular Proposition 5.3 of Ref. 5 is a consequence.

5. ESTIMATES ON THE PROPAGATOR

The objective is to prove (4.14). A proof of (4.13) can easily be constructed using the same techniques.

To begin with, write

$$\exp[-|t-u|\omega(s)] = \exp[-|t-u|\omega_D(s)] - E, \qquad (5.1)$$

where t,u are assumed to be in the same interval J, so that $\omega(s)=\omega(s,\tau)$ is constant for $\tau\in[t,u]$. $\omega_D(s)$ is defined to be the result of replacing Δ by Δ_D in (1.3). Δ_D is the Laplacian with Dirichlet conditions at every boundary point of intervals I, in the partition of the spatial axis. (see Sec. 1.) Now it will be shown, by constructing an integral representation for E, that

$$\int ds_\gamma \|E_\gamma\|_{3,\nabla_x\times\nabla_y,2}$$
$$\leq \text{const}\, M_0^{-|\gamma|/q} \exp(-M_0 d(\nabla_x,\nabla_y,\gamma)/2), \qquad (5.2)$$

where $E_\gamma = \partial^\gamma E(x,y)$ with $x=(x',t)$, $y=(y',u)$. ∇_x, the localization of x, is $I_x X J$ and $\nabla_y = I_y X J$.

The representation for E, by (1.10), is

$$E(x,y) = (i\pi)^{-1} \int_{-\infty}^{\infty} dk_0\, k_0 \exp(ik_0|t-u|) g(x',y'), \quad (5.3)$$

where $g(x',y')$ is the difference between two Green's functions with different boundary conditions for the differential operator $\omega^2(s)$. Thus, it satisfies the homogeneous equation (acting on x')

$$(\omega^2 + k_0^2) g(x',y') = 0,$$
$$g(x',y') = (k_0^2+\omega^2)^{-1}(x',y'), \quad \text{for } x'\in\partial I_x, \qquad (5.4)$$

where the s dependences have been suppressed. $(k_0^2+\omega^2)^{-1}(x',y')$ denotes the kernel of $(k_0^2+\omega^2)^{-1}$. Let $l(x')$ be the linear function of x' which coincides with $(k_0^2+\omega^2)^{-1}(x',y')$ when $x'\in\partial I_x$. By obtaining the equation for $g(x',y')-l(x')$ as a function of x', it can be seen that

$$g(x',y') = -(\Delta_D(k_0^2+\omega^2)^{-1} l)\,(x',y'), \quad \text{for } x'\in\partial I_x, \qquad (5.5)$$

where the operators act on the x' variable. Let $l_{x_-}(x')$ be the linear function that equals one at x'_-, the left-hand boundary of I_x, and equals zero at x'_+ the right end point. Set $l_{x_+} = 1 - l_{x_-}$. Define $g_\xi(x') = (-\Delta_D[k_0^2+\omega_D^2]^{-1} l_\xi)(x')$, where $\xi = x'_+$ or x'_-. Then

$$g(x',y') = \sum_{\xi=x'_+,x'_-} (k_0^2+\omega^2)^{-1}(\xi,y') g_\xi(x'). \qquad (5.6)$$

By the same argument applied to y',

$$(k_0^2+\omega^2)^{-1}(\xi,y')$$
$$= (k_0^2+\omega^2)^{-1}(\xi,y') - (k_0^2+\omega_D^2)^{-1}(\xi,y')$$
$$= g(\xi,y')$$
$$= \sum_{\zeta=y'_+,y'_-} (k_0^2+\omega^2)^{-1}(\xi,\zeta) g_\zeta(y'). \qquad (5.7)$$

Combining (5.6) and (5.7),

$$g(x',y') = \sum_{\xi,\zeta} (k_0^2+\omega^2)^{-1}(\xi,\zeta) g_\xi(x') g_\zeta(y'). \qquad (5.8)$$

Now use the identity $\beta^{-1} = \int_0^\infty \exp(-\sigma\beta)\, d\sigma$ to convert resolvents to exponentials in (5.8). Let $h_\xi(x',\sigma) = (-\Delta_D \exp[-\sigma\omega_D^2] l_\xi)(x')$, so that (5.8) becomes

$$g(x',y') = \int d\sigma_1 d\sigma_2 d\sigma_3 \exp(-k_0^2[\sigma_1+\sigma_2+\sigma_3])$$
$$\times \sum_{\xi,\zeta} \exp[-\sigma_2\omega^2](\xi,\zeta) h_\xi(x',\sigma_1) h_\zeta(y',\sigma_3). \qquad (5.9)$$

Substitute (5.9) into (5.3) and evaluate the k_0 integral. Let $d\sigma$ abbreviate $d\sigma_1 d\sigma_2 d\sigma_3$ and set $\mu_{t,u}(\sigma) = |t-u|^{-2}\mu(\sigma/|u-t|^2)$ where μ was defined in Sec. 1. Then

$$E(x,y) = \sum_{\xi,\zeta} \int d\sigma\, \mu_{t,u}(\sigma_1+\sigma_2+\sigma_3) h_\xi(x',\sigma_1)$$
$$\times \exp[-\sigma_2\omega^2](\xi,\zeta) h_\zeta(y',\sigma_3). \qquad (5.10)$$

Let $s_I = 0$ unless $I\times J$ is the support of some barrier, b

say, in which case $s_I = s_b$. The s dependence can be factored out of the hs by noting that $(\exp[-\sigma\omega_D^2]l_\xi)(x') = \exp[-\sigma s_{I_x}]\{\exp[-\sigma(M_0^2-\Delta_D)]l_\xi\}(x')$, when $x' \in I_x$. Let $k_\xi(x',\sigma) = -\{\Delta_D \exp[-\sigma(M_0^2-\Delta_D)]l_\xi\}(x')$, then

$$E(x,y) = \sum_{\xi,\zeta} \int d\sigma\, \mu_{t,u}(\sigma_1+\sigma+\sigma_3)\{(\exp-\sigma_2\omega^2)(\xi,\zeta) \times \exp[-\sigma_1 s_{I_x} - \sigma_3 s_{I_y}]\}\, k_\xi(x',\sigma_1)\, k_\zeta(y',\sigma_3). \quad (5.11)$$

All the s dependence is inside the curly brackets. All the x' dependence is in $k_\xi(x',\sigma_1)$, and the y' dependence is in $k_\zeta(y',\sigma_3)$. (Recall x',y' are localized in I_x, I_y.) Moreover, the part in curly brackets has a path space representation because

$$(\exp-\sigma\omega^2)(\xi,\zeta) = \int dP^\sigma_{\xi,\zeta} \exp\left(-\int_0^\sigma \left(\sum_b s_b\chi_b + M_0^2\right) d\sigma'\right), \quad (5.12)$$

where $dP^\sigma_{\xi,\zeta}$ is the measure of Brownian motion paths starting at ξ and ending at ζ at time σ. The functions χ_b in (5.12) abbreviate $\chi_b(X_{\sigma'},\tau)$ where $X_{\sigma'}$ is the Brownian motion and τ is any time within $[t,u]$. (The χ_b are constant as functions of τ within $[t,u]$.) (5.12) shows that the term in curly brackets and all its derivatives with respect to s are either positive functions of s, or negative functions of s, hence

$$\int ds_\gamma \|E_\gamma\|_{3,\nabla_x\times\nabla_y} \le \sum_{\xi,\zeta} \int d\sigma\, \mu_{t,u}(\sigma_1+\sigma_2+\sigma_3) \times \left|\left\{\int ds_\gamma\, \partial^\gamma \exp(-\sigma_1 s_{I_x} - \sigma_2\omega^2 - \sigma_3 s_{I_y})(\xi,\zeta)\right\}\right| \times \|k_\xi(\sigma_1)\|_{3,I_x}\, \|k_\zeta(\sigma_3)\|_{3,I_y}, \quad (5.13)$$

where $\|\ \|_{3,I}$ denotes the $L^3(I)$ norm. By the fundamental theorem of calculus and (5.12), the term in curly brackets in (5.13) is a path integral over all paths that travel from ξ to ζ in time σ_2 and visit all differentiated bonds on the way, which can be estimated as were the covariances in Ref. 5. The norms in (5.13) are easily estimated by writing $k_\xi(\sigma_1)$, $k_\zeta(\sigma_3)$ in terms of Dirichlet eigenfunctions for the intervals I_x, I_y and using the Titchmarsh theorem which says that the L^3 norm is less than the $L^{3/2}$ norm of the Fourier (and hence Dirichlet) transform. The final result is that the left-hand side of (5.13) is less than

$$\text{const}\ M_0^{-(|\gamma|-2)/q} \exp[-M_0 d(\nabla_x,\nabla_y,\gamma)/2]\ |t-u|^{-\eta}, \quad (5.14)$$

where $\eta > \frac{1}{3}$ and the constant depends on q and η. Thus, choosing $\eta < \frac{1}{2}$ gives (5.2). The $|t-u|^{-\eta}$ comes from terms in (5.13) for which $\xi = \zeta$, and the distance between ξ and ζ via bonds in γ is zero. This concludes the proof of (5.2).

The next step is to incorporate the spinors to obtain, together with an analogous estimate involving W_+, i.e.,

$$\int ds_\gamma \|(\partial^{\gamma_1}\bar{w}_-)E_{\gamma_2}(\partial^{\gamma_3}w_-)\|_{3,\nabla_x\times\nabla_y,2} \le \text{const}\ M_0^{-|\gamma|/q} \exp[-M_0 d(\nabla_x,\nabla_y,\gamma)/2], \quad (5.15)$$

where $\gamma_1\cup\gamma_2\cup\gamma_3 = \gamma$ with γ_1, γ_2, γ_3 disjoint and $\partial^{\gamma_1}\bar{w}_-$, E_{γ_2}, $\partial^{\gamma_3}w_-$ are being used to denote both the kernels given by $\partial^{\gamma_1}w_-$, E_{γ_2}, $\partial^{\gamma_3}w_-$ as functions of the spatial variables, while s,t,u are fixed, and the corresponding operators. (5.15) is proved with the aid of the following lemma.

Lemma: Let $\|T\|^{(p)}_{I,I'}$ denote the operator norm of $T: L^p(I)\to L^p(I')$. If T has spinor indices, $\|\ \|$ includes a sum over the norms of components. Then,

$$\int ds_\gamma \|\partial^\gamma w_-\|^{(p)}_{I,I'} \le \text{const}\ \exp[-M_0' d(\nabla,\nabla',\gamma)], \quad (5.16)$$

where the constant depends only on p $(1<p<\infty)$ and $M_0' \le \frac{9}{10}M_0$; $\nabla = I\times J$, $\nabla' = I'\times J$. A similar lemma holds for $\bar{w}_-$, w_+, $\bar{w}_+$. (5.16) is uniform in all s_b not integrated out, t,u and ϵ.

Proof: $\partial^\gamma w_-$ when restricted to $I\times I'$ becomes the function $\partial^\gamma f(s)$ (see Sec. 1) times the operator of convolution by $\hat{\kappa}_\epsilon \omega^{-1/2}v)^\vee$. The $\partial^\gamma f(s)$ can be taken outside the norm and $\int ds_\gamma$ evaluated by the fundamental theorem of calculus because $|\partial^\gamma f(s)| = \pm\partial^\gamma f(s)$. $\partial^\gamma f(s)$ vanishes if the distance between I and I' is less than $d(\nabla,\nabla',\gamma)$; therefore, without loss of generality, assume the distance between I and I' is greater than $d(\nabla,\nabla',\gamma)$. $\omega^{-1/2}v$ is analytic in p space and the nearest singularity to the real axis is at distance M_0, hence, by a distortion of the contour in evaluating the Fourier transform, it follows that $(\omega^{-1/2}v)^\vee(\xi)$ decays as $\exp(-M_0'|\xi|)$ and is C^∞ away from $\xi = 0$. (5.16) follows except when $d(\nabla,\nabla',\gamma) = 0$, i.e., ∇ and ∇' are contiguous. This case is easily completed by using Mihlin's theorem (Ref. 10, p. 120) to show that as a convolution operator the Fourier transform of $\omega^{-1/2}v$ maps $L^p(\mathbb{R})\to L^p(\mathbb{R})$, $1<p<\infty$, uniformly in M_0.

(5.15) follows by writing $E_{\gamma_2} = \sum_{I,I'} \chi_I E_{\gamma_2}\chi_{I'}$ in (5.15), where χ_I is the characteristic function of interval I. Then use (5.16) and control the sum over I, I' by using some of the exponential decay in (5.16). (A factor $M_0^{-(|\gamma|-1)/q}$ may be included in the right-hand side of (5.16) on replacing M_0' by $M_0'' < M_0'$.)

By Leibniz' rule, (5.15) implies the same estimate for the norm of $\partial^\gamma(w_-Ew_-)$. Therefore, by (5.1), (4.14) can now be proven in the special case where $j(\gamma)$ specifies the same time-localization for the variables x,y, in S_γ by showing that

$$\int ds_\gamma \|(\partial^{\gamma_1}\bar{w}_-)\{\partial^{\gamma_2}\exp[-|t-u|\omega_D(s)]\}(\partial^{\gamma_3}w_-)\|_{3,\nabla_x\times\nabla_y,2} \le \text{const}\ M_0^{-|\gamma|/q} \exp[-M_0 d(\nabla_x,\nabla_y,\gamma)/2], \quad (5.17)$$

together with a similar estimate with w_+. Note that either $\gamma_2 = \phi$ or $|\gamma_2| = 1$.

Proof of (5.17): Use (1.11) to express the exponential in terms of $\exp(-\sigma\omega_D^2)$. As has been done above [see (5.11) and the proof of (5.16)], take the s dependences outside the norm sign and evaluate the s integral by the fundamental theorem of calculus. The problem then reduces to estimating

$$\sum_I \|\chi_{I_x}(\partial^{\gamma_1}\bar{w}_-(0))\chi_I \exp[-\sigma(M_0^2-\Delta_D)] \times \chi_I(\partial^{\gamma_3}w_-(0))\chi_{I_y}\|_3 \quad (5.18)$$

(the sum over I reduces to one term if $\gamma_2 \neq \phi$, the term where $I \times J$ is the support of the $b \in \gamma_2$), where $\partial^{\gamma_1}\overline{w}_-(0)$ denotes the operator obtained by setting $s=0$ in $\partial^{\gamma_1}\overline{w}_-(s)$, and χ_I is the characteristic function of interval I. When $I_x \neq I$ and $I_y \neq I$, (5.18) can be estimated by writing it out in terms of the kernels of the operators and then taking absolute values of the kernels. The kernel of $\exp\sigma\Delta_D$ is less than the kernel of $\exp\sigma\Delta$ which is well-known to be a Gaussian. The kernel of $\partial^{\gamma_1}\overline{w}_-(0)$ is $(\hat{\kappa}_\epsilon \omega^{-1/2}\overline{v})^{\vee}(\xi - \zeta)$ when every barrier $b \in \gamma_1$ intersects the straight line joining (ξ, t) and (ζ, t), zero otherwise. Even when $\gamma_1 = \phi$, the singularity of $(\hat{\kappa}_\epsilon \omega^{-1/2}\overline{v})^{\vee}$ is blunted because the characteristic functions in (5.18) force ξ and η to lie in different squares. Hence it is permissible to take the absolute value and still get estimates uniform in ϵ. Using this information it is fairly easy to complete the estimates for the terms where $I \neq I_x = I_y$.

Consider next the term (if $\neq 0$) in (5.18) for which $I \neq I_x$, $I = I_y$. In this case, use L^p continuity of $\partial^{\gamma_3}w_-(0)$ [see the proof of (5.16)], and positivity to estimate by

$$\text{const}\|\chi_{I_x}(\partial^{\gamma_1}\overline{w}_-(0))\chi_{I_y}\exp[-\sigma(M_0^2 - \Delta)]\chi_{I_y}\|_3, \qquad (5.19)$$

and estimate (5.19) as in the paragraph above. The term $I = I_x \neq I_y$, if $\neq 0$, is treated the same way.

The most interesting term is $I = I_x = I_y$, if there is one. Such a term would be nonvanishing only if $\gamma_1 = \gamma_3 = \phi$. Since every propagator is differentiated at least once, $\gamma_2 \neq \phi$, which forces $I \times J$ to be the support of the differentiated barrier in γ_2. From the proof of (5.16) it can be seen that $w_-(0)$ is continuous from $L^p(\mathbb{R})$ to $L^p(I)$ uniformly in I and ϵ. Hence by taking $\epsilon = 0$, this term can be overestimated by

$$F(\sigma) = \text{const}\ \|\kappa_\epsilon \chi_I \exp[-\sigma(M_0^2 - \Delta)]\chi_I \kappa_\epsilon\|_3. \qquad (5.21)$$

Upon evaluating $\{\int_J dt\, du[\int_0^\infty d\sigma\, \mu_{t,u}(\sigma) F(\sigma)]^2\}^{1/2}$ in order to complete the estimate, it would appear by a power counting type of argument that there is a divergence as $\epsilon \to 0$. The point is that since $I \times J$ is the support of a barrier, there is a small volume factor ϵ which controls this divergence. This concludes the proof of (5.17), and hence (4.14) in this special case.

To obtain (4.14) in general, observe from (1.9) that in general $p(x,y)$ is a product of factors as on the left-hand side of (5.1) and construct decompositions analogous to (5.1) and (5.11). The estimate analogous to (5.17) is easier because the t, u integrals are more convergent when t, u are localized in different intervals.

Note added in proof: Cluster expansions for Yukawa[2] have recently been established using the method of integrating out fermions by J. Magnen and R. Seneor, preprint, Centre de Physique Théorique Ecole Polytechnique and independently by A. Cooper and L. Rosen, preprint, Department of Mathematics, University of Toronto.

APPENDIX: THE YUKAWA MODEL

When ϕ is a boson field, there are also functional derivatives with respect to ϕ coming from differentiations of the boson measure, and the integral with respect to this measure has to be performed after removing the fermions by means of the $\|\ \|_D$ norm. The procedure given in Sec. 4 fails in two serious ways: (1) the estimate (3.1) diverges logarithmically in ϵ; (2) (4.7) and (4.8) are no good because $|\phi|$ does not exist. The way around the first difficulty is to use an expansion as in Ref. 11 to lower the cutoff in the exponent. As in Ref. 11, certain divergent loops coming from the expansion have to be renormalized. To do this, a simple modification of the techniques in Sec. 5 gives control over $S(x,y,s) - S(x,y,0)$, when $x, y \in \nabla$, and ∇ is not the support of a barrier. When ∇ is the support of a barrier it is not necessary to renormalize because of the small volume of the barrier.

The other difficulty (2) may be circumvented by majorizing the norm by Hilbert Schmidt norms in a different way from (4.7) so that the majority of the ϕs are "padded" by propagators, for example, the Hilbert Schmidt norm of the operator whose kernel is $x_i = (x_i', t_i)$,

$$\int_{\substack{x_i \in \nabla_i \\ i=2,\ldots,n-1}} dx_2 \cdots dx_{n-1} S_{\gamma_1}(x_1, x_2)\,\phi(x_2) S_{\gamma_2}(x_2, x_3) \times \phi(x_3) \cdots \phi(x_{n-1}) S_{\gamma_{n-1}}(x_{n-1}, x_n) \qquad \text{(A1)}$$

at fixed $t_i, \ldots, t_n$, is quite respectable. It is the square root of a boson field polynomial and can easily be estimated by using a Cauchy Schwarz estimate with respect to the boson measure to remove the square root and then using Wick's theorem to evaluate the boson integral. The resulting estimate can be expressed in terms of L^3 norms of the boson and fermion propagators. There are not enough propagators to group every boson field into a kernel like (A1) but it can be arranged that the remaining ones have cutoffs so that their L^3 norms do exist. To see how this is done, see Ref. 11 where a similar problem arises. The idea is that one can contract ψs (the analog to integration by parts in boson theories) after the expansion to lower the cutoff in the exponent. Any ψs left are attached to vertices with a low cutoff because they arose from contractions to the exponent.

*Junior Fellow, Michigan Society of Fellows.

[1]P. Federbush, University of Michigan Preprints.
[2]O. McBryan, Rockefeller University Preprint.
[3]E. Seiler, Commun. Math. Phys. **42**, 163 (1975).
[4]E. Seiler and B. Simon, Princeton University Preprint.
[5]J. Glimm, A. Jaffe, and T. Spencer, *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman, Lecture Notes in Physics Series, No. 25 (Springer Verlag, New York, 1973).
[6]J. Glimm and A. Jaffe, in *Statistical Mechanics and Quantum Field Theory*, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1972).
[7]J. Glimm, Commun. Math. Phys. **5**, 343 (1967); **6**, 61 (1967).
[8]P. Federbush, J. Math. Phys. **14**, 1532 (1973).
[9]R. Schrader, Ann. Phys. **70**, 412 (1972).
[10]L. Hormander, Act. Math. **104**, 93 (1960).
[11]D. Brydges, "Boundedness Below for Fermion Model Theories, Part II: The Linear Lower Bound," Commun. Math. Phys. (to be published).

Canonical symmetrization for unitary bases. I. Canonical Weyl bases*

Chris W. Patterson and William G. Harter

Instituto de Fisica, Departamento de Eletronica Quantica, Universidade Estadual de Campinas, Caixa Postal 1170 Campinas, S.P. Brazil
(Received 24 September 1975)

It is shown that the Weyl basis formed by the canonical symmetrization of an n-dimensional, p-rank tensor space with canonical projection operators of S_p is a Gel'fand basis of $U(n)$. This basis may easily be generated using standard projection operator techniques.

INTRODUCTION

It is well known that an n-dimensional, p-rank tensor space forms a reducible basis for both $U(n)$, the unitary group of n dimensions, and S_p, the symmetric group of p objects. The irreducible bases of the symmetric group formed from this tensor space (the symmetrized bases) can be used to reduce the bases of $U(n)$ and $SU(n)$, the unimodular subgroup of $U(n)$. The resultant irreducible bases of $U(n)$ and $SU(n)$ are called Weyl bases. Weyl generally avoided the explicit construction of these bases while using them to enumerate the bases of the irreducible representations (IR's) of $U(n)$ and to determine the characters of these IR's.[1]

However, works by Biedenharn, Baird, Ciftan, and Louck[2–6] have shown the need for an explicit construction of the irreducible bases of $U(n)$ in order to find the matrix elements of the IR's of $U(n)$ and the Clebsch–Gordon coefficients associated with the direct products of these IR's. Up to now this has been accomplished using a boson basis developed by Schwinger[7] for $U(2)$ and extended to $U(n)$ by Biedenharn and Louck.[8] For certain applications this basis is more complicated than necessary. For example, work is presently being done in atomic and molecular orbital theory using a unitary basis to describe complex atomic and molecular configurations in a way that simplifies the evaluation of matrix elements of spin and orbital operators.[9–11] The explicit construction of a Weyl basis for $U(n)$ using the symmetrization techniques described in this paper has been instrumental in this simplification. A number of tableau "tricks" have appeared which make the use of a Weyl basis much more elegant and convenient. Already, the theory of permutation operators acting on a canonical Weyl basis has led to tableau algorithms for directly relating Slater determinant states to those states having separate spin and orbital parts and definite total angular momentum.[12] We expect more applications will be found in the future.

Attempts to construct a Weyl basis have been made in the past, notably by Littlewood.[13] Recent work has been done by Tompkins,[14] Sullivan,[15] and Lezuo.[16] Lezuo has succeeded in constructing a canonical Weyl basis for $SU(2)$ and $SU(3)$. His techniques have been simplified and extended in this work to yield a canonical Weyl basis for all $U(n)$ and $SU(n)$ as he himself had predicted. The success of the present method is a result of the choice of projection operators of S_p used to symmetrize the tensor space. Previous dependence upon Young symmetrizers to project out bases, as outlined by Hammermesh,[17] led to difficulties in the orthogonalization of the bases. These difficulties even created some doubt as to the possibility of ever constructing Weyl bases with projection operators.[18] The present choice of *canonical* projection operators of S_p not only makes orthogonality of the bases automatic, but also creates a *canonical* basis of $U(n)$ or a Gel'fand basis.

In Sec. I we review the relationship between the canonical irreducible bases of $U(n)$ and S_p. In particular, we show that the canonical bases of $U(n)$ are specified by the eigenvalues of the complete set of Gel'fand operators I_r^l for $r=1,2,\ldots,l$ and $l=1,2,\ldots,n$. Similarly, the canonical bases of S_p are specified by the eigenvalues of the complete set of r-cycle class operators K_r^l for $r=1,2,\ldots,l$ and $l=1,2,\ldots,p$.

In Sec. II we show that the Gel'fand invariant operators of $U(n)$ may be expanded in terms of the r-cycle class operators. The canonically symmetrized tensor basis transforms like a canonical basis of S_p under two different mutually commuting permutation operators which we call the particle and state permutations. In Sec. III we canonically symmetrize our tensor basis with *particle* projection operators of S_p. When acting on this basis, we show that the Gel'fand operators may be expanded in terms of the *state* r-cycle class operators of S_p. Furthermore, our symmetrized basis is an eigenfunction of all these *state* r-cycle class operators. Thus, we form canonical Weyl bases of $U(n)$ or Gel'fand bases from a canonically symmetrized tensor space. Since the canonical invariant operators of $SU(n)$ depend on the invariant operators of $U(n)$, we also form canonical bases of $SU(n)$.

In the following work (Paper II), we shall show that the boson basis is a generalization of the Weyl basis and, as a result, may be generated simply and straightforwardly using symmetrization operators. The boson basis is frequently called a "Weyl basis" even though it is constructed from a different tensor space than Weyl originally considered. Since we are constructing a basis from the same tensor space used by Weyl which may be used to generate a boson basis, we feel justified in adopting the name "Weyl basis."

I. REVIEW

A. Canonical irreducible bases of S_p

The symmetric group S_p consists of all $p!$ permutations of p objects. The IR's of S_p are labeled by the partitions $[u]=[u_1u_2\cdots u_p]$,[19] where $u_i \geq u_{i+1}$, $u_i>0$ and $\sum_{i=1}^{p} u_i = p$. There are u_i boxes in row i of the partition $[u]$.

The matrix elements of the IR's of S_p can always be put in real form so that the IR's are orthogonal matrices. The IR's of S_p obey the standard orthonormality and completeness relations of group theory:

$$\frac{l^{[u]}}{p!}\sum_{i,j,[u]} D^{[u]}_{ij}(q)\, D^{[u]}_{ij}(q') = \delta_{(q)(q')}, \tag{1.1}$$

$$\frac{l^{[u]}}{p!}\sum_{q\in S_p} D^{[u]}_{ij}(q)\, D^{[u']}_{i'j'}(q) = \delta_{ii'}\,\delta_{jj'}\,\delta_{[u][u']}, \tag{1.2}$$

where $l^{[u]}$ is the dimension of the IR $[u]$ of S_p.

Permutations with the same cyclic structure belong to the same class of S_p. For the Lth class consisting of l_1 one-cycles, l_2 two-cycles, ..., and l_p p-cycles where $\sum_{i=1}^{p} i(l_i)=p$ we write $K_L = K_{1^{l_1}2^{l_2}\cdots p^{l_p}}$. Defining N_L to be the order of the Lth class, we have[20]

$$N_L = \frac{p!}{1^{l_1}l_1!\, 2^{l_2}l_2! \cdots p^{l_p}l_p!}. \tag{1.3}$$

The character of K_L is defined by

$$\chi^{[u]}_L = \mathrm{Tr} D^{[u]}(q) \tag{1.4}$$

for all $q\in K_L$. A special case of (1.4) is

$$\chi^{[u]}_{1^p} = l^{[u]} = \mathrm{Tr} D^{[u]}(1). \tag{1.5}$$

A method of finding the dimension of IR $[u]$ using hook-lengths[21] is shown in Fig. 1.

The characters of S_p obey the standard orthonormality and completeness relations of group theory:

$$\frac{1}{p!}\sum_{[u]} N_L\, \chi^{[u]}_L \chi^{[u]}_{L'} = \delta_{LL'}, \tag{1.6}$$

$$\frac{1}{p!}\sum_{L} N_L \chi^{[u]}_L \chi^{[u']}_L = \delta_{[u][u']}. \tag{1.7}$$

We now define the projection operators of S_p by the relation:

$$P^{[u]}_{ij} = \frac{l^{[u]}}{p!}\sum_{t\in S_p} D^{[u]}_{ij}(t)\ (t). \tag{1.8}$$

Then it follows that

$$\begin{aligned}(q)\, P^{[u]}_{ij} &= \frac{l^{[u]}}{p!}\sum_{t\in S_p} D^{[u]}_{ij}(t)\ (q)(t),\\ &= \frac{l^{[u]}}{p!}\sum_{q'\in S_p} D^{[u]}_{ij}(q^{-1}q')\ (q'),\\ &= \frac{l^{[u]}}{p!}\sum_{i'}\sum_{q'\in S_p} D^{[u]}_{ii'}(q^{-1})\, D^{[u]}_{i'j}(q')\ (q').\end{aligned}$$

Using the orthogonality of the matrices,

$$D^{[u]}_{ii'}(q^{-1}) = D^{[u]}_{i'i}(q), \tag{1.9}$$

we have

$$(q)\, P^{[u]}_{ij} = \sum_{i'} P^{[u]}_{i'j} D^{[u]}_{i'i}(q). \tag{1.10}$$

Similarly,

$$P^{[u]}_{ij}(q) = \sum_{j'} P^{[u]}_{ij'} D^{[u]}_{jj'}(q). \tag{1.11}$$

Furthermore, the projection operators are orthogonal as seen by using (1.2) and (1.10),

$$\begin{aligned}P^{[u]}_{ij}P^{[v]}_{kl} &= \frac{l^{[u]}}{p!}\sum_{q\in S_p} D^{[u]}_{ij}(q)\ (q)P^{[v]}_{kl}\\ &= \delta_{jk}\delta_{[u][v]}\, P^{[v]}_{il}.\end{aligned} \tag{1.12}$$

Now consider the following idempotent:

$$P^{[u]} = \sum_i P^{[u]}_{ii}. \tag{1.13}$$

We may expand $P^{[u]}$ in terms of the classes of S_p. We have

$$\begin{aligned}P^{[u]} &= \frac{l^{[u]}}{p!}\sum_i \sum_{q\in S_p} D^{[u]}_{ii}(q)\ (q)\\ &= \frac{l^{[u]}}{p!}\sum_{q\in S_p} \mathrm{Tr} D^{[u]}(q)\ (q),\end{aligned}$$

or

$$P^{[u]} = \frac{l^{[u]}}{p!}\sum_L \chi^{[u]}_L K_L. \tag{1.14}$$

We may also expand the classes of S_p in terms of the idempotents. Using (1.14) and (1.6) we have

$$\frac{1}{p!}\sum_{[u]}\sum_L N_{L'}\, \chi^{[u]}_{L'}\chi^{[u]}_L K_L = \sum_{[u]} \frac{N^{[u]}_{L'}\chi^{[u]}_{L'}P^{[u]}}{l^{[u]}},$$

or

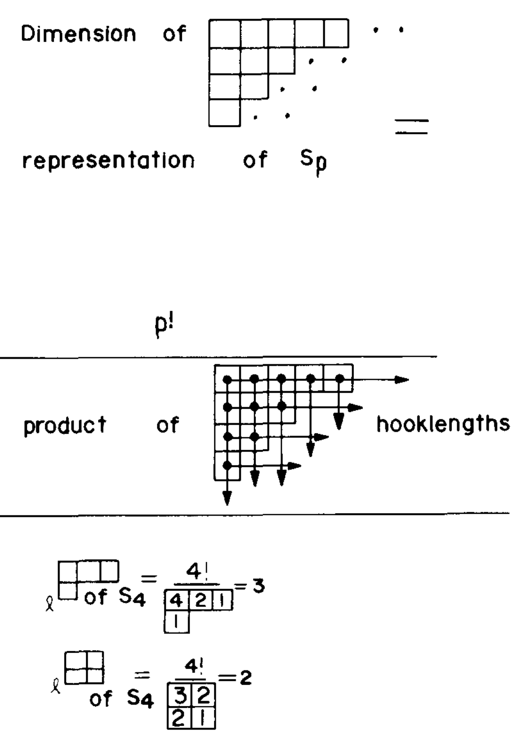


FIG. 1. Dimension of an IR of S_p. The Robinson formula for the dimension $l^{[u]}$ of an IR of S_p labeled by a partition $[u]$ is given above. A hook-length of a box in a partition is simply the number of boxes below it, to the right of it, and itself. The denominator in the dimension formula is the numerical product of all the hook-lengths of the boxes. Examples for S_4 are given below.

$$K_L = \sum_{[u]} \frac{N_L \chi_L^{[u]} P^{[u]}}{l^{[u]}}. \tag{1.15}$$

We can now show that the irreducible bases of S_p are eigenvectors of the classes of S_p. Let $|{}^{[u]}_{i}\rangle$ be an irreducible basis of the IR $[u]$ of S_p. By definition we have

$$P^{[v]} |{}^{[u]}_{i}\rangle = \sum_{i'} |{}^{[u]}_{i'}\rangle D^{[u]}_{i'i}(P^{[v]}).$$

From (1.13), (1.2), and (1.8) it follows that

$$P^{[v]} |{}^{[u]}_{i}\rangle = |{}^{[u]}_{i}\rangle \delta_{[u][v]}. \tag{1.16}$$

Operating on the irreducible basis with class K_L and using (1.15) and (1.16), we have

$$K_L |{}^{[u]}_{i}\rangle = \frac{N_L \chi_L^{[u]}}{l^{[u]}} |{}^{[u]}_{i}\rangle, \tag{1.17}$$

so the bases of the IR $[u]$ of S_p are simultaneous eigenvectors of the class operators of S_p with eigenvalues $(N_L \chi_L^{[u]})/l^{[u]}$.

The simultaneous eigenvalues of the class operators completely determine the IR's of S_p. If

$$\frac{N_L \chi_L^{[u]}}{l^{[u]}} = \frac{N_L \chi_L^{[v]}}{l^{[v]}}$$

for all L, then

$$\frac{1}{p!} \sum_L \frac{N_L \chi_L^{[u]} \chi_L^{[u]}}{(l^{[u]})^2} = \frac{1}{p!} \sum_L \frac{N_L \chi_L^{[v]} \chi_L^{[v]}}{(l^{[v]})^2},$$

and from (1.7) it follows that $l^{[u]} = l^{[v]}$. From our initial assumption we have that $\chi_L^{[u]} = \chi_L^{[v]}$ for all L. This is precisely the criteria for the two IR's $[u]$ and $[v]$ of S_p to be equivalent. However, since we only need to specify the p rows $u_1, u_2, \ldots, u_p$ to determine the IR $[u]$ of S_p, not all of the class operators are independent. Kramer[22] has shown that the eigenvalues of the r-cycle class operator, $K_r \equiv K_{1^{p-r}r^1}$ for $r = 2, 3, \ldots, p$, uniquely determine the IR's of S_p. Note that we need the $p-1$ independent operators K_r and the condition, $p = \sum_{i=1}^{p} u_i$, to determine the p unknowns $u_1, u_2, \ldots, u_p$. For this reason we adjoin to the r-cycle class operators the operator $K_1 = p$, and adopt the notation

$$K_r^p = K_r \quad \text{for } r = 2, 3, \ldots, p,$$

$$K_1^p = K_1 = p.$$

This notation makes explicit the fact that the eigenvalue p of K_1^p determines the group S_p to which the r-cycle class operators belong.

A canonical irreducible basis of S_p is defined by means of the subgroup reduction $S_p \supset S_{p-1} \supset \cdots \supset S_1$ such that an irreducible canonical basis of S_p is also an irreducible basis for all subgroups in this reduction. Thus a canonical irreducible basis of S_p is an eigenvector of all sets of class operators of $S_p, S_{p-1}, \ldots, S_1$:

$$\begin{matrix} K_p^p \; K_{p-1}^p \cdots K_1^p \\ K_{p-1}^{p-1} \cdots K_1^{p-1} \\ \ddots \quad \vdots \\ K_1^1 \end{matrix} \quad .$$

The lth row of r-cycle class operators above completely specifies an IR $[u]^l$ of S_l so that a canonical irreducible basis $|{}^{[u]}_{(r)}\rangle$ of S_p may be represented by means of a succession of partitions:

$$\begin{smallmatrix}[u]\\(r)\end{smallmatrix} = \overset{[u]}{\begin{pmatrix} [u]_r^{p-1} \\ \vdots \\ [u]_r^1 \end{pmatrix}} = \overset{[u_{1p} u_{2p} \cdots u_{pp}]}{\begin{pmatrix} [u_{1p-1} \cdots u_{p-1\,p-1}] \\ \vdots \\ [u_{11}] \end{pmatrix}}, \tag{1.18}$$

where $\sum_{i=1}^{m} u_{im} = m$ for $m = 1, 2, \ldots, p$, and $[u]^p = [u]$. We call such a succession of partitions a standard pattern of S_p. The fact that this pattern uniquely determines a canonical irreducible basis of S_p is inherent in the branching law of S_p which specifies further that[23]

$$u_{i\,m+1} \geq u_{im} \geq u_{i+1\,m+1}. \tag{1.19}$$

The standard pattern of S_p may be pictured by means of a standard tableau[24] $T^{[u]}_{(r)}$ of S_p. The standard tableau $T^{[u]}_{(r)}$ is the partition $[u]$ of S_p numbered with $1, 2, \ldots, p$ in the boxes such that the numbers increase to the right in the rows and down in the columns with no numbers repeated. There are $l^{[u]}$ such standard tableaus of $[u]$. For example, when $[u] = [210]$, then $l^{[u]} = 2$, and we have the two standard tableaus $\begin{smallmatrix}1&2\\3&\end{smallmatrix}$ and $\begin{smallmatrix}1&3\\2&\end{smallmatrix}$. For typographic convenience we have omitted the boxes surrounding the numbers in the standard tableaus.

Let $T^{[u]^m}_{(r)}$ be the standard tableau remaining after removing boxes with numbers $m+1, m+2, \ldots, p$ from $T^{[u]}_{(r)}$ and let $[u]_r^m$ be the corresponding partition remaining. The rows u_{im} of partition $[u]_r^m$ obey conditions (1.19) of the branching law of S_p so that we may associate the standard tableau $T^{[u]}_{(r)}$ with the standard pattern $\begin{smallmatrix}[u]\\(r)\end{smallmatrix}$ of S_p. The standard tableau associated with a canonical irreducible basis uniquely determines the reduction of the basis under all subgroups $S_p, S_{p-1}, \ldots, S_1$ as illustrated in Fig. 2.

We now give a prescription for finding the seminormal canonical projection operators $O^{[u]}_{rs}$ of S_p defined such that

$$O^{[u]}_{rs} = C^{[u]}_{rs} P^{[u]}_{rs} \tag{1.20}$$

where $C^{[u]}_{rs}$ is some positive constant and $P^{[u]}_{rs}$ is a canonical projection operator constructed from a canonical IR of S_p.[25] We will show later that the $O^{[u]}_{rs}$ can project out a canonical irreducible basis of S_p. Let $T^{[u]}_{(r)} = \sigma_{rs} T^{[u]}_{(s)}$ where σ_{rs} permutes the numbers in the standard tableau $T^{[u]}_{(r)}$ of S_p. For example,

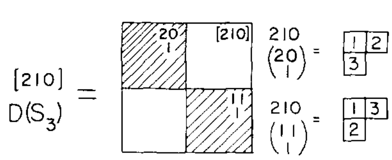


FIG. 2. Canonical reduction for S_p. The IR [210] of S_3 reduces to the IR's [20] and [11] of S_2 as shown. As a representation of S_2 or S_1 it is diagonal. Each basis of [210] corresponds to a diagonal component with a unique genealogy traced by the standard patterns or standard tableaus shown on the right.

$\begin{smallmatrix}1&2\\3\end{smallmatrix} = (23)\begin{smallmatrix}1&3\\2\end{smallmatrix}.$

Now let P_r^m be the symmetrization operator of the numbers in the rows of $\mathbf{T}_{(r)}^{[u]^m}$ and N_r^m be the antisymmetrization operator of the numbers in the columns of $\mathbf{T}_{(r)}^{[u]^m}$. In the example above we have

$$P^3_{\begin{smallmatrix}12\\3\end{smallmatrix}} = S_{12} = (1) + (12),$$

$$N^3_{\begin{smallmatrix}12\\3\end{smallmatrix}} = A_{12} = (1) - (13).$$

If we define

$$E_{rs}^p = P_r^p \sigma_{rs} N_s^p. \tag{1.21}$$

then

$$O_{rs}^{[u]} = O_{rs}^{[u]^p} = O_{rr}^{[u]^{p-1}} E_{rs}^p O_{ss}^{[u]^{p-1}}, \tag{1.22}$$

where $O^{[u]^1} = (1)$. It is interesting to note that we obtain the same $O_{rs}^{[u]}$ if we define $E_{rs}^p = N_r^p \sigma_{rs} P_s^p$. Continuing with our example, we have $E^2_{\begin{smallmatrix}12\\3\end{smallmatrix}\begin{smallmatrix}12\\3\end{smallmatrix}} = S_{12}$, $E^2_{\begin{smallmatrix}13\\2\end{smallmatrix}\begin{smallmatrix}13\\2\end{smallmatrix}} = A_{12}$, and $E^3_{\begin{smallmatrix}12\\3\end{smallmatrix}\begin{smallmatrix}13\\2\end{smallmatrix}} = S_{12}(23)A_{12}$ so that

$$\begin{aligned} O^{[210]}_{\begin{smallmatrix}12\\3\end{smallmatrix}\begin{smallmatrix}13\\2\end{smallmatrix}} &= S_{12}S_{12}(23)A_{12}A_{12} \\ &= 4S_{12}(23)A_{12} \\ &= 4((1)+(12))(23)((1)-(12)) \\ &= 4((23)-(13)+(132)-(123)). \end{aligned}$$

Let $[\tilde{u}]$ be the partition conjugate to $[u]$ formed by exchanging rows and columns of $[u]$. Similarly, $\mathbf{T}_{(\tilde{r})}^{[\tilde{u}]}$ is the conjugate standard tableau formed by exchanging rows and columns of $\mathbf{T}_{(r)}^{[u]}$. For example,

$$\widetilde{\begin{smallmatrix}123\\4\end{smallmatrix}} = \begin{smallmatrix}14\\2\\3\end{smallmatrix}.$$

Also let ϵ_q be 1 or -1 if permutation (q) is an even or odd product of bicycles respectively. We see that $\epsilon_{\sigma_{rs}} = \epsilon_{\sigma_{\tilde{r}\tilde{s}}}$ and that $O_{\tilde{r}\tilde{s}}^{[\tilde{u}]}$ may be found from $O_{rs}^{[u]}$ simply by exchanging symmetrization and antisymmetrization operators ($A \leftrightarrow S$). It follows that the coefficient of (q) in $O_{\tilde{r}\tilde{s}}^{[\tilde{u}]}$ is simply the coefficient of (q) in $O_{rs}^{[u]}$ multiplied by the factor $\epsilon_q \epsilon_{\sigma_{rs}}$. From (1.20) and the definition of the projection operators (1.8), we have the relation

$$D_{\tilde{r}\tilde{s}}^{[\tilde{u}]}(q) = \epsilon_{\sigma_{rs}} \epsilon_q D_{rs}^{[u]}(q) \tag{1.23}$$

for the canonical IR's of S_p.

B. Canonical irreducible bases of U(n) and SU(n)

The unitary group U(n) consists of the set of all n-dimensional unitary matrices. The unimodular unitary group SU(n) is the set of all n-dimensional unitary matrices with unimodular determinant (determinant equals one) and is a subgroup of U(n). The set U(n) and SU(n) form an IR of themselves which is called the self or fundamental representation.

The generators E_{ij} for $i,j = 1, 2, \ldots, n$ of U(n) obey the commutation relation[26]

$$[E_{ij}, E_{kl}] = E_{il}\delta_{jk} - E_{kj}\delta_{il}. \tag{1.24}$$

where

$$E_{ij}^\dagger = E_{ji}. \tag{1.25}$$

The generators E'_{ij} of SU(n) will obey the unimodular condition

$$\mathrm{Tr} E'_{ij} = 0, \tag{1.26}$$

if we let

$$E'_{ij} = E_{ij} \quad \text{for } i \neq j, \tag{1.27}$$

$$H_i = E'_{ii} = \frac{1}{\sqrt{i(i+1)}}\,(E_{11} + E_{22} + \cdots + E_{ii} - iE_{i+1\,i+1}) \qquad \text{for } i = 1, 2, \ldots, n-1.$$

The IR's of U(n) and SU(n) are labeled by partitions $[M]$ with the number of rows no greater than n:[27]

$$[M] = [m_{1n}\, m_{2n} \cdots m_{nn}].$$

One of the first problems is to find a complete and independent set of mutually commuting Hermitian invariant operators constructed from the generators E_{ij} or E'_{ij} which uniquely specify the IR of U(n) or SU(n) respectively. By invariant operators we mean that the operators commute with all the generators of the group.

Such a set of invariant operators has been found by Gel'fand[28] for U(n):

$$I_k^n = \sum_{i_1, i_2, \ldots, i_k}^{n} E_{i_1 i_2} E_{i_2 i_3} \cdots E_{i_k i_1} \qquad \text{for } k = 2, 3, \ldots, n, \tag{1.28}$$

$$I_1^n = \sum_{i_1}^{n} E_{i_1 i_1}.$$

From (1.25) it follows that the I_k^n are Hermitian and from the commutation relations (1.24) it is straightforward to show that

$$[E_{ij}, I_k^n] = 0 \tag{1.29}$$

for $k = 1, 2, \ldots, n$. Thus the operators I_k^n are invariant and mutually commuting. It can also be shown that the operators I_k^n are independent and complete, i.e., the eigenvalues of these operators uniquely specify the IR $[M]$ of U(n) and the eigenvectors of these operators form an orthogonal irreducible basis of $[M]$. Note that n operators are necessary to specify the n rows of partition $[M]$.

For SU(n) the dependence of the invariant operators $I_k'^n$ for $k = 2, 3, \ldots, n$ upon the generators E'_{ij} is more complicated. Such a set of mutually commuting Hermitian invariant operators has been found by Popov and Perelomov[29]:

$$I_k'^n = \sum_{r=0}^{k} \binom{k}{r} \left(-\frac{I_1^n}{n}\right)^{k-r} I_r^n, \quad I_0^n = n. \tag{1.30}$$

This is one way to prove that an irreducible basis of U(n) is also an irreducible basis of SU(n). Hence, the bases of U(n) and SU(n) may be simultaneously specified by the same partition as we have indicated. There are only $n-1$ operators needed to specify the partition $[M]$, because for SU(n) we have the equivalence of IR's[30]:

$$[m_{1n}\, m_{2n} \cdots m_{nn}] = [m_{1n} - m_{nn}\;\, m_{2n} - m_{nn} \cdots m_{n-1\,n} - m_{nn}\;\, 0], \tag{1.31}$$

or

$$[M] = [M']$$

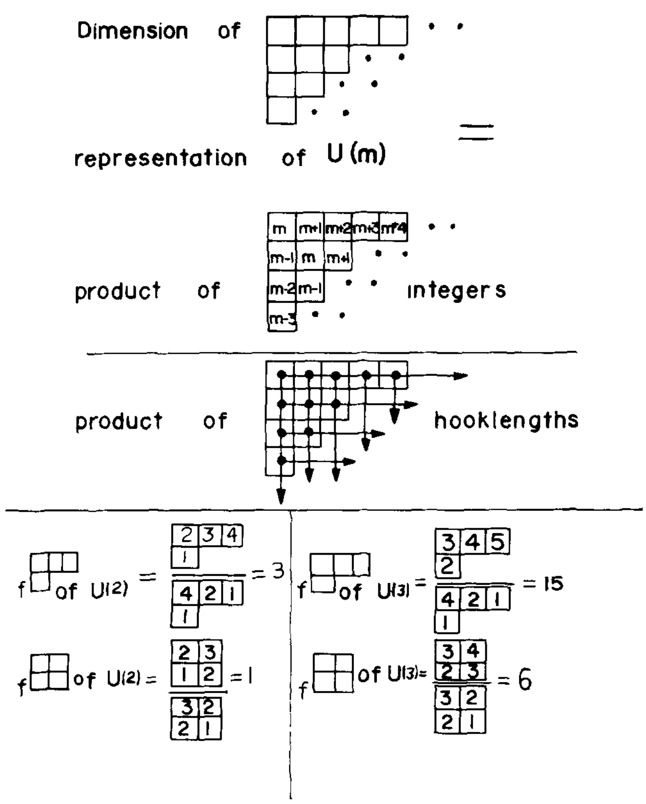


FIG. 3. Dimension of an IR of $U(m)$. The Robinson formula for the dimension $f^{[M]}$ of an IR of $U(m)$ labeled by a partition $[M]$ is given above. Examples for $U(2)$ and $U(3)$ are given below.

One may remove all the columns with n boxes from the partition and obtain the same IR of $SU(n)$.

A canonical irreducible basis of $U(n)$ is defined by means of the subgroup reduction $U(n) \supset U(n-1) \supset \cdots \supset U(1)$ such that an irreducible canonical basis of $U(n)$ is also an irreducible basis for all subgroups in this reduction. Thus, a canonical irreducible basis of $U(n)$ is an eigenvector of all the sets of invariant operators of $U(n), U(n-1), \ldots, U(1)$:

$$\begin{array}{l} I_n^n I_{n-1}^n \cdots I_1^n \\ I_{n-1}^{n-1} \cdots I_1^{n-1} \\ \quad \ddots \quad \vdots \\ \qquad I_1^1 \end{array} .$$

The lth row uniquely specifies an IR of $U(l)$ so that a canonical irreducible basis $|{}^{[M]}_{(m)}\rangle$ of $U(n)$ may be represented by means of a succession of partitions

$$ {}^{[M]}_{(m)} = \left\langle \begin{array}{c} [m_{1n}\, m_{2n} \cdots m_{nn}] \\ [m_{1n-1} \cdots m_{n-1\,n-1}] \\ \cdots \\ [m_{11}] \end{array} \right\rangle . \tag{1.32}$$

The pattern above is called a Gel'fand pattern[31] or a standard pattern of $U(n)$. The fact that this pattern uniquely determines a canonical irreducible basis of $U(n)$ is inherent in the Weyl branching law[32] which specifies further that

$$m_{i\,j+1} \geq m_{ij} \geq m_{i+1\,j+1}. \tag{1.33}$$

There are $f^{[M]}$ such patterns, where $f^{[M]}$ is the dimension of the IR $[M]$ of $U(n)$. A method of finding the dimension $f^{[M]}$ using hook-lengths is shown in Fig. 3. If we let $\lambda_l = p_l - p_{l-1}$ where $p_l = \sum_{i=1}^{l} m_{il}$, then $(\lambda) = (\lambda_1 \lambda_2 \cdots \lambda_n)$ is called the weight of the standard pattern ${}^{[M]}_{(m)}$.

The standard pattern of $U(n)$ may be pictured by means of a standard tableau $T^{[M]}_{(m)}$ of $U(n)$. The standard tableau $T^{[M]}_{(m)}$ is the partition $[M]$ of $U(n)$ numbered with $1, 2, \ldots, n$ in the boxes such that the numbers are nondecreasing to the right in the rows and are increasing down in the columns. There may be repeated numbers in the rows but not in the columns. When there are no repeated numbers in the rows of $T^{[M]}_{(m)}$ of $U(n)$ it will be equivalent to some $T^{[M]}_{(m)}$ of S_n. There are $f^{[M]}$ standard tableaus of $U(n)$ corresponding to IR $[M]$. For example, when $[M]=[210]$, $f^{[M]}=8$, and we have the eight

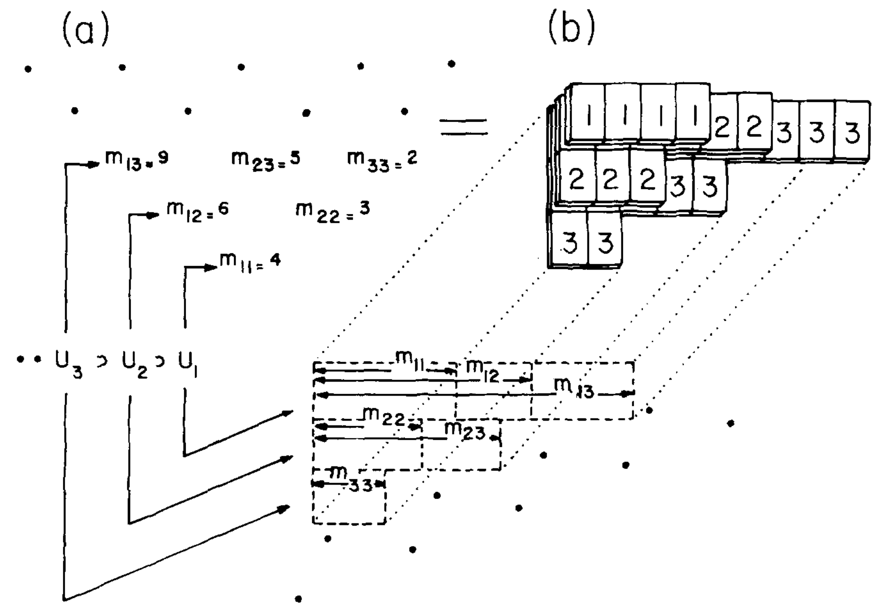


FIG. 4. Labeling for a canonical unitary basis. (a) Standard or Gel'fand pattern of $U(n)$. The lth row of integers $[m_{1l} m_{2l} \cdots m_{ll}]$ specifies to which IR of $U(l)$ the basis belongs. The $(l-1)$th row specifies to which IR of $U(l-1)$ the IR of $U(l)$ reduces. In this way each basis of $U(n)$ has a unique genealogy chain and labeling. (b) Standard tableau of $U(n)$. The standard pattern of $U(n)$ may be pictured by means of a standard tableau. (When labeled algebraically, it is just an upside down standard pattern.)

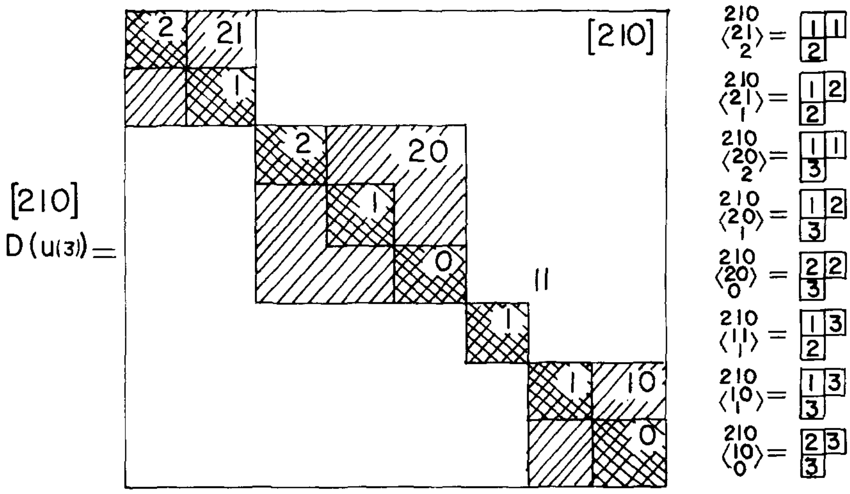


FIG. 5. Canonical reduction for U(n). The IR [210] of U(3) reduces to the IR's [21], [20], [11], and [10] of U(2) as shown. As a representation of U(1) it is diagonal. Each basis of [210] corresponds to a diagonal component with a unique genealogy traced by the standard patterns or standard tableaus shown on the right.

standard tableaus $^{11}_{2}$, $^{12}_{2}$, $^{11}_{3}$, $^{12}_{3}$, $^{22}_{3}$, $^{13}_{2}$, $^{13}_{3}$, and $^{23}_{3}$.

Let the standard tableau $T^{[M]}_{\langle m\rangle}$ contain λ_l numbers l and let $[M]^{p_l}_m$ be the partition remaining after removing boxes with the number $l+1, l+2, \ldots, n$ from $T^{[M]}_{\langle m\rangle}$. We see that the rows m_{il} of the partition $[M]^{p_l}_m$ must obey conditions (1.33) of the branching law of U(n) so that we may associate the standard tableau $T^{[M]}_{\langle m\rangle}$ with the standard pattern $^{[M]}_{\langle m\rangle}$ as shown in Fig. 4. The standard tableau associated with a canonical irreducible basis uniquely determines the reduction of the basis under all subgroups U(n), U($n-1$), ..., U(1) as illustrated in Fig. 5.

It is impossible to completely specify an irreducible representation of SU(n) by means of the subgroup reduction $SU(n)\supset SU(n-1)\supset\cdots\supset SU(1)$ as one might expect.[33] Instead one uses the canonical reduction $SU(n)\supset U_{n-1}(1)\times SU(n-1)\supset U_{n-2}(1)\times SU(n-2)\supset\cdots\supset U_2(1)\times SU(2)\supset U_1(1)$, where the generator for $U_l(1)$ is H_l. Note that H_l commutes with all generators E_{ij} of SU(l) as required for the direct product $U_l(1)\times SU(l)$. H_l is also the only invariant operator of $U_l(1)$, so the complete set of invariant operators for $U_l(1)\times SU(l)$ is H_l and I'^{l}_k for $k=2,3,\ldots,l$. A canonical irreducible basis of SU(n) is then an eigenvector of all the sets of invariant operators:

$$
\begin{array}{l}
I'^n_n\, I'^n_{n-1}\cdots I'^n_2\; 0\\
I'^{n-1}_{n-1}\cdots I'^{n-1}_2\, H_{n-1}\\
\quad\vdots\\
\quad I'^2_2\; H_2\\
\quad\; H_1
\end{array}
$$

Because of the relations (1.27) and (1.30), a canonical irreducible basis of SU(n) is specified by the simultaneous eigenvalues of the invariant operators of U(n) within the equivalence

$$\left|{}^{[M]}_{\langle m\rangle}\right\rangle = \left|{}^{[M']}_{\langle m\rangle}\right\rangle. \qquad (1.34)$$

II. U(n) INVARIANT OPERATORS AND S_p CLASS OPERATORS

A. Unitary invariant operators

If $f(h_1, h_2, \ldots, h_n)$ is any symmetric polynomial function of its arguments $(h)=(h_1,h_2,\ldots,h_n)$, and $(p)(h)=(h_p)$, where $(p)\in S_n$, then from the definition of a symmetric function we have

$$(p)f(h)=f(h_p)=f(h). \qquad (2.1)$$

It has been shown by Perelomov and Popov[34] that the eigenvalues of the invariant operators of U(n), $\langle I^n_k\rangle$, for $k=1,2,\ldots,n$, are symmetric kth degree polynomials of the partial hooks,

$$h_i=u_i+n-i \qquad (2.2)$$

for $i=1,2,\ldots,n$, where $[u]=[u_1u_2\cdots u_n]$ is any IR of U(n). Furthermore, the only kth degree terms of the (h) in $\langle I^n_k\rangle$ are

$$S_k=\sum_{i=1}^{n} h_i^k. \qquad (2.3)$$

The S_k for $k=1,2,\ldots,n$ are the Pythagorian symmetric functions. From the fundamental theorem of symmetric functions[35] any symmetric kth degree polynomial function of (h) for $k\le n$ is expressible as a polynomial function of the $S_1, S_2, \ldots, S_k$ of kth degree in (h). A symmetric kth degree polynomial function of (h) for $k>n$ is expressible as a polynomial function of only the $S_1, S_2, \ldots, S_n$ of nth degree in (h). Clearly, from the above we have

$$\langle I^n_k\rangle = S_k + F^n_{k-1}(S_1, S_2, \ldots, S_{k-1}), \qquad (2.4)$$

for $k=1,2,\ldots,n$, where F^n_{k-1} is a polynomial of the $S_1, S_2, \ldots, S_{k-1}$ of degree $k-1$ in (h) and F_0 is a constant.

We can now prove that the invariant operators of U(n) are independent and complete, and therefore uniquely specify the IR $[u]$ of U(n). To prove the independence of the I^n_k for $k=1,2,\ldots,n$ we show that the Jacobian,

$$J(\langle I^n_k\rangle)=\frac{\partial(\langle I^n_1\rangle\langle I^n_2\rangle\cdots\langle I^n_n\rangle)}{\partial(u_1u_2\cdots u_n)}, \qquad (2.5)$$

is nonvanishing.[36] From (2.4) we have

$J(\langle I^n_k\rangle)=J(S^n_k)$,

$$=n!\begin{vmatrix} h_1^{n-1} & h_1^{n-2} \cdots 1 \\ h_2^{n-1} & h_2^{n-2} \cdots 1 \\ \vdots & \vdots \\ h_n^{n-1} & h_n^{n-2} \cdots 1 \end{vmatrix}$$

$$=n!\,D(h_1,h_2,\ldots,h_n), \tag{2.6}$$

where

$$D(h_1,h_2,\ldots,h_n)=\prod_{i<j}^{n}(h_i-h_j)$$

is the Vandermonde determinant. Now $h_i \neq h_j$ for any $i \neq j$ since $u_i - i \neq u_j - j$ for any partition $[u]$. So $J(\langle I_k^n\rangle)$ is nonvanishing and the operators I_k^n for $k=1,2,\ldots,n$ are independent.

To prove completeness,[37] we must show that

$$h_i = h_i(\langle I_1^n\rangle, \langle I_2^n\rangle, \ldots, \langle I_n^n\rangle) \tag{2.7}$$

for $i=1,2,\ldots,n$. Then any invariant operator I^n where $\langle I^n\rangle = I^n(h_1,h_2,\ldots,h_n)$ may be expressed in terms of $I_1^n, I_2^n, \ldots, I_n^n$ using (2.7). This is equivalent to showing that the n independent equations,

$$\langle I_k^n\rangle = I_k^n(h_1,h_2,\ldots,h_n) \tag{2.8}$$

for $k=1,2,\ldots,n$ have only one solution (h) which corresponds to a partition.

Let (h) be such a solution of (2.8) corresponding to partition $[u]$. Then (h_p) is also a solution of (2.8) since the $I_k^n(h_1,h_2,\ldots,h_n)$ are symmetric functions of (h). Also, since $h_i \neq h_j$ for $i\neq j$ we have $n!$ distinct solutions (h_p) for all $p \in S_n$. But this is the maximum number of distinct solutions allowed from the n equations of (2.8) since $I_k^n(h_1,h_2,\ldots,h_n)$ is of kth degree in (h) and

$$\prod_{k=1}^{n}(k)=n!.$$

So the (h_p) give us all possible solutions of (2.8).

Now define $[u'] < [u]$ if the first nonzero difference $u_i - u_i'$ for $i=1,2,\ldots,n$ is positive. If $[u]$ is a partition then $(p)[u] \le [u]$. Let $R_i = n-i$ and $(R)=(R_1,R_2,\ldots,R_n)$ so that $(h)=[u]+(R)$. We note that $(p)(R) < (R)$ and $-(R) < -(p^{-1})(R)$ for $(p)\neq(1)$. Now let $[u_p]=(h_p)-(R)$ be the partition corresponding to the solution (h_p) of (2.8). Then for $(p)\neq(1)$

$$\begin{aligned}[u_p] &= (p)[u]+(p)(R)-(R) \\ &< [u]+(R)-(p^{-1})(R)=(p^{-1})[u_p]. \end{aligned} \tag{2.9}$$

So $[u_p]$ cannot be a partition unless $(p)=(1)$. We have only one solution (h) which corresponds to a partition.

B. Permutation class operators

For a number of applications in theoretical spectroscopy[38] it is convenient to recast some of the unitary operator formalism in terms of r-cycle permutation operators. In particular, we will show that the invariant operator eigenvalues of U(n) and SU(n) can be expressed in terms of the r-cycle class operator eigenvalues given in (1.17). This is convenient because the eigenvalues of the K_r are easily evaluated in terms of a hook-length formula given in Appendix A.

Let us consider the eigenvalues of the K_r of S_p in more detail. If we restrict our attention to partitions $[u]=[u_1u_2\cdots u_n]$ with n rows, it has been shown by Yamanouchi that[39]

$$\chi_r^{[u]}=(p-r)!\sum_{i=1}^{n}\frac{D(h_1,h_2,\ldots,h_i-r,\ldots,h_n)}{h_1!h_2!\cdots(h_i-r)!\cdots h_n!}, \tag{2.10}$$

for $r=0,2,3,\ldots,n$, where the sum is over all indices such that there are no negative factorials. From (2.10) it follows that for $r=0$,

$$\chi_{1^p}^{[u]}=l^{[u]}=\frac{p!}{h_1!h_2!\cdots h_n!}\,D(h_1,h_2,\ldots,h_n), \tag{2.11}$$

and from (1.3) and (1.17) that

$$\langle K_r\rangle=\frac{1}{r}\sum_{i=1}^{n}\frac{h_i!\,D(h_1,h_2,\ldots,h_i-r,\ldots,h_n)}{(h_i-r)!\,D(h_1,h_2,\ldots,h_n)}, \tag{2.12}$$

for $r=2,3,\ldots,n$.

It can now be seen that the eigenvalues of the operators K_r are symmetric rth degree polynomials of the partial hooks (h). Thus $\langle K_r\rangle$ for $r=2,3,\ldots,n$ is a polynomial function of the Pythagorian symmetric functions $S_1,S_2,\ldots,S_r$ when K_r acts on an irreducible bases of $[u]$ with n rows.

We would like to find the class operators which are complete and independent when acting on such a basis. For this purpose we adjoin to the r-cycle class operators the invariant operator I_1^n of U(n), where

$$\langle I_1^n\rangle=\sum_{i=1}^{n}u_i=p_n=p$$

and adopt the notation:

$$\begin{aligned} K_r^{p_n} &= K_r \quad \text{for } r=2,3,\ldots,n, \\ K_1^{p_n} &= I_1^n. \end{aligned} \tag{2.13}$$

This notation makes explicit the fact that the eigenvalue p_n of I_1^n determines the group S_{p_n} to which the r-cycle class operators belong.

We may expand the $\langle K_r^{p_n}\rangle$ in terms of polynomials of S_r of rth degree in (h):

$$\langle K_r^{p_n}\rangle=\frac{1}{r}\sum_{k=1}^{r}\sum_{(v)}a_k^{(v)}(S_1^{v_1}S_2^{v_2}\cdots S_k^{v_k}), \tag{2.14}$$

where we sum over those $(v)=(v_1,v_2,\ldots,v_k)$ such that $\sum_{i=1}^{k} i(v_i)=k$. We know that

$$\langle K_1^{p_n}\rangle=\sum_{i=1}^{n}u_i=S_1+n(n-1)/2. \tag{2.15}$$

We wish to prove that

$$\langle K_r^{p_n}\rangle=(1/r)S_r+f_{r-1}^n(S_1,S_2,\ldots,S_{r-1}), \tag{2.16}$$

for $r=2,3,\ldots,n$, where f_{r-1}^n is a polynomial of the $S_1,S_2,\ldots,S_{r-1}$ of degree $r-1$ in (h). This is equivalent to proving that the only rth degree terms of the (h) in $\langle K_r^{p_n}\rangle$ are $(1/r)S_r$. We follow closely the procedure of Hammermesh.[40]

From (2.12) and (2.14) we have

$$\begin{aligned}&\sum_{i=1}^{n}\frac{h_i!}{(h_i-r)!}\,D(h_1,h_2,\ldots,h_i-r,\ldots,h_n) \\ &\quad = D(h_1,h_2,\ldots,h_n)\sum_{k,(v)}a_k^{(v)}(S_1^{v_1}S_2^{v_2}\cdots S_k^{v_k}). \end{aligned} \tag{2.17}$$

To find the $a_k^{(v)}$, we equate coefficients of like monomials of (h) on both sides of Eq. (2.17). Consider only terms in which the power of h_1 is $\geq n$ and in which the power of all other h_i is $\geq n-r$. Then

$$\frac{h_1!}{(h_1-r)!}\begin{vmatrix} (h_1-r)^{n-1} & (h_1-r)^{n-2} & \cdots & (h_1-r)^{n-r} \\ h_2^{n-1} & h_2^{n-2} & \cdots & h_2^{n-r} \\ \vdots & & & \vdots \\ h_r^{n-1} & h_r^{n-2} & \cdots & h_r^{n-r} \end{vmatrix} |h_{r+1}^{n-r-1}\cdots h_{n-1}|$$

$$= \begin{vmatrix} h_1^{n-1} & h_1^{n-2} & \cdots & h_1^{n-r} \\ h_2^{n-1} & h_2^{n-2} & \cdots & h_2^{n-r} \\ \vdots & & & \vdots \\ h_r^{n-1} & h_r^{n-2} & \cdots & h_r^{n-r} \end{vmatrix} |h_{r+1}^{n-r-1}\cdots h_{n-1}|$$

$$\times \sum_{k,(v)} a_k^{(v)} S_1^{v_1} S_2^{v_2}\cdots S_k^{v_k}),$$

where $|h_{r+1}^{n-r-1}\cdots h_{n-1}|$ is the determinant left after removing the first r rows and columns from $D(h_1,h_2,\ldots,h_n)$. Equating the coefficients of the monomials of highest degree on both sides, we find

$$a_r^{(v)} = \begin{cases} 1 & \text{when } (v)=(00\cdots01), \\ 0 & \text{when } (v)\neq(00\cdots01). \end{cases} \tag{2.18}$$

The remaining terms of $\langle K_r^{p_n}\rangle$ must be symmetric polynomials of (h) of degree less than r, thus proving (2.16).

We may prove, as we did for the invariant operators I_r^n of $U(n)$, that the r-cycle class operators are complete and independent when operating on irreducible bases of $[u]$ with n rows. The operator $K_1^{p_n}$ uniquely specifies the permutation group S_{p_n} to which the r-cycle class operators belong. The remaining $K_r^{p_n}$ for $r=2,3,\ldots,n$ then uniquely specify the partition $[u]$ of S_{p_n} with n rows.

As a corollary to the above, let $[u]$ be any partition with p boxes so that it labels an IR of S_p. Since such a partition can not have over p rows it must be specified by the operators K_r^p for $r=1,2,\ldots,p$, where $K_1^p=p$. This proves the results of Kramer given in Sec. I.

We have assumed that the IR $[u]$ of $U(n)$ has n rows. Now let the IR $[u]$ of $U(n)$ have only k rows, that is, we assume $u_{k+1}=u_{k+2}=\cdots=u_n=0$. Then $\langle I_r^n\rangle$ and $\langle K_r^{p_n}\rangle$ may be expanded as a polynomial function of the $S_1,S_2,\ldots,S_k$ of kth degree in $(h)=(h_1,h_2,\ldots,h_k)$ when $k\leq r$. Obviously, in this case not all of the operators I_r^n for $r=1,2,\ldots,n$ or $K_r^{p_n}$ for $r=1,2,\ldots,n$ are independent, but we may easily choose an independent set by letting $r=1,2,\ldots,k$. However, we must use an over complete set of operators to specify the IR $[u]$ since we have no way of determining beforehand how many rows are contained in $[u]$. Unfortunately, for $r>k$ it is possible that $p_n<r$, in which case $K_r^{p_n}$ is not defined. We may remedy this situation and use the set of operators $K_r^{p_n}$ for $r=1,2,\ldots,n$ to specify any $[u]$ if we stipuate that

$$K_r^{p_n}=0, \quad \text{when } p_n<r. \tag{2.19}$$

We have shown that the IR's of $U(l)$ may be specified by either the operators I_r^l or $K_r^{p_l}$ for $r=1,2,\ldots,l$. Since both sets of operators are complete and independent, we may expand one set in terms of the other.[41,42] It follows that since the Gel'fand operators I_r^l are mutually commuting invariant Hermitian operators of $U(n)$ for $r=1,2,\ldots,l$, and $l=1,2,\ldots,n$, so are the operators:

$$\begin{matrix} K_n^{p_n}\,K_{n-1}^{p_n}\cdots K_1^{p_n} \\ K_{n-1}^{p_{n-1}}\cdots K_1^{p_{n-1}} \\ \ddots \\ K_1^{p_1} \end{matrix}$$

where $K_r^{p_l}=0$ for $p_l<r$.

Thus the set of eigenvalues of the r-cycle class operators above completely specify a canonical or Gel'fand basis of $U(n)$. Also, because of the relation (1.30), such a set of eigenvalues will also specify the canonical basis of $SU(n)$.

III. CANONICAL WEYL BASES

Let $|\phi_i^l\rangle$ be the ith state of the lth particle which is taken as the ith fundamental basis of $U(n)$ or $SU(n)$. These bases are assumed to be orthogonal,

$$\langle\phi_i^l|\phi_{i'}^{l'}\rangle=\delta_{ii'}\delta_{ll'}. \tag{3.1}$$

The single particle generators e_{ij}^l of the fundamental representation of $U(n)$ operate on the mth particle state such that

$$e_{ij}^l|\phi_k^m\rangle=\delta_{lm}\delta_{jk}|\phi_i^l\rangle. \tag{3.2}$$

Hence, the e_{ij}^l obey the commutation relations

$$[e_{ij}^m,e_{kl}^n]=\delta_{mn}(e_{il}^m\delta_{jk}-e_{kj}^m\delta_{il}), \tag{3.3}$$

as required.

The pth direct product of the fundamental bases,

$$|\phi_{(i)}\rangle\equiv|\phi_{i_1}^1\rangle|\phi_{i_2}^2\rangle\cdots|\phi_{i_p}^p\rangle=|\phi_{i_1}^1\phi_{i_2}^2\cdots\phi_{i_p}^p\rangle,$$

for $i_1,i_2,\ldots,i_p=1,2,\ldots,n$ is an n-dimensional, p-rank tensor space and forms a reducible basis of $U(n)$ with generators

$$E_{ij}=\sum_{l=1}^{p} e_{ij}^l. \tag{3.4}$$

These generators obey the required commutation relations (1.24). The tensor space also forms an orthonormal basis,

$$\langle\phi_{(i)}|\phi_{(j)}\rangle=\delta_{i_1j_1}\delta_{i_2j_2}\cdots\delta_{i_pj_p}=\delta_{(i)(j)}. \tag{3.5}$$

We may define two commuting groups of permutation operators $[q]$ and (q) which act on our tensor space. A permutation $[q]$ acts on the state labels associated with each particle and transfers them to other particles. Thus the *particle* permutation [123] means transfer the state label of particle 1 to particle 2, transfer the state label of particle 2 to particle 3, and transfer the state label of particle 3 to particle 1. For example,

$$[123]|\phi_{i_1}^1\phi_{i_3}^2\phi_{i_2}^3\rangle=|\phi_{i_2}^1\phi_{i_1}^2\phi_{i_3}^3\rangle.$$

A permutation (q) acts on the state labels themselves and transfers them to other state labels. Thus the *state* permutation (123) means transfer state label i_1 to i_2, transfer state label i_2 to i_3, and transfer state label i_3 to i_1. For example,

$$(123)\,|\phi^1_{i_1}\phi^2_{i_3}\phi^3_{i_2}\rangle = |\phi^1_{i_3}\phi^2_{i_2}\phi^3_{i_1}\rangle.$$

Because of the different multiplicative properties of $[q]$ and (q) their irreducible matrix representations are different and obey the relation

$$D^{[u]}_{ij}(q) = D^{[u]}_{ji}[q]. \tag{3.6}$$

We also note that

$$(q)\,|\phi_{(i)}\rangle = [q]\,|\phi_{(i)}\rangle. \tag{3.7}$$

When operating on the tensor space, the generators E_{ij} do not affect the ordering of the subscripts so that the E_{ij} commute with all $[q]$ and (q), i.e.,

$$E_{ij}[q] = [q]\,E_{ij}, \tag{3.8a}$$

$$E_{ij}(q) = (q)\,E_{ij}. \tag{3.8b}$$

We now reduce the tensor space under permutations of S_p. Let $[P^{[u]}_{rs}]$ operate with particle permutations and $(P^{[u]}_{rs})$ operate with state permutations. From (1.10) we have that the projected basis $[P^{[u]}_{rs}]|\phi_{(i)}\rangle$ transforms like an irreducible basis $|{}^{[u]}_{r}\rangle$ of S_p under particle permutations $[q]$,

$$[q][P^{[u]}_{rs}]\,|\phi_{(i)}\rangle = \sum_{r'} [P^{[u]}_{r's}]\,|\phi_{(i)}\rangle D^{[u]}_{r'r}[q]. \tag{3.9}$$

Also from (3.6), (1.11), and (3.7) we have

$$\begin{aligned}(q)[P^{[u]}_{rs}]\,|\phi_{(i)}\rangle &= [P^{[u]}_{rs}](q)\,|\phi_{(i)}\rangle \\ &= [P^{[u]}_{rs}][q]\,|\phi_{(i)}\rangle \\ &= \sum_{s'} [P^{[u]}_{rs'}]\,|\phi_{(i)}\rangle D^{[u]}_{ss'}[q] \\ &= \sum_{s'} [P^{[u]}_{rs'}]\,|\phi_{(i)}\rangle D^{[u]}_{s's}(q).\end{aligned} \tag{3.10}$$

Thus $[P^{[u]}_{rs}]|\phi_{(i)}\rangle$ transforms like an irreducible basis $|{}^{[u]}_{s}\rangle$ of S_p under state permutations (q). We see that the particle projection operator reduces the tensor space under both particle and state permutations. Also from (3.6) and (3.7) we have

$$[P^{[u]}_{rs}]\,|\phi_{(i)}\rangle = (P^{[u]}_{sr})\,|\phi_{(i)}\rangle.$$

One may use either particle or state projection operators to reduce the space. We choose to use particle projection operators and from now on we assume the $P^{[u]}_{rs}$ are *particle* projection operators unless otherwise denoted.

Let

$$|{}^{[u]}_{s\,r}\rangle = P^{[u]}_{rs}\,|\phi_{(i)}\rangle. \tag{3.11}$$

The different irreducible bases $|{}^{[u]}_{s\,r}\rangle$ of S_p for a fixed s and (i) are orthogonal. To show this, we first note that the particle permutations are unitary operators on the tensor space:

$$([q]\,|\phi_{(i)}\rangle)^\dagger\,|\phi_{(j)}\rangle = \langle\phi_{(i)}\,|[q^{-1}]|\,\phi_{(j)}\rangle,$$

or

$$[q]^\dagger = [q^{-1}]. \tag{3.12}$$

Using (3.12), (1.8), and (1.9) we have

$$[P^{[u]}_{rs}]^\dagger = [P^{[u]}_{sr}]. \tag{3.13}$$

From the above equation it follows that

$$\begin{aligned}\langle{}^{[u]}_{s\,r}|{}^{[u']}_{s'r'}\rangle &= \langle\phi_{(i)}\,|P^{[u]\dagger}_{rs}P^{[u']}_{r's'}|\,\phi_{(i')}\rangle \\ &= \langle\phi_{(i)}\,|P^{[u]}_{sr}P^{[u']}_{r's'}|\,\phi_{(i')}\rangle \\ &= \delta_{[u][u']}\,\delta_{rr'}\langle\phi_{(i)}\,|P^{[u]}_{ss'}|\,\phi_{(i')}\rangle.\end{aligned} \tag{3.14}$$

The number of times the IR $[u]$ of S_p is contained in the tensor space is the number of independent bases $|{}^{[u]}_{s\,r}\rangle$ for a given $[u]$ and r. This will be shown to be equal to $f^{[u]}$.

The particle projection operator also reduces the tensor space under transformations of the generators E_{ij} of U(n). We have

$$\begin{aligned}[q]E_{ij}\,|{}^{[u]}_{s\,r}\rangle &= E_{ij}[q]\,|{}^{[u]}_{s\,r}\rangle \\ &= \sum_{r'} E_{ij}\,|{}^{[u]}_{s\,r'}\rangle D^{[u]}_{r'r}[q],\end{aligned}$$

so that bases with fixed r transform among themselves under all E_{ij} of U(n) and form a representation of U(n). The dimension of this representation of U(n) for a given r and $[u]$ is the number of independent bases, $f^{[u]}$. The number of such representations contained in the tensor space is the number of independent bases $|{}^{[u]}_{s\,r}\rangle$ for a given $[u]$, s, and (i). But this is just the dimension of the IR $[u]$ of S_p, or $l^{[u]}$.

To complete the reduction of S_p and U(n) on the tensor space using particle projection operators one can determine the $f^{[u]}$ independent bases for fixed $[u]$ and r and then orthonormalize these using the standard Graham–Schmidt procedures. An alternate method, which we shall adopt, is to simultaneously diagonalize the mutually commuting Hermitian invariant operators of U(n),

$$\begin{array}{l} K^{p_n}_n\,K^{p_n}_{n-1}\cdots K^{p_n}_1 \\ K^{p_{n-1}}_{n-1}\cdots K^{p_{n-1}}_1 \\ \quad\ddots \qquad \vdots \\ \qquad\quad K^{p_1}_1 \end{array},$$

where $K^{p_l}_r = 0$ when $p_l < r$, in our bases $|{}^{[u]}_{s\,r}\rangle$. The eigenvalues of the operators must uniquely specify the canonical irreducible basis of U(n) to which the eigenvectors correspond. Hence, bases with the same set of eigenvalues must be equal within a normalization factor. Because of the Hermitian properties of the invariant operators, bases with different sets of eigenvalues must be orthogonal and correspond to different canonical irreducible bases of U(n).

An irreducible basis constructed in this manner is a Weyl basis of U(n). Because of our choice of operators the Weyl basis will also be a canonical or Gel'fand basis of U(n).

We can reduce our work by eliminating the bases which are obviously not independent. Denote by $|\phi_\lambda\rangle$ any tensor with state labels consisting of λ_1 ones, λ_2 twos, . . . , and λ_n n's in the subscripts. Using (3.10) in the following form:

$$P^{[u]}_{rs}[q]\,|\phi_\lambda\rangle = \sum_{s'} P^{[u]}_{rs'}\,|\phi_\lambda\rangle D^{[u]}_{ss'}[q],$$

we may let $|\phi_\lambda\rangle$ have any order in the individual particle states and still obtain the same independent bases. For this reason we may choose the ordering below,

$$|\phi_\lambda\rangle = |\phi^1_1\phi^2_1\cdots\phi^{p_1}_1\phi^{p_1+1}_2\cdots\phi^{p_2}_2\quad\cdots\phi^{p_{n-1}+1}_n\cdots\phi^{p_n}_n\rangle, \tag{3.15}$$

where $\sum_{i=1}^{r}\lambda_i = p_r$, and $p_n = p$, without losing generality. We now define

$$\left|{}^{[u]}_{s\,r}\right\rangle = P^{[u]}_{rs}\,|\phi_\lambda\rangle, \tag{3.16}$$

and note that

$$E_{ii}\left|{}^{[u]}_{s\,r}\right\rangle = \lambda_i\left|{}^{[u]}_{s\,r}\right\rangle. \tag{3.17}$$

The weight of the basis $|{}^{[u]}_{s\,r}\rangle$ is $(\lambda)=(\lambda_1\lambda_2\cdots\lambda_n)$. The bases $|{}^{[u]}_{s\,r}\rangle$ are already eigenvectors of the invariant operators $K_1^{p_r}$ for $r=1,2,\ldots,n$, where

$$K_1^{p_r}\left|{}^{[u]}_{s\,r}\right\rangle = p_r\left|{}^{[u]}_{s\,r}\right\rangle, \tag{3.18}$$

so bases with different weights must be orthogonal. This is also obvious from (3.5).

The question now arises as to which permutation convention to use for the class operators $K_r^{p_l}$ when operating on the tensor space bases. Because the state operators (q) and the generators E_{ij} obey the same commutation relations with respect to the particle projection operators, we use the state classes $(K_r^{p_l})$. Hence, the invariant operators I_k^l are expanded in terms of state class operators $(K_r^{p_l})$ for $r=1,2,\ldots,k$ when acting on the tensor space. From now on we assume the $K_r^{p_l}$ are *state* class operators unless otherwise denoted. Since the I_k^l act only on states with state numbers $1,2,\ldots,l$, the $K_r^{p_l}$ must act only on state labels with state numbers $1,2,\ldots,l$. We define $S_{(p_l)}$ to be the subgroup of S_p corresponding to permutations of the state labels of our tensor bases with state numbers $1,2,\ldots,l$ so that $K_r^{p_l}\subset S_{(p_l)}$. For example, if $i_1i_2i_3i_4i_5=13243$, then $S_{(p_3)}=S_{(4)}$ is the group of permutations of i_1,i_2,i_3, and i_5. Note that $S_{(4)}$ differs from S_4 where the latter is the group of permutations of the *first* four state labels i_1,i_2,i_3, and i_4.

We may now show directly that the class operators $K_r^{p_l}$ for $l=1,2,\ldots,n$ and $r=1,2,\ldots,l$ are mutually commuting invariant Hermitian operators of U(n). We have

$$[K_k^{p_m},K_r^{p_l}]=0, \tag{3.19}$$

since a class of a group commutes with all elements of that group and one of the groups $S_{(p_m)}$ or $S_{(p_l)}$ is a subgroup of the other. From (3.8b) it follows that the r-cycle class operators are invariant operators of U(n):

$$[K_r^{p_l},E_{ij}]=0. \tag{3.20}$$

If a class contains the element (q) it also contains $(q)^{-1}$. If $(K_r^{p_l})^{-1}$ is the inverse of all the permutation terms in $(K_r^{p_l})$, then

$$(K_r^{p_l})^{\dagger}=(K_r^{p_l})^{-1}=(K_r^{p_l}), \tag{3.21}$$

i.e., the r-cycle class operators are Hermitian.

The bases $|{}^{[u]}_{s\,r}\rangle$ are already eigenvectors of the operators $K_1^{p_l}$ for $l=1,2,\ldots,n$ with eigenvalues p_l. As stated before, it is just these eigenvalues that determine which canonical scheme of r-cycle class operators to use, i.e., which subgroup $S_{(p_l)}$ of S_p the r-cycle class operators act on.

We now prove that the projected tensor bases $|{}^{[u]}_{s\,r}\rangle$ are eigenvectors of $K_k^{p_n}$ for $k=1,2,\ldots,n$. From (3.10) we have that $|{}^{[u]}_{s\,r}\rangle$ transforms like an irreducible basis $|{}^{[u]}_{s}\rangle$ under permutations (q) of S_p. Then from (1.17) we have

$$K_k^{p_n}\left|{}^{[u]}_{s\,r}\right\rangle = \frac{N_k^{p_n}\chi_k^{[u]}}{l^{[u]}}\left|{}^{[u]}_{s\,r}\right\rangle \tag{3.22}$$

for $k=2,3,\ldots,n$ where $N_k^{p_n}$ is the order of the class $K_k^{p_n}$. This proves that the independent bases $|{}^{[u]}_{s\,r}\rangle$ for fixed r form an irreducible basis of U(n) corresponding to partition $[u]$ if complete.

We now let the projection operators $P^{[u]}_{rs}$ be expanded in terms of the *canonical* IR of S_p. These $P^{[u]}_{rs}$ will then be proportional to the semi-normal projection operators $O^{[u]}_{rs}$. In this case we have from (3.10) that $|{}^{[u]}_{s\,r}\rangle$ transforms like the *canonical* irreducible basis $|{}^{[u]}_{(s)}\rangle$ associated with the standard tableau $T^{[u]}_{(s)}$ under permutations (q) of S_p.

The basis $|{}^{[u]}_{s\,r}\rangle$ therefore transforms like a canonical irreducible basis $|{}^{[u]^{p_l}}_{(s)}\rangle$ of $[u]_s^{p_l}$ under permutations (q) of S_{p_l}. For $K_r^{p_l}$ to be a class operator of the subgroup S_{p_l} in the canonical reduction of the basis $|{}^{[u]}_{(s)}\rangle$, it is necessary that the subgroup $S_{(p_l)}$ of $K_r^{p_l}$ correspond to permutations of the *first* p_l state labels. Since $K_r^{p_l}$ permutes the state numbers $1,2,\ldots,l$, for $K_r^{p_l}$ to be a class of S_{p_l} in this canonical reduction these state numbers must be in the first state labels $i_1,i_2,\ldots,i_{p_l}$. This is the reason for our particular choice of $|\phi_\lambda\rangle$ with definite order such that $i_1\leqslant i_2\leqslant\cdots\leqslant i_p$. From the above considerations, we have

$$K_k^{p_l}\left|{}^{[u]}_{s\,r}\right\rangle = \frac{N_k^{p_l}\chi_k^{[u]_s^{p_l}}}{l^{[u]_s^{p_l}}}\left|{}^{[u]}_{s\,r}\right\rangle \tag{3.23}$$

for $l=1,2,\ldots,n$ and $k=1,2,\ldots,l$. Thus the canonically projected Weyl basis $|{}^{[u]}_{s\,r}\rangle$ transforms like an irreducible basis of $[u]_s^{p_l}$ under U(l). From (1.32) we see that if $|{}^{[M]}_{\langle m\rangle}\rangle$ is a canonical irreducible basis of U(n), then $|{}^{[u]}_{s\,r}\rangle = |{}^{[M]}_{\langle m\rangle}\rangle$ when

$$[u]_s^{p_l}=[m_{1l}\,m_{2l}\cdots m_{ll}] \quad \text{for } l=1,2,\ldots,n-1,$$

$$[u]_s^{p_n}=[u]=[M]. \tag{3.24}$$

Both (s) and weight $(\lambda)=(\lambda_1\lambda_2\cdots\lambda_n)$ where $\sum_{i=1}^{l}\lambda_i=p_l$ uniquely specify $\langle m\rangle$ for a given $[u]=[M]$.

We define a tableau of U(n), $T^{[u]}_{(s)}\phi_\lambda$, to be a partition $[u]$ with state label i_r in the box containing r of standard tableau $T^{[u]}_{(s)}$. For example,

$${}^{124}_{35}\,\phi_1^1\phi_2^2\phi_2^3\phi_3^4\phi_3^5={}^{123}_{23}.$$

Then $[u]_s^{p_l}$ is simply the partition remaining after removing state numbers $l+1,l+2,\ldots,n$ from $T^{[u]}_{(s)}\phi_\lambda$. We have a one to one correspondence between the tableaus $T^{[u]}_{(s)}\phi_\lambda$ and states $|{}^{[u]}_{s\,r}\rangle$.

Not every tableau $T^{[u]}_{(s)}\phi_\lambda$ of U(n) corresponds to a standard tableau of U(n), since the "betweeness conditions" (1.33),

$$m_{i\,l+1}\geqslant m_{il}\geqslant m_{i+1\,l+1},$$

where $[u]_s^{p_l}=[m_{1l}\,m_{2l}\cdots m_{ll}]$ are not necessarily satisfied for all l. However, if a tableau $T^{[u]}_{(s)}\phi_\lambda$ of U(n) contains no identical state numbers in a column, the "betweeness conditions" will always be satisfied and the resulting tableau of U(n) will correspond to a standard tableau $T^{[u]}_{(s)}$.

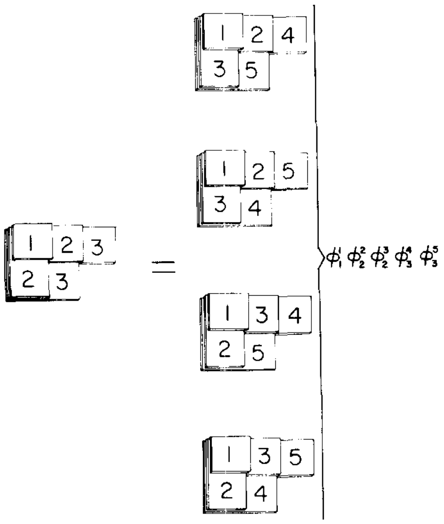


FIG. 6. Correspondence between standard tableaus of U(n) and S_p. The standard tableau of U(3) on the left corresponds to the standard tableaus of S_5 on the right when the numbers 1, 2, 3, 4, 5 are replaced by the state numbers 1, 2, 2, 3, 3 respectively. The first standard tableau of S_5 is derived from the standard tableau of U(3) by replacing the state numbers (1), (2), and (3) by the numbers (1), (2,3), and (4,5) respectively in "book order."

We have not yet shown that all standard tableaus $T^{[u]}_{\langle s\rangle}$ of U(n) for a given weight (λ) correspond to some tableau $T^{[u]}_{(s)}\phi_\lambda$, i.e., we have not yet shown that the projected bases $|^{[u]}_{s\,r}\rangle$ are complete. We need a way of replacing the numbers $1,2,\ldots,n$ in the standard tableau $T^{[u]}_{\langle s\rangle}$ of U(n) with the nonrepeated numbers $1,2,\ldots,p$ to produce a standard tableau $T^{[u]}_{(s)}$ of S_p such that $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}\phi_\lambda$. One such way is to simply replace the numbers i in $T^{[u]}_{\langle s\rangle}$ by the numbers $p_{i-1}+1, p_{i-1}+2,\ldots,p_i$ in "book order" such that they increase to the right in the rows and down in the columns. Usually there are *several* different standard tableaus $T^{[u]}_{(s)}$ of S_p which yield a given $T^{[u]}_{\langle s\rangle}$ of U(n) as shown in Fig. 6. It follows that the projected states $|^{[u]}_{s\,r}\rangle$ corresponding to *different* standard tableaus $T^{[u]}_{\langle s\rangle}$ form a complete and independent set of canonical bases of U(n). The number of such independent bases will be $f^{[u]}$ as indicated previously.

Since the invariant operators are Hermitian, eigenvectors belonging to different sets of eigenvalues are orthogonal. The eigenvalues of $K_1^{p_l}$ determine p_l and the eigenvalues of $K_r^{p_l}$ for $r=2,3,\ldots,l$ uniquely determine the partition $[u]_s^{p_l}$. So, if $[u]_s^{p_l}\neq[u]_{s'}^{p_l}$ for some $l=1,2,\ldots,n-1$, or correspondingly, if $T^{[u]}_{(s)}\phi_\lambda \neq T^{[u]}_{(s')}\phi_{\lambda'}$, then

$$\langle^{[u]}_{s\,r}|^{[u]}_{s'\,r}\rangle = 0.$$

Similarly, eigenvectors belonging to the same set of eigenvalues must be equal within a normalization factor. So, if $T^{[u]}_{(s)}\phi_\lambda = T^{[u]}_{(s')}\phi_{\lambda'}$, then

$$|^{[u]}_{s\,r}\rangle = C\,|^{[u]}_{s'\,r}\rangle,$$

and $\lambda=\lambda'$, where C is a constant. This result has been proven by Goddard using only the properties of the canonical IR's of S_p.[43]

Projected states with tableaus $T^{[u]}_{(s)}\phi_\lambda$ having identical state numbers in a column are orthogonal to projected states with standard tableaus and must therefore be null states. Again, this has been proven using only the properties of the canonical IR's of S_p.[44] As an example, we have

$$\left|\begin{smallmatrix}1&2&4\\3&5\end{smallmatrix}\ \phi_1^1\phi_1^2\phi_1^3\phi_3^4\phi_3^5\right\rangle = \left|\begin{smallmatrix}1&1&3\\1&3\end{smallmatrix}\right\rangle = 0.$$

Now if the IR $[u]$ of U(n) had more than n rows, $T^{[u]}_{(s)}\phi_\lambda$ would have at least two state numbers in a column for any λ and s. So $|^{[u]}_{s\,r}\rangle=0$ for any Weyl basis of U(n) when $[u]$ has more than n rows.

Since the canonically projected nonnull states $|^{[u]}_{s\,r}\rangle$ form a basis $|^{[u]}_{(r)}\rangle$ under permutations (q) of S_p and a basis $|^{[u]}_{\langle s\rangle}\rangle$ under generators E_{ij} of U(n), we write

$$|^{[u]}_{s\,r}\rangle \equiv |^{[u]}_{\langle s\rangle(r)}\rangle.$$

A canonically projected Weyl basis simultaneously forms a canonical basis of S_p and a canonical basis, or Gel'fand basis, of U(n). We shall show the significance of the subspace of all such states $|^{[u]}_{\langle s\rangle(r)}\rangle$ more clearly in Paper II.

To normalize the canonical Weyl basis, we let $|^{[u]}_{\langle s\rangle}\rangle = N_s^{[u]}|^{[u]}_{\langle s\rangle(r)}\rangle$, where $\langle^{[u]}_{\langle s\rangle}|^{[u]}_{\langle s\rangle}\rangle=1$. Then

$$(N_s^{[u]})^2\langle^{[u]}_{\langle s\rangle(r)}|^{[u]}_{\langle s\rangle(r)}\rangle = (N_s^{[u]})^2\langle\phi_\lambda|P^{[u]}_{ss}|\phi_\lambda\rangle = 1.$$

Finally, from (1.8) we have

$$N_s^{[u]} = (p!/l^{[u]}\sum_{q\in S_\lambda} D^{[u]}_{ss}[q])^{1/2}, \tag{3.25}$$

where $S_\lambda = S_{\lambda_1}\times S_{\lambda_2}\times\cdots\times S_{\lambda_n}$.

The canonical Weyl bases of U(n) are eigenvectors of the invariant operators of SU(n). So the canonical Weyl bases of U(n) are also canonical bases of SU(n). Thus if $|^{[u]}_{\langle s\rangle}\rangle = |^{[u]}_{\langle s\rangle(r)}\rangle$, we have that $|^{[u]}_{\langle s\rangle(r)}\rangle$ must also transform like $|^{[u']}_{\langle s\rangle}\rangle$ under SU(n), where

$$[u'] = [u_1-u_n\ u_2-u_n\cdots u_{n-1}-u_n\ 0]. \tag{3.26}$$

APPENDIX A[45]

We wish to find the eigenvalues $N_r^p\chi_r^{[u]}/l^{[u]}$ where $[u]=[u_1u_2\cdots u_m]$ is a partition labeling the IR's of S_p and satisfying the relation

$$\sum_{i=1}^m u_i = p. \tag{A1}$$

By partition we mean that the elements of $[u]$ are monotonically decreasing integers such that

$$u_1\geq u_2\geq\cdots\geq u_m. \tag{A2}$$

We first find the $[u]_{ir}$ for all i where

$$[u]_{ir} = [u_1u_2\cdots u_i - r\cdots u_m]. \tag{A3}$$

If $[u]_{ir}$ is not a partition, it may be possible to transform it into one using the following procedure. Let $[R]=[m-1\ m-2\cdots 0]$ and find the permutation (q_i) such that $(q_i)([u]_{ir}+[R])$ has monotonically decreasing elements. Then find the $[u]'_{ir}$ for all i where

$$[u]'_{ir} = (q_i)([u]_{ir}+[R]) - [R]. \tag{A4}$$

Note that $[u]'_{ir}$ is not a partition if and only if $([u]_{ir}+[R])$ contains repeated or negative elements. Also note that $[u]'_{ir}=[u]_{ir}$ if $[u]_{ir}$ is a partition. We shall need only the $[u]'_{ir}$ which are partitions.[46]

We may now find the eigenvalues of the r-cycle class operators by using the simple hook-length formula below,

$$N_r^p \chi_r^{[u]}/l^{[u]} = \frac{1}{r} \sum_{i=1}^{m} \epsilon_{(q_i)} \frac{H([u])}{H([u]'_{ir})}. \tag{A5}$$

The sum in this equation is over all i such that $[u]'_{ir}$ is a partition. Also $\epsilon_{(q_i)}$ is 1 or -1 if the permutation (q_i) is an even or odd number of bicycles respectively. $H([u])$ is the product of hook-lengths of partition $[u]$ which has been presented in Fig. 1.

As an example, we find the eigenvalue of K_3^9 for IR $[u]=[432]$ of S_9. The $[u]_{ir}$ and corresponding $[u]'_{ir}$ are:

$$[432]_{13}=[132],$$

$$[432]_{23}=[402],$$

$$[432]_{33}=[43-1],$$

$$[432]'_{13}=(12)[342]-[210]=[222],$$

$$[432]'_{23}=(23)[612]-[210]=[411],$$

$$[432]'_{33}=[64-1]-[210]=[43-1].$$

Using the partitions $[432]'_{13}$ and $[432]'_{23}$ it is now a simple matter to evaluate the eigenvalue of K_3^9 in terms of hook-lengths as shown below,

$$\frac{N_3^9\chi_3^{[432]}}{l^{[432]}} = \frac{1}{3}\left[-\frac{\begin{matrix} 6 & 5 & 3 & 1 \\ 4 & 3 & 1 & \\ 2 & 1 & & \end{matrix}}{\begin{matrix} 4 & 3 \\ 3 & 2 \\ 2 & 1 \end{matrix}} - \frac{\begin{matrix} 6 & 5 & 3 & 1 \\ 4 & 3 & 1 & \\ 2 & 1 & & \end{matrix}}{\begin{matrix} 6 & 3 & 2 & 1 \\ 2 & & & \\ 1 & & & \end{matrix}} \right]$$

$$= -15.$$

*Work supported by Centro Técnico Aerospacial.

[1]H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, New Jersey, 1946).
[2]G. E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[3]G. E. Baird and L.C. Biedenharn, J. Math. Phys. **5**, 1723 (1964).
[4]G. E. Baird and L.C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).
[5]M. Ciftan and L.C. Biedenharn, J. Math. Phys. **10**, 221 (1969).
[6]J.D. Louck, Am. J. Phys. **38**, 3 (1970).
[7]J. Schwinger, in *Quantum Theory of Angular Momentum*, edited by L.C. Biedenharn and C. Van Dam (Academic, New York, 1965), p. 229.
[8]J.D. Louck, Am. J. Phys. **38**, 36 (1970); L.C. Biedenharn, A. Giovannini, and J.D. Louck, J. Math. Phys. **8**, 691 (1967).
[9]W.G. Harter, Phys. Rev. A 8 6, 2819 (1973).
[10]J. Drake, G.W.F. Drake, and M. Schlesinger, J. Phys. B **8**, 7 (1975).
[11]J. Paldus (to appear in Adv. Theor. Chem. Vol. 2).
[12]W. G. Harter and C.W. Patterson, Phys. Rev. A **13**, 1067 (1976).
[13]D. E. Littlewood, *The Theory of Group Characters* (Clarendon, Oxford, England, 1950), p. 178.
[14]D.R. Tompkins, J. Math. Phys. **8**, 1502 (1967); J. Math. Phys. **11**, 105 (1970).
[15]J. L. Sullivan, Phys. Rev. A **5**, 29 (1972); J. Math. Phys. **14**, 387 (1973).
[16]K. J. Lezuo, J. Math. Phys. **13**, 1389 (1972).
[17]M. Hammermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).
[18]G. E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1460 (1963).
[19]See Ref. 17, p. 189.
[20]See Ref. 17, p. 27.
[21]G. de B. Robinson, *Representation Theory of Symmetric Groups* (University of Toronto Press, Toronto, Canada, 1961).
[22]P. Kramer, Z. Naturforsch. **21*a***, 657 (1966).
[23]See Ref. 13, p. 140.
[24]D. E. Rutherford, *Substitutional Analysis* (Edinburgh U.P., London, 1948), p. 24.
[25]See Ref. 24, p. 27.
[26]J.D. Louck, Am. J. Phys. **38**, 3 (1970).
[27]See Ref. 17, p. 388.
[28]I. M. Gel'fand, Mat. Sb. **26**, 103 (1950).
[29]A. M. Perelomov and V.S. Popov, J. Nucl. Phys. (USSR) **3**, 924 (1966) [Sov. J. Nucl. Phys. **3**, 676 (1966)].
[30]See Ref. 17, p. 388.
[31]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[32]H. Weyl, *The Theory of Groups and Quantum Mechanics*, translated by H.P. Robertson (Dover, New York, 1931), p. 391.
[33]G. E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[34]See Ref. 29, p. 678, Eq. (20).
[35]See Ref. 1, p. 30.
[36]This result is shown in Ref. 29.
[37]This proof parallels a proof given for the completeness of invariant operators of SU(n). See B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964). The S Theorem of this reference is equivalent to our $I_k^n(h) = I_k^n(h_p)$.
[38]C. W. Patterson and W.G. Harter, *A Unitary Calculus for Electronic Orbitals* (Lecture Notes in Physics) (Springer-Verlag, Heidelberg, 1976).
[39]T. Yamanouchi, Proc. Phys. Math. Soc. Jpn. **19**, 436 (1937). Also see Ref. 17, p. 213.
[40]See Ref. 17, p. 213.
[41]A. Partensky, J. Math. Phys. **13**, 1503 (1972).
[42]C. W. Patterson, Ph.D. thesis, Univ. of Southern California (1974).
[43]W.A. Goddard III, Phys. Rev. **157**, 73 (1967).
[44]See Ref. 43.
[45]A. J. Coleman, *Induced Representations with Applications to S_n and $GL(n)$* (Queens University, Kingston, Ontario, Canada, 1966). The results given are based on this work.
[46]This treatment is remarkably similar to that used to find the Clebsch—Gordon series for U(n). See J. D. Louck, Am. J. Phys. **38**, 18 (1970).

Canonical symmetrization for the unitary bases. II. Boson and fermion bases*

Chris W. Patterson and William G. Harter

Instituto de Fisica, Departamento de Eletronica Quantica, Universidade Estadual de Campinas, Caixa Postal 1170 Campinas, S.P. Brazil

(Received 5 November 1975)

The canonical Weyl basis described in Paper I is generalized to give a boson and fermion calculus which generates the symmetric and antisymmetric bases of $U(nm)$ respectively contained in the irreducible bases of $U(n)\times U(m)$. The boson calculus may be used to find the multiplicity free Clebsch–Gordan coefficients of $U(n)$.

I. INTRODUCTION

In works by Biedenharn, Baird, Moshinsky, Louck, and Seligman,[1–5] extensive use has been made of boson operators to generate an irreducible basis of $U(n)\times U(m)$ which is simultaneously a symmetric basis of $U(nm)$. Louck has shown that this boson basis is a basis for the n-dimensional, m-particle harmonic oscillator.[6] However, symmetrization of this boson basis using the "boson calculus" has so far failed to generate all the independent bases in this space which we shall denote as $U(n) * U(m)$. The construction of a complete basis is presently a tedious task using lowering operator techniques.[7–10] Furthermore, the boson calculus itself has no justification other than a constructional validity.

In this work, we shall show the simple relationship between the boson basis and the canonical Weyl basis described in an earlier work[11] (denoted hereafter as I). This relationship leads to a new boson calculus capable of generating all the independent bases of $U(n) * U(m)$. Furthermore, we show that the canonical Weyl basis may be considered as the "special" boson basis of subspace $U(n) * S_p$ or $S_p * U(m)$ of $U(n) * U(m)$ first noted by Moshinsky.[12] This clarifies the fact that the generators of the unitary group can be used as elements of the permutation group S_p when acting on this particular basis. We also show that the canonical Weyl bases of the subspace $S_p * S_p$ form a basis for the regular representation of S_p. As a result of this new boson calculus, it becomes a straightforward task to determine the matrix elements of the irreducible representations (IR's) of $U(n)$, and by means of the factorization lemma,[13] to calculate the multiplicity-free Clebsch–Gordan coefficients of $U(n)$.

In a similar manner, we develop a fermion calculus to generate an irreducible basis of $U(n)\times U(m)$, which is simultaneously an antisymmetric basis of $U(nm)$. We shall denote the subspace of all such irreducible bases of $U(n)\times U(m)$ as $U(n) \tilde{*} U(m)$. The fermion calculus enables us to find the antisymmetric bases of $U(nm)$ contained in the irreducible bases of $U(n)\times U(m)$. This is a very important task when dealing with fermions in atomic and nuclear physics. For example, in atomic l-shells one often needs to find the antisymmetric content of $U(4l+2)$ in symmetrized orbit–spin states of $U(2l+1)\times U(2)$. Similarly, in nuclear shells one often needs to find the antisymmetric content of $U(8l+4)$ in states of $U(2l+1)\times U(4)$ since the spin states now include isotopic spin. In the atomic case, closed form expressions already exist for the antisymmetric content in symmetrized orbit–spin states which were derived using a canonical Weyl basis.[14]

II. IRREDUCIBLE BASES FOR U(nm) AND U(n) × U(m)

Let $|\phi_i^l\rangle$ for $i=1,2,\ldots,n$ and $|\psi_l^j\rangle$ for $j=1,2,\ldots,m$ form bases for the lth particle of the fundamental representation of $U(n)$ and $U(m)$ respectively with generator relations:

$$e^l_{im}|\phi_p^k\rangle=\delta_{lk}\delta_{mp}|\phi_i^l\rangle,\qquad e_l^{jn}|\psi_k^q\rangle=\delta_{lk}\delta_{nq}|\psi_l^j\rangle. \tag{1}$$

The pth direct products

$$|\phi_{(i)}\rangle\equiv|\phi^1_{i_1}\phi^2_{i_2}\cdots\phi^p_{i_p}\rangle,\qquad |\psi^{(j)}\rangle\equiv|\psi_1^{j_1}\psi_2^{j_2}\cdots\psi_p^{j_p}\rangle \tag{2}$$

form a reducible bases of $U(n)$ and $U(m)$ respectively with the generators:

$$E_{im}=\sum_{l=1}^{p}e^l_{im},\qquad E^{jn}=\sum_{l=1}^{p}e_l^{jn}. \tag{3}$$

As we have seen in I, we may reduce direct products (2) using the canonical projection operators of S_p as in (4) to form a canonical Weyl basis for $U(n)$ and $U(m)$ respectively:

$$\left|\begin{matrix}[u]\\ \langle s\rangle\quad (m)\end{matrix}\right\rangle=N_s^{[u]}\underline{P}^{[u]}_{ms}|\phi_{(i)}\rangle, \tag{4a}$$

$$\left|\begin{matrix}[v]\\ (n)\quad \langle r\rangle\end{matrix}\right\rangle=N_r^{[v]}\bar{P}^{[v]}_{nr}|\psi^{(j)}\rangle. \tag{4b}$$

The upper bar ($^-$) denotes permutations of the superscripts and the lower bar (_) denotes permutations of the subscripts. The reason for the change in notation for the Weyl basis on the left of (4b) will become evident later.

We define the direct product basis $|\Phi_i^j(l)\rangle$ in Eq. (5):

$$|\phi_i^l\rangle\times|\psi_l^j\rangle=|\Phi_i^j(l)\rangle \tag{5}$$

Then the $|\Phi_i^j(l)\rangle$ for $i=1,2,\ldots,n$ and $j=1,2,\ldots,m$ form bases for the lth particle of the fundamental representation of $U(nm)$ with generator relations:

$$e^{jn}_{im}(l)\,|\,\Phi^{q}_{p}(k)\rangle\!\rangle = \delta_{lk}\delta_{mp}\delta_{nq}\,|\,\Phi^{j}_{i}(l)\rangle\!\rangle, \quad (6)$$

where $e^{jn}_{m}(l) = e^{l}_{im}\times e^{jn}_{l}$.

The direct product basis

$$|\,\phi_{(i)}\rangle \times |\,\psi^{(j)}\rangle = |\,\Phi^{j_1}_{i_1}(1)\Phi^{j_2}_{i_2}(2)\cdots\Phi^{j_p}_{i_p}(p)\rangle\!\rangle \equiv |\,\Phi^{(j)}_{(i)}\rangle. \quad (7)$$

is a reducible basis of $U(n)\times U(m)$ with generators $E_{im}\times E^{jn}$. We may reduce this basis using (4) and find the irreducible bases $|\,{}^{[u]}_{\langle s\rangle}\rangle \times |\,{}^{[v]}_{\langle r\rangle}\rangle$ of $U(n)\times U(m)$ are simply

$$\left|\begin{matrix}[u]\\ \langle s\rangle\quad(m)\end{matrix}\right\rangle \times \left|\begin{matrix}[v]\\ (n)\quad\langle r\rangle\end{matrix}\right\rangle = N^{[v]}_{r}N^{[u]}_{s}\bar{P}^{[v]}_{nr}\underline{P}^{[u]}_{ms}\,|\,\Phi^{(j)}_{(i)}\rangle. \quad (8)$$

The pth direct product basis $|\,\Phi^{(j)}_{(i)}\rangle$ also forms a reducible bases of $U(nm)$ with generators:

$$E^{jn}_{im} = \sum_{l=1}^{p} e^{jn}_{im}(l). \quad (9)$$

Again, we may reduce the bases of $U(nm)$ using the canonical projection operators of S_p to form a canonical irreducible Weyl basis $|\,{}_{\langle t\rangle}{}^{[\lambda]}{}_{(0)}\rangle$ as in (10):

$$\left|\begin{matrix}[\lambda]\\ \langle t\rangle\quad(o)\end{matrix}\right\rangle = N^{[\lambda]}_{t}P^{[\lambda]}_{ot}\,|\,\Phi^{(j)}_{(i)}\rangle. \quad (10)$$

III. BOSON BASES

A. Boson calculus

From (10) we find the symmetric states of $U(nm)$ are

$$|\,[p0\cdots 0]\rangle = \frac{1}{\sqrt{p!}}\sum_{q\in S_p}[q]\,|\,\Phi^{(j)}_{(i)}\rangle. \quad (11)$$

We may construct linear combinations of these symmetric states of $U(nm)$ from the irreducible bases of $U(n)\times U(m)$ in (8) using the Clebsch–Gordan coefficients of the canonical bases of S_p since the bases in (8) are also irreducible bases $|\,{}^{[u]}_{\langle m\rangle}\rangle \times |\,{}^{[v]}_{\langle n\rangle}\rangle$ of $S_p\times S_p$. We find that

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = \frac{1}{(l[u])^{1/2}}\sum_{n}\left|\begin{matrix}[u]\\ \langle s\rangle\quad(n)\end{matrix}\right\rangle \times \left|\begin{matrix}[u]\\ (n)\quad\langle r\rangle\end{matrix}\right\rangle,$$

$$= \frac{N^{[u]}_{r}N^{[u]}_{s}}{(l[u])^{1/2}}\sum_{n}\bar{P}^{[u]}_{nr}\underline{P}^{[u]}_{ns}\,|\,\Phi^{(j)}_{(i)}\rangle \quad (12)$$

is symmetric under all permutations $[q]$ of S_p as can easily be verified directly:

$$[q]\sum_{n}\bar{P}^{[u]}_{nr}\underline{P}^{[u]}_{ns} = \sum_{n}[\bar{q}]\bar{P}^{[u]}_{nr}[\underline{q}]\underline{P}^{[u]}_{ns}$$

$$= \sum_{n,t,t'}D^{[u]}_{tn}[q]D^{[u]}_{t'n}[q]\bar{P}^{[u]}_{tr}\underline{P}^{[u]}_{t's}$$

$$= \sum_{t}\bar{P}^{[u]}_{tr}\underline{P}^{[u]}_{ts}.$$

Hence, $|\,{}_{\langle s\rangle}{}^{[u]}_{*}{}_{\langle r\rangle}\rangle$ must be some linear combination of symmetric bases (11) of $U(nm)$. We shall denote the subspace of all bases $|\,{}_{\langle s\rangle}{}^{[u]}_{*}{}_{\langle r\rangle}\rangle$ of the direct product space $|\,\Phi^{(j)}_{(i)}\rangle$ as $U(n) * U(m)$. We wish to find the symmetric content of $U(nm)$ in our bases $|\,{}_{\langle s\rangle}{}^{[u]}_{*}{}_{\langle r\rangle}\rangle$ of $U(n) * U(m)$. This is equivalent to determining the symmetric canonical irreducible bases of $U(nm)$ contained in the irreducible basis $|\,{}^{[u]}_{\langle s\rangle}\rangle \times |\,{}^{[u]}_{\langle r\rangle}\rangle$ of $U(n)\times U(m)$.

We shall accomplish this task by using boson operators to generate our bases. We may write our symmetric basis in terms of boson operators as in (13) below:

$$\sum_{q\in S_p}[q]\,|\,\Phi^{(j)}_{(i)}\rangle = \frac{1}{\sqrt{p!}}\sum_{q\in S_p}[q]a^{(j)}_{(i)}\,|\,0\rangle, \quad (13)$$

where $a^{(j)}_{(i)}\,|\,0\rangle = a^{j_1}_{i_1}a^{j_2}_{i_2}\cdots a^{j_p}_{i_p}\,|\,0\rangle$, and $a^{j\dagger}_{i} = \bar{a}^{j}_{i}$. The boson operators obey the following commutation relations:

$$[a^{j}_{i}, a^{j'}_{i'}] = [\bar{a}^{j}_{i}, \bar{a}^{j'}_{i'}] = 0, \quad (14a)$$

$$[\bar{a}^{j}_{i}, a^{j'}_{i'}] = \delta_{jj'}\delta_{ii'}. \quad (14b)$$

We may now expand the generators of $U(n)$, $U(m)$, and $U(nm)$ in terms of the boson operators as follows[15]:

$$E_{im} = \sum_{l=1}^{p} a^{l}_{i}\bar{a}^{l}_{m}, \quad (15a)$$

$$E^{jn} = \sum_{l=1}^{p} a^{j}_{l}\bar{a}^{n}_{l}, \quad (15b)$$

$$E^{jn}_{im} = a^{j}_{i}\bar{a}^{n}_{m}. \quad (15c)$$

From (13) we see that $|\,{}_{\langle s\rangle}{}^{[u]}_{*}{}_{\langle r\rangle}\rangle$ may be put in terms of boson operators as in (16):

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = \frac{N^{[u]}_{r}N^{[u]}_{s}}{(l^{[u]}p!)^{1/2}}\sum_{n}\bar{P}^{[u]}_{nr}\underline{P}^{[u]}_{ns}a^{(j)}_{(i)}\,|\,0\rangle. \quad (16)$$

It follows from (14) that

$$[\bar{q}]a^{(j)}_{(i)}\,|\,0\rangle = [\underline{q}^{-1}]a^{(j)}_{(i)}\,|\,0\rangle \quad (17)$$

and, hence,

$$\bar{P}^{[u]}_{nr}a^{(j)}_{(i)}\,|\,0\rangle = \underline{P}^{[u]}_{rn}a^{(j)}_{(i)}\,|\,0\rangle. \quad (18)$$

Using (18), we may simplify the expression for $|\,{}_{\langle s\rangle}{}^{[u]}_{*}{}_{\langle r\rangle}\rangle$ in (16) to find

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = N^{[u]}_{r}N^{[u]}_{s}(l^{[u]}/p!)^{1/2}\underline{P}^{[u]}_{rs}a^{(j)}_{(i)}\,|\,0\rangle. \quad (19)$$

If we let

$$M^{[u]}_{rs} = N^{[u]}_{r}N^{[u]}_{s}(l^{[u]}/p!)^{1/2}, \quad (20)$$

then we have the following relation between our basis of $U(n) * U(m)$ and the boson operators:

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = M^{[u]}_{rs}\underline{P}^{[u]}_{rs}a^{(j)}_{(i)}\,|\,0\rangle \quad (21a)$$

or, similarly,

$$\left|\begin{matrix}[u]\\ \langle r\rangle * \langle s\rangle\end{matrix}\right\rangle = M^{[u]}_{rs}\bar{P}^{[u]}_{rs}a^{(i)}_{(j)}\,|\,0\rangle. \quad (21b)$$

Equations (21) illustrate the reciprocity between upper and lower projection operators for bases of $U(m) * U(n)$

Since $P^{[u]}_{rs} = C^{[u]}_{rs}O^{[u]}_{rs}$, where $C^{[u]}_{rs}$ is a positive constant, and since the seminormal canonical projection operators $O^{[u]}_{rs}$ may be easily generated, we now have a convenient and straightforward method of finding the symmetric content of $U(nm)$ in the irreducible bases of $U(n)\times U(m)$ which we call the boson calculus. As an example of the use of the boson calculus, we find the highest weight bases $|\,{}^{11}_{2}\rangle \times |\,{}^{11}_{2}\rangle$ of $U(3)\times U(3)$ in terms of boson operators. Using Eq. (21a) with seminormal projection operators, we have

$$\underline{O}^{[210]}_{\substack{12\\3}\ \substack{12\\3}}a^{1}_{1}a^{1}_{1}a^{2}_{2}\,|\,0\rangle = 2\underline{S}_{12}\underline{A}_{13}\underline{S}_{12}a^{1}_{1}a^{1}_{1}a^{2}_{2}\,|\,0\rangle,$$

$$= 8(a^{1}_{1}a^{1}_{1}a^{2}_{2} - a^{1}_{1}a^{1}_{2}a^{2}_{1})\,|\,0\rangle.$$

Normalizing, we find $|\,{}^{11}_{2}\rangle \times |\,{}^{11}_{2}\rangle$ in terms of the symmetric Weyl bases of U(6) with three particles

$$\left|\begin{matrix}11\\2\end{matrix}\right\rangle \times \left|\begin{matrix}11\\2\end{matrix}\right\rangle = \left(\sqrt{2}\left|\begin{matrix}1&1&2\\1&1&2\end{matrix}\right\rangle - \left|\begin{matrix}1&1&2\\1&2&1\end{matrix}\right\rangle\right)\Big/\sqrt{3}. \tag{22}$$

Similarly,

$$\begin{aligned}\underline{O}^{[210]}_{\substack{12\\3}\,\substack{12\\3}} a_1^1a_2^2a_3^2|0\rangle &= 2\underline{S}_{12}\underline{A}_{13}\underline{S}_{12}a_1^1a_2^2a_3^2|0\rangle\\ &= 2(2a_1^1a_2^2a_3^2 + 2a_2^1a_1^2a_3^2 - a_1^1a_3^2a_2^2 - a_2^1a_3^2a_1^2\\ &\quad - a_3^1a_1^2a_2^2 - a_3^1a_2^2a_1^2)|0\rangle, \end{aligned} \tag{23a}$$

$$\begin{aligned}\underline{O}^{[210]}_{\substack{12\\3}\,\substack{13\\2}} a_1^1a_2^2a_3^2|0\rangle &= 4\underline{S}_{12}[\underline{23}]\underline{A}_{12}a_1^1a_2^2a_3^2|0\rangle\\ &= 4(a_1^1a_3^2a_2^2 + a_3^1a_1^2a_2^2 - a_2^1a_3^2a_1^2 - a_3^1a_2^2a_1^2|0\rangle, \end{aligned} \tag{23b}$$

$$\begin{aligned}\underline{O}^{[100]}a_1^1a_2^2a_3^2|0\rangle &= 4\underline{S}_{123}a_1^1a_2^2a_3^2|0\rangle\\ &= (a_1^1a_2^2a_3^2 + a_1^1a_3^2a_2^2 + a_3^1a_1^2a_2^2 + a_3^1a_2^2a_1^2\\ &\quad + a_2^1a_3^2a_1^2 + a_2^1a_1^2a_3^2)|0\rangle. \end{aligned} \tag{23c}$$

Collecting terms and normalizing, we find

$$\begin{matrix}\left|\begin{matrix}12\\3\end{matrix}\right\rangle\times\left|\begin{matrix}12\\2\end{matrix}\right\rangle\\ \left|\begin{matrix}13\\2\end{matrix}\right\rangle\times\left|\begin{matrix}12\\2\end{matrix}\right\rangle\\ |123\rangle\times|122\rangle\end{matrix} = \begin{pmatrix}-2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6}\\ 0 & -1/\sqrt{2} & 1/\sqrt{2}\\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\end{pmatrix} \begin{matrix}\left|\begin{matrix}2&2&1\\1&2&3\end{matrix}\right\rangle\\ \left|\begin{matrix}2&1&2\\1&2&3\end{matrix}\right\rangle\\ \left|\begin{matrix}1&2&2\\1&2&3\end{matrix}\right\rangle\end{matrix} \tag{24}$$

The bases generated by our boson calculus must be identical to those generated by Baird and Biedenharn[16] by antisymmetrizing the columns of the "boson tableau". However, column antisymmetrization can only be applied to derive certain bases, and merely represents a simplification of the permutational content of the canonical projection operators when acting on such bases. Column antisymmetrization can be used for all bases $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[u]}_{\langle r\rangle}\rangle$ where both $|{}^{[u]}_{\langle s\rangle}\rangle$ and $|{}^{[u]}_{\langle r\rangle}\rangle$ have nondegenerate weights. For example, using column symmetrization to generate the basis in Eq. (22), we have

$$\begin{aligned}\left|\begin{matrix}11\\2\end{matrix}\right\rangle\times\left|\begin{matrix}11\\2\end{matrix}\right\rangle &= \frac{1}{\sqrt{3}}\begin{bmatrix}a_1^1 & a_1^1\\ a_2^2 & \end{bmatrix}|0\rangle\\ &= \frac{1}{\sqrt{3}}(a_1^1a_2^2 - a_2^1a_1^2)|a_1^1|0\rangle\\ &= \frac{1}{\sqrt{3}}(a_1^1a_1^1a_2^2 - a_1^1a_2^1a_1^2)|0\rangle. \end{aligned} \tag{25}$$

We have antisymmetrized with respect to the subscripts of the columns in the "boson tableau." Because of the reciprocity in Eqs. (21), we could have equally well antisymmetrized with respect to the superscripts.

In the case where one of the bases $|{}^{[u]}_{\langle s\rangle}\rangle$ or $|{}^{[u]}_{\langle r\rangle}\rangle$ has a semimaximum weight [a highest weight for $U(n-1)$ or $U(m-1)$ respectively] and the other has a nondegenerate weight, we may again use column antisymmetrization to derive a boson basis. Thus, corresponding to our previous result, we have

$$\begin{aligned}\left|\begin{matrix}13\\2\end{matrix}\right\rangle\times\left|\begin{matrix}12\\2\end{matrix}\right\rangle &= \frac{1}{\sqrt{2}}\begin{bmatrix}a_1^1 & a_3^2\\ a_2^2 & \end{bmatrix}|0\rangle\\ &= \frac{1}{\sqrt{2}}(a_1^1a_2^2 - a_2^1a_1^2)a_3^2|0\rangle\\ &= \frac{1}{\sqrt{2}}(a_1^1a_2^2a_3^2 - a_1^2a_2^1a_3^2)|0\rangle. \end{aligned} \tag{26}$$

However, in most cases where $|{}^{[u]}_{\langle s\rangle}\rangle$ or $|{}^{[u]}_{\langle r\rangle}\rangle$ have degenerate weights, column antisymmetrization fails to generate an orthonormal basis, and lowering operator techniques must be employed. Thus, to find $|{}^{12}_{3}\rangle \times |{}^{12}_{2}\rangle$, we must lower the basis $|{}^{11}_{3}\rangle\times|{}^{12}_{2}\rangle$ as shown below:

$$\begin{aligned}\left|\begin{matrix}12\\3\end{matrix}\right\rangle\times\left|\begin{matrix}12\\2\end{matrix}\right\rangle &= \frac{1}{\sqrt{2}}E_{21}\left|\begin{matrix}11\\3\end{matrix}\right\rangle\times\left|\begin{matrix}12\\2\end{matrix}\right\rangle,\\ &= \frac{E_{21}}{\sqrt{6}}\begin{bmatrix}a_1^1 & a_1^2\\ a_3^2 & \end{bmatrix}|0\rangle,\\ &= \frac{1}{\sqrt{6}}(-2a_1^2a_2^2a_3^1 + a_1^2a_2^1a_3^2 + a_1^1a_2^2a_3^2)|0\rangle. \end{aligned} \tag{27}$$

In general, this lowering technique is very tedious and our boson calculus represents a considerable simplification for deriving the boson bases.

B. Weyl bases

Let $m=p$ and $|{}_{\langle s\rangle *\langle r\rangle}^{[u]}\rangle$ be a basis of $U(n)*U(p)$, where $\langle r\rangle$ has a weight with maximum degeneracy; that is, let $\psi^{(j)} = \psi_1^1\psi_2^2\cdots\psi_p^p$ so that the standard tableau of $U(p)$, $T^{[u]}_{\langle r\rangle}$, is the same as the standard tableau of S_p, $T^{[u]}_{(r)}$. Then from Eq. (21a) we have

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = M^{[u]}_{rs}\underline{P}^{[u]}_{rs}a^1_{i_1}a^2_{i_2}\cdots a^p_{i_p}|0\rangle. \tag{28}$$

Comparing this with Eq. (4a), we see there is a one-to-one correspondence between the boson bases $|{}_{\langle s\rangle *\langle r\rangle}^{[u]}\rangle$ and Weyl bases $|{}_{\langle s\rangle}^{[u]}{}_{(r)}\rangle$. Since $|{}_{\langle s\rangle *\langle r\rangle}^{[u]}\rangle$ transforms like a Weyl basis under permutations $[\underline{q}]$ and generators E_{ij}, we have

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = \left|\begin{matrix}[u]\\ \langle s\rangle \quad (r)\end{matrix}\right\rangle \tag{29a}$$

when $T^{[u]}_{\langle r\rangle} = T^{[u]}_{(r)}$.

Also,

$$M^{[u]}_{rs} = N^{[u]}_{s} \tag{29b}$$

when $T^{[u]}_{\langle r\rangle} = T^{[u]}_{(r)}$.

The commutivity of boson operators in Eq. (28) illustrates the fact that a reordering of the notation for single particle states leaves the Weyl basis unchanged. It is evident that the Weyl bases $|{}_{\langle s\rangle *\langle r\rangle}^{[u]}\rangle$ form a subspace $U(n)*S_p \subset U(n)*U(p)$.

Now let $n=p$ and $|{}_{\langle s\rangle *\langle r\rangle}^{[u]}\rangle$ be a basis of $U(p)*U(m)$, where $\langle s\rangle$ has a weight with maximum degeneracy; that is, let $\phi_{(i)} = \phi_1^1\phi_2^2\cdots\phi_p^p$ so that $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}$. Then from Eq. (21b) we have

$$\left|\begin{matrix}[u]\\ (s) * \langle r\rangle\end{matrix}\right\rangle = M^{[u]}_{rs}\overline{P}^{[u]}_{sr}a_1^{j_1}a_2^{j_2}\cdots a_p^{j_p}|0\rangle. \tag{30}$$

Comparing this with Eq. (4b), we see there is a one-to-one corespondence between the boson bases $|{}_{(s) *\langle r\rangle}^{[u]}\rangle$ and the Weyl bases $|{}_{(s)}^{[u]}{}_{\langle r\rangle}\rangle$. Since $|{}_{(s) *\langle r\rangle}^{[u]}\rangle$ transforms like a Weyl basis under permutations $[\overline{q}]$ and generators E^{ij}, we have

$$\left|\begin{matrix}[u]\\ \langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = \left|\begin{matrix}[u]\\ (s) \quad \langle r\rangle\end{matrix}\right\rangle \tag{31a}$$

when $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}$.

Also,

$$M^{[u]}_{rs} = N^{[u]}_r \tag{31b}$$

when $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}$.

It is evident that the Weyl bases $|{}^{[u]}_{\langle s\rangle *\langle r\rangle}\rangle$ form a subspace $S_p * \mathrm{U}(m) \subset U(p) * \mathrm{U}(m)$. The reason for our choice of notation in (4b) is now clear.

Moshinsky[17] has shown that for the "special" Gel'fand bases $|{}_{\langle s\rangle}^{[u]}{}_{(r)}\rangle$ of $\mathrm{U}(n) * S_p$

$$E^{nm}E^{mn} = (\overline{1}) + (\overline{nm}), \tag{32a}$$

and similarly for the "special" Gel'fand bases $|{}_{(r)}^{[u]}{}_{\langle s\rangle}\rangle$ of $S_p * \mathrm{U}(m)$:

$$E_{nm}E_{mn} = (\underline{1}) + (\underline{nm}). \tag{32b}$$

Equation (32a) is easily verified since both generators E^{mn} of $\mathrm{U}(n)$ and *state* permutations $(\overline{q})$ of S_p commute with the *particle* operators $\underline{P}^{[u]}_{rs}$. Thus, we have

$$\begin{aligned}
E^{nm}E^{mn}\underline{P}^{[u]}_{rs}|\phi_{(i)}\rangle
&= \underline{P}^{[u]}_{rs}E^{nm}E^{mn}|\phi^{1\,2\cdots n\cdots m\cdots p}_{i_1 i_2\cdots i_n\cdots i_m\cdots i_p}\rangle\\
&= \underline{P}^{[u]}_{rs}[|\phi^{1\,2\cdots n\cdots m\cdots p}_{i_1 i_2\cdots i_n\cdots i_m\cdots i_p}\rangle + |\phi^{1\,2\cdots n\cdots m\cdots p}_{i_1 i_2\cdots i_n\cdots i_m\cdots i_p}\rangle]\\
&= \underline{P}^{[u]}_{rs}[(\overline{1}) + (\overline{nm})]|\phi^{1\,2\cdots n\cdots m\cdots p}_{i_1 i_2\cdots i_n\cdots i_m\cdots i_p}\rangle\\
&= [(\overline{1}) + (\overline{nm})]\underline{P}^{[u]}_{rs}|\phi_{(i)}\rangle
\end{aligned}$$

Equation (32b) may be verified in a similar manner. Other more complicated expressions may also be derived for the r-cycles of S_p in terms of the generators of $\mathrm{U}(n)$ or $\mathrm{U}(m)$ when operating on these "special" Gel'fand bases. However, it is more important to note that the upper and lower Gel'fand invariant operators

$$\overline{I}^m_k = \sum^m_{i_1, i_2, \ldots, i_k} E^{i_1i_2}E^{i_2i_3}\cdots E^{i_ki_1}, \tag{33a}$$

$$\underline{I}^n_k = \sum^n_{i_1, i_2, \ldots, i_k} E_{i_1i_2}E_{i_2i_3}\cdots E_{i_ki_1} \tag{33b}$$

of $\mathrm{U}(m)$ and $\mathrm{U}(n)$ may be expanded in terms of the upper and lower *state* r-cycle class operators of S_p, $\overline{K}^{pm}_r$ for $r = 1, 2, \ldots, k$ and $\underline{K}^{pn}_r$ for $r = 1, 2, \ldots, k$ respectively as has been shown in I. The boson bases $|{}^{[u]}_{\langle s\rangle *\langle r\rangle}\rangle$ in (21a) are eigenvectors of these upper and lower class operators since these bases transform like irreducible bases $|{}^{[u]}_{(s)}\rangle$ of S_p under lower permutations $(\underline{q})$, and like irreducible bases $|{}^{[u]}_{(r)}\rangle$ of S_p under upper permutations $(\overline{q})$. It is for this reason that the projected bases $|{}^{[u]}_{\langle s\rangle *\langle r\rangle}\rangle$ forms a Gel'fand bases $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[u]}_{\langle r\rangle}\rangle$ of $\mathrm{U}(n) \times \mathrm{U}(m)$ for different standard tableaus $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}\phi_{(i)}$ of $\mathrm{U}(n)$ and different standard tableaus $T^{[u]}_{\langle r\rangle} = T^{[u]}_{(r)}\psi^{(j)}$ of $\mathrm{U}(m)$.

Finally, let $m = p$, $n = p$, and $T^{[u]}_{\langle r\rangle} = T^{[u]}_{(r)}$, $T^{[u]}_{\langle s\rangle} = T^{[u]}_{(s)}$. Then the boson basis

$$\left|\begin{matrix}[u]\\(s) * (r)\end{matrix}\right\rangle = M^{[u]}_{rs}\underline{P}^{[u]}_{rs}a^1_1a^2_2\cdots a^p_p|0\rangle$$

is a canonical irreducible basis of S_p under permutations $[\underline{q}]$ and $[\overline{q}]$. The bases $|{}^{[u]}_{(s) *(r)}\rangle$ form a subspace $S_p * S_p \subset \mathrm{U}(p) * \mathrm{U}(p)$ and are a bases for the regular representation of S_p.

From Eqs. (29b) and (20) we have the relation

$$(l^{[u]}/p!)^{1/2}N^{[u]}_rN^{[u]}_s = N^{[u]}_s$$

when $T^{[u]}_{\langle r\rangle} = T^{[u]}_{(r)}$. Therefore,

$$N^{[u]}_r = (p!/l^{[u]})^{1/2} = \sqrt{H([u])}. \tag{34}$$

when $T^{[u]}_{(r)} = T^{[u]}_{\langle r\rangle}$. This also follows directly from evaluating (3.25) of I.

C. Factorization lemma[18]

One of the most important aspects of the boson calculus is that we may use it to determine the matrix elements of the IR's of the unitary group. Then, by means of the factorization lemma, we can generate the multiplicity-free Clebsch–Gordan coefficients of the unitary group.

Let $D^{[1]}(U)$ be the fundamental or self-representation of $\mathrm{U}(n)$ given by

$$D^{[1]}(U) = \begin{pmatrix} u^1_1 & u^2_1 & \cdots & u^n_1\\ u^1_2 & u^2_2 & \cdots & u^n_2\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ u^1_n & u^2_n & \cdots & u^n_n\end{pmatrix}, \tag{35}$$

and let $n = m$ so that $|{}^{[u]}_{\langle s\rangle *\langle r\rangle}\rangle$ is a basis of $\mathrm{U}(n) * \mathrm{U}(n)$. We multiply the boson bases $|{}^{[u]}_{\langle s\rangle *\langle r\rangle}\rangle$ by the constant $L([u])$ such that

$$L([u])\left|\begin{matrix}[u]\\\langle s\rangle * \langle r\rangle\end{matrix}\right\rangle = L([u])M^{[u]}_{rs}\underline{P}^{[u]}_{rs}a^{(j)}_{(i)}|0\rangle$$

contains the term $a^{(j)}_{(i)}|0\rangle$ just once when (i) and (j) have highest weight in $\mathrm{U}(n)$. Then Louck has proven that[19]

$$D^{[u]}_{\langle s\rangle\langle r\rangle}(U) = L([u])M^{[u]}_{rs}\underline{P}^{[u]}_{rs}u^{(j)}_{(i)}, \tag{36}$$

where $u^{(j)}_{(i)} = u^{j_1}_{i_1}u^{j_2}_{i_2}\cdots u^{j_p}_{i_p}$. For example, from Eq. (25) we have that $L([210]) = \sqrt{3}$. From (24) it follows that

$$\begin{matrix} D^{[210]}_{\substack{12\\3}\,\substack{12\\2}}(U)\\ D^{[210]}_{\substack{13\\2}\,\substack{12\\2}}(U)\\ D^{[300]}_{123\;122}(U)\end{matrix} = \begin{pmatrix} -\sqrt{2} & \sqrt{2}/2 & \sqrt{2}/2\\ 0 & -\sqrt{3}/\sqrt{2} & \sqrt{3}/\sqrt{2}\\ 1 & 1 & 1\end{pmatrix} \begin{matrix} u^2_1u^2_2u^1_3\\ u^2_1u^1_2u^2_3\\ u^1_1u^2_2u^2_3\end{matrix}. \tag{37}$$

Ciftan and Biedenharn have shown that[20]

$$L([u]) = \sqrt{H([u])}, \tag{38}$$

where $H([u])$ is the product of hook-lengths described in I.

We now have an explicit means of calculating the canonical Clebsch–Gordan coefficients of $\mathrm{U}(n)$ by using the factorization lemma. Let

$$D^{[v]}_{\langle q\rangle\langle p\rangle}(a) = \sqrt{H([v])}\,M^{[v]}_{pq}\underline{P}^{[v]}_{pq}a^{(j)}_{(i)}.$$

The factorization lemma can then be written as

$$\left\langle\begin{matrix}[u]\\\langle s\rangle * \langle r\rangle\end{matrix}\right|D^{[v]}_{\langle q\rangle\langle p\rangle}(a)\left|\begin{matrix}[\lambda]\\\langle n\rangle * \langle m\rangle\end{matrix}\right\rangle = \left(\frac{H([u])}{H([\lambda])}\right)^{1/2}\sum_{\delta} C^{[v][\lambda][u]^\delta}_{\langle q\rangle\langle n\rangle\langle s\rangle}C^{[v][\lambda][u]^\delta}_{\langle p\rangle\langle m\rangle\langle r\rangle} \tag{39}$$

where $[u]^\delta$ is the δth IR $[u]$ contained in the direct product $[v]\times[\lambda]$. Since the left-hand side of (39) can be calculated explicitly using the seminormal canonical projection operators, we can directly evaluate the product of Clebsch–Gordan coefficients of the unitary group on the right of (39). Because of the sum on the right on (39), only the multiplicity-free coefficients can be uniquely determined.

IV. FERMION BASES

From (10) we find the antisymmetric states of U(nm) are

$$|[11\cdots 1]\rangle = \frac{1}{\sqrt{p!}} \sum_{q\in S_p} \epsilon_q [q] \,|\Phi^{(j)}_{(i)}\rangle. \tag{40}$$

We may construct linear combinations of these antisymmetric states of U(nm) from the irreducible bases of U(n)×$U(m)$ using the Clebsch–Gordan coefficients of the canonical bases of S_p. We find that

$$\left|\begin{matrix}[u]\\ \langle s\rangle \,\tilde{*}\, \langle\tilde{r}\rangle\end{matrix}\right\rangle = \frac{1}{(l^{[u]})^{1/2}} \sum_n \epsilon_{\sigma_{nm}} \left|\begin{matrix}[u]\\ \langle s\rangle\ (n)\end{matrix}\right\rangle \times \left|\begin{matrix}[\tilde{u}]\\ (\tilde{n})\ \langle\tilde{r}\rangle\end{matrix}\right\rangle,$$

$$= \frac{N^{[\tilde{u}]}_{\tilde{r}} N^{[u]}_s}{(l^{[u]})^{1/2}} \sum_n \epsilon_{\sigma_{nm}} \bar{P}^{[\tilde{u}]}_{\tilde{n}\tilde{r}} \underline{P}^{[u]}_{ns} \,|\Phi^{(j)}_{(i)}\rangle \tag{41}$$

is antisymmetric under all permutations $[q]$ of S_p for any standard tableau $T^{[u]}_{(m)}$. This can be shown, using (1.23) of I, since

$$[q] \sum_n \epsilon_{\sigma_{nm}} \bar{P}^{[\tilde{u}]}_{\tilde{n}\tilde{r}} \underline{P}^{[u]}_{ns} = \sum_n \epsilon_{\sigma_{nm}} [\bar{q}] \bar{P}^{[\tilde{u}]}_{\tilde{n}\tilde{r}} [\underline{q}] \underline{P}^{[u]}_{ns}$$

$$= \sum_{n,t,t'} \epsilon_{\sigma_{nm}} \epsilon_{\sigma_{tn}} \epsilon_q D^{[u]}_{tn}[q] D^{[u]}_{t'n}[q] \bar{P}^{[\tilde{u}]}_{\tilde{t}\tilde{r}} \underline{P}^{[u]}_{t's}$$

$$= \epsilon_q \sum_t \epsilon_{\sigma_{tm}} \bar{P}^{[\tilde{u}]}_{\tilde{t}\tilde{r}} \underline{P}^{[u]}_{ts}.$$

Hence, $|{}^{[u]}_{\langle s\rangle * \langle\tilde{r}\rangle}\rangle$ must be some linear combination of antisymmetric bases (40) of U(nm). From (8) we see the basis $|{}^{[u]}_{\langle s\rangle \tilde{*} \langle\tilde{r}\rangle}\rangle$ is an irreducible basis $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[\tilde{u}]}_{\langle\tilde{r}\rangle}\rangle$ of U(n)×U(m). We shall denote the subspace of all bases $|{}^{[u]}_{\langle s\rangle \tilde{*} \langle\tilde{r}\rangle}\rangle$ of the direct product space $|\Phi^{(j)}_{(i)}\rangle$ as U(n) $\tilde{*}$ U(m).

It is important to determine the antisymmetric content of U(nm) in our bases $|{}^{[u]}_{\langle s\rangle \tilde{*} \langle\tilde{r}\rangle}\rangle$ of U(n) $\tilde{*}$ U(m), for this will be equivalent to determining the antisymmetric irreducible bases of U(nm) contained in the irreducible basis $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[\tilde{u}]}_{\langle\tilde{r}\rangle}\rangle$ of U(n)×U(m). In what follows we shall show a simple and straightforward means of finding this antisymmetric content.

For this purpose it is convenient to use fermion operators to generate our bases. We may write our antisymmetric basis in terms of fermion operators as in (42) below:

$$\sum_{q\in S_p} \epsilon_q [q] \,|\Phi^{(j)}_{(i)}\rangle = \frac{1}{\sqrt{p!}} \sum_{q\in S_p} \epsilon_q [q] a^{(j)}_{(i)} |0\rangle, \tag{42}$$

where $a^{(j)}_{(i)}|0\rangle = a^{j_1}_{i_1} a^{j_2}_{i_2} \cdots a^{j_p}_{i_p} |0\rangle$, and $a^{j\dagger}_i = \bar{a}^j_i$. The fermion operators obey the following anticommutation relations:

$$[a^j_i, a^{j'}_{i'}]_+ = [\bar{a}^j_i, \bar{a}^{j'}_{i'}]_+ = 0, \tag{43a}$$

$$[\bar{a}^j_i, a^{j'}_{i'}]_+ = \delta_{jj'} \delta_{ii'}. \tag{43b}$$

We may also expand the generators of U(n), U(m), and U(nm) in terms of the fermion operators as in (15).

From (42) we see that $|{}^{[u]}_{\langle s\rangle \tilde{*} \langle\tilde{r}\rangle}\rangle$ may be put in terms of fermion operators as in (44):

$$\left|\begin{matrix}[u]\\ \langle s\rangle \,\tilde{*}\, \langle\tilde{r}\rangle\end{matrix}\right\rangle = \frac{N^{[\tilde{u}]}_{\tilde{r}} N^{[u]}_s}{(l^{[u]} p!)^{1/2}} \sum_n \epsilon_{\sigma_{nm}} \bar{P}^{[\tilde{u}]}_{\tilde{n}\tilde{r}} \underline{P}^{[u]}_{ns} a^{(j)}_{(i)} |0\rangle. \tag{44}$$

By using the anticommutation relations (43), it follows that

$$[\bar{q}] a^{(j)}_{(i)} |0\rangle = \epsilon_q [\underline{q}^{-1}] a^{(j)}_{(i)} |0\rangle. \tag{45}$$

From the above relation and (1.23) of I, we have

$$\bar{P}^{[\tilde{u}]}_{\tilde{n}\tilde{r}} a^{(j)}_{(i)} |0\rangle = \epsilon_{\sigma_{nr}} \underline{P}^{[u]}_{rn} a^{(j)}_{(i)} |0\rangle. \tag{46}$$

Using (46), we may simplify the expression for $|{}^{[u]}_{\langle s\rangle \tilde{*} \langle\tilde{r}\rangle}\rangle$ in (44) to find

$$\left|\begin{matrix}[u]\\ \langle s\rangle \,\tilde{*}\, \langle\tilde{r}\rangle\end{matrix}\right\rangle = N^{[\tilde{u}]}_{\tilde{r}} N^{[u]}_s (l^{[u]}/p!)^{1/2} \epsilon_{\sigma_{rm}} \underline{P}^{[u]}_{rs} a^{(j)}_{(i)} |0\rangle.$$

If we let

$$M^{[u]}_{\tilde{r}s} = N^{[\tilde{u}]}_{\tilde{r}} N^{[u]}_s (l^{[u]}/p!)^{1/2},$$

then we have the following simple relation between our basis of U(n) $\tilde{*}$ U(m) and the fermion operators:

$$\left|\begin{matrix}[u]\\ \langle s\rangle \,\tilde{*}\, \langle\tilde{r}\rangle\end{matrix}\right\rangle = \epsilon_{\sigma_{rm}} M^{[u]}_{\tilde{r}s} \underline{P}^{[u]}_{rs} a^{(j)}_{(i)} |0\rangle, \tag{47a}$$

or similarly,

$$\left|\begin{matrix}[u]\\ \langle\tilde{r}\rangle \,\tilde{*}\, \langle s\rangle\end{matrix}\right\rangle = \epsilon_{\sigma_{rm}} M^{[u]}_{\tilde{r}s} \bar{P}^{[u]}_{rs} a^{(i)}_{(j)} |0\rangle. \tag{47b}$$

Equations (47) illustrate the reciprocity between upper and lower projection operators for bases of U(m) $\tilde{*}$ U(n) and U(n) $\tilde{*}$ U(m).

We now have a convenient and straightforward method of finding the antisymmetric content of U(nm) in the irreducible bases of U(n)×U(m) which we call the fermion calculus. From the expression $\underline{P}^{[u]}_{rs} a^{(j)}_{(i)}|0\rangle$, we may find the antisymmetric content of U(nm) in the bases $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[\tilde{u}]}_{\langle\tilde{r}\rangle}\rangle$ when the $a^{(j)}_{(i)}$ are fermion operators, or we may find the symmetric content of U(nm) in the bases $|{}^{[u]}_{\langle s\rangle}\rangle \times |{}^{[u]}_{\langle r\rangle}\rangle$ when the $a^{(j)}_{(i)}$ are boson operators. Hence, the fermion calculus can be generated by the seminormal projection operators $O^{[u]}_{rs}$ acting on fermion operators, and the boson calculus can be generated by the same seminormal projection operators acting on boson operators.

To illustrate this point, we use the results of Eqs. (23a) and (23b) to find the antisymmetric content of $|{}^{12}_{3}\rangle \times |{}^{12}_{2}\rangle$ and $|{}^{13}_{2}\rangle \times |{}^{12}_{2}\rangle$ respectively by letting the $a^{(j)}_{(i)}$ be fermion operators. Our case is somewhat special since $\langle r\rangle = \langle\tilde{r}\rangle$. The fermion basis analogous to Eq. (23c) is shown below.

$$\underline{O}^{[100]} a^1_1 a^2_2 a^2_3 |0\rangle = 4\underline{A}_{123} a^1_1 a^2_2 a^2_3 |0\rangle,$$

$$= 4(a^1_1 a^2_2 a^2_3 - a^1_1 a^2_3 a^2_2 + a^1_3 a^2_1 a^2_2 - a^1_3 a^2_2 a^2_1 + a^1_2 a^2_3 a^2_1 - a^1_2 a^2_1 a^2_3) |0\rangle.$$

For convenience we let $\psi^1 = \psi^+$, $\psi^2 = \psi^-$, $\Phi^1_1 = \Phi_{1^+}$, etc. Collecting terms and normalizing the above bases, we find

$$\begin{matrix} \left|\begin{matrix}12\\3\end{matrix}\right\rangle \times \left|\begin{matrix}+\\-\end{matrix}\right\rangle^- \\[2ex] \left|\begin{matrix}13\\2\end{matrix}\right\rangle \times \left|\begin{matrix}+\\-\end{matrix}\right\rangle^- \\[2ex] \left|\begin{matrix}1\\2\\3\end{matrix}\right\rangle \times |+\,-\,-\rangle \end{matrix} = \begin{pmatrix} 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} \begin{matrix} \left|\begin{matrix}1^-\\2^-\\3^+\end{matrix}\right\rangle \\[2ex] \left|\begin{matrix}1^-\\2^+\\3^-\end{matrix}\right\rangle \\[2ex] \left|\begin{matrix}1^+\\2^-\\3^-\end{matrix}\right\rangle \end{matrix}. \tag{48}$$

*Work supported by Centro Técnico Aeroespacial.
[1]C. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963); J. Math. Phys. 5, 1723, 1730 (1964).
[2]M. Moshinsky, J. Math. Phys. 7, 691 (1966).
[3]M. Ciftan and L. C. Biedenharn, J. Math. Phys. 10, 221 (1969).
[4]J. D. Louck, Am. J. Phys. 38, 3 (1970).
[5]T. H. Seligman, J. Math. Phys. 13, 876 (1972).
[6]J. D. Louck, J. Math. Phys. 6, 1786 (1965).
[7]G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).
[8]J. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965).
[9]M. Moshinsky and T. H. Seligman, Ann. Phys. (N. Y.) 66, 311 (1971).
[10]W. J. Holman III, Nuovo Cimento 4, 904 (1971).
[11]C. W. Patterson and W. G. Harter, J. Math. Phys. 17, 1125 (1976), preceding paper.
[12]M. Moshinsky, J. Math. Phys. 7, 691 (1966).
[13]L. C. Biedenharn, A. Giovannini, and J. C. Louck, J. Math. Phys. 8, 691 (1967).
[14]W. G. Harter and C. W. Patterson, Phys. Rev. A 13, 1067 (1976).
[15]For a good review of the boson basis see J. D. Louck, J. Am. Phys. 38, 36 (1970).
[16]G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).
[17]M. Moshinsky, J. Math. Phys. 7, 691 (1966).
[18]Ref. 13 or Ref. 15, p. 39.
[19]Ref. 15, p. 39.
[20]Ref. 3, p. 224.

Uniform bounds of the Schwinger functions in boson field models

Yong Moon Park*

Fakultät für Physik, Universität Bielefeld, 48 Bielefeld, West Germany
(Received 8 December 1975)

We study the lattice and space cutoff boson field models with periodic boundary condition in d-dimensional space–time ($d \in N^+$). We prove that if the pressures of the interaction (and also the pressures of the interaction with linear external fields) under consideration are bounded uniformly in the cutoffs, the corresponding Schwinger functions are also bounded uniformly in the cutoffs. By applying the above result we prove the uniform bounds of the space cutoff Schwinger functions for the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model and the lattice and space cutoff Schwinger functions for the exponential type interactions in d-dimensional space–time.

1. INTRODUCTION

In the Euclidean strategy of constructive quantum field theory, the uniform bounds of the Schwinger functions (Euclidean Green's functions) is the first step in completing the program of constructing relativistic quantum field theories.[1–6] It has been suggested that control of pressures gives control of the Schwinger functions.[4,7,8] We consider the lattice and space cutoff boson field models with periodic boundary condition in d-dimensional space—time, $d \in N^+$. We show that, if the pressures of the interaction under consideration (and also the pressures of the interaction with linear external fields) are bounded uniformly in the cutoffs, it follows that the corresponding Schwinger functions are also bounded uniformly in the cutoffs. We apply our result to obtain uniform bounds of the Schwinger functions in the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model with space cutoff and in the exponential type interactions with lattice and space cutoffs in d-dimensional space—time.

For the construction of the ϕ_4^4 field model, Glimm and Jaffe[9] have shown that the bounds of the two-point Schwinger functions give bounds of n-point Schwinger functions. For other possible suggestions in this subject, we refer to Fröhlich, Guerra, and Schrader.[4,7,10] In two- and three-dimensional boson field models, it is easier to control the pressures than the Schwinger functions.[5,8,11,12] Our result implies that to prove the existence of the ϕ_4^4 field theory it suffices to show the uniform bounds of the pressures in the lattice and space cutoffs.

The organization of this paper is as follows. In Sec. 2 we introduce notation and definitions for the lattice fields and interaction measures with periodic boundary condition in d-dimensional space—time. We then state our main result and give its proof. The main idea we will use is the method of transfer matrix, Nelson's symmetry and the first Griffiths inequality.[4–6,11,13,14] In Sec. 3 we show that the space cutoff Schwinger functions of the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ field models are bounded uniformly in the space cutoff. The above result will follow as a consequence of Theorem 2.1 (the method of its proof) and the results in Refs. 11 and 13. We also derive uniform bounds of the lattice and space cutoff Schwinger functions of the exponential type interactions in d-dimensional space—time, $d \in N^+$. This means that the exponential type boson self-interacting theory exists in any dimension of the space—time.

While we were undertaking this work, we have come to know the beautiful results of Seiler and Simon on the uniform bounds of the Schwinger functions of the Yukawa$_2$ and $\lambda\phi_3^4$ field model (with free boundary condition).[2] Some of the methods we use are very similar to those in Ref. 12. For the $\lambda\phi_3^4$ model we think that our method is simple and direct. We do not use any other property of the pressures than the uniform boundedness.

The method we use in Sec. 2 might be extended to give uniform bounds of correlation functions of the generalized classical ferromagnetic interactions if the interactions are of finite range. It might also be interesting to study the exponential type interactions in more detail. We plan to make studies in this direction in a forthcoming paper.

2. UNIFORM BOUNDS OF THE SCHWINGER FUNCTIONS

In this section we first introduce the lattice field models with periodic boundary conditions. We then state our main theorem and give the proof. In d-dimensional Euclidean space, R^d, let $\Lambda \subset R^d$ be a d-dimensional cube of volume $|\Lambda| = l^d$ centered at the origin, where $l = 2^n$ for some $n \in N^+$. We consider the lattice approximation with periodic boundary condition on $\partial\Lambda$. We assume that the lattice spacing parameter δ has the form $\delta = 2^{-m}$, $m \in N^+$. Our results in this section hold for more general shapes of boxes in R^d by a straight modification of our method. Let T_Λ be the torus obtained by identifying opposite sides of Λ. We then denote

$$\Lambda_\delta = \{n\delta \mid n = (n_1, \ldots, n_d) \in Z^d, n\delta \in T_\Lambda\},$$
$$Z_\Lambda^d = (2\pi/l)Z^d, \qquad (2.1)$$
$$\Delta^d(l_1, \ldots, l_d) = [-l_1/2, l_1/2] \times \cdots \times [-l_d/2, l_d/2] \subset \Lambda.$$

We note that $\Lambda = \Delta^d(l, \ldots, l)$. Following Refs. 4, 5, 15, and 16, we introduce the free lattice fields $\phi_\delta(n\delta)$ as the real Gaussian random processes indexed by the lattices in Λ_δ, with mean zero and covariance given by

$$\langle \phi_\delta(n\delta)\phi_\delta(n'\delta)\rangle = (z_{\Lambda,\delta})^{-1} S_\delta(n\delta - n'\delta).$$

Here the free two-point function is given by

$$S_\delta(n\delta - n'\delta) = (2\pi)^{-d}\left(\frac{\delta}{l}\right)^d \sum_{\substack{k_\Lambda \in Z_\Lambda^d \\ |k_{\Lambda,i}| \le \pi/\delta}} \exp[ik_\Lambda \cdot (n\delta - n\delta)]\mu_\delta^{-2}(k_\Lambda),$$

where

$$\mu_\delta^2(k_\Lambda)=\delta^{-2}\left(2d-2\sum_{i=1,\cdots,d}\cos(\delta k_{\Lambda,i})\right)+m^2$$

$$\approx k^2+m^2 \quad \text{as } \delta\to 0 \text{ and } l\to\infty.$$

For the field strength renormalization constant we have $z_{\Lambda,\delta}=1$ for $d=1, 2, 3$, and $0<z_{\Lambda,\delta}\le 1$ for $d=4$ depending on models. For a more detailed discussion of $z_{\Lambda,\delta}$, we refer the reader to Ref. 4. We now introduce the interacting action for boson field models by

$$V_{\Lambda,\delta}=\delta^d\sum_{n\delta\in\Lambda_\delta}[P(\phi_\delta(n\delta))+R_{\Lambda,\delta}(\phi_\delta(n\delta))]+E_{\Lambda,\delta},$$

where $P(x)=P_e(x)-\lambda x$, $\lambda\ge 0$. $P_e(x)$ is a given semi-bounded even function (throughout the paper we assume that $P_e(x)$ is a polynomial or exponential type so that the first Griffiths inequality holds for the model[6]) depending on the models (for example, $P(x)=x^4$ for the ϕ_d^4 model), and $R_{\Lambda,\delta}$ and $E_{\Lambda,\delta}$ are renormalization counter terms. Here we have written $E_{\Lambda,\delta}$ for all constant (scalar) counter terms. For the $P(\phi)_2$ theory only counter terms coming from Wick ordering are necessary. For the $(\lambda\phi^4-\sigma\phi^2-\mu\phi)_3$ field model[1,8,11,12,15–18] one should also introduce mass and vacuum counter terms. For ϕ_4^4 theory the detailed form of $V_{\Lambda,\delta}$ is not known.

Following Ref. 4 we define smeared fields by either

$$\phi_\delta(f)=\delta^d\left[\sum_{n\delta\in\Lambda_\delta}\sum_{n'\in\mathcal{C}(n)}(\tfrac{1}{2})^d\phi_\delta(n'\delta)\right]f(n\delta), \tag{2.2a}$$

where $\mathcal{C}(n)$ is the cube of the unit volume centered at $(n+\frac{1}{2})=((n_1+\frac{1}{2}),\ldots,(n_d+\frac{1}{2}))$, or else

$$\phi_\delta^1(f)=\delta^d\sum_{n\delta\in\Lambda_\delta}\phi_\delta(n\delta)f(n\delta) \tag{2.2b}$$

for $f\in\mathcal{S}(\Lambda)$. The above two definitions of smeared fields differ only slightly and coincide at $\delta=0$. We have introduced the definition (2.2a) for convenience in the derivation of the transfer matrix (see Lemma 2.2). The partition function and the Schwinger functions are defined by

$$Z(\Lambda,\delta)=\langle\exp(-V_{\Lambda,\delta})\rangle,$$

$$S(\Lambda,\delta;f_1,\ldots,f_n)=\langle\prod_{i=1}^n\phi_\delta^*(f_i)\rangle_{\Lambda,\delta} \tag{2.3}$$

$$\equiv Z(\Lambda,\delta)^{-1}\langle\prod_{i=1}^n\phi_\delta^*(f_i)\exp(-V_{\Lambda,\delta})\rangle,$$

where $\phi_\delta^*(f)$ is either $\phi_\delta(f)$ or $\phi_\delta^1(f)$, respectively defined in (2.2a) and (2.2b).

Finally we define the pressures by

$$\alpha_{\Lambda,\delta}(P)=(1/|\Lambda|)\log Z(\Lambda,\delta),$$

$$\alpha_{\Lambda,\delta}(P-ax)=(1/|\Lambda|)\log\langle\exp[a\phi_\delta(\chi_\Lambda)]\exp(-V_{\Lambda,\delta})\rangle, \tag{2.4}$$

where χ_B, $B\subseteq\Lambda$, is the characteristic function of the set B. We note that the definition of $\alpha_{\Lambda,\delta}(P-ax)$ is identical for (2.2a) and (2.2b).

We now give the main result of this section.

Theorem 2.1: Let $\phi_\delta^*(f)$ be defined by either (2.2a) or (2.2b). Assume that the pressures $\alpha_{\Lambda,\delta}(P)$ and $\alpha_{\Lambda,\delta}(P-x)$ are bounded uniformly in Λ and δ. Then there exists a Schwartz space norm, $|\cdot|_S$, and constant K independent of Λ and δ such that

$$|S(\Lambda,\delta;f_1,\ldots,f_n)|\le K^n n!\prod_{i=1}^n|f_i|_S$$

for $f_i\in\mathcal{S}(\Lambda)$.

Remark 2.1: The above result implies the existence of boson field theories if the corresponding pressures are bounded uniformly in Λ and δ. As in Ref. 9 one could choose convergent subsequences (as $\delta\to 0$ and $|\Lambda|\to\infty$). The limit Schwinger functions are translation invariant. Since the physical positivity condition holds for lattice fields,[10] it only remains to verify the Euclidean invariance and clustering in order to establish the Wightman axioms.

We postpone the proof of Theorem 2.1 to the end of this section. We first establish a formula of the transfer matrix for lattice fields. One may derive the transfer matrix formula by using a method similar to that in Ref. 4 (also see Ref. 19), but we use a more elementary method to derive it. Since we only need the transfer matrix for one direction (say time direction), we consider the case of $d=2$. The formula for $d\ge 2$ will follow by the same method. Let $\phi_\delta(\chi_{\Delta^2(l_1,l_2)})$ be defined as in (2.2a). Following Refs. 5 and 6 and identifying $\phi_\delta(n\delta)=q_n$ we may write

$$\langle\exp(-V_{\Lambda,\delta})+a\phi_\delta(\chi_{\Delta^2(l_1,l_2)})\rangle\exp(E_{\Lambda,\delta})$$

$$=|2\pi C_\delta^\Lambda|^{-M/2}\int\exp\Big[-\tfrac{1}{2}\delta^2 z_{\Lambda,\delta}q\cdot(-\Delta_\delta^P+m^2)\cdot q$$

$$-\delta^2\sum(P(q_n)+R_{\Lambda,\delta}(q_n))+a\delta^2\sum_{n\delta\in\Delta^2(l_1,l_2)}$$

$$\times\Big(\tfrac{1}{4}\sum_{n'\in\mathcal{C}(n)}q_{n'}\Big)\Big]d^Mq \tag{2.5}$$

where $M=(\Lambda_\delta)^\#$, Δ_δ^P is an $M\times M$ matrix given by

$$\Delta_\delta^P(n,n')=\begin{cases}4\delta^{-2}, & |n-n'|=0,\\ -\delta^{-2}, & |n-n'|=1, \quad n\delta, n'\delta\in\Lambda_\delta\\ 0, & \text{otherwise},\end{cases}$$

and $C_\delta^\Lambda=[\delta^2(-\Delta_\delta^P+m^2)]^{-1}$. For details we refer the reader to Refs. 5 and 6. Let us introduce a transfer matrix between two hyperplanes ($n^{(2)}=0, 1$) separated by δ. Let $N^2=M=(\Lambda_\delta)^\#$. We define an operator $T_\delta(P-a\chi_{\Delta^1(l_1)})$ on $L^2(R^N)$ (depending on the interaction) by its kernel

$$[T_\delta(P-a\chi_{\Delta^1(l_1)})](x,y)=\exp A_\delta(x,y)$$

where $x,y\in R^N$ and

$$A_\delta(x,y)=\tfrac{1}{2}\delta z_{\Lambda,\delta}\Big[\sum_{i=1}^N\delta^{-2}x_iy_i+\tfrac{1}{2}\sum_{\substack{i=1\\N+1\equiv1}}^N\delta^{-2}(x_ix_{i+1}+y_iy_{i+1})\Big]$$

$$-\tfrac{1}{2}\delta z_{\Lambda,\delta}\Big\{\tfrac{1}{2}\sum_{i=1}^N[4\delta^{-2}(x_i^2+y_i^2)+m^2(x_i^2+y_i^2)]\Big\}$$

$$-\tfrac{1}{2}\delta\sum_{i=1}^N[P(x_i)+R_{\Lambda,\delta}(x_i)+P(y_i)+R_{\Lambda,\delta}(y_i)]$$

$$+\tfrac{1}{2}\delta a\sum_{\substack{i\delta\in\Delta^1(l_1)\\(i+1)\delta\in\Delta^1(l_1)}}[\tfrac{1}{2}(x_i+x_{i+1})+\tfrac{1}{2}(y_i+y_{i+1})].$$

One may easily check that (2.5) can be written as

$$[\mathrm{Tr}(T_\delta(0)^{l/\delta})]^{-1}\mathrm{Tr}[(T_\delta(P))^{(l-l_2)/\delta}(T_\delta(P-a\chi_{\Delta^1(l_1)}))^{l_2/\delta}]. \quad (2.6)$$

Here we have identified $|2\pi C_\delta^\Lambda|^{-M/2}=\mathrm{Tr}(T_\delta(0)^{l/\delta})$ by setting $V_{\Lambda,\delta}=0$ and $a=0$. The cases for $d>2$ can be handled by a straightforward modification of the above method. For any $d\in N^+$ we define

$$\begin{aligned} T_0 &= T_\delta(0)^{1/\delta},\\ T(P) &= T_\delta(P)^{1/\delta}, \qquad (2.7)\\ T(P-a\chi_{\Delta^{d-1}(l_1,\dots,l_{d-1})}) &= (T_\delta(P-a\chi_{\Delta^{d-1}(l_1,\dots,l_{d-1})}))^{1/\delta}. \end{aligned}$$

From (2.5) and (2.6) we obtain that

$$\langle\exp[a\phi_\delta(\chi_{\Delta^d(l_1,\dots,l_d)})]\exp(-V_{\Lambda,\delta})\rangle\exp(E_{\Lambda,\delta}) = \frac{\mathrm{Tr}[(T(P))^{l-l_d}(T(P-a\chi_{\Delta^{d-1}(l_1,\dots,l_{d-1})}))^{l_d}]}{\mathrm{Tr}(T_0^{l_d})}. \quad (2.8)$$

The above expression is the transfer matrix formula on the lattice fields.

Lemma 2.2: Let $\phi_\delta(\chi_{\Delta^d(1,\dots,1)})$ be defined as in (2.2a). For $\tau\in C$ and $|\tau|=1$,

$$\langle\exp[\tau\phi_\delta(\chi_{\Delta^d(1,\dots,1)})]\rangle_{\Lambda,\delta}\le\exp[\alpha_{\Lambda,\delta}(P-x)-\alpha_{\Lambda,\delta}(P)].$$

Proof: We use the relation (2.4) and the Schwartz inequality $[|\mathrm{Tr}(AB)|\le(\mathrm{Tr}(A^*A)\mathrm{Tr}(B^*B))^{1/2}]$ to obtain that for $a\in R$

$$\langle\exp[a\phi_\delta(\chi_{\Delta^d(1,\dots,1)})]\rangle_{\Lambda,\delta}=\frac{\mathrm{Tr}[(T(P))^{l-1}(T(P-a\chi_{\Delta^{d-1}(1,\dots,1)}))]}{\mathrm{Tr}(T(P)^l)} \le\left(\frac{\mathrm{Tr}[(T(P))^{l-2}(T(P-a\chi_{\Delta^{d-1}(1,\dots,1)}))^2]}{\mathrm{Tr}(T(P)^l)}\right)^{1/2}. \quad (2.9)$$

If we use the Schwartz inequality $(n-1)$ more times (note that $l=2^n$), we bound the above expression by

$$\begin{aligned} &\frac{\mathrm{Tr}[(T(P-a\chi_{\Delta^{d-1}(1,\dots,1)}))^l]}{\mathrm{Tr}(T(P)^l)}^{1/l}\\ &=(\langle\exp[a\phi_\delta(\chi_{\Delta^d(1,\dots,1,l)})]\rangle_{\Lambda,\delta})^{1/l}\\ &=(\langle\exp[a\phi_\delta(\chi_{\Delta^d(l,1,\dots,1)})\rangle_{\Lambda,\delta})^{1/l}. \end{aligned}$$

Here we have used the Nelson's symmetry to obtain the last equality. The above procedures iterated $d-1$ times bounds (2.9) by

$$(\langle\exp[a\phi_\delta(\chi_\Lambda)]\rangle_{\Lambda,\delta})^{1/|\Lambda|}=\exp[\alpha_{\Lambda,\delta}(P-ax)-\alpha_{\Lambda,\delta}(P)]. \quad (2.10)$$

We note that for $|\tau|=1$ [expand $\exp(|\mathrm{Re}\tau|\phi_\delta)$ and use the first Griffiths inequality]

$$|\langle\exp[\tau\phi_\delta(\chi_{\Delta^d(1,\dots,1)})]\rangle_{\Lambda,\delta}|\le\langle\exp[\phi_\delta(\chi_{\Delta^d(1,\dots,1)})]\rangle_{\Lambda,\delta}. \quad (2.11)$$

The proof of the lemma now follows from (2.10) and (2.11).

Remark 2.2: Replacing $\Delta^d(1,\dots,1)$ by $\Delta^d(2,\dots,2)$ and following the proof of Lemma 2.1, it is easy to check that for $|\tau|\le 1$,

$$\langle\exp[\tau\phi_\delta(\chi_{\Delta^d(2,\dots,2)})\rangle_{\Lambda,\delta}\le\exp[2^d(\alpha_{\Lambda,\delta}(P-x)-\alpha_{\Lambda,\delta}(P))].$$

We now prove Theorem 2.1.

Proof of Theorem 2.1: We first prove the theorem for the Schwinger functions defined by (2.2a) and (2.3). Let $\Delta^{(j)}\subset\Lambda$ be a unit cube centered at $j\in Z^d$. The translation invariance property and Lemma 2.2 yield

$$\langle\exp[\phi_\delta(\chi_\Delta(j))]\rangle_{\Lambda,\delta}\le\exp[\alpha_{\Lambda,\delta}(P-x)-\alpha_{\Lambda,\delta}(P)]. \quad (2.12)$$

We note that $\langle\exp[\tau\phi_\delta(\chi_\Delta(j))]\rangle_{\Lambda,\delta}$ is an entire function. By using the Cauchy integral theorem (see Ref. 2) we obtain

$$\begin{aligned}\langle\phi_\delta(\chi_\Delta(j))^m\rangle_{\Lambda,\delta}&\le(\mathrm{const})^m m!\,\exp[\alpha_{\Lambda,\delta}(P-x)-\alpha_{\Lambda,\delta}(P)]\\ &\le K^m m! \qquad (2.13)\end{aligned}$$

from the assumption in Theorem 2.1, and by Lemma 2.2 and (2.12), where K is a constant independent of Λ and δ.

We now prove the theorem for $f_i\ge 0$, $i=1,\dots,n$. For general f_i we only need to decompose each f_i into its positive and negative parts to obtain in the theorem for arbitrary f_i (with a K increase to $2K$). Hence we assume that all functions f_i in the following proof are positive.

Let f_i have support in some unit cube $\Delta^{(i)}\subset\Lambda$. Then, by the first Griffiths inequality,[5,13,14,20] it follows that

$$\begin{aligned}\langle\phi_\delta(f_i)^m\rangle_{\Lambda,\delta}&=\|f_i\|_\infty^m\langle(\phi_\delta(f_i|\,\|f_i\|_\infty))^m\rangle_{\Lambda,\delta}\\ &\le\|f_i\|_\infty^m\langle\phi_\delta(\chi_\Delta(i))^m\rangle_{\Lambda,\delta}\\ &\le K^m m!\,\|f_i\|_\infty^m \qquad (2.14)\end{aligned}$$

from (2.13). Let each f_i have support in some unit cube $\Delta^{(i)}$. (2.14) yields

$$\begin{aligned}\langle\prod_{i=1}^k\phi_\delta(f_i)\rangle_{\Lambda,\delta}&\le\prod_{i=1}^k(\langle\phi_\delta(f_i)^k\rangle_{\Lambda,\delta})^{1/k}\\ &\le K^k k!\prod_{i=1}^k\|f_i\|_\infty. \qquad (2.15)\end{aligned}$$

For $f_i\in\mathcal{S}(\Lambda)$ we write $f_i=\sum_{j_i}f_i^{(j_i)}$, where $f^{(j_i)}$ has support in the unit cube centered at $j_i\in Z^d$. Then, from (2.15) it follows that

$$\begin{aligned}\langle\prod_{i=1}^k\phi_\delta(f_i)\rangle_{\Lambda,\delta}&=\sum_{j_i\in\Lambda\cap Z^d}\langle\prod_{i=1}^k\phi_\delta(f_i^{(j_i)})\rangle_{\Lambda,\delta}\\ &\le\sum_{j_i}K^k k!\prod_{i=1}^k\|f_i^{(j_i)}\|_\infty\\ &\le K^k k!\prod_{i=1}^k\sum_{j_i}\|f_i^{(j_i)}\|_\infty)\\ &\le K^k k!\prod_{i=1}^k|f_i|_S\end{aligned}$$

for some Schwartz norm $|\cdot|_S$. This proves the theorem in the case of (2.2a).

We now prove the theorem when $\phi_\delta^*(f)$ is defined by (2.2b). Let $\phi_\delta(f)$ and $\phi_\delta^1(f)$ be the lattice fields defined by (2.2a) and (2.2b), respectively. The first Griffiths inequality yields

$$\begin{aligned}\langle\phi^1(\chi_\Delta(j))^m\rangle_{\Lambda,\delta}&\le\langle\phi(\chi_{\tilde\Delta}(j))^m\rangle_{\Lambda,\delta}\\ &\le(K')^m m!, \qquad (2.16)\end{aligned}$$

where $\tilde\Delta^{(j)}$ is the cube of volume 2^d centered at $j\in Z^d$. Here we have used Remark 2.2. Replacing (2.13) by

(2.16) and following the same procedure we prove the theorem for the Schwinger functions defined by (2.2b) and (2.3). This completes the proof.

3. APPLICATIONS TO THE $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ AND EXPONENTIAL TYPE INTERACTION MODELS

In this section we first prove uniform bounds of the space cutoff Schwinger functions of the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model with periodic boundary condition. We also prove uniform bounds of the lattice and space cutoff Schwinger functions of exponential type interactions in d-dimensional space—time. $d \in N^+$. The interacting action of the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model ($\lambda \geq 0, \sigma, \mu \in R$) has the form

$$V_{\Lambda,\delta} = \delta^3 \sum_{n\delta \in \Lambda_\delta} [:(\lambda\phi_\delta(n\delta)^4 - \sigma\phi_\delta(n\delta)^2 - \mu\phi_\delta(n\delta)):] + \tfrac{1}{2}\delta m_\delta^2 : \phi_\delta(n\delta)^2 :] + E_{2,\Lambda,\delta} + E_{3,\Lambda,\delta}, \qquad (3.1)$$

where $\frac{1}{2}\delta m_\delta^2 : \phi_\delta(n\delta)^2 :$ and $E_{2,\Lambda,\delta} + E_{2,\Lambda,\delta}$ are the mass and vacuum counter terms. For the details we refer the reader to Refs. 8, 9, 15—17. We note that the linear and quadratic terms do not introduce counter terms (except Wick counter terms). As an immediate consequence of the method used in the proof of the theorem and the results in Refs. 11 and 15 we have the following theorem.

Theorem 3.1: For the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ field model with periodic boundary conditions, the space cutoff Schwinger functions are bounded uniformly in Λ by

$$|S(\Lambda; f_1, \ldots, f_n)| \leq K^n \prod_{i=1}^n |f_i|_S$$

for a suitable Schwartz space norm $|\cdot|_S$, where K is a constant independent of Λ.

Proof: In Ref. 11 we established uniform bounds of the pressures $\alpha_\Lambda(P)$ and $\alpha_{\Lambda,\delta}(P)$ for the $\lambda\phi_3^4$ model. Since we have not given the detailed proof of uniform bounds of $\alpha_{\Lambda,\delta}$ in Ref. 11, we will only use here the uniform bounds of $\alpha_\Lambda(P)$. Since the $\alpha\phi^2 + \mu\phi$ term do not introduce counter terms (except Wick ordering) in the interacting action given in (3.1), a straightforward modi fication of the method used in Ref. 11 gives us uniform bounds of the pressures $\alpha_\Lambda(p)$ and $\alpha_\Lambda(P-x)$. In Ref. 15 we have proven the convergence of the lattice approximation for the $\lambda\phi_3^4$ model: That is

$$Z(\Lambda, \delta) \to Z(\Lambda),$$

$$S(\Lambda, \delta; f_1, \ldots, f_n) \to S(\Lambda; f_1, \ldots, f_n),$$

as $\delta \to 0$, where the lattice cutoff Schwinger functions have been defined through (2.2b) and (2.3). The same result probably holds for the Schwinger functions defined through (2.2a) and (2.3) by the method in Ref. 15. The convergence of the lattice approximation for the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model follows by a straightforward modification of the method used in Ref. 15. From (2.13) and (2.16) (the first inequality in each relation) we obtain that for $\mu \geq 0$

$$\langle \phi_\delta^1(\chi_\Delta(j))^m \rangle_{\Lambda,\delta} \leq (\text{const})^m m! \exp\{2^d[\alpha_{\Lambda,\delta}(P-x) - \alpha_{\Lambda,\delta}(P)]\}.$$

By taking the limit as $\delta \to 0$ and by using the convergence of the lattice approximation together with the uniform bounds of the pressures we conclude that

$$\langle \phi(\chi_\Delta(j))^m \rangle_\Lambda \leq K^m m!, \qquad (3.2)$$

where K is a constant independent of Λ. The bound in (3.2) and the method used in the second part of the proof of theorem 2.1 gives the proof for $\mu \geq 0$. The theorem for $\mu \leq 0$ follows from the transformation $\phi(x) \to -\phi(x)$.

Remark 3.1:

(a) The point of Theorem 3.1 is that it gives bounds in a form suitable for the Osterwalder—Schrader reconstruction theorem.[21] In (3.1) one may show that the mass counter term can be chosen independent of boundary conditions. As a result the Dirichlet Schwinger functions are bounded by the periodic Schwinger functions. Furthermore the Dirichlet Schwinger functions are monotone in the region. Hence, if one proves the convergence of the lattice approximation in the $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ model with Dirichlet boundary condition, our bounds then allow the passage of an infinite volume theory.[5,6]

(b) Recently Fröhlich[22] constructed the strongly coupled $(\lambda\phi^4 - \sigma\phi^2 - \mu\phi)_3$ field models with so-called weak coupling boundary conditions by using Seiler and Simon's result.[12] The same method can be applied to construct an infinite volume theory with the weak periodic coupling boundary condition. One may be able to show that the above two constructions are equivalent by proving equivalence of the weakly coupled $\lambda\phi_3^4$ field theories with free and periodic boundary conditions.

We next consider the exponential type interactions in d-dimensional space—time with lattice and space cutoff. The corresponding interacting action is given by

$$V_{\Lambda,\delta} = \delta^d \sum_{n\delta \in \Lambda_\delta} : \int dv(\alpha) \exp[\alpha\phi_\delta(n\delta)] :, \qquad (3.3)$$

where $dv(\alpha)$ is any (positive) finite measure.[6,23] We take the field strength renormalization constant $z_{\Lambda,\delta} = 1$ for any $d \in N^+$. It is known that[6,23]

$$V_{\Lambda,\delta} \geq 0, \qquad \langle V_{\Lambda,\delta} \rangle = (\int dv(\alpha)) |\Lambda|, \qquad (3.4)$$

for $\delta > 0$. From Theorem 2.1, we have the following result.

Theorem 3.2: Let $dv(\alpha)$ be any (positive) finite even measure on R. Then the Schwinger functions of the exponential type interactions given by (3.3) are bounded by

$$|S(\Lambda, \delta; f_1, \ldots, f_n)| \leq K^n n! \prod_{i=1}^n |f_i|_S,$$

uniformly in Λ and δ for $f_i \in \mathcal{S}(R^d)$, $d \in N^+$, where $|\cdot|_S$ is a suitable Schwartz space norm.

Proof: By Theorem 2.1, it is sufficient to show uniform bounds of the pressures $\alpha_{\Lambda,\delta}(P)$, $\alpha_{\Lambda,\delta}(P-x)$. We have that for $a \in R$

$$\langle \exp[a\phi_\delta(\chi_\Lambda) - V_{\Lambda,\delta}] \rangle \leq \langle \exp[a\phi_\delta(\chi_\Lambda)] \rangle \leq \exp[O(1)|\Lambda|], \qquad (3.5)$$

from (3.4). The Jensen's inequality and (3.4) yields

$$\langle \exp[a\phi_\delta(\chi_\Lambda) - V_{\Lambda,\delta}]\rangle \geq \exp[\langle a\phi_\delta(\chi_\Lambda) - V_{\Lambda,\delta}\rangle] \geq \exp[-(\int d\nu(\alpha))|\Lambda|], \quad (3.6)$$

for $a \in R$. From (3.5) and (3.6) it follows that

$$|\alpha_{\Lambda,\delta}(P)| + |\alpha_{\Lambda,\delta}(P-x)| \leq \text{const}$$

uniformly in Λ and δ. Theorem 3.2 follows from Theorem 2.1 and the above bounds.

Note added in manuscript: Recently the author has constructed the infinite volume limit Dirichlet states by proving the convergence of lattice approximation of the model with Dirichlet boundary condition and using Theorem 3.1. See Y. Park, "Convergence of Lattice Approximation and Infinite Volume Limit in the $(\lambda\phi^4 - \sigma\phi - \mu\phi)_3$ Field Theory," University of Bielefeld, Preprint (1975).

ACKNOWLEDGMENT

The author would like to thank Professor Francesco Guerra for information about the transfer matrix with periodic boundary condition and helpful suggestions. This paper would not have been written without him.

*On leave from the Department of Mathematics, Yonsei University, Seoul, Korea,

[1]J. Feldman and K. Osterwalder, "The Wightman Axioms and Mass Gap for Weakly Coupled $(\lambda\phi^4)_3$ Quantum Field Theories," Harvard preprint (1975).

[2]J. Fröhlich, Helv. Phys. Acta **47**, 265 (1974).

[3]J. Glimm, A. Jaffe, and T. Spencer, "The Particle Structure of the Weakly Coupled $P(\phi)_2$ Model and Other Applications of High Temperature Expansions" *Springer Lecture Notes in Physics* (Springer, New York, 1973).

[4]F. Guerra, "Exponential Bounds in Lattice Field Theories," preprint (1975).

[5]F. Guerra, R. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).

[6]B. Simon, *The $p(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, N.J., 1974).

[7]J. Fröhlich, "The Pure Phase, the Irreducible Quantum Fields and ...," Princeton preprint (1975).

[8]J. Glimm and A. Jaffe, Fortschr. Phys. **21**, 327 (1973).

[9]J. Glimm and A. Jaffe, Phys. Rev. Lett. **33**, 440 (1974).

[10]R. Schrader, "A Possible Constructive Approach to ϕ_4^4," preprint (1975).

[11]Y. Park, J. Math. Phys. **17**, 1073 (1976).

[12]E. Seiler and B. Simon, "Nelson's Symmetry and All That in the Ukawa$_2$ and $\lambda\phi_3^4$ Field Theories," Princeton preprint.

[13]F. Guerra, R. Rosen, and B. Simon, "Boundary Conditions in the $P(\phi)_2$ Euclidean Quantum Field Theory," preprint (1975).

[14]F. Guerra, R. Rosen, and B. Simon, Commun. Math. Phys. **41**, 19 (1975).

[15]Y. Park, J. Math. Phys. **16**, 2183 (1975).

[16]Y. Park, J. Math. Phys. **16**, 1065 (1975).

[17]J. Feldman, Commun. Math. Phys. **37**, 93 (1974).

[18]J. Magnen and R. Seneor, "Infinite Volume Limit of the $\lambda\phi_3^4$ Model," preprint (1975).

[19]R. Höegh-Krohn, "Relativistic Quantum Statistical Mechanics in two-dimensional space—time," preprint.

[20]R. Griffiths, J. Math. Phys. **8**, 418, 484 (1967).

[21]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973); **42**, 281 (1975).

[22]J. Fröhlich, "Existence and Analyticity in the Bare Parameter of the $(\lambda(\boldsymbol{\phi}\cdot\boldsymbol{\phi})^2 - \sigma\phi_1^2 - \mu\phi_1)_3$—Quantum Field Models, I.", Princeton preprint.

[23]S. Albeverio and R. Höegh-Krohn, J. Funct. Anal. **16**, 39 (1974).

Critique of the generalized cumulant expansion method

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 28 January 1976)

The method of ordered cumulants is presented for the solution of multiplicative stochastic processes. The relationship between this method and the irreducible cluster integral method of Mayer, which is used in the theory of the imperfect gas, is elucidated. The cluster property of ordered cumulants is proved. A critique of the literature in this area is presented which exposes some errors in the formulas. Several examples are indicated for the application of the ordered cumulant method.

1. INTRODUCTION

Many physical processes may be mathematically modelled by stochastic operator (differential, matrix, commutator, etc.) equations. Linear stochastic operator equations usually are of two major types, the inhomogeneous or "additive" stochastic processes with the canonical form

$$\frac{d}{dt}\mathbf{a}(t)=\mathbf{G}\mathbf{a}(t)+\tilde{\mathbf{F}}(t), \tag{1}$$

and the homogeneous or "multiplicative" stochastic processes with the canonical form

$$\frac{d}{dt}\mathbf{a}(t)=\mathbf{A}\mathbf{a}(t)+\tilde{\mathbf{A}}(t)\mathbf{a}(t). \tag{2}$$

The vector $\mathbf{a}(t)$ is the quantity for which each equation is solved, $\mathbf{G}$ and $\mathbf{A}$ are t-independent operators, and $\tilde{\mathbf{F}}(t)$ is a stochastic vector whereas $\tilde{\mathbf{A}}(t)$ is a stochastic operator. Additive stochastic processes have a long history and are now well understood, but multiplicative stochastic processes have only been studied relatively recently and their solution will be the subject of this paper.

Before commencing with the formulation of the solution of multiplicative stochastic processes, it is useful, for purposes of comparison, to briefly review the applications of additive stochastic processes. Their prototype is Langevin's equation for Brownian motion. [1] Uhlenbeck and Ornstein, [2] and Wang and Uhlenbeck[3] extended Lagevin's equation to the Brownian motion of a harmonic oscillator and other more complicated systems. Onsager and Machlup[4] used this extended version of Langevin's equation as the basis of their theory of irreversible thermodynamics. de Groot and Mazur[5] began the development of a completely general theory of Markovian stochastic processes which was completed by Fox and Uhlenbeck. [6] This theory extended applicability of additive stochastic processes to hydrodynamics in precisely the manner suggested by Landau and Lifshitz, [7] and it also provided a basis for the stochastic Boltzmann equation. The non-Markovian extension of the theory was subsequently established by Hauge and Martin-Löf. [8] Recently achieved applications of additive stochastic processes include problems in binary mixtures, [9] light scattering, [10] "long time tails" for autocorrelation functions, [8] and laser theory. [11] Quite recently, a nonlinear version of additive stochastic processes has also been presented by Keizer. [12]

Multiplicative stochastic processes, in the special case of differential equations with stochastic coefficients, have been studied for several decades by many people, especially by probability theorists. The emphasis in this paper, however, will be placed on the ordered cumulant method which was pioneered by Kubo. [13] A detailed account of spin relaxation theory based upon Kubo's work has been presented by Freed, [14] and Fox[15] has studied the stochastic Schrödinger equation and its associated density matrix equation using Kubo's methods. An almost independent development, stemming from the work on stochastic differential equations with stochastic coefficients, has been published by van Kampen. [16]

In the remainder of this paper a review of the ordered cumulant method of solution of multiplicative stochastic processes will be presented. The relationship between this method and the irreducible cluster integral method of Mayer, used in the theory of the imperfect gas, will be elucidated. The cluster property for the nth order, ordered cumulant will be proved. A critique of the present literature on this subject will be given which exposes some correctable errors. Finally, the paper will conclude will a representative selection of examples.

2. ORDERED CUMULANTS

In Eq. (2) it is assumed that the various nth order moments of $\tilde{\mathbf{A}}(t)$ are computable from an appropriately specified characteristic functional. In addition, it is assumed that $\langle\tilde{\mathbf{A}}(t)\rangle=0$, where $\langle\cdots\rangle$ denotes averaging.

Transformation to the "interaction picture" through the equation

$$\mathbf{b}(t)\equiv\exp(-t\mathbf{A})\mathbf{a}(t)$$

leads to the equation

$$\frac{d}{dt}\mathbf{b}(t)=\tilde{\mathbf{B}}(t)\mathbf{b}(t) \tag{3}$$

in which $\tilde{\mathbf{B}}(t)$ is given by

$$\tilde{\mathbf{B}}(t)\equiv\exp(-t\mathbf{A})\tilde{\mathbf{A}}(t)\exp(t\mathbf{A}).$$

It follows that $\langle\tilde{\mathbf{B}}(t)\rangle=0$.

The solution to (3) may be written in the form

$$\mathbf{b}(t)=\underset{\leftarrow}{T}\exp\left[\int_0^t\tilde{\mathbf{B}}(s)\,ds\right]\mathbf{b}(0) \tag{4}$$

in which the symbol $\underset{\leftarrow}{T}$ denotes t-ordering and is defined by

$$\underset{\leftarrow}{T}\left[\int_0^t \tilde{\mathbf{B}}(s)\,ds\right]^n = n!\int_0^t ds_1\int_0^{s_1} ds_2\cdots\int_0^{s_{n-1}} ds_n \times\tilde{\mathbf{B}}(s_1)\mathbf{B}(s_2)\cdots\tilde{\mathbf{B}}(s_n). \tag{5}$$

The average of (4) leads to the introduction of the ordered cumulants $\mathbf{G}^{(n)}$ through the equation

$$\begin{aligned}\langle \mathbf{b}(t)\rangle &= \langle \underset{\leftarrow}{T}\exp[\textstyle\int_0^t \mathbf{B}(s)\,ds]\rangle \mathbf{b}(0)\\ &= \underset{\leftarrow}{T}\exp\left(\sum_{n=1}^{\infty}\int_0^t \mathbf{G}^{(n)}(s)\,ds\right)\mathbf{b}(0).\end{aligned} \tag{6}$$

In (6) $\mathbf{b}(0)$ is specifically statistically independent of $\tilde{\mathbf{B}}(s)$ for all $s\geq 0$. The t-ordering symbol in the second equality is defined by

$$\begin{aligned}&\underset{\leftarrow}{T}\left\{\prod_{i=1}^{k}\left[\int_0^t \mathbf{G}^{(l_i)}(s_i)\,ds_i\right]\right\}\\ &= \int_0^t ds_1\int_0^{s_1} ds_2\cdots\int_0^{s_{k-1}} ds_k \sum_p \mathbf{G}^{(l_{p(1)})}(s_1)\mathbf{G}^{(l_{p(2)})}(s_2)\\ &\quad\times\cdots\mathbf{G}^{(l_{p(k)})}(s_k),\end{aligned} \tag{7}$$

in which the sum over p is a sum over all $k!$ permutations of the integers $1,2,\ldots,k$, and l_i is some positive integer. If $l_i=m$ for $i=1,2,\ldots,k$, then all $k!$ integrands in (7) are identical and the sum over permutations produces an over-all factor of $k!$, which is identical with the behavior of $\underset{\leftarrow}{T}$ expressed in (5). If an intermediate state of degeneracy exists such that m_l of the l_i's equal l and $\sum_l m_l=k$, then the sum over permutations will generate factors of $\Pi_l m_l!$ for each of the distinct integrands which will occur in (7). For example,

$$\begin{aligned}&\underset{\leftarrow}{T}\left\{\int_0^t \mathbf{G}^{(6)}(s)\,ds\left[\int_0^t \mathbf{G}^{(2)}(s)\,ds\right]^2\right\}\\ &=2!\left\{\int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\left[\mathbf{G}^{(6)}(s_1)\mathbf{G}^{(2)}(s_2)\mathbf{G}^{(2)}(s_3)\right.\right.\\ &\quad\left.\left.+\mathbf{G}^{(2)}(s_1)\mathbf{G}^{(6)}(s_2)\mathbf{G}^{(2)}(s_3)+\mathbf{G}^{(2)}(s_1)\mathbf{G}^{(2)}(s_2)\mathbf{G}^{(6)}(s_3)\right]\right\}.\end{aligned} \tag{8}$$

The mth order moment operator $\mathbf{A}^{(m)}$ is defined through the equation

$$\begin{aligned}\left\langle \underset{\leftarrow}{T}\exp\left(\int_0^t \tilde{\mathbf{B}}(s)\,ds\right)\right\rangle &= \sum_{m=0}^{\infty}\frac{1}{m!}\left\langle \underset{\leftarrow}{T}\left\{\left(\int_0^t \tilde{\mathbf{B}}(s)\,ds\right)^m\right\}\right\rangle\\ &=\sum_{m=0}^{\infty}\int_0^t \mathbf{A}^{(m)}(s)\,ds,\end{aligned} \tag{9}$$

in which $\int_0^t \mathbf{A}^{(0)}(s)\,ds=\mathbf{1}$ by convention. The moments and the ordered cummulants are related by the identity

$$\int_0^t \mathbf{A}^{(m)}(s)\,ds = \sum_{\substack{\text{partitions}\\ \text{of } m}}\underset{\leftarrow}{T}\left\{\prod_{l=1}^{\infty}\frac{1}{m_l!}\left(\int_0^t \mathbf{G}^{(l)}(s)\,ds\right)^{m_l}\right\}, \tag{10}$$

in which the sum over *partitions of* m is specified by $\sum_{l=1}^{\infty} l m_l=m$, and which is proved by the sequence of identities

$$\begin{aligned}\sum_{m=0}^{\infty}\int_0^t \mathbf{A}^{(m)}(s)\,ds &= \sum_{m=0}^{\infty}\sum_{\substack{\text{partitions}\\ \text{of } m}}\underset{\leftarrow}{T}\left\{\prod_{l=1}^{\infty}\frac{1}{m_l!}\left(\int_0^t \mathbf{G}^{(l)}(s)\,ds\right)^{m_l}\right\}\\ &=\sum_{m=0}^{\infty}\sum_{\substack{\text{compositions}\\ \text{of } m}}\underset{\leftarrow}{T}\left\{\prod_{l=1}^{\infty}\frac{1}{m_l!}\left(\int_0^t \mathbf{G}^{(l)}(s)\,ds\right)^{m_l}\right\}\\ &=\sum_{m=0}^{\infty}\frac{1}{m!}\underset{\leftarrow}{T}\left\{\left(\sum_{l=1}^{\infty}\int_0^t \mathbf{G}^{(l)}(s)\,ds\right)^m\right\}\\ &=\underset{\leftarrow}{T}\exp\left(\sum_{n=1}^{\infty}\int_0^t \mathbf{G}^{(n)}(s)\,ds\right).\end{aligned} \tag{11}$$

The sum over *compositions of* m in the second equality is defined by $\sum_{l=1}^{\infty} m_l=m$. The second equality is the crux of the proof, and follows from the fact that following $\sum_{l=1}^{\infty}$, a sum over partitions of m or a sum over compositions of m lead to equivalent series which differ only in the sequence of occurrence of their summands. The third equality is just the multinomial expansion which still holds in the presence of $\underset{\leftarrow}{T}$.

Using (7) in (10) gives expressions for the mth order moments in terms of ordered cumulants of order less than or equal to m. It is possible to invert these expressions and express the mth order, ordered cumulant in terms of moments of order less than or equal to m. The first three terms of each type of expression are listed below:

$$\begin{aligned}\int_0^t \mathbf{A}^{(1)}(s)\,ds &= \int_0^t \mathbf{G}^{(1)}(s)\,ds,\\ \int_0^t \mathbf{A}^{(2)}(s)\,ds &= \int_0^t \mathbf{G}^{(2)}(s)\,ds + \tfrac{1}{2}\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{G}^{(1)}(s)\,ds\right]^2\right\},\\ \int_0^t \mathbf{A}^{(3)}(s)\,ds &= \int_0^t \mathbf{G}^{(3)}(s)\,ds + \underset{\leftarrow}{T}\left\{\int_0^t ds_1\int_0^t ds_2\mathbf{G}^{(2)}(s_1)\right.\\ &\quad\left.\times\mathbf{G}^{(1)}(s_2)\right\} + (1/3!)\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{G}^{(1)}(s)\,ds\right]^3\right\}.\end{aligned} \tag{12}$$

$$\begin{aligned}\int_0^t \mathbf{G}^{(1)}(s)\,ds &= \int_0^t \mathbf{A}^{(1)}(s)\,ds,\\ \int_0^t \mathbf{G}^{(2)}(s)\,ds &= \int_0^t \mathbf{A}^{(2)}(s)\,ds - \tfrac{1}{2}\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{A}^{(1)}(s)\,ds\right]^2\right\},\\ \int_0^t \mathbf{G}^{(3)}(s)\,ds &= \int_0^t \mathbf{A}^{(3)}(s)\,ds - \underset{\leftarrow}{T}\left\{\int_0^t ds_1\int_0^t ds_2\mathbf{A}^{(2)}(s_1)\mathbf{A}^{(1)}(s_2)\right.\\ &\quad\left.+\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{A}^{(1)}(s)\,ds\right]\tfrac{1}{2}\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{A}^{(1)}(s)\,ds\right]^2\right\}\right\}\right.\\ &\quad\left.-(1/3!)\underset{\leftarrow}{T}\left\{\left[\int_0^t \mathbf{A}^{(1)}(s)\,ds\right]^3\right\}\right\}.\end{aligned} \tag{13}$$

In order to arrive at a general expression for $\int_0^t \mathbf{G}^{(n)}(s)\,ds$, it is necessary to introduce a shorthand notation. This is most easily achieved if all expression are written out with all integrals explicit. For example,

$$\begin{aligned}\int_0^t \mathbf{A}^{(3)}(s)\,ds &= \int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\langle\tilde{\mathbf{B}}(s_1)\tilde{\mathbf{B}}(s_2)\tilde{\mathbf{B}}(s_3)\rangle,\\ \underset{\leftarrow}{T}&\left\{\int_0^t ds_1\int_0^t ds_2\,\mathbf{A}^{(2)}(s_1)\mathbf{A}^{(1)}(s_2)\right\}\\ &=\int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\left\{\langle\tilde{\mathbf{B}}(s_1)\tilde{\mathbf{B}}(s_2)\rangle\langle\tilde{\mathbf{B}}(s_3)\rangle\right.\\ &\quad\left.+\langle\tilde{\mathbf{B}}(s_1)\tilde{\mathbf{B}}(s_3)\rangle\langle\tilde{\mathbf{B}}(s_2)\rangle+\langle\tilde{\mathbf{B}}(s_1)\rangle\langle\tilde{\mathbf{B}}(s_2)\tilde{\mathbf{B}}(s_3)\rangle\right\},\\ \underset{\leftarrow}{T}&\left\{\left(\int_0^t \mathbf{A}^{(1)}(s)\,ds\right)\tfrac{1}{2}\underset{\leftarrow}{T}\left\{\left(\int_0^t \mathbf{A}^{(1)}(s)\,ds\right)^2\right\}\right\}\\ &=\int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\left\{2\langle\tilde{\mathbf{B}}(s_1)\rangle\langle\tilde{\mathbf{B}}(s_2)\rangle\langle\tilde{\mathbf{B}}(s_3)\rangle\right.\\ &\quad\left.+\langle\tilde{\mathbf{B}}(s_1)\rangle\langle\tilde{\mathbf{B}}(s_3)\rangle\langle\tilde{\mathbf{B}}(s_2)\rangle\right\}.\end{aligned} \tag{14}$$

A natural abbreviation for the right-hand sides of these equations is

$$\begin{aligned}&\langle 123\rangle,\\ &\langle 1\rangle\langle 23\rangle+\langle 12\rangle\langle 3\rangle+\langle 13\rangle\langle 2\rangle,\\ &2\{\langle 1\rangle\langle 2\rangle\langle 3\rangle\}+\langle 1\rangle\langle 3\rangle\langle 2\rangle,\end{aligned} \tag{15}$$

respectively. Denoting $\int_0^t \mathbf{G}^{(n)}(s)\,ds$ by $\langle 12\cdots n\rangle_c$, where c stands for cumulant, the equations in (13) may be given in shorthand by[17]

$$\begin{aligned}\langle 1\rangle_c &= \langle 1\rangle,\\ \langle 12\rangle_c &= \langle 12\rangle-\langle 1\rangle\langle 2\rangle,\\ \langle 123\rangle_c &= \langle 123\rangle-\langle 1\rangle\langle 23\rangle-\langle 12\rangle\langle 3\rangle-\langle 13\rangle\langle 2\rangle\\ &\quad+\langle 1\rangle\langle 2\rangle\langle 3\rangle+\langle 1\rangle\langle 3\rangle\langle 2\rangle.\end{aligned} \tag{16}$$

This shorthand notation permits the writing out of

$\int_0^t \mathbf{G}^{(4)}(s)\,ds$, which would otherwise be much too unwieldy:

$$\begin{aligned}\langle 1234\rangle_c = &\langle 1234\rangle - \langle 1\rangle\langle 234\rangle - \langle 123\rangle\langle 4\rangle - \langle 124\rangle\langle 3\rangle \\ &- \langle 134\rangle\langle 2\rangle - \langle 12\rangle\langle 34\rangle - \langle 13\rangle\langle 24\rangle - \langle 14\rangle\langle 23\rangle \\ &+ \langle 1\rangle\langle 2\rangle\langle 34\rangle + \langle 1\rangle\langle 3\rangle\langle 24\rangle + \langle 1\rangle\langle 4\rangle\langle 23\rangle \\ &+ \langle 1\rangle\langle 23\rangle\langle 4\rangle + \langle 1\rangle\langle 24\rangle\langle 3\rangle + \langle 1\rangle\langle 34\rangle\langle 2\rangle \\ &+ \langle 12\rangle\langle 3\rangle\langle 4\rangle + \langle 13\rangle\langle 2\rangle\langle 4\rangle + \langle 14\rangle\langle 2\rangle\langle 3\rangle \\ &+ \langle 12\rangle\langle 4\rangle\langle 3\rangle + \langle 13\rangle\langle 4\rangle\langle 2\rangle + \langle 14\rangle\langle 3\rangle\langle 2\rangle \\ &- \langle 1\rangle\langle 2\rangle\langle 3\rangle\langle 4\rangle - \langle 1\rangle\langle 2\rangle\langle 4\rangle\langle 3\rangle - \langle 1\rangle\langle 3\rangle\langle 2\rangle\langle 4\rangle \\ &- \langle 1\rangle\langle 3\rangle\langle 4\rangle\langle 2\rangle - \langle 1\rangle\langle 4\rangle\langle 2\rangle\langle 3\rangle - \langle 1\rangle\langle 4\rangle\langle 3\rangle\langle 2\rangle. \end{aligned} \tag{17}$$

The pattern that is exhibited in (16) and (17) has the general form,[16,17] known as van Kampen's rules:

$$\begin{aligned}\int_0^t \mathbf{G}^{(n)}(s)\,ds &= \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \langle 12\cdots n\rangle_c \\ &= \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \\ &\quad\times \sum_{\substack{\text{ordered}\\ \text{partitions}\\ \text{of } n}} (-1)^{k-1} \sum_P \prod_{i=1}^k \langle\{l_{P(i)}\}\rangle, \end{aligned} \tag{18}$$

in which the sum over *ordered partitions of n* involves partitions of the first n positive integers into m_l groups of l integers each, such that $\sum_{l=1}^\infty l m_l = n$ and such that in each group the l integers increase from left to right. k is defined for each particular partition by $k = \sum_{l=1}^\infty m_l$, which is the total number of groups in the partition. $\langle\{l_i\}\rangle$ denotes the shorthand for the integrand of a moment with l_i factors. $\langle\{l_1\}\rangle$ always contains the integer 1 in its group of integers. P is a permutation of the $k-1$ integers $2, 3, \ldots, k$, so that $P(1) = 1$ always holds. Equations (16) and (17) may be seen to be special cases of (18).

3. THE CLUSTER PROPERTY

In the Mayer theory of the imperfect gas,[18–23] the partition function for N gas molecules is given by

$$Q_N = N! \sum_{\substack{\text{partitions}\\ \text{of } N}} \prod_{l=1}^N \frac{1}{m_l!} (V b_l)^{m_l}, \tag{19}$$

in which the *partitions of N* are specified by $\sum_{l=1}^\infty l m_l = N$, V is the volume and b_l is determined by the linked, or irreducible, cluster integrals. The grand canonical partition function, Q, is defined by

$$Q = \sum_{N=0}^\infty \frac{1}{N!} Z^N Q_N.$$

Using an argument patterned after (11) gives

$$\begin{aligned} Q &= \sum_{N=0}^\infty \sum_{\substack{\text{partitions}\\ \text{of } N}} \prod_{l=1}^\infty \frac{1}{m_l!} (V b_l Z^l)^{m_l} \\ &= \sum_{N=0}^\infty \sum_{\substack{\text{compositions}\\ \text{of } N}} \prod_{l=1}^\infty \frac{1}{m_l!} (V b_l Z^l)^{m_l} \\ &= \sum_{N=0} \frac{1}{N!} \left(\sum_{l=1}^\infty V b_l Z^l\right)^N \\ &= \exp\left(V \sum_{l=1}^\infty b_l Z^l\right). \end{aligned} \tag{20}$$

The sum over *partitions of N* was specified for (19), and the sum over *compositions of N* is defined by $\sum_{l=1}^\infty m_l = N$. The third equality follows from the multinomial expansion. The very strong similarity between (20) and (11) suggests that the ordered cumulants possess a cluster property on analogy with the b_l's which are their counterparts. In the remainder of this section, the cluster property for $\int_0^t \mathbf{G}^{(n)}(s)\,ds$ will be proved, and the proof will use Eq. (18).

In order for the cluster property for ordered cumulants to hold, it is necessary that the moments possess the factorization property, which may be expressed by

$$\begin{aligned}&\langle \tilde{\mathbf{B}}(t_1)\tilde{\mathbf{B}}(t_2)\cdots\tilde{\mathbf{B}}(t_j)\tilde{\mathbf{B}}(t_{j+1})\cdots\tilde{\mathbf{B}}(t_n)\rangle \\ &\quad= \langle \tilde{\mathbf{B}}(t_1)\cdots\tilde{\mathbf{B}}(t_j)\rangle\langle\tilde{\mathbf{B}}(t_{j+1})\cdots\tilde{\mathbf{B}}(t_n)\rangle \end{aligned} \tag{21}$$

whenever $|t_{j+1} - t_j| \gg \tau$ and the time variables are ordered by $t_1 > t_2 > \cdots > t_n$. τ is the correlation time for $\tilde{\mathbf{B}}$, and the factorization property is strictly an identity only asymptotically. When τ is sufficiently short, however, factorization for intervals, $|t_{j+1} - t_j|$, which are not asymptotically large, may be a very good approximation.

Referring to (18), suppose that two time variables, s_j and s_k, satisfy $|s_j - s_k| \gg \tau$ and $k > j$. Because the time integrals are ordered as is specified by the limits of integration, the integrand involves time domains including $|s_j - s_{j+1}| \gg \tau$, so that if k is not $j+1$, then $|s_j - s_k| \gg \tau$ implies that $|s_j - s_{j+1}| \gg \tau$ obtains for the integrand. Therefore, consideration of the special case, $|s_j - s_{j+1}| \gg \tau$, is in fact general enough to cover all cases. For $|s_j - s_{j+1}| \gg \tau$, two cases arise in the analysis of (18): Either s_j and s_{j+1} are in the same moment, or s_j and s_{j+1} are in distinct moments. In the first case, the moment in question has the abbreviated form $\langle q\cdots j\, j+1\cdots r\rangle$ because the integers in any moment in (18) are arranged so that they increase from left to right. The factorization property implies

$$\langle q\cdots j\, j+1\cdots r\rangle = \langle q\cdots j\rangle\langle j+1\cdots r\rangle. \tag{22}$$

If $\langle q\cdots j\, j+1\cdots r\rangle$ was originally a factor in a product containing k factors, then the product contains $k+1$ factors when the right-hand side of (22) is substituted for the left-hand side. These $k+1$ factors satisfy all the conditions required for them to be a term in $\langle 12\cdots n\rangle_c$, but with the sign $(-1)^k$ rather than the sign $(-1)^{k-1}$ which attends the term containing the factor $\langle q\cdots j\, j+1\cdots r\rangle$. Therefore, these two terms cancel identically when (22) is used, and the first case is completed. In the second case, s_j is in one moment, $\langle q\cdots j\, l\cdots r\rangle$, and s_{j+1} is in another moment, $\langle g\cdots i\, j+1\cdots h\rangle$. According to the rules attending (18), $s_i > s_j > s_{j+1} > s_l$. Therefore, $|s_j - s_{j+1}| \gg \tau$ implies the two factorizations

$$\begin{aligned}\langle q\cdots j\, l\cdots r\rangle &= \langle q\cdots j\rangle\langle l\cdots r\rangle, \\ \langle g\cdots i\, j+1\cdots h\rangle &= \langle g\cdots i\rangle\langle j+1\cdots h\rangle. \end{aligned} \tag{23}$$

As in the first case, either of these factorization produces factors which are identical with factors in terms in $\langle 12\cdots n\rangle_c$. However, in this case, four terms rather than two, must be simultaneously considered. Suppose that the original term containing both $\langle q\cdots j\, l\cdots r\rangle$ and $\langle g\cdots i\, j+1\cdots h\rangle$ and k factors. It also has the sign $(-1)^{k-1}$. $\langle 12\cdots n\rangle_c$ also contains a term with $k+1$ factors, all of which are identical with the factors in the original term except that $\langle q\cdots j\ l\cdots r\rangle$ is replaced by $\langle q\cdots j\rangle\langle l\cdots r\rangle$, and $\langle g\cdots i\ j+1\cdots h\rangle$ is unchanged. This term has the sign $(-1)^k$. Similarly, there is another term with sign $(-1)^k$ and $k+1$ factors, all of which are identical with the factors in the original term except that $\langle g\cdots i\ j+1\cdots h\rangle$ is replaced $\langle g\cdots i\rangle \times\langle j+1\cdots h\rangle$, and $\langle q\cdots j\ l\cdots r\rangle$ is unchanged. $\langle 12\cdots n\rangle_c$ also has a term with $k+2$ factors and sign $(-1)^{k+1}$ in which all the factors are identical with the factors in the original term except that both $\langle q\cdots j\ l\cdots r\rangle$ and $\langle g\cdots i\ j+1\cdots h\rangle$ are replaced by $\langle q\cdots j\rangle\langle l\cdots r\rangle$ and $\langle g\cdots i\rangle\langle j+1\cdots h\rangle$ respectively. Now, if the right-hand side of (23) is used in place of the left-hand side of (23) in the original term and in these three related other terms, then four identical terms are obtained with two of one sign and two of the opposite sign. Therefore, the four terms cancel identically, and case two is completed.

The consequence of this theorem is that $\langle 12\cdots n\rangle_c$, as defined by (18), vanishes unless all n s_i's are "clustered" together relative to the scale τ. Closer examination of the details suggests that there are terms in $\langle 12\cdots n\rangle_c$ as expressed in (18) which are comprised of sufficiently many "short overlaps" that they permit s_1 and s_n to be as far apart as roughly $\frac{1}{2}n\tau$. However, τ is usually defined to be a time long enough so that $\langle 12\rangle$ falls to only a few percent of its value when $s_1=s_2$. If τ is the time for which $\langle 12\rangle$ is only $\frac{1}{10}$ its equal time value, then terms permitting s_1 and s_n to be apart by as much as $\frac{1}{2}n\tau$ are also porportional to a factor of order $(1/10)^{n/2}$.

The cluster property for $\langle 12\cdots n\rangle_c$ can be used to prove that, for $t\gg\tau$, $\int_0^t \mathbf{G}^{(n)}(s)\,ds$ has the simple form

$$\lim_{t\gg\tau}\int_0^t \mathbf{G}^{(n)}(s)\,ds=\int_0^t \exp(-t'\mathbf{A})\mathbf{M}^{(n)}\exp(t'\mathbf{A})\,dt' \qquad (24)$$

in which $\mathbf{M}^{(n)}$ is t'-independent. The exponential-of-A factors stem from the interaction picture which is being used throughout this paper. The proof of (24) goes as follows: Consider the t derivative of $\int_0^t \mathbf{G}^{(n)}(s)\,ds$ which is $\mathbf{G}^{(n)}(t)$, and consider the t derivative of the right-hand side of (18), which converts s_1 into t. The cluster property of $\langle 12\cdots n\rangle_c$ with $s_1\equiv t$ requires that all time variables be close to t relative to τ. The leading factor in every term in the t derivative of the right-hand side of (18) has the form $\langle\tilde{\mathbf{B}}(t)\cdots\rangle$. Using the definition of $\tilde{\mathbf{B}}(t)$ shows that $\langle\tilde{\mathbf{B}}(t)\cdots\rangle=\exp(-t\mathbf{A})\langle\tilde{\mathbf{A}}(t)\exp(t\mathbf{A})\cdots\rangle$. Additionally, every term also ends with a factor which is farthest to the right of the form $\langle\cdots\tilde{\mathbf{B}}(s_j)\rangle$, where $j=2,3\cdots n$ but $j\neq 1$. This factor may be written as $\langle\cdots\exp(-s_j\mathbf{A})\tilde{\mathbf{A}}(s_j)\rangle\exp(s_j\mathbf{A})$ and $\exp(s_j\mathbf{A})$ may be approximated by $\exp(t\mathbf{A})$ because of the cluster property if $\tau\mathbf{A}$ is sufficiently small. After removal of the leading exponential, $\exp(-t\mathbf{A})$, and the tailing exponential, $\exp(t\mathbf{A})$, the residual terms integrate to a t-independent expression when $t\gg\tau$. This expression will be called $\mathbf{M}^{(n)}$.

4. THE SOLUTION

It has now been demonstrated that the solution to (2), when averaged, is given by

$$\langle\mathbf{b}(t)\rangle=\underset{\leftarrow}{T}\exp\left(\sum_{n=1}^{\infty}\int_0^t \mathbf{G}^{(n)}(s)\,ds\right)\mathbf{b}(0)$$

and that

$$\mathbf{G}^{(n)}(t)\underset{t\gg\tau}{\rightarrow}\exp(-t\mathbf{A})\mathbf{M}^{(n)}\exp(t\mathbf{A}).$$

$\langle\mathbf{b}(t)\rangle$ is also equal to $\exp(-t\mathbf{A})\langle\mathbf{a}(t)\rangle$. Therefore, the above solution is equivalent to the solution to

$$\frac{d}{dt}\langle\mathbf{a}(t)\rangle=\mathbf{A}\langle\mathbf{a}(t)+\exp(t\mathbf{A})\sum_{n=1}^{\infty}\mathbf{G}^{(n)}(t)\exp(-t\mathbf{A})\langle\mathbf{a}(t)\rangle, \qquad (25)$$

which for $t\gg\tau$ is very well approximated by

$$\frac{d}{dt}\langle\mathbf{a}(t)\rangle=\mathbf{A}\langle\mathbf{a}(t)\rangle+\sum_{n=1}^{\infty}\mathbf{M}^{(n)}\langle\mathbf{a}(t)\rangle, \qquad (26)$$

which has a simple exponential solution. Because $\langle\tilde{\mathbf{B}}(t)\rangle=0$, it follows that $\mathbf{M}^{(1)}=0$, as well as $\mathbf{G}^{(1)}(t)=0$. $\mathbf{M}^{(2)}$, which is often a good approximation to $\sum_{n=1}^{\infty}\mathbf{M}^{(n)}$ by itself, is given explicitly by

$$\mathbf{M}^{(2)}=\int_0^t\langle\tilde{\mathbf{A}}(t)\exp((t-s)\mathbf{A})\mathbf{A}(s)\rangle\exp((s-t)\mathbf{A})\,ds, \qquad (27)$$

where $t\gg\tau$.

Two special cases are worth mentioning at this point. If $\tilde{\mathbf{A}}(t)$ is determined by a Gaussian characteristic functional which is consequently an even functional, then only the even order, ordered cumulants are nonvanishing. However, unlike the fully commutative situation in which the cumulants for a Gaussian process vanish after the second cumulant, the higher order, ordered cumulants do not vanish.[15] They do, of course, satisfy the cluster property as may be observed directly by writing out (18) for the Gaussian case explicitly. The second special case is the case in which the autocorrelation function for $\tilde{\mathbf{A}}(t)$ is porportional to a delta function in the time variables. In this case, the higher order, ordered cumulants all vanish after $n=2$, because of the cluster property.

5. COMMENTARY

It may be asked why van Kampen's rules[16] for the construction of the ordered cumulants were necessary when it is the case that Kubo's basic paper[13] on the subject appeared nearly twelve years earlier. Indeed, Eq. (6.9) of Kubo's paper[13] purports to be a closed form formula for ordered cumulants. The formula is an extention to ordered operators of a formula for commuting quantities which was worked out by Meeron.[24] In the commutative case Meeron's formula is unquestionably correct. In Kubo's formula there is an index j which may be restricted to just one value for the context of this paper, and the ordering operation, Q, in his formula corresponds with $\underset{\leftarrow}{T}$ here. Rewriting Kubo's general formula for just one value of j and with $\underset{\leftarrow}{T}$ replacing Q, and with a few changes of variables, gives the formula

$$\int_0^t \mathbf{G}^{(n)}(s)\,ds=\sum_{\substack{\text{partitions}\\ \text{of } n}}(-1)^{p-1}(p-1)!\times\underset{\leftarrow}{T}\left\{\prod_{l=1}^{\infty}\frac{1}{m_l!}\left(\int_0^t \mathbf{A}^{(l)}(s)\,ds\right)^{m_l}\right\}, \qquad (28)$$

in which the sum over partitions is specified by $\sum_{l=1}^{\infty} l m_l = n$ and $p = \sum m_l$, and which more recently appeared in a paper by Fox.[25] A combinational procedure for the derivation of (28) was presented, and Fox did not then know that the formula, in a somewhat more general form, had already appeared in the literature.[13] In fact the equivalent of (28) had also been used by Freed[14] for $n=1,2,3,4$ in his Eqs. (2.14a)–(2.14d). Are Eqs. (28) and (18) equivalent? The answer is no.

The nature of the error is somewhat subtle because only a slight, but very significant, difference exists between formulas (18) and (28). It is easiest to see the difference by example. If the $n=3$ case in equation (13) is compared with the $n=3$ case in (28), then two expressions are obtained for $\int_0^t \mathbf{G}^{(3)}(s)\,ds$ which contain some identical terms, and some nonidentical terms. Ignoring the identical terms, from (13) one gets

$$\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)\tfrac{1}{2}\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)^2\}\} - (1/3!)\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)^3\} \tag{29}$$

whereas from (28) one gets instead

$$2\underleftarrow{T}\{(1/3!)(\int_0^t \mathbf{A}^{(1)}(s)\,ds)^3\}. \tag{30}$$

By using the shorthand notation of Sec. 2 of this paper and working out all the integrals, (29) and (30) become (31) and (32) respectively

$$\langle 1\rangle\langle 2\rangle\langle 3\rangle + \langle 1\rangle\langle 3\rangle\langle 2\rangle, \tag{31}$$

$$2\{\langle 1\rangle\langle 2\rangle\langle 3\rangle\}. \tag{32}$$

Generally, the factor $(p-1)!$ in (28) corresponds with $(p-1)!$ permuted terms in (18). van Kampen[16] even suggests using an expression like (28) as a mnemonic device for remembering (18), provided one replaces the $(p-1)!$ by $(p-1)!$ permutations of the $p-1$ moments following the moment which contains 1. In Fox's paper[25] the error arises from the incorrect identity

$$\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)\tfrac{1}{2}\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)^2\}\} = (3/3!)\underleftarrow{T}\{(\int_0^t \mathbf{A}^{(1)}(s)\,ds)^3\}, \tag{33}$$

which would render (29) and (30) equal. An identity such as (33) is discussed by Fox,[25] and is in fact valid in the appropriate context, but that context requires a very different meaning for $\mathbf{A}^{(1)}$ than its meaning here. While the discrepancy is exhibited here for $n=3$, it does not show up until $n=6$ if $\langle\tilde{\mathbf{A}}(t)\rangle = 0$. Since both Freed[14] and Fox[25] assumed $\langle\tilde{\mathbf{A}}(t)\rangle = 0$ in their subsequent calculations, and both only looked explicitly at $n=2$ and $n=4$ results, they got the correct results.

The cluster property of the ordered cumulants, which is of such fundamental importance as far as the utility of the ordered cumulant method is concerned, also has a somewhat confused literature. van Kampen[16] barely mentions it. Freed[14] makes very interesting use of it, but bases its validity upon general results in the Kubo[13] paper. In the Kubo paper, the important equation in this regard is (6.7). It is not difficult to show that for $\underleftarrow{T}$ ordering, counter examples to Eq. (6.7) can be constructed, and that even though the cluster property is nevertheless true, Kubo's argument does not justify it. The argument in the present paper is offered to remedy this situation. In a very clear and concise paper, Terweil[26] demonstrates the connection between the ordered cumulant method and the Zwanzig projection operator technique. In that paper, he proves a cluster property for the "partial kernels" which arise in the projection operator approach. His proof should be compared with the proof here. His proof does not, however, establish directly the cluster property for the ordered cumulants.

Yoon, Deutch, and Freed[27] have also compared the ordered cumulant method with the projection operator method, and their considerations are deducible from Terweil's paper, although their context is very different, and very interesting.

6. EXAMPLES

In addition to the examples discussed by Kubo,[13] van Kampen,[16] and Freed,[14] a few other representative examples will be sketched here. The examples will be treated to the extent that they are made to have the form of (2).

Example 1: The stochastic, standing wave equation[28]:

$$\frac{d^2}{dx^2}v(x) + k^2[1+\tilde{\phi}(x)]v(x) = 0.$$

Replace $v(x)$ by

$$v(x) = A(x)\exp(ikx) + B(x)\exp(-ikx),$$

in which $A(x)$ and $B(x)$ are subject to the auxillary condition

$$\frac{dA}{dx}\exp(ikx) + \frac{dB}{dx}\exp(-ikx) = 0.$$

It then follows that

$$\frac{d}{dx}\begin{pmatrix}A(x)\\B(x)\end{pmatrix} = i\frac{k}{2}\tilde{\phi}(x)\begin{pmatrix}1 & \exp(-2ikx)\\ -\exp(2ikx) & -1\end{pmatrix}\begin{pmatrix}A(x)\\B(x)\end{pmatrix}.$$

This is the form of (2) if x is changed to t, $A=0$, and $a(t)$ and $\tilde{A}(t)$ are complex valued.

Example 2: The stochastic Schrödinger equation[29]:

$$i\hbar\frac{\partial}{\partial t}\psi = (\mathbf{H}^\circ + \tilde{\mathbf{H}}(t))\psi$$

Expand ψ in terms of the eigenstates of $\mathbf{H}^\circ$, $\mathbf{H}^\circ|\alpha\rangle = E_\alpha|\alpha\rangle$:

$$\psi(t) = \sum_\alpha C_\alpha(t)\,|\alpha\rangle.$$

The Schrödinger equation becomes

$$i\hbar\frac{d}{dt}C_\alpha(t) = E_\alpha C_\alpha(t) + \tilde{H}_{\alpha\alpha'}(t)C_{\alpha'}(t),$$

in which $\tilde{H}_{\alpha\alpha'}(t) = \langle\alpha|\tilde{\mathbf{H}}(t)|\alpha'\rangle$. This is also a complex realization of (2) with a diagonal **A**.

Example 3: The stochastic density matrix equation[29]:

By using the previous example and defining the density matrix by

$$\rho_{\alpha\beta}(t) = C_\beta^*(t)C_\alpha(t)$$

the density matrix equation is

$$i\hbar\frac{d}{dt}\rho_{\alpha\beta} = (E_\alpha - E_\beta)\rho_{\alpha\beta} + \tilde{L}_{\alpha\beta\alpha'\beta'}(t)\rho_{\alpha'\beta'},$$

where $\tilde{L}_{\alpha\beta\alpha'\beta'}(t) \equiv \delta_{\beta\beta'}\tilde{H}_{\alpha\alpha'}(t) - \delta_{\alpha\alpha'}\tilde{H}_{\beta'\beta}(t)$. This can also be written equivalently

$$i\hbar\frac{\partial}{\partial t}\rho = [\mathrm{H}^\circ, \rho] + [\tilde{\mathrm{H}}(t), \rho].$$

Example 4: Equation of motion for a magnetic moment in a stochastic magnetic field:

$$\frac{d}{dt}\mathbf{m} = -\frac{e}{2mc}[\mathbf{B} + \tilde{\mathbf{B}}(t)] \times \mathbf{m}.$$

Example 5: Reduced density matrix equation:

$$i\hbar\frac{\partial}{\partial t}\rho = [\mathrm{H}^s + \mathrm{H}^R + \mathrm{H}^I, \rho],$$

where H^s is a system Hamiltonian, H^R is a reservoir Hamiltonian, and H^I is the interaction Hamiltonian. No explicit stochasticity appears. However, reduction of the full density matrix by tracing over the reservoir states can be treated as averaging, and associated ordered cumulants can then be defined in a natural way. Another paper doing this in detail will be forthcoming.

Generally, the effect of $\mathrm{M}^{(2)}$, which appears in (26), is to create dissipative behavior for the averaged equations in these examples. Each example is in fact time reversal invariant before averaging is performed, so that averaging is the sole source of dissipation.

ACKNOWLEDGMENTS

This work was supported partially by NSF Grant MPS 73-08945 A01 and partially by the Alfred P. Sloan foundation. The author is grateful to G. W. Ford for providing a copy of his lecture notes. [17]

[1]P. Langevin, C.R. Acad. Sci. Paris 146, 530 (1908).
[2]G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. 36, 823 (1930).
[3]M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945).
[4]L. Onsager and S. Machlup, Phys. Rev. 91, 1512 (1953).
[5]S. R. de Groot and P. Mazur, *Non-equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[6]R. F. Fox and G. E. Uhlenbeck, Phys. Fluids 13, 1893, 2881 (1970).
[7]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison-Wesley, Reading, Mass., 1959), Chap. 17.
[8]E. H. Hauge and A. Martin-Löf, J. Stat. Phys. 7, 259 (1973).
[9]C. Cohen, Ph.D. thesis (Princeton University, 1971).
[10]J. Foch, Phys. Fluids 14, 893 (1971); 15, 224 (1972).
[11]M. Sargent, M. O. Scully, and W. E. Lamb, *Laser Physics* (Addison-Wesley, Reading, Mass., 1974), Chaps. 19 and 20.
[12]J. E. Keizer, J. Chem. Phys. 63, 398, 5037 (1975).
[13]R. Kubo, J. Phys. Soc. Jpn. 17, 1100 (1962).
[14]J. H. Freed, J. Chem. Phys. 49, 376 (1968).
[15]R. F. Fox, J. Math. Phys. 15, 1479 (1974).
[16]N. G. van Kampen, Physica 74, 215, 239 (1974).
[17]G. W. Ford, lecture notes, "Notes on Stochastic Differential Equations", July 1974, unpublished.
[18]J. E. Mayer, J. Chem. Phys. 5, 67 (1937).
[19]J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940), Chap. 13.
[20]B. Kahn and G. E. Uhlenbeck, Physica 5, 399 (1938).
[21]D. ter Haar, *Elements of Statistical Mechanics* (Rinehart, New York, 1954), Chap. 8.
[22]T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956), Chap. 5.
[23]G. E. Uhlenbeck and G. W. Ford, *Statistical Mechanics* (American Mathematical Society, Providence, R.I., 1963), Chap. 2.
[24]E. Meeron, J. Chem. Phys. 27, 1238 (1957), Appendix.
[25]R. F. Fox, J. Math. Phys. 16, 289 (1975), Eq. (52).
[26]R. H. Terweil, Physica 74, 248 (1974).
[27]B. Yoon, J. M. Deutch, and J. H. Freed, J. Chem. Phys. 62, 4687 (1975).
[28]G. C. Papanicolaou, SIAM J. Appl. Math. 21, 13 (1971).
[29]R. F. Fox, J. Math. Phys. 13, 1196 (1972).

Exponential decay and regularity properties of the Hartree approximation to the bound state wavefunctions of the helium atom

V. Benci

Istituto Matematica Applicata U. Dini, Università di Pisa, Pisa, Italy

D. Fortunato

Istituto Matematica Applicata, Facoltà di Ingegneria, Università di Bari, Bari, Italy

F. Zirilli

Istituto Matematico G. Castelnuovo, Università di Roma, Roma, Italy

(Received 17 December 1975; revised manuscript received 3 February 1976)

The exponential decay and regularity properties of the Hartree approximation to the bound state wavefunctions of the helium atom are proved.

1. INTRODUCTION

In quantum chemistry and theoretical physics, the Hartree equations are often used to simplify the study of the N-body Schrödinger Hamiltonian.

The Hamiltonian of the helium atom is

$$H=\sum_{i=1}^{2}-\frac{\hbar^2}{2m}\Delta_i-\sum_{i=1}^{2}\frac{2e}{r_i}+\frac{e^2}{r_{12}}, \qquad (1.1)$$

where the first term on the right-hand side of (1.1) is the sum of the kinetic energy of the two electrons and Δ_i is the Laplacian in the variables of the ith electron. The second term is the Coulomb potential energy of the electric field of the nucleus, and r_i is the distance between the nucleus and the ith electron. Finally, the third terms is the electric repulsive interaction between the two electrons and r_{12} is the distance between the two electrons, and $-e<0$ is the electric charge of the electron.

The restricted Hartree equation[1] has the following form for the helium atom:

$$-\frac{\hbar^2}{2m}\Delta u(x)-\frac{e}{|x|}u(x)+e^2u(x)\int\frac{u^2(y)}{|x-y|}\,dy=\lambda u(x) \qquad (1.2)$$

with $u\in L^2(\mathbb{R}^3)$ and $\|u\|_{L^2}=1$.

Chemists and physicists have done for a long time an enormous amount of heuristic and computer work on Eq. (1.2); only in recent years the progress of nonlinear functional analysis has made possible the discussion of nonlinear eigenvalues problems such as (1.2) on a rigorous mathematical basis.

On this last direction work has been done by Reeken,[2] Wolkowisky,[3] Reeken,[5] Stuart,[6] Lieb and Simon,[7] Micheletti and Zirilli.[8] Using bifurcation theory, topological degree techniques, and variational methods, many questions about the existence and the number of solutions of the nonlinear eigenvalue problem (1.2) have been solved.

The purpose of this paper is to prove for the helium atom restricted Hartree approximation the exponential decay and regularity properties of the bound-state wave functions solutions of the Schrödinger equation for the helium Hamiltonian (1.1).

Lieb and Simon[7] have announced the same exponential decay properties for the solutions of the Hartree—Fock problem.

Earlier not rigorous results of this type are contained in Handy— Marron—Silvestrone.[10]

The exponential decay of the eigenfunction of a Schrödinger Hamiltonian H has been considered by several authors: Hunziker,[11] Simon,[12] Ahlrichs,[13] Simon,[14] O'Connor,[15] and the regularities properties have been considered by Kato.[9]

In the following, we will make use of the results of Reeken,[4] O'Connor,[15] and Kato.[9]

2. THE RESULTS

Reeken in Refs. 4 and 5 has shown that there are at least countably many pairs (u_n,λ_n) with $u_n\in L^2(\mathbb{R}^3)$, $\|u_n\|_{L^2}=1$ and $\lambda_n<0$ such that Eq. (1.2) is verified.

Let on $L^2(\mathbb{R}^3)$

$$-(\hbar^2/2m)\Delta+V$$

be a two-body Schrödinger Hamiltonian. We recall the following theorems:

Theorem 1 (O'Connor[15]): If $V\in R+L^\infty_\epsilon$ and $[-(\hbar^2/2m)\Delta+V]\,\psi=-E\psi$, $E>0$, $\psi\in L^2(\mathbb{R}^3)$ then

$$\psi\in \mathcal{D}[\exp\theta(2mE)^{1/2}|x|]$$

for every θ, $0\leq\theta<1$.

Theorem 2 (Kato[9]): If $V\in L^2+L^\infty_\epsilon$ and $[-\hbar^2/Lm)\Delta+V]\,\psi=-E\psi$, $E>0$, $\psi\in L^2(\mathbb{R}^3)$, then ψ is Holder continuous, where R is the Rollnik class of functions. Equation (1.2) can be rewritten in the following way:

$$\left(-\frac{\hbar^2}{2m}\Delta-\frac{4e}{|x|}+q_u(x)\right)u=\lambda u, \qquad (2.1)$$

where

$$q_u(x)=2e^2\frac{u^2(y)}{|x-y|}\,dy \qquad (2.2)$$

so that in order to apply Kato and O'Connor theorems to the solutions $\{u_i\}$ of (2.3) given by Reeken's theorem is enough to show that $q_u(x) \in L^2 + L^\infty_\epsilon$ because $L^2 + L^\infty_\epsilon \subset R + L^\infty_\epsilon$ and $-2e/|x| \in L^2 + L^\infty_\epsilon$.

Theorem 3: Let $q_u(x)$ be given by (2.2) then if $u \in L^2(\mathbb{R}^3)$ then $q_u(x) \in L^2 + L^\infty_\epsilon$.

Proof: Let $\alpha_\rho(x)$ and $\beta_\rho(x)$ be defined as follows:

$$\alpha_\rho(x) = \begin{cases} 0 & \text{if } |x| < \rho \\ 2e^2/|x| & \text{if } |x| \geq \rho \end{cases}$$

and

$$\beta_\rho(x) = \begin{cases} 2e^2/|x| & \text{if } |x| < \rho \\ 0 & \text{if } |x| \geq \rho. \end{cases}$$

We have

$$q_u(x) = \int u^2(y)\,\alpha_\rho(x-y)\,dy + \int u^2(y)\,\beta_\rho(x-y)\,dy.$$

For any given $\epsilon > 0$ we take $\rho > \epsilon^{-1}\|u\|_{L^2}$, so that we have

$$\left\| \int u^2(y)\,\alpha_\rho(x-y)\,dy \right\|_{L^\infty} \leq \|u\|_{L^2}/\rho < \epsilon.$$

Moreover,

$$\int u^2(y)\,\beta_\rho(x-y)\,dy = (u^2 * \beta_\rho)(x) \in L^2(\mathbb{R}^3). \qquad (2.3)$$

In fact, (2.3) is the convolution product of an L^2 function with an L^1 function.

Concluding, we have the following result:

Theorem 4: Let $(u_i, -\lambda_i)$, $\lambda_i > 0$ be solutions of (1.2) then

$$u_i \in \mathcal{D}\,\exp[\theta(2m\lambda_i)\,|x|]$$

for every θ, $0 \leq \theta < 1$, moreover u_i is Holder continuous.

[1]D.R. Hartree, *The Calculation of Atomic Structures* (Wiley, New York, 1957).

[2]M. Reeken, J. Math. Phys. **11**, 2505 (1970).

[3]J.H. Wolkowisky, Indiana Math. J. **22**, 551 (1972).

[4]M. Reeken, in *Eigenvalues of Nonlinear Problems*, edited by G. Prodi (Centro Internazionale Matematico Estivo, C.I.M.E. III, Ciclo, 1974).

[5]M. Reekon, "Existence of Solutions for the Hartree and Hartree Fock Equations" (to appear in the Proc. R. Soc. of Edinburgh).

[6]D.A. Stuart, Archive Rat. Mech. An. **51**, 60 (1973).

[7]E. Lieb and B. Simon, J. Chem. Phys. **61**, 735 (1974).

[8]A.M. Micheletti and F. Zirilli (to be published).

[9]T. Kato, Trans. A.M.A. **10**, 95 (1951).

[10]N.C. Handy, M.T. Marron, and H.J. Silvestrone, Phys. Rev. **180**, 45 (1969).

[11]W. Hunziker, J. Math. Phys. **7**, 300 (1966).

[12]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton University Press, Princeton, 1971).

[13]R. Alhrichs, Preprint 1972 Institut für Physikalische Chemie der Universität Karlsruhe (unpublished).

[14]B. Simon, IHES preprint, Dec. 1972, Pointwise bounds on eigenfunctions and wave packets in N body quantum system.

[15]A.J. O'Connor, Commun. Math. Phys. **32**, 319 (1973).

Iterative solution of the Hartree equations

C. D. Luning and W. L. Perry

Department of Mathematics, Texas A&M University, College Station, Texas 77843
(Received 9 June 1975)

An iterative scheme based on eigenpairs of Hilbert–Schmidt operators obtained from a Green's function representation for solutions of a linearization of the Hartree equations, $d^2y_i(r)/dr^2 - [l_i(l_i+1)/r^2]y_i(r) + (2z/r)y_i(r) - (2/r)\Sigma_{j=1, j\neq i}^{N} Y_j(r)y_i(r) = \lambda_i^2 y_i(r)$, $Y_i(r) = \int_0^r y_i^2(s)ds + r\int_0^\infty s^{-1}y_i^2(s)\,ds$, $y_i(0)=0$, $y_i(\infty)=0$, $\int_0^\infty y_i^2(s)ds = 1$, $i=1,2,\ldots,N$, establishes existence of solutions to the Hartree equations and the sequence of eigenpairs generated converge subsequentially to a solution. In the case of the helium atom, for which we show some computational results, sequential convergence is obtained. Due to the Hilbert–Schmidt nature of the operators involved, the iterative method is implementable with Galerkin methods.

1. INTRODUCTION

In this paper, we demonstrate an iterative method for solving the Hartree equations. For an atom with N electrons, the Hartree equations may be written as

$$\frac{d^2y_i(r)}{dr^2} - \frac{l_i(l_i+1)}{r^2}\,y_i(r) + \frac{2z}{r}\,y_i(r) - \frac{2}{r}\sum_{\substack{j=1\\ j\neq i}}^{N} Y_j(r)y_i(r) = \lambda_i^2 y_i(r),$$

$$Y_i(r) = \int_0^r y_i^2(s)\,ds + r\int_0^\infty s^{-1}y_i^2(s)\,ds, \qquad (1.1)$$

$$y_i(0)=0,\ \ y_i(\infty)=0,\ \ \int_0^\infty y_i^2(s)\,ds = 1,$$

$$i=1,2,\ldots,N,$$

where $y_i(r) = p_{n_i,l_i}(r)$ is the radial wavefunction of the ith electron, which has quantum numbers (n_i, l_i, m_i), and where the number $z > N-1$ is the atomic number of the atom.

The Hartree equations only approximately describe the structure of an atom; however, they are used quite extensively for carrying out atomic calculations.

Hartree's method of solution is to make an approximation for the functions Y_i. This then reduces the problem to a system of linear equations. By using series techniques, two families of approximate solutions to the linear equations are found; one family satisfies the boundary conditions $y_i(0)=0$ and the other $y_i(\infty)=0$. By adjusting the parameters λ_i, it is then possible to piece these functions together smoothly and to satisfy the normalizing conditions $\int_0^\infty y_i^2(t)\,dt = 1$. A new approximation to the functions Y_i can now be made and the process can be repeated until a self-consistent field criterion is met. This technique has yielded many physically acceptable results. For more details see Hartree.[1] Slater[2] has a bibliography of calculations of this type.

Although Hartree's computational method gives physically acceptable results, it has never been proven that his method will converge to a solution. In fact, we know of no proven convergent iterative method of solution. However, the iterative method exhibited in this paper generates a sequence of successive approximations

$$\{\lambda_{1,i}, y_{1,i}; \lambda_{2,i}, y_{2,i}; \circ\circ\circ; \lambda_{k,i}, y_{k,i}; \circ\circ\circ\},$$

$i=1,2,\ldots,N$, which has a subsequence that converges to a solution of the Hartree equations.

For the case of the ground state of the Helium atom, where the uniqueness of the solution has been proved,[3] the subsequence condition can be removed, and it can be shown that the sequence of successive approximations does converge to a solution.

If uniqueness of solution were established for the general N electron case, the subsequence condition could then be removed. Physical observations and previous Hartree calculations tend to support a uniqueness conjecture, but mathematically the important question of uniqueness is still open.

In Sec. 3, we carry out some computations for the helium atom, $N=2$. Our technique is implementable by use of Galerkin methods.

Existence of solutions to the Hartree equations, first proven by Wolkowisky[4] in 1972, is also shown by our technique. Reeken[3] and Stuart[5] showed the existence of solutions for the helium atom case. None of these solutions, however, are constructive in nature. By using bifurcation techniques, other work on Hartree's equations has been done by Stuart[5,6] and Gustafson and Sather.[7,8]

2. THE ITERATIVE TECHNIQUE

The technique depends upon the construction of certain Hilbert–Schmidt integral operators which in turn depend upon Green's functions of differential operators arising from (1.1), namely the operators

$$L(y) = y'' - [l(l+1)/r^2]y - \lambda^2 y, \quad y(0) = y(\infty) = 0. \qquad (2.1)$$

By a change of variable, $L(y)=0$ is reduced to a standard form of Whittaker's differential equation.[9] Using the independent solutions of this equation, one finds that the solution to $L(y) = -f$, for appropriate f, is

$$y(r) = [-1/2\lambda(l!)^2][u_0(r)\int_r^\infty u_1(t)f(t)\,dt + u_1(r)\int_0^r u_0(t)f(t)\,dt], \qquad (2.2)$$

where

$$u_0(r) = \exp(-\lambda r)(2\lambda r)^{l+1}\int_0^1 \exp(2\lambda r t)t^l(1-t)^l\,dt,$$
$$u_1(r) = \exp(-\lambda r)(2\lambda r)^{l+1}\int_0^\infty \exp(-2\lambda r t)t^l(1+t)^l\,dt. \qquad (2.3)$$

With this Green's function representation of the solution to $L(y) = -f$ we can proceed to the development of the Hilbert–Schmidt integral operators necessary for the iteration scheme.

We let $\mathcal{G}$ be the vector space of real valued functions,

$$\mathcal{G}=\{f: \int_0^\infty x^{-1}f^2(x)\,dx<\infty\},$$

where the integral (as well as all integrals in the rest of the paper) is the Lebesgue integral. Let $\mathcal{H}$ denote the Hilbert space consisting of $\mathcal{G}$ with the inner product

$$(f,g)=\int_0^\infty x^{-1}f(x)g(x)\,dx.$$

For each continuous function ϕ defined on $[0,\infty)$, with $0\le\phi(x)\le N-1$, $\mathcal{H}_\phi$ is the Hilbert space consisting of $\mathcal{G}$ with the inner product

$$(f,g)_\phi=\int_0^\infty 2x^{-1}(z-\phi(x))f(x)g(x)\,dx.$$

We define linear operators U, U_ϕ, and $R_{\lambda,l}$ by the rules

$$U(f)=-r^{-1}f,\quad U_\phi(f)=2r^{-1}(\phi(r)-z)f$$

$$R_{\lambda,l}(f)=[-1/2(l!)^2][u_0(r)\int_r^\infty u_1(t)f(t)\,dt + u_1(r)\int_0^r u_0(t)f(t)\,dt],$$

where u_1 and u_0 are as in (2.3), $l\ge 0$, $\lambda>0$.

We now present some preliminary lemmas.

Lemma 1: The composition $R_{\lambda,l}U$ has a natural extension to a nonnegative self-adjoint Hilbert–Schmidt operator on $\mathcal{H}$.

Proof: Clearly, $R_{\lambda,l}U$ is a symmetric operator on $\mathcal{H}$. We now show a complete orthogonal set of eigenfunctions $\{f_n\}$ and corresponding eigenvalues μ_n of $R_{\lambda,l}U$ such that $\sum_{n=0}^\infty(\mu_n)^2<\infty$.

For each integer $n\ge 0$, there is a polynomial $p_n(r)$ of degree n such that

$$f_n(r)\equiv r^{l+1}p_n(r)\exp(-\lambda r)$$

is a solution of

$$y''-[l(l+1)/r^2]y-\lambda^2y+[2(n+l+1)/r]\lambda y=0.$$

Using the representation in (2.2) for the solution of $L(y)=-f$, one obtains

$$(R_{\lambda,l}U)(f_n)=[1/2(n+l+1)]f_n,\quad n\ge 0.$$

$\operatorname{span}\{f_n\}_{n=0}^\infty$ is dense in $\mathcal{H}$, for if

$$0=(r^{l+1}p_n(r)\exp(-\lambda r),\,g(r))$$
$$=\int_0^\infty p_n(r)\exp[-(\lambda/2)r](r^lg(r)\exp[-(\lambda/2)r])\,dr,$$

then, since the Laguerre functions are complete in $\mathcal{L}^2(0,\infty)$,

$$r^lg(r)\exp[-(\lambda/2)r]=0.$$

That is, $g(r)=0$. The orthogonality is a consequence of $R_{\lambda,l}U$ being symmetric. //

Lemma 2: The composition $R_{\lambda,l}U_\phi$ has a natural extension to a nonnegative self-adjoint Hilbert–Schmidt operator on $\mathcal{H}_\phi$.

Proof: It is easy to see that $R_{\lambda,l}U_\phi$ is symmetric on $\mathcal{H}_\phi$. Let $\{f_n\}_0^\infty$ be any complete orthonormal sequence in $\mathcal{H}_\phi$. Then $\{(2z-2\phi)^{1/2}f_n\}_0^\infty$ is a complete orthonormal sequence in $\mathcal{H}$. Moreover, since $R_{\lambda,l}U$ is a Hilbert–Schmidt operator in $\mathcal{H}$ and, since multiplication by $(2z-2\phi)^{1/2}$ is a continuous operator on $\mathcal{H}$, the operator

$$T\equiv(2z-2\phi)^{1/2}\circ R_{\lambda,l}U\circ(2z-2\phi)^{1/2}$$

is Hilbert–Schmidt on $\mathcal{H}$. Hence,

$$\sum_{n=0}^\infty\|(R_{\lambda,l}U_\phi)(f_n)\|_\phi^2=\sum_{n=0}^\infty\|T((2z-2\phi)^{1/2}f_n)\|^2<\infty,$$

and

$$(R_{\lambda,l}U_\phi f,f)=(R_{\lambda,l}U(2z-2\phi)f,(2z-2\phi)f)\ge 0. \ //$$

Since $R_{\lambda,l}U_\phi$ is a nonnegative self-adjoint Hilbert–Schmidt operator in $\mathcal{H}_\phi$, it has a complete orthogonal system of eigenfunctions $\{f_n\}_{n=0}^\infty$ with corresponding eigenvalues $\nu_n>0$ such that $\sum_{n=0}^\infty\nu_n^2<\infty$.

Lemma 3: If $f_n\in\mathcal{G}$ is an eigenfunction of $R_{\lambda,l}U_\phi$, then

(a) f_n, f_n', f_n'' are continuous,

(b) $\lim_{x\to 0}f_n(x)=0$,

(c) $\lim_{x\to\infty}f_n(x)=0$,

(d) $\lim_{x\to\infty}xf_n(x)=0$,

(e) $\lim_{x\to\infty}f_n'(x)=0$,

(f) $f_n\in\mathcal{L}^2(0,\infty)$,

(g) the corresponding eigenvalue ν_n is simple.

Proof: Part (a) is a consequence of the integral nature of the definition of $R_{\lambda,l}U_\phi$. Using formulas (2.3) and the equation

$$u_1(x)=\exp(-\lambda x)(2\lambda x)^{-l}\sum_{j=0}^{l}\binom{l}{j}(l+j)!\,(2\lambda x)^{l-j},$$

we have

$$u_0(x)\le(2\lambda x)^{l+1}\exp(\lambda x)\quad\text{and}\quad\int_x^\infty t^{-1}u_1^2(t)\,dt\le c(2\lambda x)^{-2l-1}.$$

For $x\ge 1$

$$u_1(x)\le c\exp(-\lambda x),\ u_0(x)\le c\exp(\lambda x),$$

$$|u_1'(x)|\le c\exp(-\lambda x),\quad |u_0'(x)|\le c\exp(\lambda x).$$

If

$$\nu_nf_n(x)=-[1/(l!)^2][u_0(x)\int_x^\infty u_1(t)t^{-1}(z-\phi(t))f_n(t)\,dt + u_1(x)\int_0^x u_0(t)t^{-1}(z-\phi(t))f_n(t)\,dt],$$

then

$$|\nu_nf_n(x)|\le K\{(2\lambda x)^{l+1}\exp(\lambda x)[\int_x^\infty t^{-1}u_1^2(t)\,dt]^{1/2}\times[\int_0^\infty t^{-1}f_n^2(t)\,dt]^{1/2}+u_1(x)[\int_0^x\exp(2\lambda t)(2\lambda t)^{2l+1}\,dt]^{1/2}\times[\int_0^\infty t^{-1}f_n^2(t)\,dt]^{1/2}\}$$
$$\le K_1[(2\lambda x)^{1/2}\exp(\lambda x)+\exp(\lambda x)u_1(x)(2\lambda x)^{l+1/2}].$$

However, $x^lu_1(x)$ is bounded as $x\to 0$; thus $\lim_{x\to 0}f_n(x)=0$. For $x\ge 1$

$$|\nu_nf_n(x)|\le K\{\exp(\lambda x)[\int_x^\infty t^{-1}\exp(-2\lambda t)\,dt]^{1/2}[\int_0^\infty t^{-1}f_n^2(t)\,dt]^{1/2} + \exp(-\lambda x)\int_0^1 u_0(t)t^{-1}|f_n(t)|\,dt + \exp(-\lambda x)[\int_1^x t^{-1}\exp(2\lambda t)\,dt]^{1/2}[\int_0^\infty t^{-1}f_n^2(t)\,dt]^{1/2}\}$$
$$\le K_1\{x^{-1/2}+\exp(-\lambda x)+\exp(-\lambda x)[\int_1^x t^{-1}\exp(2\lambda t)\,dt]^{1/2}\}.$$

Thus $\lim_{x\to\infty} f_n(x)=0$

$$|\nu_n x f_n(x)| \leq K\left(\frac{x\int_x^\infty \exp(-\lambda t)t^{-1}|f_n(t)|\,dt}{\exp(-\lambda x)}+\frac{x\int_0^x \exp(\lambda t)t^{-1}|f_n(t)|dt}{\exp(\lambda x)}\right).$$

Thus

$$\lim_{x\to\infty}|xf_n(x)|=K\lim_{x\to\infty}[2|f_n(x)|/\lambda]=0,$$

$$\nu_n f_n'(x)=-(l!)^{-2}[u_0'(x)\int_x^\infty u_1(t)t^{-1}(z-\phi(t))f_n(t)\,dt + u_1'(x)\int_0^x u_0(t)t^{-1}(z-\phi(t))f_n(t)\,dt].$$

The proof that $\lim_{x\to\infty} f_n'(x)=0$ proceeds just as in $\lim_{x\to\infty} f_n(x)=0$ using the appropriate estimates on $|u_0'(x)|$ and $|u_1'(x)|$. Of course, (f) follows immediately from (b) and (d), and (g) is proved exactly as in Coddington and Levinson.[10]

Lemma 4: For each continuous function ϕ, $0\leq\phi(x)\leq N-1$, and for each nonnegative integer n, there is an eigenfunction $f_n\in\mathcal{G}$ and at least one value of λ_n, $(z-N+1)/(n+l+1)\leq\lambda_n\leq z/(n+l+1)$ such that

$$R_{\lambda_n,l}U_\phi f_n=\lambda_n f_n.$$

Proof: Let $\mathcal{N}_n$ denote the class of all n-dimensional subspaces of $\mathcal{H}_\phi$. Then by the min–max properties of eigenvalues we have

$$\nu_n=\min_{\mathcal{L}\in\mathcal{N}_n}\max_{\substack{(f,g)_\phi=0\\ g\in\mathcal{L}}}\left(\frac{\|R_{\lambda,l}U_\phi f\|_\phi}{\|f\|_\phi}\right)$$

$$\leq\min_{\mathcal{L}\in\mathcal{N}_n}\max_{\substack{(f,g)_\phi=0\\ g\in\mathcal{L}}}\left((2z)^{1/2}\frac{\|R_{\lambda,l}U(2z-2\phi)f\|}{\|(2z-2\phi)f\|}\times\frac{\|(2z-2\phi)f\|}{\|f\|_\phi}\right)$$

$$\leq\min_{\mathcal{L}\in\mathcal{N}_n}\max_{\substack{(h,g)=0\\ g\in\mathcal{L}}}\left(2z\frac{\|R_{\lambda,l}Uh\|}{\|h\|}\right)=\frac{z}{n+l+1}.$$

Similarly since $z-\phi(x)\geq z-N+1$, we obtain $\nu_n\geq(z-N+1)/(n+l+1)$.

Since the eigenfunction f_n corresponding to the eigenvalue ν_n is a solution of

$$y''-[l(l+1)/r^2+\lambda^2]y+(\lambda/\nu_n)\{[2z-2\phi(r)]/r\}y=0$$

it follows from Courant and Hilbert[11] that ν_n depends continuously on λ. Thus, since $(z-N+1)/(n+l+1)\leq\nu_n\leq z/(n+l+1)$, there is a

$\lambda_n\in[(z-N+1)/(n+l+1),\ z/(n+l+1)]$ such that $\lambda_n=\nu_n(\lambda_n)$. //

Using the results of these lemmas, we can now present the iterative scheme for finding a solution to the Hartree equations and prove the subsequential convergence.

For $i=1,\ldots,N$ let $y_{0,i}\in\mathcal{G}$ be such that $\int_0^\infty y_{0,i}^2(r)\,dr=1$, and for $k=1,2,\cdots$ let $(y_{k,1},\lambda_{k,1};\cdots;y_{k,N},\lambda_{k,N})$ denote the solution set of the system

$$y_i''-[l_i(l_i+1)r^2]y_i+\{2[z-\phi_{k,i}(r)]/r\}y_i-\lambda_{k,i}^2y_i=0,$$

$$y_i(0)=0,\quad y_i(\infty)=0,\quad \int_0^\infty y_i^2(r)\,dr=1,$$

$$(z-N+1)/(n_i+l_i+1)\leq\lambda_{k,i}\leq z/(n_i+l_i+1),$$

where

$$\phi_{k,i}(r)=\sum_{\substack{j=1\\ j\neq 1}}^{N}[\int_0^r y_{k-1,j}(s)^2\,ds+\int_r^\infty s^{-1}y_{k-1,j}(s)^2\,ds].$$

The iterative sequence so defined behaves as follows:

Theorem: There exists a subsequence of the sequence

$$\{y_{1,i},\lambda_{1,i};y_{2,i},\lambda_{2,i};\cdots;y_{k,i},\lambda_{k,i};\cdots\},$$

$i=1,2,\cdots,N$, defined above that converges to a solution of the Hartree equations (1.1).

In order to prove the theorem, we need a lemma.

Lemma 5: For each $i=1,\ldots,N$, the sequence $\{\int_0^\infty y_{k,i}'(r)^2\,dr\}_{k=1}^\infty$ is bounded and the sequence $\{y_{k,i}(r)\}_{k=1}^\infty$ and $\{ry_{k,i}(r)\}_{k=1}^\infty$ are uniformly bounded on $[0,\infty)$.

Proof:

$$\lambda_{k,i}y_{k,i}(r)=-(l_i!)^{-2}[u_0(r)\int_r^\infty u_1(t)t^{-1}(z-\phi_{k,i}(t))y_{k,i}(t)\,dt + u_1(r)\int_0^r u_0(t)t^{-1}(z-\phi_{k,i}(t))y_{k,i}(t)\,dt].$$

Thus, by the Schwarz inequality and since $\int_0^\infty y_{k,i}(t)^2\,dt=1$,

$$|\lambda_{k,i}y_{k,i}(r)|\leq z(l_i!)^{-2}\{u_0(t)[\int_r^\infty t^{-2}u_1^2(t)\,dt]^{1/2} + u_1(r)[\int_0^r t^{-2}u_0^2(t)\,dt]^{1/2}\}.$$

However, $u_0(r)\leq(2\lambda r)^{l_i+1}\exp(\lambda r)$ and

$$u_1(r)=(2\lambda r)^{-l_i}\exp(-\lambda r)\sum_{j=0}^{l_i}\binom{l_i}{j}(l_i+j)!\,(2\lambda r)^{l_i-j}.$$

Also $\lim_{t\to 0}t^{-1}u_0(t)$ exists and is finite and for $r\geq 1$ there is a scalar c such that $u_0(r)\leq c\exp(\lambda r)$, $u_1(r)\leq c\exp(-\lambda r)$. Thus for $0\leq r\leq 1$

$$u_0(r)[\int_r^\infty t^{-2}u_1^2(t)\,dt]^{1/2} \leq\exp(\lambda r)\{\int_r^\infty\exp(-2\lambda t)[\sum_{j=0}^{l_i}\binom{l_i}{j}(l_i+j)!\,(2\lambda t)^{l_i-j}]^2\,dt\}^{1/2}$$

$$\leq e^\lambda\{\int_0^\infty\exp(-2\lambda t)[\sum_{j=0}^{l_i}\binom{l_i}{j}(l_i+j)!\,(2\lambda t)^{l_i-j}]^2\,dt\}^{1/2}=M_1$$

and

$$u_1(r)[\int_0^r t^{-2}u_0^2(t)\,dt]^{1/2} \leq(2\lambda)^2(2\lambda r)^{l_i}u_1(r)[\int_0^r\exp(2\lambda t)\,dt]^{1/2}\leq M_2,$$

$$|\lambda_{k,i}ry_{k,i}(r)|\leq|\lambda_{k,i}y_{k,i}(r)|\leq z(l_i!)^{-2}(M_1+M_2).$$

For $r\geq 1$

$$|\lambda_{k,i}ry_{k,i}(r)| \leq(l_i!)^{-2}zr\{u_0(r)[\int_r^\infty t^{-2}u_1^2(t)\,dt]^{1/2} + u_1(r)[\int_0^1 t^{-2}u_0^2(t)\,dt+\int_1^r t^{-2}u_0^2(t)\,dt]^{1/2}\}$$

$$\leq (l_i!)^{-2} z \left\{ c^2 \exp(\lambda r)[\int_r^\infty \exp(-2\lambda t)\,dt]^{1/2} + c[r^2 \exp(-2\lambda r)\int_0^1 t^{-2}u_0^2(t)\,dt + cr^2 \exp(-2\lambda r)\int_1^r t^{-2}\exp(2\lambda t)\,dt]^{1/2}\right\}.$$

Since the right-hand side is bounded as $r \to \infty$, there exists M_3 such that

$$|\lambda_{k,i} y_{k,i}(r)| \leq |\lambda_{k,i} r y_{k,i}(r)| \leq M_3 z (l_i!)^{-2}.$$

Thus, since $(z-N+1)/(n_i+l_i+1) \leq \lambda_{k,i} \leq z/(n_i+l_i+1)$, $\{y_{k,i}(r)\}_{k=1}^\infty$ and $\{ry_{k,i}(r)\}_{k=1}^\infty$ are uniformly bounded.

By a similar argument, using the fact that $|u_0'(r)| \leq c_1 \exp(-\lambda r)$ and $|u_1'(r)| \leq c_1 \exp(\lambda r)$ for $r \geq 1$, we obtain scalars M_4 and M_5, independent of k such that

$$|y_{k,i}'(r)| \leq \begin{cases} M_4 & \text{if } 0 \leq r \leq 1, \\ r^{-1}M_5 & \text{if } 1 \leq r; \end{cases}$$

thus

$$\int_0^\infty y_{k,i}'(r)^2\,dr = \int_0^1 y_{k,i}'(r)^2\,dr + \int_1^\infty y_{k,i}'(r)^2\,dr \leq M_4^2 + M_5^2. \ //$$

Proof of Theorem: For each $i=1,\ldots,N$, $\{\lambda_{k,i}\}_{k=1}^\infty$ is bounded and thus there is a λ_i, $(z-N+1)/(n_i+l_i+1) \leq \lambda_i \leq z/(n_i+l_i+1)$ and a subsequence $\{\lambda_{\hat{k},i}\}$ such that $\lambda_{\hat{k},i} \to \lambda_i$.

$$|y_{\hat{k},i}(r_2) - y_{\hat{k},i}(r_1)| = |\int_{r_1}^{r_2} y_{\hat{k},i}'(r)\,dr| \leq |r_2 - r_1| [\int_0^\infty y_{\hat{k},i}'(r)^2\,dr]^{1/2}.$$

By Lemma 5, $\{y_{\hat{k},i}(r)\}$ is equicontinuous and uniformly bounded, and, by Ascoli's lemma, there is a function y_i and a subsequence of $\{y_{\hat{k},i}\}$ which converges to y_i uniformly on every compact set. Since $|y_{k,i}(r)| \leq M_i r^{-1}$ for $r \geq 1$, it follows $|y_i(r)| \leq M_i r^{-1}$ for $r \geq 1$ and $y_i \in \mathcal{L}^2(0,\infty)$.

$y_{\hat{k},i} \to y_i$ $i=1,\ldots,N$, uniformly on bounded sets implies

$$\phi_{\hat{k},i}(r) \to \phi(r) \equiv \sum_{\substack{j=1\\ j\neq i}}^{N} \int_0^r y_j^2(t)\,dt + r\int_r^\infty t^{-1}y_j^2(t)\,dt$$

uniformly on $[0,\infty)$. Thus

$$y_{\hat{k},i}''(r) = \{r^{-2}l_i(l_i+1) - 2r^{-1}[z - \phi_{k,i}(r)] + \lambda_{\hat{k},i}^2\} y_{\hat{k},i}(r)$$

converges uniformly on all sets $[1/M, M]$ and hence it must converge to y_i''. Therefore, $\{y_i, \lambda_i\}_{i=1}^N$ is a solution set of the Hartree equations (1.1). //

From the proof, it is seen that if the Hartree equations have a unique solution, then the entire iterative sequence would converge to the solution. Indeed for each subsequence of $\{y_{k,i}, \lambda_{k,i}\}_{k=1}^\infty$ we have a convergent subsequence, so that the original sequence $\{y_{k,i}, \lambda_{k,i}\}_{k=1}^\infty$ would converge to the solution (y_i, λ_i).

3. CALCULATIONS FOR THE HELIUM ATOM CASE

We applied the technique to the ground state of helium ($N=2$, $z=2$, $l=0$, $n=1$). The calculations were performed with single precision on an IBM 360/65.

The iteration scheme was implemented by approximately calculating the largest eigenvalue and corresponding eigenfunction of $R_{\lambda,l}U_\phi$ in $\mathcal{H}_\phi$ by Galerkin's method. We used an averaging technique to find the λ such that the largest eigenvalue of $R_{\lambda,l}U_\phi$ was equal to λ.

We note that in the proof of the theorem, the eigenvalues and eigenfunctions of $R_{\lambda,l}U_\phi$ are utilized, whereas here we only obtain approximations to these. The iteration scheme still yields a valid approximation to the solution of the Hartree equations, since the necessary eigenvalues and eigenfunctions depend continuously on λ and ϕ.[11]

Galerkin's method says that if $\{f_k\}$ is a complete orthonormal system in $\mathcal{H}_\phi$ and if P_n is the orthogonal projection in $\mathcal{H}_\phi$ onto $\mathcal{H}_n = \mathrm{span}\{f_1,\ldots,f_n\}$, then, as $n \to \infty$, the largest eigenvalue and corresponding eigenfunction of $P_n R_{\lambda,l}U_\phi$ restricted to $\mathcal{H}_n$ converges to the largest eigenvalue and corresponding eigenfunction of $R_{\lambda,l}U_\phi$.[12]

The orthonormalization procedure in $\mathcal{H}_\phi$ and formulas for $R_{\lambda,l}U_\phi(f)$ are fairly easy to compute when applied to functions of the form $g_k = x^k \exp(-\alpha x)$, k a positive integer, and so we applied Galerkin's method with $\mathcal{H}_n = \mathrm{span}\{g_1, g_2, \ldots, g_n\}$.

For $\mathcal{H}_1$, we found $\lambda = 1.412$ and for $\mathcal{H}_2$, we found $\lambda = 1.381$, already within .026 of Hartree's calculated 1.355.[13] Our initial guesses at λ were 1.1, 1.5, 1.9, and the method converged to the same solution each time with no difference in computation time, which was about twenty seconds.

We implemented the method with unsophisticated techniques. Since the operator $R_{\lambda,l}U_\phi$ is compact, it is expected that as the dimension of $\mathcal{H}_n$ increases the matrix whose eigenvalues must be found will become more ill-conditioned. It would be interesting to see computational results obtained by more knowledgeable computists, using more sophisticated numerical techniques.

[1]D. Hartree, *The Calculation of Atomic Structures* (Wiley, New York, 1957).
[2]J. Slater, *Quantum Theory of Atomic Structure* (McGraw-Hill, New York, 1960), Appendix 16.
[3]M. Reeken, J. Math. Phys. **11**, 2505 (1970).
[4]J. Wolkowisky, Indiana Univ. Math. J. 22, 551 (1972).
[5]C. Stuart, Arch. Ratl. Mech. Anal. 51, 60 (1973).
[6]C. Stuart, J. Math. Anal. Appl. 49, 725 (1975).
[7]K. Gustafson and D. Sather, Rend. Mat. 4, 1 (1971).
[8]K. Gustafson and D. Sather, Arch. Ratl. Mech. Anal. 48, 109 (1972).
[9]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[10]E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), Chap. 9, Theorem 2.4.
[11]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1953), Chap. 6, Sec. 6.
[12]M. Krasnoselskii, *Topological Methods in Nonlinear Integral Equations* (Pergamon, New York, 1964).
[13]W. Wilson and R. Lindsay, Phys. Rev. 47, 684 (1935).

On the extrapolation of optical image data

G. A. Viano

Istituto di Scienze Fisiche dell'Università, Genova, Italy
and Istituto Nazionale Fisica Nucleare, Sezione di Genova, Italy
(Received 29 April 1975; revised manuscript received 10 February 1976)

In this paper we show that the extrapolation of an image's piece as well as the object-reconstruction problem are improperly posed in the sense that the solutions do not depend continuously on the data. We try to restore the stability for these problems introducing suitable additional constraints. In the present work we treat in detail only the extrapolation of the image data. At this purpose we use and illustrate two numerical methods, which are based on the doubly-orthogonality of the linear-prolate-spheroidal-functions. Finally a probabilistic approach to these questions is outlined.

I. INTRODUCTION

The problem of image restoration is quite old and it has been widely discussed in literature.[1] More precisely the problems which have been investigated are essentially two:

(a) The extrapolation of a given image piece beyond its borders, for recreating the entire image (see Ref. 1, p. 318);

(b) The reconstruction of the object from its optical image.[2]

Let us denote by $f(x)$ the complex amplitude distribution of a coherently illuminated one-dimensional object; then $f(x)$ is a space-limited function, since it vanishes outside the interval $|x| \leq \frac{1}{2}X_0$. Furthermore it can be represented as follows[2]:

$$f(x) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{+\infty} F(\omega) \exp(i\omega x)\, d\omega, \quad F(\omega) \in L^2(-\infty, +\infty). \tag{1}$$

Inverting (1), one obtains

$$F(\omega) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{+\infty} f(x) \exp(-i\omega x)\, dx, \tag{2}$$

where $F(\omega)$ is an entire function in the complex ω plane, since $f(x)$ is space-limited.[3] Then we denote by $\bar{f}(x)$ the image distribution, which is a band-limited function, since the pupil stop blocks all the waves with ω greater than a positive constant Ω (see Fig. 2 of Ref. 2). (Hereafter we shall refer to an optical system analogous to that considered in Ref. 2, Fig. 2, whose magnification is 1. However the general case is readily amenable to this system with only slight modifications.) Therefore, we can write

$$\bar{f}(x) = (\sqrt{2\pi})^{-1} \int_{-\Omega}^{\Omega} F(\omega) \exp(i\omega x)\, d\omega \tag{3}$$

and also $\bar{f}(x)$ is an entire function in the complex x plane. Now one could argue (as it has been observed by several authors) that, even if the knowledge of the function $F(\omega)$ is limited to the finite interval $|\omega| \leq \Omega$, nevertheless, thanks to the uniqueness of analytic continuation, one could uniquely determine $F(\omega)$ everywhere. Hence, one could reconstruct the object in all its details and there should be no ambiguity in interpreting the image and no loss of information in passing through the optical system.[2]

At this point we want to remark strongly that the argument above is not correct, because mere uniqueness is wholly inadequate and the problem must be reconsidered taking into account the question of stability. More generally we observe that in the procedures concerning the object-reconstruction problem, as well as in the extrapolation methods, which are in use in optics, not enough attention has been paid, up to now, to the questions concerning the continuous dependence of the solutions on the data. In fact, as is well known to mathematicians after Hadamard,[4] there is a large class of problems (usually called "ill-posed" or "improperly posed problems"), in which the solution depends uniquely but not continuously on the data. The Cauchy problem for elliptic equations, Fredholm integral equations of the first kind, elliptic continuation, and complex analytic continuation are a few of the classical problems which are not well posed in the sense of Hadamard. Now, as we shall see in Sec. II, the instability in the object-reconstruction problem as well as in the extrapolation of the image can be explicity shown; more precisely it shall be proved that an arbitrarily small error in the data can induce an arbitrarily large error in the solution.

Methods for obtaining a stable solution in the case of ill-posed problems of mathematical physics have, recently, undergone considerable development. More precisely these methods have been used in the numerical analytic continuation of scattering data in particle physics,[5] in geophysical research,[6] and more generally in many of the so-called "inverse problems".[7]

In this paper, we shall prove (see Sec. II), that the usual formulations of the problems (a) and (b) (i.e., the image-extrapolation and the object-reconstruction problems) are not correctly posed; then we shall consider in detail only the problem (a), showing how it is possible to restore the stability imposing a suitable stabilizing constraint. Then in Sec. III we shall discuss two numerical methods which give the nearly-best-possible approximations. In Sec. IV, we reconsider the problem from a probabilistic point of view and the approximations of Sec. III shall be reobtained. In Sec. V, we shall try some conclusions. Finally, the Appendix is devoted to the problem of the continuity of the analytic continuation in the specific case of band-limited functions and the stabilizing constraint of Sec. II shall be derived.

II. FORMULATION OF THE PROBLEMS (a) AND (b)

From the equality (3) combined with (2), we get

$$\bar{f}(x)=\int_{-X_0/2}^{X_0/2}\frac{\sin[\Omega(x-s)]}{\pi(x-s)}f(s)\,ds \tag{4}$$

which states that the image is represented by the convolution of the object and the diffraction image of a point source.[2]

Next we consider the integral operator given by

$$\int_{-X_0/2}^{X_0/2}\frac{\sin[\Omega(x-s)]}{\pi(x-s)}\psi(s)\,ds. \tag{5}$$

In Ref. 3, it has been proved that the kernel of (5) is positive definite. Moreover it is easy to prove that the symmetric integral operator, defined by (5), is compact[8] in the space of square-integrable functions over the interval $|x|\leq\frac{1}{2}X_0$. In fact, we have

$$\left\{\int_{-X_0/2}^{X_0/2}\left(\int_{-X_0/2}^{X_0/2}\left|\frac{\sin[\Omega(x-s)]}{\pi(x-s)}\right|^2 ds\right)dx\right\}^{1/2} \leq\frac{\Omega X_0}{\pi}. \tag{6}$$

As a consequence, the integral operator (5) admits a complete set of orthogonal eigenfunctions corresponding to a countably infinite set of real positive eigenvalues (see Ref. 9, Cap. VI). Therefore, we can write

$$\int_{-X_0/2}^{X_0/2}\frac{\sin[\Omega(x-s)]}{\pi(x-s)}\psi_n(s)\,ds=\lambda_n\psi_n(x) \tag{7}$$

with $\lambda_1>\lambda_2>\lambda_3>\cdots$.

The eigenfunctions $\psi_n(x)$ turn out to be the so-called "linear-prolate-spheroidal-functions", which have been extensively analyzed by Slepian *et al.*[1,3] Here we mention the main properties of these functions:

(i) the $\psi_n(x)$ are all band limited to Ω;

(ii) they are orthogonal and complete in the space of square integrable functions over the interval $|x|\leq\frac{1}{2}X_0$; more precisely we can write

$$\int_{-X_0/2}^{X_0/2}\psi_i(x)\psi_j(x)\,dx =\begin{cases}\lambda_i, & i=j,\\ 0, & i\neq j,\end{cases}\quad i,j=1,2,3,\ldots; \tag{8}$$

(iii) they are orthogonal and complete in the space of band-limited and square-integrable functions in the interval: $-\infty<x<+\infty$.

One remarkable property of these functions is their orthogonality over two different intervals; in Sec. III we shall make significant use of this property.

Now, recalling that $f(x)$ is a space-limited function, we write the following expansion:

$$f(x)=\sum_{n=1}^{\infty}f_n\varphi_n(x),\quad |x|\leq\tfrac{1}{2}X_0 \tag{9}$$

where we use, as a basis, the orthonormal functions: $\varphi_n(x)\equiv(\sqrt{\lambda_n})^{-1}\psi_n(x)$.

Next, following Toraldo di Francia,[2] we substitute (9) in formula (4) and using (7) we can write

$$\bar{f}(x)=\int_{-X_0/2}^{X_0/2}\frac{\sin[\Omega(x-s)]}{\pi(x-s)}\sum_{n=1}^{\infty}f_n\varphi_n(s)\,ds =\sum_{n=1}^{\infty}f_n\lambda_n\varphi_n(x), \tag{10}$$

where we have interchanged the order of summation and integration; however this exchange is legitimate since

$$\lim_{N\to\infty}\int_{-X_0/2}^{X_0/2}\left\{f(x)-\sum_{n=1}^{N}f_n\varphi_n(x)\right\}^2 dx\to 0. \tag{11}$$

In conclusion, we have the following expansion for the image distribution $\bar{f}(x)$:

$$\bar{f}(x)=\sum_{n=1}^{\infty}f_n\lambda_n\varphi_n(x)=\sum_{n=1}^{\infty}\bar{f}_n^0\varphi_n(x). \tag{12}$$

Therefore, from the experimental measurements of the image's coefficients $\bar{f}_n^0$ one could, in principle, reconstruct completely the object; i.e., to determine all the coefficients f_n. However, by the evaluation of the eigenvalues $\{\lambda_n\}$ which has been performed by many authors, one can observe that the behavior of these eigenvalues is quite similar to that of a step function (see, for instance, Ref. 2); more precisely, the values of λ_n for n sufficiently large are very near to zero. This means that the kernel of the integral equation (7) has a smoothing action and therefore arbitrarily small noise perturbations on the measurements of the image coefficients can induce arbitrarily large effects in the reconstructed object $f(x)$.

One is faced by a similar difficulty in the extrapolation of a given image piece beyond its borders. In fact, the experimental measurements of the image's coefficients are possible only over the finite interval $|x|\leq\frac{1}{2}X_0$; therefore if one wants to recreate the image over a larger domain then one must know the Fourier coefficients of the following expansion:

$$\bar{f}(x)=\sum_{n=1}^{\infty}\bar{f}_n\psi_n(x). \tag{13}$$

[Let us remark that an approximation obtained truncating the series (13) is itself band-limited, and therefore it must be preferred to an approximation, based on a Taylor series representation, which, of course, is not band-limited.]

Then the Fourier coefficients of the expansion (13) $\bar{f}_n$ are related to the measurable terms $\bar{f}_n^0$ through the following equality:

$$\bar{f}_n=(\sqrt{\lambda_n})^{-1}\bar{f}_n^0, \tag{14}$$

where each eigenvalue λ_n measures the relative amount of "energy" of functions $\psi_n(x)$ that is within interval $|x|\leq\frac{1}{2}X_0$, i.e.,

$$\lambda_n=\int_{-X_0/2}^{X_0/2}|\psi_n(x)|^2\,dx\Big/\int_{-\infty}^{+\infty}|\psi_n(x)|^2\,dx. \tag{15}$$

Therefore, also in this case, a small error on the coefficients $\bar{f}_n^0$ can induce large effects on the extrapolated image.

At this point it is necessary to distinguish between the ideal noiseless image distribution $\bar{f}(x)$, which contains all the information carried by the diffracted field, and the image distribution actually measured $\bar{h}(x)$. Then we write a first condition, which takes into account the noise of the measurements, i.e.,

$$\|\bar{f}(x) - \bar{h}(x)\|_{L^2(-X_0/2, X_0/2)} \leq \epsilon. \tag{16}$$

Of course this condition is not sufficient for the stability; we must introduce specific stabilizing constraints.

Now in order to find the most suitable stabilizing conditions, it is convenient to discuss separately the problems (a) and (b); i.e., the image extrapolation and the object-reconstruction problems. Let us start from the problem (a) (i.e., the image extrapolation). In this case we can use the following stabilizing constraint, which is very natural from the physical point of view:

$$\|\bar{f}(x)\|_{L^2(-\infty,+\infty)} \leq E. \tag{17}$$

In fact, in the Appendix, we shall prove that the bound (17) is sufficient for restoring the continuity to the analytic continuation of $\bar{f}(x)$ in any compact subdomain of the complex x plane. Then in Sec. III we shall find the approximations which are nearly-best-possible with respect to the conditions (16) and (17); as we shall see these numerical methods are largely based on the doubly orthogonality of the linear-prolate-spheroidal-functions.

Of course the bound (17) is not sufficient in order to guarantee the stability in the object-reconstruction problem [i.e., problem (b)].

In the latter case it is necessary to use a constraint like

$$\|Cf\|_{L^2(-X_0/2, X_0/2)} \leq E', \tag{18}$$

where C is stronger than the identity operator; for instance Cf gives the first or the second derivative of f. Unfortunately, the functions $\varphi_n(x)$ are no longer orthogonal with respect to C, in the sense that $(C\varphi_n, C\varphi_m)_{(-X_0/2, X_0/2)} \neq 0$ for $n \neq m$. Therefore, we cannot use a method of eigenfunction expansions (see Sec. III), for finding the nearly-best-possible approximations with respect to the constraints (16) and (18). Of course one can use a different procedure (see, for instance, the so-called least-squares method of Ref. 10); however, in any case, many appealing features of the eigenfunction expansion methods (especially those concerning the physical interpretation) are probably lost.

A second alternative is to use the natural constraint:

$$\|f\|_{L^2(-X_0/2, X_0/2)} \leq E''. \tag{19}$$

In this case, one can work with the method of eigenfunction expansion and use the orthogonality of the linear-prolate-spheroidal-functions. However, one cannot pretend of reconstructing stably the function f itself [since the bound (19) is too weak in this sense], but only the average of f over the interval $(x-\delta, x+\delta)$, where δ is a small number.

As it appears from these considerations the object reconstruction is a quite involved problem, and therefore in this paper we limit ourselves to discuss how to restore the stability in the problem (a) (i.e., the image-extrapolation). Of course we hope to return on the object-reconstruction problem elsewhere.

III. THE NEARLY-BEST-POSSIBLE APPROXIMATIONS FOR THE PROBLEM (a)

Instead of dealing with the two constraints (16) and (17) separately, we combine them quadratically into a single constraint. In fact if $\bar{f}$ satisfies (16) and (17) then it also satisfies

$$\|\bar{f}(x) - \bar{h}(x)\|^2_{L^2(-X_0/2, X_0/2)} + (\epsilon/E)^2 \|\bar{f}(x)\|^2_{L^2(-\infty,+\infty)} \leq 2\epsilon^2. \tag{20}$$

Conversely any $\bar{f}$ satisfying (20) satisfies also (16) and (17) except for a factor of at most $\sqrt{2}$. Therefore, as a first approximation, we shall look for that function which minimizes

$$\|\bar{f}(x) - \bar{h}(x)\|^2_{L^2(-X_0/2, X_0/2)} + (\epsilon/E)^2 \|\bar{f}(x)\|^2_{L^2(-\infty,+\infty)}. \tag{21}$$

At this purpose we shall closely follow a method which was first proposed by Miller (see Ref. 10) and which is called in Ref. 10 "the least-squares Method 1". Then we expand $\bar{f}(x)$ and $\bar{h}(x)$ in terms of the linear-prolate-spheroidal functions $\psi_n(x)$ as follows:

$$\bar{f}(x) = \sum_{n=1}^{\infty} \bar{f}_n \psi_n(x), \tag{22}$$

$$\bar{h}(x) = \sum_{n=1}^{\infty} \bar{h}_n \psi_n(x). \tag{23}$$

Remark: Recall that the Fourier coefficients $\bar{h}_n$ of the expansion (23) are related to the measurable terms $\bar{h}_n^0$ through the following relationship:

$$\bar{h}_n = (\sqrt{\lambda_n})^{-1} \bar{h}_n^0, \tag{24}$$

where $\bar{h}_n^0$ are the Fourier coefficients of the expansion of $\bar{h}(x)$ in terms of $\varphi_n(x)$.

Next, thanks to the expansion (22) and (23) and to the formula (8), we have

$$\|\bar{f}(x) - \bar{h}(x)\|_{L^2(-X_0/2, X_0/2)} = \left(\sum_{n=1}^{\infty} \lambda_n |\bar{f}_n - \bar{h}_n|^2\right)^{1/2}. \tag{25}$$

Analogously, for the property (iii) of the $\psi_n(x)$, we obtain

$$\|\bar{f}(x)\|_{L^2(-\infty,+\infty)} = \left(\sum_{n=1}^{\infty} |f_n|^2\right)^{1/2}. \tag{26}$$

Therefore, the expression (21) can be written as follows:

$$\left(\sum_{n=1}^{\infty} \lambda_n |\bar{f}_n - \bar{h}_n|^2\right) + \left(\frac{\epsilon}{E}\right)^2 \left(\sum_{n=1}^{\infty} |\bar{f}_n|^2\right) \tag{27}$$

and we can conclude that the coefficients of the approximation $\bar{f}^a$, which minimizes (27), are given by

$$\bar{f}_n^a = \frac{\lambda_n \bar{h}_n}{(\epsilon/E)^2 + \lambda_n} = \frac{(\sqrt{\lambda_n})\, \bar{h}_n^0}{(\epsilon/E)^2 + \lambda_n}, \tag{28}$$

and, finally,

$$\bar{f}^{a}=\sum_{n=1}^{\infty}\bar{f}_{n}^{a}\psi_{n}(x). \tag{29}$$

With this method the instability in the image extrapolation is partially removed; nevertheless, this procedure does not give a criterion which indicates the value n where one can truncate the series: $\sum_{n=1}^{\infty}\bar{f}_{n}^{a}\psi_{n}(x)$. For this reason we shall discuss a second approximation which will be quite useful in the numerical computations. With this in mind, and recalling the formulas (25) and (26) one can rewrite the conditions (26) and (17) as the square root of weighted quadratic sums:

$$\left\{\sum_{n=1}^{\infty}\left(\frac{\sqrt{\lambda_n}}{\epsilon}\right)^2\left|\bar{f}_n-\bar{h}_n\right|^2\right\}^{1/2}\leqslant 1, \tag{30}$$

$$\left\{\sum_{n=1}^{\infty}\left(\frac{1}{\epsilon}\right)^2\left|\bar{f}_n-0\right|^2\right\}^{1/2}\leqslant 1. \tag{31}$$

Remembering that the eigenvalues λ_n are decreasing and tend to zero as $n\to+\infty$, we can construct an approximation as follows: the coefficients of the approximation will be $\bar{h}_n$ for all that components where the weight $\sqrt{\lambda_n}/\epsilon$ is larger than $1/E$, and vice versa they will be zero for all those components where the weight $1/E$ is larger than $\sqrt{\lambda_n}/\epsilon$. In other words, we write

$$\bar{f}^{\alpha}=\sum_{n=1}^{\alpha}\bar{h}_n\psi_n(x)=\sum_{n=1}^{\alpha}(\sqrt{\lambda_n})^{-1}\bar{h}_n^0\psi_n(x), \tag{32}$$

where α is the largest integer such that

$$\lambda_n>(\epsilon/E)^2. \tag{33}$$

The procedure outlined above is the method which Miller[11] and Miller-Viano[5] call the "partial eigenfunction expansion Method 1".

Formula (33) shows that if $\epsilon\to 0$, then $\alpha\to\infty$. Moreover it indicates that the information which one can extract from the image's piece, contained in the interval $(-\frac{1}{2}X_0, \frac{1}{2}X_0)$ is carried in a finite number of degrees of freedom.

In this paper we have supposed that both the numbers ϵ and E are known. However, it is possible to elaborate methods which require the knowledge of only one of these numbers (see Ref. 10), but for the sake of clearness we prefer to return on this point elsewhere (see also Sec. V below).

Next we shall prove that the approximation $\bar{f}^{\alpha}$ is nearly-best-possible with respect to the conditions (16) and (17); more precisely we can show that

$$\|\bar{f}^{\alpha}-\bar{h}\|_{L^2(-X_0/2,X_0/2)}\leqslant 2\epsilon, \tag{34}$$

$$\|\bar{f}^{\alpha}\|_{L^2(-\infty,+\infty)}\leqslant 2E. \tag{35}$$

First, we shall prove the inequality (34). Let us write

$$\begin{aligned}\|\bar{f}^{\alpha}-\bar{h}\|_{L^2(-X_0/2,X_0/2)}&\leqslant\Big\|\sum_{n=\alpha+1}^{\infty}\bar{h}_n\psi_n(x)\Big\|_{L^2(-X_0/2,X_0/2)}\\&\leqslant\Big\|\sum_{n=\alpha+1}^{\infty}\bar{h}_n\psi_n(x)-\sum_{n=\alpha+1}^{\infty}\bar{f}_n\psi_n(x)\Big\|_{L^2(-X_0/2,X_0/2)}\\&\quad+\Big\|\sum_{n=\alpha+1}^{\infty}\bar{f}_n\psi_n(x)\Big\|_{L^2(-X_0/2,X_0/2)}.\end{aligned} \tag{36}$$

From (16) it follows

$$\Big\|\sum_{n=\alpha+1}^{\infty}\bar{h}_n\psi_n(x)-\sum_{n=\alpha+1}^{\infty}\bar{f}_n\psi_n(x)\Big\|_{L^2(-X_0/2,X_0/2)}\leqslant\epsilon, \tag{37}$$

moreover, thanks to the formulas (17) and (33) we have

$$\begin{aligned}\Big\|\sum_{n=\alpha+1}^{\infty}\bar{f}_n\psi_n(x)\Big\|_{L^2(-X_0/2,X_0/2)}&=\left(\sum_{n=\alpha+1}^{\infty}\lambda_n|\bar{f}_n|^2\right)^{1/2}\\&\leqslant\frac{\epsilon}{E}\left(\sum_{n=\alpha+1}^{\infty}|\bar{f}_n|^2\right)^{1/2}\leqslant\epsilon\end{aligned} \tag{38}$$

and therefore, from the formulas (37) and (38) the inequality (34) follows.

We can proceed in the same way, concerning the inequality (35). In fact, we have

$$\begin{aligned}\|\bar{f}^{\alpha}\|_{L^2(-\infty,+\infty)}&\leqslant\Big\|\sum_{n=1}^{\alpha}\bar{h}_n\psi_n(x)-\sum_{n=1}^{\alpha}\bar{f}_n\psi_n(x)\Big\|_{L^2(-\infty,+\infty)}\\&\quad+\Big\|\sum_{n=1}^{\alpha}\bar{f}_n\psi_n(x)\Big\|_{L^2(-\infty,+\infty)}\end{aligned} \tag{39}$$

then, thanks to inequality (17), we can write

$$\Big\|\sum_{n=1}^{\alpha}\bar{f}_n\psi_n(x)\Big\|_{L^2(-\infty,+\infty)}\leqslant E; \tag{40}$$

moreover, using (16) and (33), we obtain

$$\begin{aligned}&\Big\|\sum_{n=1}^{\alpha}\bar{h}_n\psi_n(x)-\sum_{n=1}^{\alpha}\bar{f}_n\psi_n(x)\Big\|_{L^2(-\infty,+\infty)}\\&=\left(\sum_{n=1}^{\alpha}|\bar{h}_n-\bar{f}_n|^2\right)^{1/2}\leqslant\frac{E}{\epsilon}\left(\sum_{n=1}^{\alpha}\lambda_n|\bar{h}_n-\bar{f}_n|^2\right)^{1/2}\leqslant E,\end{aligned} \tag{41}$$

and finally from (40) and (41) the inequality (35) follows. Thus we have proved that the method is nearly-best-possible, in the sense that it generates an approximation $\bar{f}^{\alpha}$ which satisfies nearly the same constraints as the unknown $\bar{f}$.

Let us remark that analogous methods, which are largely based on the orthogonality of the linear-prolate-spheroidal-functions, can also work in the case of the object reconstruction problem if one uses the conditions (16) and (19).

IV. THE PROBABILISTIC APPROACH

Before going to the probabilistic approach, we introduce a finite dimensional subspace of the infinite dimensional space of square-integrable function over the interval $(-\infty, +\infty)$. Let us denote with N the dimension of this subspace.

Let us observe that the errors which we make replacing the functions $\bar{f}(x)$ and $\bar{h}(x)$ by their projections on this subspace can be neglected since these errors become negligibly small by making N sufficiently large. Therefore, in place of the expansions (22) and (23) we shall write

$$\bar{f}(x)=\sum_{n=1}^{N}\bar{f}_n\psi_n(x), \tag{42}$$

$$\bar{h}(x)=\sum_{n=1}^{N}\bar{h}_n\psi_n(x), \tag{43}$$

Now, coming to the probabilistic approach, it is obvious that the coefficients which enter in the formulas (42) and (43) must be regarded as random variables. (For this type of approach see also Ref. 7 and 12.) Furthermore, let us assume that these random variables are Gaussian distributed. Then the stabilizing constraint (17) can be translated into probabilistic language, writing the following *a priori* probability density:

$$P(\bar{f}) = C_1 \exp\left(-\frac{1}{2E^2}\sum_{n=1}^{N} |\bar{f}_n|^2\right). \tag{44}$$

Analogously the following conditional-probability density of the vector $\bar{h}$ for a given vector $\bar{f}$:

$$P(\bar{h}|\bar{f}) = C_2 \exp\left(-\frac{1}{2\epsilon^2}\sum_{n=1}^{N} \lambda_n |\bar{f}_n - \bar{h}_n|^2\right) \tag{45}$$

is the probabilistic analog of the condition (16). Then using the following Bayes formula[13]

$$P(\bar{f}\,|\bar{h}) = P(\bar{f})P(\bar{h}|\bar{f}) / \int P(\bar{f})P(\bar{h}|\bar{f})\, d\bar{f} \tag{46}$$

where $\int d\bar{f}$ means integral over C^N, we obtain

$$P(\bar{f}\,|\bar{h}) = C_3 \exp - \frac{1}{2\epsilon^2}\left\{\sum_{n=1}^{N} \lambda_n |\bar{f}_n - \bar{h}_n|^2 + \left(\frac{\epsilon}{E}\right)^2 \sum_{n=1}^{N} |f_n|^2\right\}. \tag{47}$$

Now we can find a mean value for any component of the random vector $\bar{f}$; more precisely we have

$$\langle \bar{f}_n \rangle = \int \bar{f}_n P(\bar{f}\,|\bar{h})\, d\bar{f} = \frac{(\sqrt{\lambda_n})\bar{h}_n^0}{(\epsilon/E)^2 + \lambda_n} = \bar{f}_n^a \tag{48}$$

which coincides exactly with the formula (28).

Next we want to reconsider the second approximation of Sec. III, from a probabilistic point of view. We can say that the probability density (47) is essentially the product of the probability densities (44) and (45). Moreover for the first α components [α is the largest integer such that $\lambda_n > (\epsilon/E)^2$] the probability density (44) is more dispersed than the (45); vice versa for the components from $\alpha + 1$ to N, the probability density (44) is more concentrated than the (45). This consideration suggests an approximation where the first α components are given by $\bar{h}_n$ ($n = 1, 2, \ldots, \alpha$) and the components from $\alpha + 1$ to N are given by the mean values corresponding to the normal probability density (44), i.e., zero. This approximation can be written as follows:

$$\bar{f}_n^{\alpha} = \sum_{n=1}^{\alpha} \bar{h}_n \psi_n(x) = \sum_{n=1}^{\alpha} \bar{h}_n^0 (\sqrt{\lambda_n})^{-1} \psi_n(x) \tag{49}$$

and it coincides exactly with the approximation (32).

V. CONCLUSION

In this paper, we have proved that the extrapolation of an optical image's piece as well so the object-reconstruction problem are improperly posed in the sense of Hadamard. Therefore, it is necessary to introduce suitable stabilizing constraints in order to restore the stability to the problems connected with the extraction of information from an optical image. In this first paper, we have treated in detail the image's extrapolation, discussing the nearly-best-possible approximations for this problem. For the sake of simplicity we have considered the one-dimensional case, but it is reasonable to conjecture that the extension to the bidimensional case does not present serious difficulties.

Of course, there remains much work to be done; here we limit ourselves to mention the most relevant questions on which we hope to return:

(a) to treat in detail the object-reconstruction problem, considering both the constraints (18) and (19);

(b) to investigate the extraction of information supposing that the object is noncoherently illuminated;

(c) to consider the case when only one of the two numbers (the error bound ϵ and the constraint E) is known and to elaborate appropriate numerical methods (see Ref. 10);

(d) to elaborate a quantitative and numerical analysis in order to see how poor the restored stability is in the various cases.

APPENDIX

In this Appendix, we want to find a stabilizing constraint for the analytic continuation of band-limited functions. Let us denote by W_η the class of functions $f(z)$ which admit a representation of the following form:

$$f(z) \equiv f(x + iy) = (\sqrt{2\pi})^{-1} \int_{-\eta}^{\eta} \exp(iuz)\varphi(u)\, du,$$

$$\varphi(u) \in L^2(-\eta, \eta) \tag{A1}$$

We suppose that $f(z)$ is approximately known (within a certain accuracy) on a finite interval of the real axis; then we analyze the continuity of the analytic continuation from this interval to any compact subset of the complex z plane, which shall be denoted hereafter by Λ. Now let us recall that Λ is an open set, union of a sequence $\{\Lambda_j\}$ ($j = 0, 1, 2, \ldots$) of closed bounded subsets, such that Λ_j is contained in Λ_{j+1} and any compact subset of Λ lies in some Λ_j.[14]

Moreover it is necessary to introduce suitable metrics for both the data space Y (a certain space of functions on the data set) and the solution space X (see Refs. 5 and 15). Concerning the solution space, we define the distance between two functions f and g in X, as follows (see Ref. 15):

$$d(f, g) = d(f - g, 0) = \sum_{j=0}^{\infty} 2^{-j}\left(\frac{\|f - g\|_j}{1 - \|f - g\|_j}\right), \tag{A2}$$

where we denote by $\|f\|_j$ the uniform norm of f on Λ_j. The distance (A2) satisfies the triangle inequality and the sequence $f_m \to f$ subuniformly in Λ (i.e., $f_m(z) \to f(z)$ uniformly on each closed bounded subset of Λ) if and only if $d(f_m, f) \to 0$.

After these preliminaries let us recall the following theorem:

Theorem (see Ref. 16, p.141): Let σ be a continuous map on a compact topological space into a Hausdorff topological space; if σ is $1 - 1$, then its inverse map σ^{-1} is continuous.

From this theorem it follows that the compactness of the solution space is sufficient to guarantee the continuity of σ^{-1}, since the uniqueness theorem of analytic continuation guarantees that σ is $1-1$.

Now let us recall some inequalities which are satisfied by the functions which belong to the class W_η. The first is the following:

$$|f(z)| \leq C \exp(\eta |z|). \tag{A3}$$

In fact, from (A1) it follows:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \int_{-\eta}^{\eta} |\varphi(u)|^2 du < \infty. \tag{A4}$$

Moreover, we can write

$$f(z) = (\sqrt{2\pi})^{-1} \int_{-\eta}^{\eta} \exp(iux) \exp(-uy) \varphi(u)\, du \tag{A5}$$

and then, through the Schwarz inequality, we get

$$|f(z)| \leq (\sqrt{2\pi})^{-1} \left(\int_{-\eta}^{\eta} |\varphi(u)|^2 du\right)^{1/2} \left(\int_{-\eta}^{\eta} \exp(-2uy)\, du\right)^{1/2}$$
$$= C_1[(\exp(2\eta y) - \exp(-2\eta y))/y]^{1/2} \leq C \exp(\eta |y|). \tag{A6}$$

Let us denote by B_η the totality of those entire transcendental functions of exponential type whose exponent is not greater than η and which satisfy the inequality:

$$\sup_{-\infty<x<\infty} |f(x)| < \infty. \tag{A7}$$

We observe that *all* the functions in W_η belong also to B_η. Then we recall the following Bernstein theorem (for the proof of this theorem and of the inequalities with follow see Ref. 17, p. 138):

Theorem (Bernstein): If the function $f(z)$ lies in B_η then $f(z)$ satisfies the inequality

$$\sup_{-\infty<x<\infty} |f'(x)| \leq \eta \sup_{-\infty<x<\infty} |f(x)|. \tag{A8}$$

Then from (A8) it follows

$$\sup_{-\infty<x<\infty} |f^{(k)}(x)| \leq \eta^k \sup_{-\infty<x<\infty} |f(x)|, \tag{A9}$$

where k is any positive integer.

Finally it is possible to prove, through the formulas (A8) and (A9), that the inequality

$$|f'(z)| \leq C\eta \exp(\eta |y|), \quad z = x + iy \tag{A10}$$

holds for all the functions which belong to B_η.

Now, thanks to the inequalities (A3) and (A10), it is possible to prove the following: suppose that all the functions which belong to the set $I \subset W_\eta$ are uniformly bounded on the real axis, then the functions in I are equicontinuous in every closed bounded point set of the complex plane Λ (see Ref. 17, p. 139). Hence every infinite sequence of functions in I contains a partial sequence which converges uniformly in every closed bounded subset of the complex plane Λ. This is sufficient for saying that the functions belonging to I form a normal family of functions. Furthermore, the limit functions also belong to I, and therefore if the solution space is the space of functions belonging to I [with a metric given by the formula (A2)], then this solution space is certainly compact.

Now a condition like that given by formula (A4), i.e.,

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty \tag{A11}$$

is not sufficient to guarantee the uniformly boundedness on the real axis. Therefore, we shall construct the set $I \subset W_\eta$ with those functions which, besides to belong to W_η satisfy the following inequality:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx \leq K, \tag{A12}$$

where K is a prescribed constant, which does not depend on the particular function f. Analogous requirement is also necessary for the constant η. The condition (A12) gives the stabilizing constraint (14) for the image distribution $\bar{f}(x)$ if we put $(K)^{1/2} = E$. Of course the requirement on the constant η must be translated into an analogous condition on Ω. These conditions are quite natural and can be easily satisfied from the physical point of view.

ACKNOWLEDGMENTS

It is a pleasure to thank my friend Professor K. Miller for his very illuminating criticisms and suggestions. I am also deeply indebted to Professor L. Abbozzo-Ronchi and Professor G. Toraldo di Francia for very useful discussions on many topics concerning this paper. Finally, I am indebted to Professor M. Bertero, Professor A. Borsellino, and Professor C. Pontiggia for a critical reading of the manuscript.

[1]B.R. Frieden, *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1971), Vol. 9.
[2]G. Toraldo di Francia, Nuovo Cimento 1, 460 (1969), see also the papers quoted there.
[3]D. Slepian and H.O. Pollak, Bell System Tech. J. **40**, 43 (1961).
[4]J. Hadamard, *Le probleme de Cauchy* (Hermann, Paris, 1932), p. 40.
[5]K. Miller and G.A. Viano, J. Math. Phys. **14**, 1037 (1973), see also the papers quoted there.
[6]G. Backus, in *Mathematical Problems in the Geophysical Sciences*, edited by W.M. Reid (American Math. Soc. Providence, R.I., 1971), Vol. 2.
[7]V.F. Turchin, V.P. Kozlov, and M.S. Malkevich, Soc. Phys. Usp. **13**, 681 (1971).
[8]N. Dunford and J.I. Schwartz, *Linear Operators Part. I* (Interscience, New York, 1958), p. 518.
[9]F. Riesz and B.Sz. Nagy, *Functional Analysis* (Ungar, New York, 1955), Chap. VI.
[10]K. Miller, SIAM J. Math. Anal. 1, 52 (1970).
[11]K. Miller, Arch. Rat. Mech. Anal. 16, 126 (1964).
[12]G.A. Viano, Lett. Nuovo Cimento 10, 591 (1974).
[13]M. Fisz, *Probability Theory and Mathematical Statistics* (Wiley, New York, 1965).
[14]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966), p. 253.
[15]K. Miller and G.A. Viano, Nucl. Phys. B 25, 460 (1971).
[16]J. Kelley, *General Topology* (Van Nostrand, Princeton, 1955), p. 141.
[17]N.I. Akhiezer, *The Theory of Approximation* (Ungar, New York, 1956).

Integration formula for Wigner 3-j coefficients*

D. E. Winch

Department of Applied Mathematics, University of Sydney, Sydney, N.S.W., 2006, Australia
(Received 14 July 1975)

An integration formula is given for Wigner 3-j coefficients having 72 different forms associated directly with the symmetry properties of the Regge square. The integration formula permits a direct derivation of the integral of a product of three rotation matrix elements, and is shown to include an integration formula for spin projection coefficients as a special case.

1. INTRODUCTION

There are several forms for the series which is used to define Wigner 3-j coefficients, but the series given by Racah,[1] namely

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-)^{j_1-j_2-m_3}[(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)! \times(j_3-m_3)!\,(j_3+m_3)!\,(j_1+j_2-j_3)!\,(j_1-j_2+j_3)! \times(-j_1+j_2+j_3)!/(j_1+j_2+j_3+1)!]^{1/2} \times\sum_t (-)^t[(j_1-m_1-t)!\,(j_3-j_2+m_1+t)!\,(j_2+m_2-t)! \times(j_3-j_1-m_2+t)!\,t!\,(j_1+j_2-j_3-t)!]^{-1},$$

summed over all t giving nonnegative arguments for factorial expressions, is the most symmetric. The 3-j coefficient is defined for integer or half odd integer values of its parameters, and is defined to be zero unless the nine expressions $-j_1+j_2+j_3$, $j_1-j_2+j_3$, $j_1+j_2-j_3$, j_1-m_1, j_2-m_2, j_3-m_3, j_1+m_1, j_2+m_2, j_3+m_3, are nonnegative integers, and unless $m_1+m_2+m_3$ is zero. Racah noted most of the symmetry properties (or equivalent expressions) of 3-j coefficients, but it wasn't until 1958 that Regge[2] found two further symmetries, and introduced the convention now known as the Regge square for 3-j coefficients:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{vmatrix} -j_1+j_2+j_3 & j_1-j_2+j_3 & j_1+j_2-j_3 \\ j_1-m_1 & j_2-m_2 & j_3-m_3 \\ j_1+m_1 & j_2+m_2 & j_3+m_3 \end{vmatrix}.$$

The series for the Wigner 3-j coefficient remains invariant under cyclic permutations of rows or of columns of the Regge square and also under rotation of the Regge square about a diagonal. The series for the 3-j coefficient is multiplied by $(-)^{j_1+j_2+j_3}$ under noncyclic permutations of rows or of columns of the Regge square. Hence there are 72 different forms or symmetries of the Regge square or 3-j coefficient having series which differ at most by sign.

The purpose of this paper is to give an integration formula for the representation of Wigner 3-j coefficients which exhibits the various forms required by the Regge symmetries when the identities of Schendel[3] (associated with the symmetry properties of rotation matrix elements) and the process of integration by parts are applied. The integration formula includes as a special case the integration formula for spin projection coefficients discussed in a series of papers by Percus and Rotenberg,[4] Sasaki and Ohno,[5] Smith,[6] Smith and Harris,[7] and Mano.[8,9] The integration formula also allows a direct derivation of the integral of a product of three rotation matrix elements, and therefore of the Wigner–Eckart theorem.

2. THEORY

The principal result to be proven here is that

$$\frac{(-)^{2j_1}}{(j_1+j_2-j_3)!}\int_{-1}^{1}(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}\left(\frac{d}{d\mu}\right)^{j_1+j_2-j_3}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_1+m_1}]\,d\mu$$
$$=2^{j_1+j_2+j_3+1}\left[\frac{(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)!\,(j_3-m_3)!\,(j_3+m_3)!}{(j_1+j_2-j_3)!\,(j_1-j_2+j_3)!\,(-j_1+j_2+j_3)!\,(j_1+j_2+j_3+1)!}\right]^{1/2}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad (1)$$

in which the five indices of the integrand are sufficient to specify the appropriate Regge square, since the sums of rows or columns of a Regge square are all equal to $j_1+j_2+j_3$. Note that the degree of the polynomial in the integrand is also $j_1+j_2+j_3$. Integrals of the same form as that on the left of Eq. (1) have been treated by Watson[10] and Bailey[11] from the point of view of generalized hypergeometric functions ${}_3F_2(1)$. They derived "reduction" formulas for the integral, which, by virtue of Eq. (1), are equivalent to the well-known recurrence relations for Wigner 3-j coefficients.

We use I to denote the integral, thus

$$I=\int_{-1}^{1}(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}\left(\frac{d}{d\mu}\right)^{j_1+j_2-j_3}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_1+m_1}]\,d\mu, \quad (2)$$

and apply Leibnitz's theorem for the multiple differentiation of a product of two functions to the integrand, carry out the required differentiations, interchange the order of summation and integration of the finite number of terms in the series, and obtain

$$I=\sum_t \frac{(j_1-m_1)!\,(j_1+m_1)!\,(j_1+j_2-j_3)!}{(j_1+j_2-j_3-t)!\,t!\,(j_3-j_2+m_1+t)!\,(j_1-m_1-t)!}\int_{-1}^{1}(\mu-1)^{j_1+j_2+m_3-t}(\mu+1)^{j_3-m_3+t}\,d\mu.$$

Applying the integral definition of the beta function, the integral can now be evaluated to give

$$I=(-)^{j_1+j_2+m_3}\;2^{j_1+j_2+j_3+1}\frac{(j_1-m_1)!\,(j_1+m_1)!\,(j_1+j_2-j_3)!}{(j_1+j_2+j_3+1)!}\sum_t\frac{(-)^t(j_1+j_2+m_3-t)!\,(j_3-m_3+t)!}{(j_1+j_2-j_3-t)!\,(j_1-m_1-t)!\,t!\,(j_3-j_2+m_1+t)!}. \tag{3}$$

The series summed over t in Eq. (3) will now be converted to the symmetric series used to define the Wigner 3-j coefficients. By comparing coefficients of x^r on both sides of the identity $(1+x)^{n+m}=(1+x)^n(1+x)^m$, we obtain the binomial sum theorem,

$$\binom{n+m}{r}=\sum_u\binom{n}{u}\binom{m}{r-u}, \tag{4}$$

and with the following substitutions:

$$n=j_1-m_1-t,\quad m=j_2-m_2,\quad r=j_3+m_3,\quad n+m=j_1+j_2+m_3-t,$$

the last relationship making use of the condition $m_1+m_2+m_3=0$, we obtain, after rearranging the factorial expressions

$$\frac{(j_1+j_2+m_3-t)!}{(j_1+j_2-j_3-t)!\,(j_1-m_1-t)!}=\sum_u\frac{(j_2-m_2)!\,(j_3+m_3)!}{(j_1-m_1-t-u)!\,u!\,(j_2-j_3+m_1+u)!\,(j_3+m_3-u)!}. \tag{5}$$

Substitute Eq. (5) into the series of Eq. (3) to obtain

$$\sum_t\frac{(-)^t(j_1+j_2+m_3-t)!\,(j_3-m_3+t)!}{(j_1+j_2-j_3-t)!\,(j_1-m_1-t)!\,(j_3-j_2+m_1+t)!\,t!}$$

$$=\sum_u\frac{(j_2-m_2)!\,(j_3+m_3)!}{(j_3+m_3-u)!\,u!\,(j_2-j_3+m_1+u)!\,(j_1-j_2+j_3-u)!}\sum_t\frac{(-)^t(j_1-j_2+j_3-u)!\,(j_3-m_3+t)!}{t!\,(j_1-m_1-t-u)!\,(j_3-j_2+m_1+t)!}, \tag{6}$$

in which a factor $(j_1-j_2+j_3-u)!$ has been introduced into the numerator and denominator of the expression on the right of Eq. (6). By equating coefficients of x^r on both sides of the identity $(1+x)^{-n+m-1}=(1+x)^{-n-1}(1+x)^m$, $n+1>m$, we obtain the binomial sum theorem in the form

$$(-)^r\binom{n-m+r}{r}=\sum_t(-)^t\binom{n+t}{t}\binom{m}{r-t}, \tag{7}$$

and with the following values for the parameters of Eq. (7):

$$n=j_3-m_3,\quad m=j_1-j_2+j_3-u,\quad r=j_1-m_1-u,\quad n-m+r=j_2+m_2,$$

we are able to derive an expression for the series summed over t on the right of Eq. (6),

$$\sum_t\frac{(-)^t(j_1-j_2+j_3-u)!\,(j_3-m_3+t)!}{t!\,(j_1-m_1-t-u)!\,(j_3-j_2+m_1+t)!}=(-)^{j_1-m_1-u}\frac{(j_2+m_2)!\,(j_3-m_3)!}{(j_1-m_1-u)!\,(j_2-j_1-m_3+u)!}. \tag{8}$$

We now substitute Eqs. (6) and (8) into Eq. (3), replacing the sum over u by a sum over t where $u=j_1-m_1-t$, to obtain

$$I=2^{j_1+j_2+j_3+1}\frac{(j_1+j_2-j_3)!}{(j_1+j_2+j_3+1)!}\sum_t\frac{(-)^{j_1+j_2+m_3+t}(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)!\,(j_3-m_3)!\,(j_3+m_3)!}{(j_1-m_1-t)!\,(j_3-j_2+m_1+t)!\,(j_2+m_2-t)!\,(j_3-j_1-m_2+t)!\,t!\,(j_1+j_2-j_3-t)!} \tag{9}$$

and hence it follows directly from the series used to define the Wigner 3-j coefficient, that the result of Eq. (1), an integration formula for Wigner 3-j coefficients, has been established.

3. REGGE SYMMETRIES

Corresponding to the 72 Regge symmetries of the Wigner 3-j coefficients there are 72 corresponding integration formulas. To obtain these formulas, we note that Eq. (1), expressed as an integration formula for the Regge square corresponding to the 3-j coefficient, is

$$\begin{vmatrix} -j_1+j_2+j_3 & j_1-j_2+j_3 & j_1+j_2-j_3 \\ j_1-m_1 & j_2-m_2 & j_3-m_3 \\ j_1+m_1 & j_2+m_2 & j_3+m_3 \end{vmatrix}$$

$$=2^{-j_1-j_2-j_3-1}\left[\frac{(j_1+j_2-j_3)!\,(j_1-j_2+j_3)!\,(-j_1+j_2+j_3)!\,(j_1+j_2+j_3+1)!}{(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)!\,(j_3-m_3)!\,(j_3+m_3)!}\right]^{1/2}$$

$$\times\frac{(-)^{2j_1}}{(j_1+j_2-j_3)!}\int_{-1}^{1}(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}\left(\frac{d}{d\mu}\right)^{j_1+j_2-j_3}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_1+m_1}]\,d\mu.$$

There are further 71 Regge squares, whose series' expressions as given in Sec. 1 differ at most by sign. To each of these squares there corresponds an integration formula, but to give all 71 squares and the corresponding integration formulas would be an extravagant waste of space. However, one Regge symmetry is used twice in the following sections, and it is appropriate to give it explicitly,

$$\begin{array}{|ccc|}\hline j_3-m_3 & j_1+j_2-j_3 & j_3+m_3 \\ j_2-m_2 & j_1-j_2+j_3 & j_2+m_2 \\ j_1-m_1 & -j_1+j_2+j_3 & j_1+m_1 \\ \hline \end{array}$$

$$= 2^{-j_1-j_2-j_3-1}\left[\frac{(j_3+m_3)!\,(j_1+j_2-j_3)!\,(j_3+m_3)!\,(j_1+j_2+j_3+1)!}{(j_2-m_2)!\,(j_1-m_1)!\,(j_1-j_2+j_3)!\,(-j_1+j_2+j_3)!\,(j_2+m_2)!\,(j_1+m_1)!}\right]^{1/2}$$

$$\times\frac{(-)^{j_1+j_2+m_3}}{(j_3+m_3)!}\int_{-1}^{1}(\mu-1)^{j_1-j_2+j_3}(\mu+1)^{-j_1+j_2+j_3}\left(\frac{d}{d\mu}\right)^{j_3+m_3}[(\mu-1)^{j_2-m_2}(\mu+1)^{j_1-m_1}]\,d\mu .$$

The Regge squares are equivalent, since the second is obtained from the first by interchanging columns 1 and 3, then rows 1 and 2, and finally rotating about the principal diagonal. We may now equate the corresponding integration formulas to obtain

$$\frac{(-)^{2j_1}}{(j_1+j_2-j_3)!}\int_{-1}^{1}(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}\left(\frac{d}{d\mu}\right)^{j_1+j_2-j_3}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_1+m_1}]\,d\mu$$

$$=\frac{(-)^{j_1+j_2+m_3}(j_3-m_3)!}{(j_1-j_2+j_3)!\,(-j_1+j_2+j_3)!}\int_{-1}^{1}(\mu-1)^{j_1-j_2+j_3}(\mu+1)^{-j_1+j_2+j_3}\left(\frac{d}{d\mu}\right)^{j_3+m_3}[(\mu-1)^{j_2-m_2}(\mu+1)^{j_1-m_1}]\,d\mu . \quad (10)$$

It is also possible to derive the rhs of Eq. (10) directly from its lhs by applications of the Schendel identities [Eqs. (A1)—(A3)], and integration by parts. Thus, use $j=j_1$, $M=j_2-j_3$, $m=-m_1$ in Eq. (A1), integrate by parts $j_1-j_2+j_3$ times; use $j=j_3$, $M=j_1-j_2$, $m=m_3$ in Eq. (A3), and finally integrate by parts j_3+m_3 times.

4. SPIN PROJECTION COEFFICIENTS

The spin projection coefficients (SPC) of Sasaki and Ohno[5] and Smith[6] are defined by

$$C_j(S,M,n)=(-)^j(2S+1)\int_0^1 x^j(1-x)^{n-j+M}\,{}_2F_1(S+M+1,\,-S+M;\,1;\,x)\,dx, \quad (11)$$

which can be transformed into an integral of the form given in Eq. (1), and hence expressed as a 3-j coefficient. The hypergeometric function is dealt with by noting that

$$\left(\frac{d}{dx}\right)^{S+M}[x^{S+M}(1-x)^{S-M}]=(S+M)!\,{}_2F_1(S+M+1,\,-S+M;\,1;\,x),$$

so that with the substitution $x=(1-\mu)/2$, $(1-x)=(1+\mu)/2$, the Eq. (11) defining the SPC becomes

$$C_j(S,M,n)=\frac{1}{2^{n+S+1}}\,\frac{2S+1}{(S+M)!}\int_{-1}^{1}(\mu-1)^j(\mu+1)^{n-j+M}\left(\frac{d}{d\mu}\right)^{S+M}[(\mu-1)^{S+M}(\mu+1)^{S-M}(\mu+1)^{S-M}]\,d\mu . \quad (12)$$

Hence, from the integration formula for 3-j coefficients, it follows that the SPC is given by

$$C_j(S,M,n)=(-)^{2S}(2S+1)j!\left(\frac{(n-j-M)!\,(n-j+M)!}{(n-S)!\,(n+S+1)!}\right)^{1/2}\begin{pmatrix} S & (n+M)/2 & (n-M)/2 \\ -M & -j+(n+M)/2 & j-(n-M)/2\end{pmatrix}. \quad (13)$$

The Regge square corresponding to the 3-j coefficient of Eq. (13) is

$$\begin{array}{|ccc|}\hline n-S & S-M & S+M \\ S+M & j & n-j-M \\ S-M & n-j+M & j \\ \hline \end{array},$$

and the Regge symmetry about the principal diagonal of the square implies that $C_j(S,M,n)=C_j(S,-M,n)$, a result which is obvious from the generalized hypergeometric series for SPC's given by Smith and Harris.[7] Another symmetry property of SPC's is obtained by using Eq. (10) to give another form for the integral of Eq. (12) defining the SPC. The new integral also has the form of an SPC, and gives rise to the relationship

$$C_j(S,M,n)=(-)^{j-S+M}\frac{2S+1}{j+S+M+1}\,\frac{j!\,j!}{(S-M)!\,(n-S)!}\,C_{S-M}[(j+S+M)/2,\;(j-S-M)/2,\;n-(j-S+M)/2]. \quad (14)$$

It should be noted that if the parameters of the SPC are such that any of the terms of the Regge square are zero, then the corresponding SPC is given by a single term. Note also that recurrence relations between contiguous SPC's may be obtained directly from the known recurrence relations between 3-j coefficients.

5. ROTATION MATRIX ELEMENTS

Rotation matrix elements $\mathcal{D}^j_{Mm}(\alpha\beta\gamma)$ for the rotation of axes through Eulerian angles $(\alpha\beta\gamma)$ are given by Brink and Satchler[12] as

$$\mathcal{D}^j_{Mm}(\alpha\,\beta\gamma) = \exp[-i(M\alpha + m\gamma)]\,d^j_{Mm}(\beta),$$

where reduced rotation matrix elements with $\mu = \cos\beta$ are defined by the Rodrigues' formula

$$d^j_{Mm}(\mu) = \frac{(-)^{m-M}}{2^j(j+M)!}\left(\frac{(j-M)!\,(j+M)!}{(j-m)!\,(j+m)!}\right)^{1/2}(1+\mu)^{(M+m)/2}(1-\mu)^{(M-m)/2}\left(\frac{d}{d\mu}\right)^{j+M}[(\mu-1)^{j+m}(\mu+1)^{j-m}]. \tag{15}$$

The symmetry properties of the reduced rotation matrix elements,

$$d^j_{M,m}(\mu) = (-)^{M-m}d^j_{-M,-m}(\mu) = d^j_{-m,-M}(\mu) = (-)^{M-m}d^j_{m,M}(\mu),$$

follow directly from the Schendel identities, Eq. (A1)–(A3). From the integration formula, Eq. (1), and the Rodrigues' formula for reduced rotation matrix elements with appropriate parameters substituted, we obtain the following formula:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \frac{(-)^{2j_1+j_3+m_3}}{2^{j_2+j_3+1}}\left(\frac{(-j_1+j_2+j_3)!\,(j_1+j_2+j_3+1)!}{(j_2-m_2)!\,(j_2+m_2)!\,(j_3-m_3)!\,(j_3+m_3)!}\right)^{1/2}$$
$$\times\int_{-1}^{1}(1-\mu)^{(j_2+j_3+m_3-m_2)/2}(1+\mu)^{(j_2+j_3-m_3+m_2)/2}d^{j_1}_{j_2-j_3,-m_1}\,d\mu. \tag{16}$$

Application of the orthogonality properties of rotation matrix elements and 3-j coefficients to Eq. (16) leads to two further formulas which need not be given explicitly here.

The result for the integral of a product of three rotation matrix elements (including spherical harmonics and Legendre polynomials as special cases) is obtained so simply and directly by means of the integration formula for 3-j coefficients, that it would seem to be appropriate to give the derivation here. The integral of a product of three rotation matrix elements over the ranges of the three Eulerian angles, when expressed in terms of the reduced rotation matrix elements, becomes

$$D = \tfrac{1}{2}\int_{-1}^{1} d^{j_1}_{M_1m_1}\,d^{j_2}_{M_2m_2}\,d^{j_3}_{M_3m_3}\,d\mu, \tag{17}$$

in which the orthogonality of the complex exponential factors from the rotation matrix elements requires that $M_1+M_2+M_3=0$, and $m_1+m_2+m_3=0$. Substituting the Rodrigues' formula Eq. (15) for the reduced rotation matrix elements into Eq. (17) gives

$$D = \frac{2^{-j_1-j_2-j_3-1}}{(j_1+M_1)!\,(j_2+M_2)!\,(j_3+M_3)!}\left(\frac{(j_1-M_1)!\,(j_1+M_1)!\,(j_2-M_2)!\,(j_2+M_2)!\,(j_3-M_3)!\,(j_3+M_3)!}{(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)!\,(j_3-m_3)!\,(j_3+m_3)!}\right)^{1/2}E,$$

where the expression E is given by

$$E = \int_{-1}^{1}\left(\frac{d}{d\mu}\right)^{j_1+M_1}[(\mu-1)^{j_1+m_1}(\mu+1)^{j_1-m_1}]\left(\frac{d}{d\mu}\right)^{j_2+M_2}[(\mu-1)^{j_2+m_2}(\mu+1)^{j_2-m_2}]\left(\frac{d}{d\mu}\right)^{j_3+M_3}[(\mu-1)^{j_3+m_3}(\mu+1)^{j_3-m_3}]\,d\mu.$$

After integrating by parts j_3+M_3 times and applying Leibnitz's theorem on the multiple differentiation of the product of two functions, we obtain

$$E = (-)^{j_3+M_3}\int_{-1}^{1}\sum_r\binom{j_3+M_3}{r}(\mu-1)^{j_3+m_3}(\mu+1)^{j_3-m_3}\left(\frac{d}{d\mu}\right)^{j_1+j_3-M_2-r}[(\mu-1)^{j_1+m_1}(\mu+1)^{j_1-m_1}]\left(\frac{d}{d\mu}\right)^{j_2M_2+r}[(\mu-1)^{j_2+m_2}(\mu+1)^{j_2-m_2}]\,d\mu. \tag{18}$$

The undifferentiated factors in the integrand on the right of Eq. (18) are replaced, using

$$(\mu-1)^{j_3+m_3}(\mu+1)^{j_3-m_3} = (\mu-1)^{j_3-M_2-r-m_1}(\mu+1)^{j_3-M_2-r-m_1}\times(\mu-1)^{M_2+r-m_2}(\mu+1)^{M_2+r+m_2},$$

when application of the identity, Eq. (A1), gives the integral for E in the form

$$E = (-)^{j_3+M_3}\int_{-1}^{1}\sum_r\binom{j_3+M_3}{r}\frac{(j_1+j_3-M_2-r)!}{(j_1-j_3+M_2+r)!}\left(\frac{d}{d\mu}\right)^{j_1-j_3+M_2+r}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_1+m_1}]$$
$$\times\frac{(j_2+M_2+r)!}{(j_2-M_2-r)!}\left(\frac{d}{d\mu}\right)^{j_2-M_2-r}[(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}]\,d\mu. \tag{19}$$

After integrating by parts j_2-M_2-r times, the integral is broken into a product of two distinct parts. One part, a series S, contains the parameters $j_1j_2j_3M_1M_2M_3$ only, whilst the second part, an integral of the form given in Eq.

(1), contains $j_1 j_2 j_3 m_1 m_2 m_3$. Thus the Wigner-Eckart theorem has been established directly from the Rodrigues formulas for the rotation matrix elements. We find that the integral E becomes

$$E = S\int_{-1}^{1}(\mu-1)^{j_2-m_2}(\mu+1)^{j_2+m_2}\left(\frac{d}{d\mu}\right)^{j_1+j_2-j_3}[(\mu-1)^{j_1-m_1}(\mu+1)^{j_2+m_2}]\,d\mu, \tag{20}$$

where the series S is given by

$$S = (-)^{-j_2+j_3-M_1}\sum_r(-)^r\binom{j_3+M_3}{r}\frac{(j_1+j_3-M_2-r)!\,(j_2+M_2+r)!}{(j_1-j_3+M_2+r)!\,(j_2-M_2-r)!}. \tag{21}$$

The series S can be expressed as an integral of the form given on the right of Eq. (10), corresponding to a Regge symmetry of the integration formula for 3-j coefficients given in Eq. (1). We commence by using the beta function integral to represent two of the factorials in S as an integral, thus

$$(-)^r(j_1+j_3-M_2-r)!\,(j_2+M_2+r)! = (-)^{j_1+j_3-M_2}\frac{(j_1+j_2+j_3+1)!}{2^{j_1+j_2+j_3+1}}\int_{-1}^{1}(\mu-1)^{j_1+j_3-M_2-r}(\mu+1)^{j_2+M_2+r}\,d\mu,$$

and after the substitution the series S becomes

$$S = \frac{(-)^{j_1-j_2+2j_3+m_3}(j_1+j_2+j_3+1)!}{2^{j_1+j_2+j_3+1}(j_1-M_1)!\,(j_2-M_2)!}\int_{-1}^{1}(\mu-1)^{j_1-j_2+j_3}(\mu+1)^{-j_1+j_2+j_3}$$

$$\times\sum_r\binom{j_3+M_3}{r}\left[\left(\frac{d}{d\mu}\right)^r(\mu-1)^{j_2-M_2}\right]\left[\left(\frac{d}{d\mu}\right)^{j_3+M_3-r}(\mu+1)^{j_1-M_1}\right]d\mu.$$

Applying Leibnitz's theorem to the integrand we obtain

$$S = \frac{(-)^{j_1-j_2+2j_3+m_3}(j_1+j_2+j_3+1)!}{2^{j_1+j_2+j_3+1}(j_1-M_1)!\,(j_2-M_2)!}\times\int_{-1}^{1}(\mu-1)^{j_1-j_2+j_3}(\mu+1)^{-j_1+j_2+j_3}\left(\frac{d}{d\mu}\right)^{j_3+M_3}[(\mu-1)^{j_2-M_2}(\mu+1)^{j_1-M_1}]\,d\mu,$$

which is proportional to a 3-j coefficient by virtue of the Regge symmetry indicated in Eq. (10), and the integration formula of Eq. (1). It has therefore been shown directly that

$$\tfrac{1}{2}\int_{-1}^{1}d^{j_1}_{M_1m_1}d^{j_2}_{M_2m_2}d^{j_3}_{M_3m_3}\,d\mu = \begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & m_2 & m_3\end{pmatrix}\begin{pmatrix}j_1 & j_2 & j_3\\ M_1 & M_2 & M_3\end{pmatrix}. \tag{22}$$

APPENDIX: THE SCHENDEL IDENTITIES

The following identities, Eqs. (A1)–(A3), first given by Schendel,[3] are readily established by means of Leibnitz's theorem for the multiple differentiation of a product of two functions:

$$(j-M)!\,(\mu-1)^{M-m}(\mu+1)^{M+m}\left(\frac{d}{d\mu}\right)^{j+M}[(\mu-1)^{j+m}(\mu+1)^{j-m}]$$

$$= (j+M)!\left(\frac{d}{d\mu}\right)^{j-M}[(\mu-1)^{j-m}(\mu+1)^{j+m}] \tag{A1}$$

$$= (j+m)!\,(\mu-1)^{M-m}\left(\frac{d}{d\mu}\right)^{j-m}[(\mu-1)^{j-M}(\mu+1)^{j+M}] \tag{A2}$$

$$= (j-m)!\,(\mu+1)^{M+m}\left(\frac{d}{d\mu}\right)^{j+m}[(\mu-1)^{j+M}(\mu+1)^{j-M}]. \tag{A3}$$

*Part of this work was carried out while the author was on sabbatical leave at the School of Mathematics, The University, Newcastle upon Tyne, England.

[1]G. Racah, Phys. Rev. 62, 438 (1942).
[2]T. Regge, Nuovo Cimento 10, 544 (1958).
[3]L. Schendel, Crelle's Journal 82, 158 (1877).
[4]J.K. Percus and A. Rotenberg, J. Math. Phys. 3, 928 (1962).
[5]F. Sasaki and K. Ohno, J. Math. Phys. 4, 1140 (1963).
[6]V.H. Smith, Jr., J. Chem. Phys. 41, 277 (1964).
[7]V.H. Smith, Jr. and F.E. Harris, J. Math. Phys. 10, 771 (1969).
[8]K. Mano, J. Math. Phys. 12, 2361 (1971).
[9]K. Mano, J. Chem. Phys. 52, 2785 (1970).
[10]G.N. Watson, Proc. Math. Assoc. Glasgow 2, 57 (1953).
[11]W.N. Bailey, Proc. Math. Assoc. Glasgow 2, 62 (1953).
[12]D.M. Brink and G.R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1968), 2nd ed., p. 146.

Self-gravitating fluids with cylindrical symmetry

Dipankar Ray

Department of Physics, New York University, New York, New York 10003
(Received 14 October 1975)

In a recent paper [P. S. Letelier, J. Math. Phys. **16**, 1488 (1975)], the problem of self-gravitating fluid with cylindrical symmetry with $p = \rho c^2$ has been reduced to a single equation. The present note solves the equation and also rectifies an oversight in the above-mentioned paper.

1. INTRODUCTION

In a recent paper,[1] it was found that Einstein's field equations for a self-gravitating perfect fluid with pressure p, equal to rest energy ρ and 4-velocity u_μ is equivalent to the field equations

$$R_{\mu\nu} = -2\sigma_{,\mu}\sigma_{,\nu},$$
$$\Box\sigma = (\sqrt{-g}\,\sigma_{,\mu}g^{\mu\nu})_{,\nu}/\sqrt{-g} = 0, \tag{1}$$

when irrotationality is imposed, i.e.,

$$u_\mu = \sigma_{,\mu}/(\sigma_{,\nu}\sigma^{,\nu})^{1/2}, \tag{2}$$

where $c=1$ and $G = 1/(8\pi)$,

$$p = \rho = \sigma_{,\nu}\sigma^{,\nu} \tag{3}$$

$$T_{\mu\nu} = 2\sigma_{,\mu}\sigma_{,\nu} - g_{\mu\nu}\sigma_{,\alpha}\sigma^{,\alpha}. \tag{4}$$

Letelier[2] has tried to solve these equations for an axially symmetric metric

$$ds^2 = \exp[2(\omega-\lambda)](dt^2 - dr^2)$$
$$- \exp(2\mu)[r^2\exp(-2\lambda)\,d\theta^2 + \exp(2\lambda)\,dz^2] \tag{5}$$

where ω, λ, and μ are functions of r and t alone.

According to Letelier, Eq. (5) reduces the field equations for the above case to

$$\exp(2\mu) = [F(t-r) + G(t+r)]/r, \quad F, G \text{ arbitrary functions}, \tag{6}$$

$$W = \int 1/[(1/r + 2\mu_1)^2 - 4\mu_0^2]$$
$$\times\{[2\mu_0(f + \sigma_0^2 + \sigma_1^2) - (1/r + 2\mu_1)(\phi + 2\sigma_0\sigma_1)]\,dt$$
$$+ [2\mu_0(\phi + 2\sigma_0\sigma_1) - (1/r + 2\mu_1)(f + \sigma_0^2 + \sigma_1^2)]\,dr\}, \tag{7}$$

where

$$f = \mu_{00} + \mu_{11} + \mu_1/r + \mu_0^2 + \mu_1^2 + \lambda_0^2 + 2\mu_0\lambda_0 + 2\mu_1\lambda_1,$$
$$\phi = \mu_0/r + 2\mu_{10} + 2\mu_1\mu_0 + 2\lambda_1\lambda_0 + 2\lambda_0\mu_1 + 2\lambda_1\mu_0,$$
$$\lambda_{00} - \lambda_{11} + 2\mu_0\lambda_0 - 2\mu_1\lambda_1 - \lambda_1/r = 0, \tag{8}$$
$$\sigma_{00} - \sigma_{11} + 2\mu_0\sigma_0 - 2\mu_1\sigma_1 - \lambda_1/r = 0. \tag{9}$$

0 and 1 mean derivatives with respect to t and r respectively.

From (8) and (9) we note that λ and σ satisfy the same equation according to Letelier and so he says that if that equation can be solved, the field equations are solved for the case under consideration.

2. NECESSARY CORRECTION

Checking Leterlier's[2] Eq. (9) and (10), i.e.,

$$\mu_{11} - \mu_{00} + (3\mu_1 - \lambda_1)/r - \lambda_{11} + \lambda_{00} + 2(\mu_1^2 - \mu_0^2 + \mu_0\lambda_0 - \mu_1\lambda_1) = 0,$$

$$\mu_{11} - \mu_{00} + (\mu_1 + \lambda_1)/r + \lambda_{11} - \lambda_{00} + 2(\mu_1^2 - \mu_0^2 - \mu_0\lambda_0 + \mu_1\lambda_1) = 0, \tag{10}^3$$

one easily sees that (8) of the present note (which is (11b) of Letelier) should be replaced by

$$\lambda_{00} - \lambda_{11} + 2\mu_0\lambda_0 - 2\mu_1\lambda_1 + (\mu_1 - \lambda_1)/r = 0. \tag{8'}$$

λ thus does not satisfy the same differential equation as σ. However, if we put

$$\nu = \lambda - \tfrac{1}{2}\log r, \tag{10'}$$

then,

$$\nu_{00} - \nu_{11} + 2\mu_0\nu_0 - 2\mu_1\nu_1 - \nu_1/r = 0, \tag{9'}$$

which is of the same form as (9).

Thus it remains true that the solution of (9) leads to the solution of the Einstein equation for the case considered by Letelier.

3. SOLUTION

Obviously, Eq. (9) can be rewritten as

$$(re^{2\mu}\sigma_0)_0 = (re^{2\mu}\sigma_1)_1.$$

Therefore there exists a function χ such that

$$re^{2\mu}\sigma_0 = \chi_1, \quad re^{2\mu}\sigma_1 = \chi_0$$

or

$$(F+G)\frac{\partial\sigma}{\partial u} = -\frac{\partial\chi}{\partial u}, \quad (F+G)\frac{\partial\sigma}{\partial v} = \frac{\partial\chi}{\partial v}, \tag{11}$$

where

$$u = t - r, \quad v = t + r, \tag{12}$$

$$F = F(t-r) = F(u), \quad G = G(t+r) = G(v). \tag{13}$$

From (11) and (13),

$$(F+G)\frac{\partial\sigma}{\partial F} = -\frac{\partial\chi}{\partial F}, \quad (F+G)\frac{\partial\sigma}{\partial G} = \frac{\partial\chi}{\partial G},$$

if F and G are not constants

or

$$\xi\frac{\partial\sigma}{\partial\xi} = -\frac{\partial\chi}{\partial\eta}, \quad \xi\frac{\partial\sigma}{\partial\eta} = -\frac{\partial\chi}{\partial\xi}, \tag{14}$$

where

$$\xi = F + G, \quad \eta = F - G, \tag{15}$$

or from (14)

$$\frac{\partial}{\partial \xi}\left(\xi \frac{\partial \sigma}{\partial \xi}\right)=\frac{\partial}{\partial \eta}\left(\xi \frac{\partial \sigma}{\partial \eta}\right)$$

or

$$\frac{1}{\xi}\frac{\partial}{\partial \xi}\left(\xi \frac{\partial \sigma}{\partial \xi}\right)=\frac{\partial^2 \sigma}{\partial \eta^2}. \tag{16}$$

Equation (16) has the following solution:

$$\sigma=\sum_k (A_k \exp(k\eta)+B_k \exp(-k\eta))[C_k J_0(k\xi)+D_k N_0(k\xi)] + \sum_{k'}(A_k' \cos k'\eta + B_k' \sin k'\eta)[C_k' I_0(k'\xi)+D_k' K_0(k'\xi)], \tag{17}$$

where A_k, B_k, C_k, D_k, A_k', B_k', C_k', D_k', k, k' are constant, J_0 and N_0 are zeroth order Bessel functions of first and second kind respectively, and similarly I_0 and K_0 are two zeroth order modified Bessel functions. ξ and η are given by (15).

4. CONCLUSION

Proceeding with (9)′, we similarly get by using (10):

$$\lambda=\tfrac{1}{2}\log r+\sum_l (A_l \exp(l\eta)+B_l \exp(-l\eta))[C_l J_0(l\xi)+D_l N_0(l\xi)] + \sum_{l'}(A_l' \cos l'\eta + B_l' \sin l'\eta)[C_l' I_0(l'\xi)+D_l' K_0(l'\xi)], \tag{18}$$

where as before A_l, B_l, C_l, D_l, A_l', B_l', C_l', D_l', l, l' are constants of integration and J_0, N_0, I_0, K_0 are Bessel functions and modified Bessel functions as stated above.

Equations (7), (17), and (18) thus provide a complete set of solutions of the Einstein equations for the case under consideration (provided F and G and hence ξ and η are not constants).

[1]R. Tabensky and A.H. Taub, Commun. Math. Phys. 29, 61 (1973).

[2]P.S. Letelier, J. Math. Phys. 16, 1488 (1975).

[3]Equations (10) of the present note [which are Eqs. (9) and (10) of Letelier] also indicate that Letelier's Eq. (13), i.e.,

$$\mu_{00}-\mu_{11}-2\mu_1/r+(\mu_0^2-\mu_1^2)=0, \tag{i}$$

should be replaced by

$$\mu_{00}-\mu_{11}-2\mu_1/r+2(\mu_0^2-\mu_1^2)=0. \tag{ii}$$

This, however; seems to be a mere printing error, since one can easily check that (ii) and not (i) is consistent with (6) of the present note which is (14) of Letelier.

Eigenvalues of invariants of U(n) and SU(n)*

M. C. K. Aguilera-Navarro

Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

V. C. Aguilera-Navarro

Instituto de Física Teórica, São Paulo, Brazil

(Received 9 December 1975)

In this paper, we present four distinct ways of obtaining the eigenvalues of invariants of unitary groups, in any irreducible representation. The invariants are defined according to a different contraction convention. Their eigenvalues can be given in terms of special partial hooks different from those found by other authors.

I. INTRODUCTION

The eigenvalues of invariants of unitary groups have recently been discussed in the literature.[1-7] In this article, we discuss a different set of invariants and present four distinct ways to obtain their eigenvalues in any irreducible representation (irrep).

Usually, the invariants are defined by using "up—down" (UD) contractions of the generators indices, in a particular realization of the Lie algebra. Here, we shall make use of the "down—up" (DU) convention. There criteria are related to the way the indices of the generators are contracted to make up invariants and are discussed in Sec. II.

Several authors[1-7] have shown that the eigenvalues of the UD invariants can be given in terms of quantities—special partial hooks—that depend on the partition characterizing the irrep considered and the dimension n of the space on which the group transformations operate.

The DU criterion leads to similar results and the eigenvalues are given in terms of new quantities depending only on the irrep labels. Oddly enough, no explicit dependence on n appears. Those quantities are also special partial hooks.

In the next section, we define the DU invariants, and in Sec. III, we give their eigenvalues in any irrep.

Section IV contains a brief discussion of the representation of the eigenvalues in terms of power sums.

II. THE DOWN–UP INVARIANTS

The n^2 generators A_i^j of the unitary group U(n) can be realized in terms of creation η_i and annihilation ξ^j operators (boson or fermion type) as

$$A_i^j \equiv \eta_i \cdot \xi^j = \sum_{s=1}^{t} \eta_{is}\, \xi^{js}, \quad i, j = 1, 2, \ldots, n, \qquad (2.1)$$

where t is the dimension of the spaces containing the vectors η_i and ξ^j.

The generators satisfy the commutation relations

$$[A_i^j, A_k^l] = \delta_k^j A_i^l - \delta_i^l A_k^j. \qquad (2.2)$$

We can form a k-order invariant of U(n) by doing the following contractions:

$$C_k^{(n)} \equiv A_{i_1}^{i_k} A_{i_2}^{i_1} A_{i_3}^{i_2} \cdots A_{i_k}^{i_{k-1}}, \qquad (2.3)$$

where use is made of the usual convention of summing repeated indices over all their values. Let us refer to these invariants as DU invariants, and maintain the usual convention that $C_0^{(n)} = n$.

There are at most n independent invariants of U(n).[8] Here, we have choosen them as being $C_k^{(n)}$, $k = 1, 2, \ldots, n$, while other authors have considered those defined by the up—down criterion, namely,

$$\tilde{C}_k^{(n)} \equiv A_{i_k}^{i_1} A_{i_1}^{i_2} A_{i_2}^{i_3} \cdots A_{i_{k-1}}^{i_k}, \quad k = 1, 2, \cdots \qquad (2.4)$$

whose eigenvalues in any irrep of U(n) are given[1-6] in terms of the particular partial hooks

$$p_{in} = h_{in} + n - i, \qquad (2.5)$$

where h_{in} is the ith component of the partition $[h] \equiv [h_{1n} h_{2n} \cdots h_{nn}]$ characterizing the considered irrep of U(n).

In this paper, the invariants (2.4) will be named UD invariants. Indeed, the DU and UD invariants are two components of the symmetrized invariant

$$S_k^{(n)} \equiv \sum_{p \in S(k)} p C_k^{(n)}, \qquad (2.6)$$

i.e., (2.3) is the "first" term arising when $p = e$ = identity, and (2.4) is the term corresponding to some other permutation. In this sum, p is an element of the symmetric group S(k) whose effect on $C_k^{(n)}$ is the corresponding permutation of the A's, in (2.3).

No general formula for the eigenvalues of the symmetrized invariant (2.6) have been obtained up to now. Some kind of fortuitousness led authors to consider the same component of $S_k^{(n)}$ defined by (2.4), namely, the UD invariants. It seems that they were not aware of the existence of the DU invariants (2.3). The present authors have the feeling that the knowledge of the eigenvalues of both classes of invariants will pave the way to get the eigenvalue of $S_k^{(n)}$. This operator is considered more convenient due to its contraction-convention-free character. What rests to do in that direction is to examine the individual properties of the other terms appearing in (2.4).

III. THE EIGENVALUES OF THE DU INVARIANTS

Throughout this section, we shall apply extensively many of the ideas introduced in Refs. 1 and 3.

To get the eigenvalues of the DU invariants defined in (2.3), we can avoid any explicit reference to the basis for the irrep considered by noting that they can be obtained by means of a constructive procedure using, alternative and conveniently, the following relations:

$$C_1^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}) = h_{1n} + h_{2n} + \cdots + h_{nn}, \tag{3.1}$$

$$C_k^{(1)}(h_{11}) = h_{11}^k, \tag{3.2}$$

$$C_k^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}) = \sum_{l=0}^{k} \binom{k}{l} h_{nn}^{k-l} \times C_l^{(n)}(h_{1n}-h_{nn}, h_{2n}-h_{nn}, \ldots, h_{n-1n}-h_{nn}, 0), \tag{3.3}$$

$$C_k^{(n+1)}(h_{1n+1}, \ldots, h_{nn+1}, 0) = C_k^{(n)}(h_{1n+1}, h_{2n+1}, \ldots, h_{nn+1}) + \sum_{l=1}^{k-1} (-n)^{k-l-1} C_l^{(n)}(h_{1n+1}, h_{2n+1}, \ldots, h_{nn+1}), \quad k>1, \tag{3.4}$$

where $C_k^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn})$ is the eigenvalue of the k-order DU invariant $C_k^{(n)}$ in the irrep $[h_{1n} h_{2n} \cdots h_{nn}]$ of $U(n)$.

The relations (3.1) and (3.2) are well known, while (3.3) is a consequence of the result[9]

$$C_k^{(n)}(h_{1n}+\lambda, h_{2n}+\lambda, \ldots, h_{nn}+\lambda) = \sum_{l=0}^{k} \binom{k}{l} \lambda^{k-l} C_l^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}), \tag{3.5}$$

when we replace $h_{in}+\lambda$ by h_{in}, $i=1,2,\ldots,n$, and take $\lambda = h_{nn}$.

The last relation is a trivial particular case of a general relation between DU invariants of unitary groups found by the present authors.[9,10]

Indeed, (3.2) is contained in (3.1). However, here we consider them separately for computational conveniences.

Now, we shall present three closed formulas to get those eigenvalues.

First of all, let us consider the integers q_{in} defined by

$$q_{in} \equiv h_{in} + 1 - i, \tag{3.6}$$

which are also partial hooks of the type p_{i1} and the function

$$D^{(n)}(h) \equiv D^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}) \equiv \frac{1}{(n-1)\underline{!}} \prod_{i<j}^{n} (q_{in} - q_{jn}), \tag{3.7}$$

where

$$m\underline{!} \equiv m!(m-1)!(m-2)!\cdots 2!. \tag{3.8}$$

Since $q_{in} - q_{jn} = p_{in} - p_{jn}$, the function $D^{(n)}$ coincides with that of Ref. 3. When its arguments define a Young partition, $D^{(n)}(h) = \dim(h) =$ dimension of the irrep$[h]$ of $U(n)$.[11] In the following, we shall need some properties of that function which are similar to, but different from, those derived by Louck and Biedenharn.[3] The proofs follow from techniques very similar to the ones developed by those authors. The details are given in Ref. 9 which also contains details and proofs of several other results mentioned and used here.

The function $D^{(n)}$ defined in (3.7) satisfies the following relations:

$$D^{(n+1)}(h_{1n+1}, \ldots, h_{n+1\,n+1}) = \frac{1}{n!} D^{(n)}(h_{1\,n+1}, \ldots, h_{n\,n+1}) \times \prod_{l=1}^{n} (q_{l\,n+1} - q_{n+1\,n+1}), \tag{3.9a}$$

$$D^{(n+1)}(h_{1\,n+1}, \ldots, h_{i\,n+1}+1, \ldots, h_{n+1\,n+1}) = \frac{1}{n!} D^{(n)}(h_{1\,n+1}, \ldots, h_{i\,n+1}+1, \ldots, h_{n\,n+1}) \times [1 + 1/(q_{i\,n+1} - q_{n+1\,n+1})] \prod_{l=1}^{n} (q_{l\,n+1} - q_{n+1\,n+1}), \quad i=1,2,\ldots,n, \tag{3.9b}$$

$$D^{(n)}(h_{1n}, \ldots, h_{in}+1, \ldots, h_{nn}) = D^{(n)}(h) \prod_{l\neq i}^{n} [1 - 1/(q_{ln} - q_{in})], \tag{3.9c}$$

$$D^{(n)}(h+\lambda) \equiv D^{(n)}(h_{1n}+\lambda, h_{2n}+\lambda, \ldots, h_{nn}+\lambda) = D^{(n)}(h). \tag{3.9d}$$

Using these properties, it can be shown[9] that the rhs of

$$C_k^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}) = \sum_{i=1}^{n} q_{in}^k D^{(n)}(h_{1n}, \ldots, h_{in}+1, \ldots, h_{nn}) / D^{(n)}(h), \quad k=1,2,\ldots, \tag{3.10}$$

satisfies the relations (3.1) to (3.4), so that, (3.10) is a closed formula to obtain the eigenvalues of the DU invariants (2.3).

Another closed relation to get the eigenvalues of (2.3) is given by[9]

$$C_k^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn}) = \sum_{l=0}^{k} (-1)^l \varphi_l(q) \sum_{m=0}^{k-l} \binom{n-l}{k-l-m+1} \beta_m(q), \quad k=1,2,\ldots, \tag{3.11}$$

where q stands for the ordered set $(q_{1n}, q_{2n}, \ldots, q_{nn})$ made up of the quantities q_{in} defined in (3.6), and

$$\beta_m(q) \equiv \sum_{\text{all } \alpha_i \geq 0}{}' \frac{(-)^{m-\alpha}\, \alpha!}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} \varphi_1^{\alpha_1}(q) \varphi_2^{\alpha_2}(q) \cdots \varphi_n^{\alpha_n}(q). \tag{3.12}$$

The prime in the sum symbol means that the α_i are restricted to values such that $\alpha_1 + 2\alpha_2 + \cdots + n\alpha_n = m$. In (3.12), $\varphi_1, \varphi_2, \ldots, \varphi_n$ are the elementary symmetric functions

$$\begin{aligned} \varphi_1(x) &= \sum x_i = x_1 + x_2 + \cdots + x_n, \\ \varphi_2(x) &= \sum_{i<j} x_i x_j = x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n, \\ &\vdots \\ \varphi_n(x) &= \Pi x_i = x_1 x_2 \cdots x_n. \end{aligned} \tag{3.13}$$

Formula (3.11) is similar to the corresponding one derived by Louck and Biedenharn[3] for the UD invariant (2.4). We note that for the DU invariants (2.3), it was possible to factorize φ_l, thus simplifying the calculation of the eigenvalues.

Finally, another closed representation for the eigenvalues of $C_k^{(n)}$ can be obtained from a single matrix a, following an idea introduced by Perelomov and Popov,[1] who also used the UD criterion. For the DU invariant (2.3) let us consider the $n\times n$ matrix $a = a^{(n)}(h_{1n}, h_{2n}, \ldots, h_{nn})$, whose elements are given by

$$a_{ij}=q_{in}\delta_{ij}+t_{ij}, \quad i,j=1,2,\ldots,n, \tag{3.14}$$

where q_{in} was defined in (3.6), and t is the triangular matrix

$$t_{ij}=\begin{cases}1, & \text{if } i>j,\\ 0, & \text{otherwise.}\end{cases} \tag{3.15}$$

In terms of the matrix a, we simply have that

$$C_k^{(n)}(h_{1n},h_{2n},\ldots,h_{nn})=\sum_{i,j=1}^{n}(a^k)_{ij}, \tag{3.16}$$

i.e., the eigenvalue of the k-order DU invariant (2.3) in the irrep$[h_{1n}h_{2n}\cdots h_{nn}]$ of $U(n)$ is given by the sum of all matrix elements of the power k of the matrix (3.14).

To prove the last statement, we show[9] that the rhs of (3.16) fulfills all the relations (3.1)–(3.4).

A sketch of the proof follows.

It is easy to see that the rows and columns of a satisfy the following relations:

$$\sum_{i=1}^{n}a_{ij}=q_{jn}+n-j, \quad j=1,2,\ldots,n \tag{3.17}$$

and

$$\sum_{j=1}^{n}a_{ij}=h_{in}, \quad i=1,2,\ldots,n. \tag{3.18}$$

The relation (3.1) follows immediately from (3.17) or (3.18), while (3.2) follows from the definition (3.14), and it is not difficult to prove (3.3). What is not so simple is to show that (3.4) is also verified. For this purpose, it is convenient to decompose $a^{(n+1)}$ as the sum of two matrices, b and c, where b is the direct sum of the submatrix $a^{(n)}$ with the 1×1 null matrix, i.e.,

$$b=a^{(n)}(h_{1\,n+1},h_{2\,n+1},\ldots,h_{n\,n+1})\oplus 0, \tag{3.19}$$

and c is also an $(n+1)\times(n+1)$ matrix whose only nonzero row is the last one, namely,

$$c=\begin{pmatrix}0 & 0\cdots & 0 & 0\\ 0 & 0\cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0\cdots & 0 & 0\\ 1 & 1\cdots & 1 & -n\end{pmatrix}. \tag{3.20}$$

We recall that, in the present case $q_{n+1\,n+1}=-n$. Such matrices have the following properties:

$$bc=0, \tag{3.21a}$$

$$c^l=(-n)^{l-1}c, \quad l=1,2,\ldots, \tag{3.21b}$$

$$b^k=[a^{(n)}]^k\oplus 0, \tag{3.21c}$$

$$(cb^l)_{ij}=\begin{cases}\sum_{m=1}^{n}[(a^{(n)})^l]_{mj}\,\delta_{in}, & j\le n,\\ 0, & j=n+1,\end{cases} \tag{3.21d}$$

which allows us to complete the proof that (3.16) gives the eigenvalue of the DU invariant of U(n).

IV. THE POWER SUMS

Let us define power sums S_k by

$$S_k\equiv\sum_{i=1}^{n}q_{in}^k. \tag{4.1}$$

In terms of these sums, we can see that

$$C_1^{(n)}(h_{1n},\ldots,h_{nn})=\binom{n}{2}+S_1, \tag{4.2a}$$

$$C_2^{(n)}(h_{1n},\ldots,h_{nn})=\binom{n}{3}+(n-1)S_1+S_2, \tag{4.2b}$$

$$C_3^{(n)}(h_{1n},\ldots,h_{nn})=\binom{n}{4}+\binom{n-1}{2}S_1+\tfrac{1}{2}(2n-3)S_2+\tfrac{1}{2}S_1^2+S_3, \tag{4.2c}$$

$$C_4^{(n)}(h_{1n},\ldots,h_{nn})=\binom{n}{5}+\binom{n-1}{3}S_1+S_1S_2+\tfrac{1}{2}(n-2)S_1^2+\tfrac{1}{2}(n-2)^2S_2+(n-2)S_3+S_4, \tag{4.2d}$$

etc.

These expressions are as complicated as those obtained within the UD-criterion. The corresponding expressions found by Perelomov and Popov[1] are affected by mistakes. The correct ones were found by Louck and Biedenharn.[3] There is no closed formula for the eigenvalues in terms of those power sums and, in this sense, they do not constitute a convenient basis to express the eigenvalues of either $C_k^{(n)}$ or $\tilde{C}_k^{(n)}$.

Comparing the expressions for the UD invariants obtained by Louck and Biedenharn[3] with Eqs. (4.2), we note a sporadic change of sign among the terms. However, there is a deeper difference hidden by the notation involved. In their case, the power sums are defined in terms of powers of the quantities (2.5) instead of (3.6) as we did here, i.e., they define

$$\tilde{S}_k\equiv\sum_{i=1}^{n}p_{in}^k. \tag{4.3}$$

The power sums $\tilde{S}_k$ are related to the present ones through the expression

$$\tilde{S}_k=\sum_{l=0}^{k}\binom{k}{l}(n-1)^{k-l}S_l, \tag{4.4}$$

which can be obtained by noting that $p_{in}=q_{in}+n-1$.

Finally, for the sake of completeness, we recall that the DU-invariants $\overline{C}_k^{(n)}$ of SU(n) can be obtained from those of U(n) by means of the relation[1]

$$\overline{C}_k^{(n)}=\sum_{l=0}^{k}\binom{k}{l}\left(-\frac{h}{n}\right)^{k-l}C_l^{(n)}, \quad k=1,2,\ldots, \tag{4.5}$$

where

$$h=h_{1n}+h_{2n}+\cdots+h_{n-1n}, \quad (h_{nn}=0).$$

The relation (4.5) arises from the "traceless" condition imposed on the generators (2.1) in order to make them generators of SU(n). It is independent of the choice of the DU or UD contraction criterion.

ACKNOWLEDGMENT

The authors express their gratitude to the referee for valuable comments and for calling their attention to Refs. 4, 5, and 6.

*Work partially supported by FINEP, Brazil.

[1]A. M. Perelomov and V. S. Popov, Yad. Fiz. 3, 924 (1966) [Sov. J. Nucl. Phys. 3, 676 (1966)].
[2]J. D. Louck, Am. J. Phys. 38, 3 (1970).

[3]J. D. Louck and L. C. Biedenharn, J. Math. Phys. 11, 2368 (1970).
[4]R. L. Hudson, J. Math. Phys. 15, 1067 (1974).
[5]S. Okubo, J. Math. Phys. 16, 528 (1975).
[6]M. K. F. Wong and Hsin-Yang Yeh, J. Math. Phys. 16, 1239 (1975).
[7]Our list of references is not exhaustive. A better one is presented in Ref. 6.
[8]L. C. Biedenharn, J. Math. Phys. 4, 436 (1963).
[9]M. C. K. Aguilera-Navarro and V. C. Aguilera-Navarro, to appear in An. Acad. Bras. Ciênc.
[10]M. C. K. Aguilera-Navarro and V. C. Aguilera-Navarro, J. Math. Phys. 16, 72 (1975).
[11]N. Jacobson, *Lie Algebras* (Interscience, New York, 1966).

Fundamental canonical realizations of connected Lie groups

Luis Martinez Alonso

Departamento de Física Teorica, Universidad Complutense de Madrid, Madrid, Spain
(Received 19 December 1975)

An important class of transitive canonical realizations of connected Lie groups is studied by means of a general formalism. We give a simple method for the classification and the construction of these realizations.

1. INTRODUCTION

The theory we are going to describe in this paper is to answer the following question: Given a connected Lie group $\mathcal{G}$, to what extent can one find all possible transitive canonical realizations of $\mathcal{G}$?

With a notable exception[1] the consistent theory of canonical realizations of Lie groups has been largely ignored since the discovery of quantum mechanics. This is deplorable indeed. It is well-known[2] that this theory provides a most natural framework for the characterization of the elementary systems in classical mechanics. In this paper we intend to show that it is possible to complete this unfinished work and to put it in its proper perspective.

Section 2 reviews the basic concepts of canonical transformations. In Sec. 3 we reduce the problem to the case of $\mathcal{G}$ simply connected. Sections 4 and 5 review the projective covering group and the co-adjoint action of a connected Lie group. Section 6 is devoted to the construction of a method of finding all the fundamental canonical realizations of a connected Lie group. In Sec. 7 we consider the application of the method to the Gallilei group. The physical interpretation of some of the mathematical terms used here is given in the Appendix.

2. SOME BASIC NOTIONS

Let V and V' be symplectic manifolds,[3] with Poisson brackets $\{\ ,\ \}$, and $\{\ ,\ \}'$ respectively. By a canonical map on V to V' we shall mean a C^∞ map τ: $V \to V'$ such that:

$$\{f\circ\tau,\ h\circ\tau\}=\{f,h\}'\circ\tau,\ \forall\ f,\ h\in C^\infty(V'). \tag{1}$$

If τ is also a diffeomorphism[3] of V onto V' we shall say that τ is a canonical transformation.

We shall deal throughout with connected Lie groups $\mathcal{G}$. By a canonical realization $(r, \mathcal{G}, V)$ of such a group $\mathcal{G}$ on a symplectic manifold V we shall mean a C^∞ map

$$\mathcal{G}\times V\to V,\quad (g,x)\to r(g)x,$$

which has the following properties:

(i) $r(g_1g_1)x=r(g_1)r(g_2)x,$ (2a)

(ii) $r(e)x=x,$ (2b)

(iii) $r(g)$ is a canonical transformation for all g in $\mathcal{G}$. (2c)

We shall say that $(r, \mathcal{G}, V)$ is transitive if $\forall\, x, y\in V$, $\exists\, g\in\mathcal{G}/r(g)x=y$.

Let $(r_i, \mathcal{G}, V)$, $i=1,2$, be canonical realizations of $\mathcal{G}$. If a canonical transformation τ of V_1 onto V_2 such that

$$r_2(g)\circ\tau=\tau\circ r_1(g),\ \forall\ g\in\mathcal{G} \tag{3}$$

exits, the two realizations are said to be equivalent.

Let G be the Lie algebra of $\mathcal{G}$ and $(r, \mathcal{G}, V)$ a canonical realization of $\mathcal{G}$. Each $A\in G$ defines a contravariant vector field $r(A)$ on V by

$$(r(A)f)(x)\equiv\frac{d}{dt}f([\exp(-tA)]x)\bigg|_{t=0},\quad f\in C^\infty(V).$$

This vector field is locally Hamiltonian,[3,4] that is, for each $x_0\in V$ there is a neighborhood N of x_0 and $a\in C^\infty(N)$ such that

$$(r(A)f)(x)=\{a,f\}(x),\ \forall\ f\in C^\infty(N), \forall\ x\in N.$$

A canonical realization $(r, \mathcal{G}, V)$ is said to be a Hamiltonian realization if

$$\forall\ A\in G\ \exists\, a\in C^\infty(V)/r(A)f=\{a,f\},\ \forall\ f\in C^\infty(V). \tag{4}$$

For each $A\in G$ the function $a\in C^\infty(V)$ is unique up to additive constants. If $B\equiv\{A_\alpha\colon \alpha=1,\ldots,n\}$ is a basis of G with commutation relations

$$[A_\alpha,A_\beta]=\sum_\nu c^\nu_{\alpha\beta}A_\nu, \tag{5}$$

then the associated functions $\{a_\alpha=a_\alpha(x)\colon \alpha=1,\ldots,n\}$ verify[3]

$$\{a_\alpha,a_\beta\}=\sum_\nu c^\nu_{\alpha\beta}a_\nu+\eta(A_\alpha,A_\beta), \tag{6}$$

where the $\eta(A_\alpha,A_\beta)$ are constants that define an equivalence class of infinitesimal exponents of the Lie algebra G of $\mathcal{G}$.

If $(r, \mathcal{G}, V)$ is a transitive Hamiltonian realization such that the map $A\in G\to a\in C^\infty(V)$ verifies the following property:

$$a(x)=a(x'),\ \forall\ A\in G\Rightarrow x=x', \tag{7}$$

we shall say that $(r, \mathcal{G}, V)$ is a fundamental canonical realization (f. c. r.) of $\mathcal{G}$. It is our main purpose to present a general method of finding all the f. c. r. of a connected Lie group $\mathcal{G}$. The first step is to reduce the problem to the case of connected and simply connected Lie groups.

3. THE REDUCTION TO COVERING GROUP

Let $\mathcal{G}$ be a connected Lie group with simply connected covering group $\tilde{\mathcal{G}}$ and covering homomorphism p. It is well-known[5] that the kernel Ker$p$, is a discrete central subgroup of $\tilde{\mathcal{G}}$.

Lemma 1: Let $(\tilde{r}, \tilde{\mathcal{G}}, V)$ be a f.c.r. of $\tilde{\mathcal{G}}$ such that Ker$p$ acts trivially. Then the map:

$$\mathcal{G}\times V \to V, \quad (g,x) \to \tilde{r}(p^{-1}(g))x \tag{8}$$

defines a f.c.r. $(\tilde{r}\circ p^{-1}, \mathcal{G}, V)$ of $\mathcal{G}$.

Proof: The trivial action of Ker$p$ makes sure that $(\tilde{r}\circ p^{-1}, \mathcal{G}, V)$ verifies (2a) and (2b). On the other hand it is known[5] that there exists a neighborhood U of the identity element in $\mathcal{G}$ such that p^{-1} is a diffeomorphism of U onto $p^{-1}(U)$, then $(g,x)\in U\times V \to \tilde{r}(p^{-1}(g))x\in V$ is a C^∞ map.

For each $g_0 \in \mathcal{G}$ the set g_0U is a neighborhood of g_0, and the map $(g,x)\in g_0U\times V \to \tilde{r}(p^{-1}(g))x \in V$ is the composition of the following C^∞ maps:

$$g_0U\times V \to U\times V \to V \to V,$$

$$(g,x) \to (g_0^{-1}g, x) \to r(g_0^{-1}g)x \to r(g_0)r(g_0^{-1}g)x,$$

where $r \equiv \tilde{r}\circ p^{-1}$. Then (8) is a C^∞ map. Hence $(\tilde{r}\circ p^{-1}, \mathcal{G}, V)$ is a transitive canonical realization of $\mathcal{G}$. Let $\tilde{G}$, G be the Lie algebras of $\tilde{\mathcal{G}}$ and $\mathcal{G}$ respectively; since p: $\tilde{\mathcal{G}} \to \mathcal{G}$ defines a Lie algebra isomorphism, $(\tilde{r}\circ p^{-1}, \mathcal{G}, V)$ is a f.c.r. of $\mathcal{G}$ if $(\tilde{r}, \tilde{\mathcal{G}}, V)$ is a f.c.r. of $\tilde{\mathcal{G}}$. QED

If $(r, \mathcal{G}, V)$ is a f.c.r. of $\mathcal{G}$ it is easy to prove that the map

$$\tilde{\mathcal{G}}\times V \to V, \quad (\tilde{g},x) \to r(p(g))x$$

defines a f.c.r. $(r\circ p, \tilde{\mathcal{G}}, V)$ of $\tilde{\mathcal{G}}$ such that Ker$p$ acts trivially. Therefore the determination of the f.c.r. of $\mathcal{G}$ is equivalent to the determination of the f.c.r. of $\tilde{\mathcal{G}}$ such that Ker$p$ acts trivially.

4. THE PROJECTIVE COVERING GROUP OF A CONNECTED LIE GROUP[6]

Let $\mathcal{G}$ be a connected Lie group with Lie algebra G. Let $H_0^2(G, \mathbb{R})$ be the second cohomology of G and $\mathbb{R}$ relative to the trivial action of G on $\mathbb{R}$, and let $\{\eta_1, \ldots, \eta_r\}$ be a set of infinitesimal exponents of G such that the associated cohomology classes are a basis of $H_0^2(G, \mathbb{R})$. We define on the linear space $\mathbb{R}^r \oplus G$ the following product law:

$$[(\theta_1, \ldots, \theta_r; A), (\xi_1, \ldots, \xi_r; A')]$$

$$\equiv (\eta_1(A, A'), \ldots, \eta_r(A,A'); [A, A']).$$

With this law: $\hat{G} \equiv \mathbb{R}^r \oplus G$ is a Lie algebra. Let $B \equiv \{A_\alpha : \alpha = 1, \ldots, n\}$ be a basis of G with commutation relations

$$[A_\alpha, A_\beta] = \sum_\nu c^\nu_{\alpha\beta} A_\nu .$$

We shall denote $M_1 \equiv (1, 0, \ldots, 0; 0), \ldots, M_r \equiv (0, \ldots, 0, 1; 0)$, and $A_\alpha \equiv (0, \ldots, 0; A_\alpha)$. The set $\hat{B} \equiv \{M_i, A_\alpha : i = 1, \ldots, r;\ \alpha = 1, \ldots, n\}$ is a basis of $\hat{G}$ with commutation relations

$$\begin{aligned} &[A_\alpha, A_\beta] = \sum_\nu c^\nu_{\alpha\beta} A_\nu + \sum_i \eta_i(A_\alpha, A_\beta) M_i, \\ &[A_\alpha, M_i] = [M_i, M_j] = 0. \end{aligned} \tag{9}$$

The map j: $(\theta_1, \ldots, \theta_r; A) \in \hat{G} \to A \in G$ is a homomorphism of $\hat{G}$ onto G and its kernel K is the central Lie subalgebra in $\hat{G}$ generated by the set $\{M_i : i = 1, \ldots, r\}$. Then we have the following exact sequence of Lie algebras:

$$0 \to K \xrightarrow{i} \hat{G} \xrightarrow{j} G \to 0, \tag{10}$$

where i: $K \to \hat{G}$ is the inclusion map. This exact sequence defines a central extension of G by K.

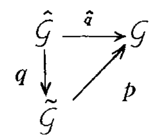
If $\hat{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are the unique (up to isomorphism) connected and simply connected Lie groups with Lie algebras $\hat{G}$ and G respectively, it is well-known that there exists a unique central extension of $\tilde{\mathcal{G}}$ by $\mathbb{R}^r$,

$$0 \to \mathbb{R}^r \xrightarrow{t} \hat{\mathcal{G}} \xrightarrow{q} \tilde{\mathcal{G}} \to 0, \tag{11}$$

such that the Lie algebras homorphism associated to t and q are equal to i and j respectively. Moreover $q: \hat{\mathcal{G}} \to \tilde{\mathcal{G}}$ is a C^∞ homomorphism of $\hat{\mathcal{G}}$ onto $\tilde{\mathcal{G}}$ with central kernel in $\hat{\mathcal{G}}$. A result due to Hochschild[7] shows the existence of a C^∞ section, i.e., a C^∞ map c: $\tilde{\mathcal{G}} \to \hat{\mathcal{G}}$ such that

$$q(c(\tilde{g})) = \tilde{g}, \ \forall\ \tilde{g} \in \tilde{\mathcal{G}}. \tag{12}$$

The group $\tilde{\mathcal{G}}$ is the universal covering group of $\mathcal{G}$, and $\hat{\mathcal{G}}$ is called the projective covering group of $\mathcal{G}$. If $p: \tilde{\mathcal{G}} \to \mathcal{G}$ is the covering homorphism of $\tilde{\mathcal{G}}$ onto $\mathcal{G}$, we define $\hat{q}: \hat{\mathcal{G}} \to \mathcal{G}$ by the composition



(13)

Evidently $\hat{q}$ is a C^∞ homomorphism of $\hat{\mathcal{G}}$ onto $\mathcal{G}$ and Ker$\hat{q} = q^{-1}$(Ker$p$). We shall say that q and $\hat{q}$ are the projective covering homomorphisms of $\hat{\mathcal{G}}$ onto $\tilde{\mathcal{G}}$ and $\mathcal{G}$ respectively.

5. THE CO-ADJOINT ACTION

Let G^* be the dual space of G. We define the co-adjoint action of $\mathcal{G}$ on G^* by the formula

$$\langle \mathrm{cad}g(A^*), B\rangle \equiv \langle A^*, \mathrm{ad}g^{-1}(B)\rangle, \quad g\in\mathcal{G},\ A^*\in G^*,\ B\in G. \tag{14}$$

This action is linear. Moreover, if $(\mathrm{ad}g)$ is the associated matrix with ad$g$ $(g \in \mathcal{G})$ in a basis $B = \{A_\alpha : \alpha = 1, \ldots, n\}$ of G and $(\mathrm{cad}g)$ is the corresponding matrix to cad$g$ $(g\in\mathcal{G})$ in the dual basis $B^* = \{A^*_\alpha : \alpha = 1, \ldots, n\}$ of B (i.e., $\langle A^*_\alpha, A_\beta\rangle = \delta_{\alpha\beta}$),

$$(\mathrm{cad}g)_{\alpha\beta} = ((\mathrm{ad}g)^{-1})_{\beta\alpha}. \tag{15}$$

The co-adjoint action of $\mathcal{G}$ on G^* is generated by a linear action of G on G^* given by

$$\langle \mathrm{cad}A(B^*), C\rangle \equiv \langle B^*, [C, A]\rangle, \quad A, C \in G, \ B^* \in G^*. \tag{16}$$

If $\{c^\nu_{\alpha\beta} : \alpha, \beta, \nu = 1, \ldots, n\}$ are the structure constants of G in the basis B and $(\mathrm{cad}A_\alpha)$ is the associated matrix of cadA_α $(A_\alpha \in B)$ in the dual basis B^*, then

$$(\mathrm{cad}A_\alpha)_{\nu\beta} = c^\beta_{\nu\alpha}. \tag{17}$$

We denote by $a \equiv (a_1, \ldots, a_n) \in \mathbb{R}^n$ the element $\sum_\alpha a_\alpha A^*_\alpha \in G^*$. Each $A \in G$ defines a contravariant vector field cad(A) on G^* by

$$(\mathrm{cad}(A)f)(a) \equiv \frac{d}{dt} f([\exp(-t\,\mathrm{cad}(A)]a)\Big|_{t=0}, \quad f \in C^\infty(G^*).$$

In particular, using the matrix expression (17) we find

$$(\mathrm{cad}(A_\alpha)f)(a)=\sum_{\beta,\nu} c^\nu_{\alpha\beta}a_\nu \frac{\partial f}{\partial a_\beta}. \quad (18)$$

Let $\mathcal{O}$ be a orbit in G^* under the co-adjoint action of $\mathcal{G}$. It is well-known[5] that $\mathcal{O}$ admits an unique structure of the C^∞ manifold such that the transitive action of $\mathcal{G}$ on $\mathcal{O}$ can be C^∞. With this structure $\mathcal{O}$ is a C^∞ submanifold of G^*. Let Ω be an open set such that $\mathcal{O}\subset\Omega$. Given a function $f:\Omega\to\mathbb{C}$ we shall denote by $\tilde f$ its restriction on $\mathcal{O}$.

Let $\mathcal{O}$ be a nontrivial orbit ($\dim\mathcal{O}\neq 0$) in G^* and $\{a_\alpha:\ \alpha=1,\dots,n\}$ the coordinate functions on G^* associated to the dual basis B^*. We now summarize the main properties of $\mathcal{O}$ in the following theorem.[8,9,10]

Theorem 1: (i) The orbit $\mathcal{O}$ has a structure of a symplectic manifold. If f_1 and f_2 are C^∞ functions over an open set $\Omega\subset G^*$ such that $\mathcal{O}\subseteq\Omega$ then

$$\{\tilde f_1,\tilde f_2\}=\widetilde{\sum_{\alpha,\beta,\nu} c^\nu_{\alpha\beta}\left(a_\nu \frac{\partial f_1}{\partial a_\alpha}\frac{\partial f_2}{\partial a_\beta}\right)}, \quad (19)$$

where $\{\ ,\ \}$ is the Poisson bracket of $\mathcal{O}$.

(ii) The co-adjoint action of $\mathcal{G}$ on $\mathcal{O}$ is a Hamiltonian realization of $\mathcal{G}$. The associated action of the Lie algebra G on $C^\infty(\mathcal{O})$ is given by

$$\mathrm{cad}(A_\alpha)f=\{\tilde a_\alpha, f\},\quad f\in C^\infty(\mathcal{O}), \quad (20)$$

where $\tilde a_\alpha (\alpha=1,\dots,n)$ are the restrictions on $\mathcal{O}$ of the coordinate functions associated to the dual basis B^*.

Note that (cad, $\mathcal{G},\mathcal{O}$) is a f.c.r. of $\mathcal{G}$.

6. THE CONSTRUCTION OF THE FUNDAMENTAL REALIZATIONS

Let $\tilde{\mathcal{G}}$ be the covering group of $\mathcal{G}$. We shall now describe a process defining a f.c.r. of $\tilde{\mathcal{G}}$.

Let (cad, $\hat{\mathcal{G}},\mathcal{O}$) be the f.c.r. of the projective covering group $\hat{\mathcal{G}}$ of $\mathcal{G}$ defined by a nontrivial orbit $\mathcal{O}$ in $\hat G^*$ under the co-adjoint action of $\hat{\mathcal{G}}$. We may construct the following action of $\tilde{\mathcal{G}}$ on $\mathcal{O}$:

$$\tilde{\mathcal{G}}\times\mathcal{O}\to\mathcal{O},\quad (\tilde g, a)\to \mathrm{cad}q^{-1}(\tilde g)a, \quad (21)$$

where q is the projective covering homomorphism of $\hat{\mathcal{G}}$ onto $\tilde{\mathcal{G}}$.

Lemma 2: Every f.c.r. (cad, $\hat{\mathcal{G}},\mathcal{O}$) of $\hat{\mathcal{G}}$ defines a f.c.r. (cad$\circ q^{-1}$, $\tilde{\mathcal{G}},\mathcal{O}$) of $\tilde{\mathcal{G}}$.

Proof: Since Ker$q$ is a central subgroup of $\hat{\mathcal{G}}$, $\tilde r \equiv \mathrm{cad}\circ q^{-1}$ is well defined. Moreover, if $c:\tilde{\mathcal{G}}\to\hat{\mathcal{G}}$ is any C^∞ section, as in Eq. (12), $\tilde r(\tilde g)=\mathrm{cad}c(\tilde g)$. Hence (21) is a C^∞ map. It is now straightforward to show that $(\tilde r,\tilde{\mathcal{G}},\mathcal{O})$ is a transitive Hamiltonian realization of $\tilde{\mathcal{G}}$. On the other hand, it is easy to prove that $(\tilde r,\tilde{\mathcal{G}},\mathcal{O})$ verifies (7). Consequently (21) is a f.c.r. of $\tilde{\mathcal{G}}$. QED

Now, our aim will be to show that the f.c.r. of $\tilde{\mathcal{G}}$ described by (21) are (up to equivalence) all the f.c.r. of $\tilde{\mathcal{G}}$.

Let $(\tilde r,\tilde{\mathcal{G}},V)$ be a f.c.r. of $\tilde{\mathcal{G}}$. We may define a f.c.r. $(\hat r\circ q,\hat{\mathcal{G}},V)$ of $\hat{\mathcal{G}}$ by

$$\hat{\mathcal{G}}\times V\to V,\ (\hat g,x)\to\tilde r(q(\hat g))x. \quad (22)$$

If $\{a_\alpha=a_\alpha(x):\ \alpha=1,\dots,n\}$ are the generators of $(\tilde r,\tilde{\mathcal{G}},V)$, there is a unique (up to equivalence) infinitesimal exponent η of the Lie algebra G of $\tilde{\mathcal{G}}$ such that

$$\{a_\alpha,a_\beta\}_V=\sum_\nu c^\nu_{\alpha\beta}a_\nu+\eta(A_\alpha,A_\beta). \quad (23)$$

Given a maximal independent set of infinitesimal exponents η_i $(i=1,\dots,r)$ of G, we may pick out the functions $a_\alpha(\alpha=1,\dots,n)$ such that there exits a linear combination $\eta=\sum_i m_i\eta_i$. Then

$$\{a_\alpha,a_\beta\}_V=\sum_\nu c^\nu_{\alpha\beta}a_\nu+\sum_i\eta_i(A_\alpha,A_\beta)m_i. \quad (24)$$

Since $(\tilde r,\tilde{\mathcal{G}},V)$ is a f.c.r. of $\tilde{\mathcal{G}}$, the map

$$x\in V\to(a_1(x),\dots,a_n(x))\in\mathbb{R}^n$$

is one to one and C^∞ over V. Hence the map

$$x\in V\to\tau(x)\equiv\sum_i m_iM_i^*+\sum_\alpha a_\alpha(x)A^*_\alpha\in\hat G^*, \quad (25)$$

where $\{M_i^*,A^*_\alpha: i=1,\dots,r;\ a=1,\dots,n\}$ is the dual basis of $\hat G^*$, is also one to one and C^∞ over V.

The following propositions connect the f.c.r. $(\hat r,\hat{\mathcal{G}},V)$ of $\hat{\mathcal{G}}$ given by (22) with the co-adjoint action.

Proposition 1: $(\mathrm{cad}\hat g)\circ\tau=\tau\circ\tilde r(\hat g),\ \forall\ \hat g\in\hat{\mathcal{G}}$.

Proof: Since $\hat{\mathcal{G}}$ is a connected Lie group, it is sufficient[5] to prove

$$\hat r(A_\alpha)(f\circ\tau)=(\mathrm{cad}(A_\alpha)f)\circ\tau,\ \forall\ f\in C^\infty(\hat G^*).$$

From (14), (15), and (24) we have

$$\hat r(A_\alpha)(f\circ\tau)=\sum_\beta\frac{\partial f\circ\tau}{\partial a_\beta}\{a_\alpha,a_\beta\}_V$$

$$=\left(\sum_\beta\left(\sum_\nu c^\nu_{\alpha\beta}a_\nu+\sum_i\eta_i(A_\alpha,A_\beta)m_i\right)\frac{\partial f}{\partial a_\beta}\right)\circ\tau$$

$$=(\mathrm{cad}(A_\alpha)f)\circ\tau.$$ QED

From this, it follows that the image of V under τ is an orbit $\mathcal{O}$ in $\hat G^*$ under the co-adjoint action of $\hat{\mathcal{G}}$, and the map τ is a diffeomorphism of V onto $\mathcal{O}$.

Proposition 2: $\tau: V\to\mathcal{O}$ is a canonical transformation.

Proof: Since $\mathcal{O}$ is a submanifold of $\hat G^*$, any C^∞ mapping defined on $\mathcal{O}$ is, locally on $\mathcal{O}$, the restriction of a C^∞ mapping on $\hat G^*$. Then, we need only to prove the equality

$$\{\tilde f\circ\tau,\ \tilde h\circ\tau\}_V=\{\tilde f,\tilde h\}_{\mathcal{O}}\circ\tau,$$

where $\tilde f$ and $\tilde h$ are restrictions on $\mathcal{O}$ of any C^∞ functions f and h over any open set Ω such that $\mathcal{O}\subset\Omega$. In this case

$$\{\tilde f\circ\tau,\ \tilde h\circ\tau\}_V(x)=\sum_{\alpha,\beta}\frac{\widetilde{\partial f}}{\partial a_\alpha}(\tau(x))\frac{\widetilde{\partial h}}{\partial a_\beta}(\tau(x))\{a_\alpha,a_\beta\}_V(x).$$

On the other hand,

$$\{\tilde f,\tilde h\}=\sum_{\alpha,\beta}\left(\sum_\nu c^\nu_{\alpha\beta}\tilde a_\nu+\sum_i\eta_i(A_\alpha,A_\beta)\tilde m_i\right)\frac{\widetilde{\partial f}}{\partial a_\alpha}\ \frac{\widetilde{\partial h}}{\partial a_\beta}.$$

From (24) and (25) follows the conclusion. QED

An immediate consequence of Propositions 1 and 2 is that $(\tilde r\circ q,\hat{\mathcal{G}},V)$ and (cad, $\hat{\mathcal{G}},\tau(V)$) are equivalent f.c.r.

of $\hat{\mathcal{G}}$. Applying now the construction (21) to $(\text{cad}, \hat{\mathcal{G}}, \tau(V))$, it is easy to prove the following lemma.

Lemma 3: Let $(\tilde{r}, \tilde{\mathcal{G}}, V)$ be a f.c.r. of $\tilde{\mathcal{G}}$, then $(\tilde{r}, \hat{\mathcal{G}}, V)$ and $(\text{cad}\circ q^{-1}, \hat{\mathcal{G}}, \mathcal{O})$ are equivalent, where $\mathcal{O}$ is the image of V under τ (25).

The following decisive theorem summarizes the content of Lemmas 1, 2, and 3.

Theorem 2: Let $\mathcal{G}$ be a connected Lie group with projective covering group $\hat{\mathcal{G}}$ and projective covering homomorphism $\hat{q}$. Then, a nontrivial orbit $\mathcal{O}$ in G^* under the co-adjoint action of $\hat{\mathcal{G}}$, such that $\ker\hat{q}$ acts trivially, defines a f.c.r. of $\mathcal{G}$ given by

$$\mathcal{G}\times\mathcal{O}\to\mathcal{O}, \quad (g,a)\to \text{cad}\hat{q}^{-1}(g)a. \tag{26}$$

Moreover every f.c.r. of $\mathcal{G}$ is equivalent to one of this form.

We can see that this theorem has many points of similarity with the well-known theorem on the projective irreducible representations of a connected Lie group.[6,11]

7. AN EXAMPLE: THE GALILEI GROUP

Let $\mathcal{G}$ be the Galilei group. The projective group $\hat{\mathcal{G}}$ of $\mathcal{G}$ is given[6] by the elements

$$(t, b, \mathbf{a}, \mathbf{v}, A), \quad t, b\in \mathbb{R}, \quad \mathbf{a}, \mathbf{v}\in\mathbb{R}^3, \quad A\in \text{SU}(2),$$

with the composition law

$$(t_1, b_1, \mathbf{a}_1, \mathbf{v}_1, A_1)(t_2, b_2, \mathbf{a}_2, \mathbf{v}_2, A_2)$$
$$=(t_1+t_2+w_{12}, b_1+b_2, \mathbf{a}_1+R(A_1)\mathbf{a}_2 + b_2\mathbf{v}_1, \mathbf{v}_1+R(A_1)\mathbf{v}_2, A_1A_2),$$

where $w_{12}\equiv \frac{1}{2}v_1^2 b_2+\mathbf{v}_1\cdot R(A_i)\mathbf{a}_2$ and $R(A)$ is the image of $A\in\text{SU}(2)$ on SO(3) under the covering homomorphism.

Let $\hat{B}=\{M, H, \mathbf{P}, \mathbf{K}, \mathbf{J}\}$ be the basis of $\hat{\mathcal{G}}$ with the commutation relations

$$\begin{aligned}
&[J_i, K_j]=\epsilon_{ijk}K_k, \quad [K_i, P_j]=-\delta_{ij}M,\\
&[J_i, P_j]=\epsilon_{ijk}P_k, \quad [K_i, H]=-P_i,\\
&[J_i, J_j]=\epsilon_{ijk}J_k.
\end{aligned} \tag{27}$$

If $(m, h, \mathbf{p}, \mathbf{k}, \mathbf{j})$ are the coordinates in the dual basis $\hat{B}^*$ of a point $a\in\hat{G}^*$, then the transformed point under the co-adjoint action of $\hat{\mathcal{G}}$ is given by

$$\begin{aligned}
&m'=m,\\
&h'=h+\tfrac{1}{2}mv^2+(R(A)\mathbf{p})\cdot\mathbf{v},\\
&\mathbf{p}'=R(A)\mathbf{p}+m\mathbf{v},\\
&\mathbf{k}'=R(A)\mathbf{k}+bR(A)\mathbf{p}+bm\mathbf{v}-m\mathbf{a},\\
&\mathbf{j}'=R(A)\mathbf{j}+\mathbf{v}\times R(A)\mathbf{k}+\mathbf{a}\times R(A)\mathbf{p}+m\mathbf{a}\times\mathbf{v}.
\end{aligned} \tag{28}$$

We have three functionally independent invariant functions over $\hat{G}^*$, which may be chosen in the following way:

$$m, \quad u\equiv 2mh-\mathbf{p}^2, \quad s\equiv|m\mathbf{j}+\mathbf{k}\times\mathbf{p}|. \tag{29}$$

There are two classes of orbits in $\hat{G}^*$:

(A) $m\neq 0$

In this case we may define the following coordinates:

$$\mathbf{q}\equiv-(1/m)\mathbf{k}, \quad \mathbf{s}\equiv\mathbf{j}-\mathbf{q}\times\mathbf{p}. \tag{30}$$

The action of $\hat{\mathcal{G}}$ is given by

$$\begin{aligned}
&\mathbf{q}'=R(A)(\mathbf{q}-(b/m)\mathbf{p})-b\mathbf{v}+\mathbf{a},\\
&\mathbf{p}'=R(A)\mathbf{p}+m\mathbf{v}, \quad \mathbf{s}'=R(A)\mathbf{s}.
\end{aligned} \tag{31}$$

The orbit of a point $(\mathbf{q}_0, \mathbf{p}_0, \mathbf{s}_0)$ is given by the manifold $\mathbb{R}^6(\mathbf{q},\mathbf{p})\times S^2(\mathbf{s})$, where $S^2(\mathbf{s})$ is the sphere $|\mathbf{s}|=|\mathbf{s}_0|$. One can check the following Poisson bracket relations:

$$\begin{aligned}
&\{q_i, q_j\}=\{p_i, p_j\}=\{q_i, s_j\}=\{p_i, s_j\}=0,\\
&\{q_i, p_j\}=\delta_{ij}, \quad \{s_i, s_j\}=\epsilon_{ijk}s_k.
\end{aligned} \tag{32}$$

(B) $m=0$

Now, we find four types:

(B_1) $p\equiv|\mathbf{p}|\neq 0, \quad \nu\equiv|\mathbf{k}\times\mathbf{p}|\neq 0$

In terms of the coordinates

$$v\equiv(1/p^2)\mathbf{k}\mathbf{p}, \quad \mathbf{w}\equiv\mathbf{k}\times\mathbf{p}, \tag{33}$$

the orbit of a point $(h_0, \mathbf{j}_0, v_0, \mathbf{p}_0, \mathbf{w}_0)$ is given by:

$$\mathbb{R}^5(h, \mathbf{j}, v)\times\{(\mathbf{p}, \mathbf{w})/|\mathbf{p}|=|\mathbf{p}_0|, |\mathbf{w}|=|\mathbf{w}_0|, \mathbf{p}\cdot\mathbf{w}=0\}. \tag{34}$$

The transformation properties of v and $\mathbf{w}$ are

$$v'=v+b, \quad \mathbf{w}'=R(A)\mathbf{w}. \tag{35}$$

(B_2) $p\neq 0, \ \nu=0$

We can introduce

$$\lambda\equiv(1/p)\mathbf{j}\cdot\mathbf{p}, \quad u\equiv\mathbf{j}\times\mathbf{p}. \tag{36}$$

The orbit of a point $(h_0, v_0, \mathbf{p}_0, \mathbf{u}_0)$ is the manifold

$$\mathbb{R}^2(h, v)\times\{(\mathbf{p}, \mathbf{u})\,|\,|\mathbf{p}|=|\mathbf{p}_0|, \ \mathbf{p}\cdot\mathbf{u}=0\}. \tag{37}$$

The function λ is invariant, i.e., $\lambda'=\lambda\equiv\pm s$ $(s\geq 0)$.

(B_3) $p=\nu=0, \quad r\equiv|\mathbf{k}|\neq 0$

If we define $\mathbf{z}\equiv\mathbf{j}\times\mathbf{k}$, the orbits are the manifolds

$$\{(\mathbf{k}, \mathbf{z})\,|\,|\mathbf{k}|=|\mathbf{k}_0|, \ \mathbf{k}\cdot\mathbf{z}=0\}. \tag{38}$$

The functions h and $\xi\equiv(1/r)\mathbf{j}\cdot\mathbf{k}$ are invariants on these orbits.

(B_4) $\mathbf{p}=\mathbf{k}=0$

Now the orbits are the spheres $|\mathbf{j}|=|\mathbf{j}_0|\equiv s$.

Finally, all realizations $(\text{cad}, \hat{\mathcal{G}}, \mathcal{O})$ of $\hat{\mathcal{G}}$ are labeled in the following form:

$$\begin{aligned}
&\text{I. } [m, u, s], \quad m\neq 0, \quad s\geq 0,\\
&\text{II. } [p, \nu], \quad p>0, \quad \nu>0,\\
&\text{III. } [p, \pm s], \quad p>0, \quad s\geq 0,\\
&\text{IV. } [h, r, \xi], \quad r>0,\\
&\text{V. } [h, s], \quad s\geq 0.
\end{aligned} \tag{39}$$

The projective covering homomorphism $\hat{q}:\hat{\mathcal{G}}\to\mathcal{G}$ is given by

$$(t, b, \mathbf{a}, \mathbf{v}, A)\xrightarrow{\hat{q}}(b, \mathbf{a}, \mathbf{v}, R(A)). \tag{40}$$

Since $\ker\hat{q}=\{(t, 0, 0, 0, \pm 1): t\in\mathbb{R}\}$ is a central subgroup in $\hat{\mathcal{G}}$, then it has a trivial co-adjoint action. Therefore, from Theorem 2 we conclude that all the f. c. r. of the Galilei group are given by (39).

ACKNOWLEDGMENT

I would like to thank L. Abellanas for stimulating discussions.

APPENDIX

In this Appendix we shall show the physical significance of some of the mathematical terms used in this paper.

In the Hamiltonian formulation of classical mechanics, the basic states of a physical system are points in a symplectic manifold called the "phase space" of the system. The natural automorphisms of this mathematical structure are the canonical transformations. Given an invariance group $\mathcal{G}$, the action of $\mathcal{G}$ on the states of a classical system is determinated by a canonical realization of $\mathcal{G}$ on the phase space of the system. From the group theoretical point of view, the elementary objects are the transitive canonical realizations of $\mathcal{G}$. We are interested in a special class of these elementary objects, the fundamental canonical realizations (f. c. r.) of $\mathcal{G}$. The states of an elementary system described by a f. c. r. are completely determined by the measurement of the observables described by the generators of the f. c. r. in Eq. (7).

[1]J. M. Souriau, *Structure des systemes dynamiques* (Dunod, Paris, 1970).
[2]M. Pauri and G. M. Prosperi, J. Math. Phys. 16, 1503 (1975); see also E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[3]R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
[4]R. Jost, Rev. Mod. Phys. 36, 572 (1964).
[5]P. Tondeur, *Lecture Notes in Mathematics*, *Vol.* 7 (Springer, New York, 1969), Vol. 7.
[6]J. Cariñena and M. Santander, J. Math. Phys. 16, 1416 (1975).
[7]G. Hochschild, Ann. of Math. 54, 96 (1951).
[8]A. A. Kirillov, Usp. Mat. Nauk 17, 57 (1962).
[9]A. A. Kirillov, Funkcional Nyj. Analiz 3, 36 (1969).
[10]M. Vergne, Bull. Soc. Math. Fr. 100, 301 (1972).
[11]V. Bargmann, Ann. Math. 59, 1 (1954).

Nth order perturbation theory for hydrogen*

Edward J. Kelsey and Joseph Macek

Behlen Laboratory of Physics, University of Nebraska, Lincoln, Nebraska 68588
(Received 24 September 1975)

This paper presents a simple reformulation of second order perturbation theory for hydrogen. We calculate first the perturbed wavefunction and then the perturbed matrix element. This procedure is repeated to obtain the nth order matrix element in terms of $n-1$ integrals. The parameters in the nth order matrix element are defined recursively. Schwinger's representation of the Coulomb Green's function follows immediately from our expression for the second order matrix element.

I. INTRODUCTION

The use of integral representations of the nonrelativistic Coulomb Green's function[1] to calculate second-order matrix elements[2] has become a standard technique of obtaining information about the hydrogen atom in second-order perturbation theory. This paper presents a simple reformulation of these methods which avoids much of the mathematical complexity of past approaches and offers a way to extend these techniques to higher order perturbation theory.

Previous methods[2] evaluate the matrix directly employing various representations of the Coulomb Green's function. In contrast we direct our attention to the evaluation of the perturbed wavefunction which we find by solving the corresponding inhomogeneous Schrödinger equation directly. The perturbed wavefunction is found with a minimum of algebra; furthermore, it has a simple form so that the integrals in the second-order matrix element are readily evaluated. In fact, the integrals over atomic coordinates needed to evaluate the second-order matrix elements are identical to those used to evaluate the first Born term.

We seek a closed form expression for the quantity $M(\mu_1, \mathbf{p}_1; \mu_2, \mathbf{p}_2)$:

$$M(\mu_1, \mathbf{p}_1; \mu_2, \mathbf{p}_2) = \sum_n \langle u_2 | \exp(i\mathbf{p}_2 \cdot \mathbf{r}) \frac{|n\rangle\langle n|}{\mathcal{E} - \mathcal{E}(n) + i\epsilon} \exp(i\mathbf{p}_1 \cdot \mathbf{r}) | u_1 \rangle, \tag{1}$$

where

$$u_i = \exp(-\mu_i r). \tag{2}$$

The sum over n goes over all hydrogen wavefunctions, continuum states as well as bound states. $\mathcal{E}(n)$ is the energy of the nth state of hydrogen.

Complicated expressions can be created from M by parametric differentiation with respect to μ_1, μ_2, p_{1x}, p_{1y}, p_{1z}, p_{2x}, p_{2y}, and p_{2z}. In this way linear combinations of partial derivatives of M can produce matrix elements of the type used in second-order perturbation theory for hydrogen and hydrogenlike ions.

An example we explore is $M(\mu_1, \mathbf{p}_1; \mu_2, -\mathbf{p}_2)|_{\mu_1=\mu_2=0}$ which up to a factor of $(2\pi)^{-3}$ is the Coulomb Green's function in momentum space. We obtain the same representation which was first derived by Schwinger in a much different way. Further, we show that the techniques developed for the second-order perturbation problem are easily extended to higher order perturbation theory.

II. COMPUTATION

To calculate M in Eq. (1), we first construct an integral expression for the auxiliary function $\chi(\mathbf{p}_1, \mathbf{r})$ which is formally the perturbed part of the first-order wavefunction[3]:

$$\chi(\mathbf{p}_1, \mathbf{r}) = \sum_n \frac{\langle \mathbf{r} | n \rangle \langle n | \exp(i\mathbf{p}_1 \cdot \mathbf{r}) | u_1 \rangle}{\mathcal{E} - \mathcal{E}(n) + i\epsilon}. \tag{3}$$

The function $\chi(\mathbf{p}_1, \mathbf{r})$ is a solution of the inhomogeneous Schrödinger equation

$$(H - \mathcal{E} - i\epsilon)\chi(\mathbf{p}_1, \mathbf{r}) = -\exp(i\mathbf{p}_1 \cdot \mathbf{r}) u_1(\mathbf{r}), \tag{4}$$

where H is the nonrelativistic Hamiltonian for hydrogen or a hydrogenlike ion of atomic number Z. Because the inhomogeneous term singles out direction $\hat{p}_1$ in space, we write the Hamiltonian in parabolic coordinates (ξ, η, ϕ) with z axis along $\hat{p}_1$. The choice of this coordinate system removes the ϕ dependence from $\chi(\mathbf{p}_1, \mathbf{r})$, and Eq. (4) becomes

$$\left(-\frac{1}{2m}\right)\left(\frac{4}{\xi + \eta}\right)\left[\frac{\partial}{\partial \xi}\left(\xi \frac{\partial \chi}{\partial \xi}\right) + \frac{\partial}{\partial \eta}\left(\eta \frac{\partial \chi}{\partial \eta}\right)\right] - 2\frac{Ze^2}{\xi + \eta}\chi - (\mathcal{E} + i\epsilon)\chi = -\exp[\tfrac{1}{2} i p_1(\eta - \xi) - \tfrac{1}{2}\mu_1(\xi + \eta)]. \tag{5}$$

We define $\gamma = (Ze^2)m$ and $X^2 = -2m(\mathcal{E} + i\epsilon)$ and rearrange factors:

$$\frac{\partial}{\partial \xi}\left(\xi \frac{\partial \chi}{\partial \xi}\right) + \frac{\partial}{\partial \eta}\left(\eta \frac{\partial \chi}{\partial \eta}\right) + \gamma\chi - \tfrac{1}{4}X^2(\xi + \eta)\chi = \tfrac{1}{2}m(\xi + \eta)\exp[-\tfrac{1}{2}(\mu_1 + ip_1)\xi + \tfrac{1}{2}(-\mu_1 + ip_1)\eta]. \tag{6}$$

Equation (6) is solved in a series of three steps. First we define a new function $\chi^{(a)}$ by the equation

$$\chi^{(a)} = -\int d\mu_1 \chi, \tag{7}$$

which is easily inverted by differentiation to find χ once $\chi^{(a)}$ is known. The differential equation for $\chi^{(a)}$ is obtained by integrating Eq. (6) over μ_1. We have

$$\frac{\partial}{\partial \xi}\left(\xi \frac{\partial \chi^{(a)}}{\partial \xi}\right) + \frac{\partial}{\partial \eta}\left(\eta \frac{\partial \chi^{(a)}}{\partial \eta}\right) + \gamma\chi^{(a)} - \tfrac{1}{4}X^2(\xi + \eta)\chi^{(a)} = m\exp[-\tfrac{1}{2}(\mu_1 + ip_1)\xi + \tfrac{1}{2}(-\mu_1 + ip_1)\eta]. \tag{8}$$

In the second step we remove the term involving $X^2(\xi + \eta)$ on the left-hand side of Eq. (8) by the transformation

$$\chi^{(b)} = \exp[\tfrac{1}{2}X(\xi + \eta)]\chi^{(a)}. \tag{9}$$

The branch of $X = \pm[-2m(\mathcal{E} + i\epsilon)]^{1/2}$ is chosen such that $\mathrm{Re}(X) > 0$.

The corresponding equation for $\chi^{(b)}$ is

$$\xi\frac{\partial^2\chi^{(b)}}{\partial\xi^2}+\eta\frac{\partial^2\chi^{(b)}}{\partial\eta^2}+(1-\xi X)\frac{\partial\chi^{(b)}}{\partial\xi}+(1-\eta X)\frac{\partial\chi^{(b)}}{\partial\eta}+(\gamma-X)\chi^{(b)}$$

$$=m\exp[-\tfrac{1}{2}(\mu_1+ip_1)\xi+\tfrac{1}{2}(-\mu_1+ip_1)\eta+\tfrac{1}{2}X(\xi+\eta)] \quad (10)$$

which, upon dividing by X, becomes

$$(X\xi)\frac{\partial^2\chi^{(b)}}{\partial(X\xi)^2}+(X\eta)\frac{\partial^2\chi^{(b)}}{\partial(X\eta)^2}+[1-(X\xi)]\frac{\partial\chi^{(b)}}{\partial(X\xi)}+[1-(X\eta)]$$

$$\times\frac{\partial\chi^{(b)}}{\partial(X\eta)}+\left(\frac{\gamma}{X}-1\right)\chi^{(b)}=\frac{m}{X}\exp[a(X\xi)+b(X\eta)], \quad (11)$$

where

$$a=\tfrac{1}{2}(-ip_1/X-\mu_1/X+1) \quad (12)$$

and

$$b=\tfrac{1}{2}(ip_1/X-\mu_1/X+1). \quad (13)$$

In the final step we write the inhomogeneous term as a sum of Laguerre polynomials. Since these polynomials are eigenfunctions of the operator on the left-hand side of Eq. (11), the solution for $\chi^{(b)}$ is readily found as a sum over Laguerre polynomials.

The inhomogeneous term is written in a double series using the generating function for Laguerre polynomials, i.e.,

$$\exp[a(x\xi)+b(x\eta)]$$

$$=(1-s)\sum_{k=0}^{\infty}\frac{s^k}{k!}L_k(X\xi)(1-t)\sum_{l=0}^{\infty}\frac{t^l}{l!}L_l(X\eta), \quad (14)$$

where $s=a/(a-1)$ and $t=b/(b-1)$ with the condition $|s|$, $|t|<1$. This condition on s and t may be met by restricting the parameters X, μ_1, and p_1 to be real positive quantities. The final result for M may be analytically continued to other regions of its parameters.

Inserting Eq. (14) in Eq. (11), we obtain the final form for the inhomogeneous differential equation,

$$(X\xi)\frac{\partial^2\chi^{(b)}}{\partial(X\xi)^2}+(X\eta)\frac{\partial^2\chi^{(b)}}{\partial(X\eta)^2}+[1-(X\xi)]\frac{\partial\chi^{(b)}}{\partial(X\xi)}$$

$$+[1-(X\eta)]\frac{\partial\chi^{(b)}}{\partial(X\eta)}+\left(\frac{\gamma}{X}-1\right)\chi^{(b)}$$

$$=\frac{m}{X}(1-s)(1-t)\sum_{k,l=0}^{\infty}\frac{s^kt^l}{k!\,l!}L_k(X\xi)L_l(X\eta). \quad (15)$$

Comparison of the left-hand side of Eq. (14) with the defining equation for Laguerre polynomials, Eq. (16),

$$vL_q''(v)+(1-v)L_q'(v)+qL_q(v)=0, \quad (16)$$

shows that the solution of Eq. (14) is given by the infinite series

$$\chi^{(b)}=-\frac{m}{X}(1-s)(1-t)\sum_{k,l=0}^{\infty}\frac{s^kt^lL_k(X\xi)L_l(X\eta)}{(k+l-\gamma/X+1)k!\,l!}. \quad (17)$$

We sum the series employing the integral relation

$$\frac{1}{k+l-\tau+1}=\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\int_C d\rho\rho^{-\tau+k+l}, \quad (18)$$

where $\tau=\gamma/X$. The contour C begins at $\rho=1+0i$ where the phase is zero and terminates at $1-0i$ after encircling the origin within the unit circle. The condition that C stays inside the unit circle is needed for the series in Eq. (19) to converge.

Substituting Eq. (18) into Eq. (17) gives the following form for $\chi^{(b)}$:

$$\chi^{(b)}=-\frac{m}{X}(1-s)(1-t)\left(\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\right)$$

$$\times\int_C d\rho\rho^{-\tau}\sum_{k,l=0}^{\infty}\frac{(s\rho)^k(t\rho)^l}{k!\,l!}L_k(X\xi)L_l(X\eta). \quad (19)$$

The double series in Eq. (19) is summed using Eq. (14). We find

$$\chi^{(b)}=-\frac{m}{X}(1-s)(1-t)\left(\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\right)\int_C d\rho\frac{\rho^{-\tau}}{(1-\rho s)(1-\rho t)}$$

$$\times\exp\left[\left(\frac{\rho s}{\rho s-1}\right)(X\xi)+\left(\frac{\rho t}{\rho t-1}\right)(X\eta)\right]. \quad (20)$$

Substituting Eq. (7) and Eq. (9) in Eq. (20), we obtain a closed form expression for $\chi(\mathbf{p}_1,\mathbf{r})$, i.e.,

$$\chi(\mathbf{p}_1,\mathbf{r})=\frac{m}{X}\frac{\partial}{\partial\mu_1}\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\int_C d\rho\,\rho^{-\tau}\frac{(1-s)(1-t)}{(1-\rho s)(1-\rho t)}$$

$$\times\exp\left[\left(\frac{s\rho}{s\rho-1}-\frac{1}{2}\right)(X\xi)+\left(\frac{t\rho}{t\rho-1}-\frac{1}{2}\right)(X\eta)\right]. \quad (21)$$

Recalling our choice of $\hat{p}_1$ to be the $\hat{z}$ direction of our coordinate system, we perform a few algebraic manipulations to rewrite Eq. (21) in a form independent of the coordinate system, i.e.,

$$\chi(\mathbf{p}_1,\mathbf{r})=4mX\frac{\partial}{\partial\mu_1}\left(\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\right)\int_C d\rho\frac{\rho^{-\tau}}{(F_1\rho^2+2E_1\rho+D_1)}$$

$$\times\exp\left(\frac{X(F_1\rho^2-D_1)r+4iX^2\rho\mathbf{r}\cdot\mathbf{p}_1}{(F_1\rho^2+2E_1\rho+D_1)}\right), \quad (22)$$

where

$$D_i=(X+\mu_i)^2+p_i^2,$$
$$E_i=X^2-\mu_i^2-p_i^2, \quad (23)$$
$$F_i=(X-\mu_i)^2+p_i^2.$$

$\chi(\mathbf{p}_1,\mathbf{r})$ given in Eq. (22) is the auxiliary function we sought to evaluate $M(\mu_1,\mathbf{p}_1;\ \mu_2,\mathbf{p}_2)$ in Eq. (1). Note that

$$M(\mu_1,\mathbf{p}_1;\ \mu_2,\mathbf{p}_2)$$
$$=\langle u_2(\mathbf{r})|\exp(i\mathbf{p}_2\cdot\mathbf{r})|\chi(\mathbf{p}_1,\mathbf{r})\rangle. \quad (24)$$

The rest of the calculation of $M(\mu_1,\mathbf{p}_1;\ \mu_2,\mathbf{p}_2)$ is quite straightforward since the $\mathbf{r}$ dependence in $\chi(\mathbf{p}_1,\mathbf{r})$ is contained in the exponential term alone and D_1, E_1, and F_1 are independent of $\mathbf{r}$, i.e.,

$$M(\mu_1,\mathbf{p}_1;\ \mu_2,\mathbf{p}_2)=-2^4\pi Xm\frac{\partial}{\partial\mu_1}\frac{\partial}{\partial\mu_2}\left(\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\right)$$

$$\times\int_C d\rho\frac{\rho^{-\tau}}{[D_1D_2-2(E_1E_2-4X^2(\mathbf{p}_1\cdot\mathbf{p}_2))\rho+F_1F_2\rho^2]}. \quad (25)$$

The small amount of algebra involved in going from Eq. (1) to Eq. (25) suggests that this approach is a more natural way of obtaining the result than the previous method. Gavrila and Costescu derive an expression equivalent to Eq. (25) starting from Schwinger's representation for the Coulomb Green's function. In the process they twice use the most general form of the three denominator integrals given by Lewis.[4] The transformations needed to employ these integrals are quite cumbersome. In contrast, the approach of this paper avoids complicated integrals and transformations, and the starting point is Schrödinger's equation, not the Coulomb Green's function.

Gavrila has pointed out that the integrals over ρ in M and all possible parametric derivatives of M are integral representations of Appell hypergeometric functions. It is also true that M and its derivatives may be expressed in Gaussian hypergeometric functions. As a result any second-order matrix element of hydrogen may be expressed in terms of hypergeometric functions.

III. APPLICATIONS

A. Coulomb Green's function

A straightforward application of Eq. (25) is the construction of the Coulomb–Green's function in momentum space, $G(\mathbf{p}_1, \mathbf{p}_2;\ X^2)$. This Green's function relates to our matrix element M according to

$$G(\mathbf{p}_1, \mathbf{p}_2;\ X^2) = \frac{1}{(2\pi)^3} M(\mu_1, \mathbf{p}_1;\ \mu_2; -\mathbf{p}_2)\big|_{\mu_1=\mu_2=0}, \tag{26}$$

where $X^2 = -2m(\mathcal{E} + i\epsilon)$.

To obtain a simple form for $G(\mathbf{p}_1, \mathbf{p}_2;\ X^2)$ we need the derivative with respect to μ_1 and μ_2 evaluated at $\mu_1 = \mu_2 = 0$ of the denominator inside the contour integral in Eq. (25), i.e.,

$$\frac{\partial}{\partial\mu_1}\frac{\partial}{\partial\mu_2}\left[\frac{1}{D_1D_2 - 2(E_1E_2 + 4(\mathbf{p}_1\cdot\mathbf{p}_2)X^2)\rho + F_1F_2\rho^2}\right]\Bigg|_{\mu_1=\mu_2=0} = \frac{X^2}{2^2}\frac{d}{d\rho}\left[\left(\frac{1-\rho^2}{\rho}\right)\frac{1}{[X^2(\mathbf{p}_1-\mathbf{p}_2)^2 + (p_1^2+X^2)(p_2^2+X^2)((1-\rho)^2/4\rho)]^2}\right]. \tag{27}$$

Substituting Eq. (27) into Eq. (25), and the result into Eq. (26) gives the Coulomb–Green's function in the form first obtained by Schwinger,

$$G(\mathbf{p}_1, \mathbf{p}_2;\ X^2) = -\frac{X^3 m}{2\pi^2}\left(\frac{i\exp(i\pi\tau)}{2\sin\pi\tau}\right)\int_C d\rho\, \rho^{-\tau}\frac{d}{d\rho}\left[\left(\frac{1-\rho^2}{\rho}\right)\frac{1}{[X^2(\mathbf{p}_1-\mathbf{p}_2)^2 + (p_1^2+X^2)(p_2^2+X^2)((1-\rho)^2/4\rho)]^2}\right]. \tag{28}$$

B. Application to higher order perturbation theory

In Sec. II we computed the auxiliary function $\chi(\mathbf{p}_1, \mathbf{r})$ which when multiplied by $(2\pi)^{-3/2}$ with μ_1 set equal to zero, is the mixed representation of the Coulomb Green's function. The surprising result, Eq. (22) shows us that $\chi(\mathbf{p}_1, \mathbf{r})$ may be written as a single contour integral over a simple exponential function similar to the one which we started with as the inhomogeneous term in Eq. (4). This good fortune allows us to repeat the whole procedure to obtain matrix elements in higher order perturbation theory. For such a third-order matrix element we would require one more contour integral. In general, the nth order perturbation matrix element requires $n-1$ contour integrals.

We extend our notation to write down the master integral M_n for nth order perturbation theory. Expressions for nth order perturbation theory may be written as linear combinations of parametric derivatives of M_n, i.e.,

$$M_n = \sum_{j_1\cdots j_{n-1}} \langle u_n|\exp(i\mathbf{p}_n\cdot\mathbf{r})\frac{|j_{n-1}\rangle\langle j_{n-1}|\exp(i\mathbf{p}_{n-1}\cdot\mathbf{r})}{\mathcal{E}_{n-1} - \mathcal{E}(j_{n-1}) + i\epsilon}\, u_{n-1}\frac{|j_{n-2}\rangle\langle j_{n-2}|}{\mathcal{E}_{n-2} - \mathcal{E}(j_{n-2}) + i\epsilon}\times\cdots\times u_2\frac{|j_1\rangle\langle j_1|\exp(i\mathbf{p}_1\cdot\mathbf{r})|u_1\rangle}{\mathcal{E}_1 - \mathcal{E}(j_1) + i\epsilon}. \tag{29}$$

We rewrite M_n in the following form:

$$M_n = \langle u_n|\exp(i\mathbf{p}_n\cdot\mathbf{r})|\chi_{n-1}\rangle. \tag{30}$$

As in second-order perturbation theory the auxiliary function χ_{n-1} is computed first:

$$\chi_{n-1}(\mathbf{p}'_{n-1}, \mu'_{n-1}) = (4m)^{n-1}\prod_{i=1}^{n-1} X_i\frac{\partial}{\partial\mu_i}\left(\frac{i\exp(i\pi\tau_i)}{2\sin\pi\tau_i}\right)\int_C d\rho_i\frac{\rho_i^{-\tau_i}}{(F'_i\rho_i^2 + 2E'_i\rho_i + D'_i)}\exp\left(\frac{X_{n-1}[(F'_{n-1}\rho_{n-1}^2 - D'_{n-1})r + 4iX_{n-1}\rho_{n-1}\mathbf{r}\cdot\mathbf{p}'_{n-1}]}{(F'_{n-1}\rho_{n-1}^2 + 2E'_{n-1}\rho_{n-1} + D'_{n-1})}\right), \tag{31}$$

where

$$\mu'_i = -\frac{X_{i-1}(F'_{i-1}\rho_{i-1}^2 - D'_{i-1})}{(F'_{i-1}\rho_{i-1}^2 + 2E'_{i-1}\rho_{i-1} + D'_{i-1})} + \mu_i,\quad \mathbf{p}'_i = \frac{4X_{i-1}^2\rho_{i-1}\mathbf{p}'_{i-1}}{(F'_{i-1}\rho_{i-1}^2 + 2E'_{i-1}\rho_{i-1} + D'_{i-1})} + \mathbf{p}_i, \tag{32}$$

$$D'_i = (X_i + \mu'_i)^2 + p_i'^2,\quad E'_i = X_i^2 - \mu_i'^2 - p_i'^2,\quad F'_i = (X_i - \mu'_i)^2 + p_i'^2,$$

for $i \neq 1$, and

$$\mu'_1 = \mu_1,\quad \mathbf{p}'_1 = \mathbf{p}_1,\quad D'_1 = D_1,\quad E'_1 = E_1,\quad F'_1 = F_1. \tag{33}$$

The branch of $X_i = \pm[-2m(\mathcal{E}_i + i\epsilon)]^{1/2}$ is chosen such that $\mathrm{Re}(X_i) > 0$.

Finally, using the notation in Eq. (32) we write out a general expression for M_n.

$$M_n = -4^n (m)^{n-1} \pi \left(\prod_{i=1}^{n-1} X_i \frac{\partial}{\partial \mu_i} \left(\frac{i \exp(i\pi\tau_i)}{2 \sin \pi\tau_i} \right) \int_C d\rho_i \frac{\rho_i^{-\tau_i}}{(F_i' \rho_i^2 + 2E_i' \rho_i + D_i')} \right) \frac{\partial}{\partial \mu_n} \frac{1}{(\rho_n'^2 + \mu_n'^2)} . \tag{34}$$

*Supported in part by the National Science Foundation under Grant No. NSF MPS75-07805.

[1]J. Schwinger, J. Math. Phys. **5**, 1606 (1964); E.H. Wichmann and C.H. Woo, J. Math. Phys. **2**, 178 (1961); L. Hostler, J. Math. Phys. **5**, 591 (1964).

[2]M. Gavrila and A. Costescu, Phys. Rev. A **2**, 1752 (1970); C. Fronsdal, Phys. Rev. **179**, 1513 (1969); V.G. Gorshkov and B.S. Polikhanov, Zh. Eksp. Teor. Fiz. Pis. Red. **9**, 464 (1964) [JETP Lett. **9**, 279 (1969)]; S. Klarsfeld, Nuovo Cimento Lett. **1**, 682 (1969).

[3]In the context of calculating the nonrelativistic part of the Lamb shift, Schwinger uses a similar approach to the one presented here to obtain auxiliary functions and matrix elements in second order for *free* particle scattering in intermediate states. He expands the nonrelativistic Lamb shift propagator for hydrogen in a Born series of free particle scatterings of the Coulomb field and evaluates the second Born term using an approach similar to the one presented in this paper. See J. Schwinger's *Particle. Sources and Fields* (Addison-Wesley, Reading, Mass., 1973), Vol. II, p. 166.

[4]R.R. Lewis, Phys. Rev. **102**, 537 (1956).

Propagation through an anisotropic random medium. An integro–differential formulation

M. J. Beran* and J. J. McCoy†

Naval Research Laboratory, Washington, D.C. 20375
(Receiced 14 January 1976)

In a previous paper [J. Math. Phys. **15**, 1901 (1975)] we derived an equation governing the mutual coherence function, $\{\hat{\Gamma}\}$, of an initial plane wave signal propagating in an anisotropic random medium. In Eq. (49) of that paper we made a strong narrow-angle approximation which allowed us to derive an ordinary differential equation for $\{\hat{\Gamma}\}$. Here we derive an integro–differential equation for $\{\hat{\Gamma}\}$ in which this approximation is not made. We show that for the horizontal scattering cases considered in the previous paper the approximation is valid. The vertical spectrum, however, is changed somewhat.

1. INTRODUCTION

In a previous paper[1] we derived an equation governing the propagation of the mutual coherence function, $\{\hat{\Gamma}(x_{12}, y_{12}, z)\}$, of a plane wave in an anisotropic medium. In the interests of brevity we refer the reader to this paper for a full discussion of the problem and the relevant definitions.

We found that $\{\hat{\Gamma}(x_{12}, y_{12}, z)\}$ satisfied the equation

$$\frac{d\{\hat{\Gamma}(x_{12}, y_{12}, z)\}}{dz} = -\{\hat{\Gamma}(x_{12}, y_{12}, z)\}\bar{\sigma}_2(0,0) + \{\hat{\Gamma}(x_{12}, 0, z)\}\bar{\sigma}_2(x_{12}, y_{12}). \tag{1}$$

Here $\{\hat{\Gamma}(x_{12}, y_{12}, z)\} = \{\hat{p}(x_1, y_1, z)\hat{p}^*(x_2, y_2, z)\}$ and

$$\bar{\sigma}_2(x_{12}, y_{12}) = \left(\frac{2}{\pi}\right)^{1/2}\frac{\bar{k}^3}{4}\int_0^\infty \cos\left(\frac{\bar{k}y_{12}^2}{2s_z} - \frac{\pi}{4}\right)\frac{\sigma_2(x_{12}, s_z)}{(\bar{k}s_z)^{1/2}}\,ds_z,$$

$$\sigma_2(x_{12}, s_z) = \int_0^\infty \sigma(x_{12}, s_y, s_z)\,ds_y, \tag{2}$$

where $\sigma(x_{12}, s_y, s_z)$ is the correlation function associated with the fluctuations in the random medium.

In the course of deriving Eq. (1) we made an approximation in Eq. (49) of that paper. This approximation made it possible to derive Eq. (1) as an ordinary differential equation. Here we should like to return to Eq. (49) and derive an equation for $\{\hat{\Gamma}\}$ which does not require this approximation. The new equation for $\{\hat{\Gamma}\}$ will be an integro–differential equation, Eq. (21). Subsequent to deriving this equation we shall discuss the approximation which led to Eq. (1) and show that it was appropriate for the horizontal scattering problems considered in Sec. 4 of that paper. The vertical spectrum, however, is changed somewhat.

2. CONSIDERATION OF EQ. (49) IN REF. 1

The mutual coherence of the scattered radiation in the interval $(j\Delta z, (j+1)\Delta z)$ is given in Eq. (49) in Ref. 1 by the following expression:

$$\{\hat{\Gamma}_S(x_{12}, y_{12}, z)\} = \left(\frac{2}{\pi}\right)^{1/2}\frac{\bar{k}^3}{8}\int_0^z\int_{-\infty}^\infty \exp\left[\pm i\left(\frac{\bar{k}y_{12}^2}{2|s_z|} - \frac{\pi}{4}\right)\right] \times \frac{\sigma_2(x_{12}, s_z)}{(\bar{k}|s_z|)^{1/2}}\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\}\,ds_z\,dp_z \tag{3}$$

(the plus sign corresponds to $s_z > 0$ and the minus sign to $s_z < 0$).

$\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\}$ is the coherence function in the interval $(j\Delta z, (j+1)\Delta z)$ assuming that there is no scattering in this interval. Because of spatial homogeneity $\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\}$ is independent of p_z. $\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, 0, 0)\}$ is the mutual coherence function on the plane $z = j\Delta z$.

In terms of the field $\hat{p}(x)$ we have

$$\{\Gamma_{j\Delta z}(x_{12}, 0, s_z, p_z)\} = \{\hat{p}(x_1, y_1, z_1')\hat{p}^*(x_2, y_2, z_2')\}, \tag{4}$$

where

$$x_{12} = x_2 - x_1, \quad y_1 = y_2, \quad s_z = z_2' - z_1', \quad p_z = z_1'. \tag{5}$$

The independent variable p_z should not be confused with the pressure field $\hat{p}(\mathbf{x})$. In the expression for $\{\hat{\Gamma}_S(x_{12}, y_{12}, z)\}$ we have $z_1 = z_2 = z$.

To obtain Eq. (1), the approximation was made that

$$\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\} \approx \{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, 0, 0)\}. \tag{6}$$

To explore the nature of this approximation, we use the fact that $\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}, s_z, p_z)\}$ is a solution of the wave equation, with no scattering, in the interval $(j\Delta z, (j+1)\Delta z)$ and can be expressed as a combination of plane waves. We have

$$\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\} = \int_{-\infty}^\infty\int \{\tilde{\tilde{\Gamma}}_{j\Delta z}(k_1, k_2, 0, p_z)\} \times \exp(i\bar{k}s_z - ik_1x_{12} - ik_2y_{12} - ik_zs_z)\,dk_1\,dk_2, \tag{7}$$

where $\bar{k}^2 = k_1^2 + k_2^2 + k_z^2$. The third argument in $\tilde{\tilde{\Gamma}}_{j\Delta z}$ refers to the condition that the function may be found from $\hat{\Gamma}_{j\Delta z}$ on the plane $s_z = 0$. The left-hand side of Eq. (6) is obtained from Eq. (7) by setting $y_{12} = 0$. Further, as stated above, $\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z, p_z)\}$ is independent of p_z when the solution depends only on x_{12} and y_{12}, i.e., the solution is spatially homogeneous in the x_1, y_1 plane. For convenience we shall subsequently drop the p_z index.

The projected angles between the plane wave directions and the principal propagation direction (i.e., the z axis) are given by $k_1/\bar{k}$ and $k_2/\bar{k}$. For a narrow angle spectrum, $(k_1^2 + k_2^2)/\bar{k}^2 \ll 1$, and we may expand k_z as follows:

$$k_z \approx \bar{k} - \tfrac{1}{2}[(k_1^2 + k_2^2)/\bar{k}]. \tag{8}$$

We find then

$$\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z)\}$$

$$= \int_{-\infty}^{\infty} \int \exp\left[-ik_1x_{12} + i\frac{s_z}{2\bar{k}}(k_1^2 + k_2^2)\right]$$
$$\times\{\tilde{\tilde{\Gamma}}_{j\Delta z}(k_1, k_2, 0)\}dk_1\,dk_2. \quad (9)$$

The spread of the plane wave spectrum is caused in our problem by scattering. Using a single scatter calculation, we showed[1] that

$$k_1/\bar{k}_1 \approx O(1/\bar{k}l_x), \quad (10)$$

$$k_2/\bar{k} \approx O(1/(\bar{k}l_x)^{1/2}). \quad (11)$$

From Eq. (10) we conclude that if

$$\bar{k}l_{xM}/\bar{k}^2 l_{xm}^2 \ll 1 \quad (12)$$

as we have assumed in that paper, then

$$\exp(is_z k_1^2/2\bar{k}) \approx 1 \quad (13)$$

and

$$\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z)\}$$
$$= \int_{-\infty}^{\infty} \exp(is_z k_2^2/2\bar{k})\{\tilde{\Gamma}_{j\Delta z}(x_{12}, k_2, 0)\}dk_2 \quad (14)$$

where

$$\{\tilde{\Gamma}_{j\Delta z}(x_{12}, k_2, 0)\} = \int_{-\infty}^{\infty} \exp(-ik_1x_{12})\{\tilde{\tilde{\Gamma}}_{j\Delta z}(k_1, k_2, 0)\}dk_1. \quad (15)$$

In Ref. 1 the assumption was also made (without adequate justification) that

$$(s_z/2\bar{k})\, k_2^2 \ll 1. \quad (16)$$

From Eq. (11) we see that this is not true in general, and thus it is desirable to have an equation governing $\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}, z)\}$ that does not assume Eq. (16) is valid. We shall next derive such an equation in Sec. 3. We shall, however, show in Sec. 4 that, for the case of horizontal scattering treated in Ref. 1, Eq. (1) is nevertheless a valid approximation and the $\exp(is_z k_2^2/2\bar{k})$ term may be set equal to unity in Eq. (14). The vertical scattering results will be seen to be somewhat different.

3. DERIVATION OF IN INTEGRO-DIFFERENTIAL EQUATION FOR $\Gamma(x_{12}, y_{12}, z)$

From Eq. (14) we may write, using the inverse transform relation for $\{\tilde{\Gamma}_{j\Delta z}(x_{12}, k_2, 0)\}$,

$$\exp(i\bar{k}s_z)\{\hat{\Gamma}_{j\Delta z}(x_{12}, 0, s_z)\}$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int \exp\left(\frac{is_z k_2^2}{2\bar{k}} + ik_2 y_{12}'\right)$$
$$\times\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}', 0)\}dk_2\,dy_{12}'. \quad (17)$$

Substituting Eq. (17) into Eq. (3) yields then (integrating over p_z)

$$\{\hat{\Gamma}_S(x_{12}, y_{12}, z)\}$$
$$= \left(\frac{2}{\pi}\right)^{1/2}\frac{\bar{k}^3}{16\pi}\, z'\int_{-\infty}^{\infty}\int\int ds_z\,dk_2\,dy_{12}'$$
$$\times\left\{\exp\left[\pm i\left(\frac{\bar{k}y_{12}^2}{2|s_z|} - \frac{\pi}{4}\right)\right]\exp\left(\frac{is_z k_2^2}{2\bar{k}} + ik_2 y_{12}'\right)\right\}$$
$$\times\frac{\sigma_2(x_{12}, s_z)}{(k|s_z|)^{1/2}}\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}', 0)\}, \quad (18)$$

where

$$z' = z - j\Delta z.$$

If we now interchange orders of integration and define

$$\bar{\sigma}_3(x_{12}, y_{12}, k_2)$$
$$\equiv\left(\frac{2}{\pi}\right)^{1/2}\frac{\bar{k}^3}{8\pi}\int_0^{\infty}\cos\left(\frac{\bar{k}y_{12}^2}{2s_z} - \frac{\pi}{4} + \frac{s_z k_2^2}{2\bar{k}}\right)\frac{\sigma_2(x_{12}, s_z)}{(\bar{k}/s_z)^{1/2}}\,ds_z, \quad (19)$$

we have

$$\{\hat{\Gamma}_S(x_{12}, y_{12}, z)\} = z'\int_{-\infty}^{\infty}\int\bar{\sigma}_3(x_{12}, y_{12}, k_2)\exp(ik_2 y_{12}')$$
$$\times\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}', z' = 0)\}dk_2\,dy_{12}'. \quad (20)$$

In $\{\hat{\Gamma}_{j\Delta z}(x_{12}, y_{12}', z' = 0)\}$ we have dropped the $s_z = 0$ index and reintroduced an index to denote that this function is evaluated on the plane $z' = 0$ $(z = j\Delta z)$.

With Eq. (20) replacing Eq. (50) of Ref. 1, the theory development proceeds as contained therein by taking account of the energy conserving terms. The result is that Eq. (1) must be replaced by the integro—differential equation

$$\frac{d}{dz}\{\hat{\Gamma}(x_{12}, y_{12}, z)\}$$
$$= -\int_{-\infty}^{\infty}\int\bar{\sigma}_3(0, 0, k_2)\exp[ik_2(y_{12}' - y_{12})]$$
$$\times\{\hat{\Gamma}(x_{12}, y_{12}', z)\}dk_2\,dy_{12}' + \int_{-\infty}^{\infty}\int\bar{\sigma}_3(x_{12}, y_{12}, k_2)$$
$$\times\exp(ik_2 y_{12}')\{\hat{\Gamma}(x_{12}, y_{12}', z)\}dk_2\,dy_{12}'. \quad (21)$$

If the expression $s_z k_2^2/2\bar{k}$, may be approximated by zero, Eq. (21) reduces to Eq. (1).

The solution of Eq. (21) may in principle be obtained by fixing x_{12} in this equation and solving the integro—differential equation in the independent variables y_{12} and z. In terms of the y_{12} spatial transform

$$\{\tilde{\Gamma}(x_{12}, k_2', z)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(iy_{12}k_2')\{\hat{\Gamma}(x_{12}, y_{12}, z)\}dy_{12} \quad (22)$$

this equation assumes the simpler form

$$\frac{d}{dz}\{\tilde{\Gamma}(x_{12}, k_2', z)\}$$
$$= -2\pi\bar{\sigma}_3(0, 0, k_2')\{\tilde{\Gamma}(x_{12}, k_2', z)\}$$
$$+ 2\pi\int_{-\infty}^{\infty}\tilde{\sigma}_3(x_{12}, k_2', k_2)\{\tilde{\Gamma}(x_{12}, k_2, z)\}dk_2, \quad (23)$$

where $\tilde{\sigma}_3$ is the y_{12} spatial transform of $\bar{\sigma}_3$.

In the next section we shall point out that as $z \to \infty$ (suitably nondimensionalized) we expect $\{\hat{\Gamma}(0, y_{12}, z)\}$ to approach an asymptotic form independent of z. In this case $\{\hat{\Gamma}(0, y_{12}, \infty)\}$ would satisfy the equation

$$\{\tilde{\Gamma}(0, k_2', \infty)\} = \int_{-\infty}^{\infty}\left(\frac{\tilde{\sigma}_3(0, k_2', k_2)}{\bar{\sigma}_3(0, 0, k_2')}\right)\{\tilde{\Gamma}(0, k_2, \infty)\}dk_2. \quad (24)$$

In the approximation where $\bar{\sigma}_3(x_{12}, y_{12}, k_2)$ is independent of k_2 we found the simple result

$$\{\tilde{\Gamma}(0, k_2', \infty)\} = \tilde{\sigma}_3(0, k_2')\,\hat{I}/\bar{\sigma}_3(0, 0). \quad (25)$$

Here, however, we must solve the integral equation, Eq. (24), to determine the correct vertical scattering in the multiple scatter region.

4. APPROXIMATION OF EQ. (21) BY EQ. (1) WHEN $y_{12}=0$ (THE CASE OF HORIZONTAL SCATTERING)

A. Asymptotic form for $\Gamma(0, y_{12}, z)$

In order to consider the validity of replacing Eq. (21) by Eq. (1) when $y_{12}=0$ we return to some results we obtained by perturbation theory in Ref. 1. We found that if a plane wave in the z direction is scattered by an anisotropic random medium in which $\bar{k}l_y \le 0(1)$ and $\bar{k}l_x = \bar{k}l_z \gg 1$, then the characteristic angular spread of the scattered radiation in the y direction, θ_y, is of order $1/(\bar{k}l_z)^{1/2}$. Using this same type of analysis, McCoy[2] points out that if a plane wave is travelling at an angle θ_{yI} to the z axis, where θ_{yI} is of order $1/(\bar{k}_{lz})^{1/2}$ then the scattering about $\theta_{yI}, \Delta\theta_{yI}$, is again of order $1(\bar{k}l_z)^{1/2}$. If, however, θ_{yI} is of the order $\alpha/(\bar{k}l_z)^{1/2}$, where $\alpha \gg 1$, then $\Delta\theta_{yI}$ is of order $(1/\alpha)[1/(\bar{k}l_z)^{1/2}]$, that is, $\Delta\theta_{yI}/\theta_{yI} \ll 1$.

On the basis of this type of analysis we infer that in the multiple scatter region the angular distribution in the y direction reaches an asymptotic form of order $\beta/(\bar{k}l_z)^{1/2}$, where β may be a number significantly greater than unity, but is independent of $\bar{k}l_z$. This result is in contradistinction to the result for θ_x which continues to grow as $z \to \infty$.

The above result was obtained for a correlation function with a single length scale l_z. If there are two length scales l_{zM}, l_{zm} denoting respectively the maximum and minimum characteristic scales, then we find that θ_y is determined by the large scale variations, l_{zM}, where generally most of the fluctuation energy resides. In most of our calculations in Ref. 1 [e.g., Eq. (37)] we used the overly conservative condition that θ_y is determined by the small scale variations, l_{zm}. Here, however, it is important that we make use of the dependence of θ_y on the large scale variations.

When we made the approximation of neglecting the k_2 dependence in $\bar{\sigma}_3(x_{12}, y_{12}, k_2)$, we found that indeed an asymptotic form was reached and the result is given in Eq. (25). In the context of the present theory we must instead solve the integral equation given Eq. (24). For this paper, however, we accept the above physical argument and assume that Eq. (24) also has a solution and that θ_y is of order $\beta/(\bar{k}l_{zM})^{1/2}$. To obtain the solution, we may, for example, iterate the equation and take Eq. (25) as the first trial solution. That is,

$$\{\tilde{\Gamma}(0, k_2', \infty)\} \approx \frac{\hat{I}}{\bar{\sigma}_3(0,0)} \frac{1}{\bar{\sigma}_3(0,0,k_2')} \times \int_{-\infty}^{\infty} \tilde{\sigma}_3(0, k_2', k_2)\, \bar{\sigma}_3(0, k_2)\, dk_2. \quad (26)$$

B. Neglect of the k_2 dependence in $\bar{\sigma}_3(x_{12}, y_{12}, k_2)$

The assumption that θ_y reaches an asymptotic value as $z \to \infty$ may be used to show that in a number of physically important cases the expression

$$s_z k_2^2/2\bar{k}$$

may be taken equal to zero in this limit. Because $k_2/\bar{k}$ is of order θ_y the exponent, is of order $\bar{k}s_z\theta_y^2/2$.

Initially, after scattering, $\bar{k}s_z\theta_y^2/2$ is of order unity because s_z (for the scattered radiation) is of order l_{zM} and θ_y^2 is of order $1/\bar{k}l_{zM}$. As z increases, however, the θ_y^2 dependence on l_{zM} remains of the same order (with perhaps a change in the multiplicative constant), but the s_z values that are of importance for horizontal scattering become smaller and smaller. Thus the value of the expression decreases as z increases and eventually becomes much less than unity. Finally then Eq. (21) may be replaced by Eq. (1) when $y_{12}=0$ (horizontal scattering).

The physical basis for the conclusion that as z increases smaller values of s_z (i.e., smaller scale variations in the index of refraction field) become more important is the effectiveness of small scale variations in scattering. Although usually there is more energy in the larger scales, the smaller scales scatter over a larger angle (the angular spread resulting from a scale variation of characteristic size l, is of order $1/\bar{k}l$). In the small perturbation regime, (z small) we find from Eq. (21) that for an initial plane wave (i.e., $\{\hat{\Gamma}(x_{12}, y_{12}, 0)^{\wedge}\} = \hat{I}$)

$$\{\hat{\Gamma}(x_{12}, 0, z)\} = z\hat{I}[\bar{\sigma}_2(x_{12}, 0) - \bar{\sigma}_2(0,0)], \quad (27)$$

where

$$\bar{\sigma}_2(x_{12}, 0) = \text{const} \int_{-\infty}^{\infty} \frac{\sigma_2(x_{12}s_z)}{(\bar{k}|s_z|)^{1/2}}\, ds_z. \quad (28)$$

The integral in Eq. (28) is dominated by the large scale variations and θ_x is of order $1/kl_{zM}$. [Incidentally we note that in this single scatter region Eq. (21) reduces to Eq. (1) if the initial radiation is a plane wave. In this case $\{\hat{\Gamma}(x_{12}, y_{12}', z)\}$ is independent of y_{12}' and the integral over y_{12}' is proportional to $\delta(k_2)$.] The quantity of importance is the relative energy in the various scale sizes multiplied by the scale size. The fact that smaller scales scatter over wider angles is not of importance here since it is assumed that each scale size scatters only once in the distance z and larger scales have a higher probability of scatter.

In the region where multiple scatter takes place the fact that all scales scatter many times in the same propagation distance becomes of importance. In this case the smaller eddies are weighted by an additional $1/\bar{k}l$ term. The effect of this is to cause smaller values of s_z to become important in evaluating $\bar{\sigma}_3(x_{12}, 0, k_2)$ than would be the case in evaluating $\bar{\sigma}_2(x_{12}, 0)$ in Eq. (28) or $\bar{\sigma}_3(0, y_{12}, k_2)$. [Mathematically the effect occurs because in Eq. (21) the right-hand side is the difference of two terms and as z increases the important values of x_{12} contributing to $\{\hat{\Gamma}(x_{12}, 0, z)\}$ decrease.] Because $k_2^2/\bar{k}^2 = 0(kl_{zM})$ for all z the expression $s_z k_2^2/2\bar{k}$ approaches zero as z increases, and Eq. (21) may be replaced by Eq. (1) when $y_{12}=0$. For very large z the characteristic scale sizes responsible for scattering do reach an asymptotic form, but in this region $s_z k_2^2/2\bar{k}$ should be near zero.

The above argument may also be supported by a mathematical consistency argument. If we assume that for large z the exponential term is unity, we find the solutions given in Ref. 1. Analysis of these solutions, as given in Eq. (74), show that for $n < 5/2$ the solution is independent of l_{zM} while for $n > 5/2$ the dependence on l_{zM} is much weaker than in the single scatter region.

C. Summary

We have shown by a physical argument and a mathematical consistency argument that in the multiple scatter region the expression $s_z k_2^2/2\bar{k}$ may be approximated by zero and as a consequence when $y_{12}=0$ Eq. (21) may be approximated by Eq. (1). This justifies the use of Eq. (1) in Ref. 1 and also in Beran and McCoy[3] in which scattering of finite beams is discussed. It is also true that in the small perturbation region Eq. (1) is a valid approximation for an initial plane wave. Only in the transition region from the single scatter to the multiple scatter region is it necessary to solve Eq. (21).

For the vertical scattering case, $x_{12}=0$, the asymptotic solution given in Ref. 1 [Eq. (25)] should be replaced by the solution of Eq. (24). If the complete solution for $x_{12}\neq 0$, $y_{12}\neq 0$ is desired for all z, then Eq. (21) rather than Eq. (1) must be solved.

*Tel-Aviv University, Tel-Aviv, Israel.
†The Catholic University of America, Washington, D.C. 20064.

[1]M. Beran and J. J. McCoy, J. Math. Phys. 15, 1901 (1975).
[2]J. J. McCoy, *Teoria Ośrodków Wielofazowych*, edited by C. Eimer (Polskiej Akademii Nauk, Warzawa, 1974), p. 5.
[3]M. Beran and J. J. McCoy, J. Acoust. Soc. Am. 56, 1667 (1975).

A crossed-channel expansion of a conformal invariant scattering amplitude

R. Meuldermans*

Instituut voor Theoretische Fysica, Celestijnenlaan 200 D, B-3030 Heverlee, Belgium
(Received 7 November 1975; revised manuscript received 1 December 1975)

We assume that a state out of a representation space of a irreducible representation of SO(4,2), in particular the E^+(massless) and D^+(massive) series of representation, fully describes a particle. Then, considering conformal invariant scattering amplitudes, we set up a crossed-channel expansion of the t matrix for elastic two-body scattering, by calculating the Clebsch–Gordan coefficients involved. The essential point is the reduction of the direct products $E^+ \times E^-$ and $D^+ \times E^-$.

INTRODUCTION

The idea of the conformal group SO(4, 2) or its universal covering group SU(2, 2) being an exact symmetry group of scattering phenomena has gained more and more interest[1] in spite of the lack of experimental evidence.[2] Both phenomena can be explained easily. First there is the exciting new theoretical idea with an undoubtedly new physical content and aspects.[3] On the other hand the theory at present has not given any practical formula to test its validity. However, few exceptions do exist.[4,9] One practical thing which can be done is to set up a conformal invariant phase-shift analysis. This program has been started by some authors with different intentions and methods.[5,6] It might be of some interest however to have an expansion which can easily be extrapolated to asymptotic regions of the energy, because most of the people do believe the symmetry will manifest itself only in those kinematical regions. A crossed-channel expansion usually is more appropriate to do this job than a direct channel one.

Section 1 is devoted to the definition of generators of SO(4, 2) with the involved commutation relation. The observables are indicated with their action on definite states of Hilbert space.

Section 2 will consist of the calculation of the Clebsch–Gordon coefficients involved in the elastic scattering of two massless particles (spinless). This part consists in essence to the reduction of the direct product of $E^{(+)} \times E^{(-)}$, where $E^{\pm}$ refer to representation of the exceptional degenerate series.[7]

Finally, the last section consists of the same procedure for scattering of a massless particle on a massive one, which involves the reduction of $E^- \otimes D^+$ representation.

I. PRELIMINARIES

The generators of the full conformal group SO(4, 2) will be denoted by L_{ab}, $a, b = 0, 1, 2, 3, 5, 6$. Their commutation relations are given by

$$[L_{ab}, L_{cd}] = i(g_{ac}L_{bd} - g_{ad}L_{bc} - g_{bc}L_{cd} + g_{bd}L_{ac}). \qquad (1.1)$$

The metric is chosen to be $g_{11} = g_{22} = g_{33} = g_{55} = -g_{00} = -g_{66} = 1$.

We select out of this the Poincaré generators (Greek letters)

$$L_{\mu\nu} = L_{ab}, \quad \mu, \nu = 1, 2, 5, 6, \qquad (1.2)$$

$$P_\mu = L_{0\mu} + L_{3\mu}. \qquad (1.3)$$

The generators for the dilatations and special conformal transformations are given by

$$D = L_{03}, \quad K_\mu = L_{0\mu} - L_{3\mu} \qquad (1.4)$$

Commutation relations in terms of these combinations are

$$\begin{aligned}
&[L_{\mu\nu}, P_\alpha] = i(g_{\mu\alpha}P_\nu - g_{\nu\alpha}P_\mu),\\
&[L_{\mu\nu}, K_\alpha] = i(g_{\mu\alpha}K_\nu - g_{\nu\alpha}K_\mu),\\
&[L_{\mu\nu}, D] = 0, \quad [K_\mu, K_\nu] = 0,\\
&[P_\alpha, K_\beta] = -2i(L_{\alpha\beta} + g_{\alpha\beta}D),\\
&[D, P_\alpha] = -iP_\alpha, \quad [D, K_\alpha] = iK_\alpha.
\end{aligned} \qquad (1.5)$$

The rotation and boosts are defined as follows:

$$L_i = (L_{25}, L_{51}, L_{12}), \quad i = 1, 2, 5 \qquad (1.6)$$

$$N_i = -(L_{01}, L_{02}, L_{03}). \qquad (1.7)$$

States in Hilbert spaces of representations of the group will be labelled with respect to the Poincaré subgroup $SO(3,1) \times T_4$. The observables therefore are

$$C_4, C_3, C_2, P^2, W_\mu W^\mu, \mathbf{P}, W_6. \qquad (1.8)$$

The first three are the Casimir invariants of resp. order two, three, and four whereas $P^2 = -m^2$ en $W^2 = +m^2 s(s+1)$ are Einsteinian mass and spin ($W_\mu = +\frac{1}{2}\epsilon_{\mu\nu\alpha\beta}P^\alpha L^{\alpha\beta}$). The last one is proportional to the helicity λ of the particle.

The first two Casimirs will be omitted as whole our interest will be devoted to the second order Casimir operator C_2 (in all cases considered C_2 already distinguishes the representations):

$$C_2 = \tfrac{1}{2}L_{\mu\nu}L^{\mu\nu} + 4iL_{03} - L_{03}^2 - K_\mu P^\mu, \qquad (1.9)$$

and a state will get the labels

$$|C_2; ms, \mathbf{p}\lambda\rangle_\epsilon \quad (\epsilon = \text{sign of the energy}). \qquad (1.10)$$

It is often much more appropriate to use the four components of the momentum as observables. Thus in the following we will use

$$|C_2; P_\mu s\lambda\rangle. \qquad (1.11)$$

The physical ansatz therefore is that there exists a one-

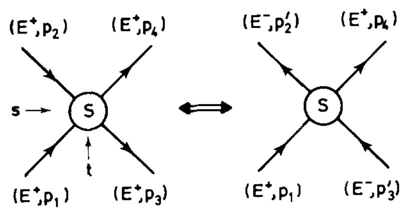


FIG. 1.

to-one relation between the physical object particle and a particular state like (1.11). It is suggested by the work of several authors[7,8] to put a massless particle into a representation of the E^+ series, which is characterized by

$$C_2 = 3(\lambda^2 - 1). \tag{1.12}$$

This representation contains only one helicity value and only admits a positive energy spectrum.[8] States of a D^+ representation (most degenerate discrete series) will be used for massive particle, they admit one spin value, positive energies, and

$$C_2 = 2(s-1)(s+2) + \nu, \quad \nu = 0, 1, 2, 3, \cdots \quad \text{(see Ref. 3b)}. \tag{1.13}$$

Two similar representations exist, $[E^-, D^-]$, which differ from the earlier in the sign of the energy spectrum. The notation therefore is clear.

States for these representations will have the shorthand notation

$$\begin{aligned} &|(E^+c_2); \mathrm{p}_\mu s\lambda\rangle \equiv |p_\mu, \lambda\rangle, \quad p_6 > 0,\\ &|(E^-c_2); \mathrm{p}_\mu s\lambda\rangle \equiv |p'_\mu, s\lambda\rangle, \quad p'_6 < 0 \quad (p = -p'),\\ &|(D^+c_2); \mathrm{p}s\lambda\rangle = |\nu, p_\mu, s, \lambda\rangle, \quad p_6 > 0,\\ &|(D^-c_2): \mathrm{p}s\lambda\rangle = |\nu, p'_\mu, s, \lambda\rangle, \quad p'_6 < 0 \quad (p_\mu = -p'_\mu). \end{aligned} \tag{1.14}$$

The normalization of the states is chosen to be

$$\langle p_2\lambda_2|p_1\lambda_1\rangle = \delta_{\lambda_2\lambda_1} E\delta^3(p_2 - p_1), \tag{1.15}$$

$$\langle \nu_2 p_2 s_2 \lambda_2 | \nu_1 p_1 s_1 \lambda_2\rangle = \delta_{\lambda_2\lambda_1}\delta_{s_2 s_1}\delta_{\nu_2\nu_1}\delta^4(p_2 - p_1). \tag{1.16}$$

II. REDUCTION OF A $E^+ \times E^-$ REPRESENTATION

Consider the matrix element of a conformal invariant scattering operator corresponding to the scattering of two massless particles (spinless):

$$s = -(p_1 + p_2)^2, \quad t = -(p_1 - p_3)^2, \quad u = -(p_1 - p_4)^2.$$

The matrix element $\langle E^+, p_4, E^+p_3|S-1|E^+p_1, E^+p_2\rangle$ is equivalent to $\langle E^+, p_4; E^-p'_2|S-1|E^+p_1, E^-p'_2\rangle$.

Introducing a complete set of intermediate states, we obtain

$$\begin{aligned} &\langle p_4, p'_2|S-1|p_1, p'_3\rangle\\ &= \sum_{\alpha,,\lambda} \int d^4p \langle p_4 p'_2|\alpha; ps\lambda\rangle\langle\alpha ps\lambda|S-1|\alpha ps\lambda\rangle\\ &\times\langle\alpha ps\lambda|p_1, p'_3\rangle. \end{aligned} \tag{2.1}$$

The sum sign is rather a formal thing because it depends upon the nature of the spectrum of the involved quantum number, which we have first to determine. Furthermore, α stands for all quantum numbers needed to specify the intermediate state.

To obtain (2.1), use has been made of the invariance of S. By the Wigner–Eckart theorem, the reduced S-matrix element

$$\langle\alpha ps\lambda|S|\alpha ps\lambda\rangle = S(\alpha) \tag{2.2}$$

is only a function of the Casimir labels, which are α. The main problem is to determine the two Clebsch–Gordan coefficients, which is equivalent to the reduction of the $E^+ \otimes E^-$ representation into irreducible parts.

In order to achieve this, we apply the following formalism. We define the action of a generator D on a state by the matrix elements

$$\langle p'|D|p\rangle = \mathcal{D}(p)\langle p'|p\rangle, \tag{2.3}$$

where the script letter stands for a differential expression. Therefore, for the successive action of two Hermitian operators it follows that

$$\langle p'|D_2 D_1|p\rangle = \mathcal{D}_1(p)\mathcal{D}_2(p)\langle p'|p\rangle, \tag{2.4}$$

which indicates, because the expression on the rhs do not depend upon p', that we can define

$$D|p\rangle = \mathcal{D}(p)|p\rangle. \tag{2.5}$$

Successive operation reverses the order of the two differential expressions

$$D_2 D_1|p\rangle = \mathcal{D}_1(p)\mathcal{D}_2(p)|p\rangle. \tag{2.6}$$

Remark: If discrete labels are involved in the rhs, there should be a sum over all accessible discrete labels. Notice also that the differential expressions do not form in general a representation of the generators; e.g.,

$$[D_1, D_2]|p\rangle = -[\mathcal{D}_1, \mathcal{D}_2]|p\rangle.$$

The commutation relations together with the hermiticity condition enables us to calculate all generators. Straightforward calculations give for the E^+ representation

$$\begin{aligned} L_1|p,\lambda\rangle &= \left(i(\mathrm{p}\times\nabla_p)_1 + \lambda\frac{p_1}{E-\alpha p_5}\right)|p,\lambda\rangle,\\ L_2|p,\lambda\rangle &= \left(i(\mathrm{p}\times\nabla_p)_2 + \lambda\frac{p_2}{E-\alpha p_5}\right)|p,\lambda\rangle,\\ L_3|p,\lambda\rangle &= [i(\mathrm{p}\times\nabla_p)_2 - \lambda\alpha]|p,\lambda\rangle,\\ N_1|p,\lambda\rangle &= \left(-iE\frac{\partial}{\partial p_1} - \lambda\frac{p_2}{E-\alpha p_5}\right)|p,\lambda\rangle,\\ N_2|p,\lambda\rangle &= \left(-iE\frac{\partial}{\partial p_2} + \lambda\frac{p_1}{E-\alpha p_5}\right)|p,\lambda\rangle,\\ N_3|p,\lambda\rangle &= \left(-iE\frac{\partial}{\partial p_5}\right)|p,\lambda\rangle,\\ D|p,\lambda\rangle &= i(\mathrm{p}\cdot\nabla + 1)|p,\lambda\rangle,\\ K_1|p,\lambda\rangle &= -p_1\Delta + 2\frac{\partial}{\partial p_1}(\mathrm{p}\cdot\nabla)\\ &\quad - 2i\lambda\left(\frac{p_2}{E-\alpha p_5}\frac{\partial}{\partial p_5} + \frac{\partial}{\partial p_2}\right)|p,\lambda\rangle, \end{aligned} \tag{2.7}$$

$$K_5|p,\lambda\rangle=-p_2\Delta+2\frac{\partial}{\partial p_2}(\mathbf{p}\cdot\nabla)$$

$$+2i\lambda\left(\frac{p_1}{E-\alpha p_5}\frac{\partial}{\partial p_5}+\frac{\partial}{\partial p_1}\right)|p,\lambda\rangle,$$

$$K_5|p,\lambda\rangle=-p_5\Delta+2\frac{\partial}{\partial p_5}(\mathbf{p}\cdot\nabla)$$

$$-\frac{2i\lambda}{E-\alpha p_5}\left(p_1\frac{\partial}{\partial p_2}-p_2\frac{\partial}{\partial p_1}\right)+2\alpha\frac{\lambda^2}{E-\alpha p_5}|p,\lambda\rangle,$$

$$K_6|p,\lambda\rangle=-E\Delta-\frac{2i\lambda}{E-\alpha p_5}\;p_1\frac{\partial}{\partial p_2}-p_2\frac{\partial}{\partial p_1}$$

$$+\frac{2\lambda^2}{E-\alpha p_5}|p,\lambda\rangle\quad(E=|\mathbf{p}|).$$

The generators for the E^- representation are obtained by replacing in (2.5) E by $-E=-|\mathbf{p}|$.

A comment seems necessary: As is known for the Poincaré group, these manipulations do not determine a unique solution. This is related to the fact that states are not specified uniquely. A choice of solution has been made in terms of one parameter α. This parameter is not determined by enlarging the algebra to that of the conformal group; however, it is restricted to satisfy $\alpha^2=1$. Specification of the representation of SO(4, 2) (C_2 being a C-number) restricts α to be equal to $+1$.

Equating the Clebsch–Gordan coefficient to

$$\delta(p-p_1-p'_3)f(s,\lambda;\mathbf{p},\mathbf{q}),\quad\text{where }\mathbf{q}=\mathbf{p}_1-\mathbf{p}_3{}',$$

we see that $f(s,\lambda;\mathbf{p},\mathbf{q})$ should be a simultaneous eigenfunction of C_2, W^2, W_6:

$$C_2f(\;)=c_2f(\;),$$

$$W^2f(\;)=-\omega^2s(s+1)f(\;),\tag{2.8}$$

$$W_6f(\;)=-m|P|f(\;)\quad(P^2=\omega^2>0\text{ spacelike}).$$

Putting these operators into differential form, we have

$$(C_2+10)f=\left(-2t\nabla_q^2+2(q_i+p_i)(q_i-p_j)\frac{\partial^2}{\partial q_i\partial q_j}\right.$$

$$\left.-4q_i\frac{\partial}{\partial q_i}\right)f,\tag{2.9}$$

$$W_\mu W^\mu f=\left(-\omega^4\nabla_q^2-\omega^2(q_i+p_i)(q_j-p_j)\frac{\partial^2}{\partial q_i\partial q_j}\right.$$

$$\left.-2q_i\omega^2\frac{\partial}{\partial q_i}\right)f,\tag{2.10}$$

$$W_6f=-i(\mathbf{p}\times\mathbf{q})_i\frac{\partial}{\partial q_i}f.\tag{2.11}$$

We restricted ourself to the spinless case.

Out of (2.9) and (2.10) one derives the relation

$$C_2+10=-2W^2/\omega^2,\tag{2.12}$$

from which it follows that

$$C_2=-10+2s(s+1).\tag{2.13}$$

Equation (2.12) reduces the problem to the calculation of a pure Poincaré Clebsch–Gordan coefficient. Equations (2.9), (2.10), (2.12) together with the square integrability of the function $f(\;)$ determine the spectrum to be as follows:

$$m:\pm0,1,2,\ldots,\quad s=-\tfrac{1}{2}+i\rho\quad(\rho\text{ real and positive});$$

no discrete spectrum exists.

The resulting expansion for the t matrix, defined by

$$\langle p_4p_3|S-1|p_1p_2\rangle=\delta(p_4+p_3-p_1-p_2)\langle p_4p_3|T|p_1p_2\rangle$$

becomes

$$\langle p_4p_3|T|p_1p_2\rangle=\frac{1}{4\pi}\int_0^\infty d\rho\,a(\rho)\rho\tanh\rho\,B_{0,0}^{-1/2+i\rho}(2\eta-1);\tag{2.14}$$

here the notations stand for

$$\eta=-s/t,\quad a(\rho)=[S(\rho)-1]16\pi^3/\rho^2\tanh\rho,\tag{2.15}$$

$B_{0,0}^{-1/2+i\rho}(z)=P_{-1/2+i\rho}(z)$, the conical functions.[10]

III. REDUCTION OF A $D^+\times E^-$ REPRESENTATION (SPINLESS)

A more interesting case involves the scattering of a massless particle on a massive one. Figure 2 shows which direct products are involved. $P=(p_1+p_4')$ now can be time-, space-, and lightlike. We consider the case of P being spacelike, e.g.,

$$P^2=\omega^2\quad(\omega^2>0).$$

The operators for the D^+ representation without spin can be calculated by a similar procedure as in Sec. II. This gives

$$L_i|\nu,p\rangle=i(\mathbf{p}\times\nabla)_i|\nu,p\rangle,$$

$$N_i|\nu,p\rangle=-i\left(E\frac{\partial}{\partial p_i}+p_i\frac{\partial}{\partial E}\right)|\nu,p\rangle,\tag{3.1}$$

$$D|\nu,p\rangle=i(p_\mu\partial^\mu+2)|\nu,p\rangle,$$

$$K_\mu|\nu,p\rangle=\left(-p_\mu\partial_\nu\partial^\nu+2(p_\nu\partial^\nu+2)_\mu-\frac{\nu^2p_\mu}{p^2}\right)|\nu,p\rangle.$$

We then introduce the variables

$$\mathbf{P}=\mathbf{p}+\mathbf{p}',$$

$$\mathbf{Q}=\mathbf{p}-\mathbf{p}',\tag{3.2}$$

$$y=m^2/\omega^2=-m^2/\nu;$$

the analog of (2.9) for this case then is

$$(C_2+3)f(\;)=\left\{4\left[\frac{\omega^2(1+y)}{2}\nabla_Q^2-2p_ip_j'\frac{\partial^2}{\partial Q_i\partial Q_j}\right.\right.$$

$$\left.\left.+\left(\frac{3Q_i}{2}-\frac{P_i}{2}\right)\cdot\frac{\partial}{\partial Q_i}\right]+\frac{\nu^2(y-1)}{2y}\right\}f.\tag{3.3}$$

In order to make a separable differential equation, we introduce a set of new variables:

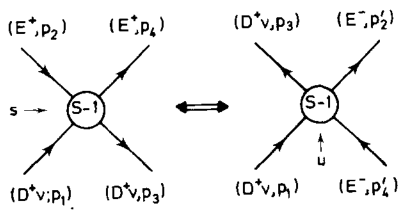


FIG. 2.

$$\mathrm{P}=\mathrm{p}+\mathrm{p}', \tag{3.4a}$$

$$x=p_6-|p'| \quad \text{(the total energy)}, \tag{3.4b}$$

$$y=-m^2/u, \tag{3.4c}$$

$$e_\mu=L_\mu^{-1\nu}(P)q_\nu. \tag{3.4d}$$

Here

$$q_\mu=(\omega/\sqrt{k})(Q_\mu+yP_\mu) \quad \text{with } \sqrt{k}=\omega^2(1+y), \tag{3.5}$$

The 2-vector e can be parametrized by

$$e=(\sinh\theta\cos\beta, \sinh\theta\sin\beta, 0; \cosh\theta) \quad (\cosh\theta=z). \tag{3.6}$$

If we put $f(y,z,\beta)=Y(y)Z(z)B(\beta)$, (3.3) will be equivalent with the following system of equations:

$$y(y+1)^2\frac{\partial^2 Y}{\partial y^2}+(1+3y)(y+1)\frac{\partial Y}{\partial y}+\left(\frac{\nu^2(y^2-1)}{4y}+p-(y+1)\right)Y=0, \tag{3.7}$$

which means that q is a 2-vector ($P\cdot q=0$ and $q^2=-1$). $L(P)$ is a Lorentz transformation which transforms $P_R(0,0,\omega;0)$ to $P=\omega(\cosh a\sin\tau\sin\Psi, \cosh a\sin\tau\cos\Psi, \cosh a\cos\tau, \sinh a)$,

$$L_\mu^\nu(P)=\begin{vmatrix} \cos^2\Psi+\cos\tau\sin^2\Psi & -(1-\cos\tau)\sin\Psi\cos\Psi & \cosh a\sin\tau\sin\Psi & \sinh a\sin\tau\sin\Psi \\ -(1-\cos\tau)\cos\Psi\sin\Psi & \sin^2\Psi+\cos\tau\cos^2\Psi & \cosh a\sin\tau\cos\Psi & \sinh a\sin\tau\cos\Psi \\ -\sin\tau\cos\Psi & -\sin\tau\cos\Psi & \cosh a\cos\tau & \sinh a\cos\tau \\ 0 & 0 & \cosh a & \cosh a \end{vmatrix}.$$

$$\frac{\partial^2 Z}{\partial z^2}+\frac{2z}{z^2-1}\frac{\partial Z}{\partial z}-\left(\frac{p}{z^2-1}+\frac{n^2}{(z^2-1)^2}\right)Z=0, \tag{3.8}$$

$$\frac{\partial^2 B(\beta)}{\partial\beta^2}+n^2B(\beta)=0, \tag{3.9}$$

where λ stands for $\lambda=(C_2+3)/2$ and p and n^2 are two other separation variables.

We need to determine the spectrum. In fact, expressing the operator W^2 in the new variables, we obtain a system which is equivalent to (3.8)—(3.9), where (3.9) is the eigenvalue equation for $(W_6)^2$.

So the spectrum of p and n^2 is not hard to obtain:

$$n: 0,1,2,\ldots, \quad p=s(s+1)=-\rho^2-\tfrac{1}{4}, \quad \text{where } s=-\tfrac{1}{2}+i\rho. \tag{3.10}$$

They will have some influence on the spectrum of λ. Equation (3.7) is of the Sturm—Liouville type with measure $\rho(y)=y+1$. The solution can be catalogued into a Riemanian scheme:

$$Y(y)=P\begin{bmatrix} 0 & 1 & \infty & \\ 0 & 0 & 1+s+\nu/2+[1-(\nu^2/4-\lambda)]^{1/2} & -y \\ -\nu & -1-2s & 1+s+\nu/2-[1-(\nu^2/4-\lambda)]^{1/2} & \end{bmatrix} y^{\nu/2}(1+y)^s. \tag{3.11}$$

The range of y for fixed but asumptotic energies $s=-(p_1+p_2)^2$ is $[0,\infty]$. So we are in the double singular case according to Titchmarsh.[11] Two appropriate solutions for the discussion therefore are

$$Y^A(y)=y^{\nu/2}(y+1)^s F_{21}(a,b,c:-y), \tag{3.12}$$

$$Y^B(y)=y^{\nu/2}(y+1)^s(y)^{-a}F_{21}(a,a+1-c,a+1-b;-1/y), \tag{3.13}$$

where

$$c=\nu+1, \quad a=1+s+\nu/2+[1-(\nu^2/4-\lambda)]^{1/2}, \quad b=1+s+\nu/2-[1-(\nu^2/4-\lambda)]^{1/2}.$$

In order to determine the spectrum of Eq. (3.7), we put this equation into standard form (no linear derivatives), giving

$$\frac{d^2Y_1(\alpha)}{d\alpha^2}+[\xi-q(\alpha)]Y_1(\alpha)=0. \tag{3.14}$$

Here we have put

$$y+1=\cosh\alpha, \quad Y(y)=Y_1(\alpha)(\sinh\alpha)^{-1/2}[\cosh(\alpha/2)]^{-1},$$
$$q(\alpha)=\tfrac{1}{4}[\cosh\alpha(\cosh\alpha-2)/\sinh^2\alpha]+[\nu^2/4\sinh^2(\alpha/2)]-s(s+1)/\cosh^2(\alpha/2), \quad \xi=\nu^2/4-\lambda-\tfrac{3}{4}=-t(t+1). \tag{3.15}$$

We divide the interval into two pieces by introducing an arbitrary intermediate point $y=b$.

Square integrable solution with respect to α over the interval $[0, b]$ and $[b, \infty]$ are given resp. by

$$Y_1^B(\alpha)=\sqrt{2}[\sinh(\alpha/2)]^{\nu+1/2-2a}[\cosh(\alpha/2)]^{2s+3/2}{}_2F_1(a, a+1-c, a+1-b; -1/\sinh^2(\alpha/2)),$$
$$Y_1^A(\alpha)=\sqrt{2}[\sinh(\alpha/2)^{\nu+1/2}[\cosh(\alpha/2)]^{2s+3/2}{}_2F_1(a, b, c; -\sinh^2(\alpha/2)). \tag{3.16}$$

$Y_1^A(\alpha)$ always is square integrable for all values of ν, whereas Y_1^B is square integrable only for the region $\mathrm{Re}t>-\frac{1}{2}$. Because of symmetry, this is exactly one of the complex half-planes in t we only had to consider.

Because $Y_1^A(\alpha, \xi)$ is real for real values of ξ, the spectrum entirely consists of those points which give contribution to

$$\lim_{\mathrm{Im}\xi\to 0} \mathrm{Im}[Y_1^A(\xi, b)Y_1^B(\xi, b)/W_\alpha(Y_1^A, Y_1^B)]. \tag{3.17a}$$

The denominator consists of the Wronskion of Y_1^A and Y_1^B with respect to α. Calculation shows that contribution only comes from a continuous region for which $t^*=-t-1$; this means $t=-\frac{1}{2}+ix$; no discrete spectrum appears. For $t=-\frac{1}{2}+ix$,

$$\mathrm{Im}[Y_1^A(\xi, b)Y_1^C(\xi, b)/W_\alpha(Y_1^A, Y_1^B)]=(1/2x)\left|\Gamma(a)\Gamma(b)/\Gamma(c)\Gamma(a-b)\right|^2\left|Y_1^A(\xi, b)\right|^2. \tag{3.17b}$$

Therefore, any square integrable function $f(\alpha)$ over the range $(0, \infty)$ can be written as

$$f(\alpha)=(1/\pi)\int_0^\infty Y_1^A(\alpha, x)\left|\Gamma(a)\Gamma(b)/\Gamma(c)\Gamma(a-b)\right|dx\int_0^\infty Y_1^A(\alpha', x)\left|\Gamma(a)\Gamma(b)/\Gamma(c)\Gamma(a-b)\right|f(\alpha')\,d\alpha; \tag{3.18}$$

from which we get the important relation

$$\int_0^\infty E(\alpha, x)E(\alpha, x')\,d\alpha=\delta(x-x'), \tag{3.19}$$

where we have defined

$$E(\alpha, x)=(1/\sqrt{\pi})\left|\Gamma(a)\Gamma(b)/\Gamma(c)\Gamma(a-b)\right|Y_1^A(\alpha, x). \tag{3.20}$$

The result of these manipulations can be gathered into two points:

(1) The spectrum of C_2 consists of the continuous range

$$C_2=\nu^2/2-2x^2-5. \tag{3.21}$$

(2) The orthogonal eigenfunctions [with respect to the measure $(1+y)\,dy$] are proportional to

$$Y^A(y)=y^{\nu/2}(1+y)^{-1/2+i\rho}F_{21}(a, b, c; -y). \tag{3.22}$$

The Clebsch—Gordan coefficients $\langle C_2; Ps\lambda|\nu p, p'\rangle=\langle x; P, \rho, n|\nu p, p'\rangle$ satisfying the orthogonality condition

$$\langle x''P''s''\lambda''|x', P', \rho'\lambda'\rangle=\delta^4(P''-P')\delta_{\lambda''\lambda'}\delta(\rho''-\rho')\delta(x''-x') \tag{3.23}$$

are therefore equal to

$$\langle xP\rho n|\nu p, p'\rangle=(1/\omega)\delta^4(P-p-p')b(\rho, x, \nu)\exp(-in\beta)(\sqrt{4\pi/\rho\tanh\rho})B_{n,0}^{-1/2+i\rho}(\cosh\theta)Y^{A(y,x)}. \tag{3.24}$$

Here we defined

$$B_{n,0}^{-1/2+i\rho}(\cosh\theta)=[\Gamma(+\tfrac{1}{2}+i\rho)/\Gamma(\tfrac{1}{2}+i\rho+n)]P^n_{-1/2+i\rho}(\cosh\theta),^{10}$$
$$b(\rho, x)=(\sqrt{2\pi}/4)\Gamma((\nu+1)/2+i(\rho+x))\Gamma((\nu+1)/2+i(\rho-x))/\Gamma(\nu+1)\Gamma(2ix). \tag{3.25}$$

The factor $1/\omega$ in (3.24) is a pure result of the condition (3.23), where use has been made of the completeness of intermediate states:

$$\int\frac{d^3p'}{|p'|}d^4p\,d\nu\,|\nu p, p'\rangle\langle\nu p, p'|=I. \tag{3.26}$$

As a final result we now write down the t-matrix element expansion for elastic scattering of a massless particle on a massive one with mass m and quantum number ν.

$$\langle\nu p_4, p_3|T|\nu p_1, p_2\rangle=[8\pi^2/(-u)]\int d\rho\,dx\,d(\rho, x)\rho\tanh\rho P_{-1/2+i\rho}(\cosh\theta)\left|b(\nu, \rho, x)\right|^2$$
$$\times(-m^2/u)^\nu(1-m^2/u)^{-1+2i\rho}\left|F_{21}(a, b, c; m^2/u)\right|^2, \tag{3.27}$$

in which we have defined

$$d(\rho, x)=[S(\rho, x)-1]/\rho^2\tanh^2\rho, \quad a=(\nu+1)/2+i(\rho+x), \quad b=(\nu+1)/2+i(\rho-x),$$
$$c=\nu+1, \quad \text{and} \quad \cosh\theta=(1+m^2/u)/(1-m^2/u)-[2(1+s/u)]/(1-m^2/u)^2. \tag{3.28}$$

CONCLUSIONS

By taking asymptotic energies at fixed scattering angles (exclusing the backward direction)

$$s/m^2 \gg 1, \quad -u/m^2 \gg 1, \quad \text{and} \quad (-u/s) \ \text{fixed},$$

the transition matrix (3.27) gives rise to a differential cross section which behaves like

$$\frac{d\sigma}{du} \sim \frac{1}{s^2}\left(-\frac{m^2}{u}\right)^{2\nu+2} F\left(\frac{u}{s}\right); \tag{3.29}$$

therefore, the expression

$$s^{2\nu+4}\frac{d\sigma}{du} = G\left(\frac{u}{s}\right) \tag{3.30}$$

exhibits a manifest scaling behavior. If no further information is given on the value which determine the used representation of D^+, we can only state as $\nu \geq 0$ that the differential cross section will drop off faster than s^{-4}

$$\frac{d\sigma}{du} \leq s^{-4} G\left(\frac{u}{s}\right). \tag{3.31}$$

We could, however, try to fit the value ν with the known experimental data. The appropriate adjusted expression for elastic scattering of two massive particles (no spin) is easily seen to be

$$\frac{d\sigma}{du} \sim \frac{1}{s^2}\left(-\frac{m_1^2}{u}\right)^{2\nu+2} \cdot \left(-\frac{m_2^2}{u}\right)^{2\nu+2} \cdot G\left(\frac{u}{s}\right). \tag{3.32}$$

Consider elastic proton—proton scattering. The experimental data seem to show a scaling behavior,[12] for $-t/s < 0.18$ and s ranging between 9.5 and 46.8 $(\text{GeV})^2$, going like

$$s^n \frac{d\sigma}{du} = G\left(\frac{u}{s}\right), \tag{3.33}$$

where the best fit is given a value of $n \approx 10$. This would suggest to attribute a proton to a D^+ representation with $\nu = 1$.

On the other hand, less reliable experimental data[13] on $\pi + p \to \pi + p$ show a similar scaling behavior with $n \approx 8$, which means that the pionic conformal quantum number would be $\nu = 0$. These numbers should be compared with the estimations of L. Castell,[14] who attributes the pion to a direct product representation of two photon representations and the proton to a direct product representation of a neutrino and a photon, obtained a scaling behavior for $\pi\pi \to \pi\pi$ with a value $n = 14$ (whereas we would suggest $n = 6$) and a scaling behavior for $pp \to pp$ with a value $n = 10$ (in perfect agreement with experiment). Finally, note that the appropriate differential cross-section for scattering of massless particles

$$\frac{d\sigma}{dt} \sim \frac{1}{s^2} F\left(\frac{t}{s}\right)$$

will not in general be obtained as a limit of (3.29) when putting the mass equal to zero.

ACKNOWLEDGMENTS

We wish to thank Dr. L. Castell for many useful discussion and Professor F. Cerulus for his interest and relevant suggestions.

*Navorser I.I.K.W.

[1]C. Garrod, Phys. Rev. D 1, 1524 (1970); A.O. Barut and R.B. Haugen, *Theory of the Conformally Invariant Mass* (University of Colorado Press, Boulder, 1971); J. Meckelson and J. Niederle, J. Math. Phys. 13, 23 (1972).

[2]E.D. Bloom *et al.*, Phys. Rev. Lett. 23, 930 (1969); M. Breidenbach *et al.*, Phys. Rev. Lett. 23, 935 (1969); H.W. Kendall: Report of the International Symposium on Electron and Photon Interactions, New York (1971).

[3](a) H.A. Kastrup, Ann. Physik 7 (1962); L. Castell, "Analysis of space—time structure in elementary particle physics," thesis (Oxford, 1965); Commun. Math. Phys. 17, 127 (1970); (b) Nucl. Phys. B 4, 344 (1967); B 13, 291 (1969); X 70A, 564 (1970).

[4]L. Castell, "Restrictions on inelastic channels from conformal invariance," in *De Sitter and Conformal Groups and Their Physical Applications*, edited by A.O. Barut and W.E. Brittin (Assoc. Univ. of Colorado Press, Boulder, 1971); F. Chan and H. Jones, Phys. Rev. D 10, 1321, 1356 (1974).

[5]L. Castell and K. Ringhofer, "Conformal-invariant phase shift analysis" (Proceedings of the 1970 Boulder Symposium).

[6]P. Winternitz, Preprint RPP/T/3, Rutherford Lab (1969); F. Bali, D. Coon, and A. Katz, J. Math. Phys. 10, 1939 (1969).

[7]G. Mack and I. Todorov, J. Math. Phys. 10, 2078 (1969).

[8]Tsu-Yao, J. Math. Phys. 8, 1931 (1967); 9, 1615 (1968); 12, 315 (1971).

[9]L. Castell, Phys. Rev. D 2, 1161 (1970).

[10]N.J. Vilenkin, *Fonctions speciales et théorie de la representation des groupes* (transl. from Russian) (Dunod, Paris, 1964).

[11]E.C. Titchmarsh, *Eigenfunction Expansions. I* (Oxford U.P., Oxford, 1962).

[12]D. Horn and F. Moshe, Nucl. Phys. B 48, 557 (1972); F. Halzen and M. Waldrop, Nucl. Phys. B 55, 157 (1973).

[13]P. Cornillon *et al.*, Phys. Rev. Lett. 30, 403, 1010 (1973); C. Baglin *et al.*, Phys. Rev. Lett. B 47, 85 (1973).

[14]L. Castell, Lett. Nuovo Cimento 6, 115 (1973).

Modulational instability of cnoidal wave solutions of the modified Korteweg–de Vries equation

C. F. Driscoll and T. M. O'Neil

Department of Physics, University of California, San Diego, La Jolla, California 92093
(Received 8 December 1975)

The stability of cnoidal wavetrain solutions of the modified Korteweg–de Vries equation is analyzed using Whitham's modulational theory. The cnoidal waves are solutions of an oscillator equation obtained by twice integrating the modified Korteweg–de Vries equation. The stability of the cnoidal waves is determined by the roots of the polynomial in the oscillator equation. For real roots the waves are stable, whereas for complex roots the waves are unstable.

I. INTRODUCTION

The Korteweg–de Vries (KdV) equation,[1]

$$u_t + 6uu_x + u_{xxx} = 0 \tag{1}$$

characterizes the evolution of many systems with weak dispersion and quadratic nonlinearity.[2] Likewise, the modified Korteweg–de Vries (mKdV) equation,[3,4]

$$v_t \pm 12v^2 v_x + v_{xxx} = 0 \tag{2}$$

characterizes the evolution of systems with weak dispersion and cubic nonlinearity. For example, long wavelength disturbances on a one-dimensional lattice are described by the KdV equation when the restoring forces have a small quadratic nonlinearity, and by the mKdV equation when the restoring forces have a small cubic nonlinearity.[4]

The numerical coefficients in Eqs. (1) and (2) are arbitrary, since they may be changed by a change of scale (i.e., $x \to \alpha x$, $t \to \beta t$, $u \to \gamma u$, or $v \to \gamma v$). The coefficients 6 and 12 in front of the nonlinear terms in the two equations will be convenient for our purposes. The sign in front of the nonlinear term in Eq. (1) is arbitrary since it may be changed by the transformation $u \to (-u)$. The sign in front of the nonlinear term in Eq. (2) may not be changed by a real transformation, so we include the + or − possibilities explicitly. For the case of a nonlinear lattice this + or − sign corresponds to the sign of the cubic term in the restoring force.

Exact wavetrain solutions may be obtained for both equations.[1,5] By setting $u = u(x - Ct)$ in Eq. (1) and integrating twice with respect to x, one obtains

$$\tfrac{1}{2}u_x^2 + u^3 - \tfrac{1}{2}Cu^2 - Bu + A = 0, \tag{3}$$

where A and B are constants of integration. By following the same procedure with Eq. (2), one obtains

$$\tfrac{1}{2}v_x^2 \pm v^4 - \tfrac{1}{2}Cv^2 - Bv + A = 0, \tag{4}$$

where the constants A, B, and C do not necessarily have the same values in the two equations. These equations may be viewed as oscillator equations. The variable u oscillates back and forth between two roots of the polynomial in Eq. (3), and v oscillates between two roots of the polynomial in Eq. (4). Of course, these two roots must be real and adjacent, that is, not separated by another real root. Since the polynomials are cubic and quartic respectively, the equations can be integrated in terms of elliptic functions. The solutions are often called cnoidal waves, since they can be expressed in terms of the Jacobian elliptic function cn. When the modulus of the elliptic function is much smaller than unity the cnoidal waves reduce to sinusoidal waves, and when the modulus is near unity the cnoidal waves reduce to sequences of solitons.

The question of the stability of cnoidal waves was considered by Whitham for the case of the KdV equation.[6,7] He showed that these waves are stable to long wavelength perturbations, by applying his modulational theory.

Here, we apply Whitham's modulational theory to the case of the mKdV equation. We find that the question of the stability of a particular cnoidal wave depends on the values of the constants A, B, and C for that wave. The wave is stable if the polynomial in the associated oscillator equation [i.e., Eq. (4)] has four real roots and unstable if the polynomial has two real roots and two complex roots. A cnoidal wave can exist only if at least two roots are real, since in a cnoidal wave v oscillates back and forth between two real roots. From this perspective, one can understand Whitham's conclusion of stability for cnoidal wave solutions of the KdV equation. The polynomial in Eq. (3) is a real cubic, and the existence of two real roots implies that all three roots are real. The stability criterion may be stated in its most general form for the case of the generalized Korteweg–de Vries (gKdV) equation,

$$w_t + (6w \pm 12\mu^2 w^2)w_x + w_{xxx} = 0, \tag{5}$$

where μ is an arbitrary real constant determining the relative amount of quadratic and cubic nonlinearity. A cnoidal wave solution of this equation is stable if the roots of the polynomial in the associated oscillator equation,

$$\tfrac{1}{2}w_x^2 \pm \mu^2 w^4 + w^3 - \tfrac{1}{2}Cw^2 - Bw + A = 0, \tag{6}$$

are all real, and unstable if two roots are real and two are complex.

In Sec. II, we develop Whitham's modulational theory for the case of the mKdV equation. To be specific, we develop partial differential equations governing the temporal evolution of slow spatial modulations of the three parameters determining a cnoidal wave. In Sec. III, we find the Riemann invariants for the modulational equations. When the characteristic speeds for all three Riemann invariants are real the cnoidal wave is stable, and when the characteristic speeds are complex the cnoidal wave is unstable. The characteristic speeds are expressed in terms of the roots of the polynomial in the oscillator equation, and the real or complex nature of the characteristic speeds follows from that of the roots. In Sec. IV, we extend our results to the gKdV equation. In Sec. V, we discuss the relation of our results to the Miura transformation[3] between the KdV equation and the mKdV equation. The interpretation of this transformation will be seen to depend on the choice of the sign in the mKdV equation.

It is rather surprising that one can find the Riemann invariants for the modulational equations, that is, for three nonlinear coupled partial differential equations. Apparently, this is another example of the surprising degree to which problems associated with the KdV (or mKdV) equation yield to analytic methods.

II. MODULATIONAL EQUATIONS

Following Whitham[6,7] we derive the modulational equations by averaging conservation equations over a spatial oscillation of the cnoidal wave. The first three conservation equations for the mKdV equation are[8]

$$\begin{aligned}&\frac{\partial}{\partial t}(v)+\frac{\partial}{\partial x}(\pm 4v^3+v_{xx})=0,\\&\frac{\partial}{\partial t}(v^2)+\frac{\partial}{\partial x}(\pm 6v^4+2vv_{xx}-v_x^2)=0,\\&\frac{\partial}{\partial t}(v^4\mp\tfrac{1}{2}v_x^2)+\frac{\partial}{\partial x}(\pm 8v^6+4v^3v_{xx}-12v^2v_x^2\mp v_xv_{xxx}\pm\tfrac{1}{2}v_{xx}^2)=0,\end{aligned}\tag{7}$$

where the sign choice corresponds to that of Eq. (2).

Calculation of the average of quantities appearing in these equations is facilitated by introduction of the function

$$\begin{aligned}W(A,B,C)&\equiv-\oint v_x\,dv\\&=-\sqrt{2}\oint(-A+Bv+\tfrac{1}{2}Cv^2\mp v^4)^{1/2}\,dv,\end{aligned}\tag{8}$$

where we have used Eq. (4) to find v_x for the cnoidal wave. The integral is defined to be over one complete cycle of the cnoidal wave. Since in a complete cycle v passes back and forth between two roots of the polynomial, the integral may be interpreted as a loop around the branch cut between the two roots. In terms of $W(A,B,C)$ the wavelength may be expressed as

$$\frac{1}{k}\equiv\lambda=\oint\frac{dv}{v_x}=\frac{\partial W}{\partial A}\equiv W_A,\tag{9}$$

and the average of v, v^2, and v_x^2 may be expressed as

$$\bar{v}=-kW_B,\quad\overline{v^2}=-2kW_C,\quad\overline{v_x^2}=-kW.\tag{10}$$

With the aid of Eq. (4) the average of all quantities in Eqs. (7) may be expressed in terms of the simple averages in Eqs. (9) and (10). The result is

$$\begin{aligned}&\frac{\partial}{\partial t}(kW_B)+\frac{\partial}{\partial x}(kCW_B-B)=0,\\&\frac{\partial}{\partial t}(kW_C)+\frac{\partial}{\partial x}(kCW_C-A)=0,\\&\frac{\partial}{\partial t}[k(AW_A+BW_B+CW_C-W)]\\&\quad+\frac{\partial}{\partial x}[kC(AW_A+BW_B+CW_C-W)-\tfrac{1}{2}B^2-AC]=0.\end{aligned}\tag{11}$$

Simple algebraic reduction brings these equations to the more symmetric form

$$\begin{aligned}&\frac{D}{Dt}W_A=W_A\frac{\partial C}{\partial x},\\&\frac{D}{Dt}W_B=W_A\frac{\partial B}{\partial x},\\&\frac{D}{Dt}W_C=W_A\frac{\partial A}{\partial x},\end{aligned}\tag{12}$$

where

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+C\frac{\partial}{\partial x}$$

is the convective derivative. These equations are the modulational equations for the mKdV equation. They differ from the corresponding equations for the KdV equation only in that W is defined in terms of the polynomial in Eq. (4) rather than that in Eq. (3).

III. RIEMANN INVARIANTS AND STABILITY CRITERION

We shall find that the Riemann invariants of the modulational equations take a simple form when expressed in terms of the roots of the polynomial in Eq. (4). From the relation

$$\pm v^4-\tfrac{1}{2}Cv^2-Bv+A\equiv\pm(v-a)(v-b)(v-c)(v-d)\tag{13}$$

we find that

$$\begin{aligned}&0=a+b+c+d,\\&\tfrac{1}{2}C=\mp(ab+ac+ad+bc+bd+cd),\\&B=\pm(abc+abd+acd+bcd),\\&A=\pm abcd.\end{aligned}\tag{14}$$

By replacing the variables (A,B,C) by the variables (a,b,c), with d given by the first of Eqs. (14), the modulational equations take the form

$$\begin{aligned}&W_{A,a}\frac{Da}{Dt}+W_{A,b}\frac{Db}{Dt}+W_{A,c}\frac{Dc}{Dt}\\&\quad=\mp 2W_A[(d-a)a_x+(d-b)b_x+(d-c)c_x],\\&W_{B,a}\frac{Da}{Dt}+W_{B,b}\frac{Db}{Dt}+W_{B,c}\frac{Dc}{Dt}\\&\quad=\pm W_A[(b+c)(d-a)a_x+(a+c)(d-b)b_x+(a+b)(d-c)c_x],\\&W_{C,a}\frac{Da}{Dt}+W_{C,b}\frac{Db}{Dt}+W_{C,c}\frac{Dc}{Dt}\\&\quad=\pm W_A[bc(d-a)a_x+ac(d-b)b_x+ab(d-c)c_x],\end{aligned}\tag{15}$$

where

$$\begin{aligned}&\frac{\partial W_A}{\partial a}\equiv W_{A,a}=\frac{i}{\sqrt{\pm 8}}\oint\frac{(a-d)\,dv}{[(v-a)^3(v-d)^3(v-b)(v-c)]^{1/2}},\\&W_{B,a}=\frac{-i}{\sqrt{\pm 8}}\oint\frac{(a-d)v\,dv}{[(v-a)^3(v-d)^3(v-b)(v-c)]^{1/2}},\\&W_{C,a}=\frac{-i}{\sqrt{\pm 8}}\oint\frac{(a-d)\frac{1}{2}v^2\,dv}{[(v-a)^3(v-d)^3(v-b)(v-c)]^{1/2}}.\end{aligned}\tag{16}$$

The quantities $W_{A,b}$, $W_{B,b}$, etc. are given by interchanging a and b in $W_{A,a}$, $W_{B,a}$, etc.

For the KdV case, the Riemann invariants are the various sums of roots taken two at a time (i.e., $a+b$, $b+c$, $a+c$). We now show that these quantities are Riemann invariants for the mKdV case as well. To show that $b+c$ is a Riemann invariant, we multiply the first of Eqs. (15) by $-(da+bc)(b+c)$, the second by $2(bc-da)$, the third by $-4(b+c)$, and add the three. The result is

$$\frac{i}{\sqrt{\pm 2}}\oint\left[\frac{v-a}{(v-d)^3(v-b)(v-c)}\right]^{1/2} dv \frac{D}{Dt}(b+c)$$

$$\pm 2W_A(a-b)(a-c)\frac{\partial}{\partial x}(b+c)=0, \tag{17}$$

where we have simplified the rhs of the equations with the identities

$$2(da+bc)(b+c)+2(b+c)(bc-da)-4bc(b+c)=0,$$
$$(d-b)[2(da+bc)(b+c)+2(a+c)(bc-da)-4ac(b+c)]$$
$$=-2(a-b)(b-d)(a-c)(c-d), \tag{18}$$

and the lhs with the identities

$$\frac{-(da+bc)(b+c)-2(bc-da)v+2(b+c)v^2}{[(v-a)^3(v-d)^3(v-b)(v-c)]^{1/2}}$$

$$=2\frac{d}{dv}\left[\frac{(v-b)(v-c)}{(v-d)(v-a)}\right]^{1/2},$$

$$\frac{(b-d)[-(da+bc)(b+c)-2(bc-da)v+2(b+c)v^2]}{[(v-b)^3(v-d)^3(v-a)(v-c)]^{1/2}} \tag{19}$$

$$=-2(b-d)\frac{d}{dv}\left[\frac{(v-a)(v-c)}{(v-b)(v-d)}\right]^{1/2}$$

$$+2(b-d)(c-d)\left[\frac{(v-a)}{(v-d)^3(v-b)(v-c)}\right]^{1/2}.$$

Note that b and c may be interchanged in all of these identities. Finally, we may rewrite Eq. (17) in the standard form

$$\frac{\partial}{\partial t}(b+c)+P\frac{\partial}{\partial x}(b+c)=0, \tag{20}$$

where the characteristic speed P is given by

$$P=C\pm\frac{2W_A(a-b)(a-c)}{(i/\sqrt{\pm 2})\oint[(v-a)/(v-d)^3(v-b)(v-c)]^{1/2}\,dv}$$

$$=C\pm\frac{2W_A(a-b)(a-c)}{W_A+2(d-a)(\partial/\partial d)(W_A)}. \tag{21}$$

Here, the partial derivative $\partial/\partial d$ must be taken before the first of Eqs. (14) is used to express d in terms of the other roots. By cyclic permutation of (a,b,c), one obtains the other two equations

$$\frac{\partial}{\partial t}(a+c)+Q\frac{\partial}{\partial x}(a+c)=0,$$
$$Q=C\pm\frac{2W_A(b-c)(b-a)}{W_A+2(d-b)(\partial/\partial d)(W_A)}, \tag{22}$$

and

$$\frac{\partial}{\partial t}(b+a)+R\frac{\partial}{\partial x}(b+a)=0,$$
$$R=C\pm\frac{2W_A(c-a)(c-b)}{W_A+2(d-c)(\partial/\partial d)(W_A)}. \tag{23}$$

We note that W_A may be expressed in terms of the elliptic integral of the first kind.[9] For the case of four real roots $(a>b>c>d)$, W_A is given by

$$W_A=\sqrt{8}K(r)/[(a-c)(b-d)]^{1/2}, \tag{24}$$

where $r^2=(a-b)(c-d)/(a-c)(b-d)$ for the upper sign choice in Eq. (2) and $r^2=(b-c)(a-d)/(a-c)(b-d)$ for the lower sign choice. For the case of two real roots $(c>d)$ and two complex roots $(b=a^*)$, W_A is given by

$$W_A=\sqrt{8}K(s)/\sqrt{pq}, \tag{25}$$

where

$$p^2=[c-(a+b)/2]^2-(a-b)^2/4,$$
$$q^2=[d-(a+b)/2]^2-(a-b)^2/4,$$
$$s^2=\frac{(c-d)^2-(p-q)^2}{4pq}.$$

This case is obtained only for the upper sign in Eq. (2).

To obtain the stability predictions of these equations, we follow the evolution of a small initial modulation [e.g., $b+c=b_0+c_0+(\delta b+\delta c)\cos(\kappa x)$], linearizing in the amplitude of the modulation [i.e., neglecting terms of order $(\delta b)^2$ or $(\delta c)^2$]. For the case of four real roots, the equations predict stability. For any solution, the wavelength W_A is real. Consequently, P, Q, and R are each real, and the small amplitude modulations oscillate rather than grow [e.g., $b+c=b_0+c_0+(\delta b+\delta c)\times\cos\{\kappa(x-Pt)\}$]. For the case of two real and two complex roots, the equations predict instability. Let c and d $(c>d)$ be the real roots, and a and b $(b=a^*)$ the complex roots. W_A is still real, but P and Q are complex, with $P=Q^*$. For a modulation with complex characteristic speed, either the component proportional to $\exp(i\kappa x)$ or the component proportional to $\exp(-i\kappa x)$ grows [e.g., $b+c=b_0+c_0+\frac{1}{2}(\delta b+\delta c)\exp\{i\kappa(x-Pt)\}+\frac{1}{2}(\delta b+\delta c)\exp\{-i\kappa(x-Pt)\}$]. This exponential growth will continue until the perturbation significantly modifies the wave parameters (a,b,c) and thus modifies the speeds (P,Q,R).

One could have anticipated that R is real and $Q=P^*$ from general considerations. Since the roots must occur in complex conjugates and since c and d are initially real and unequal, c and d must remain real during the initial evolution. This requires that R be real, since $a+b=-(c+d)$. Also, the evolution must preserve the relations $b=a^*$ or, since c is real, the relation $b+c=(a+c)^*$. This requires that $Q=P^*$. Of course, we could turn the argument around and show that the relations $R=R^*$ and $Q=P^*$ imply that the evolution preserves the relations $a=b^*$, $c=c^*$, $d=d^*$, and $a+b+c+d=0$.

For the case of complex characteristic speed, the equations take a more familiar form when rewritten in terms of real variables. If we let $X\equiv\mathrm{Re}(b+c)$, $Y\equiv\mathrm{Im}(b+c)$, $D\equiv\mathrm{Re}(P)$, and $E\equiv\mathrm{Im}(P)$, the real and imaginary parts of Eq. (20) are

$$\left(\frac{\partial}{\partial t}+D\frac{\partial}{\partial x}\right)X-E\frac{\partial}{\partial x}Y=0,$$
$$\left(\frac{\partial}{\partial t}+D\frac{\partial}{\partial x}\right)Y+E\frac{\partial}{\partial x}X=0. \tag{26}$$

Equation (22) leads to the same result, since $Q=P^*$. Eqs. (26) can be rewritten as the elliptic equations

$$\left(\frac{\partial}{\partial t}+D\frac{\partial}{\partial x}\right)^2 X+E^2\frac{\partial^2}{\partial x^2}X=0,$$
$$\left(\frac{\partial}{\partial t}+D\frac{\partial}{\partial x}\right)^2 Y+E^2\frac{\partial^2}{\partial x^2}Y=0, \qquad (27)$$

and it is well known that elliptic equations are unstable for Cauchy boundary conditions.

For the simple example of small amplitude cnoidal waves, P, Q, and R may be explicitly evaluated as expansions in $c-d$, as shown in the Appendix. In this small amplitude limit, stability predictions may also be obtained from mode coupling theory,[10] for comparison with the modulational results. In the Appendix, we demonstrate that the characteristic speeds P, Q, and R agree with the stability results of mode coupling theory, to first order in $c-d$.

Finally, we note that the modulational equations describe the evolution of long wavelength perturbations only. For the small amplitude example, we are able to obtain higher order dispersive corrections from mode coupling theory. It is seen in the Appendix that these corrections tend to stabilize shorter wavelength perturbations.

IV. GENERALIZED KORTEWEG-DE VRIES EQUATION

In this section, we extend the results of the previous section to the gKdV equation. The first step is to note that the mKdV equation [i.e., Eq. (2)] is transformed into the gKdV equation [i.e., Eq. (5)] by the transformation

$$v=\mu w+1/(4\mu), \quad x\to x-3t/(4\mu^2). \qquad (28)$$

Consequently, to every cnoidal wave solution of the gKdV equation there corresponds a cnoidal wave solution of the mKdV equation, and the stability (or instability) of the former may be inferred from that of the latter. By applying the same transformation to the oscillator equations for the two waves [i.e., Eqs. (4) and (6)], one can see that the roots of the polynomials in the two oscillator equations are also related by the transformation. Since this is a real transformation, we conclude that the cnoidal wave solution of the gKdV equation is stable when all four roots of the polynomial in the associated oscillator equation are real, and unstable when two roots are real and two are complex. Of course, the characteristic speeds for modulations and growth rates for instabilities are easily inferred from the transformation.

V. RELATION TO THE MIURA TRANSFORMATION

Miura's transformation[3] relates solutions of the mKdV equation, or gKdV equation, and solutions of the KdV equation. By setting $u=\pm 2v^2+\sqrt{\mp 2}\,v_x$ one can see by direct substitution that

$$u_t+6uu_x+u_{xxx}=\left(\pm 4v+\sqrt{\mp 2}\,\frac{\partial}{\partial x}\right)(v_t\pm 12v^2v_x+v_{xxx}). \qquad (29)$$

Consequently, to every solution of the mKdV equation there corresponds a (possibly complex) solution of the KdV equation. The inverse does not follow because of the operator $(\pm 4v+\sqrt{\mp 2}\,\partial/\partial x)$ on the rhs.

For the lower choice of sign, the Miura transformation is real, and the stability properties of real solutions of the two equations should correspond. Consider the mKdV equation with negative nonlinear term. One can see from the associated oscillator polynomial [i.e., Eq. (4)] that bounded, real solutions exist only if all four roots are real. Thus our stability analysis shows that all real solutions of the mKdV equation with negative nonlinear term are stable, and this corresponds to the known stability of real solutions of the KdV equation.

For the upper choice of sign, the transformation is complex. The mKdV equation with positive nonlinear term has real, unstable solutions, obtained from oscillator polynomials with two real and two complex roots. These unstable mKdV solutions transform into complex, unstable solutions of the KdV equation. Of course, Whitham's stability analysis for the KdV equation was restricted to real solutions (as is ours for the mKdV equation), so the two results need not agree under a complex transformation.

ACKNOWLEDGMENTS

We would like to thank G.B. Whitham and J.W. Miles for several enlightening discussions. This work was supported by the National Science Foundation under Grant MPS-73-05125-A01 and by the Alfred P. Sloan Foundation.

APPENDIX

In this appendix, we consider the simple example of small amplitude cnoidal waves; for clarity, we consider only the mKdV Eq. (2) with positive sign choice. We first obtain the stability results of mode coupling theory,[10] valid to first order in the wave amplitude and second order in the perturbation wavenumber κ. We then evaluate the modulational speeds P, Q, and R to first order in the wave amplitude. The two theories are seen to agree for long wavelength perturbations. For shorter wavelength perturbations, mode coupling theory gives corrections which tend to stabilize the growth of these components.

The small amplitude cnoidal wave is approximated by a mean value β, a fundamental mode A_1, and a single harmonic A_2. The perturbations are seen as sideband modes, with wavenumbers differing from the main modes by κ,

$$\begin{aligned}v(x,t)=\beta&+A_1\exp(ikx)+A_2\exp(2ikx)+A_+\exp(i\kappa x)\\&+A_{1-}\exp[i(k-\kappa)x]+A_{1+}\exp[i(k+\kappa)x]\\&+A_{2-}\exp[i(2k-\kappa)x]+A_{2+}\exp[i(2k+\kappa)x]\\&+\text{complex conjugate.}\end{aligned} \qquad (A1)$$

The evolution of a modal amplitude is determined by the appropriate spatial Fourier component of the mKdV equation. For component k this gives

$$\frac{\partial A_1}{\partial t}+i\,\omega_1A_1+12ik(|A_1|^2A_1+2\beta A_2A_1^*)=0, \qquad (A2)$$

where the linear frequency is $\omega_1 = -k^3 + 12\beta^2 k$. Solving the analogous evolution equations for the driven modes near $2k$ gives $A_2 = 4\beta A_1^2/k^2$, $A_{2-} = 8\beta A_1 A_{1-}/k^2$, $A_{2+} = 8\beta A_1 A_{1+}/k^2$. The nonlinear frequency of mode k is then seen to be $\Omega_1 = \omega_1 + 12k|A_1|^2(1 + 8\beta^2/k^2)$. The coupled evolution equations for the remaining perturbations are

$$\frac{\partial A_+}{\partial t} + i\omega_+ A_+ + 12i\kappa(2\beta A_{1+}A_1^* + 2\beta A_1 A_{1-}^*) = 0,$$

$$\frac{\partial A_{1-}}{\partial t} + i\omega_{1-}A_{1-} + 12i(k-\kappa)(2\beta A_{2-}A_1^* + 2\beta A_2 A_{1+}^* + 2\beta A_1 A_+^* + A_1^2 A_{1+}^* + 2|A_1|^2 A_{1-}) = 0, \quad \text{(A3)}$$

$$\frac{\partial A_{1+}}{\partial t} + i\omega_{1+}A_{1+} + 12i(k+\kappa)(2\beta A_{2+}A_1^* + 2\beta A_2 A_{1-}^* + 2\beta A_1 A_+ + A_1^2 A_{1-}^* + 2|A_1|^2 A_{1+}) = 0.$$

We take $A_1 \propto \exp(-i\Omega_1 t)$, $A_+ \propto \exp(-i\nu t)$, $A_{1-} \propto \exp(-i\Omega_1 t + i\nu t)$, $A_{1+} \propto \exp(-i\Omega_1 t - i\nu t)$, and solve Eqs. (A3) for the three roots ν. Two roots are seen to be near $\nu \approx \omega'\kappa = (-3k^2 + 12\beta^2)\kappa$; this approximation can be used to solve for $A_+ = -8\beta(A_{1+}A_1^* + A_1 A_{1-}^*)/k^2$. The resulting second order secular equation is

$$(\nu - \omega'\kappa)^2 = 12k|A_1|^2(1 - 8\beta^2/k^2)\omega''\kappa^2 + (\tfrac{1}{2}\omega''\kappa^2)^2, \quad \text{(A4)}$$

where $\omega'' = -6k$. A similar procedure gives the third root

$$\nu = 12\beta^2\kappa. \quad \text{(A5)}$$

The perturbation grows exponentially when one of the roots is complex, i.e., when the rhs of Eq. (A4) is negative.

We now evaluate P, Q, and R to order $c-d$, where $c > v > d$. To this order, Eqs. (24) and (25) for W_A are equivalent; we use Eq. (24) for simplicity,

$$W_A = \sqrt{2}\pi\,[(a-c)(b-d)]^{-1/2}(1 + r^2/4 + 9r^4/64).$$

Expressing all quantities in terms of $(a, c, c-d)$ gives

$$\frac{\partial}{\partial d}\ln(W_A) = \frac{c}{(a-c)(a+3c)} - \frac{(c-d)(10a^2 + 42c^2 + 4ac)}{16(a-c)^2(a+3c)^2},$$

$$k^2 \equiv \left(\frac{2\pi}{W_A}\right)^2 = -2(a-c)(a+3c) - 2(c-d)(3c-a),$$

$$|A_1| = \tfrac{1}{4}(c-d),$$

$$\beta = c - \tfrac{1}{2}(c-d).$$

The characteristic speeds may then be expressed as

$$\begin{aligned} P &= 6a^2 + 12ac - 6c^2 + (c-d)(3c-9a) \\ &= -3k^2 + 12\beta^2 - 6\sqrt{2}\,|A_1|(8\beta^2 - k^2)^{1/2}, \\ Q &= 6a^2 + 12ac - 6c^2 + (c-d)(9c-3a) \\ &= -3k^2 + 12\beta^2 + 6\sqrt{2}\,|A_1|(8\beta^2 - k^2)^{1/2}, \\ R &= 12c^2 - 12c(c-d) \\ &= 12\beta^2. \end{aligned} \quad \text{(A6)}$$

The two speeds P and Q, when multiplied by κ, correspond to the two roots ν in Eq. (A4); similarly, R times κ corresponds to the third root in Eq. (A5). The term $(\frac{1}{2}\omega''\kappa^2)^2$ in Eq. (A4) is a dispersive correction not found in modulational theory, and it decreases the instability for perturbations with large κ. Indeed, the small amplitude wavetrain is stable with respect to perturbations satisfying $\kappa^2 \geq 8|A_1|^2(1 - 8\beta^2/k^2)$. Thus modulational theory, valid for small κ, agrees with mode coupling theory, valid for small amplitude, in their range of overlap. Furthermore, mode coupling theory indicates that shorter wavelength perturbation tend to be stabilized.

[1]D. J. Korteweg and G. de Vries, Philos. Mag. 39, 422 (1895).
[2]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974), pp. 13, 460.
[3]R. M. Miura, J. Math. Phys. 9, 1202 (1968).
[4]N. J. Zabusky, "A Synergetic Approach to Problems of Nonlinear Dispersive Wave Propagation and Interaction," in *Proceedings of the Symposium on Nonlinear Partial Differential Equations*, edited by W. Ames (Academic, New York, 1967), p. 223.
[5]B. B. Kadomtsev and V. I. Karpman, Sov. Phys. Usp. 14, 40 (1971).
[6]G. B. Whitham, Proc. R. Soc. London A 283, 238 (1965).
[7]G. B. Whitham, Ref. 2, p. 565.
[8]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. 9, 1204 (1968).
[9]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York 1965), pp. 240, 242.
[10]G. B. Witham, Ref. 2, p. 527.

Boundary conditions and singular potentials in diffusion theory

H. Narnhofer* and J. R. Klauder

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 8 December 1975)

The Feynman–Kac integral and its associated stochastic differential equations are generalized to incorporate the analog of arbitrary boundary conditions in the equivalent differential equation formulation of diffusion problems. The influence of additional singular potentials is considered, especially the interplay between the boundary conditions and various regularizations of the singular potential. The basic results are translated into d-dimensional processes and compared with standard results known for differential operators.

I. INTRODUCTION

In Ref. 1 the effect of singular potentials in the diffusion in one-dimension was studied. We want to generalize the results to the diffusion in the half-space with different boundary conditions. The purpose is twofold. On one hand we have the feeling that it gives a deeper insight to the effect of the regularization. On the other hand it leads to the diffusion problem in higher dimensions with a rotation invariant potential. It is a well known fact that half the radial part of the Laplace operator

$$\frac{1}{2}\left(\frac{d^2}{\partial r^2}+\frac{d-1}{r}\frac{\partial}{\partial r}\right)$$

acting in the Hilbert space $L^2(R^d)$ corresponds to an operator

$$\frac{1}{2}\frac{\partial^2}{\partial r^2}-\frac{\lambda}{r^2}$$

with adjusted boundary conditions for $\psi(r)\in L^2(R^+)$. Namely, with $\lambda=(d-1)(d-3)/8$ for

$$d=1,\quad \lambda=0,\quad \psi'(0)=0,$$

$$d=2,\quad \lambda=-\tfrac{1}{8},\quad \psi(0)=0,$$

$$d=3,\quad \lambda=0,\quad \psi(0)=0,$$

$$d=4,\quad \lambda=\tfrac{3}{8},\quad \psi(0)=0.$$

It turns out that all these values of λ are critical such that the diffusion process changes significantly, corresponding to the fact that recurrence time and probability for a particle to escape to infinity are different in different dimensions.

We can reproduce also some properties that are already known for operators and forms. We consider the class of operators defined by the boundary condition $\gamma\psi(0)=\psi'(0)$. Varying the boundary condition means that we are closer to the operator framework since for forms the boundary conditions are generally regarded as fixed.

Following Ref. 1 very closely we will use several methods to treat the possible regularizations, mainly differential equation techniques and the path space viewpoint. Comparing singularities that are due to the presence of singular potentials to those arising from the boundary condition we will conclude: For $-\infty<\gamma<\infty$ potentials with a singularity of the order $r^{-\alpha}$, $\alpha<1$ leads to an unambiguous diffusion process, for $1\leq\alpha\leq 2$ we obtain several diffusion processes, mixing between different boundary conditions.

For $\gamma=\infty$ the diffusion process is unique for $\alpha<2$. For $\alpha=2$ it depends on the strength of the potential. With $\lambda\geq\frac{3}{8}$ the diffusion process is unique and for $-\frac{1}{8}\leq\lambda<\frac{3}{8}$ we have various choices, regardless of the value of γ. For $\lambda<-\frac{1}{8}$ our methods do not give any information.

For $\alpha>2$ positive potentials give a unique diffusion process, but negative potentials cannot be treated by our methods. We want to translate this result into d-dimensional diffusion processes. $d=1$ corresponds to $\gamma=0$ and we obtain processes depending on regularization already for $\alpha\geq 1$.

For $d=2,3$ we can insist that $\gamma=\infty$. Therefore for $\alpha<2$ we have only one diffusion process. This result corresponds to the existence of the Friedrichs extension of the operator. But, as we need not necessarily choose the Friedrichs extension for the operator, we can vary the boundary condition, and this will give us different solutions for the diffusion process.

For $d\geq 4$ the answer is again different. We know that the operator is already essentially self-adjoint on the C^∞ functions. Analogously, in the diffusion problem the ever present potential λ/r^2 with $\lambda\geq\frac{3}{8}$ selects the boundary condition $\gamma=\infty$—otherwise the process would not be well defined—so that the diffusion process is always unique for additional singular potentials $r^{-\alpha}$, for $\alpha<2$. If $\alpha=2$, then the region of uniqueness depends on the strength and dimension. If $\alpha>2$, then the dimension plays no role in the existence and uniqueness problem for the diffusion.

II. CONTINUITY IN THE BOUNDARY CONDITION

Following Ref. 2 we define reflecting Brownian motion by considering the reflected Brownian path

$$X^+(t)=|X(t)|$$

with corresponding transition density

$$\rho_0(x,y,t)=(1/\sqrt{2\pi t})[\exp(-(x-y)^2/2t)+\exp(-(x+y)^2/2t)].$$

Let $t^*(x,T)$ be the local time at the point x defined formally as

$$t^*(x,t)=\int_0^t \delta(X(s)-x)\,ds,$$

then we can consider elastic Brownian motion characterized by the transition probability

$$\rho_\gamma(x,y,t)=E[\exp(-\gamma t^*(0,t))\,|X^*(t)=y]\rho_0(x,y,t).$$

The corresponding differential equation is the heat flow problem

$$\frac{\partial u}{\partial t}=\frac{1}{2}\frac{\partial^2 u}{\partial x^2}$$

with boundary condition

$$\gamma u(t,0)=\frac{\partial}{\partial x}u(t,0).$$

Reference 2 restricts its interest to $\gamma>0$ though the considerations remain true also for negative γ. Of course for negative γ the killing interpretation of these authors is not applicable any more. We want to consider continuity properties in γ, i.e., if

$$\lim_{\gamma\to\pm\infty} u_\gamma(t,x)=\lim_{\gamma\to\pm\infty}E_{\gamma,x}[f(x_t)]$$
$$=\lim_{\gamma\to\pm\infty}E_{0,x}[\exp(-\gamma t^*(0,t))f(x_t)]$$

exists. Calculating the limit we use the Laplace transform

$$\tilde u_\gamma(\alpha,x)=E_{\gamma,x}\left[\int_0^\infty dt\,\exp(-(\alpha/2)t)f(x_t)\right].$$

It exists provided $\gamma>0$ and $\mathrm{Re}\,\alpha>0$ or $\gamma<0$ and $\mathrm{Re}\,\alpha>|\gamma|^2$.

Let us concentrate first on positive γ:

$$\lim_{\gamma\to+\infty}u_\gamma(t,x)=\lim_{\gamma\to\infty}\int_{-\infty}^{+\infty}\exp[(\nu+i\mu)t/2]\,d\mu\,\tilde u_\gamma(\nu+i\mu,x)$$
$$=\lim_{\gamma\to\infty}\int_{-\infty}^{+\infty}\exp[(\nu+i\mu)t/2]\,d\mu$$
$$\times\Bigg\{\int_0^\infty[\exp(-\sqrt{\nu+i\mu}\,|y-x|)$$
$$+\exp(-\sqrt{\nu+i\mu}\,|y+x|)]\frac{f(y)}{\sqrt{\nu+i\mu}}dy$$
$$+\exp(-\sqrt{\nu+i\mu}\,x)\Bigg[\tilde u_\gamma(\nu+i\mu,0)-2$$
$$\times\int_0^\infty\frac{\exp(-\sqrt{\nu+i\mu}\,y)}{\sqrt{\nu+i\mu}}f(y)\,dy\Bigg]\Bigg\},\quad \nu>0,$$

where

$$\tilde u_\gamma(\nu+i\mu,0)=\frac{2}{\sqrt{\nu+i\mu}+\gamma}\int_0^\infty\exp(-y\sqrt{\nu+i\mu})f(y)\,dy.$$

We concentrate our interest on the second term which is the only γ-dependent one,

$$\lim_{\gamma\to\infty}\int_{-\infty}^{+\infty}d\mu\,\exp[(\nu+i\mu)t/2]\frac{2\gamma}{\sqrt{\nu+i\mu}\,(\gamma+\sqrt{\nu+i\mu})}$$
$$\times\int_0^\infty\exp(-y\sqrt{\nu+i\mu})f(y)\,dy.$$

Assume $f(y)=\delta(y-y_0)$ (which is equivalent to changing the order of limit and integration and that is allowed if the limit is obtained uniformly in y). To calculate

$$\lim_{\gamma\to\infty}\int_{-\infty}^{+\infty}d\mu\,\exp[(\nu+i\mu)t/2]\frac{2\gamma}{\sqrt{\nu+i\mu}(\gamma+\sqrt{\nu+i\mu})}$$
$$\times\exp(-y\sqrt{\nu+i\mu})$$

we change the path of integration according to Fig. 1. Therefore, only the integration around the cut contributes which can be written with $b=\sqrt{-i\mu-\nu}$,

$$\int_{-\infty}^{+\infty}2\gamma\,db\frac{\exp(-b^2t/2-iby)}{(\gamma+ib)}\quad\underset{\gamma\to\infty}{\to 2}\quad\underset{\gamma\to 0}{\to 0}.$$

Therefore

$$\lim_{\gamma\to 0}u_\gamma(x,t)=u_0(x,t),$$
$$\lim_{\gamma\to\infty}u_\gamma(x,t)=\int_0^\infty\frac{1}{\sqrt{2\pi t}}$$
$$\times[\exp(-(x-y)^2/2t)-\exp(-(x+y)^2/2t)]$$
$$\times f(y)\,dy,$$

corresponding to absorbing Brownian motion.

We turn to the limit $\gamma\to-\infty$. Here the Laplace transform exists only for $\mathrm{Re}\,\alpha>|\gamma|^2$, therefore $\sqrt{\nu+i\mu}$ has to be replaced by $\sqrt{\gamma^2+\nu+i\mu}$ and γ by $-\gamma$. We consider again the contribution of the second term:

$$\lim_{\gamma\to\infty}\int_{-\infty}^{+\infty}\exp[(\gamma^2+\nu+i\mu)t/2]\,d\mu\frac{\gamma}{(\gamma^2+\nu+i\mu)^{1/2}-\gamma}$$
$$\times\exp(-y\sqrt{\gamma^2+\nu+i\mu})\frac{1}{(\gamma^2+\nu+i\mu)^{1/2}}.$$

Again we change the integration path in the previous manner but have to take into account that a pole is contained in the integration region; see Fig. 2. The integration around the cut is essentially the same,

$$\int_{-\infty}^{+\infty}2\gamma\,db\frac{\exp(-b^2t/2-iby)}{-ib+\gamma}\quad\underset{\gamma\to\infty}{\to 2}.$$

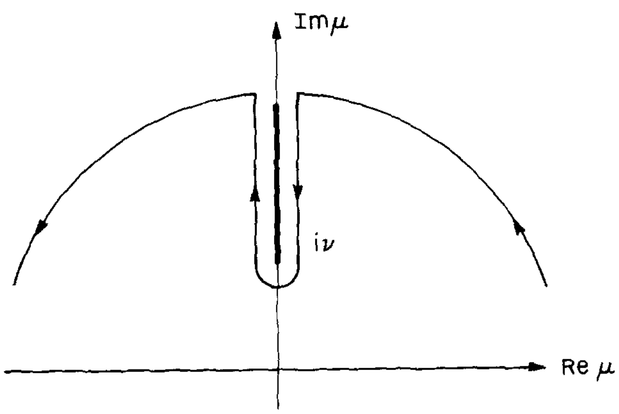


FIG. 1. Path of integration relevant to calculating the integral. There is a cut beginning at $\mu=i\nu$. The pole lies in a different sheet.

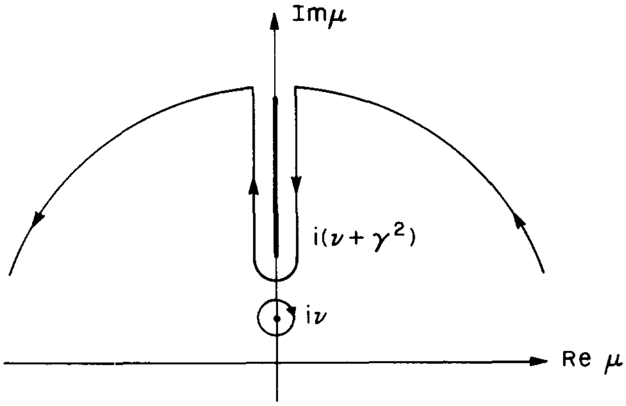


FIG. 2. Path of integration relevant to the integral. There is a cut beginning at $\mu=i(\nu+\gamma^2)$. The pole is on the first sheet at $\mu=i\nu$.

But the contribution of the pole diverges for $\gamma\to\infty$,

$$\gamma\,\exp(\gamma^2 t/2-\gamma y)\quad \underset{\gamma\to\infty}{\to\infty}\quad \underset{\gamma\to 0}{\to 0}.$$

Therefore we have no limit for elastic Brownian motion in this direction.

It seems worthwhile to consider whether we have the same behavior if we examine the corresponding operators. The operators are defined on different domains so we can only compare the resolvents of the corresponding self-adjoint operators. For the resolvents and fixed $\alpha=\nu+i\mu$ evidently the operators converge for $\gamma\to\pm\infty$ to the same operator and the same is true for the unitary operators, where we have only to replace t by it in our calculations. Especially the contribution of the second term vanishes for all $y\neq 0$, therefore also for all functions $\psi(y)$ with $\psi(0)=0$. But these functions are dense in the Hilbert space, therefore the unitary operators $\exp(iH_\gamma t)$ converge strongly (compare the general Theorem in Ref. 3, Chap. VIII, pp. 20, 21).

We turn now to the behavior of the form. The form corresponding to H_γ is

$$h(\psi,\psi)=\tfrac{1}{2}\int_0^\infty dx\left|\frac{d\psi}{dx}\right|^2+\gamma\psi(0)^2$$

and the form domain is the same for all $\gamma\neq\pm\infty$. Evidently these forms do not converge on the whole domain either for $\gamma\to+\infty$ or for $\gamma\to-\infty$. But they converge (and coincide) on the smaller domain of functions satisfying $\psi(0)=0$. This is the form domain corresponding to the self-adjoint operator H_∞. Again a general theorem[4] tells us that this is sufficient for the strong resolvent convergence of the operators.

III. BOUNDARY CONDITIONS AND ADDITIONAL POTENTIALS

To every boundary condition (except $\gamma=-\infty$) there belongs a probability measure $d\mu_\gamma$ on the path space and we can consider now the influence of an interaction potential V along with different boundary conditions.

Defining the Brownian functional

$$\hat V(t)=\int_0^t V[X(s)]\,ds,$$

we consider the expectation functional

$$u(t,x)=E^{\lambda V}_{x,\gamma}[f(x_t)]=E_{x,\gamma}[\exp(-\lambda\hat V(t))f(x_t)]$$
$$=E_{x,0}[\exp(-\gamma t^*(0,t)-\lambda\hat V(t)f(x_t)]$$

and its Laplace transform

$$\tilde u(\alpha,x)=E_{x,0}\Big[\int_0^\infty\exp(-\alpha t-\gamma t^*(0,t)-\lambda\hat V(t))\times f(x_t)\,dt\Big].$$

Assume that constants c_1 and c_2 exist such that $c_2\geq V(x)\geq c_1$. Then the Laplace transform exists for all α with $\mathrm{Re}\,\alpha>\mathrm{Max}(0,-\gamma|\gamma|,-\gamma|\gamma|-c_1)$ and

$$\tilde u(\alpha,x)-(\alpha-H_\gamma)^{-1}f=(\alpha-H_\gamma)^{-1}\lambda V\tilde u,$$

i.e.,

$$(H_\gamma+\lambda V-\alpha)\tilde u=f.$$

So $u(t,x)$ satisfies the differential equation

$$\frac{\partial u}{\partial t}=\frac{1}{2}\frac{\partial^2 u}{\partial x^2}-\lambda V u.$$

We check that the boundary conditions are satisfied,

$$\tilde u'(\alpha,0)=[(\alpha-H_\gamma)^{-1}f]'+[(\alpha-H_\gamma)^{-1}\lambda V\tilde u]'$$
$$=\gamma(\alpha-H_\gamma)^{-1}f+\gamma(\alpha-H_\gamma)^{-1}\lambda V\tilde u$$
$$=\gamma\tilde u(\alpha,0).$$

We had to assume V bounded so that $(\alpha-H_\gamma)^{-1}V\tilde u$ is well defined.

We ask again about the continuity properties in γ as well as in the coupling parameter λ. For fixed γ, $\tilde u(\alpha,x)$ and $u(t,x)$ are differentiable in λ. But they are also jointly continuous in γ and λ, independent of the sign of γ and λ. Considering the limit $\gamma\to\pm\infty$ we do not have the explicit structure to deal with. Resolvent convergence can easily be seen using the series

$$(H_\gamma+\lambda V-\alpha)^{-1}=(H_\gamma-\alpha)^{-1}\sum_{n=0}^{\infty}[\lambda V(H_\gamma-\alpha)^{-1}]^n,$$

which converges uniformly for $(H_\gamma-\alpha)^{-1}<1/\lambda|V|$. Due to the gap between the continuous spectrum and the eigenvalue, we can easily find an α_0 such that the inequality is satisfied for all γ with $|\gamma|>\gamma_0(\lambda V)$ and all $\alpha=\alpha_0+i\beta$. Nevertheless the convergence will not be uniform in α_0 just as was the case in the absence of a potential. So adding a bounded potential to the Hamiltonian does not change the continuity properties with respect to the boundary condition.

IV. SINGULAR POTENTIALS

So far we have only considered the effect of bounded potentials. But the main purpose of this paper is to discuss singular potentials, especially if different boundary conditions can weaken the influence of a singularity. We expect also that continuity properties with respect to the boundary condition will change, if we consider potentials with their singularity at the origin (other singularities are completely analogous to the case $\gamma=0$ independent of the chosen boundary condition), because the wavefunction may be singular at the origin so that the condition $\gamma u(0)=u'(0)$ does not make sense.

We will regularize our potentials by bounded potentials $V_\epsilon(x)$ satisfying $\lim_{\epsilon\to0}V_\epsilon(x)=V(x)$ pointwise with the possible exception of $x=0$. For these potentials the measure $d\mu_{\gamma\epsilon}$ on the path space is well defined and we will consider whether there exists one or more weak limit points of these measures, how they depend on γ, and whether varying γ together with ϵ can lead to new results.

We have different possibilities to proceed. We can examine directly the differential equation, but this is only useful if the solution is a well known function, where we can discuss the ϵ dependence explicitly. On the other hand we can try to discuss $E_{x,y}^{\lambda V_\epsilon}$ directly which can only be done conveniently if the potential is not too singular. The last, most general possibility, is to consider how the paths themselves are changed by the influence of the potential.

A. The differential equation

Let us consider the example $V(r)=1/r$ with the regularization $V_\epsilon(r)=1/(r+\epsilon)$. Then the solutions of the differential equation

$$\left(-\frac{1}{2}\frac{d^2}{dr^2}+\frac{\lambda}{r+\epsilon}+\alpha\right)\varphi(\alpha,r)=0$$

are the Whittaker functions $M_{k,1/2}(r+\epsilon)$ and $W_{k,1/2}(r+\epsilon)$ with $k=-\lambda/\sqrt{2}\alpha$.

We have to consider the ϵ and γ dependence of $\tilde{u}(\alpha,a)$. The general theory[2,5] tells us that

$$\tilde{u}(\alpha,a)=\int dr f(r)\varphi_r(\alpha,a),$$

where $\varphi_r(\alpha,a)$ satisfies the above differential equation for $a\lessgtr r$ with the additional condition

$$\lim_{\epsilon\to0}\varphi_r(\alpha,r-\epsilon)=\lim_{\epsilon\to0}\varphi_r(\alpha,r+\epsilon),$$

$$\lim_{\epsilon\to0}\varphi_r'(\alpha,r-\epsilon)=\lim_{\epsilon\to0}\varphi_r'(\alpha,r+\epsilon)+2,$$

provided the Laplace transform exists, which is guaranteed if $\lambda\geq0$. In order that the above integral exists $\varphi_r(\alpha,a)$ has to vanish for $r\to\infty$ and it has to satisfy the boundary conditions for $a=0$.

The behavior of $M_{k,1/2}$ and $W_{k,1/2}$ at the origin is

$$M_{k,1/2}(a+\epsilon)=a+\epsilon+O((a+\epsilon)^2)$$

$$\begin{aligned}W_{k,1/2}(a+\epsilon)=&[1/\Gamma(1-k)][1-(a+\epsilon)/2-k(a+\epsilon)\\&\times\{\psi(1-k)-2\psi(1)-1\}-k(a+\epsilon)\log(a+\epsilon)\\&+O((a+\epsilon)^2)\log(a+\epsilon)]\end{aligned}$$

with $\psi(\cdot)$ the digamma function.

The ratio of the contribution of $M_{k,1/2}$ and $W_{k,1/2}$ depends on γ. If we suppose γ is fixed, then this ratio tends to a limit as $\epsilon\to0$, namely only M contributes, independent of the γ we start with. On the other hand we can imagine that this ratio f is fixed and γ varies as a function of ϵ,

$$\gamma(\epsilon)\approx+k\Gamma(1-k)\,|\log\epsilon|.$$

Therefore γ tends to $+\infty$ or $-\infty$, depending whether the potential is attractive or repulsive. Of course the limit $\gamma\to-\infty$ is dangerous (since we have no convergence of the unperturbed measure) and must not be included, the other limit is already excluded because we can only consider repulsive potentials. Let us assume γ fixed such that f is an equicontinuous function of γ and ϵ with respect to k, i.e., with respect to λ and α. We want to check whether this continuity is preserved for $\tilde{u}$. We have to calculate φ, knowing that only $W_{k,1/2}$ decays at infinity. Therefore

$$\begin{aligned}\varphi_r(\alpha,a)=&\theta(r-a)A[M_{k,1/2}(a+\epsilon)+fW_{k,1/2}(a+\epsilon)]\\&+\theta(a-r)BW_{k,1/2}(a+\epsilon),\end{aligned}$$

A and B satisfying

$$A[M_{k,1/2}(r+\epsilon)+fW_{k,1/2}(r+\epsilon)]=BW_{k,1/2}(r+\epsilon),$$

$$A[M_{k,1/2}'(r+\epsilon)+fW_{k,1/2}'(r+\epsilon)]=BW_{k,1/2}'(r+\epsilon)+2,$$

or

$$A=\frac{2W}{M'W-MW'},$$

$$B=\frac{2[M+fW]}{M'W-MW'}.$$

Evaluating $\tilde{u}(\alpha,a)$ we notice that W and M have no singularities in the considered region. So the only singularity that could lead to trouble belongs to $M'W-MW'=0$ independent of f or γ. But for $\gamma=+\infty$ (which corresponds to the $1/r$ potential in three dimensions) we know that the resolvent is bounded and the strong limit of the resolvent corresponding to regularized potentials, so A and B are well defined and this singularity does not exist.

Taking into account that for fixed γ and $\lambda\geq0$ the Hamiltonian is bounded from below we can conclude that for fixed γ the measures $d\mu_{\gamma\epsilon}^\lambda$ converge weakly for $\epsilon\to0$ to a unique measure $d\mu_\infty^\lambda$ [since $\lim_{\epsilon\to0}f(\gamma,\epsilon)=0$] and if we now consider the limit $\lambda\to0$ then $d\mu_\infty^\lambda$ converges to absorbing Brownian motion.

B. The expectation functional

We turn now to estimates that are more tailormade for the diffusion problem since we deal directly with the expectation functional. Following the ideas of Ref. 1 we have to evaluate

$$u(t,x)=E_x[\exp(-\gamma t^*(0,t)-\lambda\int_0^\infty V_\epsilon(y)t^*(y,t)\,dy)f(x_t)],$$

where $V_\epsilon(y)$ is some regularization of the singular potential $V(y)$, not yet specified.

As in Ref. 1, we can estimate

$$\begin{aligned}\int_{y<1}V_\epsilon(y)t^*(y,t)\,dy=&\int_{y<1}V_\epsilon(y)[t^*(y,t)-t^*(0,t)]\,dy\\&+t^*(0,t)\int_{y<1}V_\epsilon(y)\,dy\end{aligned}$$

and use such a regularization that the last integral tends to a finite limit as $\epsilon \to 0$. As an estimate for the first term we use the fact that for all $\eta < \frac{1}{2}$ there exists a nonnegative random variable R_η with

$$|t^*(y,t) - t^*(0,t)| \le y^\eta R_\eta.$$

If therefore the potential is less singular then $r^{-3/2}$ we see that a limit is obtained as $\epsilon \to 0$ and this limit $d\mu_\gamma^\lambda$ is continuous in γ and λ and equivalent to $d\mu_{\bar\gamma}$, as long as $\gamma \ne \infty$ and $\bar\gamma \ne \infty$. For $\gamma \to +\infty$ it tends to a limit $d\mu_\infty^\lambda$ which is equivalent to $d\mu_\infty$. Nevertheless $d\mu_\gamma^\lambda$ will depend on the special choice of regularization, i.e., on $\lim_{\epsilon\to 0}\int V_\epsilon(y)\,dy$. The only exception is absorbing Brownian motion. Here the expectation of $t^*(0,t) = 0$, since paths that come to the origin do not contribute. Therefore the special value of the limit cannot change the result and we can even omit the restriction on the regularization, so that it must be finite. This uniqueness of the regularization corresponds to the well known fact that in two and higher dimensions potentials smoother than $r^{-3/2}$ are Kato-tiny and give therefore a unique self-adjoint Hamiltonian.[6]

We ask for the existence of the Laplace transform which corresponds to the boundedness from below of the Hamiltonian. We have the following contributions.

(a) $\lambda \int_{y>1} V_\epsilon(y) t^*(y,t)\,dy$

This can give a shift to α proportional to λ, corresponding to a bounded potential.

(b) $\lambda t^*(0,t) \int_{y<1} V_\epsilon(y)\,dy$.

This is a term similar to the boundary condition and we know already that such a term gives a shift proportional to λ^2.

(c) $\lambda \int_{y<1} V_\epsilon(y)[t^*(y,t) - t^*(0,t)]\,dy$.

Here we use again the estimate with R_η with the additional information that $R_\eta < [t^*(y,t)]^{1/2}$. Therefore, this last term only rises with $\sqrt{t}$ so that we know that $-\alpha t$ will always win provided Reα is big enough and the Laplace transform exists.

Finally we want to find the connection between the two different approaches for the $1/r$ potential. Evidently our regularization of Sec. IV, Part A does not satisfy the requirements of Part B, namely $\int_{y<1} V_\epsilon(y)\,dy$ does not converge. For $\gamma = \infty$ this requirement was not essential and both approaches lead to the same result. But for $\gamma \ne \infty$ this is no longer true and $\int_{y<1} V_\epsilon(y)\,dy$ tends to infinity for $\epsilon \to 0$. But in Part A we had to restrict our interest to positive λ and in this case also in our present approach the integral converges to the one corresponding to absorbing Brownian motion.

V. THE PATH SPACE VIEWPOINT

As already mentioned the most general information about the influence of singular potentials can be obtained by considering the paths directly. The general theory can be found in Ref. 7 and the references cited there. We recall the main facts: Consider the two normalized measures $d\mu_Y$ and $d\mu_W$ on the path space in one dimension with $d\mu_W$ corresponding to free Brownian motion and $d\mu_Y$ defined by the Radon–Nikodym derivative

$$\frac{d\mu_Y}{d\mu_W} = N(T)\exp\left[-\lambda \int_0^T V(x(t))\,dt\right].$$

Then the path Y starting at c at $t=0$ is obtained by the stochastic differential equation

$$dY = a(Y(t),t)\,dt + dW(t),$$

with

$$a(x,t) = \frac{\partial B(x,t)/\partial x}{B(x,t)}$$

and $B(x,t)$ satisfying

$$\frac{\partial B}{\partial t} = -\frac{1}{2}\frac{\partial^2 B}{\partial x^2} + \lambda V(x) B,$$

$$B(x,T) \equiv 1$$

uniquely, provided $a(x,s)$ is continuous and $|a(x,s)|^2 \le \text{const}(1+|x|^2)$ for $0 \le s \le T$. Furthermore $N(T) = B(c,0)^{-1}$.

In order to use this general theory we have to introduce a connection between diffusion in one dimension and diffusion in half-space; namely a path $\tilde{Y}(c,t)$ in half-space starting at point $c>0$, can be associated with a path $Y(c,t)$ in one dimension according to

$$\tilde{Y}(c,t) = |Y(c,t)|.$$

A. Boundary condition

We consider the path $Y_\gamma(c,t)$ in one dimension. Formally the differential operator can be written as

$$-\frac{1}{2}\frac{d^2}{dx^2} + \gamma\delta(x)$$

and the δ function can be approximated by regularized potentials $\Delta_\delta(x)$. Define $B_\delta(x,t)$ as the solution of

$$\frac{\partial B_\delta}{\partial t} = -\frac{1}{2}\frac{\partial^2 B_\delta}{\partial x^2} + \gamma\Delta_\delta(x) B_\delta,$$

with $B_\delta(x,T) \equiv 1$. Then for $\delta \to 0$, B_δ converges to B_γ satisfying

$$\frac{\partial B_\gamma}{\partial t} = -\frac{1}{2}\frac{\partial^2 B_\gamma}{\partial x^2}, \quad x \ne 0,$$

$$\gamma B_\gamma(0,t) = \frac{\partial B_\gamma}{\partial x}(0,t), \quad t \ne T,$$

$$B_\gamma(x,T) = 1, \quad x \ne 0.$$

Explicitly

$$\begin{aligned} B_\gamma(x,t) &= \int_0^\infty \frac{dy}{\sqrt{2(T-t)}}[\exp(-(x-y)^2/2(T-t)) \\ &\quad + \exp(-(x+y)^2/2(T-t))] + \int_0^\infty dy \\ &\quad \times \int_{-\infty}^{+\infty} dk \frac{2i\gamma}{k-i\gamma}\exp[ik(x+y) - (k^2/2)(T-t)] \\ &\quad + \theta(-\gamma)\sqrt{2|\gamma|}\exp[-|\gamma|x + (\gamma^2/2)(T-t)] \end{aligned}$$

for $x > 0$ and $B(x) = B(-x)$ which can be obtained by using the generalized eigenfunctions of the differential equation. Then we have to find the solution of the stochastic differential equation

$$dY_\gamma(c, t) = a_\gamma(Y_\gamma, s)\, ds + dW(c, t)$$

with $a_\gamma = (\partial/\partial x)\ln B_\gamma$. For $\gamma \neq \infty$, a_γ has no singularity and satisfies the inequality $a_\gamma^2(x, t) \leq K(1 + x^2)$ and the general theory applies. The measures $d\mu_\gamma$ are equivalent to the Wiener measure with Radon—Nikodym derivative $\neq 0$. For $\gamma \to \infty$ we obtain absorbing Brownian motion with

$$B_\infty(x, t) = \int_0^\infty \frac{dy}{\sqrt{2(T-t)}} [\exp(-(x-y)^2/2(T-t)) - \exp(-(x+y)^2/2(T-t))] = \lim_{\gamma \to \infty} B_\gamma(x, t), \quad x > 0.$$

$a_\infty(x, t)$ has a singularity of the form x^{-1} at the origin, since $B_\infty(x, t) \approx x$ and $(\partial/\partial x)B_\infty(x, t)$ is a nonzero function of t for $x \to 0$. Nevertheless the differential equation can be solved uniquely (corresponding in fact to three-dimensional Brownian motion, see Ref. 1). It follows that paths reaching the origin have zero probability with respect to the new measure [that can be seen also from the normalization factor $N(x) = B(x, 0)$, which is $= 0$ for $x = 0$]. Therefore it is impossible to return from absorbing Brownian motion to reflecting Brownian motion and the probability measures are not equivalent. The limit $\gamma \to -\infty$ does not exist, a_γ going to infinity proportional to γ (only the bound state contributes) in complete agreement with our previous discussion.

B. Boundary condition and singular potential

Considering the effect of singular potentials we have to deal with two regularizations, that of the potential and that of the δ function. The effect of the second one is not changed by an additional potential. So we have to solve the differential equation with the boundary conditions

$$\frac{\partial B_{\gamma,\epsilon}}{\partial t} = -\frac{1}{2}\frac{\partial^2 B_{\gamma,\epsilon}}{\partial x^2} + \lambda V_\epsilon B_{\gamma,\epsilon},$$

$$\gamma B_{\gamma,\epsilon}(0, t) = \frac{\partial}{\partial x} B_{\gamma,\epsilon}(0, t), \quad t \neq T$$

$$B_{\gamma,\epsilon}(x, T) = 1, \quad x \neq 0.$$

As in Ref. 1 we do not really try to find the solution but are satisfied with an approximation for the singularity of the potential. In Ref. 1 one can find the complete discussion for $\gamma = 0$. There the approximation had the form

$$B_{0,\epsilon}(x, t) = \exp[\mu(t) W(x)],$$

where $\mu(t) = \text{const}$ for $t \leq 0.99T$ going smoothly to zero for $0.99T \leq t < T$.

We assume further that

$$\frac{\partial}{\partial x} B_{0,\epsilon}(0, t) = 0$$

so that the boundary condition is satisfied.

Assume further that $B_{0,\epsilon}(0, t) = 1$ [which fixes $W(0)$]. Then we can get an approximation for arbitrary γ by

$$B_{\gamma,\epsilon}(x, t) = B_\gamma(x, t) B_{0,\epsilon}(x, t).$$

The $B_{\gamma,\epsilon}(x, t)$ satisfies the differential equation with a potential

$$2\lambda V(x, t) = \mu W'' + \mu^2 W'^2 + 2\dot{\mu} W + \mu W' \frac{\partial B_\gamma}{\partial x} B_\gamma^{-1}$$

together with the boundary conditions at $x = 0$. If $B_{0,\epsilon}$ was a good approximation for V, so $B_{\gamma,\epsilon}$ should be, because we get only the last term in addition to the previous one and this term is for $\gamma \neq \infty$ less singular than W''. If $\gamma = \infty$ we have an additional singularity of the kind $W'x^{-1}$ which will be of the same order as that belonging to W''. To argue why this approximation really gives the correct answer and the behavior in the time interval $0.99T < t < T$ cannot change the behavior of the paths significantly, we use the following result (partly contained in Ref. 7):

Theorem: If the potential is such that the Hamiltonian has one ground state $u_0(x)$ with eigenvalue 0, isolated from the rest of the spectrum, then the Y process has a limit as $T \to \infty$, which is a temporarily homogeneous Markov process obtained by choosing

$$B(x, t) = u_0(x), \quad a(x, t) = u_0'(x)/u_0(x).$$

If the potential is such, that the Hamiltonian has generalized eigenfunctions $u(k, x)$ that are continuous in the energy-eigenvalue k, then $a(x, t)$ converges to $u'(0, x)/u(0, x)$, where $u(0, x)$ satisfies $[-\frac{1}{2} d^2/dx^2 + V(x)]u(0, x) = 0$.

Proof: $B(x, t, T)$ can always be written as (neglecting degeneracy)

$$B(x, t, T) = \int_0^\infty u(k, x) u^*(k, y) \exp[-k(T-t)]\, dk\, dy + \sum \int u_i(x) u_i^*(y) \exp[-E_i(T-t)]\, dy,$$

where the $u(k, x)$ are the generalized eigenfunctions and the u_i are the eigenfunctions for the eigenvalues E_i. We are really interested in a,

$$a = \frac{B'}{B} = \lim_{T \to \infty} \frac{\int_0^\infty u'(k, x) u^*(k, y) \exp[-k(T-t)]\, dk\, dy + \sum \int u_i'(x) u_i^*(y) \exp[-E_i(T-t)]\, dy}{\int_0^\infty u(k, x) u^*(k, y) \exp[-k(T-t)]\, dk\, dy + \sum \int u_i(x) u_i^*(y) \exp[-E_i(T-t)]\, dy}.$$

Now either H has a separated ground state, then $u_0 \exp[|E_0|(T-t)]$ will dominate for $T \to \infty$ and

$$a = u_0'(x)/u_0(x)$$

or the last term is not present. In this case we can transform it to

$$a = \lim_{T \to \infty} \frac{\int_0^\infty u'(k/T-t, x) u^*(k/T-t, y) \exp(-k)\, dk\, dy}{\int_0^\infty u(k/T-t, x) u^*(k/T-t, y) \exp(-k)\, dk\, dy} = \frac{u'(0, x)}{u(0, x)},$$

using the continuity property of $u(k, x)$ in k. The condition of the above theorem is satisfied[8] if

$$\int_0^\infty dx\, x\,|V(x)| < \infty,$$

i.e., if the potential vanishes sufficiently at ∞ and is less singular than $1/r^2$ at the origin. The first condition is not really a restriction, since we are only interested in singularities at the origin and can therefore assume that the potential has finite range. The condition at the origin is satisfied, if we consider the regularized potentials.

We will not really use the above theorem since we are mainly interested in finite T, but here the continuity property of the generalized eigenfunctions tells us that $u(0, x)$ deals with the singularity at the origin in the same way as the other eigenfunctions do. Therefore it is sufficient to examine the path space for $B(x, t) = u(0, x)$ for $t \leqslant 0.99T$ provided the ground state energy of the sequence of Hamiltonians does not go to $-\infty$. In this case the only dominating contribution is given by the ground state and has to be examined.

VI. SPECIAL TYPES OF SINGULAR POTENTIALS

We turn now to the discussion of the special types of singular potentials and follow completely the discussion in Ref. 1.

A. Behavior for $\alpha < 1$

Since

$$B_0(x, t) = \exp[\mu(|x| + \epsilon)^{1+\beta}/(1+\beta)], \quad \beta > 0,$$

then

$$a(x, t) = \mu\, \mathrm{sgn}(x)(|x| + \epsilon)^{\beta}$$

and

$$2\lambda V_\epsilon(x) = \mu\beta(|x| + \epsilon)^{\beta-1} + 2\mu\,\delta(x)\epsilon^{\beta} + \mu^2(|x| + \epsilon)^{2\beta}.$$

Therefore choose $\beta - 1 = -\alpha$ and adjust μ. Notice that $\beta > 0$, so the term proportional to the δ function vanishes and the regularization does not mix between different boundary conditions. Also μ is uniquely determined (we have to adjust it in accordance to λ and the boundary condition), therefore we have only one limit point.

B. Behavior for $1 \leqslant \alpha < \frac{3}{2}$

$$B_0(x, t) = \exp[-\mu(|x| + \epsilon)^{1-\beta}/(1-\beta)], \quad \alpha = 1 + \beta$$

and

$$a(x, t) = -\mu\, \mathrm{sgn}(x)(|x| + \epsilon)^{-\beta}.$$

Again we can find regularizations for all γ such that for every ϵ the boundary conditions are satisfied. But if we vary now the boundary condition in addition to the potential the term which arises is smoother than the highest singularity already contained in B_γ. So it seems natural to vary the boundary conditions and we can really say that different regularizations lead to different boundary conditions. For $\gamma = \infty$ things are different. For every ϵ and μ, $\mu\, \mathrm{sgn}(x)(|x| + \epsilon)^{-\beta}$ is smoother than $\mathrm{sgn}(x)(|x|)^{-1}$ which is already present due to the boundary condition. Therefore we cannot get to $\gamma = \infty$ if we start from $\gamma \neq \infty$ regarding this kind of regularization. On the other hand starting already with $\gamma = \infty$ the singularity corresponding to the boundary condition dominates so that we cannot obtain any other boundary condition by a suitable regularization.

C. Behavior for $\frac{3}{2} \leqslant \alpha < 2$

Here we have to deal with regularizations depending on the strength of the potential. The detailed form can be found in Ref. 1. The behavior with respect to the boundary condition is analogous to the region B.

D. Behavior for $\alpha = 2$

This point is the critical point in the operator theory. On the other hand we have rather explicit solutions, so that it is worth discussing it in detail. We choose for $t < 0.99T$

$$B(x) = a(\gamma, \epsilon)(|x| + \epsilon)^{\theta_+} + b(\gamma, \epsilon)(|x| + \epsilon)^{\theta_-},$$

with

$$\theta_\pm = \tfrac{1}{2} \pm \sqrt{2\lambda + 1/4}\,.$$

We can assume that a and b are adjusted to fit the boundary conditions. First we have to consider if we are really allowed to restrict our interest on the beginning of the continuous spectrum. For $\lambda \geqslant -\frac{1}{8}$ we know that the form corresponding to the potential is bounded by the form corresponding to the Hamiltonian and so is the form corresponding to a regularized potential. Therefore we will not have eigenvalues for $\gamma > 0$ and for $\gamma < 0$ eigenvalues can occur, but they are bounded from below by $-\gamma^2$. (The corresponding fact holds for A, B, and C above, though in the present case the ground state eigenvalue may also depend on λ.)

(a) First we consider the region $-\frac{1}{8} < \lambda < \frac{3}{8}$.

$$a(x, t) = \mathrm{sgn}(x)\frac{a\theta_+(|x| + \epsilon)^{\theta_+-1} + b\theta_-(|x| + \epsilon)^{\theta_--1}}{a(|x| + \epsilon)^{\theta_+} + b(|x| + \epsilon)^{\theta_-}}$$

and

$$V_\epsilon(x) = \lambda(|x| + \epsilon)^{-2} + (a\theta_+\epsilon^{\theta_+-\theta_-} + b\theta_-)\delta(x)\epsilon^{-1}.$$

To satisfy the boundary condition

$$\frac{a\theta_+\epsilon^{\theta_+} + b\theta_-\epsilon^{\theta_-}}{a\epsilon^{\theta_+} + b\epsilon^{\theta_-}} = \gamma\epsilon$$

or

$$b = \frac{a(\gamma\epsilon^{\theta_++1} - \theta_+\epsilon^{\theta_+})}{\theta_-\epsilon^{\theta_-} - \gamma\epsilon^{\theta_-+1}} = \frac{a(\epsilon\gamma - \theta_+)}{\theta_- - \epsilon\gamma}\,\epsilon^{\theta_+-\theta_-}.$$

If we therefore fix the boundary condition, then b tends to zero for all γ and the differential equation for the paths converges to one corresponding to a $(2\theta + 1)$-dimensional Bessel process[2] (notice that $\theta_+ > \frac{1}{2}$ and $\frac{1}{2}$ corresponds to a two-dimensional Bessel process),

$$a(x, t) \approx \theta_+(|x| + \epsilon)^{-1}.$$

On the other hand the effect of changing the boundary condition leads to a singularity that is of the same or smaller order of magnitude as the singularity belonging to the potential, so we can regard it as belonging to the regularization and fix, e.g., a and b. Then, near $x = 0$,

$$a(x, t) \approx \mathrm{sgn}(x)[(a/b)\theta_+(|x| + \epsilon)^{\theta_+-\theta_--1} + \theta_-(|x| + \epsilon)^{-1}]$$

for $b \neq 0$ ($b=0$ corresponds completely to the previous discussion) and

$$\gamma(\epsilon) \approx (a/b)\theta_+\epsilon^{\theta_+-\theta_--1} + \theta_-\epsilon^{-1}.$$

In the region of λ we consider the second term is the dominating one for a as well as γ. If $\lambda > 0$ then θ_- is negative, i.e., we regularize with an additional attractive potential whereas for $\lambda < 0$, θ_- is positive and we regularize with a repulsive potential so that if we think of regularization as adding an additional potential concentrated at the singularity then the additional potential works in the opposite direction to the initial one.

We have neglected the contribution of the ground state which may exist. For $\lambda < 0$ and $\gamma(\lambda, \epsilon)$ the form is positive definite, so we need not worry. For $\lambda > 0$ $\gamma(\lambda, \epsilon)$ goes to $-\infty$, but it can easily be shown that in the considered region the ground state-eigenvalue remains bounded from below. For the eigenfunction of the free Hamiltonian

$$\langle\psi| V_\epsilon |\psi\rangle = \int_0^\infty \exp(-2\gamma r) 2\gamma\, dr \frac{\lambda}{(r+\epsilon)^2}$$
$$= \int_0^\infty \exp(-x)\, dx \frac{\lambda(2\gamma)^2}{(x+|\theta_-|)^2}$$

cancels the effect of the ground state ($0 \leq |\theta_-| \leq 1$) and combinations $\psi + \varphi$, $\varphi \in H_{\text{cont}}$ can be estimated by

$$|\langle\varphi| V_\epsilon |\psi\rangle| \leq \left\langle\varphi\left|\frac{1}{(r+\epsilon)^2}\right|\varphi\right\rangle \left\langle\psi\left|\frac{1}{(r+\epsilon)^2}\right|\psi\right\rangle^{1/2}$$

and

$$\left\langle\varphi\left|\frac{1}{(r+\epsilon)^2}\right|\varphi\right\rangle \leq \frac{1}{4}\left\langle\frac{\partial\varphi}{\partial r}\left|\frac{\partial\varphi}{\partial r}\right.\right\rangle,$$

so they stay bounded from below too, and our considerations give the correct answer.

(b) $\lambda = -\frac{1}{8}$

Here $\theta_+ = \theta_- = \frac{1}{2}$ and the most general solution is

$$B(x,t) = a(\gamma,\epsilon)(|x|+\epsilon)^{1/2} + b(\gamma,\epsilon)(|x|+\epsilon)^{1/2}\ln(|x|+\epsilon)$$

and

$$a(x,t) = \text{sgn}(x)\frac{a + b\ln(|x|+\epsilon) + 2b}{2[a(|x|+\epsilon) + b(|x|+\epsilon)\ln(|x|+\epsilon)]}.$$

As in the previous case the term proportional to b is the dominant one, so that for fixed γ and $\epsilon \to 0$,

$$b = a\frac{1/2 - \epsilon\gamma}{(1/2)\ln\epsilon - b\gamma\epsilon\ln\epsilon} \to 0$$

for any choice of γ. On the other hand we can also fix the ratio of a and b and obtain different results as in (a).

(c) For $\lambda \geq \frac{3}{8}$ we realize that we have only one B that satisfies the requirements. $\rho(x,y,t)$, the probability to find a particle at point x at time t, if it was at point y at time 0 should be an integrable function of x. As is discussed in Ref. 1

$$\rho(x,y,t) = B(x,t)\psi_y(x,t)/B(0,0),$$

where $\psi_y(x,t)$ satisfies

$$\frac{\partial\psi}{\partial t} = \frac{1}{2}\frac{\partial^2\psi}{\partial x^2} - \lambda V\psi,$$

$$\psi_y(x,0) = \delta(x-y),$$
$$\gamma\psi(0,t) = \frac{\partial\psi}{\partial x}(0,t), \quad t \neq 0.$$

Evidently B and ψ are singular at the same point, namely $x=0$. If therefore $\rho \in L^1$, B and ψ have to belong to L^2 and this holds only for the B corresponding to θ_+ since $\theta_- \leq -\frac{1}{2}$ if $\lambda \geq \frac{3}{8}$. Thus one must choose $b=0$ in the solution for B and no arbitrariness remains. Closer to the path space viewpoint is the following argument: Consider the stochastic differential equation

$$dY(t) = dW(t) + \theta Y(t)^{-1}\, dt$$

subject to the initial condition that $W(0) = Y(0) = c > 0$. In order that this equation make sense $Y(t)$ has to be bigger than zero. What happens afterwards has to be defined. We restrict our interest on the paths till they hit the origin and calculate

$$E\left[\int_0^{\min(T,\tau_0)} Y dY(t)\right] = \theta E\left[\int_0^{\min(T,\tau_0)} dt\right] + E\left[\int_0^{\min(T,\tau_0)} Y dW\right],$$

where τ_0 is the time when Y hits the origin. The last expectation value is zero, since we are dealing with martingales. Calculating the first expectation value we use

$$Y dY = \tfrac{1}{2}(dY^2 - dt)$$

and obtain

$$0 \leq E[\tfrac{1}{2}Y^2(\min(T,\tau_0))] = \tfrac{1}{2}c^2 + (\theta + \tfrac{1}{2})E[\min(T,\tau_0)].$$

Therefore we realize that for $\theta \leq -\frac{1}{2}$ the last value is nonpositive and so $E[Y^2(\min(T,\tau_0))] \leq c^2$ which means that almost certainly all paths come to the origin in a finite time and something serious happens. Consequently, we cannot find a useful solution of the stochastic equation for $\theta \leq -\frac{1}{2}$. This situation reflects the fact that for $2\lambda V = (\theta^2 - \theta)|x|^{-2}$, and $\lambda \geq \frac{3}{8}$, the solution with $\theta \leq -\frac{1}{2}$ is unacceptable, and only one solution is allowed just as in the formulation of the problem as a differential equation.

(d) We turn to the remaining region $\lambda < -\frac{1}{8}$.

Here θ_+ and θ_- become complex which already indicates that something fundamental happens. In fact we have already mentioned that the Hamiltonian does not remain bounded from below. There already exists a large amount of literature about the possible extensions of the Hamiltonian, self-adjoint or not self-adjoint.[9,10] Nelson's[11] method is closest to the spirit of this paper, because he uses Feynman integrals, the imaginary correspondence to Wiener integrals. But since Nelson stays on the imaginary time axis he is not affected by the divergence of the ground state energy.

Let us concentrate on the problem corresponding to absorbing Brownian motion and consider a regularization $V_\epsilon(x) = \lambda(|x|+\epsilon)^{-2}$. Then the eigenfunctions corresponding to the eigenvalue $-c^2(\epsilon)$ are[12]

$$\psi(x) = (|x|+\epsilon)^{1/2}[I_\nu((|x|+\epsilon)c) - I_{-\nu}((|x|+\epsilon)c)]$$

with $\nu = i\sqrt{2|\lambda| - 1/4}$.

The eigenvalue c^2 is defined by the boundary condition $\psi(0)=0$, i.e.,

$$I_\nu(\epsilon c)=I_{-\nu}(\epsilon c).$$

Therefore we conclude $c(\epsilon)=c_0\epsilon^{-1}$, $c_0\neq 0$.

Now the ground state gives the leading contribution to

$$a_\epsilon(x,t)=\frac{(|x|+\epsilon)^{-1/2}[I_\nu-I_{-\nu}]+(|x|+\epsilon)^{1/2}c_0\epsilon^{-1}[I'_\nu-I'_{-\nu}]}{(|x|+\epsilon)^{1/2}[I_\nu-I_{-\nu}]}.$$

Qualitatively, this expression is given by $(|x|+\epsilon)^{-1}$ times a term involving (for small ϵ) $\tan[\sqrt{2|\lambda|-1/4}\times\ln(1+|x|/\epsilon)c_0]$, and thus has no unique limit as $\epsilon\to 0$ for any x value. This behavior is characteristic for other boundary conditions: $c(\epsilon)$ is always going to infinity such that $J_\nu(i|x|c)$ is not uniformly continuous in $|x|$ for all ϵ, corresponding to the fact that the bound states become more and more concentrated at the origin. Therefore the path space viewpoint fails here for this type of regularization. Similarly Meetz[13] has shown that such regularizations do not lead to a self-adjoint (or to any) extension of the operator.

E. Behavior for $\alpha > 2$

Here we have to distinguish between $\lambda\geq 0$ (corresponding to the region $\lambda\geq\frac{3}{8}$ for $\alpha=2$) and $\lambda<0$ (corresponding to the region $\lambda<-\frac{1}{8}$ for $\alpha=2$). For $\lambda>0$ and any γ there is only one choice of regularization (see Ref. 1) leading to absorbing Brownian motion.

For $\lambda<0$ the contribution of the ground state, the energy of which goes to $-\infty$, leads to nonconvergence of a_ϵ.

F. Behavior for $\lambda x^{-2}+\mu x^{-\alpha}$, $\alpha>2$

We can choose $B\approx B_{-2}B_{-\alpha}$. If μ is bigger than zero then $B_{-\alpha}$ has the winning singularity. Therefore different choices for B_{-2} do not change the result and we have only one solution that corresponds to absorbing Brownian motion.

If μ is smaller than zero we have to determine whether λ big enough can save us so that we can control the ground state. But a scaling argument shows that this is impossible. The $1/r^2$ potential is not strong enough to keep the particles away the origin so we cannot treat the problem on the level of Brownian motion.

ACKNOWLEDGMENTS

In the context of differential equations arising in singular quantum mechanical perturbations, F. A. Berezin and L. D. Faddeev[14] have emphasized the nonuniqueness of self-adjoint extensions of the Laplace operator in three-dimensions as defined on smooth functions all of which vanish at one point. A discussion with Professor Faddeev by one of us (JRK) on the dimensional dependence of such questions has in part prompted us to investigate corresponding properties in the context of the Feynman–Kac integral and the theory of stochastic differential equations.

*On leave from the Institute for Theoretical Physics, University of Vienna.

[1] H. Ezawa, J. R. Klauder, and L. A. Shepp, J. Math. Phys. 16, 783 (1975).
[2] K. Itô and H. P. McKean, *Diffusion Processes and their Sample Paths* (Springer, New York, 1965).
[3] M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1973).
[4] H. Narnhofer, Nuovo Cimento B 30, 254 (1975).
[5] M. Kac, *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California, Berkeley, Calif., 1951), p. 189.
[6] T. Kato, *Perturbation Theory of Linear Operators* (Springer, New York, 1966).
[7] H. Ezawa, J. R. Klauder, L. A. Shepp, Ann. of Phys. 88, 588 (1974).
[8] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[9] H. Narnhofer, Acta Phys. Austriaca 40, 306 (1974).
[10] Ch. Radin, J. Math. Phys. 16, 544 (1975).
[11] E. Nelson, J. Math. Phys. 5, 332 (1964).
[12] K. M. Case, Phys. Rev. 80, 797 (1950).
[13] K. Meetz, Nuovo Cimento 34, 690 (1964).
[14] F. A. Berezin and L. D. Faddeev, Dokl. Sov. Math. 2, 372 (1961).

On canonical quantum field theories*

Ira W. Herbst

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 5 December 1975)

We consider time reversal invariant canonical quantum field theories in which $\pi = \dot{\phi}$. We show that if the theory has a mass gap, the vacuum is an analytic vector for the time zero field $\phi(f)$. With the additional assumptions of Poincaré covariance, cyclicity of the vacuum for the time zero fields, and a domain condition on the Hamiltonian, we show that the Schwinger functions of the theory determine a Euclidean covariant Markov field theory. We also consider the implications of a bound of the form $\pm\phi(f)^j \leq H + \gamma(f)$ for the behavior of the ground state at large field strength. We show that such a bound implies that the vacuum is an analytic vector for $|\phi(f)|^{j/2+1}$.

I. INTRODUCTION

In his 1960 thesis (see also Ref. 1), Araki studied canonical, time reversal invariant quantum field theories in which $\pi = i[H, \phi]$. Such theories have Hamiltonians which are infinite dimensional analogs of $-\frac{1}{2}\Delta + V$, with V a real valued function of finitely many variables. As Araki (implicitly) shows, the infinite dimensional analog of $-\frac{1}{2}\Delta + V$ in a Hilbert space where the ground state, Ω_0, is the function identically equal to one, is a Hamiltonian which is a Dirichlet form. We use this result as our starting point.

In finite dimensions with appropriate restrictions on the potential, V, it is known[2–6] that the eigenstates of $-\frac{1}{2}\Delta + V$ decay exponentially. In Sec. II we prove a result of this type for the ground state of an infinite dimensional theory: Under the assumption that $H\Omega_0 = 0$, $H \upharpoonright \Omega_0^\perp \geq m > 0$, we show that Ω_0 is an analytic vector for $\phi(f)$, with bounds dependent only on m, f, and $(\Omega_0, \phi(f)\Omega_0)$ (see Theorem 2.2).

The exponential decay of the eigenstates of $-\frac{1}{2}\Delta + V$ (analogous to what is found in Theorem 2.2) is to be expected when V does not grow at infinity. If, however, V behaves like $|x|^j$ at infinity it is known[7] that the eigenstates of $-\frac{1}{2}\Delta + V$ are analytic vectors for $|x|^{j/2+1}$. We recover a result of this type in the infinite dimensional case: Assuming a bound of the form $\pm\phi(f)^j \leq H + \gamma(f)$, we show that the vacuum is an analytic vector for $|\phi(f)|^{j/2+1}$ (see Theorem 2.5).

The above results require neither cyclicity of the vacuum for the time zero fields nor any consequences of Poincaré covariance. If, in addition to the existence of a mass gap, we impose the requirement of translational invariance and demand the spectral condition, the Glimm–Jaffe $\nabla\phi$ bound[8] is shown to imply that for $f \in L^2 \cap L^1$, Ω_0 is an entire vector for $\phi(f)$. In Sec. III, we assume that Ω_0 is a cyclic vector for the time zero fields and that the theory is Poincaré covariant. We discuss a condition under which the Hamiltonian generates a positivity preserving semigroup, $\{\exp(-tH)\}_{t>0}$. All that is needed there is that the Dirichlet form actually determines the Hamiltonian. A generalization of a theorem of Gross[9] (in whose proof we correct a gap) then gives the fact that $\exp(-tH)$ is positivity preserving. (This is not surprising in view of the results known for $-\frac{1}{2}\Delta + V$.) It is then shown, using an idea of Fröhlich's[10] that the resulting Euclidean region field theory (shown to exist by Simon[11]) is actually Euclidean invariant.

We conclude by discussing some open questions.

II. THE BEHAVIOR OF THE VACUUM FOR LARGE FIELD STRENGTH

In the following, $\mathcal{K}$ will denote a real Hilbert space and $\mathcal{L} \subseteq \mathcal{K}$ a linear subset of $\mathcal{K}$, dense in $\mathcal{K}$. The symbol $\langle\ ,\ \rangle$ will denote the inner product in $\mathcal{K}$.

We assume there exists a complex Hilbert space $\mathcal{H}$ such that, for each $f \in \mathcal{L}$, $\phi(f)$ is a self-adjoint operator in $\mathcal{H}$ satisfying

$$\exp[i\phi(\lambda f + g)] = \exp[i\lambda\phi(f)]\exp[i\phi(g)], \quad f, g \in \mathcal{L}, \ \lambda \in R. \tag{2.1}$$

In addition, we assume the existence of a nonnegative self-adjoint operator H in $\mathcal{H}$ and a distinguished vector $\Omega_0 \in \mathcal{H}$ [which we normalize so that $(\Omega_0, \Omega_0) = 1$] such that

(a) for each $f \in \mathcal{L}$, $\exp[i\phi(f)]\Omega_0 \in \mathcal{D}(\sqrt{H})$,

(b) for each f, $g \in \mathcal{L}$, (2.2)

$$(\exp[i\phi(f)]\Omega_0, H\exp[i\phi(g)]\Omega_0) = \tfrac{1}{2}\langle f, g\rangle(\exp[i\phi(f)]\Omega_0, \exp[i\phi(g)]\Omega_0). \tag{2.3}$$

Equation (2.3) was given by Araki in his 1960 Princeton thesis and in Ref. 1. He derived this result in the framework of canonical quantum field theory, invariant under time reversal, with $\pi = i[H, \phi]$. We sketch his proof for the convenience of the reader:

Instead of the relation $\pi(f) = [iH, \phi(f)]$ which involves three unbounded operators, assume the equation

$$[\exp[-i\phi(f)], [H, \exp[i\phi(g)]]] = \langle f, g\rangle \exp[-i\phi(f)]\exp[i\phi(g)], \tag{2.4}$$

which follows formally from $\pi(f) = [iH, \phi(f)]$ and the canonical commutation relations $[\pi(f), \phi(g)] = -i\langle f, g\rangle$. Assume the existence of a vector Ω_0 and an antiunitary operator T such that

$$H\Omega_0 = 0, \quad T\Omega_0 = \Omega_0, \quad THT^{-1} = H, \quad T\exp[i\phi(f)]T^{-1} = \exp[-i\phi(f)]. \tag{2.5}$$

We then can compute

$$(\exp[i\phi(f)]\Omega_0, H\exp[i\phi(g)]\Omega_0)$$
$$= (\Omega_0, \exp[-i\phi(f)]H\exp[i\phi(g)]\Omega_0)$$
$$= (T\exp[-i\phi(f)]H\exp[i\phi(g)]\Omega_0, \Omega_0)$$
$$= (\exp[i\phi(f)]H\exp[-i\phi(g)]\Omega_0, \Omega_0)$$
$$= (\Omega_0, \exp[i\phi(g)]H\exp[-i\phi(f)]\Omega_0).$$

Thus

$$(\exp[i\phi(f)]\Omega_0, H\exp[i\phi(g)]\Omega_0)$$
$$= \tfrac{1}{2}(\Omega_0, \exp[-i\phi(f)]H\exp[i\phi(g)]\Omega_0)$$
$$+ \tfrac{1}{2}(\Omega_0, \exp[i\phi(g)]H\exp[-i\phi(f)]\Omega_0)$$
$$= \tfrac{1}{2}(\Omega_0, [\exp(-i\phi(f)), [H, \exp(i\phi(g))]]\Omega_0),$$

which because of Eq. (2.4) gives Eq. (2.3).

Although Araki's derivation of Eq. (2.3) (which we henceforth call the Araki formula) is formal, it indicates that the Araki formula should be true in many interesting cases including some where H is not of the form $\frac{1}{2}\int \pi^2(\mathbf{x})\,d\mathbf{x} + V(\phi)$. We state below some of the situations in which it is easy to prove the Araki formula along with the domain condition Eq. (2.2):

1. Schrödinger operators: Let $H = H_0 + Z$ be defined as a sum of the quadratic forms $H_0 = -\frac{1}{2}\Delta$ and $Z = V + T$, where V is a nonnegative measurable function with $\mathcal{D}(V^{1/2}) \cap \mathcal{D}(H_0^{1/2})$ dense in $L^2(R^N)$ and T is a real distribution with $\pm T \leqslant aH_0 + b$, $a < 1$. Let $\mathcal{H} = L^2(R^N, d^N x)$, $\mathcal{K} = R^N$, and suppose $\phi(a)$ is multiplication by the function $\langle x, a\rangle = \sum_{i=1}^N x_i a_i$. Then if Ω_0 is any *real* eigenfunction of H with eigenvalue E_0, the Araki formula holds with H replaced by $H - E_0$, for any $f, g \in \mathcal{L} = \mathcal{K} = R^N$.

2. Suppose $H = H_0 + \int :(\phi):g(x)\,dx - E(g)$ is the spatially cut-off $P(\phi)_2$ Hamiltonian or H is the infinite volume $P(\phi)_2$ Hamiltonian in the small coupling theories of Glimm, Jaffe, and Spencer.[12] Then the Araki formula holds with Ω_0 the vacuum state. In the spatially cut-off case we can take $\mathcal{K} = \mathcal{L} = L^2_{\text{real}}(R^1, dx)$ and in the infinite volume case $\mathcal{L}$ = real C^∞ functions with compact support (this can certainly be improved). In both cases we have of course $\langle f, g\rangle = \int f(x) g(x)\,dx$.

We expect that the Araki formula will hold in all $P(\phi)_2$ infinite volume theories as well as the ϕ_3^4 theory. For the more interesting (!) ϕ_4^4 interaction, perturbation theory "predicts" an infinite field strength renormalization is necessary and this would invalidate the Araki formula. (However, whether or not the field strength renormalization is actually infinite is not yet known.)

We also expect the Araki formula to hold in the scalar and pseudoscalar Yukawa field theories (in two space-time dimensions). This is because the Schwinger functions of the theory are real[13] and thus on the boson Hilbert space $\mathcal{H}_b$ [$\mathcal{H}_b$ is the smallest subspace containing the vacuum and invariant under $\exp(-tH)$ and $\exp(i\phi(f))$] we can define an antiunitary operator T satisfying Eq. (2.5). [Note that here H should be replaced by $H_b = H\upharpoonright\mathcal{H}_b$. The operator H_b still satisfies Eq. (2.4)]. The reason for going to $\mathcal{H}_b$ instead of working directly on $\mathcal{H}$ is that in pseudoscalar Yukawa, the usual T operator satisfies $T\phi(f)T^{-1} = -\phi(f)$.[14]

The above remarks and Araki's derivation of Eq. (2.3) are given for motivational purposes only. In the following we assume only the structure discussed at the beginning of this section which is summarized by the Eqs. (2.1)–(2.3). We do not assume the existence of a self-adjoint operator $\pi(f)$.

Our present purpose is to generalize the Araki formula.

Notation: If $\{f_i\}_{i=1}^N$ is an orthonormal sequence from $\mathcal{L}$ and $F: R^N \to \mathbb{C}$ is a Borel function we write $\hat{F} \equiv F(\phi(f_1), \ldots, \phi(f_N))\Omega_0$ if the right-hand side is in $\mathcal{H}$. The dependence of $\hat{F}$ on $\{f_i\}_{i=1}^N$ will not be indicated but will be clear in context.

Proposition 2.1: Suppose $\{f_i\}_{i=1}^N$ is an orthonormal sequence from $\mathcal{L}$. Suppose $F_1, F_2: R^N \to \mathbb{C}$ are continuously differentiable with $\hat{F}_i$ and $\widehat{\partial_j F_i}$ in $\mathcal{H}$. Then $\hat{F}_i \in \mathcal{D}(\sqrt{H})$ and

$$(\hat{F}_1, H\hat{F}_2) = \tfrac{1}{2}(\widehat{\nabla F_1}, \widehat{\nabla F_2}), \tag{2.6}$$

where here $(\widehat{\nabla F}, \widehat{\nabla G})$ is shorthand for $\sum_{i=1}^N (\widehat{\partial_i F}, \widehat{\partial_i G})$.

Remark: It is because of Eq. (2.6) that we call H a Dirichlet form.

The proof of Proposition 2.1 is a simple exercise in approximating the functions F_i by exponentials for which Eq. (2.6) reduces to the Araki formula. It is given in Appendix A.

We would now like to use Eq. (2.6) to bound $\|\exp[\phi(f)]\Omega_0\|^2 = (\exp[\phi(f)]\Omega_0, \exp[\phi(f)]\Omega_0)$ when there is a gap in the spectrum of H above zero:

Theorem 2.2: Assume Eqs. (2.1)–(2.3) and suppose in addition,

$$H \upharpoonright \Omega_0^\perp \geqslant m > 0. \tag{2.7}$$

Then if $f \in \mathcal{L}$ and $\langle f, f\rangle/2m < 1$, we have $\Omega_0 \in \mathcal{D}(\exp[\phi(f)])$ and

$$(\Omega_0, \exp[2\phi(f)]\Omega_0)$$
$$\leqslant \gamma \exp[2(\Omega_0, \phi(f)\Omega_0)](1 - \langle f, f\rangle/2m)^{-1}, \tag{2.8}$$

where γ is a universal constant.

Remark: This result is essentially the best possible if nothing more is assumed. For if we consider the one-dimensional Schrödinger Hamiltonian $H = -\frac{1}{2}(d^2/dx^2) + V(x)$ with $V(x) = -\lambda\delta(x - a)$ there is one bound state Ω_0 if $\lambda > 0$ with $H\Omega_0 = -m\Omega_0$. We can explicitly compute for $\alpha^2 < 2m$,

$$(\Omega_0, \exp(2\alpha x)\Omega_0) = \exp[2\alpha(\Omega_0, x\Omega_0)](1 - \alpha^2/2m)^{-1}$$

in agreement with Eq. (2.8).

In the following we prove Theorem 2.2 assuming that for $\rho = \langle f, f\rangle/2m < 1$, $\Omega_0 \in \mathcal{D}(\exp(\phi(f)))$. This assumption is proved in Appendix B.

Making this assumption, we know from Eqs. (2.6) and (2.7) that

$$(\exp[\phi(f)]\Omega_0, H\exp[\phi(f)]\Omega_0) = \tfrac{1}{2}\langle f, f\rangle(\Omega_0, \exp[2\phi(f)]\Omega_0)$$
$$\geqslant m\{\|\exp[\phi(f)]\Omega_0\|^2 - (\Omega_0, \exp[\phi(f)]\Omega_0)^2\}$$

and thus

$$(\Omega_0, \exp[2\phi(f)]\Omega_0) \le (1-\rho)^{-1}(\Omega_0, \exp[\phi(f)]\Omega_0). \quad (2.9)$$

Iterating Eq. (2.9) gives

$$(\Omega_0, \exp[2\phi(f)]\Omega_0) \le \left(\prod_{n=0}^{N}(1-\rho/4^n)^{-2^n}\right)(\Omega_0, \exp[+2^{-N}\phi(f)]\Omega_0)^{2^{N+1}}. \quad (2.10)$$

In the limit $N\to\infty$ we get Eq. (2.8) with $\gamma = \Pi_{n=1}^{\infty}(1-4^{-n})^{-2^n}$.

We remark that a technique similar to the above was used by Ahlrichs[5] to find explicit bounds on eigenfunctions of atomic Hamiltonians.

Using a cut-off version of $\exp[\phi(f)]$ this method could be used to prove the bound Eq. (2.8) if one knew in advance only that $(\Omega_0, \phi(f)\Omega_0)$ was finite. However, in Appendix B we modify a technique of Combes and Thomas,[2] which immediately gives the fact that Ω_0 is an analytic vector for $\phi(f)$.

We remark that the bound we have just derived allows the extension of the field $\psi(f) \equiv \phi(f) - (\Omega_0, \phi(f)\Omega_0)$ from $\mathcal{L}$ to $\mathcal{K}$ as a self-adjoint operator on the sub-Hilbert space spanned by $\{\exp[i\phi(f)]\Omega_0 : f\in\mathcal{L}\}$. The details are left to the interested reader.

Before leaving the subject of Theorem 2.2 we give one corollary whose easy proof (which we omit) employs the Cauchy integral formula for the derivatives of $F(z_1,\ldots,z_n) = \{\Omega_0, \exp[\sum z_i\psi(f_i)]\Omega_0\}$.[15]

Corollary 2.3: Under the assumptions of Theorem 2.2 the field

$$\psi(f) \equiv \phi(f) - (\Omega_0, \phi(f)\Omega_0), \quad (2.11)$$

which we can extend to the complexification of $\mathcal{L}$ by linearity obeys the bound

$$|(\Omega_0, \psi(f_1)\cdots\psi(f_n)\Omega_0)| \le \beta^n n! \prod_{i=1}^{n}(\langle f_i, f_i\rangle/2m)^{1/2} \quad (2.12)$$

for some universal constant β, and f_i in the complexification of $\mathcal{L}$ for each i.

We now continue our study of the decay properties of the ground state to see what can be said when the potential grows at infinity. We first prove a lemma which shows that in our formalism certain growth conditions imply others.

Lemma 2.4: Assume Eqs. (2.1)–(2.3) hold and suppose that for some fixed $f\in\mathcal{L}$ and some positive integer j,

$$\pm[\phi(f)]^j \le C(H+1) \quad (2.13)$$

as forms on the set $\mathcal{D}_f = \{F(\phi(f))\Omega_0 : F\in\mathcal{S}(R^1)\}$. Then again as forms on $\mathcal{D}_f$ we have

$$|\phi(f)|^j \le C_1(H+1), \quad (2.14)$$

and if $\lambda \ge 0$, $j>1$,

$$\lambda|\phi(f)| \le C_2(H+1+\lambda^{j/(j-1)}). \quad (2.15)$$

Proof: It is enough to consider the case $\langle f,f\rangle = 1$. We prove (2.14) first and thus assume that j is odd. Let $\theta: R\to[0,1]$ be C^∞ with $\theta(x)=0$ if $x\le\frac{1}{2}$, $\theta(x)=1$ if $x\ge 1$. Define $\hat\theta(x) = \theta(-x)$. We denote multiplication by $\theta(\phi(f))$ by θ and similarly for $\hat\theta$. Thus as forms on $\mathcal{D}_f$

$$|\phi(f)|^j = \theta\phi(f)^j\theta - \hat\theta\phi(f)^j\hat\theta + (1-\theta^2-\hat\theta^2)|\phi(f)|^j \le C(\theta H\theta + \hat\theta H\hat\theta + 1) + 1.$$

We will have proved (2.14) if we can show that

$$\theta H\theta \le C'(H+1), \quad (2.16)$$

and similarly for $\hat\theta$. If $F\in\mathcal{S}(R^1)$, then

$$(\hat F, \theta H\theta\hat F) = \tfrac{1}{2}\|\theta' F + \theta F'\|^2 \le \|\widehat{\theta F'}\|^2 + \|\widehat{\theta' F}\|^2 \le b(\tfrac{1}{2}\|\hat F'\|^2 + \|\hat F\|^2) = b((\hat F, H\hat F) + (\hat F, \hat F)).$$

This proves (2.16) and thus (2.14). To show (2.15), note that

$$C_1H - \lambda|\phi(f)| = C_1H - |\phi(f)|^j + |\phi(f)|^j - \lambda|\phi(f)| \ge -C_1 + \inf_{x\ge 0}(x^j - \lambda x) = -C_1 - C_3\lambda^{j/(j-1)}.$$

This completes the proof of the lemma.

Theorem 2.5: Assume Eqs. (2.1)–(2.3) hold.

(A) If for some fixed $f\in\mathcal{L}$ and some positive integer j

$$\pm\lambda\phi(f) \le C(H+1+\lambda^{j/(j-1)}) \text{ if } j>1, \quad (2.17)$$

$$\pm\phi(f) \le C(H+1) \text{ if } j=1, \quad (2.18)$$

as forms on $\mathcal{D}_f$. Then Ω_0 is an analytic vector for $|\phi(f)|^{(j/2+1)}$ or in other words for some $\alpha>0$ (and f dependent)

$$(\Omega_0, |\phi(f)|^n\Omega_0) \le \alpha^n (n!)^{(j/2+1)^{-1}}. \quad (2.19)$$

(B) If for all $f\in\mathcal{L}$ and some seminorm $|\cdot|$ on $\mathcal{L}$

$$\pm\phi(f) \le H + |f|^2 \quad (2.20)$$

as forms on $\mathcal{D}_f$, then for some universal constant ξ

$$(\Omega_0, |\phi(f)|^n\Omega_0) \le (\xi^n n!(|f|\langle f,f\rangle^{1/2})^n)^{1/2} \quad (2.21)$$

for all $f\in\mathcal{L}$.

Remarks: (1) Bounds of the type Eqs. (2.13), (2.14), and (2.17) have been known in certain field theory models for quite some time. In the $P(\phi)_2$ theories such bounds were first given by Glimm and Jaffe,[16] who derived a bound analogous to (2.13) for the more singular Wick polynomial. A simplified proof was later given by Guerra, Rosen, and Simon[17] (see also Ref. 18), and based on their work Fröhlich[19] derived bounds of the form (2.17) with an explicit dependence on the smearing function f. Although none of these authors seem to consider the non-Wick ordered $\phi(f)^j$ explicitly (except for $j=1,2$), the relevant bounds follow easily from their methods. More recently bounds of this type for ϕ and $:\phi^2:$ have been proved for the two-dimensional Yukawa model,[20–22] and in the ϕ_3^4 model (see Ref. 22) and references given there).

(2) We have not kept track of the dependence of the constant α in Eq. (2.19) on the smearing function f although with more work this could have been done. We have isolated the bound (2.20) because the Glimm–Jaffe $\nabla\phi$ bound[8] is of this form [i.e. $\pm\phi(\partial_j f) \le H + \frac{1}{2}\langle f,f\rangle$].

Proof of Theorem 2.5: By the spectral theorem we can write $[\Omega_0, F(\phi(f))\Omega_0] = \int d\mu(x)F(x)$ for some probability measure μ on R. We use the abbreviation $\langle F\rangle = \int d\mu(x)F(x)$. Consider the bound (2.17) which we write for $F\in\mathcal{S}(R^1)$ as

$$\pm\lambda\langle x|F|^2\rangle \le C\{(\hat{F},(H+1)\hat{F})+\lambda^{j/(j-1)}\langle|F|^2\rangle\}.$$

Optimizing with respect to λ, we find

$$|\langle x|F|^2\rangle|^j \le d(\hat{F},(H+1)\hat{F})\langle|F|^2\rangle^{j-1}, \quad (2.22)$$

which also holds if we are in the case $j=1$. We now proceed under the assumption that all the moments $(\Omega_0, |\phi(f)|^n\Omega_0)$ are finite and show later how to justify this. Choose θ as in the proof of Lemma 2.4 and let $F(x)=\theta(x)x^{(n-1)/2}$. Then (2.22) yields for $n\ge 3$

$$\langle\theta^2(x)|x|^n\rangle^j \le d_1\{n^2\langle\theta^2(x)|x|^{n-3}\rangle + \langle\theta^2(x)|x|^{n-1}\rangle+1\}\cdot\langle\theta^2(x)|x|^{n-1}\rangle^{j-1}. \quad (2.23)$$

If $y_n \equiv \langle|x|^n\theta^2(x)\rangle n^{-n\beta}$, $\beta=(\frac{1}{2}j+1)^{-1}$, then (2.23) implies

$$y_n^j \le d_2\{y_{n-3}+y_{n-1}+1\}y_{n-1}{}^{j-1}.$$

Using the inequality $\alpha^{1/j}\beta^{1-1/j}\le(1/j)\alpha+(1-1/j)\beta$, we find the linear recursion relation

$$y_n \le d_3(y_{n-3}+y_{n-1}+1),$$

which yields $y_n\le\rho^n$. The same procedure works for $\langle\theta^2(-x)|x|^n\rangle$, and thus we get Eq. (2.19). To show that the moments are finite, one can first replace $|x|^n$ by $|x|^n\exp(-\epsilon x^2)$. The recursion relation analogous to (2.23) yields bounds independent of $\epsilon>0$ and thus one can take $\epsilon\searrow 0$.

The proof of part B is similar and is omitted.

Remarks: (1) Similar theorems have been proved in the finite dimensional case by Simon[7] who used a generalization of the Combes–Thomas technique.[2] To carry over into the infinite dimensional theory, Simon's method seems to require a bound of the form $\pi^2(f)\le C(H+1)$ and more information about the domain of $\sqrt{H}$ than we have assumed. However, his method would work in the $P(\phi)_2$ model where a π^2 bound is available from estimates of Spencer.[23]

(2) A similar procedure using Wick ordered monomials $:\phi^j:(f)$ fails because $:\phi^j:(f)\Omega_0$ is not in the domain of $\sqrt{H}$.

We now discuss the $\nabla\phi$ bound of Glimm and Jaffe.[8] In order to do so, we need to introduce some additional structure appropriate to a translation invariant quantum field theory. We assume Eqs. (2.1)–(2.3) with $\mathcal{L}=\mathcal{S}_{\rm real}(R^s)$, $\mathcal{K}=L^2_{\rm real}(R^s,d^sx)$. In addition we impose the following conditions:

(T_1) $U(a_0,\mathbf{a})=\exp(ia_0H)\exp(-i\mathbf{a}\cdot\mathbf{P})$ is a continuous unitary representation of the translation group in $s+1$ dimensions.

(T_2) The joint spectrum of $(H,\mathbf{P})$ is contained in the forward light cone, $\{p\in R^{s+1}: p_0\ge|\mathbf{p}|\}$.

(T_3) $U(0,\mathbf{a})\exp(i\phi(f))U(0,-\mathbf{a})=\exp(i\phi(f_{\mathbf{a}}))$, where $f_{\mathbf{a}}(\mathbf{x})=f(\mathbf{x}-\mathbf{a})$.

(T_4) $(\Omega_0,\phi(f)\phi(g)\Omega_0)$ as a function on $\mathcal{S}\times\mathcal{S}$ is continuous in each variable separately in the topology of $\mathcal{S}$.

We remark that if we are given (T_1)–(T_3) and a mass gap, then (T_4) is satisfied. [In fact, all the moments $(\Omega_0,\phi(f_1)\cdots\phi(f_n)\Omega_0)$ are tempered.] This follows from Corollary 2.3 and the result that $(\Omega_0,\phi(f)\Omega_0)=c\int f(x)\,dx$. The proof of this formula is nontrivial because we do not know *a priori* that $(\Omega_0,\phi(f)\Omega_0)$ is tempered. Our proof in Appendix C makes use of a theorem of Liouville in number theory.

In any case given (T_1)–(T_4) we have

Theorem 2.6 (*Glimm–Jaffe*[8]): For any unit vector $\mathbf{a}$,

$$\pm\phi(\mathbf{a}\cdot\nabla f)\le H+\tfrac{1}{2}\langle f,f\rangle$$

as forms on $\{\hat{F}\Omega_0: F\in\mathcal{S}(R^N),\ N\ge 0\}$.

Since our formalism differs slightly from that of Glimm and Jaffe, we sketch a proof: We call $\mathbf{a}$ the j direction and consider the expression

$$\tfrac{1}{2}(\exp[-i\phi(f)]\hat{F},(H-P_j)\exp[i\phi(f)]\hat{F}) + \tfrac{1}{2}(\exp[-i\phi(f)]\hat{F},(H+P_j)\exp[-i\phi(f)]\hat{F}),$$

which by assumption is nonnegative. By explicit calculation, using Eq. (2.6) and (T_3) above, we find this expression is equal to

$$\tfrac{1}{2}(\hat{F},\hat{F})\langle f,f\rangle+(\hat{F},H\hat{F})+\lim_{a\to 0}(1/a)(\sin\phi(f_a-f)\hat{F},U(0,a\mathbf{e}_j)\hat{F}),$$

where $f_a(\mathbf{x})=f(\mathbf{x}-a\mathbf{e}_j)$ and $\mathbf{e}_j$ is a unit vector in the j direction. Using assumption (T_4) above, it is easily shown that the last term above is just $-(\hat{F},\phi(\nabla_j f)\hat{F})$ and thus the result follows.

Corollary 2.7: Suppose Eqs. (2.1)–(2.3) hold with $\mathcal{L}=\mathcal{S}_{\rm real}(R^s)$ and in addition (i) $H\restriction_{\Omega_0^\perp}\ge m>0$, (ii) ($T_1$), ($T_2$), and ($T_3$) hold. Then for any $f\in\mathcal{L}$ the vacuum is an analytic vector for $\phi(f)$.

Proof: Consider the field $\psi(f)\equiv\phi(f)-(\Omega_0,\phi(f)\Omega_0)$. By Corollary 2.3, $|(\Omega_0,\psi(f)\psi(g)\Omega_0)|\le C\|f\|_2\|g\|_2$ so that assumption (T_4) above holds. This we have by Theorem 2.6

$$\|\exp\{t[\psi(\nabla_j f)]^2\}\Omega_0\|<\infty \quad\text{for some } t>0,$$

and hence $\|\exp[\psi(\nabla_j f)]\Omega_0\|<\infty$ for all $f\in\mathcal{L}$. [In fact one can show that $(\Omega_0,\exp(\psi(\nabla_j f))\Omega_0)\le\exp(\frac{1}{4}\|f\|_2\|\nabla_j f\|_2)$.] Using a C^∞ partition of unity in momentum space, we can write for any $\epsilon>0$

$$f=g+\sum_{j=1}^{s}\nabla_j f_j,$$

with $f_j\in\mathcal{L}$ and $\|g\|_2<\epsilon$. Thus

$$(\Omega_0,\exp[\psi(f)]\Omega_0) \le(\Omega_0,\exp[p\psi(g)]\Omega_0)^{1/p}\prod_{j=1}^{s}(\Omega_0,\exp[p\psi(\nabla_j f_j)]\Omega_0)^{1/p}$$

with $p=s+1$. The first term is finite by Theorem 2.2 (if ϵ is small enough) and thus the result is proved.

We remark that if $\psi(f)$ is extended to $f\in L^2_{\rm real}$ (see the remark after Theorem 2.2), the bounds we have proved show that Ω_0 is analytic for $\psi(f)$ for all $f\in L^2_{\rm real}$.

III. CANONICAL WIGHTMAN FIELD THEORIES

In this section we consider some implications of the structure discussed in the last section in a field theory satisfying the Wightman axioms. We will also impose additional restrictions. We first assume:

(A) ϕ,H,Ω_0 satisfy Eqs. (2.1)–(2.3) with $\mathcal{L}=\mathcal{S}_{\rm real}(R^s)$ and $\langle f,g\rangle=\int d^sx\,f(x)g(x)$.

(B) $H\upharpoonright\Omega_0^\perp \geq m > 0$.

(C) For convenience of notation we assume $(\Omega_0, \phi(f)\Omega_0) \equiv 0$.

In addition we assume that the vacuum is cyclic for the time zero fields:

(D) The linear span, $\mathcal{D}_0$, of $\{\exp[i\phi(f)]\Omega_0 : f \in \mathcal{L}\}$ is dense in $\mathcal{K}$.

We remark that this last condition is unproved as of this writing in the infinite volume $P(\phi)_2$ theories, although we expect it to hold.

In addition to the conditions (A)–(D), we want to make an assumption which guarantees that the matrix elements of H between vectors in $\mathcal{D}_0$ already determine H. We make this precise by assuming:

(E) $\mathcal{D}_0$ is a form core for H.

Note that if H does not satisfy (E), the closure of the Dirichlet form defined by H on $\mathcal{D}_0 \times \mathcal{D}_0$ gives a self-adjoint operator, H_0, which satisfies (A), (B), and (E) above, but H_0 may be totally irrelevant to the physical theory.

The set $\mathcal{D}_0$ may look rather small to be a form core for H, but in fact condition (E) is equivalent to assuming a much larger set of vectors is a form core. To better explain this, we introduce some more notation: Let

$$\|\psi\|_{+1} = \|\sqrt{H+1}\,\psi\|, \quad \psi \in \mathcal{D}(\sqrt{H})$$

and note that with this norm, $\mathcal{D}(\sqrt{H})$ becomes a Hilbert space which we denote $\mathcal{H}_{+1}$. By definition, a linear set $Q \subseteq \mathcal{D}(\sqrt{H})$ is a form core for H if and only if Q is $\|\cdot\|_{+1}$ dense in $\mathcal{H}_{+1}$.

As we mentioned after the proof of Theorem 2.2, the bounds of that theorem allow the extension of $\phi(\cdot)$ from $\mathcal{L}$ to $\mathcal{K} = L^2_{\text{real}}(R^s)$. We assume this to have been done. Define

$$\mathcal{D}_1 = \bigcup_{N=1}^{\infty} \{F(\phi(f_1), \ldots, \phi(f_N))\Omega_0 \in \mathcal{H} : \{f_i\}_{i=1}^N \text{ an orthonormal sequence from } \mathcal{K};\ F \in C^1(R^N),\ \widehat{\partial_i F} \in \mathcal{H},\ i = 1, \ldots, N.$$

Lemma 3.1: Assume (A)–(D) above. Then $\mathcal{D}_1 \subseteq \mathcal{D}(\sqrt{H})$ and $\mathcal{D}_0$ is $\|\cdot\|_{+1}$ dense in $\mathcal{D}_1$. If $\hat{F}, \hat{G} \in \mathcal{D}_1$, then

$$(\hat{F}, H\hat{G}) = \tfrac{1}{2}(\widehat{\nabla F}, \widehat{\nabla G}).$$

Proof: The proof is exactly the proof of Proposition 2.1 after we have established the following facts: For each $f \in \mathcal{K}$,

$$\exp[i\phi(f)]\Omega_0 \in \mathcal{D}(\sqrt{H}), \text{ and if } f_n \to f, \text{ then } \|\exp[i\phi(f)]\Omega_0 - \exp[i\phi(f_n)]\Omega_0\|_{+1} \to 0.$$

The proof of the latter involves a simple computation which we omit.

The basic content of this lemma is that instead of (E) we could have equivalently assumed that $\mathcal{D}_1$ was a form core for H.

The main reason we have assumed (E) is that with it we can prove that $\exp(-tH)$ is positivity preserving. [Here and in the following we assume that we have already made a unitary transformation so that $\mathcal{H} = L^2(M, d\omega)$ with $\phi(f)$ a multiplication operator on $L^2(M, d\omega)$ and Ω_0, the function identically equal to 1. Then $\exp(-tH)$ is positivity preserving means that if $f \in L^2(M, d\omega)$ and $f \geq 0$, then $\exp(-tH)f \geq 0$. This definition as is readily checked, does not depend on (M, ω).] The basic idea behind our proof of the fact that $\exp(-tH)$ is positivity preserving is due to Gross[9] who proved this result using stronger assumptions. (Gross' proof in Ref. 9 contains a gap which we will fill below.)

Theorem 3.2: Assume (A)–(E) above. Then $\{\exp(-tH)\}_{t \geq 0}$ extends to a positivity preserving contraction semigroup on $L^p(M, d\omega)$, $p \in [1, \infty]$, strongly continuous for $p \in [1, \infty)$.

Remarks: 1. All the conclusions of the theorem are independent of the measure space (M, ω). An easy way to see this is to notice that they all follow from the fact that $\exp(-tH)$ is positivity preserving. (See Ref. 24 for example.) In our proof we make a special choice of (M, ω).

2. Assumption (B) is far stronger than is needed in the proof of this theorem.

3. In a recent preprint, Albeverio and Höegh-Krohn[25] have also studied Dirichlet forms on infinite dimensional spaces and their relationship to Markov processes. They have proved a version of Theorem 3.2 under the assumptions that the canonical commutation relations are satisfied in Weyl form and that $\pi(f)\Omega_0 \in L^\infty(M, d\omega)$.

Proof of Theorem 3.2: We first note that since H commutes with complex conjugation we can assume that all L^p spaces are real. The basic idea is to show that for each $F \in \mathcal{D}(H)$, the inequality

$$(F_p, HF) \geq 0 \tag{3.1}$$

holds for $p \in (1, 2]$, with $F_p = (\text{sgn}F)|F|^{p-1}$. The result follows from this inequality and general theorems. Formally Eq. (3.1) is trivial, for $\nabla F_p = (p-1)|F|^{p-2}\nabla F$ and thus $\nabla F_p \cdot \nabla F \geq 0$. However, there are two problems with this formal manipulation. The first is that even for $F \in C_0^\infty(R^1)$, ∇F_p may not be in $L^2(M, d\omega)$ if $p \in (1, 2]$. This problem will be overcome by an analytic continuation argument which allows us to deal with $p < 2$. Secondly, we have yet made sense out of ∇ for arbitrary $F \in \mathcal{D}(H)$. (In his proof, Gross avoided the second problem by making stronger assumptions, but his proof contains a gap because he did not consider the first problem at all.)

Choose an orthonormal basis $\{e_i\}_{i=1}^\infty$ for $\mathcal{K}$ with $e_i \in \mathcal{L}$ and define $V \subseteq \mathcal{L}$ as the set of all finite linear combinations of the e_i's. Let $M = \Pi_{i=1}^\infty \mathring{R}$, the Cartesian product of countably many copies of the one point compactification of the real line. Let

$$C_0 = \text{linear span}\{\exp[i\phi(f)]\Omega_0 : f \in V\}$$

and note that C_0 is dense in $\mathcal{H}$ (and in fact $\|\cdot\|_{+1}$ dense in $\mathcal{H}_{+1}$). Thus by a standard version of the spectral theorem there is a Borel probability measure ω on M and a unitary map U from $\mathcal{H}$ onto $L^2(M, d\omega)$ so that $U\Omega_0 = 1$ and $U\phi(e_i)U^{-1} = M_{x_i}$ (multiplication by the ith coordinate function). In the following we suppress U and identify $\mathcal{H}$ with $L^2(M, d\omega)$.

If $F \in L^2(M, d\omega)$ and $G: R^n \to R$ satisfies $F(x_1, \ldots, x_n, \ldots) = G(x_1, \ldots, x_n)$ for almost every $x = (x_1, \ldots, x_n, \ldots) \in M$, will not distinguish between F and G. Thus define

$$C_1 = \{F \in L^2(M, d\omega) : F \in C^1(R^n) \text{ for some } n \geq 1, \text{ and } F,\ \partial_i F \text{ are bounded for } i = 1, \ldots, n\}.$$

If $F, G \in C_1$ then we have

$$(F, HG) = \tfrac{1}{2} \int d\omega(x) \langle (\nabla F)(x), (\nabla G)(x) \rangle_{l_2},$$

where we have introduced the notation $\langle\ ,\ \rangle_{l_2}$ for the inner product in l_2 and where ∇ is the ordinary gradient. In the following it will be convenient to think of ∇F as a vector in $L^2(M, d\omega; l_2)$. Thus ∇ is a mapping from C_1 into $L^2(M, d\omega; l_2)$. We would like to extend this mapping from C_1 to $\mathcal{D}(\sqrt{H})$. This is easy because the equality

$$\|\nabla F\|^2_{L^2(M, d\omega; l_2)} = 2 \|\sqrt{H} F\|^2_{L^2(M, d\omega)} \tag{3.2}$$

says that ∇ is a continuous map from the dense subset C_1 of $\mathcal{H}_{+1}$ into $L^2(M, d\omega; l_2)$ and thus has a unique continuous extension to $\mathcal{H}_{+1}$.

We need to know that in some sense ∇ acts like differentiation. In particular, let $g \in C^1(R)$, with $|g'(x)| \leq C$. We claim that if $F \in \mathcal{D}(\sqrt{H})$, then $g \circ F$ ($\circ$ means composition) is in $\mathcal{D}(\sqrt{H})$ and

$$\nabla g \circ F = (g' \circ F) \nabla F. \tag{3.3}$$

To prove this, choose $F_n \in C_1$ with $\|F_n - F\|_{+1} \to 0$ and $F_n \to F$ pointwise a.e. Then since $|g(x)| \leq C|x| + |g(0)|$, $g \circ F_n \in \mathcal{H}$ and since $F_n \in C^1(R^m)$ for some $m = m(n)$, we have $g \circ F_n \in C_1$ and

$$\nabla g \circ F_n = (g' \circ F_n) \nabla F_n.$$

Since $g' \circ F$ is bounded, $(g' \circ F) \nabla F \in L^2(M, d\omega; l_2)$ and as is easily seen

$$\begin{aligned}
&\|\nabla g \circ F_n - (g' \circ F) \nabla F\|_{L^2(M, d\omega; l_2)} \\
&\quad \leq \|(g' \circ F_n - g' \circ F) \nabla F\|_{L^2(M, d\omega; l_2)} \\
&\qquad + C \|\nabla (F_n - F)\|_{L^2(M, d\omega; l_2)} \to 0,
\end{aligned}$$

as $n \to \infty$. (The first term $\to 0$ because $|g' \circ F_n - g' \circ F| \to 0$ a.e. and is dominated by $2C$.) Thus $\nabla g \circ F_n$ converges to $(g' \circ F) \nabla F$ in $L^2(M, d\omega; l_2)$. In particular $\nabla g \circ F_n$ is Cauchy and thus by Eq. (3.2) so is $\sqrt{H}\, g \circ F_n$. Since $\sqrt{H}$ is closed $g \circ F$ is in $\mathcal{D}(\sqrt{H})$ and by the definition of ∇ on $\mathcal{H}_{+1}$, Eq. (3.3) holds.

We now prove a version of the inequality Eq. (3.1). Suppose $F \in \mathcal{D}(H)$. Define for $\epsilon > 0$, $p \in (1, \infty)$,

$$g_1(x) = x(1 + \epsilon x^2)^{-1/2}, \quad g_2(x) = (\mathrm{sgn} x) |x|^{p-1}, \quad g = g_2 \circ g_1,$$

$$F_\epsilon = g_1 \circ F, \quad (F_\epsilon)_p = g_2 \circ F_\epsilon = g \circ F,$$

$$\mathrm{sgn} x = \begin{cases} x/|x|, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Note that $g_1'(x) = (1 + \epsilon x^2)^{-3/2}$, $g_2'(x) = (p-1)|x|^{p-2}$ are continuous for $p \geq 2$, and g' is bounded. Thus since $\mathcal{D}(\sqrt{H}) \supseteq \mathcal{D}(H)$, $g \circ F = (F_\epsilon)_p \in \mathcal{D}(\sqrt{H})$ for $p \geq 2$, and

$$\nabla (F_\epsilon)_p = (p-1) |F_\epsilon|^{p-2} (1 + \epsilon F^2)^{-3/2} \nabla F, \quad p \geq 2.$$

Thus for $p \geq 2$,

$$((F_\epsilon)_p, HF) = \tfrac{1}{2}(p-1) \int d\omega |F_\epsilon|^{p-2} (1 + \epsilon F^2)^{-3/2} \langle \nabla F, \nabla F \rangle_{l_2}. \tag{3.4}$$

Define $h(p) = (p-1)^{-1} ((F_\epsilon)_p, HF)$ for $p \in (1, \infty)$ and notice that because $(F_\epsilon)_p$ is an $L^2(M, d\omega)$ valued analytic function of p in $\mathrm{Re}\, p > 1$, h also has this property. We want to show that $h(p)$ is nonnegative for $p \in (1, 2]$. For this purpose note that for $p \geq 2$

$$2(\sqrt{\epsilon})^{p-2} h(p) = \int d\omega |Q|^{p-2} R, \tag{3.5}$$

with $Q = |\sqrt{\epsilon} F_\epsilon| \leq 1$ and $R = (1 + \epsilon F^2)^{-3/2} \langle \nabla F, \nabla F \rangle_{l_2}$.

The right-hand side of Eq. (3.5) is analytic in $\mathrm{Re}\, p > 2$ while the left is analytic in $\mathrm{Re}\, p > 1$. Thus to calculate the left-hand side for $p \in (1, 2]$, we can choose $p_0 > 2$ and expand the right side in a power series around p_0. This gives

$$2(\sqrt{\epsilon})^{p-2} h(p) = \sum_{n=0}^{\infty} \int \frac{[(p - p_0) \ln Q]^n}{n!} |Q|^{p_0 - 2} R\, d\omega.$$

Each term in the series is nonnegative for $p \in (1, 2]$ so that $h(p) \geq 0$ in this region. Thus again for $p \in (1, 2]$, $F \in \mathcal{D}(H)$ and $\lambda \geq 0$,

$$\|(F_\epsilon)_p\|_{p'} \|(H + \lambda) F\|_p \geq ((F_\epsilon)_p, (H + \lambda) F) \geq \lambda ((F_\epsilon)_p, F).$$

Taking the limit $\epsilon \downarrow 0$, yields for $p \in (1, 2]$, $\lambda \geq 0$,

$$\|(H + \lambda) F\|_p \geq \lambda \|F\|_p \quad \forall F \in \mathcal{D}(H).$$

With $R(z) = (H - z)^{-1}$ we conclude that

$$\|R(-\lambda) G\|_p \leq \lambda^{-1} \|G\|_p, \quad \text{all } G \in L^2(M, d\omega), \quad \lambda > 0.$$

Thus the closure of $R(-\lambda)$ in L^p (which we denote by $R_p(-\lambda)$) satisfies the same bound. Since $R(-\lambda)$ and $R_p(-\lambda)$ agree on L^2, R_p is a pseudo-resolvent. We now refer to an argument of Yosida[26] which shows that R_p is the resolvent of a closed linear operator, H_p. The main point here is that one can show that the null space of $R_p(-\lambda)$ is empty, a fact which is equivalent to the condition that H is closable in L^p, with closure H_p. The Hille–Yosida theorem then shows that $\exp(-tH_p)$ is a contraction semigroup on L^p [which agrees with $\exp(-tH)$ on L^2]. By duality $\exp(-tH) \upharpoonright_{L^p}$, $p \in [2, \infty)$ is a contraction semigroup and by a limiting argument the same is true on L^1 and L^∞. If $1 \geq f \geq 0$ we have

$$\exp(-tH)(1 - f) \leq 1 \quad \text{a.e.},$$

and thus $\exp(-tH)\, 1 = 1$ implies $\exp(-tH) f \geq 0$. Thus $\exp(-tH)$ is positivity preserving. The strong continuity of the semigroups on L^p, $p \in [1, \infty)$ follows from the strong continuity on L^2. This completes the proof.

As was first noted by Simon[11] in the context of quantum field theory (see also Klein and Landau[24] for another proof) the fact that $\exp(-tH)$ is positivity preserving means that it is the transition function for a Markov process and that one is thus dealing with a "Euclidean" field theory. Specifically, we have the following theorem.

Theorem 3.3: Assume (A)–(E). Then there exists a probability space (Q, Σ, μ) with $Q = \mathcal{S}'(R^{s+1})$, Σ the σ-algebra generated by the Borel cylinder sets of $\mathcal{S}'(R^{s+1})$ which satisfies:

(a) Let T_t and R be respectively the point transformations on $\mathcal{S}'(R^{s+1})$ implementing time translations and reflections in the $t=0$ hyperplane. Then R and T_t are measure preserving and μ is ergodic with respect to T_t.

(b) If $\delta_n(t)=n\delta_1(nt)$ is a delta converging sequence [we take $\delta_1\in \mathcal{S}(R^1)$, $\delta_1\geq 0$, $\|\delta_1\|_1=1$], then for $f\in \mathcal{S}(R^s)$ the limit

$$\lim_{n\to\infty}\Phi(f\otimes\delta_n)\equiv\Phi(f\otimes\delta_0)\quad[\Phi\in\mathcal{S}'(R^{s+1})]$$

exists in every $L^p(\mathcal{S}'(R^{s+1}),\Sigma,d\mu)$ for $p<\infty$. We denote by Σ_0 the σ-algebra generated by $\{\Phi(f\otimes\delta_0):f\in\mathcal{S}(R^s)\}$ and by E_0 the conditional expectation relative to Σ_0.

(c) Let $(U(t)F)(\Phi)=F(T_t\Phi)$. Then there exists a measure preserving isometry, j, from $L^p(M,d\omega)$ onto $L^p(\mathcal{S}'(R^{s+1}),\Sigma_0,d\mu)$, $p\in[1,\infty]$ with $j(\exp(-|t|H))j^{-1}=E_0U(t)\restriction\mathcal{H}_0$, $\mathcal{H}_0=L^2(\mathcal{S}'(R^{s+1}),\Sigma_0,d\mu)$ and $j\phi(f)=\Phi(f\otimes\delta_0)$.

(d) Suppose $F_1,\ldots,F_n\in\bigcap_{p<\infty}L^p(M,d\omega)$ and $t_1\leq t_2\leq\cdots\leq t_n$. Then

$$(\Omega_0,F_1\exp[-(t_2-t_1)H]F_2\cdots\exp[-(t_n-t_{n-1})H]F_n\Omega_0)$$
$$=\int d\mu\,(U(t_1)jF_1)\cdots(U(t_n)jF_n).$$

In particular with $\Phi(f\otimes\delta_t)=U(t)\,\Phi(f\otimes\delta_0)$ we have

$$(\Omega_0,\phi(f_1)\exp[-(t_2-t_1)H]\,\phi(f_2)\cdots\exp[-(t_n-t_{n-1})H]\,\phi(f_n)\,\Omega_0)$$
$$=\int d\mu\,\Phi(f_1\otimes\delta_{t_1})\cdots\Phi(f_n\otimes\delta_{t_n}).\tag{3.6}$$

(e) The characteristic function $\int d\mu(\Phi)\exp[i\Phi(F)]$, $F\in\mathcal{S}(R^{s+1})$, is continuous in the topology of $\mathcal{S}(R^{s+1})$ and uniquely determines μ.

The proof of this theorem will not be given since in all essentials it is contained in Ref. 11 (and also in Ref. 24). We remark that the reason we can take $\mathcal{S}'$ for our measure space and the reason that (b) is true is the fact that the left-hand side of (3.6) is jointly continuous in $(t_1,\ldots,t_n)$ and in $(f_1,\ldots,f_n)$. Thus we can apply Minlos' theorem.[27] The continuity in turn follows from the strong continuity of $\exp(-tH)$ on $L^p(M,d\omega)$ and the fact that $\|\phi(f)\|_p\leq c_p\|f\|_2$ for $p\in[1,\infty)$.

As we have just remarked, the sharp time "Euclidean" Green's functions

$$S_N(\underline{f};\underline{t})\equiv\int d\mu\,\Phi(f_1\otimes\delta_{t_1})\cdots\Phi(f_N\otimes\delta_{t_N})$$

are continuous in $\underline{t}$ $(\underline{t}\in R^N)$. They also satisfy the bound

$$|S_N(\underline{f}:\underline{t})|\leq\prod_{i=1}^N\|\Phi(f_i\otimes\delta_0)\|_N\leq C^N N!\prod_{i=1}^N\|f_i\|_2.$$

Thus, if $F_1,\ldots,F_N\in\mathcal{S}(R^{s+1})$, we have

$$\int d\mu\,\Phi(F_1)\cdots\Phi(F_N)\equiv S_N(\underline{F})$$

with

$$|S_N(\underline{F})|\leq C^N N!\int d^Nt\prod_{i=1}^N\|F_{it_i}\|_2,$$

where $F_t(\mathbf{x})=F(\mathbf{x},t)$. Thus

$$|S_N(\underline{F})|\leq C_1{}^N N!\prod_{i=1}^N|F_i|_{\mathcal{S}},\tag{3.7}$$

where $|F|_{\mathcal{S}}=\|(t^2+1)^{1/2}F\|_{L^2(R^{s+1})}$. In particular S_N extends (by the nuclear theorem) to a tempered distribution on $\mathcal{S}(R^{N(s+1)})$.

We have used the word "Euclidean" in quotation marks because so far no assumption of Euclidean covariance has been introduced. The following assumption (our last) should be true in any theory of a Hermitian scalar Wightman field:

(f) Let $\Omega_N=\{(x_1,\ldots,x_N)\in R^{N(s+1)}:x_i-x_j\neq 0 \text{ if } i\neq j\}$ and suppose $F\in C_0^\infty(\Omega_N)$. Then

$$S_N(F)=S_N(F_{(\Lambda,a)}),\tag{3.8}$$

where $F_{(\Lambda,a)}(x_1,\ldots,x_N)=F(\Lambda x_1+a,\ldots,\Lambda x_N+a)$ and Λ is a proper rotation.

Theorem 3.4: Assume (A)–(F). Then (3.4) holds for any $F\in\mathcal{S}(R^{N(s+1)})$.

Proof: We write $S_N(F_{(\Lambda,a)})=T_{(\Lambda,a)}(F)$ defining the tempered distribution $T_{(\Lambda,a)}$. To show that $T_{(\Lambda,a)}$ is independent of (Λ,a) it is enough to show that the derivatives of $T_{(\Lambda,a)}$ with respect to the group parameters all vanish, and because of the group property, this need only be done at the identity, $(I,0)$. (A similar technique has been used by Nelson.[28]) We illustrate with a rotation in the (x_1,t) plane where $(d/d\lambda)\,T_{(R_\lambda,0)}(F)\equiv T'(F)\equiv T(F')$ with

$$F'(\mathbf{x}_1,t_1,\ldots,\mathbf{x}_N,t_N)=\sum_{j=1}^N\left(t_j\frac{\partial}{\partial x_{1j}}-x_{1j}\frac{\partial}{\partial t_j}\right)F.$$

We need only show $T'(F)=0$ for F of the form $f_1\otimes\cdots\otimes f_N\otimes f$ where $f_i\in C_0^\infty(R^s)$ is a function of the space variables and $f\in C_0^\infty(R^N)$ is a function of the time variables, since this set is total in $\mathcal{S}(R^{N(s+1)})$ in the topology of $\mathcal{S}$.

Choose a function $g\in C_0^\infty(R)$ with the properties

(a) $0\leq g\leq 1$,

(b) $g(t)=0$ for $|t|\leq 1$, $g(t)=1$ for $|t|\geq 2$,

and define

(i) $g_\lambda^{ij}(t_1,\ldots,t_N)=g(\lambda(t_i-t_j))$.

(ii) $F_{\{\lambda\}}=\left(\prod_{i>j}g_{\lambda_{ij}}^{ij}\right)f_1\otimes\cdots\otimes f$,

with $\{\lambda\}$ denoting the collection $\{\lambda_{ij}:N\geq i>j\geq 1\}$. Note that since $F_{\{\lambda\}}\in C_0^\infty(\Omega_N)$,

$$T'(F_{\{\lambda\}})=0.$$

We will show that taking one λ_{ij} to ∞ at a time gives

$$\lim T'(F_{\{\lambda\}})=T'(F),$$

with $F=f_1\otimes\cdots\otimes f_N\otimes f$, and this will complete the proof. We thus consider

$$T'(F_{\{\lambda\}})-T'(F_{\{\lambda\}'}),$$

where $F_{\{\lambda\}'}$ is the same as $F_{\{\lambda\}}$ except that $g_{\lambda_{lm}}^{lm}$ is replaced by 1. The above expression is a sum of two terms of the form

(i) $\int dt_1\cdots dt_N\,G(t_1,\ldots,t_N)(g_{\lambda_{lm}}^{lm}-1)$,

(ii) $\int dt_1\cdots dt_N\,H(t_1,\ldots,t_N)(g_{\lambda_{lm}}^{lm})'$,

where G and H are continuous functions of compact support and the prime denotes differentiation with respect to t_l. The expression (i) tends toward zero as $\lambda_{lm}\to\infty$ by dominated covergence. The expression (ii) can be written

$$\int dt K(t) g'(\lambda t)\lambda,$$

where K is continuous and has compact support and $\lambda = \lambda_{Im}$. However, a change of variables shows that the above is just

$$\int g'(t)[K(t/\lambda) - K(0)]\,dt,$$

which has limit zero as $\lambda \to \infty$ because K is continuous. This concludes the proof.

The idea for the following theorem is due to Fröhlich.[10]

Theorem 3.5: Assume (A)–(F). Then the measure μ of Theorem 3.3 is Euclidean invariant.

Proof: Consider the functional

$$\int d\mu \exp[it\Phi(F_{(\Lambda,a)})] = g(t, (\Lambda, a)).$$

Then because of the bound (3.7), $g(t,(\Lambda,a))$ and $g(t,(I,0))$ are analytic in t for $|\mathrm{Im}t| < C$ with $C > 0$. Expanding $\exp(it(F_{(\Lambda,a)})$ in a power series around $t = 0$, we see from Theorem 3.4 that

$$g(t,(\Lambda,a)) = g(t,(I,0)) \quad \text{for } |t| < C. \tag{3.9}$$

Thus by the analyticity of both sides of Eq. (3.9), this equation holds for all real t.

Define the measure $\mu_{(\Lambda,a)}$ by the equation

$$\mu_{(\Lambda,a)}(A) = \mu((\Lambda,a)A),$$

where $(\Lambda,a)A = \{\Phi_{(\Lambda,a)} : \Phi \in A\}$ with $\Phi_{(\Lambda,a)}(F) = \Phi(F_{(\Lambda,a)})$. Then $\mu_{(\Lambda,a)}$ is a measure on Σ satisfying

$$\int d\mu_{(\Lambda,a)} \exp[i\Phi(F)] = \int d\mu \exp[i\Phi(F)].$$

By the uniqueness part of Minlos' theorem [part (e) of Theorem 3.3] $\mu_{(\Lambda,a)} = \mu$ and thus μ is Euclidean invariant.

We have not yet shown that our assumptions (A)–(F) imply the Wightman axioms. It is clear from the bound (3.7) and Theorem 3.3 that the S_n obey the Osterwalder–Schrader axioms[29] for Euclidean Green's functions, and thus the analytic continuation of the S_n to real time leads to a field theory satisfying all of the Wightman axioms. However, what is not clear is the relationship of the relativistic field to the time zero field, $\phi(f)$. In the following we will show exactly how the smeared relativistic field is related to $\phi(f)$ and in the process explicitly make the analytic continuation to real time. This analytic continuation can be done in one step thanks to a theorem of Stein[30] (based on the Stein interpolation theorem):

Theorem 3.6 (*Stein*[30]): Suppose $P(t) = \exp(-tH)$ is a self-adjoint strongly continuous positivity preserving contraction semigroup on $L^2(M, d\omega)$. Then as an operator on L^p, $P(t)$ has an analytic continuation, $P(z)$, to the sector $S_p = \{z : |\arg z| < \pi/2(1 - |2/p - 1|)\}$ satisfying

(a) $P(z_1)P(z_2) = P(z_1 + z_2), \quad z_1, z_2 \in S_p$,

(b) $\|P(z)\|_{p,p} \le 1, \quad z \in S_p$.

Here and in the following we denote by $\|\ \|_{p,q}$ the norm of an operator as a map from $L^p(M,d\omega)$ to $L^q(M,d\omega)$. We will also write $P(z) = \exp(-zH)$.

Let $\underline{f} = (f_1, \ldots, f_n)$, $\underline{z} = (z_1, \ldots, z_n)$, and for $\mathrm{Re} z_j > 0$; $j = 2, \ldots, n$, $\mathrm{Re} z_1 \ge 0$ define

$$F_n(\underline{f};\underline{z}) \equiv \exp(-z_1H)\phi(f_1)\exp(-z_2H)\phi(f_2)\cdots \times \exp(-z_nH)\phi(f_n)\Omega_0. \tag{3.10}$$

Proposition 3.7: Assume (A)–(F). Then in the region $S = \{\underline{z} : \mathrm{Re} z_j > 0,\ j = 1, 2, \ldots, n\}$, $F_n(\underline{f};\underline{z})$ is an $L^2(M,d\omega)$ valued analytic function satisfying the bound

$$\|F_n(\underline{f};\underline{z})\|_2 \le (A^n n! \prod_{i=1}^n \|f_i\|_2)(1 + \xi^{n-1}), \tag{3.11}$$

where

$$\xi = \max_{j\ge 2}\{|t_j|/\tau_j\}, \quad z_j = \tau_j + it_j.$$

Proof: The stated analyticity follows from Theorem 3.6 and the fact that $\phi(f)$ is a bounded map from L^p to L^q if $\infty \ge p > q$. To prove (3.11), note that

$$\|F_n(\underline{f};\underline{z})\|_2 \le \|\phi(f_1)\|_{p_2,2}\|\phi(f_2)\|_{p_3,p_2}\cdots \times \|\phi(f_{n-1})\|_{p_n,p_{n-1}}\|\phi(f_n)\|_{\infty,p_n},$$

if $|\arg z_2| < \pi/p_2, \ldots, |\arg z_n| < \pi/p_n$. We also have $\|\phi(f_j)\|_{p_{j+1},p_j} \le \|\phi(f_j)\|_{r_j} \le Cr_j\|f_j\|_2$ with $1/r_j = 1/p_j - 1/p_{j+1}$. This follows from Hölder's inequality and the bound Eq. (2.12) (we set $p_1 = 2, p_{n+1} = \infty$). We consider only the case $n > 1$ since the result is otherwise trivial.

We choose p_j for $j = 2, \ldots, n$ by requiring

$$1 - 2/p_j = \gamma_j\lambda, \quad \lambda = \min_{j\ge 2}(1 - (\tfrac{1}{2}\pi)^{-1}|\arg z_j|),$$

where the $\{\gamma_j\}_{j=2}^n$ will be specified in a moment. For now, notice that if $\gamma_j < 1$, then

$$1 - 2/p_j < \lambda \le 1 - (\tfrac{1}{2}\pi)^{-1}|\arg z_j|, \quad j = 2, \ldots, n,$$

so that the requirement $|\arg z_j| < \pi/p_j$ is satisfied. The requirement that r_j be in the interval $[1,\infty)$ can also be checked for our choice of the γ_j. We choose

$$\gamma_j = \frac{1}{2}\left[1 - \beta\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n-j+1}}\right)\right], \quad j = 2, \ldots, n,$$

with

$$\beta = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n-1}}\right)^{-1}.$$

Then

$$\prod_{j=1}^n r_j = 2^n\lambda^{-(n-1)}\gamma_2^{-1}(\gamma_3 - \gamma_2)^{-1}\cdots(\gamma_n - \gamma_{n-1})^{-1}(1 - \lambda\gamma_n)^{-1}$$

and

$$\gamma_2 = \tfrac{1}{4}, \quad (\gamma_{j+1} - \gamma_j)^{-1} = 4\sqrt{n-j+1}\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n-1}}\right)$$

$$\text{for } j = 2, \ldots, n-1, \quad \gamma_n = \tfrac{1}{2}(1 - \beta).$$

Thus

$$\prod_{j=1}^n r_j \le C^n\lambda^{-(n-1)}\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n-1}}\right)^{n-1}\sqrt{(n-1)!}$$

$$\le d^n n!\,\lambda^{-(n-1)}.$$

Thus $\|F_n(\underline{f};\underline{z})\|_2 \le \alpha^n n!\,\lambda^{-(n-1)}\prod_{j=1}^n\|f_j\|_2$. If $\lambda = 1 - |\arg z_l|(\tfrac{1}{2}\pi)^{-1} \le \tfrac{1}{2}$, then $\lambda = (\tfrac{1}{2}\pi)^{-1}\tan^{-1}(\tau_l/|t_l|) \ge C_1(\tau_l/|t_l|)$ or $\lambda^{-1} \le C_2\max_{j=2,\ldots,n}(|t_j|/\tau_j)$. This completes the proof.

We make two remarks about the above proof. Firstly, if one is willing to settle for an unknown n dependent constant C_n instead of the factor $A^n n!$, then the proof simplifies a great deal. Secondly, if one only knows that $|(\Omega_0, \phi(f)^n\Omega_0)| \le (n!)^L |f|^n$, for some norm $|\cdot|$, then the same argument provides the bound $\|F_n(\underline{f};\underline{z})\|_2 \le A^n(n!)^L(1+\xi^{(n-1)L})\prod_{j=1}^n |f_j|$ which is certainly good enough to construct the Wightman field.

We now define the smeared relativistic field $\theta(f)$ for $f\in \mathcal{S}_{\text{real}}(R^{s+1})$. Define

$$D_1 = \text{linear span} \bigcup_{n=0}^{\infty} \{F_n(\underline{f};\underline{z}) : f_i \in \mathcal{S}_{\text{real}}(R^s),$$

$$i=1,\ldots,n \text{ and } \underline{z}\in S\}.$$

For $f\in \mathcal{S}_{\text{real}}(R^{s+1})$, let $f_t(\mathbf{x}) = f(\mathbf{x},t)$ and define for $\psi\in D_1$,

$$\theta_1(f)\,\psi = \int dt \exp(itH)\,\phi(f_t)\exp(-itH)\,\psi. \tag{3.12}$$

Note that if $\psi = F_n(\underline{f};\underline{z})$ then

$$\phi(f_t)\exp(-itH)\,\psi = F_{n+1}(f_t,\ldots,f_n; 0, z_1+it,\ldots,z_n)$$

so that $\phi(f_t)\exp(-itH)\psi$ is continuous in t and by Proposition 3.7 satisfies the bound

$$\|\phi(f_t)\exp(-itH)\,\psi\|_2 \le C(1+|t|^n)\,\|f_t\|_2.$$

Thus the operator $\theta_1(f)$ is well defined on D_1 and as one easily checks is symmetric. Let $\theta_2(f)$ be its closure.

Theorem (3.8): Let $f_1,\ldots,f_n\in\mathcal{S}_{\text{real}}(R^{s+1})$. Then Ω_0 is in the domain of $\theta_2(f_1)\cdots\theta_2(f_n)$. Let D_0 be the linear span of

$$\bigcup_{n=0}^{\infty}\{\theta_2(f_1)\cdots\theta_2(f_n)\,\Omega_0 : f_i\in\mathcal{S}_{\text{real}}(R^{s+1})\}.$$

Then the field

$$\theta(f) = \theta_2(f)\restriction D_0$$

satisfies the Wightman axioms for a Hermitian scalar field.

Proof: For $f_i\in\mathcal{S}_{\text{real}}(R^{s+1})$, $i=1,\ldots,n$, define the $L^2(M,d\omega)$ valued functions

$$\begin{aligned}G_n(\underline{\tau},\underline{t};f_t) &\equiv \exp[-(\tau_1 - it_1)H]\,\phi(f_{1t_1})\\ &\quad\times\exp[-(\tau_2+i(t_1-t_2))H]\,\varphi(f_{2t_2})\cdots\\ &\quad\times\exp[-(\tau_n+i(t_{n-1}-t_n))H]\,\phi(f_{nt_n})\,\Omega_0\\ &= F_n(\underline{f_t};\tau_1-it_1,\tau_2+i(t_1-t_2),\ldots,\tau_n+i(t_{n-1}-t_n))\end{aligned}$$

and the functions

$$G_n(\underline{\tau};\underline{f}) \equiv \int dt_1\cdots dt_n\, G_n(\underline{\tau},\underline{t};\underline{f_t})$$

where again $f_t(\mathbf{x}) = f(\mathbf{x},t)$.

Note that from Proposition 3.7 we have

$$\|G_n(\underline{\tau};\underline{f})\|_2 \le \frac{B^n n!}{(\tau_{\min})^{n-1}}\int dt_1\cdots dt_n(1+|t|)^{n-1}\prod_{j=1}^n\|f_{jt_j}\|_2$$

with $|t| = (t_1^2+\cdots+t_n^2)^{1/2}$, $\tau_{\min} = \min_{j\ge 2}\{\tau_j\}$, and $\tau_{\min}\le 1$. By standard techniques of distribution theory[31] G_n extends to a continuous function of $(\tau_1,\ldots,\tau_n)$ for $\tau_i\in[0,1]$. We will show that

$$G_n(\underline{0};\underline{f}) = \theta_2(f_1)\cdots\theta_2(f_n)\,\Omega_0.$$

First note that $G_n(\underline{\tau},t;f_t)\in D_1$ for $\tau_i>0$ and that we can find a linear combination of such vectors to form a Riemann sum approximation, G_n^N, to the vector $G_n(\underline{\tau};\underline{f})$ ($\tau_i>0$ for all i) with the property

$$G_n^N \to G_n(\underline{\tau};\underline{f}),$$

$$\theta_1(g)\,G_n^N \to G_{n+1}(0,\tau_1,\ldots,\tau_n;g,f_1,\ldots,f_n),$$

$$\tau_i>0,\ i=1,\ldots,n.$$

The fact that $\theta_2(g)$ is the closure of $\theta_1(g)$ then implies $G_n(\underline{\tau};\underline{f})\in\mathcal{D}(\theta_2(g))$ and

$$\theta_2(g)\,G_n(\underline{\tau};\underline{f}) = G_{n+1}(0,\tau_1,\ldots,\tau_n;g,f_1,\ldots,f_n)\quad\text{if } \tau_i>0. \tag{3.13}$$

We now proceed by induction. First note that $\theta_2(f_n)\,\Omega_0 = G_1(0;f_n)$. Suppose we have already shown

$$\theta_2(f_{j+1})\cdots\theta_2(f_n)\,\Omega_0 = G_{n-j}(\underline{0};f_{j+1},\ldots,f_n). \tag{3.14}$$

Using Eq. (3.13), we have for $\tau_{j+1},\ldots,\tau_n>0$

$$\begin{aligned}&\theta_2(f_j)\,G_{n-j}(\tau_{j+1},\ldots,\tau_n;f_{j+1},\ldots,f_n)\\ &\quad= G_{n+1-j}(0,\tau_{j+1},\ldots,\tau_n;f_j,\ldots,f_n).\end{aligned}$$

Thus the continuity of G_{n-j} and G_{n+1-j} in τ imply

$$\begin{aligned}&\lim_{\tau_{j+1},\ldots,\tau_n\downarrow 0}\theta_2(f_j)\,G_{n-j}(\tau_{j+1},\ldots,\tau_n;f_{j+1},\ldots,f_n)\\ &\quad= G_{n+1-j}(\underline{0};f_j,\ldots,f_n)\end{aligned}$$

and

$$\begin{aligned}&\lim_{\tau_{j+1},\ldots,\tau_n\downarrow 0} G_{n-j}(\tau_{j+1},\ldots,\tau_n;f_{j+1},\ldots,f_n)\\ &\quad= G_{n-j}(\underline{0};f_{j+1},\ldots,f_n).\end{aligned}$$

Since $\theta_2(f_j)$ is closed, we conclude $G_{n-j}(\underline{0};f_{j+1},\ldots,f_n)\in\mathcal{D}(\theta_2(f_j))$ and

$$\theta_2(f_j)\,G_{n-j}(\underline{0};f_{j+1},\ldots,f_n) = G_{n+1-j}(\underline{0};f_j,\ldots,f_n),$$

which when (3.14) is taken into account reads

$$\theta_2(f_j)\cdots\theta_2(f_n)\,\Omega_0 = G_{n+1-j}(\underline{0};f_j,\ldots,f_n).$$

This concludes the proof of the first statement in the theorem.

The Poincaré covariance of the distributions determined by $(\Omega_0, \theta(f_1)\cdots\theta(f_n)\,\Omega_0)$ follows from the Euclidean covariance of the Euclidean Green's functions in a standard way as does locality from the symmetry of these functions. The uniqueness of the vacuum is obvious from our assumptions. Cyclicity of the vacuum is a consequence of the fact that the vectors $F_n(\tau_1,\ldots,\tau_n;f_1,\ldots,f_n)$ $\tau_i>0$ are limits of linear combinations of $\theta(g_1)\cdots\theta(g_n)\,\Omega_0$. Finally the spectral condition follows from $H\ge 0$ and Lorentz invariance.

IV. CONCLUDING REMARKS

We have investigated some of the consequences of the Araki formula in quantum field theory. Here we will mention some problems which we have not considered.

Because it was not necessary for the results which we have obtained, we did not assume the existence of a self-adjoint operator, π. However, the structure of a theory satisfying the Wightman axioms is quite rich, and thus it is not clear that the axiom scheme presented in Sec. III does not already imply that the theory is canonical in the Weyl sense. If the theory is canonical in the Weyl sense, there is an interesting question which can be asked: The Weyl relations imply that the measure ω [which can be thought of as a measure on $\mathcal{S}'(R^s)$] is quasi-invariant (and conversely). Is the corresponding Euclidean measure, μ, $\mathcal{S}(R^{s+1})$ quasi-invariant? If this is so, is the Radon—Nikodym derivative $d\mu(\Phi+f)/d\mu(\Phi)$ measurable relative to the σ-algebra generated by $\{\Phi(g): \mathrm{supp} g \subseteq \mathrm{supp} f\}$? Positive answers to these questions could lead to a parametrization of such theories in terms of an interaction density. (This has been discussed by Fröhlich.[16])

Finally another interesting problem not considered here is the question of self-adjointness and locality (in the sense of commuting spectral projections) of the Wightman field, $\theta(f)$.

APPENDIX A: PROOF OF PROPOSITION 2.1

Let μ be the probability measure on R^N (guaranteed to exist by the spectral theorem) for which

$$(\hat{F}, \hat{G}) = \int d\mu(x_1, \ldots, x_N)\, \bar{F}(x_1, \ldots, x_N)\, G(x_1, \ldots, x_N). \quad \text{(A1)}$$

Assume first that the functions F_i of the proposition are in the Schwartz space $\mathcal{S}(R^N)$ with Fourier transforms $\tilde{F}_i \in C_0^\infty(R^N)$. The Riemann approximations to the integrals

$$F_i(x) = (2\pi)^{-N/2} \int d^N p\, \tilde{F}_i(p) \exp(ix\cdot p)$$

are just finite sums of exponentials for which Eq. (2.6) is the same as Eq. (2.3). We can find sequences F_{in} of Riemann sums with $\|F_{in} - F_i\|_\infty$ and $\|\partial_j(F_{in} - F_i)\|_\infty \to 0$ and thus by dominated convergence in Eq. (A1), $\widehat{F_{in}} \to \widehat{F_i}$, $\widehat{\partial_j F_{in}} \to \widehat{\partial_j F_i}$. Because $\sqrt{H}$ is closed, Eq. (2.6) follows for $\tilde{F}_i \in C_0^\infty$. By using the same reasoning it is easy to extend this result to all $F_i \in \mathcal{S}(R^N)$ by multiplying by a suitable sequence of smooth cutoff functions in p-space. If now F_i is continuously differentiable with F_i and ∇F_i bounded, convolution with a suitable sequence of approximate delta functions ρ_n produces $\|\cdot\|_\infty$ convergence of $\rho_n * F_i$ and $\nabla\rho_n * F_i$ and thus leads to Eq. (2.6) for such functions. Finally, if F_i and ∇F_i are not bounded, but are continuous, then multiplication by $\chi_n(x) = \chi(x/n)$ with $\chi \in C_0^\infty(R^N)$, $\chi(x) = 1$ for $|x| \le 1$ yields (2.6) by a dominated convergence argument if $\widehat{F_i}$ and $\widehat{\partial_j F_i}$ are in $\mathcal{H}$.

APPENDIX B: PROOF OF THEOREM 2.2

We will show that if $f_0 \in \mathcal{L}$ with $\langle f_0, f_0\rangle = 1$, then $\|\exp[t\phi(f_0)]\Omega_0\| < \infty$ if $t^2/2m < 1$.

By the spectral theorem we have

$$(F(\phi(f_0))\Omega_0, G(\phi(f_0))\Omega_0) = \int d\mu(x)\, \bar{F}(x)\, G(x)$$

for some Borel probability measure μ on R.

Let $Q = \{f \in L^2(R, d\mu): f, f' \text{ are continuous}; f' \in L^2(R, d\mu)\}$. Then for $f, g \in Q$ the equation

$$(\hat{f}, H\hat{g}) = \tfrac{1}{2} \int \bar{f}'(x) g'(x)\, d\mu(x) \equiv h(f, g) \quad \text{(B1)}$$

defines a sesquilinear form h on $Q \times Q$ with $h(f,f) \ge 0$. It is clear that since $\sqrt{H}$ is closed, h is closable and its closure $\bar{h}$ (with form domain $\bar{Q}$) satisfies

$$\bar{h}(f,g) = (\hat{f}, H\hat{g}) \quad \text{for all } f, g \in \bar{Q}.$$

Thus there exists a nonnegative self-adjoint operator K in $L^2(R, d\mu)$ with $\mathcal{D}(\sqrt{K}) = \bar{Q}$ such that for all $f, g \in \bar{Q}$

$$(f, Kg) \equiv (\sqrt{K} f, \sqrt{K} g) = (\hat{f}, H\hat{g}). \quad \text{(B2)}$$

Note that since $\|\sqrt{H}\,\hat{f}\|^2 \ge m(\|\hat{f}\|^2 - |(\Omega_0, \hat{f})|^2)$ for all $f \in \bar{Q}$, Eq. (B2) implies

$$\|\sqrt{K} f\|_2^2 \ge m(\|f\|_2^2 - |(1, f)|^2), \quad f \in \mathcal{D}(\sqrt{K})$$

and thus $K \restriction_1^{\perp} \ge m > 0$.

We now define another form on $Q \times Q$ [it is just $\pi(f_0)$ in disguise] by the equation

$$\rho(f,g) = \tfrac{1}{2} i \int d\mu\, (\bar{f}' g - \bar{f} g'). \quad \text{(B3)}$$

Note that ρ is symmetric. Also note that for $f, g \in Q$, $\lambda \in R$ we have

$$(\exp(i\lambda x) f, K \exp(+i\lambda x) g) = (f, Kg) + \lambda\rho(f,g) + (\lambda^2/2)(f,g) \quad \text{(B4)}$$

so that

$$\pm\lambda\rho \le h + \lambda^2/2 \quad \text{(B5)}$$

(which is just the Glimm—Jaffe[8] π-bound). Equation (B5) implies that ρ is a small form perturbation of $\bar{h}$ and thus can be extended to a form $\bar{\rho}$ with form domain $\bar{Q} = \mathcal{D}(\sqrt{K})$.

From here on our proof follows Combes and Thomas.[2] We define the form (for all complex z)

$$h(z) \equiv \bar{h} + z\bar{\rho} + z^2/2 \quad \text{(B6)}$$

with form domain $\bar{Q}$. Then there is an m-sectorial operator $K(z)$ such that

$$(f, K(z) g) = h(z)(f, g)$$

and the form domain of $K(z) = \mathcal{D}(\sqrt{K}) = \bar{Q}$ (see Kato[32]).

A calculation similar to that leading to (B4) yields

$$\exp(-i\lambda x) K(z) \exp(i\lambda x) = K(z+\lambda), \quad \lambda \in R.$$

The family of operators $K(z)$ is a holomorphic family of type B [32] so that the eigenprojection

$$P(z) = (2\pi i)^{-1} \int_{|s|<\delta} (s - K(z))^{-1}\, ds$$

is analytic in z for sufficiently small $|z|$. Let $f(x) = \exp(-x^2)$. Then for λ real

$$\exp(-i\lambda x) P(0) f = P(\lambda) \exp(-i\lambda x) f. \quad \text{(B7)}$$

However the right-hand side of (B7) has an analytic continuation [$L^2(R, d\mu)$ valued] to a disk $|\lambda| < r$ and thus, for sufficiently small $|t|$, $\exp(tx) \in L^2(R, d\mu)$.

As we mentioned after Theorem 2.2, the above argument is only necessary to prove the finiteness of $(\Omega_0, \phi(f_0)\Omega_0)$. We illustrate the technique by showing that if $t \in R$, $t^2/2m < 1$, and $\exp(tx/2) \in L^2(R, d\mu)$, then $\exp(tx) \in L^2$. (Repeated application of this result will then show that if $t^2/2m < 1$, then $\exp[t\phi(f_0)\Omega_0 \in \mathcal{H}$ and hence prove the theorem.) Thus let

$$G_\epsilon(x) = \exp(tx) \exp(-\epsilon x^2/2).$$

A computation similar to the one leading to Eq. (2.9) gives

$$\|G_\epsilon\|_2^2 = (1-(t^2/2m)\,\alpha)^{-1}\|G_\epsilon\|_1^2 \tag{B3}$$

with

$$\alpha = \left(\int (1-\epsilon x/t)^2 G_\epsilon(x)^2\, d\mu\right)/\|G_\epsilon\|_2^2.$$

If we can show that $(1-(t^2/2m)\,\alpha)^{-1}$ is bounded in ϵ as $\epsilon \downarrow 0$, the monotone convergence theorem will then show that $\exp(tx) \in L^2$.

Note that

$$\alpha = 1 + (\epsilon/t^2)\ \left\{\int d\mu(x)\, g(x)\exp[-g(x)]\right\}/\|G_\epsilon\|_2^2$$

with $g(x) = tx(\epsilon x/t - 2)$. Now $g\exp(-g) \le \exp(-1)$ and by Jensen's inequality

$$\|G\|_2^2 \ge \exp\left[\int d\mu\ \ (2tx - \epsilon x^2)\right] \ge C \quad \text{for } \epsilon \in [0,1].$$

and thus $\alpha \le 1 + \beta\epsilon/t^2$. This completes the proof.

We mention that if the set of vectors, span of $\{\exp[i\phi(f)]\Omega_0 : f \in \mathcal{L}\}$, is assumed to be a form core for H, then the Combes–Thomas technique gives results for all eigenvectors of H which are finitely degenerate with isolated eigenvalues.

APPENDIX C: TRANSLATION INVARIANT LINEAR FUNCTIONALS ON $\mathcal{S}(R^N)$

We say that a linear functional $l : \mathcal{S}(R^N) \to \mathbb{C}$ is translation invariant if

$$l(f_a) = l(f) \quad \forall a \in R^N, \ \ \forall f \in \mathcal{S}(R^N)$$

where $f_a(x) = f(x-a)$.

The main result of this appendix is the following theorem.

Theorem C: Suppose l is a translation invariant linear functional on $\mathcal{S}(R^N)$. Then

$$l(f) = C\int f(x)\,dx, \quad \forall f \in \mathcal{S}(R^N). \tag{C1}$$

We remind the reader that the above result is trivial if l is assumed to be continuous.

We begin the proof by showing that the theorem is equivalent to the equality of two subspaces of $\mathcal{S}(R^N)$.

Let

$$\mathcal{V} = \text{linear span } \{([\exp(ip\cdot a)]-1)f : f \in \mathcal{S}(R^N),\ a \in R^N\}$$

$$\mathcal{V}_0 = \{f \in \mathcal{S} : f(0) = 0\},$$

and define $\tilde{l}(f) \equiv l(\tilde{f})$, where $\tilde{f}$ is the Fourier transform of f. Then clearly l is translation invariant if and only if $\tilde{l}$ vanishes on $\mathcal{V}$. Note also that l is of the form (C1) if and only if $\tilde{l}(f) = Cf(0)$. Thus since we can write $f(p) = f(0)\exp(-p^2) + (f(p) - f(0)\exp(-p^2))$, l is of the form (C1) if and only if $\tilde{l}$ vanishes on $\mathcal{V}_0$. Hence the theorem will be proved if we show $\mathcal{V} = \mathcal{V}_0$. (This equality is false for the analogous subspaces if in Theorem C, $\mathcal{S}$ is replaced by L^1. An application of Zorn's lemma then shows that the analogous theorem is also false.)

To show that $\mathcal{V} = \mathcal{V}_0$, we first prove two lemmas.

Lemma C1: Suppose $f \in \mathcal{S}(R^N)$ with $f(0) = 0$. Then

$$f(p) = \sum_{j=1}^{N} p_j f_j(p) \quad \text{with } f_j \in \mathcal{S}(R^N).$$

Proof: Write

$$f(p) = \sum_{j=1}^{N} g_j(p_1, \ldots, p_j)\exp[-(p_{j+1}^2 + \cdots + p_N^2)]$$

with

$$g_j(p_1,\ldots,p_j) = f(p_1,\ldots,p_j,0,\ldots,0) - f(p_1,\ldots,p_{j-1},0,0,\ldots,0)\exp(-p_j^2).$$

We need only show that $p_j^{-1}g_j \in \mathcal{S}(R^j)$. If we call $(p_1,\ldots,p_{j-1}) = q$ and $p_j = p$ then $g_j(p_1,\ldots,p_j) = g(q,p)$. It is clearly enough to show that for $|p| \le 1$ and all n, m

$$|\partial_p^n p^{-1} g(q,p)| \le C_{nm}(1+q^2)^{-m}.$$

However, if $\tilde{g}$ is the Fourier transform of g we have

$$(-iq)^{(m)}\,\partial_p^n p^{-1} g(q,p) = \int\ [\partial_x^{(m)}\tilde{g}(x,y)]\exp(ix\cdot q)\,\partial_p^n\left(\frac{\exp(iyp)-1}{p}\right)dx\,dy$$

and thus the result follows from $|\partial_p^n(\exp(iyp)-1)/p)| \le C_n|y|^{n+1}$.

Now denote by $\mathcal{O}_M$ the set of multipliers for $\mathcal{S}(R^1)$, i.e., $\mathcal{O}_M$ is the set of all C^∞ functions, f, such that for each n there exists an m with

$$|(D^n f)(p)| \le c_{nm}(1+p^2)^{m}.$$

Our next task is to construct certain functions in $\mathcal{O}_M$.

Suppose $\lambda \in (0,1]$. Let $\varkappa_0 \in C_0^\infty(R^1)$ with $\varkappa_0(p) = 1$ for $|p| \le \lambda$ and $\varkappa_0(p) = 0$ for $|p| \ge 2\lambda$. Define

$$\varkappa_n(p) = \varkappa_0(|n|(p - 2\pi n)) \quad n = \pm 1, \pm 2, \ldots$$

$$\varkappa = \sum_{|n|=1}^{\infty} \varkappa_n$$

Lemma C2: $p(1-\varkappa)(1-\exp(ip))^{-1}$ and $\varkappa(1-\exp(ip\sqrt{2}))^{-1}$ are in $\mathcal{O}_M$ if λ is small enough.

Proof: We consider the more interesting function $\varkappa(1-\exp(ip\sqrt{2}))^{-1}$. The proof that the first function is in $\mathcal{O}_M$ is easier. First note a special case of a theorem of Liouville.[33]

$$\inf_m |n\sqrt{2} - m| \ge c/n, \quad n \ge 1.$$

(For a proof, consider the case $|n\sqrt{2} - m| \le 1$. Then $|n\sqrt{2} - m| = |2n^2 - m^2|\,(n\sqrt{2}+m)^{-1} \ge (n\sqrt{2}+m)^{-1} \ge dn^{-1}$ where the first inequality follows from the fact that $|2n^2 - m^2|$ is a positive integer.)

To prove the lemma, it is clearly enough to show

$$|(D^m\varkappa)(1-\exp(ip\sqrt{2}))^{-l}| \le C_{ml}\,(1+p^2)^{N_{lm}}.$$

Suppose for some $n_0 \ne 0$, $|n_0|$ an integer, $|p - 2\pi n_0| \le 2\lambda|n_0|^{-1}$. Then from Liouville's theorem we have

$$|p\sqrt{2} - 2\pi m| \ge 2\pi|n_0\sqrt{2} - m| - 2\sqrt{2}\,\lambda|n_0|^{-1} \ge \gamma|n_0|^{-1}$$

for small enough λ. Thus, if $|p - 2\pi n_0| \le 2\lambda|n_0|^{-1}$,

$$\inf_m |p\sqrt{2} - 2\pi m| \ge \gamma'(1+|p|)^{-1}$$

and hence $|(1-\exp(ip\sqrt{2}))^{-1}| \le \beta(1+|p|)$.

However, if $|p - 2\pi n_0| > 2\lambda|n_0|^{-1}$ for all $|n_0| \ne 0$, then $D^m\varkappa = 0$

$$|(D^m\varkappa)(1-\exp(ip\sqrt{2}))^{-l}| \le |D^m\varkappa|\,\beta^l(1+|p|)^l.$$

It is easy to see that $|(D^m \varkappa)(p)| \le C_m(1+|p|)^m$ and thus the proof is complete.

We remark that a similar construction will work for any irrational algebraic number.

The proof of the equality $\mathcal{V} = \mathcal{V}_0$ is now easy. If $f \in \mathcal{V}_0$ write $f = \sum p_j f_j$, $f_j \in \mathcal{S}(R^N)$. Now note that

$$p_j f_j = (1-\exp(ip_j))[(1-\varkappa^j)(1-\exp(ip_j))^{-1} p_j f_j]$$
$$+ (1-\exp(ip_j\sqrt{2}))[\varkappa^j(1-\exp(ip_j\sqrt{2}))^{-1} p_j f_j],$$

with $\varkappa^j(p) = \varkappa(p_j)$. The functions in square brackets are in $\mathcal{S}(R^N)$ by Lemma C2. This completes the proof of Theorem C.

Note added in proof: The results of Theorem 3.2 also follow from the methods of A. Beurling and J. Deny, Acta Math. **99**, 203–24 (1958) and of M. Fukushima "On the generation of Markov Processes by Symmetric Forms." *Proc. of the Second Japan–USSR Symposium on Probability Theory* (Springer-Verlag, Berlin, 1973). The method of proof is quite different from ours.

ACKNOWLEDGMENTS

It is a pleasure to thank Jürg Fröhlich, Erhard Seiler, and Barry Simon for helpful discussions and to thank Jürg Fröhlich for reading part of the manuscript.

*Research supported in part by the National Science Foundation under Grant Number MPS74-22844.

[1]H. Araki, J. Math. Phys. 1, 492 (1960).
[2]J. Combes and L. Thomas, Commun. Math. Phys. **34**, 251 (1973).
[3]A. O'Connor, Commun. Math. Phys. **32**, 319 (1973).
[4]E. L. Slaggie and E. H. Wichmann, J. Math. Phys. **3**, 946 (1962).
[5]R. Ahlrichs, J. Math. Phys. **14**, 1860 (1973).
[6]B. Simon, Proc. Am. Math. Soc. **42**, 395 (1974).
[7]B. Simon, Trans. Am. Math. Soc. **208**, 317 (1975).
[8]J. Glimm and A. Jaffe, J. Math. Phys. **11**, 3335 (1970).
[9]L. Gross, "Logarithmic Sobolev Inequalities," Am. J. Math. (to be published).
[10]J. Fröhlich, Ann. Inst. H. Poincaré **21**, 271 (1974).
[11]B. Simon, Helv. Phys. Acta. **46**, 686 (1973).
[12]J. Glimm, A. Jaffe, and T. Spencer, "The Particle Structure of the Weakly Coupled $P(\phi)_2$ Model and Other Applications of High Temperature Expansion, II: The Cluster Expansion," in *Constructive Quantum Field Theory*, edited by G. Velo and A. S. Wightman (Springer-Verlag, Berlin, 1973), pp. 199–249.
[13]E. Seiler, Commun. Math. Phys. **42**, 163 (1975).
[14]A. S. Wightman and S. S. Schweber, Phys. Rev. **98**, 812 (1955).
[15]J. Fröhlich, Helv. Phys. Acta. **47**, 265 (1974).
[16]J. Glimm and A. Jaffe, J. Math. Phys. **13**, 1568 (1972).
[17]F. Guerra, L. Rosen, and B. Simon, Commun. Math. Phys. **27**, 10 (1972).
[18]B. Simon, "The Glimm–Jaffe ϕ-Bound: A Markov Proof," in *Constructive Quantum Field Theory*, edited by G. Velo and A. S. Wightman (Springer-Verlag, Berlin, 1973), pp. 125–131.
[19]J. Fröhlich, "Schwinger Functions and Their Generating Functionals, II: Markovian and Generalized Path Space Measures on $\mathcal{S}'$," Adv. Math., to appear.
[20]O. McBryan, "Convergence of the Vacumm Energy Density, ϕ-bounds and Existence of Wightman Functions for the Yukawa_2 Model," Rockefeller University preprint.
[21]E. Seiler and B. Simon, "Bounds in the Yukawa_2 Quantum Field Theory: Upper Bound on the Pressure, Hamiltonian Bound and Linear Lower Bound," Princeton University preprint.
[22]E. Seiler and B. Simon, "Nelsen's Symmetry and All That in the Yukawa_2 and $(\phi^4)_3$ Field Theories," Princeton University preprint.
[23]T. Spencer, J. Math. Phys. **14**, 823 (1973).
[24]A. Klein and L. J. Landau, J. Funct. Anal. **20**, 44 (1975).
[25]S. Alberverio and R. Höegh-Krohn, "Quasi Invariant Measures, Symmetric Diffusion Processes and Quantum Fields," University of Oslo preprint.
[26]K. Yosida, *Functional Analysis* (Springer, New York, 1971), 3rd ed., p. 216–18.
[27]R. A. Minlos, Trudy Mosk. Mat. Obs. **8**, 497 (1959).
[28]E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[29]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).
[30]E. Stein, *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory*, Ann. of Math. Study 63, (Princeton U. P., Princeton, N.J., 1970).
[31]V. S. Vladimirov, *Methods of the Theory of Many Complex Variables* (MIT Press, Cambridge, Mass., 1966), p. 235.
[32]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
[33]G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford, Clarendon, 1960), 4th ed., p. 161.

The energy levels of the one-dimensional potential well $V(X) = a|X|$ calculated by means of certain phase-integral approximations

Bo Thidé

Institute of Theoretical Physics, University of Uppsala, Uppsala, Sweden
(Received 19 January 1976)

The eigenvalue problem of the one-dimensional potential well $V(X) = a|X|$ is solved by means of certain higher-order phase-integral approximations. The purpose of this paper is to demonstrate numerically the applicability and accuracy of these approximations (which are related to, but not identical to, the higher-order JWKB approximations) and, therefore, a comparison is made with exact results. An upper bound for the absolute error in the first-order approximation is calculated analytically and found to be in accordance with the actual numerical results which are displayed in a table.

1. FORMULATION OF THE PROBLEM

In order to determine the accuracy of certain higher order phase-integral approximations introduced by N. Fröman,[1,2] these approximations are used here for calculating the energy levels of a particle with mass m and energy E moving in the one-dimensional potential $V(X) = a|X|$, where a is a real, positive parameter. We thus consider the Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dX^2} + a|X|\psi = E\psi. \qquad (1)$$

By introducing the new independent variable

$$x = \left(\frac{2m}{\hbar^2}a\right)^{1/3} X \qquad (2)$$

and the real parameter

$$\alpha = \left(\frac{2m}{\hbar^2 a^2}\right)^{1/3} E, \qquad (3)$$

we can transform (1) into

$$\frac{d^2\psi}{dx^2} + (\alpha - |x|)\psi = 0. \qquad (4)$$

The eigenfunctions for bound states fulfill the boundary conditions

$$\psi \to 0, \quad |x| \to \infty. \qquad (5)$$

Furthermore, since the potential is symmetric the eigenfunctions have either odd parity, i.e., satisfy the condition

$$\psi = 0, \quad x = 0, \qquad (6a)$$

or even parity, i.e., satisfy the condition

$$\frac{d\psi}{dx} = 0, \quad x = 0. \qquad (6b)$$

The boundary conditions (5) can therefore be replaced by condition (6a) or (6b), together with the boundary condition

$$\psi \to 0, \quad x \to +\infty. \qquad (7)$$

Consequently, we can restrict ourselves to a consideration of the region where $x \geq 0$. Equation (4) can then be written

$$\frac{d^2\psi}{dx^2} + Q^2(x)\psi = 0, \qquad (8)$$

where

$$Q^2(x) = \alpha - x. \qquad (9)$$

The physical importance of Eq. (1) has been pointed out by Bell[3] and Abrikosov.[4] One encounters, e.g., the potential $V(X) = aX$ for $X \geq 0$, $V(X) = \infty$ for $X < 0$, in problems regarding magnetic surface states, in which case condition (6a) is valid.

2. THE EXACT EIGENVALUES α

The exact solution of (8) tending to zero at $+\infty$ is given by the Airy function[5]

$$\psi = \mathrm{Ai}(x - \alpha) \qquad (10)$$

apart from an arbitrary constant factor. According to (6) the exact quantization conditions are therefore

$$\mathrm{Ai}(-\alpha) = 0 \quad \text{for the odd states,} \qquad (11a)$$

and

$$\mathrm{Ai}'(-\alpha) = 0 \quad \text{for the even states.} \qquad (11b)$$

These conditions are given by Bell[3] in another form.

Sherry[6] has calculated the first hundred exact eigen-

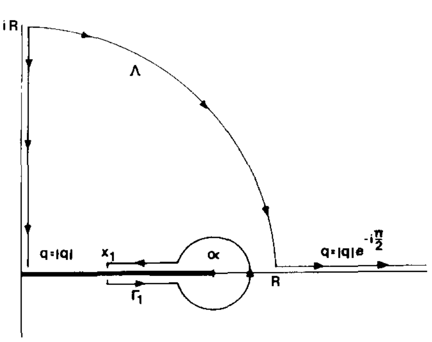


FIG. 1. Contours Γ_1 and Λ of integration for obtaining the integrals $w(x_1)$ and μ, respectively. On Λ the arrows indicate the directions in which $|\exp\{iw(z)\}|$ increases. The complex plane is cut along the real axis from $x = \alpha$ to $x = 0$ (heavy line). The choice of phase of q above the real axis on the first Riemann sheet is also given.

values α satisfying (11) by utilizing convergent series for the Airy function in the computation of the first twenty of these eigenvalues, but asymptotic formulas when computing the remaining ones. Some of these exact eigenvalues are included in Table I.

3. THE APPROXIMATE EIGENVALUES α CALCULATED BY MEANS OF THE PHASE-INTEGRAL APPROXIMATIONS

Approximate phase-integral solutions of (8) for a general form of Q^2 are given in Refs. 1 and 2, to which we refer the reader for the general background. The parameter λ, introduced there merely as a formal mathematical tool, will in the present treatment be set equal to unity.

Henceforth we shall treat x as being the real part of a complex variable z. We introduce a cut from the point α on the positive real z axis to the origin (cf. Fig. 1) and define $w(x)$ according to Eqs. (17a), (17b) in Ref. 2 for the case corresponding to Fig. 2(a) in Ref. 2, where x' is chosen to be equal to α. Then, since $q^2(z)$ is real on the real axis, the approximate wavefunction vanishing at infinity in the classically forbidden region is given by

$$\psi = |q(x_2)|^{-1/2} \exp\{-|w(x_2)|\}, \quad x_2 > \alpha, \tag{12}$$

except for a normalization factor. From the connection formula (21) in Ref. 2 we realize that in the classically allowed region the corresponding solution ψ to (8) is approximately given by

$$\psi = 2|q(x_1)|^{-1/2} \cos[|w(x_1)| - \tfrac{1}{4}\pi], \quad 0 \le x_1 < \alpha. \tag{13}$$

The condition (6a), valid for the odd eigenfunctions, yields, when applied to (13), the approximate quantization condition

$$|w(0)| = (s + \tfrac{1}{2})\pi/2, \quad s = 1, 3, 5, \cdots. \tag{14a}$$

For the choice of phase of $q(z)$ indicated in Fig. 1, the condition (6b), valid for the even eigenfunctions, when used in (13), yields the approximate quantization condition

$$|w(0)| + \arctan\left[\frac{1}{2}\left(\frac{d}{dz}\frac{1}{q(z)}\right)_{z=0}\right] = (s + \tfrac{1}{2})\frac{\pi}{2}, \quad s = 0, 2, 4, \cdots. \tag{14b}$$

Using the explicit expressions for the functions Y_{2n} given by Campbell[7] and recalling (9), we obtain

$$Y_{2n} = \frac{(-1)^{n+1}}{2^{6n-1}} \frac{b_{2n}}{(\alpha - z)^{3n}}, \tag{15}$$

where the first six coefficients b_{2n} are

$$b_0 = -\tfrac{1}{2}, \tag{16a}$$

$$b_2 = 5, \tag{16b}$$

$$b_4 = 1105, \tag{16c}$$

$$b_6 = 828\,250, \tag{16d}$$

$$b_8 = 1\,282\,031\,525, \tag{16e}$$

$$b_{10} = 3\,366\,961\,243\,7\,50. \tag{16f}$$

From Eq. (7c) in Ref. 1 [or Eq. (10) in Ref. 2] and Eqs. (9) and (15) in the present paper, we obtain, for the phase-integral approximation of order $2N+1$, the formula

$$q(z) = (\alpha - z)^{1/2} \sum_{n=0}^{N} \frac{(-1)^{n+1}}{2^{6n-1}} \frac{b_{2n}}{(\alpha - z)^{3n}}. \tag{17}$$

By inserting (17) into formula (17a) in Ref. 2, where the integration contour Γ_1 is that of our Fig. 1 and $x_1 (\ge 0)$ is a point on the upper lip of the cut along the real axis in the classically allowed region, we find that

$$w(x_1) = \tfrac{1}{2}\int_{\Gamma_1} q(z)\,dz$$

$$= \tfrac{2}{3}(\alpha - x_1)^{3/2} \sum_{n=0}^{N} \frac{(-1)^{n+1}}{2^{6n-1}(2n-1)} \frac{b_{2n}}{(\alpha - x_1)^{3n}}. \tag{18}$$

After introducing (17) and (18) with $x_1 = 0$ into the quantization conditions (14), taking (16a)—(16f) into account, we have by numerical means computed the approximate eigenvalues α when $2N+1 = 1, 3, 5, 7, 9, 11$ for several quantum numbers s. These eigenvalues are given in the right-hand column of Table I. As exceptional cases the eigenvalue of the ground state, corresponding to $s = 0$, cannot be found in the first-, fifth-, and ninth-order phase-integral approximation. A graphical investigation of (14b) will clearly demonstrate this.

4. ERROR ESTIMATES

According to Sec. 4 of Ref. 2 an essential condition for the validity of the connection formula yielding (13) in the present paper is that there is only one extremum of $|\exp\{iw(z)\}|$ on the path of integration Λ for the quantity μ, which is defined by Eq. (18) in Ref. 2 [cf. also Eq. (10) in Ref. 1] and which determines an upper bound for the error involved in the connection. If μ is much smaller than unity and we include error terms, Eqs. (14a) and (14b) read, in the first-order approximation,

$$\tfrac{2}{3}\alpha^{3/2} + O(\mu) = (s + \tfrac{1}{2})\pi/2, \quad s = 1, 3, 5, \cdots \tag{19a}$$

and

$$\tfrac{2}{3}\alpha^{3/2} + \arctan(\tfrac{1}{4}\alpha^{-3/2}) + O(\mu) = (s + \tfrac{1}{2})\pi/2, \quad s = 0, 2, 4, \cdots, \tag{19b}$$

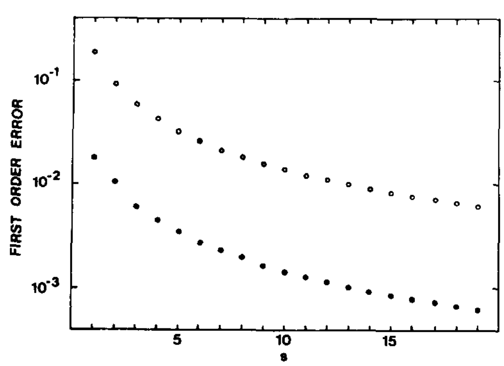


FIG. 2. Plot of the magnitude of the actual absolute error of the eigenvalues α (filled circles) as well as the estimate (22) of this error (open circles) vs the quantum number s for the first-order phase-integral approximation. In the plot of the actual error the "unsmooth" behavior detectable for the very lowest quantum numbers is due to the fact that the eigenvalues for odd and for even s are computed from two different expressions, namely (14a) and (14b), respectively.

TABLE I. Approximate and exact eigenvalues α, related to the energy E through Eq. (3), for various quantum numbers s. For each value of s this table gives (from top to bottom) the eigenvalue obtained in first-, third-, fifth-, seventh-, ninth-, and eleventh-order phase-integral approximations, and finally, the exact eigenvalue found in Ref. 6. Dashes indicate that the eigenvalues cannot be found.

s	Approximate and exact α	s	Approximate and exact α
0	---	6	6.160 5
	1.13		6.163 321
	---		6.163 306 92
	1.43		6.163 307 389
	---		6.163 307 350 8
	1.79		6.163 307 356 7
	1.019		6.163 307 355 64
1	2.320	7	6.784 5
	2.339 3		6.786 716 2
	2.337 63		6.786 707 91
	2.338 56		6.786 708 099 5
	2.337 30		6.786 708 089 16
	2.340 4		6.786 708 090 21
	2.338 107		6.786 708 090 072
2	3.238	8	7.370 3
	3.248 58		7.372 182 9
	3.248 109		7.372 177 15
	3.248 247		7.372 177 259 5
	3.248 148		7.372 177 254 68
	3.248 276		7.372 177 255 097
	3.248 1976		7.372 177 255 0478
3	4.081 8	9	7.942 5
	4.088 05		7.944 137 3
	4.087 940 6		7.944 133 536
	4.087 951 4		7.944 133 588 8
	4.087 948 66		7.944 133 587 01
	4.087 949 94		7.944 133 587 131
	4.087 949 444		7.944 133 587 1209
4	4.815 5	19	12.828 14
	4.820 148		12.828 777 10
	4.820 095 9		12.828 776 751 7
	4.820 099 75		12.828 776 752 875 0
	4.820 099 05		12.828 776 752 865 618
	4.820 099 292		12.828 776 752 865 760 5
	4.820 099 2112		12.828 776 752 865 757 20
5	5.517 2	99	38.020 937
	5.520 582		38.021 008 678 8
	5.520 558 94		38.021 008 677 255 06
	5.520 559 913		38.021 008 677 255 254 494
	5.520 559 813		38.021 008 677 255 254 433 097
	5.520 559 832 2		38.021 008 677 255 254 433 132 50
	5.520 559 828 10		38.021 008 677 255 254 433 132 47

respectively. Here $O(\mu)$ denotes a quantity which is at most of the order of magnitude μ.

Using Eqs. (19a) and (19b) and some simple algebra one finds that the magnitude of the absolute error of the eigenvalue α, when calculated in the first-order phase-integral approximation, is at most of the order of magnitude $\alpha^{-1/2}\mu$. The μ-integral is performed from $z=0+i0$ to $z=+\infty$ along the path Λ shown in Fig. 1, where the arrows on this path indicate the directions in which $|\exp\{iw(z)\}|$ increases. Letting R in Fig. 1 tend to infinity, we obtain for the first-order approximation

$$\mu=\frac{5\sqrt{\pi}}{32}\,\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{5}{4})}\,\alpha^{-3/2}\approx 0.4\alpha^{-3/2}. \tag{20}$$

Since the quantities μ and $\alpha^{-3/2}/4$ are both small, Eqs. (19a) and (19b) obviously yield

$$\alpha\approx\left[(s+\tfrac{1}{2})\frac{3\pi}{4}\right]^{2/3},\quad s=0,1,2,\cdots. \tag{21}$$

If we use the previously mentioned error expression $\alpha^{-1/2}\mu$, Eq. (20) with unity replacing the numerical constant (≈ 0.4) multiplying $\alpha^{-3/2}$, and Eq. (21), we find that the magnitude of the absolute error of the eigenvalue α, when calculated in the first-order phase-integral approximation, is at most of the order of magnitude

$$\left[(s+\tfrac{1}{2})\frac{3\pi}{4}\right]^{-4/3},\quad s=0,1,2,\cdots. \tag{22}$$

Over a wide range of quantum numbers s this estimate (22) of the error is approximately a constant factor times the actual error. In Fig. 2 this can be seen to hold for $s=1,2,\ldots,19$. We also see that the first-order error decreases for increasing quantum number, as should one expect. If we go to higher orders of approximation, we generally improve the accuracy of the eigenvalue for a given quantum number until a certain finite, optimum order is reached where the error is at minimum (cf. Table I).

ACKNOWLEDGMENTS

This problem was suggested by Dr. P.O. Fröman, whom I should like to thank for guidance and encouragement during the time this work was carried out. I am also grateful to Dr. N. Fröman, the head of our institute, for the time spent in stimulating discussions.

[1]N. Fröman, Ark. Fys. 32, 541 (1966).
[2]N. Fröman, Ann. Phys. (N.Y.) 61, 451 (1970).
[3]R.P. Bell, Philos. Mag. 35, 582 (1944).
[4]A.A. Abrikosov, Solid State Phys. Suppl. 12, 211 (1972).
[5]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (National Bureau of Standards, Washington, D.C., 1968), p. 446.
[6]M.E. Sherry, AFCRC-TR-59-135, "The zeros and maxima of the Airy function and its first derivative to 25 significant figures" (Air Force Cambridge Research Center, Bedford, Mass., 1959).
[7]J.A. Campbell, J. Comput. Phys. 10, 308 (1972).

Coupled gravitational and electromagnetic perturbations around a charged black hole

Chul Hoon Lee

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 6 June 1975; final revised manuscript received 1 March 1976)

Coupled gravitational and electromagnetic perturbation equations are derived in the electrovacuum space around a charged rotating black hole using the Newman–Penrose formalism. When restricted to the nonrotating case, the equations separate in Schwarzschild coordinates. The asymptotic solutions at infinity and on the event horizon are obtained. In the coupled scattering of the electromagnetic and the gravitational waves on a Reissner–Nordström black hole, it is shown that the net energy flux is radially inward.

I. INTRODUCTION

It appears that a collapsed rotating object could possess a net charge.[1,2] For such an object the Kerr–Newman metric would be the best approximation. Thus the study of the electromagnetic and gravitational perturbation away from the Kerr–Newman metric is physically interesting. The resulting perturbation equations can be used to study both the stability of such an object and its scattering of electromagnetic and gravitational waves. This paper derives these perturbation equations using the Newman–Penrose formalism.[3]

In the study of an uncharged black hole, since the electromagnetic stress-energy tensor is second order in the electromagnetic field, one can treat the electromagnetic perturbation separately keeping the background metric unchanged to first order of the perturbation. However, for a charged black hole the change in the stress-energy tensor is first order in the electromagnetic perturbation and thus a perturbation of the electromagnetic field inevitably accompanies a metric perturbation and vice versa.

In Sec. II, equations for the coupled metric and electromagnetic perturbations are derived. This is done first for the Kerr–Newman metric. When restricted to the nonrotating case, these equations can be separated in Schwarzchild coordinates (a special case of Boyer–Lindquist coordinates.) In Sec. III, the asymptotic solutions at infinity and on the event horizon are obtained. Also, the connections between the coefficients in the asymptotic solutions are obtained. In Sec. IV, using the conservation of a "Wronskian" of the coupled equations, we show explicitly that the net energy flux for the coupled electromagnetic and gravitational waves is radially inward at infinity. This result points to the stability of Reissner–Nordström black holes.

In Boyer–Lindquist coordinates the Kerr–Newman metric has the form[4]

$$ds^2=\left(1-\frac{2Mr-Q^2}{\Sigma}\right)dt^2+2(2Mr-Q^2)\frac{a\sin^2\theta}{\Sigma}\,dt\,d\phi - (\Sigma/\Delta)\,dr^2-\Sigma\,d\theta^2-\sin^2\theta\Big[r^2+a^2 + (2Mr-Q^2)\frac{a^2\sin^2\theta}{\Sigma}\Big]d\phi^2, \qquad (1.1)$$

where $\Sigma\equiv r^2+a^2\cos^2\theta$, $\Delta\equiv r^2-2Mr+a^2+Q^2$, M is the mass, a is the angular momentum per unit mass, and Q is the charge of the black hole.

We follow Teukolsky's notation[5] denoting unperturbed quantities by superscript A and the perturbation by superscript B.

We use the tetrad whose components are

$$\overset{A}{l}{}^{\mu}=\left(\frac{r^2+a^2}{\Delta},\ 1,\ 0,\ \frac{a}{\Delta}\right),$$
$$\overset{A}{n}{}^{\mu}=\left(\frac{r^2+a^2}{2\Sigma},\ -\frac{\Delta}{2\Sigma},\ 0,\ \frac{a}{2\Sigma}\right),$$
$$\overset{A}{m}{}^{\mu}=\frac{1}{\sqrt{2}(r+ia\cos\theta)}\left(ia\sin\theta,0,1,\frac{i}{\sin\theta}\right). \qquad (1.2)$$

The resulting tetrad components of the Weyl tensor are[6]

$$\overset{A}{\psi}_0=\overset{A}{\psi}_1=\overset{A}{\psi}_3=\overset{A}{\psi}_4=0,\quad \overset{A}{\psi}_2=-\frac{M(r+ia\cos\theta)-Q^2}{(r-ia\cos\theta)^3(r+ia\cos\theta)} \qquad (1.3)$$

and of the electromagnetic field tensor are

$$\overset{A}{\phi}_0=\overset{A}{\phi}_2=0,\quad \overset{A}{\phi}_1=\frac{Q}{2(r-ia\cos\theta)^2}. \qquad (1.4)$$

In electro-vacuum space,

$$\phi_{mn}=2\phi_m\phi_n^*,\quad m,n=0,1,2. \qquad (1.5)$$

The spin coefficients are

$$\overset{A}{\kappa}=\overset{A}{\sigma}=\overset{A}{\lambda}=\overset{A}{\nu}=\overset{A}{\epsilon}=0,$$
$$\overset{A}{\rho}=\frac{-1}{r-ia\cos\theta},\quad \overset{A}{\tau}=\frac{-ia\sin\theta}{\sqrt{2}}\,\overset{A}{\rho}\overset{A}{\rho}{}^*,$$
$$\overset{A}{\pi}=\frac{ia\sin\theta}{\sqrt{2}}\,\overset{A}{\rho}{}^2,\quad \overset{A}{\beta}=\frac{-\cot\theta}{2\sqrt{2}}\,\overset{A}{\rho}{}^*,$$
$$\overset{A}{\alpha}=\overset{A}{\pi}-\overset{A}{\beta}{}^*,\quad \overset{A}{\mu}=\overset{A}{\rho}{}^2\overset{A}{\rho}{}^*\Delta/2,$$
$$\overset{A}{\gamma}=\overset{A}{\mu}+\frac{r-M}{2}\,\overset{A}{\rho}\overset{A}{\rho}{}^*. \qquad (1.6)$$

The perturbation equations we derive couple ψ_0^B with $\chi_1^B\equiv 3\psi_2^A\phi_0^B-2\phi_1^A\psi_1^B$ and ψ_4^B with $\chi_{-1}^B\equiv 3\psi_2^A\phi_2^B-2\phi_1^A\psi_3^B$. In the asymptotic flat region ψ_0^B is related to the ingoing gravitational wave and ψ_4^B to the outgoing gravitational wave. Similarly, ϕ_0^B is related to the ingoing electromagnetic wave and ϕ_2^B to the outgoing electromagnetic wave. The expressions for energy flux in terms of ϕ^B's and ψ^B's are listed in Appendix A.

II. COUPLED GRAVITATIONAL AND ELECTROMAGNETIC PERTURBATION EQUATIONS

To derive the equations for the perturbed quantities, we start with the sourceless Maxwell's equations,[3]

$$(D-2\rho)\phi_1-(\delta^*+\pi-2\alpha)\phi_0+\kappa\phi_2=0, \tag{2.1}$$

$$(\delta-2\tau)\phi_1-(\Delta+\mu-2\gamma)\phi_0+\sigma\phi_2=0, \tag{2.2}$$

and the following Newman–Penrose equation and Bianchi identities[7]:

$$(\delta-3\beta-\alpha^*+\pi^*-\tau)\kappa-(D-3\epsilon+\epsilon^*-\rho-\rho^*)\sigma+\psi_0=0, \tag{2.3}$$

$$(\delta^*-4\alpha+\pi)\psi_0-(D-4\rho-2\epsilon)\psi_1+(D-2\rho^*-2\epsilon)\phi_{01}$$
$$-(\delta+\pi^*-2\alpha^*-2\beta)\phi_{00}+\kappa^*\phi_{02}-2\sigma\phi_{10}=(3\psi_2-2\phi_{11})\kappa, \tag{2.4}$$

$$(\Delta-4\gamma+\mu)\psi_0-(\delta-4\tau-2\beta)\psi_1-(\delta+2\pi^*-2\beta)\phi_{01}$$
$$+(D-2\epsilon+2\epsilon^*-\rho^*)\phi_{02}+2\kappa\phi_{12}+\lambda^*\phi_{00}=(3\psi_2+2\phi_{11})\sigma, \tag{2.5}$$

where $D=l^\mu\partial/\partial x^\mu$, $\Delta=n^\mu\partial/\partial x^\mu$, $\delta=m^\mu\partial/\partial x^\mu$, and $\delta^*=m^{*\mu}\partial/\partial x^\mu$.

Operating by $(\delta-2\tau-\alpha^*-\beta+\pi^*)$ on Eq. (2.1) and by $(D-\epsilon+\epsilon^*-2\rho-\rho^*)$ on Eq. (2.2), subtracting and making use of relevant Newman–Penrose equations we obtain

$$[(\delta-2\tau-\alpha^*-\beta+\pi^*)(\delta^*+\pi-2\alpha)$$
$$-(D-\epsilon+\epsilon^*-2\rho-\rho^*)(\Delta+\mu-2\gamma)]\phi_0$$
$$=[2(\Delta-3\gamma-\gamma^*-\mu+\mu^*)\kappa-2$$
$$\times(\delta^*-3\alpha+\beta^*-\tau^*-\pi)\sigma+4\psi_1-\kappa\Delta+\sigma\delta^*]\phi_1$$
$$+[(\delta-2\tau-\alpha^*-\beta+\pi)\kappa-(D-\epsilon+\epsilon^*-2\rho-\rho^*)\sigma]\phi_2. \tag{2.6}$$

From Eq. (2.6) we obtain a first order perturbation equation

$$[(\delta-2\tau-\alpha^*-\beta+\pi^*)^A(\delta^*+\pi-2\alpha)^A$$
$$-(D-\epsilon+\epsilon^*-2\rho-\rho^*)^A(\Delta+\mu-2\gamma)^A]\phi_0^B$$
$$=2\phi_1^A[(\Delta-3\gamma-\gamma^*-2\mu+\mu^*)^A\kappa^B$$
$$-(\delta^*-3\alpha+\beta^*-\tau^*-2\pi)^A\sigma^B+2\psi_1^B]. \tag{2.7}$$

Here we have used the unperturbed Maxwell's equations

$(\delta^*+2\pi)^A\phi_1^A=0$ and $(\Delta+2\mu)^A\phi_1^A=0$.

Henceforth we will drop the superscript A from unperturbed quantities.

The perturbation equations resulting from Eqs. (2.3), (2.4), and (2.5) are

$$(\delta-3\beta-\alpha^*+\pi^*-\tau)\kappa^B-(D-\rho-\rho^*)\sigma^B+\psi_0^B=0, \tag{2.8}$$

$$(\delta^*-4\alpha+\pi)\psi_0^B-(D-4\rho)\psi_1^B+2\phi_1^*D\phi_0^B=(3\psi_2-2\phi_{11})\kappa^B, \tag{2.9}$$

$$(\Delta-4\gamma+\mu)\psi_0^B-(\delta-4\tau-2\beta)\psi_1^B-2\phi_1^*(\delta-2\beta)\phi_0^B$$
$$=(3\psi_2+2\phi_{11})\sigma^B. \tag{2.10}$$

Operating by $(\delta-3\beta-\alpha^*-\pi^*-4\tau)$ on Eq. (2.9) and by $(D-4\rho-\rho^*)$ on Eq. (2.10) and subtracting with the use of the Newman–Penrose equations for the unperturbed metric one obtains

$$[(\delta-3\beta-\alpha^*+\pi^*-4\tau)(\delta^*-4\alpha+\pi)$$
$$-(D-4\rho-\rho^*)(\Delta-4\gamma+\mu)+3\psi_2]\psi_0^B$$
$$+2\phi_1^*[(\delta-3\beta-\alpha^*-\pi^*-4\tau)D$$
$$+(D-4\rho+\rho^*)(\delta-2\beta)]\phi_0^B$$
$$=2\phi_{11}[(\delta-3\beta-\alpha^*-\pi^*+\tau)\kappa^B+(D+\rho+\rho^*)\sigma^B]. \tag{2.11}$$

Similarly if one acts with $(\Delta-3\gamma-\gamma^*+\mu+\mu^*)$ on Eq. (2.9) and with $(\delta^*-3\alpha+\beta^*-\tau^*+\pi)$ on Eq. (2.10) and subtracts one obtains

$$[-(\Delta-3\gamma-\gamma^*+\mu+\mu^*)(D-4\rho)$$
$$+(\delta^*-3\alpha+\beta^*-\tau^*+\pi)(\delta-4\tau-2\beta)]\psi_1^B$$
$$+2\phi_1^*[(\Delta-3\gamma-\gamma^*+\mu-\mu^*)D$$
$$+(\delta^*-3\alpha+\beta^*+\tau^*+\pi)(\delta-2\beta)]\phi_0^B$$
$$=3\psi_2[(\Delta-3\gamma-\gamma^*-2\mu+\mu^*)\kappa^B$$
$$-(\delta^*-3\alpha+\beta^*-\tau^*-2\pi)\sigma^B]$$
$$-2\phi_{11}[(\Delta-3\gamma-\gamma^*+2\mu-\mu^*)\kappa^B$$
$$+(\delta^*-3\alpha+\beta^*+\tau^*+2\pi)\sigma^B]. \tag{2.12}$$

Now using Eqs. (2.8), (2.9), and (2.10) and thus eliminating κ^B and σ^B in Eq. (2.11), a perturbation equation which couples ψ_0^B and $\chi_1^B\equiv 3\psi_2\phi_0^B-2\phi_1\psi_1^B$ results in

$$\Bigg[\left(D-4\rho-\rho^*-\frac{4\phi_{11}(\rho+\rho^*)}{3\psi_2+2\phi_{11}}\right)(\Delta-4\gamma+\mu)-(3\psi_2+2\phi_{11})$$
$$-\frac{3\psi_2+2\phi_{11}}{3\psi_2-2\phi_{11}}\left(\delta-3\beta-\alpha^*-4\tau+\pi^*-\frac{4\phi_{11}(\pi^*-\tau)}{3\psi_2-2\phi_{11}}\right)$$
$$\times(\delta^*-4\alpha+\pi)\Bigg]\psi_0^B$$
$$=\frac{4\phi_1^*}{3\psi_2-2\phi_{11}}\Bigg[\left(D-5\rho-\frac{4\phi_{11}(\rho+\rho^*)}{3\psi_2+2\phi_{11}}\right)$$
$$\times\left(\delta-2\beta-5\tau-\pi^*-\frac{4\phi_{11}(\pi^*-\tau)}{3\psi_2-2\phi_{11}}\right)$$
$$-\frac{3\psi_2+2\phi_{11}}{3\psi_2-2\phi_{11}}\left(2\rho\tau-2\rho^*\pi^*+\frac{4\phi_{11}(\pi^*-\tau)(2\rho-\rho^*)}{3\psi_2-2\phi_{11}}\right)\Bigg]\chi_1^B. \tag{2.13}$$

Similarly using Eqs. (2.7), (2.9), and (2.10) in Eq. (2.12) we obtain

$$\Bigg[\left(\Delta-3\gamma-\gamma^*+3\mu+\mu^*+\frac{4\phi_{11}(2\mu-\mu^*)}{3\psi_2-2\phi_{11}}\right)$$
$$\times(D-6\rho)-2(3\psi_2-2\phi_{11})$$
$$-\frac{3\psi_2-2\phi_{11}}{3\psi_2+2\phi_{11}}\left(\delta^*+4\beta^*-\tau^*+\frac{4\phi_{11}(2\pi+\tau^*)}{3\psi_2+2\phi_{11}}\right)$$
$$\times(\delta-2\beta-6\tau)\Bigg]\chi_1^B$$
$$=\frac{-8\phi_1\circ\phi_{11}}{3\psi_2+2\phi_{11}}\Bigg[\left(\Delta-3\gamma-\gamma^*+3\mu+\frac{4\phi_{11}(2\mu-\mu^*)}{3\psi_2-2\phi_{11}}\right)$$
$$\times\left(\delta^*+4\beta^*+\tau^*-\pi-\frac{4\phi_{11}(2\pi+\tau^*)}{3\psi_2+2\phi_{11}}\right)$$

$$-\frac{3\psi_2-2\phi_{11}}{3\psi_2+2\phi_{11}}\bigg\{(\Delta\tau^*)+2\pi\mu+\tau^*(\gamma-\gamma^*)+\frac{4\phi_{11}(2\pi+\tau^*)(\mu+\mu^*)}{3\psi_2+2\phi_{11}}\bigg\}\bigg]\psi_0^B. \tag{2.14}$$

The equations corresponding to Eqs. (2.13) and (2.14) for ψ_4^B and $\chi_{-1}^B\equiv 3\psi_2\phi_2^B-2\phi_1\psi_3^B$ are

$$\bigg[\bigg(\Delta+3\gamma-\gamma^*+4\mu+\mu^*+\frac{4\phi_{11}(\mu+\mu^*)}{3\psi_2+2\phi_{11}}\bigg)(D-\rho)-(3\psi_2+2\phi_{11})$$
$$-\frac{3\psi_2+2\phi_{11}}{3\psi_2-2\phi_{11}}\bigg(\delta^*+3\alpha+\beta^*-\tau^*+4\pi+\frac{4\phi_{11}(\tau^*-\pi)}{3\psi_2-2\phi_{11}}\bigg)$$
$$\times(\delta+4\beta-\tau)\bigg]\psi_4^B$$
$$=\frac{4\phi_1^*}{3\psi_2-2\phi_{11}}\bigg[\bigg(\Delta+3\gamma-\gamma^*+5\mu+\frac{4\phi_{11}(\mu+\mu^*)}{3\psi_2+2\phi_{11}}\bigg)$$
$$\times\bigg(\delta^*+2\alpha+5\pi+\tau^*+\frac{4\phi_{11}(\tau^*-\pi)}{3\psi_2-2\phi_{11}}\bigg)$$
$$-\frac{3\psi_2+2\phi_{11}}{3\psi_2-2\phi_{11}}\bigg\{(\Delta\tau^*)+\tau^*(\gamma-\gamma^*)+2\pi\mu$$
$$+\frac{4\phi_{11}(\tau^*-\pi)(2\mu-\mu^*)}{3\psi_2-2\phi_{11}}\bigg\}\bigg]\chi_{-1}^B, \tag{2.15}$$

$$\bigg[\bigg(D-3\rho-\rho^*-\frac{4\phi_{11}(2\rho-\rho^*)}{3\psi_2-2\phi_{11}}\bigg)(\Delta+6\mu+2\gamma)-2(3\psi_2-2\phi_{11})$$
$$-\frac{3\psi_2-2\phi_{11}}{3\psi_2+2\phi_{11}}\bigg(\delta-\alpha^*+3\beta-3\tau+\pi^*-\frac{4\phi_{11}(\pi^*+2\tau)}{3\psi_2-2\phi_{11}}\bigg)$$
$$\times(\delta^*+2\alpha+6\pi)\bigg]\chi_{-1}^B$$
$$=\frac{-8\phi_1{}^\circ\phi_{11}}{3\psi_2+2\phi_{11}}\bigg[\bigg(D-3\rho-\frac{4\phi_{11}(2\rho-\rho^*)}{3\psi_2-2\phi_{11}}\bigg)$$
$$\times\bigg(\delta-\alpha^*+3\beta-3\tau+\frac{4\phi_1(2\tau+\pi^*)}{3\psi_2+2\phi_{11}}\bigg)$$
$$+\frac{3\psi_2-2\phi_{11}}{3\psi_2+2\phi_{11}}\bigg\{(D\pi^*)-2\tau\rho-\frac{4\phi_{11}(2\tau+\pi^*)(\rho+\rho^*)}{3\psi_2+2\phi_{11}}\bigg\}\bigg]\psi_4^B. \tag{2.16}$$

The equations are invariant under gauge transformations and infinitesimal tetrad rotations. The transformations of the N-P quantities under the most general infinitesimal tetrad rotation are[5]

$$\psi_0^B\to\psi_0^B,\quad \psi_1^B\to\psi_1^B+3a\psi_2,\quad \psi_3^B\to\psi_3^B+3b\psi_2,$$
$$\psi_4^B\to\psi_4^B,\quad \phi_0^B\to\phi_0^B+2a\phi_1,\quad \phi_2^B\to\phi_2^B+2b\phi_1, \tag{2.17}$$

where a and b are complex small parameters. Thus, though ϕ_0^B, ψ_1^B, ϕ_2^B, and ψ_3^B are not invariant individually, the combinations $\chi_1^B\equiv 3\psi_2\phi_0^B-2\phi_1\psi_1^B$ and $\chi_1^B\equiv 3\psi_2\phi_2^B-2\phi_1\psi_3^B$ are.

It does not appear that Eqs. (2.13), (2.14), (2.15), and (2.16) separate in Boyer–Lindquist coordinates. However, for the case where $a=0$, the equations simplify and can be separated by writing

$$\psi_0^B=\exp(-i\omega t)\exp(im\phi)\,{}_2Y_l^m(\theta)R_l^{(2)}(r),$$
$$\chi_1^B=\exp(-i\omega t)\exp(im\phi)\,{}_1Y_l^m(\theta)R_l^{(1)}(r), \tag{2.18}$$
$$\chi_{-1}^B=\exp(-i\omega t)\exp(im\phi)\,{}_{-1}Y_l^m(\theta)\frac{\Delta}{2r^2}R_l^{(-1)}(r),$$
$$\psi_4^B=\exp(-i\omega t)\exp(im\phi)\,{}_{-2}Y_l^m(\theta)\frac{\Delta^2}{4r^4}R_l^{(-2)}(r),$$

where the angular functions, ${}_sY_l^m(\theta)$ are the spin-weighted spherical harmonics[8,9] which satisfy

$$\bigg(\frac{d}{d\theta}-\frac{m}{\sin\theta}-s\frac{\cos\theta}{\sin\theta}\bigg){}_sY_l^m=-\sqrt{(l-s)(l+s+1)}\,{}_{s+1}Y_l^m,$$
$$\bigg(\frac{d}{d\theta}+\frac{m}{\sin\theta}+s\frac{\cos\theta}{\sin\theta}\bigg){}_sY_l^m=\sqrt{(l+s)(l-s+1)}\,{}_{s-1}Y_l^m, \tag{2.19}$$
$$l=|s|,|s|+1,\cdots,\ -l\le m\le l$$

The above forms of the radial parts of χ_{-1}^B and ψ_4^B are chosen becuase $R_l^{(-1)*}$ and $R_l^{(-2)*}$ satisfy the same equations as $R_l^{(1)}$ and $R_l^{(2)}$ do.

The radial parts of Eqs. (2.13) and (2.14) then satisfy

$$\bigg[-\omega^2\frac{r^4}{\Delta}+4i\omega r\bigg(-2+\frac{r(r-M)}{\Delta}+\frac{Q^2}{3Mr-4Q^2}\bigg)$$
$$-\Delta\frac{d^2}{dr^2}-\bigg\{6(r-M)-\frac{4Q^2\Delta}{r(3Mr-4Q^2)}\bigg\}\frac{d}{dr}-4-\frac{2Q^2}{r^2}$$
$$+\frac{4Q^2(r^2+2Mr-3Q^2)}{r^2(3Mr-4Q^2)}+\frac{3Mr-4Q^2}{3Mr-2Q^2}(l-1)(l+2)\bigg]R_l^{(2)}$$
$$=\frac{2\sqrt{2}Q\sqrt{(l-1)(l+2)}r^3}{3Mr-2Q^2}\bigg(-i\omega\frac{r^2}{\Delta}+\frac{d}{dr}$$
$$+\frac{4}{r}-\frac{4Q^2}{r(3Mr-4Q^2)}\bigg)R_l^{(1)}, \tag{2.20}$$

$$\bigg[-\omega^2\frac{r^4}{\Delta}+2i\omega r\bigg(-2+\frac{r(r-M)}{\Delta}-\frac{Q^2}{3Mr-2Q^2}\bigg)$$
$$-\Delta\frac{d^2}{dr^2}-\bigg\{\frac{6\Delta}{r}+4(r-M)-\frac{2Q^2\Delta}{r(3Mr-2Q^2)}\bigg\}\frac{d}{dr}$$
$$-\frac{18r^2-24Mr+2Q^2}{r^2}$$
$$+\frac{12Q^2\Delta}{r^2(3Mr-2Q^2)}+\frac{3Mr-2Q^2}{3Mr-4Q^2}(l-1)(l+2)\bigg]R_1^{(1)}$$
$$=\frac{-\sqrt{2}Q^3\sqrt{(l-1)(l+2)}\Delta}{r^3(3Mr-4Q^2)}\bigg(i\omega\frac{r^2}{\Delta}+\frac{d}{dr}-\frac{2}{r}+\frac{4(r-M)}{\Delta}$$
$$-\frac{2Q^2}{r(3Mr-2Q^2)}\bigg)R_l^{(2)}. \tag{2.21}$$

III. ASYMPTOTIC SOLUTIONS

Given coupled equations for two quantities f and g in the form

$$\bigg(\frac{d^2}{dx^2}+A\frac{d}{dx}+B\bigg)f=O_1g, \tag{3.1}$$

$$\bigg(\frac{d}{dx}+C\bigg)f=O_2g, \tag{3.2}$$

one has to apply a first order differential operator, $d/dx+D$, on the first equation and a second order differential operator, $d^2/dx^2+E(d/dx)+F$, on the second, and subtract one from the other to get a completely decoupled equation for g. f is eliminated if D,E, and F

satisfy

$$D = C - \frac{1}{H}\frac{dH}{dx},$$

$$E = A - \frac{1}{H}\frac{dH}{dx}, \tag{3.3}$$

$$F = (A - C)(D - C) + \frac{dA}{dx} + B - 2\frac{dC}{dx},$$

where

$$H = B - AC + C^2 - \frac{dC}{dx}.$$

The decoupled equations obtained using the above method for $R_l^{(2)}$ and $R_l^{(1)}$ have the following asymptotic forms. At infinity,

$$\left[\frac{d^4}{dr'^4} + \frac{17}{r}\frac{d^3}{dr'^3} + \left(2\omega^2 + \frac{7i\omega}{r}\right)\frac{d^2}{dr'^2} + \frac{17\omega^2}{r}\frac{d}{dr'} + \omega^4 + \frac{7i\omega^3}{r}\right] R_l^{(2)} \approx 0,$$

$$\left[\frac{d^4}{dr'^4} + \frac{17}{r}\frac{d^3}{dr'^3} + \left(2\omega^2 + \frac{5i\omega}{r}\right)\frac{d^2}{dr'^2} + \frac{17\omega^2}{r}\frac{d}{dr'} + \omega^4 + \frac{5i\omega^3}{r}\right] R^{(1)} \approx 0, \tag{3.4}$$

and on the event horizon $(r \to r_+ = M + (M^2 - Q^2)^{1/2}$,

$$\left[\frac{d^4}{dr'^4} + \frac{4(r_+ - M)}{r_+^2}\frac{d^3}{dr'^3} + \left(2\omega^2 - 8i\omega\frac{r_+ - M}{r_+^2} - \frac{4(r_+ - M)^2}{r_+^4}\right)\frac{d^2}{dr'^2} + \left(4\omega^2\frac{r_+ - M}{r_+^2} - 16i\omega\frac{(r_+ - M)^2}{r_+^4} - 16\frac{(r_+ - M)^3}{r_+^6}\right)\frac{d}{dr'} + \omega^4 - 8i\omega^3\frac{r_+ - M}{r_+^2} - 20\omega^2\frac{(r_+ - M)^2}{r_+^4} + 16i\omega\frac{(r_+ - M)^3}{r_+^6}\right] R_l^{(2)} \approx 0, \tag{3.5}$$

$$\left[\frac{d^4}{dr'^4} + 4\frac{r_+ - M}{r_+^2}\frac{d^3}{dr'^3} + \left(2\omega^2 - 4i\omega\frac{r_+ - M}{r_+^2} + \frac{4(r_+ - M)^2}{r_+^4}\right)\frac{d^2}{dr'^2} + \left(4\omega^2\frac{r_+ - M}{r_+^2} - 8i\omega\frac{(r_+ - M)^2}{r_+^4}\right)\frac{d}{dr^1} + \omega^4 - 4i\omega^3\frac{r_+ - M}{r_+^2} - 4\omega^2\frac{(r_+ - M)^2}{r_+^4}\right] R_l^{(1)} \approx 0,$$

where

$$\frac{d}{dr'} = \frac{\Delta}{r^2}\frac{d}{dr}.$$

The asymptotic solutions obtained from these equations are, at infinity,

$$R_l^{(2)} \approx \frac{\exp(-i\omega r')}{r}, \quad \frac{\exp(-i\omega r')}{r^3}, \quad \frac{\exp(i\omega r')}{r^5}, \quad \frac{\exp(i\omega r')}{r^6},$$

$$R_l^{(1)} \approx \frac{\exp(-i\omega r')}{r^4}, \quad \frac{\exp(-i\omega r')}{r^5}, \quad \frac{\exp(i\omega r')}{r^6}, \quad \frac{\exp(i\omega r')}{r^8}, \tag{3.6}$$

on the event horizon,

$$R_l^{(2)} \approx \frac{\exp(-i\omega r')}{\Delta^2}, \quad \frac{\exp(-i\omega r')}{\Delta}, \quad \exp(i\omega r'), \quad \Delta\exp(i\omega r'),$$

$$R_l^{(1)} \approx \frac{\exp(-i\omega r')}{\Delta}, \quad \exp(-i\omega r'), \exp(i\omega r'), \Delta\exp(i\omega r'). \tag{3.7}$$

Thus $R_l^{(1)}$ and $R_l^{(2)}$ each have four coefficients only two of which are independent since $R_l^{(1)}$ and $R_l^{(2)}$ have to satisfy Eqs. (2.20) and (2.21). One can also obtain the next highest terms by putting the asymptotic solutions in Eqs. (2.20) and (2.21), the results of which are needed in our later calculation in Appendix B. One group of quantities (ψ_0^B, χ_1^B) is related to the other (ψ_4^B, χ_{-1}^B) through the Bianchi identities which have not been used in Sec. II [Eqs. (B1), (B2), and (B3)]. Since these equations involve the complex conjugate of some quantities, we take the following forms for ψ_0^B, χ_1^B, ψ_4^B, and χ_{-1}^B for a given ω:

$$\psi_0^B = \exp(-i\omega t)\exp(im\phi)_2Y_l^m R_l^{(2)} - \exp(i\omega t) \times \exp(-im\phi)_2Y_l^{-m} P_l^{(1)},$$

$$\chi_1^B = \exp(-i\omega t)\exp(im\phi)_1Y_l^m R_l^{(1)} - \exp(i\omega t) \times \exp(-im\phi)_1Y_l^{-m} P_l^{(1)}, \tag{3.8}$$

$$\chi_{-1}^B = \exp(-i\omega t)\exp(im\phi)_{-1}Y_l^m \frac{\Delta}{2r^2} R_l^{(-1)} - \exp(i\omega t) \times \exp(-im\phi)_{-l}Y_l^{-m} \frac{\Delta}{2r^2} P_l^{(-1)},$$

$$\psi_4^B = \exp(-i\omega t)\exp(im\phi)_{-2}Y_l^m \frac{\Delta^2}{4r^4} R_l^{(-2)} - \exp(i\omega t) \times \exp(-im\phi)_{-2}Y_l^{-m} \frac{\Delta^2}{4\gamma^4} P_l^{(-2)}.$$

Each of the sets $(R_l^{(1)}, R_l^{(2)})$, $(P_l^{(1)*}, P_l^{(2)*})$, $(R^{(-1)*}, R^{(-2)*})$ and $(P_l^{(-1)}, P_l^{(-2)})$ satisfies the same equations [Eqs. (2.20) and (2.21)]. The asymptotic forms at infinity are

$$\psi_0^B \approx \exp(-i\omega t)\exp(im\phi)_2Y_l^m \left(A^{(2)}\frac{\exp(-i\omega r')}{r} + B^{(2)}\frac{\exp(i\omega r')}{r^5}\right) - \exp(i\omega t)\exp(-im\phi)_2Y_l^{-m}\left(C^{(2)}\frac{\exp(-i\omega r')}{r} + D^{(2)}\frac{\exp(-i\omega r')}{r^5}\right),$$

$$\chi_1^B \approx \exp(-i\omega t)\exp(im\phi)_1Y_l^m \left(A^{(1)}\frac{\exp(-i\omega r')}{r^4} + B^{(1)}\frac{\exp(i\omega r')}{r^6}\right) - \exp(i\omega t)\exp(-im\phi)_1Y_l^{-m}\left(C^{(1)}\frac{\exp(i\omega r')}{r^4} + D^{(1)}\frac{\exp(-i\omega r)}{r^6}\right),$$

$$\chi_{-1}^B \approx \exp(-i\omega t)\exp(im\phi)_{-1}Y_l^m \left(\frac{A^{(-1)}}{2}\frac{\exp(i\omega r')}{r^4} + \frac{B^{(-1)}}{2}\frac{\exp(-i\omega r')}{r^6}\right) - \exp(i\omega t)\exp(-im\phi)_{-1}Y_l^{-m}\left(\frac{C^{(-1)}}{2}\frac{\exp(-i\omega r')}{r^4}\right.$$

$$+\frac{D^{(-1)}}{2}\frac{\exp(i\omega r')}{r^6}\bigg), \tag{3.9}$$

$$\psi_4^B \approx \exp(-i\omega t)\exp(im\phi)_{-2}Y_l^m\bigg(\frac{A^{(-2)}}{4}\frac{\exp(i\omega r')}{r} + \frac{B^{(-2)}}{4}\frac{\exp(-i\omega r')}{r^5}\bigg) - \exp(i\omega t)\exp(-im\phi)_{-2}Y_l^{-m}\bigg(\frac{C^{(-2)}}{4}\frac{\exp(-i\omega r')}{r} + \frac{D^{(-2)}}{4}\frac{\exp(i\omega r')}{r^5}\bigg).$$

On the event horizon one cannot use Boyer–Lindquist coordinates to discuss the boundary conditions. However Teukolsky's argument[5] using Kerr "ingoing" coordinates for the Kerr metric can easily be generalized to the Kerr–Newman metric. Using a nonsingular tetrad in Kerr "ingoing" coordinates the boundary condition is that the solutions be nonspecial or that the group velocity be negative. This condition yields the following asymptotic forms on the event horizon:

$$\begin{aligned}
\psi_0^B &\approx \exp(-i\omega t)\exp(im\phi)_2Y_l^m\alpha^{(2)}\frac{\exp(-i\omega r')}{\Delta^2} - \exp(i\omega t)\exp(-im\phi)_2Y_l^{-m}\beta^{(2)}\frac{\exp(i\omega r')}{\Delta^2},\\
\chi_1^B &\approx \exp(-i\omega t)\exp(im\phi)_1Y_l^m\alpha^{(1)}\frac{\exp(-i\omega r')}{\Delta} - \exp(i\omega t)\exp(-im\phi)_1Y_l^{-m}\beta^{(1)}\frac{\exp(i\omega r')}{\Delta},\\
\chi_{-1}^B &\approx \exp(-i\omega t)\exp(im\phi)_{-1}Y_l^m\alpha^{(-1)}\frac{\Delta}{2r_+^2}\exp(-i\omega r') - \exp(i\omega t)\exp(-im\phi)_{-1}Y_l^{-m}\beta^{(-1)}\frac{\Delta}{2r^2}\exp(i\omega r'),\\
\psi_4^B &\approx \exp(-i\omega t)\exp(im\phi)_{-2}Y_l^m\alpha^{(-2)}\frac{\Delta^2}{4r_+^4}\exp(-i\omega r') - \exp(i\omega t)\exp(-im\phi)_{-2}Y_l^{-m}\beta^{(-2)}\frac{\Delta^2}{4r_+^4}\exp(i\omega r').
\end{aligned} \tag{3.10}$$

The procedure to get the connections between the coefficients is described in Appendix B and the final results are listed below:

$$\begin{aligned}
B^{(1)} &= -\frac{l(l+1)}{4\omega^2}A^{(-1)} - \frac{Q^2}{3i\omega M}C^{(-1)*} + \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}}{12\omega^2 M}C^{(-2)*},\\
D^{(1)} &= \frac{Q^2}{3i\omega M}A^{(-1)*} - \frac{l(l+1)}{4\omega^2}C^{(-1)} + \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}}{12\omega^2 M}A^{(-2)*},\\
B^{(2)} &= \frac{(l-1)l(l+1)(l+2)}{16\omega^4}A^{(-2)} - \bigg(\frac{3M}{4i\omega^3} + \frac{Q^2(l-1)(l+2)}{12i\omega^3 M}\bigg)C^{(-2)*} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{6\omega^2 M}C^{(-1)*},\\
D^{(2)} &= \bigg(\frac{3M}{4i\omega^3} + \frac{Q^2(l-1)(l+2)}{12i\omega^3 M}\bigg)A^{(-2)*} + \frac{(l-1)l(l+1)(l+2)}{16\omega^4}C^{(-2)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{6\omega^2 M}A^{(-1)*},
\end{aligned} \tag{3.11}$$

$$\begin{aligned}
B^{(-1)} &= -\frac{l(l+1)}{4\omega^2}A^{(1)} + \frac{Q^2}{3i\omega M}C^{(1)*} + \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}}{12\omega^2 M}C^{(2)*},\\
D^{(-1)} &= -\frac{Q^2}{3i\omega M}A^{(1)*} - \frac{l(l+1)}{4\omega^2}C^{(1)} + \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}}{12\omega^2 M}A^{(2)*},\\
B^{(-2)} &= \frac{(l-1)l(l+1)(l+2)}{16\omega^4}A^{(2)} + \bigg(\frac{3M}{4i\omega^3} + \frac{Q^2(l-1)(l+2)}{12i\omega^3 M}\bigg)C^{(2)*} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{6\omega^2 M}C^{(1)*},\\
D^{(-2)} &= -\bigg(\frac{3M}{4i\omega^3} + \frac{Q^2(l-1)(l+2)}{12i\omega^3 M}\bigg)A^{(2)*} + \frac{(l-1)l(l+1)(l+2)}{16\omega^4}C^{(2)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{6\omega^2 M}A^{(1)*},
\end{aligned} \tag{3.12}$$

$$\begin{aligned}
[(r_+ - M) - i\omega r_+^2]\alpha^{(-1)} &= \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}\,l(l+1)}{8i\omega r_+^5(3Mr_+ - 2Q^2)(r_+ - M + i\omega r_+^2)}\alpha^{(2)}\\
&\quad + \frac{\sqrt{2}\,Q^3\sqrt{(l-1)(l+2)}}{4r_+^4(3Mr_+ - 2Q^2)(r_+ - M + i\omega r_+^2)}\beta^{(2)*} - \frac{3Ml(l+1)}{4i\omega r_+(3Mr_+ - 2Q^2)}\alpha^{(1)} - \frac{Q^2}{r_+(3Mr_+ - 2Q^2)}\beta^{(1)*},\\
[(r_+ - M) + i\omega r_+^2]\beta^{(-1)} &= \frac{\sqrt{2}Q^3\sqrt{(l-1)(l+2)}}{4r_+^4(3Mr_+ - 2Q^2)(r_+ - M - i\omega r_+^2)}\alpha^{(2)*}\\
&\quad - \frac{\sqrt{2}\,Q^3\sqrt{(l-1)(l+2)}\,l(l+1)}{8i\omega r_+^5(3Mr_+ - 2Q^2)(r_+ - M - i\omega r_+^2)}\beta^{(2)} - \frac{Q^2}{r_+(3Mr_+ - 2Q^2)}\alpha^{(1)*} + \frac{3Ml(l+1)}{4i\omega r_+(3Mr_+ - 2Q^2)}\beta^{(1)},\\
[2(r_+ - M) - i\omega r_+^2]\alpha^{(-2)} &= \frac{(l-1)l(l+1)(l+2)(3Mr_+ - 4Q^2)^2(1 - 4Q^4/(3Mr_+ - 4Q^2)^2)}{16i\omega r_+^2(3Mr_+ - 2Q^2)^2|r_+ - M + i\omega r_+^2|^2}\alpha^{(2)}\\
&\quad - \bigg(\frac{3Mr_+ - 4Q^2}{4r_+^3|r_+ - M + i\omega r_+^2|^2} + \frac{Q^2(l-1)(l+2)}{4r_+(3Mr_+ - 2Q^2)|r_+ - M + i\omega r_+^2|^2}\bigg)\beta^{(2)*}
\end{aligned} \tag{3.13}$$

$$+\frac{Q\sqrt{(l-1)(l+2)l(l+1)}r_+}{\sqrt{2}\,i\omega(3Mr_+-2Q^2)(r_+-M-i\omega r_+^2)}\alpha^{(1)}+\frac{Q\sqrt{(l-1)(l+2)}r_+^2}{\sqrt{3}\,(3Mr_+-2Q^2)(r_+-M-i\omega r_+^2)}\beta^{(1)*},$$

$$[2(r_+-M)+i\omega r_+^2]\beta^{(-2)}=-\left(\frac{3Mr_+-4Q^2}{4r_+^3|r_+-M+i\omega r_+^2|^2}+\frac{Q^2(l-1)(l+2)}{4r_+(3Mr_+-2Q^2)|r_+-M+i\omega r_+^2|^2}\right)\alpha^{(2)*}$$
$$-\frac{(l-1)l(l+1)(l+2)(3Mr_+-4Q^2)^2(1-4Q^4/(3Mr_+-4Q^2)^2)}{16i\omega r_+^2(3Mr_+-2Q^2)^2|r_+-M+i\omega r_+^2|^2}\beta^{(2)}$$
$$+\frac{Q\sqrt{(l-1)(l+2)}r_+^2}{\sqrt{2}(3Mr_+-2Q^2)(r_+-M-i\omega r_+^2)}\alpha^{(1)*}-\frac{Q\sqrt{(l-1)(l+2)}l(l+1)r_+}{\sqrt{2}i\omega(3Mr_+-2Q^2)(r_+-M-i\omega r_+^2)}\beta^{(1)}.$$

IV. "WRONSKIAN" OF COUPLED PERTURBATION EQUATIONS[10]

From the fact that each of the sets $(R^{(1)},R^{(2)})$, $(P^{(1)*},P^{(2)*})$, $(R^{(-1)*},R^{(-2)*})$, and $(P^{(-1)},P^{(-2)})$ satisfies the same coupled equations, one can derive a "Wronskian" which is conserved (i.e., independent of r). It is useful to make the transformations

$$X^{(2)}=\frac{Qr^{3/2}\Delta}{(3Mr-4Q^2)^{1/2}}R^{(2)},\quad T^{(2)*}=\frac{Qr^{3/2}\Delta}{(3Mr-4Q^2)^{1/2}}P^{(2)*},$$
$$X^{(1)}=\frac{\sqrt{2}r^{9/2}\Delta^{1/2}}{(3Mr-2Q^2)^{1/2}}R^{(1)},\quad T^{(1)*}=\frac{\sqrt{2}r^{9/2}\Delta^{1/2}}{(3Mr-2Q^2)^{1/2}}P^{(1)*},$$
$$X^{(-1)*}=\frac{\sqrt{2}r^{9/2}\Delta^{1/2}}{(3Mr-2Q^2)^{1/2}}R^{(-1)*},\quad T^{(-1)}=\frac{\sqrt{2}r^{9/2}\Delta^{1/2}}{(3Mr-2Q^2)^{1/2}}P^{(-1)},$$
$$X^{(-2)*}=\frac{Qr^{3/2}\Delta}{(3Mr-4Q^2)^{1/2}}R^{(-2)*},\quad T^{(-2)}=\frac{Qr^{3/2}\Delta}{(3Mr-4Q^2)^{1/2}}P^{(-2)},\tag{4.1}$$

and

$$\frac{d}{dr'}=\frac{\Delta}{r^2}\frac{d}{dr}.$$

Then Eq. (2.20) becomes

$$O_1\begin{pmatrix}X^{(2)}\\T^{(2)*}\\X^{(-2)*}\\T^{(-2)}\end{pmatrix}=O_2\begin{pmatrix}X^{(1)}\\T^{(1)*}\\X^{(-1)*}\\T^{(-1)}\end{pmatrix},\tag{4.2}$$

where

$$O_1=\frac{d^2}{dr'^2}+\Bigg\{\omega^2-4i\omega\left(-\frac{2\Delta}{r^3}+\frac{r-M}{r^2}+\frac{Q^2\Delta}{r^3(3Mr-4Q^2)}\right)$$
$$-\frac{8(r-M)^2}{r^4}$$
$$-\frac{\Delta(3r^2-4Mr-3Q^2)}{r^6}+\frac{4Q^2\Delta(4r^2-7Mr+3Q^2)}{r^6(3Mr-4Q^2)}$$
$$-\frac{16Q^4\Delta^2}{r^6(3Mr-4Q^2)^2}-\frac{3Mr-4Q^2}{3Mr-2Q^2}\frac{\Delta}{r^4}(l-1)(l+2)\Bigg\},$$

$$O_2=\frac{2Q^2\sqrt{(l-1)(l+2)}}{(3Mr-2Q^2)^{1/2}(3Mr-4Q^2)^{1/2}}\frac{\Delta^{1/2}}{r^2}\left(-\frac{d}{dr'}+i\omega\right.$$
$$\left.+\frac{r-M}{r^2}-\frac{Q^2\Delta}{r^3(3Mr-2Q^2)}+\frac{4Q^2\Delta}{r^3(3Mr-4Q^2)}\right).$$

Similarly Eq. (2.21) becomes

$$O_3\begin{pmatrix}X^{(1)}\\T^{(-1)*}\\X^{(-1)*}\\T^{(-1)}\end{pmatrix}=O_4\begin{pmatrix}X^{(2)}\\T^{(2)*}\\X^{(-2)*}\\T^{(-2)}\end{pmatrix},\tag{4.3}$$

where

$$O_3=\frac{d^2}{dr'^2}+\Bigg\{\omega^2-2i\omega\left(-\frac{2\Delta}{r^3}-\frac{Q^2\Delta}{r^3(3Mr-2Q^2)}\right)-\frac{2(r-M)^2}{r^4}$$
$$-\frac{\Delta(25r^2-38Mr+18Q^2)}{r^6}+\frac{6Q^2\Delta(r-M)}{r^6(3Mr-2Q^2)}$$
$$-\frac{4Q^4\Delta^2}{r^6(3Mr-2Q^2)^2}-\frac{3Mr-2Q^2}{3Mr-4Q^2}\frac{\Delta}{r^4}(l-1)(l+2)\Bigg\},$$

$$O_4=\frac{2Q^2\sqrt{(l-1)(l+2)}}{(3Mr-2Q^2)^{1/2}(3Mr-4Q^2)^{1/2}}\frac{\Delta^{1/2}}{r^2}$$
$$\times\left(\frac{d}{dr'}+i\omega-\frac{3\Delta}{r^3}+\frac{2(r-M)}{r^2}\right.$$
$$\left.-\frac{2Q^2\Delta}{r^3(3Mr-2Q^2)}+\frac{2Q^2\Delta}{r^3(3Mr-4Q^2)}\right).$$

Now from Eqs. (4.2) and (4.3),

$$T^{(-2)}O_1X^{(2)}-X^{(2)}O_1T^{(-2)}+T^{(-1)}O_3X^{(1)}-X^{(1)}O_3T^{(-1)}$$
$$=T^{(-2)}O_2X^{(1)}-X^{(2)}O_2T^{(-1)}+T^{(-1)}O_4X^{(2)}-X^{(1)}O_4T^{(-2)}\tag{4.4}$$

or explicitly

$$\frac{d}{dr'}W_1\equiv\frac{d}{dr'}\Bigg[-T^{(-2)}\frac{dX^{(2)}}{dr'}+X^{(2)}\frac{dT^{(-2)}}{dr'}-T^{(-1)}\frac{dX^{(1)}}{dr'}$$
$$+X^{(1)}\frac{dT^{(-1)}}{dr'}$$
$$-\frac{2Q^2\sqrt{(l-1)(l+2)}}{(3Mr-2Q^2)^{1/2}(3Mr-4Q^2)^{1/2}}\frac{\Delta^{1/2}}{r^2}$$
$$\times(X^{(1)}T^{(-2)}-X^{(2)}T^{(-1)})\Bigg]=0,$$

hence, a conserved quantity, W_1.

Another combination similar to Eq. (4.4) but with different members results in another conserved quantity,

$$X^{(-2)}O_1^*T^{(2)}-T^{(2)}O_1^*X^{(-2)}+X^{(-1)}O_3^*T^{(1)}-T^{(1)}O_3^*X^{(-1)}$$
$$=X^{(-2)}O_2^*T^{(1)}-T^{(2)}O_2^*X^{(-1)}+X^{(-1)}O_4^*T^{(2)}-T^{(1)}O_4^*X^{(-2)}\tag{4.5}$$

or explicitly

$$\frac{d}{dr'} W_2 \equiv \frac{d}{dr'} \Big[-X^{(-2)} \frac{dT^{(2)}}{dr'} + T^{(2)} \frac{dX^{(-2)}}{dr'} + T^{(2)} \frac{dX^{(-2)}}{dr'}$$

$$- X^{(-1)} \frac{dT^{(1)}}{dr'} \; T^{(1)} \frac{dX^{(-1)}}{dr'}$$

$$- \frac{2Q^2\sqrt{(l-1)(l+2)}}{(3Mr-2Q^2)^{1/2}(3Mr-4Q^2)^{1/2}*} \frac{\Delta^{1/2}}{r^2}$$

$$\times (X^{(-2)}T^{(1)} - T^{(2)}X^{(-1)} \Big] = 0,$$

which gives another conserved quantity, W_2.

Now, using the asymptotic solutions at infinity [Eq. (3.9)] and the connections between the coefficients [Eqs. (3.11) and (3.12)], one obtains the sum of W_1 and W_2,

$$W = W_1 + W_2$$

$$= \frac{2i\omega}{3M}(Q^2A^{(2)}D^{(-2)} - Q^2B^{(2)}C^{(-2)} + 2A^{(1)}D^{(-1)} - 2B^{(1)}C^{(-1)}$$

$$+ \frac{2i\omega}{3M}(Q^2A^{(-2)}D^{(2)} - Q^2B^{(-2)}C^{(2)} + 2A^{(-1)}D^{(1)} - 2B^{(-1)}C^{(1)})$$

$$= -\frac{4Q^2}{9M^2}\left[\frac{9M^2}{8\omega^2}|A^{(2)}|^2 + \left|A^{(1)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}A^{(2)}\right|^2\right.$$

$$\left. + \frac{9M^2}{8\omega^2}|C^{(2)}|^2 + \left|C^{(1)} + \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}C^{(2)}\right|^2\right]$$

$$+ \frac{4Q^2}{9M^2}\left[\frac{9M^2}{8\omega^2}|A^{(-2)}|^2 + \left|A^{(-1)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}A^{(-2)}\right|^2\right.$$

$$\left. + \frac{9M^2}{8\omega^2}|C^{(-2)}|^2 + \left|C^{(-1)} + \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}C^{(-2)}\right|^2\right], \tag{4.6}$$

which is exactly the same as the energy flux at infinity up to a constant factor. [Compare Eq. (4.6) with Eq. (A12).]

To prove that the ingoing energy flux is greater than the outgoing energy flux, W is evaluated on the event horizon using Eqs. (3.10) and (3.13),

$$W = \frac{2Q^2r_+}{3Mr_+ - 4Q^2}\left[\{2(r_+ - M) + i\omega r_+^2\}\alpha^{(2)}\beta^{(-2)}\right.$$

$$\left. + \{2(r_+ - M) - i\omega r_+^2\}\alpha^{(-2)}\beta^{(2)}\right]$$

$$+ \frac{4r_+^4}{3Mr_+ - 2Q^2}\left[(r_+ - M + i\omega r_+^2)\alpha^{(1)}\beta^{(-1)}\right.$$

$$\left. + (r_+ - M - i\omega r_+^2)\alpha^{(-1)}\beta^{(1)}\right]$$

$$+ \frac{2\sqrt{2}Q^3\sqrt{(l-1)(l+2)}r_+^4}{(3Mr_+ - 2Q^2)(3Mr_+ - 4Q^2)}\left[\alpha^{(2)}\beta^{(-1)} + \beta^{(2)}\alpha^{(-1)}\right]$$

$$= -\frac{4Q^2}{(3Mr_+ - 2Q^2)^2}\left[\left|r_+^3\alpha^{(1)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4(r_+ - M + i\omega r_+^2)}\alpha^{(2)}\right|^2\right.$$

$$+ \left|r_+^3\beta^{(1)} - \frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4(r_+ - M - i\omega r_+^2)}\beta^{(2)}\right|^2$$

$$\left. + \frac{(3Mr_+ - 2Q^2)^2}{8r_+^2|r_+ - M + i\omega r_+^2|^2}(|\alpha^{(2)}|^2 + |\beta^{(2)}|^2)\right]. \tag{4.7}$$

So, W is always negative which guarantees that the net energy flow is radially inward. We also see that there exists no solution with $A^{(1)} = C^{(1)} = A^{(2)} = C^{(2)} = 0$ for ω real, that is with zero incoming waves—a first step in an analytic proof that the Reissner–Nordström metric is stable. However, using the Hamiltonian formalism, Moncrief[11] obtained equations for coupled gravitational and electromagnetic perturbations of a nonrotating (Reissner–Nordström) black hole and has established the stability of Reissner–Nordström black holes. Also, Chitre[12] has developed a gravitational perturbation equation of a Kerr–Newman black hole in the limit of small charge.

The coupled equations [Eqs. (2.20) and (2.21)] can be solved numerically to investigate, for instance, the scattering of electromagnetic waves on a charged black hole and the resulting generation of outgoing gravitational wave.

ACKNOWLEDGMENTS

The author is greatly indebted to Professor W.D. McGlinn for his continuous encouragement and many helpful suggestions. He also thanks Professor S.K. Bose for many helpful discussions.

APPENDIX A: ELECTROMAGNETIC AND GRAVITATIONAL ENERGY FLUX

The expression of the electromagnetic energy flux in terms of ϕ's and the gravitational energy flux in terms of ψ's for an uncharged background metric are given in Ref. 5,

$$\frac{d^2E_{el}^{out}}{dt\,d\Omega} = \lim_{r\to\infty} \frac{r^2}{2\pi} |\phi_2^B|^2, \tag{A1}$$

$$\frac{d^2E_{el}^{in}}{dt\,d\Omega} = \lim_{r\to\infty} \frac{r^2}{8\pi} |\phi_0^B|^2, \tag{A2}$$

$$\frac{d^2E_{gr}^{out}}{dt\,d\Omega} = \lim_{r\to\infty} \frac{r^2}{4\pi\omega^2} |\psi_4^B|^2, \tag{A3}$$

$$\frac{d^2E_{gr}^{in}}{dt\,d\Omega} = \lim_{r\to\infty} \frac{r^2}{64\pi\omega^2} |\psi_0^B|^2. \tag{A4}$$

These expressions also hold for a charged background metric if one imposes the boundary condition that the perturbed tetrad reduces to the same flat space–time tetrad at infinity as the unperturbed tetrad does.

Then, the first order perturbation of Eq. (2.1),

$$(D^B - 2\rho^B)\phi_1 + (D - 2\rho)\phi_1^B - (\delta^* - 2\alpha)\phi_0^B = 0, \tag{A5}$$

shows that the asymptotic form of ϕ_1^B at infinity is

$$\phi_1^B \sim \exp(-i\omega r')/r^2. \tag{A6}$$

Using Eq. (A6) and the boundary condition on the tetrad and Eqs. (1.2) and (1.4) for the unperturbed quantities, one can see that the electromagnetic energy flux has the form at infinity for both charged and uncharged background metrics,

$$\lim_{r\to\infty}(r^2T^{01}) = \lim_{r\to\infty}\left(\frac{r^2}{2\pi}|\phi_2^B|^2 - \frac{r^2}{8\pi}|\phi_0^B|^2\right). \tag{A7}$$

For the gravitational energy flux, it is the way ψ_0 and ψ_4 are defined[3] that guarantees that ψ_0^B and ψ_4^B are related to the gravitational energy flux in the same way for both charged and uncharged background metrics,

$$\begin{aligned}\psi_0 &\equiv -C_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\delta \\ &= -[R_{\alpha\beta\gamma\delta}+\tfrac{1}{2}(g_{\alpha\gamma}R_{\beta\delta}-g_{\alpha\delta}R_{\beta\gamma}+g_{\beta\delta}R_{\alpha\gamma}-g_{\beta\gamma}R_{\alpha\delta}) \\ &\quad +(R/6)(g_{\alpha\delta}R_{\beta\gamma}-g_{\alpha\gamma}R_{\beta\delta})]l^\alpha m^\beta l^\gamma m^\delta \\ &= -R_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\beta l^\gamma m^\delta. \end{aligned} \tag{A8}$$

The last step of Eq. (A8) is done by the orthogonality of the tetrad. For a pure gravitational perturbation in an uncharged background metric $R_{\mu\nu}=0$ anyway. But even for a perturbation in a charged background metric the terms involving $R_{\mu\nu}$ drop automatically as shown in Eq. (A8). Then, with the boundary condition on the tetrad, one can show that

$$\lim_{r\to\infty}\frac{r^2}{64\pi\omega^2}|\psi_0|^2=\lim_{r\to\infty}\frac{r^2}{64\pi\omega^2}|R_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\delta|^2$$

represents the ingoing gravitational energy flux.[5,13] To find the asymptotic form of ϕ_0^B one first obtains the asymptotic form of ψ_1^B by noting the dominant terms in Eq. (2.9) with the boundary condition on the tetrad.

Using the asymptotic form of ψ_0^B [Eq. (3.10)] in Eq. (2.9) one obtains

$$\begin{aligned}\psi_1^B \approx & -\exp(-i\omega t)\exp(im\phi)\,{}_1Y_l^m\frac{\sqrt{(l-1)(l+2)}}{2\sqrt{2}i\omega}A^{(2)}\frac{\exp(-i\omega r')}{r^2} \\ & -\exp(i\omega t)\exp(-im\phi)\,{}_1Y_l^{-m}\frac{\sqrt{(l-1)(l+2)}}{2\sqrt{2}i\omega}C^{(2)}\frac{\exp(i\omega r')}{r^2}.\end{aligned} \tag{A9}$$

Then,

$$\begin{aligned}\phi_0^B &= \frac{\chi_1^B+2\phi_1\psi_1^B}{3\psi_2} \\ &\approx \exp(-i\omega t)\exp(im\phi)\,{}_1Y_l^m\left(\frac{A^{(1)}}{3M}-\frac{Q\sqrt{(l-1)(l+2)}}{6\sqrt{2}i\omega M}A^{(2)}\right) \\ &\quad\times\frac{\exp(-i\omega r')}{r} \\ &\quad -\exp(i\omega t)\exp(-im\phi)\,{}_1Y_l^{-m}\left(\frac{C^{(1)}}{3M}+\frac{Q\sqrt{(l-1)(l+2)}}{6\sqrt{2}i\omega M}C^{(2)}\right) \\ &\quad\times\frac{\exp(i\omega r')}{r}.\end{aligned} \tag{A10}$$

A similar procedure for ϕ_2^B results in

$$\begin{aligned}\phi_2^B &= \frac{\chi_{-1}^B+2\phi_1\psi_3^B}{3\psi_2} \\ &\approx \exp(-i\omega t)\exp(im\phi)\,{}_{-1}Y_l^m\left(\frac{A^{(-1)}}{6M}-\frac{Q\sqrt{(l-1)(l+2)}}{12\sqrt{2}i\omega M}A^{(-2)}\right) \\ &\quad\times\frac{\exp(i\omega r')}{r} \\ &\quad -\exp(i\omega t)\exp(-im\phi)\,{}_{-1}Y_l^{-m}\left(\frac{C^{(-1)}}{6M}\right. \\ &\quad\times\left.\frac{Q\sqrt{(l-1)(l+2)}}{12\sqrt{2}i\omega M}C^{(-2)}\right)\frac{\exp(-i\omega r')}{r}.\end{aligned} \tag{A11}$$

Now we have all the necessary asymptotic forms to get the energy flux. Using Eqs. (A1)–(A4) (the angular functions Y, are normalized so that $\int Y^2 d\Omega=1$),

$$\begin{aligned}\frac{dE^{\rm total}}{dt} &= \frac{dE^{\rm out}_{\rm gr}}{dt}-\frac{dE^{\rm in}_{\rm gr}}{dt}+\frac{dE^{\rm out}_{\rm el}}{dt}-\frac{dE^{\rm in}_{\rm el}}{dt} \\ &= \frac{1}{64\pi\omega^2}[|A^{(-2)}|^2+|C^{(-2)}|^2-|A^{(2)}|^2-|C^{(2)}|^2] \\ &\quad+\frac{1}{72\pi M^2}\Big[\Big|A^{(-1)}-\frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega} \\ &\quad\times A^{(-2)}\Big|^2+\Big|C^{(-1)}+\frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}C^{(-2)}\Big|^2 \\ &\quad-\Big|A^{(1)}-\frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega} \\ &\quad\times A^{(2)}\Big|^2-\Big|C^{(1)}+\frac{\sqrt{2}Q\sqrt{(l-1)(l+2)}}{4i\omega}C^{(2)}\Big|^2\Big].\end{aligned} \tag{A12}$$

APPENDIX B: RELATIONS BETWEEN (ψ_0^B, χ_1^B) and (ψ_4^B, χ_{-1}^B)

To find the relations between the two groups we use the following first order perturbation equations of the Bianchi identities:

$$\begin{aligned}&3(D^B-3\rho^B)\psi_2+3(D-3\rho)\psi_2^B-3(\delta^*-2\alpha)\psi_1^B \\ &\quad-2(D^B-2\rho^{*B}+\rho^B)\phi_{11} \\ &\quad-2(D-\rho)\phi_{11}^B-2\phi_1(\delta^*-2\alpha)\phi_0^B+4\phi_1(\delta-2\alpha)\phi_0^{*B}=0,\end{aligned} \tag{B1}$$

$$\begin{aligned}&3(\delta^B-3\tau^B)\psi_2+3\delta\psi_2^B-3(\Delta-2\gamma+2\mu)\psi_1^B+2(\delta^B+\tau^B+2\pi^{*B})\phi_{11} \\ &\quad+2\delta\phi_{11}^B+2\phi_1(\Delta-2\mu-2\gamma)\phi_0^B-4\phi_1(D+2\rho)\phi_2^{*B}=0,\end{aligned} \tag{B2}$$

$$\begin{aligned}&3(\delta^{*B}+3\pi^B)\psi_2+3\delta^*\psi_2^B-3(D-2\rho)\psi_3^B+2(\delta^{*B}-\pi^B-2\tau^{*B})\phi_{11} \\ &\quad+2\delta^*\phi_{11}^B+2\phi_1(D+2\rho)\phi_2^B-4\phi_1(\Delta-2\mu-2\gamma)\phi_0^{*B}=0.\end{aligned} \tag{B3}$$

Subtracting the complex conjugate of Eq. (B1) from Eq. (B1) itself, one obtains

$$\begin{aligned}&\frac{1}{3\psi_2+2\phi_{11}}[(D-3\rho)Z^B-(\delta^*-2\alpha)\psi_1^B+(\delta-2\alpha)\psi_1^{*B} \\ &\quad-2\phi_1(\delta^*-2\alpha)\phi_0^B+2\phi_1(\delta-2\alpha)\phi_0^{*B}]=\rho^B-\rho^{*B},\end{aligned} \tag{B4}$$

where $Z^B=2i\,{\rm Im}(\psi_2^B)$, and subtraction of the complex conjugate of Eq. (B3) from Eq. (B2) gives

$$\begin{aligned}&\frac{1}{3\psi_2-2\phi_{11}}[\delta Z^B-(\Delta-2\gamma+2\mu)\psi_1^B+(D-2\rho)\psi_3^{*B} \\ &\quad+2\phi_1(\Delta-2\gamma-2\mu)\phi_0^B-2\phi_1(D+2\rho)\phi_2^{*B}]=\tau^B+\pi^{*B}.\end{aligned} \tag{B5}$$

Now operating $(D-2\rho)$ on Eq. (B4) and using the following N-P equations:

$$(D-2\rho)\rho^B-(\delta^*+2\beta)\kappa^B=0, \tag{B6}$$

$$(\delta^*-4\alpha)\psi_0^B-(D-4\rho)\psi_1^B+2\phi_1D\phi_0^B=(3\psi_2-2\phi_{11})\kappa^B, \tag{B7}$$

one finally obtains an equation involving only the quantities the asymptotic solutions of which we have and Z^B.

$$\frac{3\psi_2 - 2\phi_{11}}{3\psi_2 + 2\phi_{11}}\left(D - 5\rho - 2\rho\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(D - 3\rho)Z^B$$
$$= (\delta^* - 2\alpha)(\delta^* - 4\alpha)\psi_0^B - (\delta - 2\alpha)(\delta - 4\alpha)\psi_0^{*B}$$
$$+ \frac{4\phi_1}{3\psi_2 + 2\phi_{11}}\left(D - 5\rho - 2\rho\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(\delta^* - 2\alpha)\chi_1^B$$
$$- \frac{4\phi_1}{3\psi_2 + 2\phi_{11}}\left(D - 5\rho - 2\rho\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(\delta - 2\alpha)\chi_1^{*B}. \tag{B8}$$

Similarly, by operating $(\delta + 2\alpha)$ on Eq. (B5) and using

$$(\delta + 2\alpha)\tau^B - (\Delta + \mu - 2\gamma)\sigma^B - \rho\lambda^{*B} = 0, \tag{B9}$$

$$(\delta + 2\alpha)\pi^{*B} - (D - \rho)\chi^{B} + \mu\sigma^B = 0, \tag{B10}$$

$$(\Delta - 4\gamma + \mu)\psi_0^B - (\delta + 2\alpha)\psi_1^B - 2\phi_1(\delta + 2\alpha)\phi_0^B = (3\psi_2 + 2\phi_{11})\sigma^B, \tag{B11}$$

$$(D - \rho)\psi_4^B - (\delta^* + 2\alpha)\psi_3^B - 2\phi_1(\delta^* + 2\alpha)\phi_2^B = -(3\psi_2 + 2\phi_{11})\lambda^B, \tag{B12}$$

one gets another equation,

$$\frac{3\psi_2 + 2\phi_{11}}{3\psi_2 - 2\phi_{11}}(\delta + 2\alpha)\delta Z^B$$
$$= \left(\Delta - 2\gamma + 3\mu + 2\mu\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(\Delta - 4\gamma + \mu)\psi_0^B$$
$$- \left(D - 3\rho - 2\rho\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(D - \rho)\psi_4^{*B}$$
$$- \frac{4\phi_1}{3\psi_2 - 2\phi_{11}}\left(\Delta - 2\gamma + 3\mu + 2\mu\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(\delta + 2\alpha)\chi_1^B$$
$$+ \frac{4\phi_1}{3\psi_2 - 2\phi_{11}}\left(D - 3\rho - 2\rho\frac{4\phi_{11}}{3\psi_2 + 2\phi_{11}}\right)(\delta + 2\alpha)\chi_{-1}^{*B}. \tag{B13}$$

Equations (B8) and (B13) are separated by introducing Eq. (3.8) and

$$Z^B = \exp(-i\omega t)\exp(im\phi)_0Y_l^m(\theta)R_l^{(0)}(r) - \exp(i\omega t) \times \exp(-im\phi)_0Y_l^m(\theta)R_l^{(0)*}(r). \tag{B14}$$

With ${}_sY_l^{-m} = (-1)^s{}_{-s}Y_l^m$, the resulting radial equations are, from Eq. (B8),

$$\frac{3Mr - 2Q^2}{3Mr - 4Q^2}\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega - \frac{2(r-M)}{r^2} + \frac{7\Delta}{r^3} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{3\Delta}{r^3}\right)R_l^{(0)}$$
$$= \frac{\sqrt{(l-1)l(l+1)(l+2)}\Delta^2}{2r^6}(R_l^{(2)} + P_l^{(2)*})$$
$$- \frac{\sqrt{2}Q\sqrt{l(l+1)}\Delta}{(3Mr - 4Q^2)r}\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{4\Delta}{r^3} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times (R_l^{(1)} + P^{(1)*}), \tag{B15}$$

and from Eq. (B13),

$$\frac{3Mr - 4Q^2}{3Mr - 2Q^2}\frac{\sqrt{(l-1)l(l+1)(l+2)}}{2r^2}R_l^{(0)}$$
$$= \frac{1}{4}\left(\frac{\Delta}{r^2} + i\omega + \frac{2(r-M)}{r^2} + \frac{\Delta}{r^3} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times\left(\frac{\Delta}{r^2}\frac{d}{dr} + i\omega + \frac{4(r-M)}{r^2} - \frac{3\Delta}{r^3}\right)R_l^{(2)}$$
$$+ \frac{1}{4}\left(\frac{d}{r^2}\frac{d}{dr} - i\omega + \frac{2(r-M)}{r^2} + \frac{\Delta}{r^3} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{4(r-M)}{r^2} - \frac{3\Delta}{r^3}\right)P_l^{(-2)*}$$
$$+ \frac{Q\sqrt{(l-1)(l+2)}r}{\sqrt{2}(3Mr - 2Q^2)}\left(\frac{\Delta}{r^2}\frac{d}{dr} + i\omega + \frac{2(r-M)}{r^2} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)R_l^{(1)}$$
$$+ \frac{Q\sqrt{(l-1)(l+2)}r}{\sqrt{2}(3Mr - 2Q^2)}\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{2(r-M)}{r^2} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)P_l^{(-1)*}, \tag{B16}$$

$$\frac{3Mr - 4Q^2}{3Mr - 2Q^2}\frac{\sqrt{(l-1)l(l+1)(l+2)}}{2r^2}R_l^{(0)}$$
$$= \frac{1}{4}\left(\frac{\Delta}{r^2}\frac{d}{dr} + i\omega + \frac{2(r-M)}{r^2} + \frac{\Delta}{r^3} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times\left(\frac{\Delta}{r^2}\frac{d}{dr} + i\omega + \frac{4(r-M)}{r^2} - \frac{3\Delta}{r^3}\right)P_l^{(2)*}$$
$$+ \frac{1}{4}\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{2(r-M)}{r^2}\ \frac{\Delta}{r^3}\ \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)$$
$$\times\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{4(r-M)}{r^2} - \frac{3\Delta}{r^3}\right)R_l^{(-2)}$$
$$+ \frac{Q\sqrt{(l-1)(l+2)}r}{\sqrt{2}(3Mr - 2Q^2)}\left(\frac{\Delta}{r^2}\frac{d}{dr} + i\omega + \frac{2(r-M)}{r^2} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)P_l^{(1)*}$$
$$+ \frac{Q\sqrt{(l-1)(l+2)}}{\sqrt{2}(3Mr - 2Q^2)}\left(\frac{\Delta}{r^2}\frac{d}{dr} - i\omega + \frac{2(r-M)}{r^2} - \frac{4Q^2\Delta}{r^3(3Mr - 4Q^2)}\right)R_l^{(-1)}. \tag{B17}$$

From these equations to the final results is a very long but straightforward calculation. One gets a set of six equations by taking the first two highest terms with $\exp(-i\omega r')$ in each of the three equations and another set of six equations by taking the two highest terms with $\exp(i\omega r')$.

On the event horizon one set is automatically eliminated by the boundary condition that the group velocity of the wave be negative.

The final results are given in Sec. III [Eqs. (3.11), (3.12), and (3.13)].

[1]R. Ruffini and A. Treves, Astrophys. Lett. 13, 109 (1973).
[2]R. Wald, Phys. Rev. D 6, 1680 (1974).
[3]E. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[4]R. Boyer and R. Lindquist, J. Math. Phys. 8, 265 (1967).
[5]S. A. Teukolsky, Astrophys. J. 185, 635 (1973).
[6]S. K. Bose, J. Math. Phys. 16, 772 (1975).
[7]F. A. E. Pirani, *Lectures on General Relativity*, *Brandeis Summer Institute in Theoretical Physics* (Prentice-Hall, Englewood Cliffs, N. J., 1964).
[8]E. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).
[9]J. N. Goldberg, A. J. Macfarlane, E. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. 8, 2155 (1967).
[10]For an uncharged background metric case see S. A. Teukolsky and W. H. Press, Astrophys. J. 193, 443 (1974).
[11]V. Moncrief, Phys. Rev. D 9, 2707 (1974). V. Moncrief, Phys. Rev. D 10, 1057 (1974).
[12]D. M. Chitre, Phys. Rev. D 11, 760 (1975).
[13]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

A new functional equation in the plasma inverse problem and its analytic properties*

H. H. Szu,[a†] C. E. Carroll[a, b], C. C. Yang,[a] and S. Ahn[a, b]

[a]*Applied Mathematics Staff, Naval Research Laboratory, Washington, D.C. 20375*
and [b]*Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19174*
(Received 19 January 1976)

In the one-dimensional form of the plasma inverse problem, reflection of transverse electromagnetic waves is used to determine the electron density in a cold, collisionless, unmagnetized plasma. We extend the applicable Gel'fand–Levitan integral equation so that it is valid for all times. Laplace transformation of the extended equation gives a linear functional equation containing the complex reflection coefficient. We solve the functional equation analytically in special cases, and classify reflection coefficients by their analytic properties.

I. INTRODUCTION

Recent work by Case and Kac[1] and by Dyson[2] promises to cause a revival of interest in the one-dimensional form of the inverse scattering problem. This problem was solved some years ago by Gel'fand and Levitan,[3] and Faddeev[4] gave a comprehensive review of the Gel'fand—Levitan and related techniques. In this paper, we shall study the one-dimensional form of the plasma inverse problem,[5] in which the density distribution in a plasma is to be determined by the reflection of transverse electromagnetic waves. The electric field in a cold, collisionless, unmagnetized plasma obeys a partial differential equation in which the plasma density appears as the nonconstant coefficient. Kay[6] and Balanis[7] have proposed this model as a subject for mathematical study, and also for possible application to radar studies of the ionosphere.

Thus we study a stratified plasma whose electron density is $N(x)$. We assume

$$N(x)=0 \quad \text{for } x<0, \tag{1}$$

and seek to determine $N(x)$ for $x>0$. Solution of the Gel'fand—Levitan integral equation gives a function from which $N(x)$ can easily be obtained. But solution of the Gel'fand—Levitan equation by iteration of the kernel does not always converge, and is unsuitable for analytical work on the plasma inverse problem. The convergence of the iteration scheme is considered in the Appendix. In Sec. II, we extend the Gel'fand—Levitan equation to make it applicable for all times. Laplace transforms are introduced in Sec. III, and used to derive our linear functional equation. $A(s)$ is the Laplace transform of the reflection, or the complex reflection coefficient. It appears in our functional equation, which is linear in two unknown functions. The asymptotic form of either unknown function determines $N(x)$, the plasma density. In Sec. IV, this method will be applied to a simple example.

The Laplace transform $A(s)$ is a function of s, the complex variable. Various kinds of singularities are possible; they can be used to classify these reflection coefficients, and can be related to the behavior of $N(x)$ as $x\to+\infty$. Branch points of $A(s)$ are considered in Sec. V, where the complex transmission coefficient is introduced and a simple example is solved analytically. If $N(x)$ decreases exponentially as $x\to+\infty$, then $A(s)$ can have branch points; but if it decreases faster than any exponential, then $A(s)$ must be meromorphic. In Sec. VI, we apply the Nevanlinna theory[8] to this meromorphic function. Rational functions $A(s)$ are treated in Sec. VII. Kay has given a general solution of the inverse problem for this case[9]; but many rational functions, such as the Butterworth functions,[10] appear as inconvenient limits of Kay's solution. For this reason, we shall solve the plasma inverse problem for the case in which $A(s)$ is a Butterworth function. The application of our results to studies of the ionosphere is briefly discussed in Sec. VIII, the conclusion.

II. EXTENSION OF GEL'FAND–LEVITAN EQUATION

In this section, we write the wave equation for transverse electromagnetic waves in a cold, collisionless, unmagnetized plasma. The one-dimensional equation has a symmetry between space and time, and allows a spacelike solution, as well as the retarded or physical or timelike solution. This spacelike solution is closely related to the Marchenko function, which is determined by the Gel'fand—Levitan integral equation. We shall combine the spacelike and timelike solutions, to obtain an extended integral equation, which is valid for all time and is amenable to Laplace transformation.

The electromagnetic waves propagate in the $\pm x$ directions. We assume that no external magnetic field is applied; and a collisionless plasma can produce no change in the polarization of an incident electromagnetic wave. The electric field is $E(x,t)$, and we may assume that it is always parallel to the z axis. Since the electron density depends only on x, we have no separation of charges in the plasma; this means that

$$\nabla\cdot E=\nabla\cdot j=0, \tag{2}$$

where j is the current density. We neglect reflection of electromagnetic waves by the ions; this means that j is the electron current. For a cold, collisionless plasma we have

$$j=(e^2/m)N(x)\int_{-\infty}^{t}E(x,t')\,dt', \tag{3}$$

where e and m are the charge and mass of an electron. We seek to determine $N(x)$ from the knowledge of the incident and reflected waves. Because of (1), we must have

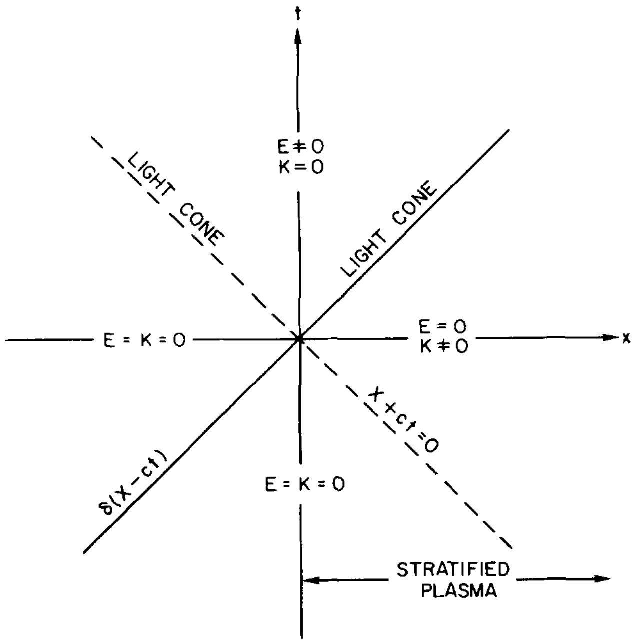


FIG. 1. The light cone divides the $x-t$ plane into four quadrants. We consider a timelike and a spacelike solution of the wave equation; each solution is the sum of $\delta(x-ct)$ and a bounded part.

$$E(x,t)=I(x-ct)+R_I(x+ct) \quad \text{when } x\leq 0. \tag{4}$$

Here c is the speed of light. Kay[9] and Balanis[5] introduce a simplifying assumption at this point; they write

$$E(x,t)=\delta(x-ct)+R(x+ct) \quad \text{when } x\leq 0. \tag{5}$$

The general form (4) can be recovered by a process of superposition, because (3) and Maxwell's equations are linear in $E(x,t)$. The function $R(y)$ is defined by (5). From causality and assumption (1), $R(y)=0$ for $y<0$; this is a statement that the reflection from the plasma is found inside the forward light cone (Fig. 1). The plasma inverse problem is the determination of $N(x)$ from $R(y)$, the reflection. To solve it, we use a reflectionless solution which vanishes inside the forward light cone, but not outside (Fig. 1).

In either case, the electric field must satisfy a partial differential equation which is derived from Maxwell's equations. Let

$$q(x)=4\pi r_e N(x), \tag{6}$$

where $r_e=e^2/mc^2$ is the classical electron radius. Then (2) and (3) lead to

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}\right)E(x,t)=q(x)E(x,t). \tag{7}$$

Here and henceforth we put $c=1$. Incidentally, we obtain plasma oscillations of long wavelength by considering the case in which $N(x)$ is nearly constant and $\partial^2/\partial x^2$ is negligible. This means that $q(x)$ is the square of the plasma frequency.

The retarded electric field, $E(x,t)$, satisfies (5) and (7). It can be written as

$$E(x,t)=\delta(x-t)+R(x+t)$$
$$+\int_{\max(-x,-t)}^{x}K(x,y)[\delta(y-t)+R(y+t)]\,dy. \tag{8}$$

The Marchenko function $K(x,y)$ appears here as a kernel which extends the known solution (5) into the interior of the plasma. In the case (4), we have

$$E(x,t)=I(x-t)+R_I(x+t)$$
$$+\int_{-x}^{x}K(x,y)[I(y-t)+R_I(y+t)]\,dy.$$

Here $R_I(x,t)$ is a function whose shape depends on that of $I(x-t)$, but $K(x,y)$ is the same function as in (8). In fact $K(x,y)$ is determined by $N(x)$ and vice versa. To see this connection, we define $K(x,y)$ as follows. We ask for a reflectionless, spacelike solution of (7). This solution is

$$\delta(x-t)+K(x,t). \tag{9}$$

It must satisfy

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}-q(x)\right)[\delta(x-t)+K(x,t)]=0. \tag{10}$$

This solution is reflectionless because we require

$$K(x,t)=0 \quad \text{when } x\leq 0. \tag{11}$$

It is called spacelike because it vanishes inside the light cone (Fig. 1); we require

$$K(x,t)=0 \quad \text{when } |t|>x>0. \tag{12}$$

The singular part of this solution is the δ function which appears explicitly in (9). We demand that $K(x,t)$ is bounded everywhere, and we can use (6) and (10) to construct $K(x,t)$ if $N(x)$ is known. A similar separation into a δ function and a bounded part is possible in the timelike solution, which satisfies (5) and (7). We can construct $E(x,t)$ if $N(x)$ is known, and we require that $E(x,t)-\delta(x,t)$ be bounded everywhere.

If either the spacelike or the timelike solution is known, we can recover (6) and the plasma density. Equation (10) can be written in the form

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}-q(x)\right)K(x,t)=q(x)\delta(x-t), \tag{13}$$

which states that the first partial derivatives of $K(x,t)$ are discontinuous at $x=t$. We find that

$$\frac{\partial}{\partial x}K(x,t)=\frac{\partial}{\partial t}K(x,t)=q(x),$$

where the partial derivatives are evaluated at $t=x^-$, just outside the light cone (Fig. 1). This result can be written as

$$q(x)=2\frac{d}{dx}K(x,x), \tag{14}$$

where the limit $t\to x^-$ is implicit. Also, (13) tells us that $K(x,t)$ and its first partial derivatives vanish at $t=-x$. Similar considerations apply to the timelike solution. We find

$$q(x)=-2\frac{d}{dx}\lim_{t\to x^+}[E(x,t)-\delta(x,t)]. \tag{15}$$

If $t<x$, the retarded electric field must vanish, and (8) gives

$$R(x+t)+K(x,t)+\int_{-t}^{x}K(x,y)R(y+t)\,dy=0, \tag{16}$$

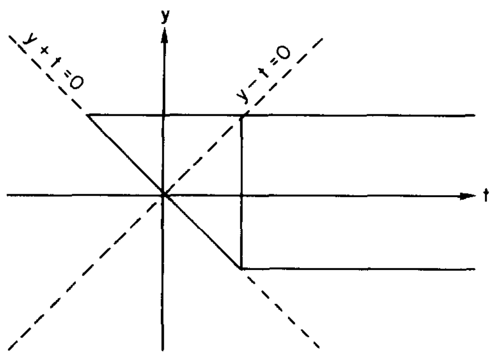


FIG. 2. Domains of integration for Gel'fand—Levitan equation and extended Gel'fand—Levitan equation are plotted as functions of t. For the Gel'fand—Levitan equation, $t<x$ and the domain appears as a triangle. For the extended Gel'fand—Levitan equation, all values of t are allowed, and the domain appears as a semi-infinite trapezoid.

which is the Gel'fand—Levitan integral equation. The range of integration appears as a triangle in Fig. 2. The solution of (16) is unique; this can be shown because $q(x)$ is nonnegative. See Faddeev[4] for a comprehensive review of the Marchenko function and the Gel'fand—Levitan integral equation. If (16) can be solved for $K(x,t)$, we obtain $N(x)$ from (6) and (14). Integral equations such as (16) are often solved by iterating the kernel. This means that we set

$$K(x,t)=\sum_{n=0}^{\infty}K_n(x,t), \tag{17}$$

where

$$K_0(x,t)=-R(x+t) \tag{18a}$$

and

$$K_n(x,t)=-\int_{-t}^{x}K_{n-1}(x,y)R(y+t)\,dy \tag{18b}$$

for $n\geq 1$. The function $R(y)$ is bounded, and we may assume that

$$\int_0^{\infty}|R(y)|\,dy<\infty, \tag{19}$$

but we are not assured that (17) converges. In the Appendix we show that (17) converges when

$$0\leq |t|<x<\pi/4\max|R(y)|. \tag{20}$$

Our method of solution uses (8) directly, without the assumption that $t<x$. We introduce the "entire" electric field

$$\mathcal{E}(x,t)=K(x,t)+\delta(x-t)-E(x,t). \tag{21}$$

This is a bounded function which vanishes when $x+t<0$; see Fig. 1. Indeed, the considerations leading to (14) and (15) show that (21) is a continuous function of t near $t=x$, and that

$$q(x)=2\frac{d}{dx}\mathcal{E}(x,x). \tag{22}$$

This function incorporates the solution of the direct scattering problem, for $E(x,t)=-\mathcal{E}(x,t)$ when $t>x$. Also, this function satisfies the integral equation

$$R(x+t)+\mathcal{E}(x,t)+\int_{\max(-x,-t)}^{x}\mathcal{E}(x,y)R(y+t)\,dy=0, \tag{23}$$

which is valid for all values of t. The range of integration is plotted in Fig. 2.

The plasma density is independent of time, and we can eliminate t from (23) by taking Laplace transforms.

III. LAPLACE TRANSFORMS

In this section, we reformulate the plasma inverse problem in terms of Laplace transforms and the complex frequency variable. The Laplace transform of $R(y)$ is

$$A(s)=\int_0^{\infty}R(y)\exp(-sy)\,dy,$$

and we classify reflection coefficients according to the analytic properties of $A(s)$. We shall obtain a functional equation that determines two unknown functions in terms of $A(s)$, and the asymptotic form of either function will give $\mathcal{E}(x,x)$. Then we use (6) and (22) to complete the solution of the plasma inverse problem. Incidentally, Laplace transforms will give us a derivation of (8) from (9) and (10), thus closing a gap in Sec. II. The Laplace transform of the retarded electric field is

$$\hat{E}(x,s)=\int_{-\infty}^{\infty}E(x,t)\exp(-st)\,dt.$$

The integral converges when the real part of s is positive, because $E(x,t)$ vanishes in the backward light cone (Fig. 1). Then (5) becomes

$$\hat{E}(x,s)=\exp(-sx)+A(s)\exp(sx)\quad\text{when } x\leq 0. \tag{24}$$

Since $R(y)$ appears in (5) as the reflected part of the electric field, we can assume certain smoothness properties. We assume that $R(y)$ and $R'(y)$ belong to the class $L^1(-\infty,+\infty)$, and we have assumed that $|R(y)|$ is a bounded function satisfying (19). With these assumptions, we can show that $|A(s)|<\epsilon$ whenever $|s|$ is sufficiently large and the real part of s is nonnegative. We choose M such that $\int_M^{\infty}|R(y)|\,dy<\frac{1}{2}\epsilon$, and then use partial integration to estimate the integral from 0 to M. In this way, we prove that $A(s)\to 0$ uniformly as $s\to\infty$ in the closed right half of the complex plane.

The time like electric field contains a δ function and a bounded part. We define

$$B(x,s)=\int_{-\infty}^{\infty}[E(x,t)-\delta(x-t)]\exp(-st)\,dt. \tag{25}$$

Then $B(x,s)\exp(sx)\to 0$ uniformly as $s\to\infty$ in the closed right half of the complex plane, by the same reasoning that we applied to $A(s)$. Since $E(x,t)-\delta(x-t)$ is bounded, we have

$$|B(x,s)|\leq(\text{const})\exp(-\sigma t)/\sigma\quad\text{for }\sigma>0, \tag{26}$$

where σ is the real part of s.

The timelike and spacelike electric fields satisfy (7) and (10). Laplace transformation gives

$$\left(\frac{\partial^2}{\partial x^2}-s^2-q(x)\right)\hat{E}(x,s)=0 \tag{27}$$

and

$$\left(\frac{\partial^2}{\partial x^2}-s^2-q(x)\right)[\exp(-sx)+F(x,s)]=0, \tag{28}$$

where

$$F(x,s)=\int_{-x}^{x}K(x,t)\exp(-st)\,dt. \tag{29}$$

The integration in (29) extends over the region where

$K(x,t)$ need not vanish; see (12). $F(x,s)=0$ when $x\leq 0$, because of (11). The finite range of integration in (29) implies that $F(x,s)$ is an entire function of s. Since $K(x,t)$ is bounded, we have

$$|F(x,s)|\leq (\text{const})(\sinh x\sigma)/\sigma. \qquad (30)$$

We can now write useful relations among these Laplace transforms. A change of sign in (28) gives

$$\left(\frac{\partial^2}{\partial x^2}-s^2-q(x)\right)[\exp(sx)+F(x,-s)]=0. \qquad (31)$$

Equation (29) shows that this change of sign corresponds to time reversal. According to (27), (28), and (31), we now have three solutions to an ordinary, second-order differential equation. There must be a linear relation among them. If $s\neq 0$, $\exp(-sx)+F(x,s)$ and $\exp(sx)+F(x,-s)$ are independent solutions, and it is possible to write $\hat{E}(x,s)$ as a linear combination of them. The coefficients may depend on s. Since $F(x,s)=0$ for $x\leq 0$, we can use (24) to determine the coefficients; we find

$$\hat{E}(x,s)=\exp(-sx)+F(x,s)+A(s)[\exp(sx)+F(x,-s)].$$

This equation is the Laplace transform of (8), which has thus been derived from (9) and (10). By continuity, it holds when $s=0$. Since $\hat{E}(x,s)=\exp(-sx)+B(x,s)$, we obtain

$$A(s)[\exp(sx)+F(x,-s)]+F(x,s)-B(x,s)=0, \qquad (32)$$

our functional equation for $B(x,s)$ and $F(x,s)$. It should have been derived directly from (23), using

$$F(x,s)-B(x,s)=\int_{-\infty}^{\infty}\mathcal{E}(x,t)\exp(-st)\,dt.$$

If $A(s)$ can be continued analytically into the left half of the complex plane, (32) can be solved by analytic methods, using the growth conditions (26) and (30).

After solving for $B(x,s)$ and $F(x,s)$, we can easily recover (21). If $\partial K/\partial t$ is bounded for small positive values of $x-t$, then partial integration of (29) gives

$$F(x,s)\sim -\mathcal{E}(x,x)\exp(-sx)/s \quad \text{as } s\to -\infty.$$

Alternatively, we can integrate (25) by parts. We have

$$\lim_{s\to+\infty} s\exp(sx)B(x,s)=\lim_{s\to-\infty} s\exp(sx)F(x,s)=-\mathcal{E}(x,x). \qquad (33)$$

Finally, $N(x)$ is found from (6) and (22).

IV. A SIMPLE EXAMPLE

As an example of analytic solution of the plasma inverse problem, we consider the case

$$R(y)=-\lambda H(y-2\alpha)\exp[-\lambda(y-2\alpha)].$$

Here α and λ are positive constants, and $H(x)$ is the Heaviside step function.

$$A(s)=-[\lambda/(s+\lambda)]\exp(-2\alpha s) \qquad (34)$$

is a meromorphic function of s. The functional equation (32) gives

$$-\lambda[\exp(sx)+F(x,-s)]\exp(-2\alpha s)+(s+\lambda)F(x,s)$$
$$=(s+\lambda)B(x,s). \qquad (35)$$

Since $F(x,s)$ is an entire function of s, $(s+\lambda)B(x,s)$ must be an entire function of s.

If $x\leq\alpha$, we write (35) as

$$(s+\lambda)F(x,s)\exp(sx)$$
$$=\lambda\exp(2sx)\exp(-2\alpha s)+\lambda[F(x,-s)\exp(-sx)]$$
$$\times\exp(2sx)\exp(-2\alpha s)+(s+\lambda)B(x,s)\exp(sx).$$

Equation (30) implies that the entire function on the left is bounded by a linear function of s if $s\to\infty$ while σ (the real part of s) is constant. Also, the function on the left is bounded as $\sigma\to-\infty$. From (30) and (26), the entire function on the right is bounded as $\sigma\to+\infty$. Then Liouville's theorem on entire functions requires $(s+\lambda)F(x,s)\exp(sx)$ to be a linear function of s. Since it is bounded as $\sigma\to-\infty$, $(s+\lambda)F(x,s)\exp(sx)$ must be a constant. Since $F(x,s)$ has no pole at $s=-\lambda$, $(s+\lambda)\times F(x,s)\exp(sx)$ must vanish identically. Therefore, $F(x,s)=0$ and $B(x,s)=A(s)\exp(sx)$. We learn that $\mathcal{E}(x,x)=0$ if $x<\alpha$.

If $x>\alpha$, we have

$$(s+\lambda)F(x,s)$$
$$=\lambda\exp(sx)\exp(-2\alpha s)+\lambda F(x,-s)\exp(-2\alpha s)$$
$$+(s+\lambda)B(x,s).$$

We use (30) and (26) to obtain a bound for the right-hand side as $\sigma\to+\infty$; it is dominated by a multiple of the first term. We have

$$|F(x,s)|\leq(\text{const}/\sigma)\exp[\sigma(x-2\alpha)] \quad \text{when } \sigma>0.$$

We now write (35) as

$$-\lambda\exp(2sx)\exp(-2\alpha s)-\lambda F(x,-s)\exp(sx)\exp(-2\alpha s)$$
$$+(s+\lambda)F(x,s)\exp(sx)=(s+\lambda)B(x,s)\exp(sx).$$

The function on the left is bounded as $\sigma\to-\infty$. The entire function on the right is bounded by a linear function of s if $s\to\infty$ while σ is constant, and it is bounded as $\sigma\to+\infty$. Liouville's theorem requires $(s+\lambda)B(x,s)\times\exp(sx)$ to be a linear function of s. Since it is bounded as $\sigma\to+\infty$, it must be equal to C, a constant. We have

$$B(x,s)=[C/(s+\lambda)]\exp(-sx)$$

and

$$(s+\lambda)F(x,s)\exp(\alpha s)-\lambda F(x,-s)\exp(-\alpha s)$$
$$=\lambda\exp(sx)\exp(-\alpha s)+C\exp(-sx)\exp(\alpha s). \qquad (36)$$

Let us replace s by $-s$ and add the resulting equation to (36). We obtain

$$s[F(x,s)\exp(\alpha s)-F(x,-s)\exp(-\alpha s)]$$
$$=2(\lambda+C)\cosh(sx-\alpha s).$$

Since $F(x,s)$ is an entire function of s, this quantity must vanish at $s=0$. Therefore, $C=-\lambda$, and $F(x,s)\exp(\alpha s)-F(x,-s)\exp(-\alpha s)$ must vanish identically. We obtain

$$B(x,s)=-[\lambda/(s+\lambda)]\exp(-sx)$$

and

$$F(x,s)=2\lambda[\sinh(sx-\alpha s)/s]\exp(-\alpha s).$$

Then (33) gives $\mathcal{E}(x,x)=\lambda$.

We conclude that

$$\mathcal{E}(x,x)=\lambda H(x-\alpha),$$

and hence

$$q(x)=2\lambda\delta(x-\alpha)$$

is the unique solution of this inverse problem.

Note that $\mathcal{E}(x,x)$ is a discontinuous function of x, and that $K(x,t)$ is a discontinuous function of t. The partial integrations leading to (33) are not justified at $x=\alpha$. Our method of obtaining $\mathcal{E}(x,x)$ fails at discontinuities of this function, but we may refrain from evaluating $\mathcal{E}(x,x)$ at such points.

If we simplify the foregoing example by letting $\alpha \to 0^+$, then $K(x,t)$ is not continuous at $t=-x$; but this causes no difficulty in our calculations.

V. BRANCH POINTS

The function $A(s)$ was defined as a Laplace transform, and it must be analytic in the right half of the complex plane. In many cases, it can be continued analytically into the left half of the complex plane, and only isolated singularities are encountered. If there are no such singularities, $A(s)$ is an entire function of s; but we conjecture that this occurs only in the trivial case $A(s)=q(x)=0$. In this section, we present examples to show that $A(s)$ can have branch points, and we relate the analytic properties of $A(s)$, the reflection coefficient, to those of the transmission coefficient and the Jost function.

If $q(x)$ approaches a positive limit as $x\to+\infty$, then $A(s)$ can have a pair of branch points on the imaginary axis. For example, let

$$A(s)=\frac{s(r-s+b)-(b+c)r-(a^2+b^2+bc)}{s(r+s+b)+(b+c)r+(a^2+b^2+bc)}, \tag{37}$$

where $r=(s^2+a^2)^{1/2}$. We assume that a, b, and c are positive constants, and assert that (37) is analytic in the right half-plane. However, (37) has branch points at $s=\pm ia$. To define (37) in the left half-plane, we draw a straight branch line from $s=-ia$ to $s=ia$, and define r as an odd function of s. We can now solve (32). We replace $F(x,s)$ by two other entire functions:

$$F_e(x,s)=\tfrac{1}{2}[F(x,s)+F(x,-s)]+\cosh sx$$

is an even entire function, and

$$F_0(x,s)=\tfrac{1}{2}[F(x,s)-F(x,-s)]-\sinh sx$$

is an odd entire function. Then (32) becomes

$$F_e(x,s)+\left(\frac{1-A(s)}{1+A(s)}\right)F_0(x,s)=\frac{B(x,s)+\exp(-sx)}{1+A(s)}.$$

Multiplication of this equation by $r+b$ gives

$$(r+b)F_e(x,s)+[s^2+(b+c)r+a^2+b^2+bc]\,F_0(x,s)/s$$
$$=G(x,s),$$

where $G(x,s)$ is an unknown function; it must satisfy

$$|G(x,s)|<(\text{const})|s|\exp(-\sigma x) \tag{38}$$

when σ is positive and $|s|$ is sufficiently large. The bound (30) implies that (38) is also valid when $\sigma\le 0$ and $|s|$ is sufficiently large. In terms of the entire functions

$$\psi(s)=bF_e(x,s)+(s^2+a^2+b^2+bc)F_0(x,s)/s$$

and

$$\phi(s)=F_e(x,s)+(b+c)F_0(x,s)/s,$$

we have

$$\psi(s)+r\phi(s)=G(x,s).$$

Since $\psi(s)$ and $\phi(s)$ are even entire functions,

$$\psi(s)-r\phi(s)=G(x,-s).$$

The product of these two equations is

$$\psi^2-(s^2+a^2)\phi^2=G(x,s)G(x,-s), \tag{39}$$

a quadratic functional equation. The right-hand side must be an entire function, and the bound (38) implies that it is a polynomial. Hence (39) is a functional equation of the type encountered by Lebowitz and Zomick.[11] Because of (38), the right-hand side of (39) must be an even quadratic function of s. Also, the bound (30) implies that $\psi(s)$ and $\phi(s)$ are entire functions of exponential type; they must satisfy

$$\psi(s)\sim -s\sinh sx$$

and

$$\phi(s)\sim\cosh sx$$

as $s\to\pm\infty$. The theory of Gross, Osgood, and Yang[12] now tells us that the solution of (39) is unique, apart from two constants that we proceed to determine. We rewrite (39) as

$$\psi^2-(s^2+a^2)\phi^2=A^2(s^2+a^2)+B^2, \tag{40}$$

where A and B are constants to be determined; they may depend on x. To solve (40), we write it as

$$(B\psi+iAr^2\phi)^2+r^2(A\psi-iB\phi)^2=(A^2r^2+B^2)^2.$$

Here $B\psi+iAr^2\phi$ and $A\psi-iB\phi$ are even entire functions. The desired solution is

$$B\psi+iAr^2\phi=\pm(A^2r^2+B^2)\cos ixr \tag{41}$$

and

$$A\psi-iB\phi=\pm(A^2r^2+B^2)(\sin ixr)/r. \tag{42}$$

The known asymptotic forms of $\psi(s)$ and $\phi(s)$ fix the coefficient ix, and require the ambiguous sign in (42) to be the same as that in (41). Since the signs of A and B are yet undetermined, the ambiguous signs can be dropped. Then we have

$$\psi(s)=iAr\sinh rx+B\cosh rx$$

and

$$\phi(s)=-iA\cosh rx-B(\sinh rx)/r.$$

Since

$$F_0(x,s)/s=[\psi(s)-b\phi(s)]/r^2$$

must be an even entire function, we have $B=-iAb/(1+bx)$. Finally, to satisfy (30), we choose $A=i$. Then

$$F(x,s)=\cosh rx+(b+c-s)\frac{\sinh rx}{r}-\frac{b\sinh rx}{r(1+bx)}$$

$$+b^2(b+c-s)\frac{rx\cosh rx-\sinh rx}{r^3(1+bx)}-\exp(-sx).$$

From this function, we find

$$\mathcal{E}(x,x)=\frac{a^2x}{2}+\frac{b^2x}{1+bx}+\frac{c}{2};$$

hence

$$q(x)=\left(a^2+\frac{2b^2}{(1+bx)^2}\right)H(x)+c\delta(x)$$

is the unique solution of this inverse problem.

In this example, $q(x)\to a^2$ as $x\to+\infty$, and $A(s)$ has branch points at $s=\pm ia$. Branch points at $s=\pm ia$ will appear if $|q(x)-a^2|$ decreases exponentially as $x\to+\infty$. In this case, we can demand

$$\hat{E}(x,s)\sim T(s)\exp(-rx) \tag{43}$$

as $x\to+\infty$. This condition introduces the transmission coefficient coefficient $T(s)$. The boundary conditions (24) and (43) serve to select the desired solution of the differential equation (27). It is often convenient to divide this solution by $T(s)$, and to consider a function $f(x,s)$ which satisfies (27) and

$$f(x,s)\sim\exp(-rx) \tag{44}$$

as $x\to+\infty$. This is essentially the Jost function.[13] It satisfies a convenient integral equation:

$$f(x,s)=\exp(-rx)-\frac{1}{r}\int_x^\infty\{\sinh[r(x-y)]\}$$
$$\times[q(y)-a^2]f(y,s)\,dy.$$

To show that $f(x,s)$ is analytic when the real part of r is positive, we use part (a) of the method of Bargmann[14]; this involves the uniform convergence of an infinite series of analytic function. Since $|q(x)-a^2|$ decreases exponentially as $x\to+\infty$, part (b) of the method of Bargmann shows that $f(x,s)$ is analytic when r is sufficiently close to the imaginary axis. In this way, we show that $f(x,s)$ is an analytic function of r in the neighborhood of $r=0$; hence $f(x,s)$ has branch points at $s=\pm ia$. We can also show that $\partial f/\partial x$ has branch points at $s=\pm ia$. To recover $A(s)$ and $T(s)$, we note that (24) implies

$$\left[sf(x,s)+\frac{\partial f}{\partial x}\right]_{x=0}=2s\frac{A(s)}{T(s)} \tag{45a}$$

and

$$\left[sf(x,s)-\frac{\partial f}{\partial x}\right]_{x=0}=\frac{2s}{T(s)}. \tag{45b}$$

This means that $A(s)$, or both, have branch points at $s=\pm ia$. We do not prove that $A(s)$ itself must have branch points at $s=\pm ia$.

Such a pair of branch points is not expected if $xq(x)\to 0$ as $x\to+\infty$. In this case we replace (43) and (44) by

$$\hat{E}(x,s)\sim T(s)\exp(-sx) \tag{46}$$

and

$$f(x,s)\sim\exp(sx). \tag{47}$$

The conditions (24) and (46) now serve as boundary conditions for the differential equation (27). At $s=0$, these conditions give $\partial\hat{E}/\partial x=0$ for $x<0$ and $\partial\hat{E}/\partial x\to 0$ as $x\to+\infty$. Because $q(x)$ is nonnegative, these conditions give $\hat{E}(x,0)=0$ for all x, unless $q(x)$ vanishes identically. Therefore, we have

$$1+A(0)=T(0)=0 \tag{48}$$

unless $q(x)$ vanishes identically.

Assuming that $\int_0^\infty xq(x)\,dx$ exists, the method of Bargmann[14] shows that $f(x,s)$ is analytic in the right half-plane. If $f(x,s)$ can be continued analytically into the left half-plane, then (45) gives $A(s)$ and $T(s)$ there. The differential equation (27) is satisfied by both $f(x,s)$ and $f(x,-s)$, and we may compute the Wronskian of these two solutions. Using (45) and (47), we obtain

$$\frac{2s}{T(s)T(-s)}-2s\frac{A(s)A(-s)}{T(s)T(-s)}=2s.$$

Therefore, we have

$$1-A(s)A(-s)=T(s)T(-s) \tag{49}$$

if $s\neq 0$ and all quantities in the equation are defined. By (48), this equation holds also when $s=0$. If s^2 is real and negative, (49) can be derived from the conservation of energy for electromagnetic waves of frequency is. This derivation involves the absolute squares of $A(s)$ and $T(s)$; these absolute squares must be bounded when s^2 is real and negative.

The assumption that $q(x)$ decreases exponentially as $x\to+\infty$ is not sufficient to exclude branch points of $A(s)$. Suppose that

$$q(x)=(x+a)^{-1}\exp(-bx)\quad\text{when } x>0,$$

where a and b are positive constants. An argument devised by Regge[15] applies to this example, and it shows that $f(x,s)$ has a branch point at $s=-\frac{1}{2}b$. Also, $\partial f/\partial x$ has a branch point at $s=-\frac{1}{2}b$. Then either $A(s)$ or $T(s)$ must have a branch point at $s=-\frac{1}{2}b$. From (45), both $A(s)$ and $T(s)$ are meromorphic in a neighborhood of $s=+\frac{1}{2}b$. Then (49) requires that both $A(s)$ and $T(s)$ have branch points at $s=-\frac{1}{2}b$.

VI. MEROMORPHIC FUNCTIONS

With stronger assumptions about the asymptotic form of $q(x)$, we can show that $A(s)$ and $T(s)$ are meromorphic functions. Furthermore, the Nevanlinna theory[8] can be used to classify these functions and to establish a connection between $A(s)$, $T(s)$, and the asymptotic density of their poles. In this section, we start with the assumption that

$$q(x)\exp(Nx)\to 0\quad\text{as } x\to+\infty, \tag{50}$$

where N is any real number. This will be sufficient to show that $s/T(s)$ is an entire function of s. Later we shall strengthen the assumption (50), and show that $s/T(s)$ is an entire function of finite order, in the Nevanlinna scheme as well as in the usual classification of entire functions.[16,17] The Nevanlinna theory has not been previously applied in plasma theory, although Sartori[18] has considered entire functions of various orders in connection with scattering theory.

Using the condition (47), we find that the Jost function satisfies an integral equation:

$$f(x,s)=\exp(-sx)-\frac{1}{s}\int_x^\infty\{\sinh[s(x-y)]\}\times q(y)f(y,s)\,dy.$$

The Jost function can now be found from a series, similar to (17), used by Jost.[13] We write

$$f(x,s)=\sum_{n=0}^\infty f_n(x,s), \tag{51}$$

where $f_0(x,s)=\exp(-sx)$ and

$$f_n(x,s)=-\frac{1}{s}\int_x^\infty\{\sinh[s(x-y)]\}q(y)f_{n-1}(y,s)\,dy$$

for $n\geq 1$. The series converges uniformly in any finite region of the s plane, and each term is an entire function of s. This shows that $f(x,s)$ and $\partial f/\partial x$ are entire functions of s. From (45), we see that $A(s)$ and $T(s)$ are meromorphic functions, and that $s/T(s)$ is an entire function.

Entire functions are classified by their order and type,[16,17] which guage the growth of the absolute value as $|s|\to\infty$. We can now show that the order of $s/T(s)$ is ≥ 1, unless

$$T(s)=s/(s+\lambda) \tag{52}$$

and $A(s)$ has the form (34). To prove this, we assume that the order of $s/T(s)$ is <1. This means that

$$\left|\frac{s}{T(s)}\frac{1}{s-1}\right|<a\exp(b|s|^c) \tag{53}$$

when σ (the real part of s) is negative and $|s|$ is sufficiently large. Here a, b, and c are positive constants, and $c<1$. When $\sigma>0$, $B(x,s)\exp(sx)\to 0$ uniformly as $s\to\infty$, as we noted in Sec. III. Since this holds for all positive values of x, $T(s)-1\to 0$ uniformly as $s\to\infty$ in the closed right half-plane. In particular, $T(s)\to 1$ as $s\to\infty$ along the imaginary axis. We can now apply the Phragmén–Lindelöf principle,[16] to show that the left-hand side of (53) is bounded when $\sigma<0$. Hence

$$\left|\frac{s}{T(s)}\right|<(\text{const})|s-1|$$

when $\sigma<0$. An inequality of this form also holds when $\sigma\geq 0$ and $|s|$ is sufficiently large. Using Liouville's theorem on entire functions, we can show that $T(s)$ must have the form (52), where λ is a constant. We also have to prove that $A(s)$ is given by (34). From (45a), $A(s)$ can have no poles except at $s=-\lambda$; hence we write

$$A(s)=-\lambda(s+\lambda)^{-1}f(s),$$

where $f(s)$ is an entire function. Since $A(s)$ is analytic in the right half-plane, and bounded when $\sigma=0$, the real part of λ is positive. From (49) and (52),

$$A(s)A(-s)=-\lambda^2/(s^2-\lambda^2) \tag{54}$$

is a function with no zeroes. Hence $f(s)$ is an entire function with no zeroes, and we can write

$$A(s)=-\lambda(s+\lambda)^{-1}\exp[g(s)], \tag{55}$$

where $g(s)$ is another entire function. From (48), $g(0)$ must be a multiple of $2\pi i$; we choose $g(0)=0$. Then (54) requires that $g(s)$ is an odd function. The real part of $g(s)$ is $h(s)$, an odd harmonic function. Since $A(s)\to 0$ as $s\to\infty$ in the right half-plane,

$$h(s)<\log|s|$$

when σ is positive and $|s|$ is sufficiently large. This inequality holds also when $\sigma=0$; indeed $\sigma=0$ implies

$$|h(s)|<\log|s|$$

when $|s|$ is sufficiently large, because $h(s)$ is an odd function. We can now appeal to Levin's theory of functions harmonic in a half-plane.[16] Lemma 2 gives

$$h(s)=\frac{\sigma}{\pi}\int_{-\infty}^\infty\frac{h(it)\,dt}{(t-\tau)^2+\sigma^2},$$

where τ is the imaginary part of s, and Lemma 3 gives

$$|h(s)|<(\text{const})|s|^2$$

when $|s|$ is sufficiently large. Since $h(s)$ is an odd function of s, it must be a linear function of s. Its conjugate harmonic function is also linear. Therefore, $g(s)$ is a linear function, and (55) reduces to the form (34). Since $A(s)$ must be real on the real axis, α and λ must be real; α cannot be negative and λ must be positive.

We have shown that $s/T(s)$ is an entire function function of order ≥ 1, except in the case of (34) and (52). We now strengthen the assumption (50). We assume that there are positive numbers δ and ϵ such that

$$q(x)<(\text{const})\exp(-\delta x^{1+\epsilon}) \tag{56}$$

when x is positive. We can now use Sartori's estimate[18] for terms in the sum (51). The integral

$$I_q=\int_0^\infty q(x)\,dx \tag{57}$$

exists, and

$$|f_n(x,s)|<\frac{I_q^{n-1}}{|s|^n}\exp(-|s|x)\int_0^\infty\exp(2|s|y)q(y)\,dy$$

when $n>0$ and $x\geq 0$. This can be proved by induction, and it implies

$$|f(x,s)|<\exp(|s|x)+\frac{\exp(-|s|x)}{|s|-I_q}\times\int_0^\infty\exp(2|s|y)q(y)\,dy$$

when $|s|>I_q$. A similar estimate for $|\partial f_n/\partial x|$ yields

$$\left|\frac{\partial}{\partial x}f(x,s)\right|<|s|\exp(|s|x)+\frac{|s|\exp(-|s|x)}{|s|-I_q}\times\int_0^\infty\exp(2|s|y)q(y)\,dy$$

when $|s|>I_q$. Because of the condition (56), the logarithm of $\int_0^\infty\exp(2|s|y)q(y)\,dy$ is bounded by a constant times $|s|^{(1+\epsilon)/\epsilon}$; this is shown by comparing the integral with $\int_0^\infty\exp(xz-\delta x^{1+\epsilon})\,dx$, and entire function of order $(1+\epsilon)/\epsilon$ and finite type.[18] We conclude that $f(x,s)$ and $\partial f/\partial x$ are entire functions of order $(1+\epsilon)/\epsilon$, at most. Furthermore, we can extend this proof to functions $q(x)$ such that (56) holds only at sufficiently large values of x; we then demand that (57) exists, and make minor changes in the proof. We find that $f(x,s)$,

$\partial f/\partial x$, $s/T(s)$, and $sA(s)/T(s)$ are entire functions of finite order.

In order to show that entire functions of any order greater than 1 can occur, one assumes that

$$q(x) = \exp(-x^p) \tag{58}$$

when $x > 0$ and that $p > 1$. Sartori shows that $f(x, s)$ is an entire function of order $p/(p-1)$, and a slight extension of his calculation shows that $\partial f/\partial x$ is an entire function of the same order. We can show that $sA(s)/T(s)$ is an entire function of order $p/(p-1)$, which can be any real number greater than 1.

In the remainder of this section, we assume that $s/T(s)$ and $sA(s)/T(s)$ are entire functions of finite order, and apply the Nevanlinna theory[8] to the meromorphic functions $A(s)$ and $T(s)$. In the Nevanlinna theory, the concept of order is generalized so that it is applicable to meromorphic functions as well as to entire functions; see the original paper[8] and Hayman's monograph.[19] Let $\rho(f)$ denote the Nevanlinna order of a function $f(s)$. The elementary properties of the Nevanlinna order give

$$\rho(s/T) = \rho(1/T) = \rho(T),$$

and the order of $s/T(s)$ is the same as was calculated above with the simple definition applicable to entire functions. We shall show that

$$\rho(A) = \rho(T) \geq 1, \tag{59}$$

except in the case of (34) and (52). The proof that $\rho(T) \geq 1$ has already been given. We can see that equality does not always hold in (59), because the example (58) shows that $\rho(sA/T) = \rho(A/T)$ can be any real number greater than 1, and $\rho(A/T) > 1$ implies $\rho(A) > 1$ or $\rho(T) > 1$.

Any meromorphic function of finite order can be written as the quotient of two Weierstrass products times an exponential function. Thus

$$A(s) = \frac{N(s)}{D(s)} \exp[P(s)], \tag{60}$$

where $N(s)$ and $D(s)$ are Weierstrass products formed with the zeroes and poles of $A(s)$, and $P(s)$ is a polynomial. For $T(s)$, we can write a similar formula. Since $T(s)$ has only one zero, which must be simple, we have

$$T(s) = \frac{s}{d(s)} \exp[p(s)],$$

where $d(s)$ is a Weierstrass product and $p(s)$ is a polynomial. Since $A(s)$ and $T(s)$ are analytic in the right half-plane, and bounded on the imaginary axis, all the zeroes of $d(s)$ and $D(s)$ are in the left half-plane. Then (49) implies that $d(s) = 0$ if and only if $D(s) = 0$. Since $d(s)$ and $D(s)$ are Weierstrass products, they are the same. Therefore,

$$T(s) = \frac{s}{D(s)} \exp[p(s)]. \tag{61}$$

The denominator $D(s)$ is an entire function having order $\rho(D)$. Let $s_1, s_2, s_3, \cdots$ be its zeroes. Except in the case (52), $\rho(D) \geq 1$ and the sum

$$\sum_{n=1}^{\infty} \frac{1}{|s_n|^{\alpha}} \tag{62}$$

converges or diverges depending on the value of α, which is real. The exponent of convergence is the greatest lower bound of the α's for which the series (62) converges.[16,17] Since $D(s)$ is a Weierstrass product, the exponent of convergence is equal to $\rho(D)$.[16,17] Clearly this exponent of convergence gauges the density of poles of $A(s)$ and $T(s)$ as $|s| \to \infty$. For this reason, we want to show that

$$\rho(D) = \rho(T). \tag{63}$$

This equation and (59) are our new results for meromorphic reflection coefficients.

We begin the proof by showing that the polynomial $p(s)$ which appears in (61) has degree $\leq \rho(D)$. In the case (52), this inequality becomes $0 \leq 0$, which is valid. In all other cases, $\rho(D) \geq 1$ and we assume that the degree of $p(s)$ is $> \rho(D)$; we have to show that this leads to a contradiction. Since $D(s)$ is an entire function, $\rho(D)$ gives the maximum rate of growth of $|D(s)|$ as $s \to \infty$ along a ray in the right half-plane. Since the degree of $p(s)$ is greater than 1 and greater than $\rho(D)$, there is a ray in the right half-plane along which $|\exp[p(s)]|$ increases faster than $|D(s)|$ as $s \to \infty$. Hence $|T(s)| \to \infty$ along this ray; but we know that $T(s) \to 1$ as $s \to \infty$ along this ray. The contradiction shows that $p(s)$ has degree $\leq \rho(D)$. From (61) and the inequality for the Nevanlinna order of a product, we obtain $\rho(T) \leq \rho(D)$. To complete the proof of (63), we use the connection between the Nevanlinna order and the exponent of convergence of the poles of $T(s)$. Since $\rho(D)$ is equal to the exponent of convergence of these poles, we must have $\rho(T) \geq \rho(D)$. This and the previous inequality imply (63).

The remaining question is the order of $A(s)$; we want to complete the proof of (59). From (61),

$$T(s)T(-s) = \frac{-s^2 \exp[p(s) + p(-s)]}{D(s)D(-s)}. \tag{64}$$

The denominator which appears here is the product of two functions having the same Nevanlinna order. From the inequality for the order of a product, we find that the order of the denominator is $\leq \rho(D)$. Furthermore, the zeroes of $D(s)D(-s)$ have the same exponent of convergence as the zeroes of $D(s)$; this is $\rho(D)$. The order of an entire function is at least equal to the exponent of convergence of its zeroes[16,17]; this means that the order of the denominator is $\geq \rho(D)$. Therefore, the denominator in (64) has order $\rho(D)$. The degree of $p(s) + p(-s)$, being less than or equal to the degree of $p(s)$, must be $\leq \rho(D)$. Therefore, $T(s)T(-s)$ has the same Nevanlinna order as $D(s)$. Using (49) and the inequality for the order of a product, we obtain

$$\rho(A) \geq \rho(D) = \rho(T). \tag{65}$$

The next step is to prove that the orders of the functions in (60) satisfy

$$\max[\rho(N), \rho(D)] \geq \rho(A), \tag{66}$$

except in the simple case (34). To prove this, assume that $\rho(N) < \rho(A)$ and $\rho(D) < \rho(A)$. Then $\rho(A)$ must be equal to the degree of $P(s)$. If the degree of $P(s)$ is less than 2, our assumptions lead to $\rho(D) < 1$, which is possible only in the case of (34) and (52). Hence we may assume that the degree of $P(s)$ is greater than 1. Then there is a ray in the right half-plane along which $|\exp[P(s)]| \to \infty$ as $s \to \infty$. Although $N(s)$ may have infinitely many zeroes, there is a sequence of circles $|s| = r_n$ on which $|N(s)|$ is not very small; see Chapter 3 of Boas[17] for estimates of $|N(s)|$. These circles interesect the ray in the right half-plane, giving us an infinite sequence of points on which $s \to \infty$ and $|N(s)\exp[P(s)]| \to \infty$, because $\rho(N) < \rho(A)$. In fact, $|N(s)\exp[P(s)]|$ increases faster than $|D(s)|$ on this sequence, because $\rho(D) < \rho(A)$. Hence $A(s) \to \infty$. However, $A(s) \to 0$ as $s \to \infty$ in the right half-plane. The contradiction means that our assumptions must be false; this proves (66).

We now consider the product

$$A(s)A(-s) = \frac{N(s)N(-s)}{D(s)D(-s)} \exp[P(s) + P(-s)]. \tag{67}$$

The exponent of convergence of the zeroes of $N(s)N(-s)$ is equal to $\rho(N)$; the proof of this was given below (64). The zeroes of (67) have $\rho(N)$ as their exponent of convergence. From the Nevanlinna theory, (67) has order $\geq \rho(N)$. But we used (64) to show that $A(s)A(-s)$ has order $\rho(D)$. Therefore,

$$\rho(D) \geq \rho(N).$$

By combining this inequality with (66), we obtain

$$\rho(D) \geq \rho(A). \tag{68}$$

Finally we prove (59) by combining (65) and (68).

This concludes our general discussion of meromorphic functions, in which we have traced the consequences of (50) and (56). Either assumption implies that $A(s)$ and $T(s)$ have infinitely many poles, except in the special case of (34) and (52). $A(s) = -\lambda(s+\lambda)^{-1}$ is a limiting case of (34), but all other rational functions remain to be considered.

VII. RATIONAL REFLECTION COEFFICIENTS

In this section, we assume that $A(s)$ is a rational function of s. We present a general method for finding $B(x, s)$ and $F(x, s)$, whose asymptotic forms give $q(x)$. Kay's solution[9] for $q(x)$ is much more explicit than this, but it depends on a simplifying assumption about the zeroes of (49). We shall derive this solution from our functional equation (32). If $A(s)$ is a Butterworth function, Kay's assumption about the zeroes of (49) is grossly violated; hence we shall solve the plasma inverse problem for this Butterworth case.

In the most general rational case,

$$A(s) = N(s)/D(s),$$

where $N(s)$ and $D(s)$ are polynomials; they must be relatively prime. Suppose that $D(s)$ is a polynomial of degree n. Since $A(s) \to 0$ as $s \to +\infty$, n must be a positive integer, and $N(s)$ must have degree less than n. From (32), we obtain

$$N(s)[\exp(sx) + F(x, -s)] + D(s)F(x, s) = D(s)B(x, s). \tag{69}$$

Since the left-hand side is an entire function,

$$D(s)B(x, s)\exp(sx) \tag{70}$$

must be an entire function. The bound (30) can be used to show that (70) is bounded by a polynomial if $s \to \infty$ while the real part of s is bounded above. If the real part is not bounded above, (26) can be used to show that (70) is bounded by a polynomial. Application of Liouville's theorem on entire functions gives

$$D(s)B(x, s) = P(x, s)\exp(-sx),$$

where $P(x, s)$ is a polynomial in s. The bound (26) implies that $P(x, s)$ has degree $n-1$, at most. Thus the coefficient of s^{n-1} is an unknown function of x, the coefficient of s^{n-2} is another unknown function, and so forth. To determine these n unknown functions, we rewrite (69) as

$$D(s)F(x, s) + N(s)F(x, -s) = -N(s)\exp(sx) + P(x, s)\exp(-sx).$$

If we replace s by $-s$, we obtain another equation in $F(x, s)$ and $F(x, -s)$; thus we can solve for $F(x, s)$. We obtain

$$F(x, s) = \frac{-N(s)[D(-s) + P(x, -s)]\exp(sx) + [P(x, s)D(-s) + N(s)N(-s)]\exp(-sx)}{D(s)D(-s) - N(s)N(-s)} \tag{71}$$

when $x \geq 0$. The denominator which appears here is an even polynomial of degree $2n$; its zeroes are called

$$\kappa_1, \kappa_2, \ldots, \kappa_n \tag{72a}$$

and

$$-\kappa_1, -\kappa_2, \ldots, -\kappa_n. \tag{72b}$$

Of course $F(x, s)$ must be an entire function of s. This requirement will give $2n$ inhomogeneous linear equations satisfied by the n unknown functions in $P(x, s)$. We claim that these $2n$ equations determine $P(x, s)$ uniquely, except at isolated values of x. To show this, suppose that there are two solutions, $P_1(x, s)$ and $P_2(x, s)$. The difference of these two polynomials is $\sum_{j=0}^{n-1} p_j(x)s^j$, and the n functions $p_j(x)$ satisfy $2n$ homogeneous linear equations. We shall select n of these equations and shall find that the $n \times n$ determinant of the system does not vanish, except at isolated values of x. The nonvanishing determinant implies that all the $p_j(x)$ must vanish; then $P(x, s)$ and $B(x, s)$ are unique. If the determinant vanishes at positive real values of x, the requirement that $P(x, s)$ is a continuous function of x may perhaps serve to make the solution unique. After finding $B(x, s)$, we use (33) and (22) to find $q(x)$.

The explicit form of the $n \times n$ determinant will be obtained under the simplifying assumptions (1) that all the numbers (72a) are distinct, and (2) that the real part of

κ_i is positive for $i=1,2,\ldots,n$. With these assumptions, the condition that $F(x,s)$ is entire leads to

$$\sum_{j=0}^{n-1}[D(-\kappa_i)\exp(-\kappa_i x)(\kappa_i)^j - N(\kappa_i)\exp(\kappa_i x)(-\kappa_i)^j]p_j(x)=0$$

and

$$\sum_{j=0}^{n-1}[D(\kappa_i)\exp(\kappa_i x)(-\kappa_i)^j - N(-\kappa_i)\exp(-\kappa_i x)(\kappa_i)^j]p_j(x)=0. \tag{73}$$

The determinant of the system (73) is

$$V\left[\prod_{i=1}^{n} D(\kappa_i)\exp(\kappa_i x)\right]+\cdots, \tag{74}$$

where the terms not written explicitly are relatively small as $x\to+\infty$, and V is the Vandermonde determinant formed from the numbers (72b). Clearly the determinant (74) is an entire function of x. If the leading term (written explicitly) does not vanish, then (74) cannot vanish except at isolated points in the complex x plane. To show that the leading term cannot vanish, we notice that $V\neq 0$, because of assumption (1). Also, $D(s)$ cannot vanish when the real part of s is nonnegative, because $A(s)$ has poles only in the left half-plane. Hence (74) cannot vanish identically. Since it is an entire function, it vanishes only at isolated values of x.

Assumption (2) can now be dropped. In any case, the zeroes of

$$D(s)D(-s)-N(s)N(-s) \tag{75}$$

are listed as (72), and we can label them so that the real part of κ_j is nonnegative. If the real part of κ_j is zero, we can demand that the imaginary part of κ_j be nonnegative. Then we can choose positive number θ, sufficiently small so that the real part of $\kappa_j\exp(-i\theta)$ is always positive, unless $\kappa_j=0$. We now let $x=|x|\times\exp(-i\theta)$, and keep θ fixed as $|x|\to+\infty$. Again the determinant (74) cannot vanish identically, unless one of the numbers (72a) vanishes. In fact, one of these numbers must vanish, because $A(0)=-1$. Let $\kappa_1=0$. Then the argument used above does not apply to the first row of the determinant. The determinant of the system (73) has $D(0)-N(0)$ in the upper left corner, and zeroes elsewhere in the first row. Since $A(0)=-1$, $D(0)-N(0)\neq 0$. To correct the leading term in (74), we replace $D(0)$ by $D(0)-N(0)$. This leading term cannot vanish, and assumption (2) is quite superfluous.

Assumption (1) can also be removed. If (75) has multiple zeroes, the numerator in (71) must vanish at each such zero, together with one or more of its derivatives. One can set up an $n\times n$ determinant, and show that it does not vanish identically. Hence $P(x,s)$ and $B(x,s)$ are determined uniquely.

In order to obtain an explicit formula for $q(x)$, we retain assumption (1) and introduce two other simplifying assumptions. Assumption (1) is slightly weaker than the corresponding assumption in Kay's paper,[9] where the other two assumptions appear implicitly. The first of these other assumptions is that $A(s)$ has only simple poles. The method of partial fractions gives

$$A(s)=\sum_{j=1}^{n}\frac{\alpha_j}{s+\lambda_j},$$

where $\alpha_j\neq 0$. Since $A(s)$ has poles only in the left half-plane, the real part of λ_j is positive. Since $B(x,s)$ must have the same singularities as $A(s)$, we write

$$B(x,s)=\sum_{j=1}^{n}\frac{\beta_j(x)}{s+\lambda_j}\exp(-sx), \tag{76}$$

where the n unknown functions $\beta_j(x)$ are to be determined. We know that $F(x,s)$ is a linear combination of $\exp(sx)$ and $\exp(-sx)$, with rational coefficients. The poles of the coefficients are the zeroes of (75). Suppose that zeroes of $A(s)$ never coincide with poles of $A(-s)$; this is the last of the simplifying assumptions. It implies that the κ's and λ's are disjoint sets of complex numbers, and that (49) has the same zeroes as (75). Hence $A(\kappa_j)A(-\kappa_j)=1$. We can expect to write $F(x,s)$ as a linear combination of the entire functions

$$\{\sinh[x(s\pm\kappa_j)]\}/(s\pm\kappa_j),$$

with coefficients depending only on x. To abbreviate the calculation, we write

$$F(x,s)=2\sum_{j=1}^{n}f_j(x)\left(A(\kappa_j)\frac{\sinh[x(s-\kappa_j)]}{(s-\kappa_j)}-\frac{\sinh[x(s+\kappa_j)]}{(s+\kappa_j)}\right), \tag{77}$$

where the functions $f_j(x)$ are to be determined. Substitute this into (32) and separate the terms in $\exp(sx)$; they are

$$A(s)+A(s)\sum_{j=1}^{n}f_j(x)\left(A(\kappa_j)\frac{\exp(\kappa_j x)}{(s+\kappa_j)}-\frac{\exp(-\kappa_j x)}{(s-\kappa_j)}\right)$$
$$+\sum_{j=1}^{n}f_j(x)\left(A(\kappa_j)\frac{\exp(-\kappa_j x)}{(s-\kappa_j)}-\frac{\exp(\kappa_j x)}{(s+\kappa_j)}\right). \tag{78}$$

This is a rational function of s, and it vanishes as $s\to\infty$. It must vanish identically. To achieve this, we set the residue equal to zero at each possible pole. The expression (77) was contrived so that (78) has zero residue at $s=\kappa_j$ and at $s=-\kappa_j$. We must have zero residue at $s=-\lambda_j$; this gives

$$\sum_{k=1}^{n}\left(A(\kappa_k)\frac{\exp(\kappa_k x)}{\lambda_j-\kappa_k}-\frac{\exp(-\kappa_k x)}{\lambda_j+\kappa_k}\right)f_k(x)=1 \tag{79}$$

for $j=1,2,\ldots,n$. We now have n linear equations to determine the n unknown functions $f_j(x)$. We could determine the functions $\beta_j(x)$ by examining the terms in $\exp(-sx)$ which appear in (32); but in fact (76) is no longer needed. To determine $q(x)$, we need only $F(x,s)$; this requires that we solve (79). To simplify the problem, we introduce an $n\times n$ matrix $M(x)$ with elements

$$M_{jk}(x)=A(\kappa_k)\frac{\exp(-\lambda_j x+\kappa_k x)}{\lambda_j-\kappa_k}-\frac{\exp(-\lambda_j x-\kappa_k x)}{\lambda_j+\kappa_k}.$$

Then (79) becomes

$$\sum_{k=1}^{n}M_{jk}(x)f_k(x)=\exp(-\lambda_j x).$$

In order to complete this calculation, we assume that the inverse matrix exists. Then

$$f_j(x)=\sum_{k=1}^{n}[M(x)^{-1}]_{jk}\exp(-\lambda_k x)$$

and

$$\mathcal{E}(x,x)=-\lim_{s\to\infty}s\exp(sx)F(x,s)$$

$$= \sum_{j=1}^{n} [A(\kappa_j)\exp(\kappa_j x) - \exp(-\kappa_j x)] f_j(x)$$

$$= -\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{d}{dx} M_{kj}(x)\right)[M(x)^{-1}]_{jk}$$

$$= -\operatorname{Tr}[M(x)^{-1}]\frac{d}{dx} M(x).$$

Using a convenient identity and inserting a constant of integration, we obtain

$$\mathcal{E}(x,x) = -\frac{d}{dx}\log\frac{\det[M(x)]}{\det[M(0)]}.$$

The denominator is chosen so that $\mathcal{E}(x,x)$ is continuous at $x=0$; $q(x)$ and $\mathcal{E}(x,x)$ must vanish for $x<0$. Finally we have

$$q(x) = -2\frac{d^2}{dx^2}\log\frac{\det[M(x)]}{\det[M(0)]},$$

which is derived from the functional equation (32) and which agrees with Kay's solution of the problem.[9] Each element of $M(x)$ increases or decreases exponentially as $x\to+\infty$, except in the simple case of $A(s) = -\lambda(s+\lambda)^{-1}$.

If $\kappa_2 = 0$, $q(x)$ cannot be expected to decrease exponentially as $x\to+\infty$; this can be treated as a limiting case of Kay's solution. If all the κ's vanish, we expect $q(x)$ to be a rational function when x is positive; a δ function may appear at $x=0$. This problem cannot easily be treated as a limiting case of Kay's solution. Hence we treat it separately. We assume that all the κ's vanish; this is the opposite of assumption (1). Also, we make the simplifying assumption that $A(s)$ has no zeroes. These two assumptions imply that $A(s)$ is one of the Butterworth functions; these functions first appeared in a different context.[10] The first assumption implies

$$A(s)A(-s) = \frac{D(s)D(-s) - (is)^{2n}}{D(s)D(-s)},$$

and the second assumption implies that the numerator which appears here is a constant. By changing the scale of s, we can arrange to have $A(s)A(-s) = [1+(is)^{2n}]^{-1}$. Using (48), we obtain

$$A(s) = -\left(\sum_{j=0}^{n} a_j s^j\right)^{-1}.$$

The constants $a_0, a_1, \ldots, a_n$ are real and positive, because the poles of $A(s)$ are in the left half-plane and are arranged in complex conjugate pairs. With this notation, $a_0 = a_n = 1$. Equation (71) leads us to

$$F(x,s) = \frac{1}{s^{2n}}\left(\sum_j g_j(x)s^j\exp(sx) + \sum_j h_j(x)s^j\exp(-sx)\right).$$

The condition (30) implies that $j\le 2n-1$ in both these sums. Hence we have $4n$ unknown functions of x, instead of the unknown $f_j(x)$ in (77). The last of the h's gives the function we want:

$$h_{2n-1}(x) = -\mathcal{E}(x,x).$$

We can write $B(x,s)$ in terms of other unknown functions of x, because (70)is a polynomial. We use (32) to connect the various unknown functions of x. The terms in $\exp(sx)$ give

$$s^{2n} + \sum_{j=0}^{2n-1} h_j(x)(-s)^j + \frac{1}{A(s)}\left(\sum_j g_j(x)s^j\right) = 0. \qquad (80)$$

The last term on the left is a product of two polynomials. Equation (80) requires that $g_j(x)=0$ when $j>n$ and that $g_n(x)=1$. Moreover, $F(x,s)\exp(sx)$ can have no pole at $s=0$; this means that

$$\left(\sum_{j=0}^{n} g_j(x)s^j\right)\exp(2sx) + \sum_{j=0}^{2n-1} h_j(x)s^j$$

vanishes at $s=0$, together with its first $2n-1$ derivatives. Let us define $\xi_k = (2x)^k/k!$, and write these two conditions on the g's and h's as

$$\sum_{k=j}^{n-1} a_{n+j-k}g_k(x) - (-1)^{n+j}h_{n+j}(x) = -a_j$$

and

$$\sum_{k=0}^{n-1} \xi_{n+j-k}g_k(x) + h_{n+j}(x) = -\xi_j,$$

for $j = 0,1,2,\ldots,n-1$. We have eliminated s, and obtained $2n$ linear inhomogeneous equations. Solution by Cramer's rule gives

$$\mathcal{E}(x,x) = \frac{\begin{vmatrix} 1 & a_{n-1} & \cdots & a_1 & (-1)^{n+1} & 0 & \cdots & a_0 \\ 0 & 1 & \cdots & a_2 & 0 & (-1)^{n+2} & \cdots & a_1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & a_{n-1} \\ \xi_n & \xi_{n-1} & \cdots & \xi_1 & 1 & 0 & \cdots & \xi_0 \\ \xi_{n+1} & \xi_n & \cdots & \xi_2 & 0 & 1 & \cdots & \xi_1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \xi_{2n-1} & \xi_{2n-2} & \cdots & \xi_n & 0 & 0 & \cdots & \xi_{n-1} \end{vmatrix}}{\begin{vmatrix} 1 & a_{n-1} & \cdots & a_1 & (-1)^{n+1} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & a_2 & 0 & (-1)^{n+2} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & (-1)^{2n} \\ \xi_n & \xi_{n-1} & \cdots & \xi_1 & 1 & 0 & \cdots & 0 \\ \xi_{n+1} & \xi_n & \cdots & \xi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \xi_{2n-1} & \xi_{2n-2} & \cdots & \xi_n & 0 & 0 & \cdots & 1 \end{vmatrix}}$$

The $2n\times 2n$ determinants which appear here could easily be reduced to $n\times n$ determinants. Finally we compute

$$q(x) = 2\frac{d}{dx}\mathcal{E}(x,x),$$

which gives the plasma density. It is a rational function of x when $x>0$. The case of $n=1$ appears as the limiting case at the end of Sec. IV, and the case of $n=2$ has recently been solved by Jordan and Kritikos.[20]

VIII. CONCLUSION

A reformulation of the plasma inverse problem has been presented. In order to obtain $q(x)$ from $R(y)$, we want to solve the integral equation (23). Laplace transformation gives the functional equation (32), which determines $q(x)$ in terms of $A(s)$, the Laplace transform of the reflection. Analytic solutions of (32) have been obtained in some simple cases. If $q(x)$ decreases sufficiently rapidly as $x \to +\infty$, then the reflection coefficient $A(s)$ is meromorphic, the Nevanlinna order $\rho(A)$ serves to classify reflection coefficients, and we have the results (59) and (63). The discussion presented here is not immediately applicable to the reflection of electromagnetic waves by the ionosphere, because of various effects omitted in the simple model. However, the simple model used here is adapted to radar frequencies near the plasma frequency. The anomalously small Doppler broadening of ionospheric reflections[21] suggests that the approximation of a cold plasma, as used here, is justified. Further work in the direction indicated here is expected to yield a method of interpreting the amplitude and phase of radar reflections from the ionosphere.

ACKNOWLEDGMENTS

We wish to thank J. M. Cohen, A. K. Jordan, B. Lepson, C. F. Osgood, A. J. Skalafuris, and P. B. Richards for help with this work. This paper was supported by the Office of Naval Research.

APPENDIX

Here we consider the convergence of the series (17). We want to show that (17) converges absolutely, if $R(y)$ is bounded and (20) is satisfied. More generally, we may assume that

$$|R(y)| \le R\exp(-\lambda y),$$

where R and λ are real parameters. Then (18) gives

$$|K_0(x,t)| \le R\exp[-\lambda(x+t)]$$

and

$$|K_1(x,t)| \le \frac{R^2}{2\lambda}\exp[-\lambda(x-t)] - \frac{R^2}{2\lambda}\exp[-\lambda(3x+t)].$$

In general,

$$\begin{aligned}|K_{2n}(x,t)| \le{}& R\left(\frac{-R^2}{4\lambda^2}\right)^n \sum_{m=0}^{n}(-1)^m[L_{n-m}^{(-2n)}(2\lambda(4mx+x+t)) \\ &+ L_{n-m-1}^{(-2n)}(2\lambda(4mx+x+t))]\exp[-\lambda(4mx+x+t)] \\ &- R\left(\frac{-R^2}{4\lambda^2}\right)^n \sum_{m=1}^{n}(-1)^m[L_{n-m}^{(-2n)}(2\lambda(4mx-x-t)) \\ &+ L_{n-m-1}^{(-2n)}(2\lambda(4mx-x-t))]\exp(-\lambda(4mx-x-t)]\end{aligned} \qquad \text{(A1)}$$

and

$$\begin{aligned}|K_{2n+1}(x,t)| \le{}& \frac{R^2}{2\lambda}\left(\frac{-R^2}{4\lambda^2}\right)^n \sum_{m=0}^{n}(-1)^m[L_{n-m}^{(-2n-1)}(2\lambda(4mx+x-t)) \\ &+ L_{n-m-1}^{(-2n-1)}(2\lambda(4mx+x-t))] \\ &\times\exp[-\lambda(4mx+x-t)] \\ &+ \frac{R^2}{2\lambda}\left(\frac{-R^2}{4\lambda^2}\right)^n \sum_{m=1}^{n+1}(-1)^m[L_{n-m+1}^{(-2n-1)}(2\lambda(4mx-x+t)) \\ &+ L_{n-m}^{(-2n-1)}(2\lambda(4mx-x+t))] \\ &\times\exp[-\lambda(4mx-x+t)],\end{aligned} \qquad \text{(A2)}$$

where $L_n^{(\alpha)}$ is the Laguerre polynomial of degree n. The polynomial of degree -1 is defined as zero, and the other polynomials are given implicitly by

$$(1-z)^{-\alpha-1}\exp[xz/(z-1)] = \sum_{n=0}^{\infty} z^n L_n^{(\alpha)}(x).$$

An explicit formula for the Laguerre polynomial can be written, using contour integration. We use this formula to construct an inductive proof of (A1) and (A2), and to derive the sum formulas:

$$\sum_{n=m}^{\infty}(-\tfrac{1}{4}\sin^2 2B)^n[L_{n-m}^{(-2n)}(x) + L_{n-m-1}^{(-2n)}(x)] = (-\tan^2 B)^m\exp(x\sin^2 B)$$

and

$$\sum_{n=m}^{\infty}(-\tfrac{1}{4}\sin^2 2B)^n[L_{n-m}^{(-2n-1)}(x) + L_{n-m-1}^{(-2n-1)}(x)] = (-\tan^2 B)^m(\cos B)^{-2}\exp(x\sin^2 B).$$

From these we obtain

$$\sum_{n=0}^{\infty}|K_n(x,t)| \le R\cosh(x+t)(\lambda^2-R^2)^{1/2} - R\left(\frac{1-\Gamma}{1+\Gamma}\right)\sinh(x+t)(\lambda^2-R^2)^{1/2},$$

where

$$\Gamma = (1/R)[\lambda - (\lambda^2-R^2)^{1/2}]\exp[-2x(\lambda^2-R^2)^{1/2}].$$

The infinite sum converges if $\lambda \ge R$. Also, it converges if $|\lambda| < R$ and

$$0 \le |t| < x < \frac{\pi - \arccos(\lambda/R)}{2(R^2-\lambda^2)^{1/2}}.$$

Finally we set $\lambda = 0$ and obtain (20).

*A preliminary version of this work has appeared as NRL Report 7946.

†Present address: Institute for Advanced Study, Princeton, New Jersey 08540.

[1]K. M. Case and M. Kac, J. Math. Phys. 14, 594 (1973); K. M. Case, J. Math. Phys. 14, 916 (1973), and later papers.
[2]F. J. Dyson, in *Studies in Mathematical Physics—Essays in Honor of V. Bargmann*, edited by A. S. Wightman (Princeton U. P., Princeton, N. J., 1976).
[3]I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Mat. 15, 309 (1951) [Amer. Math. Soc. Transl. 1, 253 (1955)].
[4]L. D. Faddeev, Usp. Mat. Nauk 14, 57 (1959) [J. Math. Phys. 4, 72 (1963)].
[5]G. N. Balanis, J. Math. Phys. 13, 1001 (1972).
[6]I. Kay, NYU Res. Report No. EM-74 (1955).
[7]G. N. Balanis, thesis, California Institute of Technology (1972).
[8]R. Nevanlinna, Acta Math. 46, 1 (1925).
[9]I. Kay, Comm. Pure Appl. Math. 13, 371 (1960).
[10]S. Butterworth, Exptl. Wireless and Wireless Engineer 7, 536 (1930).
[11]J. L. Lebowitz and D. Zomick, J. Chem. Phys. 54, 3335 (1971).
[12]F. Gross, C. F. Osgood, and C. C. Yang, J. Math. Phys. 16, 2142 (1975).
[13]R. Jost, Helv. Physica Acta 20, 256 (1947).
[14]Appendix in V. Bargmann, Rev. Mod. Phys. 21, 488 (1949).
[15]T. Regge, Nuovo Cimento 9, 295 (1958).
[16]B. Ya. Levin, *Raspredelenie Korneĭ Tselykh Funktsii* (Gosudarstvennoe, 1956) [transl: (Amer. Math. Soc., Providence, R. I., 1964)].
[17]R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[18]L. Sartori, J. Math. Phys. 4, 1408 (1963).
[19]W. K. Hayman, *Meromorphic Functions* (Oxford U. P., Oxford, 1964).
[20]A. K. Jordan and H. N. Kritikos, IEEE Trans. Antennas Propagation AP-21, 909 (1973).
[21]K. L. Bowles, Phys. Rev. Lett. 1, 454 (1958); J. P. Dougherty and D. T. Farley, Proc. Roy. Soc. A 259, 79 (1960); J. A. Fejer, Can. J. Phys. 38, 1114 (1960); E. E. Salpeter, Phys. Rev. 120, 1528 (1960).

Multiplicative renormalization of composite operators

H. J. de Vega* and F. Schaposnik*

Laboratoire de Physique Théorique, Ecole Normale Supérieure, Paris, France
(Received 24 June 1975)

We study composite operators with the aid of dimensional regularization. Using Zimmermann "normal product algorithm," we obtain a simple formula for any normal product in renormalized field theory and discuss some of its properties.

I. INTRODUCTION

The problem of defining renormalized composite operators in QFT has been treated by several authors using different approaches.[1] Among them, Zimmermann has developed a powerful formalism called the "normal product algorithm,"[2] in which composite operators are defined in Green's functions, and renormalized afterwards with the BPHZ method.[3]

With the development of the dimensional regularization method[4]—which as it is well-known, presents noteworthy properties—comes the possibility of introducing composite operators (CO) in the conventional renormalization approach.[5] They are defined using the number of dimensions (ν) as a regulator, and coincide with the CO introduced by Zimmermann when $\nu = 4$.

This new approach simplifies explicit calculations with CO and provides a new method to study different properties in QFT such as generalized Ward identities anomalies, etc. We present here an investigation in that direction. We establish (following Zimmermann's proof of the so-called "Zimmermann identities,"[2] but in the context of dimensional regularization) the possibility of obtaining any normal product $N[\Theta(x)]$ as a linear combination of a finite number of regularized CO. Those operators share the quantum numbers of $\Theta(x)$ and have a dimension that, at most, equals that of $\Theta(x)$.

The relation between normal products and regularized CO that we obtain, are related to Zimmermann identities between normal products of different degrees.[2] In our case the renormalization constants and the regularized Green's functions have poles at $\nu = 4$. The objects in Zimmermann's identities between normal products $N_d[\Theta]$ with $d \geq \dim\Theta$ are all finite in four dimensions. Actually, the expressions that we obtain correspond to a Zimmermann identity, in ν dimensions, between a normal product $N_{-1}[\Theta]$ and $N_d[\Theta]$ ($d \geq \dim\Theta$) sufficiently renormalized.

The above mentioned expression for a normal product is obtained in Sec. II. In Sec. III we discuss some properties of the new renormalization constants and state some concluding remarks.

II. REGULARIZED AND RENORMALIZED OPERATORS

The aim of this section is to present a derivation of the normal product expansion for any CO in terms of ν-dimensional regularized operators.

For the sake of clarity, we begin by considering a model of a scalar field with a $g\phi^4$ coupling, but all our results will be valid for any other renormalizable quantum field theory. We use the dimensional regularization method,[4] which as is well known, shows remarkable advantages over the other previous methods. In particular, it was chosen due to its property of preserving symmetries and invariances. Nevertheless, it is worthwhile mentioning that all that follows might have been done with any other regularization method sharing those properties.

We start by considering the standard Lagrangian that includes all counterterms

$$\mathcal{L}(x) = \tfrac{1}{2}(Z_2 + z_2)(\partial\phi)^2 - \tfrac{1}{2}(M^2 + \delta M^2 + \delta m^2)\phi^2 - (g/4!)(Z_1 + z_1)\phi^4; \tag{2.1}$$

here, Z_2, δM^2, and Z_1 are computed from the one-particle irreducible (1 PI) diagrams with two and four external legs, taken at zero external momenta. In other words, these counterterms eliminate all UV divergencies from the Green's functions (GF). In Eq. (2.1), z_2, δm^2, and z_1 are finite renormalizations.

The connected GF are then specified, in perturbation theory by the Gell-Mann–Low formula

$$\begin{aligned} G^{(N)}&(x_1, x_2, \ldots, x_N; M, g, \nu) \\ &= \langle T\phi(x_1)\cdots\phi(x_N)\rangle^{\text{conn}} \\ &= \langle 0|T\phi^{(0)}(x_1)\cdots\phi^{(0)}(x_N)\exp[i\textstyle\int\mathcal{L}_i^{(0)}(x)\,d^\nu x]|0\rangle^{\text{conn}}_{\otimes}. \end{aligned} \tag{2.2}$$

Here $\phi^{(0)}(x)$ is the free scalar field, $|0\rangle$ the free field vacuum state, and with the symbol $\otimes$ we are indicating that we omit all vacuum–vacuum diagrams. The interaction Lagrangian $\mathcal{L}_i(x)$ is

$$\mathcal{L}_i(x) = \mathcal{L}(x) - [\tfrac{1}{2}(\partial\phi)^2 - \tfrac{1}{2}M^2\phi^2_{(x)}]. \tag{2.3}$$

As it was noticed above, all preceeding expressions are considered in a ν-dimensional space.

Now, these GF, order by order in perturbation theory, are meromorphic functions of ν and its limit when $\nu \to 4$ exists and is finite.[4,6] As it was stressed in the Introduction, an alternative way of obtaining GF has been given by Zimmermann[2] without using regulators (obviously working in a four-dimensional space). This method gives

$$G^{(N)}(x_1, \ldots, x_N; M, g, 4) = \mathrm{PF}\langle 0 | T\phi^{(0)}(x_1)\cdots\phi^{(0)}(x_N) \times \exp[i\int \mathcal{L}^{(0)}_{\mathrm{eff}}(x)\, d^4x] | 0\rangle^{\mathrm{conn}}_{\otimes}, \qquad (2.4)$$

with

$$\mathcal{L}_{\mathrm{eff}}(x) = \tfrac{1}{2}(z_2 - 1)\,(\partial\phi)^2 - (\delta m^2/2)\phi(x)^2 - (g/4!)(1+z_1)\phi^4(x). \qquad (2.5)$$

Here PF means that we must consider the BPHZ finite part diagram contributions computed at zero external momenta.

For our purposes, we must have the previous BPHZ formula (2.4) extended to a ν-dimensional space. This is a straightforward generalization and reads

$$G^{(N)}(x_1, \cdots, x_N; M, g, \nu) = \mathrm{PF}\langle 0 | T\phi^{(0)}(x_1)\cdots\phi^{(0)}(x_N) \times \exp[i\int \mathcal{L}^{(0)}_{\mathrm{eff}}(x)\, d^\nu x] | 0\rangle^{\mathrm{conn}}_{\otimes}. \qquad (2.6)$$

Here z_1, z_2, and δm^2, the finite counterterms of the $\mathcal{L}_{\mathrm{eff}}$ should be considered as functions of ν. Obviously, expressions (2.2) and (2.6) are free from UV divergencies in $\nu = 4$. But once we have taken the limit $\nu \to 4$, if in such renormalized GF, we make $x_1 = x_2 = \cdots = x_j$ ($j = 2, 3, \ldots, N$), it is well known that the resulting expression diverges. This disease shows the necessity of giving renormalized expressions for the so-called composite operators—monomials of the basic field $\phi(x)$ and its derivatives. As we have mentioned, there exists a definition of a GF with a CO, due to Zimmermann, that we will use as the starting point of our derivation. Zimmermann's expressions when written in ν dimensions read

$$\langle TN[\Theta(x)]\phi(x_1)\cdots\phi(x_N)\rangle = \mathrm{PF}\langle 0 | \Theta^{(0)}(x)\phi^{(0)}(x_1)\cdots\phi^{(0)}(x_N) \times \exp[i\int \mathcal{L}_{\mathrm{eff}}(x)\, d^\nu x] | 0\rangle^{\mathrm{conn}}_{\otimes}, \qquad (2.7)$$

where $\Theta^{(0)}(x)$ is obtained from the CO $\Theta(x)$ by the substitution $\phi(x) \to \phi^{(0)}(x)$.

Let us point out that if we take $x_1 = x_2 = \cdots = x_j$ ($j = 2, 3, \ldots, N$) in a renormalized GF maintained in ν dimensions, the resulting expression remains mathematically well defined, provided ν is sufficiently negative. The second step will be to derive an expression for a nonrenormalized CO, meromorphic function of the dimension ν.[7]

We define the regularized CO in terms of its expression in momentum space

$$\Theta_{\{\mu\}}(x) = \prod_{i=1}^{m} \partial_{(\mu)_i}\phi(x), \qquad (2.8)$$

$$\partial_{(\mu)_i} \equiv \partial_{\mu_{j_1}}\cdots\partial_{\mu_{j_{m(j)}}};\ \partial_\mu = \frac{\partial}{\partial x^\mu};\ \{\mu\} \equiv \{(\mu)_1, \cdots, (\mu)_m\},$$

by the expression

$$\langle T\Theta_{\{\mu\}}(x)\phi(x_1)\cdots\phi(x_N)\rangle^{\mathrm{conn}} = \int \prod_{i=1}^{m}\frac{d^\nu q_i}{(2\pi)^\nu}\prod_{j=1}^{N}\frac{d^\nu p_j}{(2\pi)^\nu} \times \exp[-i(\sum_{j=1}^{N} p_j\cdot x_j + x\sum_{i=1}^{m} q_i)]P_{\{\mu\}} \times G^{(N+m)}(q_1, \ldots, q_m, p_1, \ldots, p_N; M, g, \nu), \qquad (2.9)$$

where

$$P_{\{\mu\}} = \prod_{i=1}^{m} q_{i,(\mu)_i}(-i)^{\sum_{j=1}^{m} m(j)},\quad q_{i,(\mu)_i} = q_{i,\mu_j}\cdots q_{i,\mu_{j_{m(j)}}}.$$

The perturbative expansion of the rhs of Eq. (2.9) contains diagrams with $(m+N)$ legs; m of those legs have been identified forming the composite vertex and also the factor $P_{\{\mu\}}$ has been added. Note that even the mentioned diagrams are different from those contributing to an ordinary GF; in the context of dimensional regularization, they have UV divergencies which, like for ordinary GF, appear as poles in $\nu = 4$. [See for example Ref. 6.] It is clear that these poles are not eliminated by the Lagrangian counterterms, because they come from the fact that arguments have been identified in a renormalized GF in configuration space.

We shall now prove that the renormalized operator can be obtained from the following linear combination of regularized operators:

$$\langle TN[\Theta_{\{\mu\}}]\phi(x_1)\cdots\phi(x_N)\rangle = \sum_{\Theta_{\{\rho\}}} (Z^{-1})^{\Theta_{\{\rho\}}}_{\Theta_{\{\mu\}}}\langle T\Theta_{\{\rho\}}\phi(x_1)\cdots\phi(x_N)\rangle, \qquad (2.10)$$

where the sum $\sum_{\Theta_{\{\rho\}}}$ extends over all operators $\Theta_{\{\rho\}}$ with

$$D_{\Theta_{\{\rho\}}} = r + \sum_{j=1}^{r} r(j) < m + \sum_{j=1}^{m} m(j) = D_{\Theta_{\{\mu\}}}. \qquad (2.11)$$

That is to say, the sum extends over all those CO $\Theta_{\{\mu\}}$ which, in $\nu = 4$, have a dimension—in mass units—less than or equal to the dimension of $\Theta_{\{\mu\}}$. We also notice that those operators must have the same transformation properties as $\Theta_{\{\mu\}}$ under the symmetries of the theory.

On the other hand, the renormalization constants $(Z^{-1})^{\Theta_{\{\rho\}}}_{\Theta_{\{\mu\}}}$ are invariant tensors, coordinate-independent, which in general have poles, order by order in perturbation theory, at $\nu = 4$. They are responsible for the renormalization of the CO.

One can derive Eq. (2.10) starting with a Zimmermann identity between nonlocal product of operators, in ν dimensions, and then taking the limit when all points become identical. However, we sketch another demonstration in which it is not necessary to make use of point splitting because one deals with ν-regularized expressions from the beginning. This derivation is analogous to that corresponding to Zimmermann identities between operators of different substraction degree.[2]

Let us consider a diagram Δ which corresponds to the perturbative expansion of the GF, as given by the lhs of Eq. (2.10). In momentum space, this diagram has N external legs corresponding to the field ϕ and a composite to vertex V_0 corresponding to the CO $\Theta_{\{\mu\}}$. This diagram Δ can be obtained from a connected diagram Γ, belonging to the connected GF $G^{(N+m)}$, when we collapse m of its legs, though they form the composite vertex V_0. (We have also to multiply the resulting expression by the factor $P_{\{\mu\}}$.)

In the BPHZ scheme, the renormalized integrand corresponding to Δ is

$$R_\Delta = S_\Delta \sum_{\mathbf{U}\in U(\Delta)} \prod_{\gamma\in\mathbf{U}} (-t^\gamma S_\gamma) S_{\mathbf{U}} I_\Delta. \qquad (2.12)$$

Here, I_Δ is the unsubstracted integrand and $S_{\mathbf{U}}$, S_γ, and S_Δ are the substitution operators (see Ref. 8). $U(\Delta)$ is the set of all Δ-forests and the t^γ are the Taylor operators at order $d(\gamma)$ in the external momenta of γ. Finally, the degree function $d(\gamma)$ is given by

$$d(\gamma) = \begin{cases} 4 - r(\gamma), & \text{if } V_0 \notin \gamma, \\ D_{\Theta_{\{\mu\}}} - r(\gamma), & \text{if } V_0 \in \gamma. \end{cases} \qquad (2.13)$$

[We take $t^\gamma \equiv 0$ for $d(\gamma) < 0$.] In this last equation, $r(\gamma)$ denotes the number of external lines of γ.

The renormalized integrand corresponding to the above mentioned diagram Γ is

$$R_\Gamma = S_\Gamma \sum_{\mathbf{U}\in U(\Gamma)} \prod_{\gamma\in\mathbf{U}} (-t^\gamma S_\gamma) S_{\mathbf{U}} I_\Gamma. \qquad (2.14)$$

The renormalized integrand R_Δ can be expressed as

$$R_\Delta = S_\Delta P_{\{\mu\}} R_\Gamma + X, \qquad (2.15)$$

where X can be written as[2]

$$X = -S_\Delta \sum_{\tau\in T\alpha} \sum_{\{\rho\}} [(-i)^{\sum_{j=1}^r r(j)} / \prod_{j=1}^r r(j)!]$$
$$\times R_{\Delta/\tau\{\rho\}} \{\partial_{p_1^\tau}^{(\rho)_1} \cdots \partial_{p_r^\tau}^{(\rho)_r} R'_\tau\}_{p_j^\tau=0}. \qquad (2.16)$$

Here τ is the smallest element belonging to U, containing V_0, T_α is the set of all proper subdiagrams τ of Δ, μ_τ is the set of all Δ-forests $\mathbf{U}_1$ which display the property of each $\gamma_1 \in \mathbf{U}_1$ satisfying

$$\tau \subset \gamma_1 \ (\tau \neq \gamma) \ \text{ or } \ \gamma \cap \gamma_1 = \emptyset,$$

and $U(\tau)$ is the set of all τ-forest not containing τ. The sum extends to

$$\sum_{j=1}^r r(j) \le D_{\Theta_{\{\mu\}}} - r; \quad r \equiv r(\tau) \qquad (2.17)$$

and

$$R_{\Delta/\tau\{\rho\}} = \sum_{\mathbf{U}_1\in\mu_\tau} \prod_{\gamma_1\in\mathbf{U}_1} (-t^{\gamma_1} S_{\gamma_1})$$
$$\times \prod_{j=1}^r (-i l^\mu_{a;b;\sigma_j})_{(\rho)_j} S_{\mathbf{U}_1} I_{\Delta/\tau}, \qquad (2.18)$$

$$R'_\tau = \sum_{\mathbf{U}_2\in U(\tau)} \prod_{\gamma_2\in\mathbf{U}_2} (-t^{\gamma_2} S_{\gamma_2}) S_{\mathbf{U}_2} I_\tau, \qquad (2.19)$$

where the $l_{a;b;\sigma_j}$ are the momenta arriving to τ.

Here it may be noticed that the integrand given by Eq. (2.19) is not renormalized. Actually, forests containing τ are not subtracted and therefore, when internal momenta integrations are performed, poles will appear at $\nu = 4$. The integration of Eq. (2.16) over the internal momenta, followed by the sum over all Δ and the Fourier transformation with respect to $p, p_1, \ldots, p_N$ lead to

$$\langle T\Theta_{\{\mu\}}(x)\phi(x_1)\cdots\phi(x_N)\rangle$$
$$= \sum_{\Theta_{\{\mu\}}} Z_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}} \langle TN[\Theta_{\{\rho\}}]\phi(x_1)\cdots\phi(x_N)\rangle, \qquad (2.20)$$

where

$$Z_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}} = \delta_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}} + (-i)^{\sum_{j=1}^r r(j)} / \prod_{j=1}^r r(j)!$$
$$\times \sum_{\tau\in F_r} \partial_{q_1}^{(\rho)_1} \cdots \partial_{q_r}^{(\rho)_r} \langle TN'[\Theta_{\{\mu\}}]\tilde\phi(q_1)\cdots\tilde\phi(q_1)\rangle^{1PI}_{q_j^\tau=0}. \qquad (2.21)$$

Here, F_r is the set of all nontrivial diagrams contributing to the proper part and N' means that the integrand contributing to this expression are not the renormalized ones but those given by Eq. (2.19). The sum in Eq. (2.20) extends to all $\Theta_{\{\rho\}}$ with the dimension $D_{\Theta_{\{\rho\}}} \le D_{\Theta_{\{\mu\}}}$.

The renormalization matrix Z is triangular and at zero order in perturbation theory, $\det Z = 1$. Therefore, we can invert Eq. (2.20) and thus obtain Eq. (2.10).

In the case of oversubtracted normal products, $N_\alpha[\Theta]$, we have similarly

$$\langle TN_a[\Theta_{\{\mu\}}(x)]\phi(x_1)\cdots\phi(x_N)\rangle$$
$$= \sum_{\Theta_{\{\rho\}}} (Z_a^{-1})_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}} \langle T\Theta_{\{\rho\}}\phi(x_1)\cdots\phi(x_N)\rangle, \qquad (2.22)$$

where the sum $\sum_{\Theta_{\{\rho\}}}$ extends to all CO with dimension $D_{\Theta_{\{\rho\}}} \le a$.

It is obvious that Eqs. (2.20) show the case in which a composite operator is renormalizable in a multiplicatively way: when there are no operators with the same quantum numbers and dimension less than or equal to it.

III. PROPERTIES OF THE RENORMALIZATION CONSTANTS

In this section we will study some properties of the renormalization constants $Z_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}}$ from the normal product properties.

The constants $Z_{\Theta_{\{\mu\}}}^{\Theta_{\{\rho\}}}$ may either be determined directly from Eq. (2.21), or more easily, by application of the normalization condition of the normal products.

For example, in the case of the operator $\phi^4(x)$

$$Z_{\phi^4}^{\phi^2_2} = \tfrac{1}{2}\langle T\phi^4(0)\tilde\phi(0)\tilde\phi(0)\rangle^{1PI},$$

$$Z_{\phi^4}^{\phi^4} = \tfrac{1}{4}\langle T\phi^4(0)\tilde\phi(0)\tilde\phi(0)\tilde\phi(0)\tilde\phi(0)\rangle$$
$$- (1/4!) Z_{\phi^4}^{\phi^2_2} / Z_{\phi^2}^{\phi^2_2} \langle T\phi^2(0)\tilde\phi(0)\tilde\phi(0)\tilde\phi(0)\tilde\phi(0)\rangle.^{1PI}.$$

One also gets expressions for $Z_{\phi^2}^{\phi^2_2}$—working with $N[\phi^2]$

$$Z_{\phi^2}^{\phi^2_2} = \tfrac{1}{2}\langle T\phi^2(0)\tilde\phi(0)\tilde\phi(0)\rangle^{1PI}$$

and similarly for the remaining constants which renormalize ϕ^4 ($Z_{\phi^4}^{\phi\Box\phi}$ and $Z_{\phi^4}^{(\partial\phi)^2}$). From the linearity of normal products and the differentiation formula, one derives

$$(Z^{-1})^{\theta}_{\theta_1+\theta_2} = (Z^{-1})^{\theta}_{\theta_1} + (Z^{-1})^{\theta}_{\theta_2},$$
$$(Z^{-1})^{\theta_1}_{\theta_2} = (Z^{-1})^{\partial_\mu \theta_1}_{\partial_\mu \theta_2}. \tag{3.1}$$

Finally, we shall obtain relations between CO renormalization constants and the renormalization constants appearing in the Lagrangian (2.1). In order to do this, we shall study the equation of motion satisfied by connected GF. Making use of the generalized Wick's theorem[10]

$$\langle 0 | T\phi^{(0)}(x_1) \circ\circ\circ \phi^{(0)}(x_N)\phi^{(0)}(x) \exp[i \int \mathcal{L}_i^{(0)}(z)\, d^\nu z] | 0\rangle_{\otimes}^{\text{conn}}$$
$$= i \int d^\nu w \langle 0 | T\phi^{(0)}(x_1) \bullet\bullet\bullet \phi^{(0)}(x_N)\, \overbrace{\phi^{(0)}(x)\, \mathcal{L}^{(0)}}(w)$$
$$\times \exp[i \int \mathcal{L}_i^{(0)}(z)\, d^\nu z] | 0\rangle_{\otimes}^{\text{conn}} + i\delta_{N_1}\Delta_F(x-x_1). \tag{3.2}$$

With the use of $\mathcal{L}_i$ given by Eqs. (2.4)—(2.1) we obtain

$$G^{(N+1)}(x, x_1, x_2, \ldots, x_N; M, g, \nu) = i\delta_{N_1}\Delta_F(x-x_1)$$
$$+ \int d^\nu w\, \Delta(x-w) \langle 0 | T\{(Z_2 + z_2 - 1)\Box\phi(w) + (\delta m^2 + \delta M^2)\,\phi(w)$$
$$+ (g/3!)(Z_1 + z_1)\phi^3(w)\}\phi(x_1) \bullet\bullet\bullet \phi(x_N)$$
$$\exp[i \int \mathcal{L}_i^{(0)}(z)\, d^\nu z] | 0\rangle_{\otimes}^{\text{conn}} \tag{3.3}$$

where

$\langle 0 | T\phi^{(0)}(x)\phi^{(0)}(y) | 0\rangle = i\Delta_F(x-y) = \overbrace{\phi(x)\phi}(y).$

Applying the Klein—Gordon differential operator, we finally have

$$(\Box_x + m^2)\langle T\phi(x)\phi(x_1) \bullet\bullet\bullet \phi(x_N)\rangle = -\langle T\{(\delta M^2 + \delta m^2)\phi(x)$$
$$+ (Z_2 + z_2 - 1)\Box\phi(x) + (g/3!)(Z_1 + z_1)\phi^3(x)\}\phi(x_1) \bullet\bullet\bullet \phi(x_N)\rangle$$
$$- i\delta_{N1}\delta(x-x_1). \tag{3.4}$$

Subtracting the equation of motion derived with the aid of the BPHZ algorithm from Eq. (3.4), we obtain

$$Z^{\phi^3}_{\phi^3} = Z_1, \quad gZ^{\phi}_{\phi^3} = 3!\,\delta M^2, \quad gZ^{\Box\phi}_{\phi^3} = 3!\,(Z_2 - 1). \tag{3.5}$$

ACKNOWLEDGMENT

We are indebted to C. Garcia Canal for a critical reading of the manuscript.

*Fellow of the Consejo Nacional de Investigaciones Cientificas y Têcnicas, Argentina.

†Laboratoire propre du C.N.R.S. associé à l'Ecole Normale Supérieure et à l'Université de Paris Sud.

[1]H. Epstein and V. Glaser, in *Statistical Mechanics and QFT*, Les Houches (Gordon and Breach, New York, 1970); K. Wilson (unpublished) quoted in Ref. 8; W. Zimmermann, Refs. 2 and 8.

[2]W. Zimmermann, in *Brandeis Lectures on Elementary Particles and QFT*, edited by S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, Mass., 1971), Vol. I. See also Ref. 8.

[3]N.N. Bogoliubov and O.S. Parasiuk, Acta Math. 97, 227 (1957); K. Hepp, Commun. Math. Phys. 2, 301 (1966); W. Zimmermann, Commun. Math. Phys. 15, 208 (1969).

[4]C.G. Bollini and J.J. Giambiagi, Phys. Lett. B 40, 566 (1972); Nuovo Cimento B 12, 20 (1972); G. 't Hooft and M. Veltman, Nucl. Phys. B 44, 189 (1972).

[5]See for example J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).

[6]H. de Vega and F. Schaposnik, J. of Math. Phys. 15, 1998 (1974).

[7]Some particular cases of CO in dimensional regularization have already been considered in Ref. 9.

[8]W. Zimmermann, Ann. Phys. 77, 536 (1973).

[9]C.G. Bollini and J.J. Giambiagi, Acta Phys. Austriaca 38, 210 (1972); H. Girotti, Nucl. Phys. B 64, 244 (1973); H. de Vega and H. Girotti, Nucl. Phys. B 79, 77 (1974).

[10]N.N. Bogoliubov and D.V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959).

Theory of irradiance distribution function in turbulent media—cluster approximation

K. Furutsu

Radio Research Laboratories, Koganei-shi, Tokyo 184, Japan
(Received 18 November 1975)

The equations of the correlation functions of irradiance in turbulent media are formally solved for arbitrary order by use of an operator method. A cluster approximation is then applied to the fourth and higher order moments of irradiance to express those in terms of the lower order moments. The values of moments predicted by this approximation show a very good agreement with the recent experimental values observed up to the fourth order moment of irradiance. The irradiance distribution function is then analytically derived from the obtained expression of the moments of irradiance and is found to have the following features: (1) It is close to the Gaussian distribution with respect to the logarithm of irradiance, (2) it has a threshold value for irradiances giving nonvanishing probability, and (3) it has a very small but sharp distribution of the form of the δ function at the threshold value. The spectrum of medium as well as the conditions of the initial wave, e.g., of whether it is a plane wave or is a beam wave do not directly enter in the expression but appear only through the first three moments of irradiance. The condition of applicability of the cluster approximation is also discussed in some details based on the Kolmogorov spectrum of turbulence.

1. INTRODUCTION

A predominant feature of the scattering of light waves in turbulent media is that it is essentially the forward scattering so that, if the coordinate system $(x)=(z,\mathbf{x})$ is introduced with the z axis taken in the main direction of wave propagation, the (scalar) wavefunction $\psi(x)\exp(-ikz)$ is sufficiently described by the parabolic wave equation of the form

$$\left[i2k\frac{\partial}{\partial z}-\left(\frac{\partial}{\partial \mathbf{x}}\right)^2-k^2\epsilon(x)\right]\psi(x)=0. \tag{1.1}$$

Here, $\epsilon(x)$ is the fluctuating part of the square of the refractive index and is usually very small compared to the unity. Thus, the mathematical procedure is reduced first to assume the Gaussian statistics for $\epsilon(x)$ with a suitably given correlation function

$$D(x,x')=\langle\epsilon(x)\epsilon(x')\rangle,\quad \langle\epsilon(x)\rangle=0, \tag{1.2}$$

and then to ask the various statistical informations of waves in terms of the moments of wavefunction $M_{\nu\mu}$, defined by

$$M_{\nu\mu}(x_1,\ldots,x_\nu;\,y_1,\ldots,y_\mu)=\langle\psi^*(x_1)\circ\circ\circ\,\psi^*(x_\nu)\psi(y_1)\circ\circ\circ\,\psi(y_\mu)\rangle. \tag{1.3}$$

Here (y) denotes the space coordinates of the original wavefunction whereas (x) denotes those of the complex-conjugate wavefunction.

The equations satisfied by the moment $M_{\nu\mu}$ have been obtained by various authors[1–7] and have been given by various expressions, most of them being equivalent. In the special case of $\nu=\mu$ in (1.3), it is particularly convenient to introduce the relative coordinate system, e.g.,

$$\mathbf{r}_n=\mathbf{y}_n-\mathbf{x}_n,\quad \rho_n=\tfrac{1}{2}(\mathbf{y}_n+\mathbf{x}_n); \tag{1.4}$$

then the equation of $M_{\nu\nu}$ is found to be given in the form[1]

$$\left[k^{-1}\frac{\partial}{\partial z}+\sum_{n=1}^{\nu}\left(\frac{i}{k^2}\frac{\partial}{\partial \mathbf{r}_n}\circ\frac{\partial}{\partial \rho_n}+V(\mathbf{r}_n)\right)+\sum_{n>m=1}^{\nu}V_I(\mathbf{r}_n,\mathbf{r}_m,\rho_{nm})\right]\times M_{\nu\nu}(\mathbf{r}_1,\ldots,\mathbf{r}_\nu;\,\rho_1,\ldots,\rho_\nu;\,z)=0, \tag{1.5}$$

where

$$V(\mathbf{x})=\tfrac{1}{2}k\int_0^\infty dz[D(z,0)-D(z,\mathbf{x})], \tag{1.6}$$

$$\begin{aligned}V_I(\mathbf{r}_1,\mathbf{r}_2,\rho_{12})&=V(\mathbf{x}_1-\mathbf{y}_2)+V(\mathbf{y}_1-\mathbf{x}_2)-V(\mathbf{x}_1-\mathbf{x}_2)-V(\mathbf{y}_1-\mathbf{y}_2)\\&=V(\rho_{12}-\tfrac{1}{2}(\mathbf{r}_1+\mathbf{r}_2))+V(\rho_{12}+\tfrac{1}{2}(\mathbf{r}_1+\mathbf{r}_2))\\&\quad-V(\rho_{12}+\tfrac{1}{2}(\mathbf{r}_2-\mathbf{r}_1))-V(\rho_{12}+\tfrac{1}{2}(\mathbf{r}_1-\mathbf{r}_2)).\end{aligned} \tag{1.7}$$

Equation (1.5) is valid on the condition that the wave amplitude will not appreciably change within the range of correlation distance of the fluctuating medium.

If the Kolmogorov spectrum[8,9] is assumed

$$V(\mathbf{r})\sim\begin{cases}c'(k\mathbf{r})^2, & |k_m\mathbf{r}|\ll 1, \quad (1.8a)\\ c|k\mathbf{r}|^{5/3}, & |k_m\mathbf{r}|\gg 1, \quad (1.8b)\end{cases}$$

where c' and c are nondimensional constants and k_m^{-1} is the minimum length associated with the index of refraction fluctuation [refer to (1.22)].

The solutions of Eq. (1.5) have also been rigorously[2,10,11] obtained for the first order moment of irradiance M_{11} and also for the second order moment M_{22} by a perturbative method[12–14] or by a numerical method.[15,16] As the result, the saturation phenomenon of the irradiance scintillation seems to have been clarified qualitatively at least. In order to find the irradiance distribution function, on the other hand, it is necessary to obtain the moments of irradiance for all orders. In the present stage, however, the exact solutions of Eq. (1.5) for arbitrary order ν were obtained only for the model (1.8a) and the result is that,[1] if the initial Gaussian beam at $z=0$ is given in the form

$$I(z,\rho)=I_0 f^{-2}\exp[-(bf)^{-2}\rho^2],$$
$$f^2=1+(z/kb^2)^2, \tag{1.9}$$

I_0 and b being constants, then the νth order moment of irradiance $M_{\nu\nu}$ is given by

$$M_{\nu\nu}(z)=(I_0 f^{-2})^{\nu}(1+\nu\sigma_E)^{-1}$$
$$\times\exp[-\nu(1+\nu\sigma_E)^{-1}(bf)^{-2}\rho^2],$$
$$\sigma_E=\tfrac{4}{3}c'k(bf)^{-2}z^3, \tag{1.10}$$
$$\mathbf{r}_1=\mathbf{r}_2=\cdots=\mathbf{r}_\nu=0,\quad \rho_1=\rho_2=\cdots=\rho_\nu=\rho.$$

The probability density function $P(I)$ of the irradiance $I=\psi^*\psi$ can be expressed in terms of the moments of irradiance $M_{\nu\nu}$ as follows: If $M_{\nu\nu}$ has the asymptotic form

$$M_{\nu\nu}\lesssim\exp(\nu E_c),\quad \nu\sim\infty, \tag{1.11}$$

E_c being a real constant, and analytic on the right half-plane of ν, then

$$P(I)=(1/2\pi i)\int_{-i\infty-\epsilon}^{i\infty-\epsilon}d\nu\, I^{-\nu-1}M_{\nu\nu},\quad \epsilon>0, \tag{1.12}$$

where ϵ is an infinitesimal positive number.[17,18]

In terms of the notation

$$E=\log(If^2/I_0),\quad E_0=(\rho/bf)^2,$$

the substitution of (1.10) in (1.12) yields

$$P(E)=P(I)\frac{dI}{dE}=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}d\nu(1+\nu\sigma_E)^{-1}$$
$$\times\exp[-\nu(1+\nu\sigma_E)^{-1}E_0-\nu E], \tag{1.13}$$

whose integrand has an essentially singular pole at $\nu=-\sigma_E^{-1}$. The above integral can be evaluated exactly and is expressed by the Rice–Nakagami distribution with respect to the log irradiance E,

$$P(E)=\begin{cases}\sigma_E^{-1}\exp[-(-E_0+E)/\sigma_E]I_0[2(-EE_0)^{1/2}/\sigma_E], & E<0\\ 0,\quad E>0,\end{cases} \tag{1.14}$$

where $I_0(x)$ is the modified Bessel function of zeroth order, and of course

$$\int_{-\infty}^{0}dE\,P(E)=1. \tag{1.15}$$

The assumption (1.8a) corresponds to an atmospheric turbulence model consisting of thin refractive wedges with infinite extent, which gives rise to a random bending of narrow optical beams in the course of wave propagation. Thus, the irradiance distribution (1.14) exactly gives the distribution when the optical beam is in the state of "spot dancing."[19-22]

When the moment $M_{\nu\nu}$ satisfies the asymptotic condition (1.11) and is analytic on the right half-plane of ν, there is no probability for the irradiances in the range of $\log I>E_c$, as it follows from the integral representation (1.12). Therefore, for arbitrary $E_0>E_c$, the nth order moment of log irradiance defined by

$$\langle(\log I-E_0)^n\rangle=\int_0^{\exp(E_0)}dI\,P(I)(\log I-E_0)^n \tag{1.16}$$

is expressed, by use of the new variable of integration $x=\log I-E_0$, as

$$(1/2\pi i)\int_{-i\infty-\epsilon}^{i\infty-\epsilon}d\nu\,M_{\nu\nu}\exp(-\nu E_0)\int_{-\infty}^{0}dx\,x^n\exp(-\nu x)\qquad \mathrm{Re}[\nu]=-\epsilon<0$$
$$=n!\,\frac{(-)}{2\pi i}\int_{-i\infty-\epsilon}^{i\infty-\epsilon}d\nu\,\nu^{-n-1}M_{\nu\nu}\exp(-\nu E_0),\qquad E_0>E_c \tag{1.17}$$

which can be reduced to the infinitesimal (clockwise) contour integral around the origin of $\nu=0$. Thus

$$\langle(\log I-E_0)^n\rangle=(n!/2\pi i)\oint d\nu\,\nu^{-n-1}M_{\nu\nu}\exp(-\nu E_0), \tag{1.18}$$

$$M_{\nu\nu}\exp(-\nu E_0)=\sum_{n=0}^{\infty}\frac{1}{n!}\langle(\log I-E_0)^n\rangle\nu^n. \tag{1.19}$$

Although the relation (1.18) was derived on the condition $E_0>E_c$, both its sides are polynomials of E_0 of order n and therefore the relation holds also for arbitrary $E_0\gtrless E_c$. Applying the expression (1.10) in (1.19), the first few moments of the log irradiance E are found to be

$$\langle E\rangle=-(E_0+\sigma_E),\quad \langle(E-\langle E\rangle)^2\rangle=2E_0\sigma_E+\sigma_E^2,$$
$$\langle(E-\langle E\rangle)^3\rangle=-(6E_0\sigma_E^2+2\sigma_E^3),\ \text{etc.} \tag{1.20}$$

Here, since $\sigma_E\propto c'z^3\to\infty$ as $z\to\infty$ or $c'\to\infty$, the variance of E infinitely grows with the distance of wave propagation and therefore does not saturate.

As is stated above, the moment equation (1.5) has been solved exactly only for the model (1.8a) which merely gives rise to a tilt of the phase front of a wave and consequently to the spot dancing of narrow optical beams, but not to the production of irradiance fluctuations. Therefore, for many actual conditions of optical beams which are not too narrow, the behavior of $V(\mathbf{r})$ in the range of (1.8b) is expected to make important contributions to the moments of irradiance for all orders.

In this paper, we shall develop the theory first for $V(\mathbf{r})$ in the general form

$$V(\mathbf{r})=c|k\mathbf{r}|^{\alpha},\quad 2>\alpha>1, \tag{1.21}$$

rather than (1.8b). Here the constant c can be expressed in terms of the structure constant C_n^2 characterizing the strength of the turbulence, as[12]

$$-c=\pi 2^{-\alpha-1}\Gamma(1+\alpha)\Gamma^{-2}(1+\alpha/2)\cot(\pi\alpha/2)k^{1-\alpha}C_n^2. \tag{1.22}$$

The pronounced features of the irradiance fluctuations resulted from the model (1.21) are expected to be the saturation of the variance of irradiance and also the irradiance distribution which is close to the Gaussian distribution with respect to the logarithm of irradiance. The first point can be confirmed by solving the second order moment equation of (1.5) and this has been done numerically and analytically to such an extent that no doubt exists for the saturation itself, finding that $\langle I^2\rangle/\langle I\rangle^2\to 2$ as the strength of turbulence increases in the case of plane waves. On the other hand, for the second point, the moment equation (1.5) must be solved for all orders and nothing has been successfully tried based on the moment equations.

2. OPERATOR METHOD FOR THE FIRST- AND SECOND-ORDER MOMENTS OF IRRADIANCE

We shall first consider the simplest case of M_{11}, whose equation, according to (1.5), has the form[23]

$$k^{-1}\frac{\partial}{\partial z}M_{11}(z)=H_1M_{11}(z),\quad H_1=T_1-V_1,$$
$$T_1=-ik^{-2}\frac{\partial}{\partial \mathbf{r}_1}\cdot\frac{\partial}{\partial \rho_1},\qquad V_1=V(\mathbf{r}_1), \tag{2.1}$$

where the coordinates other than z have been suppressed in $M_{11}(z)$. The formal solution of (2.1) is expressed by

$$M_{11}(z)=\exp(kH_1z)M_{11}(0),$$
$$\exp(kH_1z)=\exp(kT_1z)U_1(z), \tag{2.2}$$

where the last equation defines the operator $U_1(z)$. The substitution of (2.2) in (2.1) yields the equation of $U_1(z)$ of the form

$$k^{-1}\frac{\partial}{\partial z}U_1(z)=-\hat{V}(z)U_1(z), \tag{2.3}$$

where, in terms of the notation $[A,B]=AB-BA$,

$$\hat{V}(z)\equiv\exp(-kT_1z)V(\mathbf{r}_1)\exp(kT_1z)=V(\hat{r}_1(z)),$$
$$\hat{r}_1(z)=\exp(-kT_1z)\mathbf{r}_1\exp(kT_1z)$$
$$=\mathbf{r}_1-kz[T_1,\mathbf{r}_1]+\frac{1}{2!}(kz)^2[T_1,[T_1,\mathbf{r}_1]]+\cdots$$
$$=\mathbf{r}_1+ik^{-1}z\frac{\partial}{\partial\rho_1}. \tag{2.4}$$

The solution of (2.3) for the initial condition $U_1(0)=1$ is, obviously,

$$U_1(z)=\exp[-k\int_0^z dz'\hat{V}(z')]. \tag{2.5}$$

Thus, in terms of the moment in free space, $M_{11}^0(z)$, given by

$$M_{11}^0(z)=\exp(kT_1z)M_{11}(0), \tag{2.6}$$

Eq. (2.2) provides the expression of moment as

$$M_{11}(z)=\exp(kT_1z)U_1(z)\exp(-kT_1z)M_{11}^0(z)$$
$$=\exp[-k\int_0^z dz'\hat{V}(z'-z)]M_{11}^0(z), \tag{2.7}$$

where use has been made of the definition of $\hat{r}(z)$ in (2.4). The integrand of the last expression in (2.7) is a functional only of $\hat{r}(z)$ and therefore the evaluation of the moment is simple by means of the Fourier transformation with respect to ρ_1.

The second order moment equation of (1.5) can be given in the form

$$k^{-1}\frac{\partial}{\partial z}M_{22}(z)=(H_1+H_2-V_{12})M_{22}(z). \tag{2.8}$$

Here

$$H_j=T_j-V_j=-ik^{-2}\frac{\partial}{\partial r_j}\circ\frac{\partial}{\partial\rho_j}-V(\mathbf{r}_j),\quad j=1,2,$$
$$V_{12}=V_{21}=V_I(\mathbf{r}_1,\mathbf{r}_2,\rho_{12}),\quad \rho_{12}=\rho_1-\rho_2. \tag{2.9}$$

As in (2.2), if the solution of (2.8) is given in the form

$$M_{22}(z)=\exp[k(H_1+H_2)z]S_{12}(z)M_{22}(0), \tag{2.10}$$

then, the operator $S_{12}(z)$ is the solution of

$$k^{-1}\frac{\partial}{\partial z}S_{12}(z)=-\tilde{V}_{12}(z)S_{12}(z),\quad S_{12}(0)=1. \tag{2.11}$$

Here

$$\tilde{V}_{12}(z)=\exp[-k(H_1+H_2)z]V_I(\mathbf{r}_1,\mathbf{r}_2,\rho_{12})\exp[k(H_1+H_2)z]$$
$$=V_I(\tilde{\mathbf{r}}_1(z),\ \tilde{\mathbf{r}}_2(z),\ \tilde{\rho}_{12}(z)) \tag{2.12}$$

with

$$\tilde{\mathbf{r}}_j(z)=\exp(-kH_jz)\mathbf{r}_j\exp(kH_jz),$$
$$\tilde{\rho}_j(z)=\exp(-kH_jz)\rho_j\exp(kH_jz). \tag{2.13}$$

We shall see, in the following, that the operators $\tilde{V}_{12}(z)$ and $\tilde{V}_{12}(z')$ are not commutable for $z\neq z'$, and therefore the solution of (2.11) should be given, instead of (2.5), in the form

$$S_{12}(z)=P\exp[-k\int_0^z dz'\tilde{V}_{12}(z')], \tag{2.14}$$

where the symbol P means that, for any two z-dependent operators $A(z')$ and $B(z'')$,

$$P[A(z')B(z'')]=\begin{cases}A(z')B(z''), & z'>z'',\\ B(z'')A(z'), & z'<z''.\end{cases} \tag{2.15}$$

Using the expression of $\exp(kH_jz)$ in (2.2), $\tilde{\rho}_j(z)$ defined by (2.13) is expressed as

$$\tilde{\rho}_j(z)=U_j^{-1}(z)\exp(-kT_jz)\rho_j\exp(kT_jz)U_j(z)$$
$$=U_j^{-1}(z)\hat{\rho}_j(z)U_j(z), \tag{2.16}$$

where, as in (2.4),

$$\hat{\rho}_j(z)=\exp(-kT_jz)\rho_j\exp(kT_jz)$$
$$=\rho_j+ik^{-1}z\frac{\partial}{\partial\mathbf{r}_j}, \tag{2.17}$$

with the commutation relations

$$[\hat{\rho}_j(z),\ \hat{\rho}_k(z')]=[\hat{r}_j(z),\ \hat{r}_k(z')]=0,$$
$$[\hat{r}_j(z),\ \hat{\rho}_k(z')]=ik^{-1}(z-z')\delta_{jk}. \tag{2.18}$$

Here, the different space components of $\hat{r}_j(z)$ and $\hat{\rho}_j(z)$ are also commutable.

Thus, by use of the commutation relations (2.18) and the expression (2.5) for $U_j(z)$, Eq. (2.16) gives the expression

$$\tilde{\rho}_j(z)=\hat{\rho}_j(z)+k\int_0^z dz'[V(\hat{r}_j(z')),\ \hat{\rho}_j(z)]$$
$$=\hat{\rho}_j(z)+i\int_0^z dz'(z'-z)\frac{\partial}{\partial\mathbf{r}_j}V(\hat{r}_j(z')). \tag{2.19}$$

On the other hand, since $U_j(z)$ is a functional only of $\hat{r}_j(z)$ which are commutable to each other for all values of z,

$$\tilde{r}_j(z)=U_j^{-1}(z)\hat{r}_j(z)U_j(z)=\hat{r}_j(z). \tag{2.20}$$

As in (2.16) for $\tilde{\rho}_j(z)$, we have another expression of $\tilde{V}_{12}(z)$,

$$\tilde{V}_{12}(z)=U_2^{-1}U_1^{-1}(z)\hat{V}_{12}(z)U_1U_2(z)$$

$$= \exp[k\int_0^z dz'\{\hat{V}_1(z')+\hat{V}_2(z')\}]\hat{V}_{12}(z) \times \exp[-k\int_0^z dz'\{\hat{V}_1(z')+\hat{V}_2(z')\}], \tag{2.21}$$

which is often convenient for actual evaluations [see Appendix A].

Here we introduce the notation

$$I_j(z)=\psi_j^*(z)\psi_j(z)\equiv\psi^*(z,\mathbf{x}_j)\psi(z,\mathbf{y}_j),\quad j=1,2, \tag{2.22}$$

$\mathbf{x}_j$ and $\mathbf{y}_j$ being expressed by the relative coordinates $(\mathbf{r}_j,\rho_j)$ according to (1.4), and also for any functional Q_{12} of the operators $\tilde{r}_j(z)$ and $\tilde{\rho}_j(z)$ $(j=1,2)$, the notation

$$\langle Q_{12}\rangle_L=\frac{\exp[k(H_1+H_2)L]Q_{12}I_1(0)I_2(0)\,|_{\mathbf{r}_1=\mathbf{r}_2=\rho_{12}=0}}{\exp[k(H_1+H_2)L]I_1(0)I_2(0)\,|_{\mathbf{r}_1=\mathbf{r}_2=\rho_{12}=0}}. \tag{2.23}$$

Then, $M_{11}(0)=I_1(0)$ and therefore, according to (2.2),

$$\exp[k(H_1+H_2)L]I_1(0)I_2(0)\,|_{\mathbf{r}_1=\mathbf{r}_2=\rho_{12}=0}=\langle I(L)\rangle^2. \tag{2.24}$$

Also, since $M_{22}(0)=I_1(0)I_2(0)$, Eq. (2.10) for $\mathbf{r}_1=\mathbf{r}_2=\rho_{12}=0$ is expressed as

$$M_{22}(L)\,|_{\mathbf{r}_1=\mathbf{r}_2=\rho_{12}=0}=\langle S_{12}(L)\rangle_L\langle I(L)\rangle^2. \tag{2.25}$$

In order to find the equation of $\langle S_{12}(z)\rangle_L$, we put, with reference to (2.14),

$$S_{12}(z)=S'_{12}(z)\exp[-k\int_0^z dz'\langle\tilde{V}_{12}(z')\rangle_L], \tag{2.26}$$

and introduce the quantity

$$\Delta\tilde{V}_{12}(z)=\tilde{V}_{12}(z)-\langle\tilde{V}_{12}(z)\rangle_L,\quad \langle\Delta\tilde{V}_{12}(z)\rangle_L=0. \tag{2.27}$$

Then, from (2.11), the equation of $S'_{12}(z)$ is found to be

$$k^{-1}\frac{\partial}{\partial z}S'_{12}(z)=-\Delta\tilde{V}_{12}(z)S'_{12}(z). \tag{2.28}$$

Again, by putting

$$S'_{12}(z)=\langle S'_{12}(z)\rangle_L+\Delta S'_{12}(z), \tag{2.29}$$

Eq. (2.28) gives

$$k^{-1}\frac{\partial}{\partial z}\langle S'_{12}(z)\rangle_L=-\langle\Delta\tilde{V}_{12}(z)\Delta S'_{12}(z)\rangle_L,$$

$$\left\{k^{-1}\frac{\partial}{\partial z}+\Delta\tilde{V}_{12}(z)\right\}\Delta S'_{12}(z)-\langle\Delta\tilde{V}_{12}(z)\Delta S'_{12}(z)\rangle_L =-\Delta\tilde{V}_{12}(z)\langle S'_{12}(z)\rangle_L. \tag{2.30}$$

To the first order of approximation,

$$\Delta S'_{12}(z)\sim-k\int_0^z dz'\Delta\tilde{V}_{12}(z')\langle S'_{12}(z')\rangle_L, \tag{2.31}$$

and therefore the substitution into the first equation of (2.30) yields

$$k^{-1}\frac{\partial}{\partial z}\langle S'_{12}(z)\rangle_L\sim k\int_0^z dz'\langle\Delta\tilde{V}_{12}(z)\Delta\tilde{V}_{12}(z')\rangle_L\langle S'_{12}(z')\rangle_L, \tag{2.32}$$

which provides an integro-differential equation for the ordinary function $\langle S'_{12}(z)\rangle_L$.

The integral representation of $\langle\tilde{V}_{12}(z)\rangle_L$ was already obtained with the explicit expression for small values of z as well as the asymptotic expression for large values,[12] and these are also given in Appendices A, B, and C. The corresponding expression for the function $\langle\tilde{V}_{12}(z)\tilde{V}_{12}(z')\rangle_L$ in the case of plane waves is also treated.

A formally exact equation corresponding to Eq. (2.32) can be obtained as follows: We first introduce the symbolic vectors $|I\rangle$ and $\langle 0|$ defined by

$$|I\rangle=I_1(0)I_2(0),\quad\langle 0|\mathbf{r}_j=\langle 0|\rho_{12}=0,\quad j=1,2, \tag{2.33}$$

and also an "averaging" operator A by

$$A=|I\rangle\langle 0|\exp[k(H_1+H_2)L]/\langle 0|\exp[k(H_1+H_2)L]|I\rangle,$$
$$A^2=A,\quad\langle Q_{12}\rangle_L=\mathrm{Tr}[AQ_{12}], \tag{2.34}$$

where, in the last equation, Tr denotes diagonal summation with respect to all the coordinates $\mathbf{r}_j$ and ρ_j $(j=1,2)$. Then, the operation of $|I\rangle$ from the right-hand side of Eq. (2.30) yields the expression

$$\left[k^{-1}\frac{\partial}{\partial z}+(1-A)\Delta\tilde{V}_{12}(z)\right]\Delta S'_{12}(z)|I\rangle =-\Delta\tilde{V}_{12}(z)|I\rangle\langle S'_{12}(z)\rangle_L. \tag{2.35}$$

In terms of the function $S_A(z_2,z_1)$ defined by

$$S_A(z_2,z_1)=P\exp[-k\int_{z_1}^{z_2}dz'(1-A)\Delta\tilde{V}_{12}(z')],$$
$$AS_A(z_2,z_1)=A, \tag{2.36}$$

the solution of (2.35) is expressed as

$$\Delta S'_{12}(z)|I\rangle=-k\int_0^z dz' S_A(z,z')\Delta\tilde{V}_{12}(z')|I\rangle\langle S'_{12}(z')\rangle_L. \tag{2.37}$$

Thus, the use of the expression (2.37) in the first equation of (2.30) yields

$$k^{-1}\frac{\partial}{\partial z}\langle S'_{12}(z)\rangle_L=k\int_0^z dz'\langle\Delta\tilde{V}_{12}(z)S_A(z,z')\Delta\tilde{V}_{12}(z')\rangle_L \times\langle S'_{12}(z')\rangle_L, \tag{2.38}$$

which is an exact version of (2.32).

When solving Eq. (2.38), various methods used in other fields of physics may be referred[24-27] but we shall not go into further in this paper.

On the other hand, directly from Eqs. (2.26) and (2.28), we obtain

$$\langle S_{12}(L)\rangle_L=\exp[-k\int_0^L dz\langle\tilde{V}_{12}(z)\rangle_L +k^2\int_0^L dz_2\int_0^{z_2}dz_1\langle\Delta\tilde{V}_{12}(z_2)\Delta\tilde{V}_{12}(z_1)\rangle_L -k^3\int_0^L dz_3\int_0^{z_3}dz_2\int_0^{z_2}dz_1\langle\Delta\tilde{V}_{12}(z_3)\Delta\tilde{V}_{12}(z_2)\tilde{V}_{12}(z_1)\rangle_L +\cdots]. \tag{2.39}$$

3. THIRD- AND HIGHER ORDER MOMENTS OF IRRADIANCE AND THE CLUSTER APPROXIMATION

In the general case of order ν, the moment equation (1.5) is expressed as

$$k^{-1}\frac{\partial}{\partial z}M_{\nu\nu}(z)=\left[\sum_{j=1}^{\nu}H_j-\sum_{j>k=1}^{\nu}V_{jk}\right]M_{\nu\nu}(z), \tag{3.1}$$

where H_j and V_{jk} are defined by (2.9). As in the previous section, we put

$$M_{\nu\nu}(z) = \exp\left[\sum_{j=1}^{\nu} kH_j z\right] U_\nu(z) \prod_{j=1}^{\nu} I_j(0). \qquad (3.2)$$

Then, since V_{jk} depends only on the variables $\mathbf{r}_j$, $\mathbf{r}_k$, and ρ_{jk}, the substitution of (3.2) in (3.1) gives the equation of $U_\nu(z)$ as

$$k^{-1}\frac{\partial}{\partial z} U_\nu(z) = -\sum_{j>k=1}^{\nu} \tilde{V}_{jk}(z) U_\nu(z), \qquad (3.3)$$

where $\tilde{V}_{jk}$ is defined by (2.12). Now, the formal solution of (3.3) is expressed, in terms of S_{jk} in (2.14), by

$$U_\nu(z) = P\exp\left[-\sum_{j>k=1}^{\nu}\int_0^z dz'\tilde{V}_{jk}(z')\right]$$

$$= P\prod_{j>k=1}^{\nu} S_{jk}. \qquad (3.4)$$

As in (2.23), if we introduce the notation

$$\langle Q\rangle_L = \frac{\prod_{j=1}^{\nu}\exp(kH_jL)Q\prod_{k=1}^{\nu}I_k(0)\,|_{\mathbf{r}_1=\cdots=\mathbf{r}_\nu=0,\,\rho_{12}=\cdots=\rho_{jk}=0}}{\prod_{j=1}^{\nu}\exp(kH_jL)I_j(0)\,|_{\mathbf{r}_1=\cdots=\mathbf{r}_\nu=0,\,\rho_{12}=\cdots=\rho_{jk}=0}}, \qquad (3.5)$$

we have from (3.2) and (3.4) the expression of $M_{\nu\nu}(L)$ for $\mathbf{r}_j = \rho_{jk} = 0$ $(j, k = 1, 2, \cdots)$

$$M_{\nu\nu}(L) = \left\langle P\prod_{j>k=1}^{\nu} S_{jk}\right\rangle_L \langle I(L)\rangle^\nu, \quad P^2 = P,$$
$$\mathbf{r}_1 = \mathbf{r}_2 = \cdots = \mathbf{r}_\nu = 0, \quad \rho_1 = \rho_2 = \cdots = \rho_\nu = \rho. \qquad (3.6)$$

Here, we first define the quantity $K_1(jk)$ by

$$\langle S_{jk}(L)\rangle_L = \exp[K_1(jk)] \qquad (3.7)$$

with the symmetry

$$K_1(12) = K_1(23) = \cdots = K_1(jk) = K_1, \qquad (3.8)$$

which holds because of the symmetry in the definitions of $\langle\cdots\rangle_L$ and of $\tilde{V}_{jk}$. We then define the quantity $K_2(12,23)$ by

$$\langle PS_{12}S_{23}\rangle_L = \exp[K_1(12) + K_1(23) + K_2(12,23)], \qquad (3.9)$$

with the symmetry corresponding to (3.8). Here we note that, when any two "pair bonds" are disconnected, e.g.,

$$\langle PS_{12}S_{34}\rangle_L = \langle S_{12}\rangle_L\langle S_{34}\rangle_L = \exp[K_1(12) + K_1(34)], \qquad (3.10)$$

it follows that $K_2(12,34) = 0$.

In the same way,[28] $K_3(12,23,34)$ is defined by

$$\langle PS_{12}S_{23}S_{34}\rangle_L = \exp[K_1(12) + K_1(23) + K_1(34) + K_2(12,23) + K_2(23,34) + K_3(12,23,34)],$$
$$K_2(12,34) = 0, \qquad (3.11)$$

and therefore $K_3(12,23,34)$ is that term which depends on the three connected pair bonds 12, 23, and 34.

Generally, if, for any two sets A and B of products of S_{jk}, $(j, k = 1, 2, \cdots)$,

$$\langle PA\rangle_L = \exp[K(A)], \quad \langle PB\rangle_L = \exp[K(B)],$$

then, $K(A,B)$ is defined by

$$\langle P[AB]\rangle_L = \exp[K(A) + K(B) + K(A,B)].$$

Therefore $K(A,B) = 0$ whenever the sets A and B are not correlated, i.e.,

$$\langle P[AB]\rangle_L = \langle PA\rangle_L\langle PB\rangle_L = \exp[K(A) + K(B)].$$

In this way, we find from the expression (3.2) and (3.4) for $M_{\nu\nu}$ and the definition (3.5) that

$$m_2 \equiv \langle I^2(L)\rangle/\langle I(L)\rangle^2 = \langle S_{12}\rangle_L = \exp[K_1(12)],$$
$$m_3 = \langle I^3(L)\rangle/\langle I(L)\rangle^3 = \langle PS_{12}S_{23}S_{31}\rangle_L$$
$$= \exp[3K_1(12) + 3K_2(12,23) + K_3(12,23,31)], \qquad (3.12)$$
$$m_\nu = \langle I^\nu(L)\rangle/\langle I(L)\rangle^\nu = \langle P\prod_{j>k=1}^{\nu} S_{jk}\rangle_L$$
$$= \exp[\tbinom{\nu}{2}K_1(12) + \tbinom{\nu}{3}\{3K_2(12,23) + K_3(12,23,31)\}$$
$$+ \tbinom{\nu}{4}\{12K_3(12,23,34) + 3K_4(12,23,34,41)$$
$$+ 6K_5(12,23,34,41,31) + K_6(12,23,34,41,31,24)\}$$
$$+ \cdots], \qquad (3.13)$$

where we have used the existing symmetries corresponding to (3.8) and the fact that K_ν vanishes whenever the involved pair bonds are disconnected. The essential points are to express m_ν in terms of the lower order ones by use of (3.12) and (3.13), i.e.,

$$m_\nu = \exp[\tbinom{\nu}{2}\log m_2 + \tbinom{\nu}{3}\log(m_3/m_2^3)$$
$$+ \tbinom{\nu}{4}\log(m_4 m_2^6/m_3^4) + \cdots], \qquad (3.14)$$

and to expect much better convergence of series than that of the series corresponding to (2.39). Here, although the improvement of convergence depends on the particular natures of H_j and V_{jk} in (3.1) and is to be proved mathematically, it is anticipated in most cases by a physical intuition. In the present case, we shall keep only the first two terms in the series (3.14) and compare the results first with the recent experimental values to see its validity. Thus, from (3.14), we have the approximate expression

$$m_\nu \sim \exp[\tfrac{1}{2}\nu(\nu-1)K_1\{1 - (\nu-2)\Delta'\}], \qquad (3.15)$$

$$K_1 = \log m_2,$$
$$\Delta' = -\{K_2(12,23) + \tfrac{1}{3}K_3(12,23,31)\}/K_1(12)$$
$$= 1 - \tfrac{1}{3}\log m_3/\log m_2. \qquad (3.16)$$

Here, if $|\Delta'| \ll 1$, as is assumed, Eq. (3.15) is also expressed by

$$m_\nu \sim \exp[\tfrac{1}{2}\nu(\nu-1)K_1\{1 + (\nu-2)\Delta\}^{-1}],$$
$$\Delta = \Delta'(1-\Delta')^{-1} \sim \Delta', \qquad (3.17)$$

which gives the exact values as those given by (3.15) for the orders of $\nu = 1$, 2, and 3 and, to the first order of approximation of Δ, also for the higher orders.

In Table I, the recent experimental values of m_ν observed by Gracheva *et al.*[29] are shown along with the corresponding theoretical values by (3.17). The parameter Δ critically depends on the value of m_3, and therefore two values of Δ are prepared, one being determined by the value of m_3 and the other by that of m_4. The first value gives rise to the error of 7.2% for m_4 whereas the second one to the error of only 1.8% for m_3. Note that $\Delta > 0$ and $\Delta \ll 1$. The first point can be

TABLE I. Comparison of the experimental values of $m_\nu = \langle I^\nu\rangle / \langle I\rangle^\nu$ observed by Gracheva *et al.*[29] for $\beta_0^2 = 0.31 c_\epsilon^2 k^{7/6} L^{11/6} \gtrsim 25$ with the corresponding theoretical values given by (3.17). $K_1 = \log m_2 = 0.7793$ whereas the parameter Δ has the two different values according to whether it is determined by the experimental value of m_3 or of m_4. The resulted errors of m_4 or m_3 are shown in parenthesis in %. The two values in the last line and column are the ratios of the predicted theoretical values to the log-normal one.

ν \ m_ν	Experimental	Theoretical & Error (%) $\Delta = 0.0262$	$\Delta = 0.0346$	Log-normal ($\Delta = 0$)	Experimental/ Log-normal
2	2.18	2.18	2.18	2.18	1.0
3	9.76	9.76	9.58 (1.8%)	10.36	0.94
4	79.32	85.0 (7.2%)	79.32	107.32	0.74
5	xxxxx	1374	1165	2424	(0.57;0.48)

confirmed also theoretically to the first order of approximation which is valid when the distance of wave propagation is short enough [see Eqs. (3.20) and (3.21)].

It is also worthwhile to check the third term in the series of (3.14): according to the experimental values given in Table I,

$$\log(m_4 m_2^6/m_3^4) = 4.3735 + 6\times 0.7793 - 4\times 2.2783 = -0.064,$$

which is just 0.7% of the cancelled values. Therefore the above value will be certainly within the range of experimental error, and the large cancellation will mean that the third term is very small to be neglected or $m_4 \sim m_3^4/m_2^6$ which is exactly the expression obtained by the first two terms of the series or by (3.15). Thus, we may conclude, to the extent of the existing experimental data, that the agreement of the experimental and theoretical values is very good.

The explicit expression of $K_1(12)$ is given by the series in (2.39) and therefore, to the second order of $\tilde{V}_{12}$,

$$K_1(12) \sim -k\int_0^L dz\,\langle \tilde{V}_{12}(z)\rangle_L + k^2\int_0^L dz_2\int_0^{z_2} dz_1\,\langle\Delta\tilde{V}_{12}(z_2)\Delta\tilde{V}_{12}(z_1)\rangle_L. \tag{3.18}$$

In the same way,

$$K_2(12,23) \sim 2k^2\int_0^L dz_2\int_0^{z_2} dz_1\,\langle\Delta\tilde{V}_{12}(z_2)\Delta\tilde{V}_{23}(z_1)\rangle_L,$$
$$K_3(12,23,31) \sim -6k^3\int_0^L dz_3\int_0^{z_3} dz_2\int_0^{z_2} dz_1 \times\langle\Delta\tilde{V}_{12}(z_3)\Delta\tilde{V}_{23}(z_2)\Delta\tilde{V}_{31}(z_1)\rangle_L. \tag{3.19}$$

In Appendix C, $K_2(12,23)$ is evaluated for plane waves in the approximation of (3.19) and, according to (C14),

$$K_2(12,23) \sim -0.04\phi^2 < 0, \quad \alpha = 5/3, \quad \phi \ll 1, \tag{3.20}$$

where ϕ is a numerical distance defined by

$$\phi/2 = kLV[(L/k)^{1/2}] = c(kL)^{1+\alpha/2}. \tag{3.21}$$

Here, since $K_3 = 0$ to the second order of $\tilde{V}_{12}$ and $K_1(12) > 0$ according to (C3), it follows from (3.16) that $\Delta' > 0$.

4. IRRADIANCE DISTRIBUTION FUNCTION

The expression (3.17) of m_ν is equivalent to that of (3.15) in that both expressions give the correct values of the moments for the orders of $\nu = 1$, 2, and 3 and, to the first order of Δ for the higher orders. Therefore it follows from (1.19) with $M_{\nu\nu} = m_\nu\langle I\rangle^\nu$ that, to the first order of Δ, both expressions also give the same moments for the logarithm of irradiance of all orders. The essential difference is that, as $\nu\to\infty$, the expression (3.15) tends to $\exp[-\frac{1}{2}\nu^3 K_1\Delta']$ whereas the expression (3.17) tends to $\exp[\frac{1}{2}\nu K_1/\Delta]$ and therefore satisfies the condition of applicability (1.11) for the integral representation (1.12) of the irradiance distribution function $P(I)$. The immediate consequences which resulted from the latter property and the representation (1.12) are: $P(I) = 0$ for $\log(I/\langle I\rangle) > K_1/(2\Delta) \gg 1$, or that the expression (3.17) gives rise to a definite distribution function with a threshold value for the irradiances giving nonvanishing probability, whereas the expression (3.15) gives rise to the unphysical result that the moment of irradiance tends to vanish as $\nu\to\infty$.

In terms of the log irradiance E defined by

$$E = \log(I/\langle I\rangle), \tag{4.1}$$

the substitution of the expression (3.17) with $M_{\nu\nu} = m_\nu\langle I\rangle^\nu$ in (1.12) yields the integral

$$P(E) = P(I)\frac{dI}{dE} = \frac{1}{2\pi i}\int_{-i\infty-\epsilon}^{i\infty-\epsilon} d\nu \times\exp[\tfrac{1}{2}\nu(\nu-1)K_1\{1+(\nu-2)\Delta\}^{-1} - \nu E], \tag{4.2}$$

which provides the distribution function with respect to E. The integrand tends to $\exp(K_1/2\Delta - E)\nu]$ as $\nu\to\infty$ and is analytic on the entire complex plane of ν except at the essentially singular pole at $\nu = 2-\Delta^{-1} < 0$ on the negative real axis, as was exactly the case of the integral (1.13).

By use of the notation

$$\nu_0 = \Delta^{-1} - 2 > 0, \quad a^2 = (1+\nu_0)K_1/\Delta, \quad a > 0,$$
$$E_c = K_1/(2\Delta), \quad b^2 = 2\nu_0(E_c - E) \gtrless 0, \tag{4.3}$$

the integral (4.2) can be expressed in the form

$$P(E) = (1/2\pi i)\int_{-i\infty}^{i\infty} d\nu\,\exp[\tfrac{1}{2}\{b^2(\nu/\nu_0+1) + a^2(\nu/\nu_0+1)^{-1} - a^2 - b^2\}], \tag{4.4}$$

which integrand tends, as $\nu\to\infty$, to

$$\exp[\tfrac{1}{2}\{b^2\nu/\nu_0 - a^2\}] = \exp[-\tfrac{1}{2}a^2 + (E_c - E)\nu]. \tag{4.5}$$

Therefore, upon referring to the result

$$(1/2\pi i)\int_{-i\infty}^{i\infty} d\nu \exp[-\tfrac{1}{2}a^2 + (E_c - E)\nu] = \exp(-\tfrac{1}{2}a^2)\delta(E - E_c), \qquad (4.6)$$

it follows that the integral representation of the new function

$$P'(E) = P(E) - \exp(-\tfrac{1}{2}a^2)\delta(E - E_c) \qquad (4.7)$$

has an integrand which is analytic on the entire ν plane except at the pole of $\nu = -\nu_0$, and has the asymptotic expression proportional to $\nu^{-1}\exp[(E_c - E)\nu]$ for $|\nu| \sim \infty$. Thus, $P'(E)$ vanishes for $E > E_c$ whereas, for $E < E_c$, it is given by $2\pi i$ times the residue value at the pole. Thus, by the introduction of the new variable of integration

$$t = (b/a)(\nu/\nu_0 + 1) \qquad (4.8)$$

in (4.4) and by the aid of the formula

$$(1/2\pi i)\oint dt\, t^{n-1}\exp[\tfrac{1}{2}x(t + t^{-1})] = I_n(x), \quad I_{-n} = I_n, \qquad (4.9)$$

$I_n(x)$ being the modified Bessel function of nth order, we find that

$$P'(E) = \begin{cases} \nu_0(a/b)\exp[-\tfrac{1}{2}(a^2 + b^2)]I_1(ab), & E < E_c \\ 0, & E > E_c \end{cases} \qquad (4.10)$$

where, from (4.3),

$$a = \Delta^{-1}(1-\Delta)^{1/2}K_1^{1/2} \gg 1, \quad \nu_0 = \Delta^{-1} - 2,$$
$$b = \Delta^{-1}(1-2\Delta)^{1/2}K_1^{1/2}(1 - E/E_c)^{1/2} \geq 0, \quad E_c = K_1/(2\Delta). \qquad (4.11)$$

Equation (4.7) with (4.10) provides the explicit expression of the distribution function $P(E)$, and of course

$$\int_{-\infty}^{\infty} dE\, P(E) = 1. \qquad (4.12)$$

In the important range of E where $E/E_c \ll 1$ and $a \sim b \gg 1$, the use of the asymptotic expression

$$I_1(x) \sim (2\pi x)^{-1/2}e^x, \quad x \gg 1 \qquad (4.13)$$

in (4.10) gives the approximate expression

$$P(E) = P'(E) \sim (2\pi K_1)^{-1/2}\exp(-\tfrac{1}{2}F^2), \quad a \sim b \gg 1, \qquad (4.14)$$

where, to the first order of $\Delta \ll 1$,

$$F \equiv a - b \sim K_1^{1/2}[E/K_1 + \tfrac{1}{2} + \tfrac{1}{2}\Delta(E/K_1 - \tfrac{1}{2})(E/K_1 - \tfrac{3}{2})]. \qquad (4.15)$$

The expression (4.14) becomes the log normal distribution $P_0(E)$ for $\Delta = 0$ and, in Fig. 1, the ratio $P(E)/P_0(E)$ is shown against the variable $F|_{\Delta=0} = K_1^{1/2}(E/K_1 + \tfrac{1}{2})$ for $K_1 = 0.7793$ and $\Delta = 0.05$, 0.0346, and 0.015. Here, for curve 2, the parameters have the same values as those given in Table I, and the expression (4.10) was used for the computation.

On the other hand, in the vicinity of the threshold value E_c of the log irradiance, Eq. (4.10) gives

$$P(E) \sim \tfrac{1}{2}(\Delta^{-1} - 2)a^2\exp[-\tfrac{1}{2}(a^2 + b^2)], \quad E \sim E_c, \qquad (4.16)$$

which is very small, however, because of $a \gg 1$.

According to formula (1.19), the central moments of the log irradiance are found to be, to the first order of Δ,

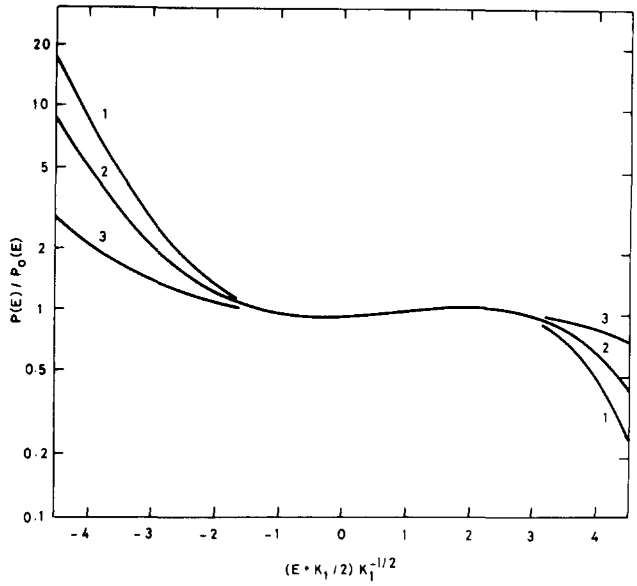


FIG. 1. Ratio of the distribution function of log-irradiance $P(E)$ to the log-normal distribution function $P_0(E)$. $K_1 = 0.7793$ and 1, $\Delta = 0.05$; 2, $\Delta = 0.0346$; 3, $\Delta = 0.015$.

$$\langle E\rangle = -\tfrac{1}{2}(1 + 2\Delta)K_1, \quad \langle(E - \langle E\rangle)^2\rangle = (1 + 3\Delta)K_1,$$
$$\langle(E - \langle E\rangle)^3\rangle = -3\Delta K_1 < 0, \quad \text{etc.} \qquad (4.17)$$

Here, we note that the third moment is again negative as in (1.20) for the previous distribution (1.14).

By use of (4.2), the cumulative probability distribution $P_c(E)$ is expressed by

$$P_c(E) = \int_{-\infty}^{E} dE'\, P(E') = \frac{1}{2\pi i}\int_{-i\infty-\epsilon}^{i\infty-\epsilon} d\nu\,(-\nu)^{-1} \times \exp[\tfrac{1}{2}\nu(\nu - 1)K_1\{1 + (\nu - 2)\Delta\}^{-1} - \nu E]. \qquad (4.18)$$

In reference to Eq. (4.4), the use of the integration variable (4.8) brings (4.18) into

$$P_c(E) = \exp[-\tfrac{1}{2}(a^2 + b^2)]\frac{1}{2\pi i} \times \int_{-i\infty+\epsilon}^{i\infty+\epsilon} dt\,(b/a - t)^{-1}\exp[\tfrac{1}{2}ab(t + t^{-1})]$$
$$= \exp[-\tfrac{1}{2}(a^2 + b^2)]\sum_{n=0}^{\infty}\left(\frac{a}{b}\right)^{n+1}\frac{1}{2\pi i} \times \oint dt\, t^n\exp[\tfrac{1}{2}ab(t + t^{-1})], \qquad (4.19)$$

where the last expression is obtained by the expansion of the factor $(b/a - t)^{-1}$ in the power series of t. Thus, by use of the formula (4.8) and also by the replacement $n + 1 \to n$, we obtain

$$P_c(E) = \exp[-\tfrac{1}{2}(a^2 + b^2)]\sum_{n=1}^{\infty}(a/b)^n I_n(ab), \quad E < E_c, \qquad (4.20)$$

which is convenient to use for $a/b < 1$. On the other hand, the use of the formula

$$\sum_{n=-\infty}^{+\infty} (a/b)^n I_n(ab) = \exp[\tfrac{1}{2}(a^2+b^2)] \quad (4.21)$$

brings (4.20) into the expression

$$P_c(E) = 1 - \exp[-\tfrac{1}{2}(a^2+b^2)] \sum_{n=0}^{\infty} (b/a)^n I_n(ab), \quad E < E_c, \quad (4.22)$$

which is convenient for $b/a < 1$. The average of the above two expressions, i.e.,

$$P_c(E) = \tfrac{1}{2} - \tfrac{1}{2}\exp[-\tfrac{1}{2}(a^2+b^2)] \times [I_0(ab) + \sum_{n=1}^{\infty} \{(b/a)^n - (a/b)^n\} I_n(ab)] \quad (4.23)$$

is also useful for the intermediate range of $a/b \sim 1$.

As $E \to E_c$ or $b \to 0$ in (4.22),

$$P_c(E) \to 1 - \exp(-\tfrac{1}{2}a^2), \quad E \to E_c, \quad (4.24)$$

whose second term on the right-hand side is to be cancelled by the contribution from the term of $\delta(E - E_c)$ in (4.7).

In Fig. 2, the cumulative probability function $P_c(E)$ is illustrated against $K_1^{1/2}(E/K_1 + \frac{1}{2})$ for the same values of the parameters as in Fig. 1.

It is noticed that, besides the features mentioned in the beginning of this section, the distribution $P(E)$ has

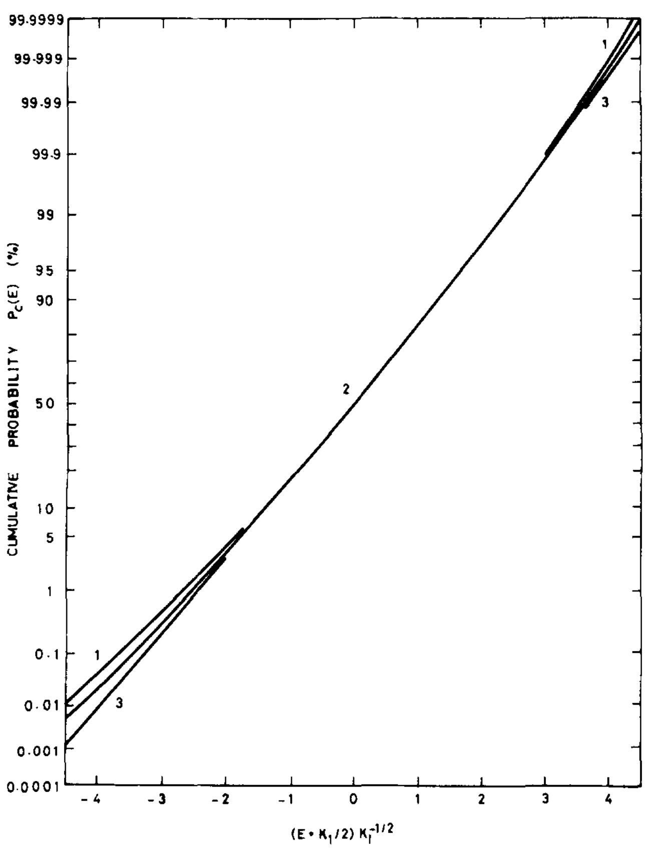


FIG. 2. Cumulative-probability distribution function $P_c(E)$. The values of the parameters are the same as those in Fig. 1 for each curve.

a very small but sharp distribution caused by the term of $\delta(E - E_c)$ at the threshold value of the irradiance. This feature will primarily be of the qualitative nature but, together with the existence of the threshold value, seems to be physically admissible and significant.

5. SUMMARY AND DISCUSSION

The equation (1.5) for the νth order correlation function of irradiance is formally solved by use of an operator method and is given by (3.6) in terms of the notation $\langle \cdots \rangle_L$ introduced in (3.5). Then the solution or the normalized moment of irradiance m_ν is expanded in the series according to a cluster expansion, as given by (3.13) or (3.14). If the convergence of the series is good enough so that the first two terms are sufficient, then the moments of irradiance can be approximated by (3.15) with (3.16) or by (3.17). The two expressions (3.15) and (3.17) are equivalent in that they give the correct moments of irradiance for the first three orders of $\nu = 1$, 2, and 3, and also, to the first order of Δ, give the same moments for both irradiance and logarithm of irradiance of all orders.

The essential difference is that the expression (3.15) leads to the unphysical result that the moment of the irradiance tends to vanish as its order ν increases, whereas the expression (3.17) tends to $\exp[\frac{1}{2}(K_1/\Delta)\nu]$ as $\nu \to \infty$ and therefore satisfies the condition of applicability (1.11) of the integral representation (1.12) for the irradiance distribution function. In Table I, the values predicted by (3.17) are compared with the recent experimental values observed by Gracheva *et al.* to show a very good agreement. The irradiance distribution function derived from (3.17) is positive definite and is given (4.7) and (4.10), whereas the corresponding cumulative distribution function is given by (4.20), (4.22), and (4.23). The features of this distribution are that (1) it is close to the log normal distribution, (2) it has a threshold value of irradiances giving nonvanishing probability, and (3) it has a very small but sharp distribution of the form of the δ function at the threshold value. The last two points are directly connected to the asymptotic behavior of the moments for large value of the order ν and seem to be physically significant, although they will primarily be of the qualitative nature.

The spectrum of medium as well as the conditions of the initial wave, e.g., whether it is a plane wave or a beam wave, do not directly enter in the expression of the distribution function but appear only through the first three moments of irradiance. Of course, in any case, the condition of applicability of the cluster approximation must be satisfied, which condition, however, is not always quite clear, not only in the present case but also in many other applications.[28] It is clear that the cluster approximation is definitely wrong for the special model (1.8a), but the model (1.8a) is drastically different from (1.8b) or (1.21) in that, in the latter, the "interaction potential" $V_I(\mathbf{r}_1, \mathbf{r}_2, \rho_{12})$ defined by (1.7) becomes of the order of magnitude of $|\rho_{12}|^{\alpha-2}$, $2 > \alpha > 1$ for large values of $|\rho_{12}|$ and tends to vanish as the "distance of separation" $|\rho_{12}|$ increases, whereas this is not the case in the former, i.e., for $|\rho_{12}| \sim \infty$,

$$V_I(\mathbf{r}_1, \mathbf{r}_2, \rho_{12}) \sim c\alpha\,|k\rho_{12}|^{\alpha-2} \times k^2[\mathbf{r}_1\cdot\mathbf{r}_2 + (\alpha-2)|\rho_{12}|^{-2}(\mathbf{r}_1\cdot\rho_{12})(\mathbf{r}_2\cdot\rho_{12})], \quad \alpha<2$$
$$= 2c'k^2\,\mathbf{r}_1\cdot\mathbf{r}_2, \quad \alpha=2. \tag{5.1}$$

Here, since the domain of $|\rho_{12}|$ (which makes important contributions to the moments of irradiance) increases with the distance of wave propagation, as may be inferred from the last term in (2.19) of $\tilde{\rho}_j$ the above fact implies that, for $\alpha<2$, V_I is effective only in a limited domain of comparatively small distances of $|\rho_{12}|$ whereas, for $\alpha=2$, V_I is effective in the entire domain of $|\rho_{12}|$. As the result, the correlation of $\tilde{V}_{12}$ and $\tilde{V}_{23}$ is expected to be small for $\alpha<2$ and the become smaller as $\alpha\to 1$.

In order to see the correlation of $\tilde{V}_{12}$ and $\tilde{V}_{23}$ when the distance of wave propagation is sufficiently large or when $\phi \gg 1$ in terms of a numerical distance ϕ of (3.21), parameters $C(12,23)$, $C(12,21)$, and $C(12)$ may be conveniently introduced according to

$$k^2\int_0^L dz_2\int_0^{z_2} dz_1\,\langle\Delta\tilde{V}_{12}(z_2)\Delta\tilde{V}_{23}(z_1)\rangle_L \sim C(12,23)\phi^{4(1-2/\alpha)},$$
$$k^2\int_0^L dz_2\int_0^{z_2} dz_1\,\langle\Delta\tilde{V}_{12}(z_2)\Delta\tilde{V}_{21}(z_1)\rangle_L \sim C(12,21)\phi^{4(1-2/\alpha)}, \tag{5.2}$$
$$k^2\int_0^L dz_2\int_0^{z_2} dz_1\,\langle\tilde{V}_{12}(z_2)\rangle_L\langle\tilde{V}_{21}(z_1)\rangle_L = \tfrac{1}{2}[k\int_0^L dz_1\,\langle\tilde{V}_{12}(z_2)\rangle_L]^2 \sim \tfrac{1}{2}C^2(12)\phi^{4(1-2/\alpha)}, \quad 2>\alpha>1,$$

which are the asymptotic expressions for $\phi \gg 1$ and are given by use of (C16) and (C5). The integration in (C16) was made by a numerical method and the results are shown in Table II as a function of α. It is found from this Table that the normalized correlation $C(12,23)/C(12,21)$ is of the same order of magnitude as of the parameter Δ in Table I for $\alpha=5/3$, and its magnitude tends to decrease as $\alpha\to 1$. It also decreases as $\alpha\to 2$, but note that both $C(12,23)$ and $C(12,21)$ tends to zero as $\alpha\to 2$ whereas these quantities rapidly increase as $\alpha\to 1$.

Note added in proof: In the case of plane waves and $\alpha=2$, it follows from (1.10) that $b=\infty$, $\delta_E=0$, and therefore $\langle I^\nu\rangle = \langle I\rangle^\nu$ or $m_\nu=1$ for all orders. In Table I, $\beta_0^2 = K_1$ for $\phi \ll 1$ and is given by (C3); thus $\beta_0^2 = 0.42\phi$, $\alpha=5/3$. The asymptotic form (1.11) may be analogous to the wavefunction of the ground states of complex nuclei under the so-called saturation condition; it has been shown by E. P. Wigner (1936) that the binding energy of nucleus, consisting of ν nucleons, increases linearly with ν (instead of ν^2) when ν is sufficiently large and the two-body potential between nucleons satisfies an appropriate condition.

APPENDIX A: ORDERED OPERATOR REPRESENTATION OF $V_{12}(z)$ and $V_{12}(z)$

The function $V(\mathbf{r})$ introduced in (1.21) can be represented by an integral of the form

$$V(\mathbf{r}) = c\,|k\mathbf{r}|^\alpha = \int_c d\mathbf{t}\,\bar{V}(\mathbf{t})\exp(-ik\mathbf{r}\cdot\mathbf{t}), \quad 2>\alpha>1, \tag{A1}$$

where

$$(\mathbf{t}) = (t\cos\varphi,\ t\sin\varphi), \quad \int_c d\mathbf{t} = \int_{(0+)}^\infty dt\,t\int_0^{2\pi} d\varphi,$$
$$\bar{V}(\mathbf{t}) = c\,(-i)(2\pi)^{-2}2^\alpha\Gamma^2(1+\alpha/2)\sec(\pi\alpha/2)(-t)^{-\alpha-2}, \tag{A2}$$

and the path of integration starts at $t=+\infty$ and goes on the infinitesimal upper side of the real axis where $\arg(-t)=-\pi$; then it encircles the origin in the counter clockwise direction and returns to the starting point. Equation (A1) may be proved by use of the formula[30]

$$\int_{(0+)}^\infty dt(-t)^{-\alpha-1}J_0(at) = -i\pi 2^{1-\alpha}\cos(\pi\alpha/2)\Gamma^{-2}(1+\alpha/2)|a|^\alpha. \tag{A3}$$

According to (1.7) and (2.9), the operator $\hat{V}_{12}(z)$ contains the terms

$$V[\hat{\rho}_{12}(z) \pm \tfrac{1}{2}(\hat{r}_1-\hat{r}_2)(z)] = \int_c d\mathbf{t}\,\bar{V}(\mathbf{t})\exp[-ik\mathbf{t}\cdot\{\hat{\rho}_{12}(z) \pm \tfrac{1}{2}(\hat{r}_1-\hat{r}_2)(z)\}], \tag{A4}$$

where the operators $\hat{\rho}_j(z)$ and $\hat{r}_j(z)$ are defined by (2.17) and (2.4).

The exponent in (A4) can be divided into the two parts A and B in such a way that A contains only the coordinates $\mathbf{r}_1$, $\mathbf{r}_2$, and ρ_{12}, while B contains only the differential operators $\partial/\partial\rho_j$ and $\partial/\partial\mathbf{r}_j$ ($j=1,2$). Hence, by use of the formula

$$\exp(A+B) = \exp(A)\exp(B)\exp(\tfrac{1}{2}[B,A]), \tag{A5}$$

the commutator $[B,A]$ being a constant in the present case, (A4) gives the ordered operator expression

$$V[\hat{\rho}_{12}(z) \pm \tfrac{1}{2}(\hat{r}_1-\hat{r}_2)(z)] = \exp(-ik\mathbf{t}\cdot\rho_{12}) \times\exp[\mp i\tfrac{1}{2}k\mathbf{t}\cdot(\hat{r}_1-\hat{r}_2)(z)] \times\exp\left[z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right]\exp(\mp ikz\mathbf{t}^2), \tag{A6}$$

where the last factor has been caused by the commutator $[B,A]$ in (A5).

For the other two terms in $\hat{V}_{12}$, the same ordering does not give rise to any additional factor because the corresponding commutator vanishes. Thus the operator $\hat{V}_{12}(z)$ is found to be given by the integral of the form

$$\hat{V}_{12}(z) = \int_c d\mathbf{t}\bar{V}(\mathbf{t})\exp(-ik\mathbf{t}\cdot\rho_{12})g[\mathbf{t},\hat{r}_1(z),\hat{r}_2(z)] \times\exp\left[z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right], \tag{A7}$$

where

$$g[\mathbf{t},\hat{r}_1(z),\hat{r}_2(z)] = 2[\cos\{\tfrac{1}{2}k\mathbf{t}\cdot(\hat{r}_1+\hat{r}_2)(z)\} - \cos\{\tfrac{1}{2}k\mathbf{t}\cdot(\hat{r}_1-\hat{r}_2)(z)+kz\mathbf{t}^2\}]$$
$$= 2^2\sin[\tfrac{1}{2}k\{\mathbf{t}^2 z - \mathbf{t}\cdot\hat{r}_2(z)\}]\sin[\tfrac{1}{2}k\{\mathbf{t}^2 z + \mathbf{t}\cdot\hat{r}_1(z)\}], \tag{A8}$$

and is of the order of magnitude of $\mathbf{t}^2$ for $|\mathbf{t}|\to 0$. Therefore, it follows from (A2) that the integrand of (A7) is of the order of $t^{-\alpha+1}$ for $t\sim 0$ and is integrable at the origin to give

$$\int_{(0+)}^\infty dt(-t)^{-\alpha-1} = 2i\sin\pi\alpha\int_0^\infty dt\,t^{-\alpha-1}. \tag{A9}$$

Thus, by use of the operator notation

$$\mathcal{J}(\mathbf{t}) \equiv \pi^{-2}2^\alpha\Gamma^2(1+\alpha/2)\sin(\pi\alpha/2)\int_{-\infty}^\infty d\mathbf{t}\,|\mathbf{t}|^{-\alpha-2}, \tag{A10}$$

TABLE II. Correlation parameters defined by (5.2) as a function of α in the case of plane waves.

α	1.9	5/3	1.5	1.45
$C(12,21)$	0.18	0.99	6.1	36.3
$C(12,23)$	−0.0033	−0.033	−0.053	−0.046
$C^2(12)/2$	0.014	0.185	0.59	0.84
$C(12,23)/C(12,21)$	−0.018	−0.033	−0.0087	−0.0013

the ordered operator $\hat{V}_{12}(z)$ is finally expressed by

$$\hat{V}_{12}(z) = -c\,\mathcal{J}(\mathbf{t})\exp(-ik\mathbf{t}\cdot\rho_{12})g[\mathbf{t},\hat{r}_1(z),\hat{r}_2(z)]$$
$$\times\exp\left[z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right]. \qquad (A11)$$

Here, $g[\mathbf{t},\hat{r}_1,\hat{r}_2]$ is given by (A8) and all the differential operators $\partial/\partial\mathbf{r}_j$ and $\partial/\partial\rho_j$ $(j=1,2)$ are ordered to the right of all the coordinates $\mathbf{r}_1$, $\mathbf{r}_2$, and ρ_{12}.

By use of the expression (A11) of $\hat{V}_{12}(z)$, it is straightforward to construct the ordered operator expression of $\tilde{V}_{12}(z)$ defined by (2.21); by use of the relations

$$\exp(ik\mathbf{t}\cdot\rho_{12})\hat{r}_j(z')\exp(-ik\mathbf{t}\cdot\rho_{12}) = \hat{r}_j(z') \pm z'\mathbf{t},$$
$$\exp\left[z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right]\hat{r}_j(z')\exp\left[-z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right]$$
$$=\hat{r}_j(z')\pm z\mathbf{t}, \qquad (A12)$$

the signs (±) being for $j=1,2$, respectively, it is found that

$$\tilde{V}_{12}(z) = -c\,\mathcal{J}(\mathbf{t})\exp(-ik\mathbf{t}\cdot\rho_{12})g[\mathbf{t},\hat{r}_1(z),\hat{r}_2(z)]$$
$$\times\exp\left[k\int_0^z dz'\{V[\hat{r}_1(z')+z'\mathbf{t}]+V[\hat{r}_2(z')-z'\mathbf{t}]\}\right]$$
$$\times\exp\left[-k\int_0^z dz'\{V[\hat{r}_1(z')+z\mathbf{t}]+V[\hat{r}_2(z')-z\mathbf{t}]\}\right]$$
$$\times\exp\left[z\mathbf{t}\cdot\left(\frac{\partial}{\partial\mathbf{r}_1}-\frac{\partial}{\partial\mathbf{r}_2}\right)\right]. \qquad (A13)$$

APPENDIX B: INTEGRAL REPRESENTATIONS OF K_1 (12) AND K_2 (12, 23) FOR PLANE WAVES TO THE SECOND ORDER OF APPROXIMATION

In the case of plane waves, the symbolic vector $|I\rangle$ introduced in (2.33) satisfies

$$\frac{\partial}{\partial\mathbf{r}_j}|I\rangle = \frac{\partial}{\partial\rho_j}|I\rangle = T_j|I\rangle = 0, \quad j=1,2, \qquad (B1)$$

T_j being defined in (2.1), and hence, by (2.2) and (2.7),

$$\exp[kH_jz]|I\rangle = \exp[kT_jz]U_j(z)|I\rangle$$
$$=\exp\{-k\int_0^z dz'V[\hat{r}_j(z'-L)]\}|I\rangle$$
$$=\exp[-kzV(\mathbf{r}_j)]|I\rangle,$$
$$\exp[k(H_1+H_2)L]|I\rangle = \exp[-k\{V(\mathbf{r}_1)+V(\mathbf{r}_2)\}L]|I\rangle,$$
$$\langle 0|\exp[k(H_1+H_2)L]|I\rangle = \langle 0|I\rangle. \qquad (B2)$$

Therefore, it is more convenient to use, instead of the notation $\langle Q_{12}\rangle_L$ in (2.23), the new notation

$$\langle Q_{12}\rangle'_L = \langle 0|Q_{12}\exp[k(H_1+H_2)L]|I\rangle/\langle 0|I\rangle. \qquad (B3)$$

Here, from the definition of $\tilde{V}_{12}(z)$ in (2.12),

$$\exp[k(H_1+H_2)L]\tilde{V}_{12}(z)\exp[-k(H_1+H_2)L] = \tilde{V}_{12}(z-L), \qquad (B4)$$

and therefore $\langle Q_{12}\rangle_L$ is connected to $\langle Q_{12}\rangle'_L$ by the relation

$$\langle Q_{12}(z)\rangle_L = \langle Q_{12}(z-L)\rangle'_L. \qquad (B5)$$

Thus, by the replacement of $z-L$ to $-z$, Eqs. (3.18) and (3.19) are then expressed by

$$K_1(12)\sim -k\int_0^L dz\langle\tilde{V}_{12}(-z)\rangle'_L$$
$$+k^2\int_0^L dz_2\int_0^{z_2}dz_1\langle\Delta\tilde{V}_{12}(-z_1)\Delta\tilde{V}_{12}(-z_2)\rangle'_L,$$
$$K_2(12,23)\sim 2k^2\int_0^L dz_2\int_0^{z_2}dz_1\langle\Delta\tilde{V}_{12}(-z_1)\Delta\tilde{V}_{23}(-z_2)\rangle'_L, \qquad (B6)$$

whose last expression has been obtained by the straightforward extension of the dimension of variables to include $\mathbf{r}_3$ and ρ_3 as in (3.5).

Here, since $V(\hat{r}_j)$ is an even function of the $\hat{r}_j(z)$ which are commutable with each other for all values of z, the use of the expression (A13) or $\tilde{V}_{12}(z)$ and the definition of $\langle\cdots\rangle'_L$ in (B3) with (B2) gives, on replacing $z\to -z$,

$$\langle\tilde{V}_{12}(-z)\rangle'_L = -c\,\mathcal{J}(\mathbf{t})g[\mathbf{t},0,0]$$
$$\times\exp[-2k\int_0^z dz'\{V(z'\mathbf{t})-V(z\mathbf{t})\}-2kLV(z\mathbf{t})], \qquad (B7)$$

where use is made of the relation (A12) and the equations of $|I\rangle$ and $\langle 0|$ given by (B1) and (2.33), particularly the formula

$$\langle 0|\hat{r}_j(z)|I\rangle = 0. \qquad (B8)$$

By using (A8) and the notation

$$\gamma(\mathbf{t},z)_L = k[L-\alpha(1+\alpha)^{-1}z]V(z\mathbf{t}), \qquad (B9)$$

Eq. (B7) is expressed by

$$\langle\tilde{V}_{12}(-z)\rangle'_L = -c\,\mathcal{J}(\mathbf{t})2^2\sin^2\{\tfrac{1}{2}kz\mathbf{t}^2\}\exp[-2\gamma(\mathbf{t},z)_L]. \qquad (B10)$$

In exactly the same way, the use of the ordered operator expression (A13) of $\tilde{V}_{12}(z)$ leads to

$$\langle\tilde{V}_{12}(-z_1)\tilde{V}_{23}(-z_2)\rangle'_L = c^2\mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)2^4\sin\{\tfrac{1}{2}kz_1\mathbf{t}_1^2\}\sin\{\tfrac{1}{2}kz_2\mathbf{t}_2^2\}$$
$$\times\sin\{\tfrac{1}{2}kz_1\mathbf{t}_1\cdot(\mathbf{t}_1-\mathbf{t}_2)\}\sin\{\tfrac{1}{2}k\mathbf{t}_2\cdot(z_2\mathbf{t}_2-z_1\mathbf{t}_1)\}$$
$$\times\exp[-\gamma(\mathbf{t}_1,z_1)_L-\gamma(\mathbf{t}_2,z_2)_L-\gamma(\mathbf{t}_1,z_1;\mathbf{t}_2,z_2)_L], \qquad (B11)$$

where

$$\gamma(\mathbf{t}_1,z_1;\mathbf{t}_2,z_2)_L = (1+\alpha)^{-1}kz_1V[z_1(\mathbf{t}_1-\mathbf{t}_2)]$$
$$+k\int_{z_1}^{z_2}dz\,V(z_1\mathbf{t}_1-z\mathbf{t}_2)+k(L-z_2)V(z_1\mathbf{t}_1-z_2\mathbf{t}_2) \qquad (B12)$$

with the relations

$$\gamma(\mathbf{t}_1,0;\mathbf{t}_2,z_2)_L = \gamma(0,z_1;\mathbf{t}_2,z_2)_L = \gamma(\mathbf{t}_2,z_2)_L,$$
$$\gamma(\mathbf{t}_1,z;\mathbf{t}_2,z)_L = \gamma(\mathbf{t}_1-\mathbf{t}_2,z)_L, \quad \gamma(\mathbf{t}_1,z_1;0,z_2)_L = \gamma(\mathbf{t}_1,z_1)_L. \qquad (B13)$$

For the integrations in (B6), it is convenient to introduce the variable $\zeta = z/L$ and replace $(kL)^{1/2}\mathbf{t}\to\mathbf{t}$; then

$$c\,\mathcal{J}((kL)^{-1/2}\mathbf{t}) = c(kL)^{\alpha/2}\mathcal{J}(\mathbf{t}) = V[(L/k)^{1/2}]\mathcal{J}(\mathbf{t}),$$
$$\gamma[(kL)^{-1/2}\mathbf{t},z]_L \doteq \phi\gamma(\zeta)|\mathbf{t}|^\alpha/2, \qquad (B14)$$

where ϕ is defined by (3.21) and

$$\gamma(\zeta) = [1-\alpha(1+\alpha)^{-1}\zeta]\zeta^\alpha. \qquad (B15)$$

For example, the use of (B10) with (B14) gives, to the first order of $\tilde{V}_{12}$,

$$K_1(12)\sim -k\int_0^L dz\langle\tilde{V}_{12}(-z)\rangle'_L$$
$$=2\phi\int_0^1 d\zeta\,\mathcal{J}(\mathbf{t})\sin^2\{\tfrac{1}{2}\zeta\mathbf{t}^2\}\exp[-\phi\gamma(\zeta)|\mathbf{t}|^\alpha]. \qquad (B16)$$

In the same way, the use of (B11) and a similar expression for $\langle\tilde{V}_{12}\tilde{V}_{21}\rangle'_L$ leads to

$$k^2\int_0^L dz_2\int_0^{z_2}dz_1\langle\tilde{V}_{12}(-z_1)\tilde{V}_{23}(-z_2)\rangle'_L$$

$$=(2\phi)^2\int_0^1 d\zeta_2\int_0^{\zeta_2}d\zeta_1\mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)$$
$$\times\sin\{\tfrac{1}{2}\zeta_1\mathbf{t}_1^2\}\sin\{\tfrac{1}{2}\zeta_2\mathbf{t}_2^2\}\sin\{\tfrac{1}{2}\zeta_1\mathbf{t}_1\cdot(\mathbf{t}_1-\mathbf{t}_2)\}$$
$$\times\sin\{\tfrac{1}{2}\mathbf{t}_2\cdot(\zeta_2\mathbf{t}_2-\zeta_1\mathbf{t}_1)\}$$
$$\times\exp[-(\phi/2)\{\gamma(\zeta_1)|\mathbf{t}_1|^\alpha+\gamma(\zeta_2)|\mathbf{t}_2|^\alpha$$
$$+\gamma(\zeta_1,\mathbf{t}_1;\ \zeta_2,\mathbf{t}_2)\}],\tag{B17}$$

$$k^2\int_0^1 dz_2\int_0^{z_2}dz_1\langle\tilde{V}_{12}(-z_1)\tilde{V}_{21}(-z_2)\rangle'_L$$

$$=(2\phi)^2\int_0^1 d\zeta_2\int_0^{\zeta_2}d\zeta_1\mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)\sin^2\{\tfrac{1}{2}\mathbf{t}_2\cdot(\zeta_2\mathbf{t}_2-\zeta_1\mathbf{t}_1)\}$$
$$\times\sin^2\{\tfrac{1}{2}\zeta_1\mathbf{t}_1\cdot(\mathbf{t}_1-\mathbf{t}_2)\}\exp[-\phi\gamma(\zeta_1,\mathbf{t}_1;\ \zeta_2,\mathbf{t}_2)],\tag{B18}$$

in terms of the notation

$$\gamma(\zeta_1,\mathbf{t}_1;\zeta_2,\mathbf{t}_2)=(1+\alpha)^{-1}\zeta_1^{\alpha+1}|\mathbf{t}_1-\mathbf{t}_2|^\alpha$$
$$+\int_{\zeta_1}^{\zeta_2}d\zeta\ |\zeta_1\mathbf{t}_1-\zeta\mathbf{t}_2|^\alpha+(1-\zeta_2)|\zeta_1\mathbf{t}_1-\zeta_2\mathbf{t}_2|^\alpha.\tag{B19}$$

Equations (B16)—(B18) provide the necessary expressions of all the terms of $K_1(12)$, $K_2(12,23)$ in (B6).

APPENDIX C: EVALUATION OF INTEGRALS OF (*B*10), (*B*17), AND (*B*18)

Since $\gamma(\zeta)\sim 1$ according to (B15), the exponential factor in the integrand of (B16) can be neglected when $\phi\ll 1$ and hence, to the first order of V_{12},

$$K_1(12)\sim 2\phi\int_0^1 d\zeta\,\mathcal{J}(\mathbf{t})\sin^2\{\tfrac{1}{2}\zeta\mathbf{t}^2\},\quad \phi\ll 1,$$

$$=2\phi\int_0^1 d\zeta\,\zeta^{\alpha/2}\mathcal{J}(\mathbf{t})\sin^2(\tfrac{1}{2}\mathbf{t}^2),\tag{C1}$$

where the last expression is obtained by the replacement of $\zeta^{1/2}\mathbf{t}\to\mathbf{t}$.

Thus, with reference to (A10) and also to the formula

$$\int_{-\infty}^{\infty}d\mathbf{t}\,|\mathbf{t}|^{-\alpha-2}\sin^2(\tfrac{1}{2}\mathbf{t}^2)$$

$$=2\pi\int_0^\infty dt\,t^{-\alpha-1}\sin^2(\tfrac{1}{2}t^2)$$

$$=(\pi/2)^2\operatorname{cosec}(\pi\alpha/4)\Gamma^{-1}(1+\alpha/2),\tag{C2}$$

Eq. (C1) becomes

$$K_1(12)\sim 2^\alpha\Gamma(1+\alpha/2)(1+\alpha/2)^{-1}\cos(\pi\alpha/4)\phi,\quad \phi\ll 1.\tag{C3}$$

On the other hand, when $\phi\gg 1$ in (B16),

$$\sin\{\tfrac{1}{2}\zeta\mathbf{t}^2\}\sim\tfrac{1}{2}\zeta\mathbf{t}^2,\tag{C4}$$

and the integration becomes straightforward by use of the new variable of integration $t'=\phi\gamma(\zeta)|\mathbf{t}|^\alpha$ to give the asymptotic expression of the integral (B16):

$$\pi^{-1}2^\alpha\alpha^{-1}(\alpha-1)^{-1}\Gamma(4/\alpha-1)\Gamma^2(1+\alpha/2)\sin(\pi\alpha/2)$$
$$\times{}_2F_1(\alpha-1,\ 4/\alpha-1;\ \alpha;\ \alpha(1+\alpha)^{-1})\phi^{2(1-2/\alpha)},\quad \phi\gg 1.\tag{C5}$$

Both results (C3) and (C5) are the same as the corresponding ones given in Ref. 12.

In the same way, when $\phi\ll 1$, the use of (B16) and (B17) yields the expression of $K_2(12,23)$ as

$$\tfrac{1}{2}K_2(12,23)\sim k^2\int_0^L dz_2\int_0^{z_2}dz_1[\langle\tilde{V}_{12}(-z_1)\tilde{V}_{23}(-z_2)\rangle'_L$$
$$-\langle\tilde{V}_{12}(-z_1)\rangle'_L\langle\tilde{V}_{23}(-z_2)\rangle'_L]$$
$$\approx(2\phi)^2\int_0^1 d\zeta_2\int_0^{\zeta_2}d\zeta_1\ \mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)\sin\{\tfrac{1}{2}\zeta_1\mathbf{t}_1^2\}\sin\{\tfrac{1}{2}\zeta_2\mathbf{t}_2^2\}$$
$$\times[\sin\{\tfrac{1}{2}\zeta_1\mathbf{t}_1\cdot(\mathbf{t}_1-\mathbf{t}_2)\}\sin\{\tfrac{1}{2}\mathbf{t}_2\cdot(\zeta_2\mathbf{t}_2-\zeta_1\mathbf{t}_1)\}$$
$$-\sin\{\tfrac{1}{2}\zeta_1\mathbf{t}_1^2\}\sin\{\tfrac{1}{2}\zeta_2\mathbf{t}_2^2\}],\tag{C6}$$

where the exponential factors have been omitted as in (C1). The last square bracket factor in (C6) is also expressed, after some manipulations, by

$$[\ ]=\cos\{\tfrac{1}{2}(\zeta_1\mathbf{t}_1^2+\zeta_2\mathbf{t}_2^2)\}\sin^2\{\tfrac{1}{2}\zeta_1\mathbf{t}_1\cdot\mathbf{t}_2\}\tag{C7}$$

plus an odd function of $\mathbf{t}_1\cdot\mathbf{t}_2$ which is cancelled in the integration. Thus, by use of (C7) in (C6),

$$K_2(12,23)\sim-\phi^2\int_0^1 d\zeta_2\int_0^{\zeta_2}d\zeta_1\mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)[1-\cos(\zeta_1\mathbf{t}_1^2)$$
$$-\cos(\zeta_2\mathbf{t}_2^2)-\cos(\zeta_1\mathbf{t}_1\cdot\mathbf{t}_2)-\cos(\zeta_1\mathbf{t}_1^2+\zeta_2\mathbf{t}_2^2+\zeta_1\mathbf{t}_1\cdot\mathbf{t}_2)$$
$$+\cos(\zeta_1\mathbf{t}_1^2+\zeta_2\mathbf{t}_2^2)+\cos\{\zeta_1(\mathbf{t}_1^2+\mathbf{t}_1\cdot\mathbf{t}_2)\}$$
$$+\cos\{\mathbf{t}_2\cdot(\zeta_2\mathbf{t}_2+\zeta_1\mathbf{t}_1)\}].\tag{C8}$$

In order to evaluate the integrals in (C8), it is convenient to introduce the cylindrical coordinate systems

$$(\mathbf{t}_1)=(|\mathbf{t}_1|,\varphi_1),\quad (\mathbf{t}_2)=(|\mathbf{t}_2|,\varphi_2),\tag{C9}$$

with the further change of variables

$$|\mathbf{t}_1|=t\cos\theta,\quad |\mathbf{t}_2|=t\sin\theta,\quad \varphi_1-\varphi_2=\varphi.\tag{C10}$$

Then, from (A10),

$$\mathcal{J}(\mathbf{t}_1)\mathcal{J}(\mathbf{t}_2)=\pi^{-2}2^{2\alpha+2}\Gamma^4(1+\alpha/2)\sin^2(\pi\alpha/2)$$
$$\times(2\pi)^{-1}\int_0^{2\pi}d\varphi\int_0^{\pi/2}d\theta\,(\sin\theta\cos\theta)^{-\alpha-1}\int_0^\infty dt\,t^{-2\alpha-1},\tag{C11}$$

and hence it follows that the integral with respect to t in (C8) is expressed in terms of integrals of the form

$$\int_0^\infty dt\,t^{-2\alpha-1}(1-\cos ft^2)=(\pi/4)\csc(\pi\alpha/2)\Gamma^{-1}(1+\alpha)|f|^\alpha,\quad \alpha<2.\tag{C12}$$

Thus, by the introduction of the variable $s=\zeta_1/\zeta_2\le 1$, Eq. (C8) is expressed by

$$K_2(12,23)\sim-\pi^{-1}2^{2\alpha}\Gamma^4(1+\alpha/2)\Gamma^{-1}(1+\alpha)(\alpha+2)^{-1}$$
$$\times\sin(\pi\alpha/2)\phi^2\int_0^1 ds\,(2\pi)^{-1}\int_0^{2\pi}d\varphi$$
$$\times\int_0^{\pi/2}d\theta(\sin\theta\cos\theta)^{-\alpha-1}[s^\alpha\cos^{2\alpha}\theta+\sin^{2\alpha}\theta$$
$$+s^\alpha|\sin\theta\cos\theta\cos\varphi|^\alpha+|\sin^2\theta$$
$$+s\cos\theta(\cos\theta+\sin\theta\cos\theta)|^\alpha-|s\cos^2\theta+\sin^2\theta|^\alpha$$
$$-s^\alpha|\cos\theta(\cos\theta+\sin\theta\cos\varphi)|^\alpha$$
$$-|\sin\theta(\sin\theta+s\cos\theta\cos\varphi)|^\alpha],\quad \phi\ll 1.\tag{C13}$$

Here, the integration with respect to s is elementary, and the remaining integration with respect to θ and φ may be made by a numerical method. The result for $\alpha = \frac{5}{3}$ is found to be

$$K_2(12,23) \sim -0.04\phi^2, \quad \alpha = \tfrac{5}{3}, \quad \phi \ll 1. \tag{C14}$$

Also for the integral (B18), the expression similar to (C13) is obtained by neglecting the exponential factor on the right-hand side, but is found to be reduced to a logarithmically divergent integral. This means that, as $\phi \to 0$, the integral (B18) over ϕ^2 tends to infinity and therefore, by (B16) with (C3), also that

$$\int_0^L dz_2 \int_0^{z_2} dz_1 \langle \Delta\tilde{V}_{12}(-z_1)\Delta\tilde{V}_{12}(-z_2)\rangle_L' / \Big[\int_0^L dz \langle \tilde{V}_{12}(-z)\rangle_L'\Big]^2 \to \infty, \tag{C15}$$

as $L \to 0$ or $\phi \to 0$.

On the other hand, when $\phi \gg 1$, the following asymptotic expressions are obtained by use of the approximation similar to (C4):

$$k^2 \int_0^L dz_2 \int_0^{z_2} dz_1 \begin{Bmatrix} \langle \tilde{V}_{12}(-z_1)\tilde{V}_{21}(-z_2)\rangle_L' \\ \langle \tilde{V}_{12}(-z_1)\tilde{V}_{23}(-z_1)\rangle_L' \end{Bmatrix}$$

$$\sim \pi^{-2} 2^{2\alpha} \Gamma^4\Big(1+\frac{\alpha}{2}\Big)\Gamma\Big(\frac{8}{\alpha}-2\Big)\alpha^{-1}$$

$$\times \sin^2(\pi\alpha/2)\phi^{4(1-2/\alpha)} \int_0^1 d\zeta_2 \zeta_2^5 \int_0^1 ds\, s^2 \pi^{-1} \int_0^\pi d\varphi \int_0^{\pi/2} d\theta$$

$$\times \begin{Bmatrix} (\sin\theta\cos\theta)^{-\alpha+1}(\sin\theta - s\cos\theta\cos\varphi)^2(\cos\theta - \sin\theta\cos\varphi)^2 \\ (\sin\theta\cos\theta)^{-\alpha+2}(\sin\theta - s\cos\theta\cos\varphi)(\cos\theta - \sin\theta\cos\varphi) \end{Bmatrix}$$

$$\times [\gamma(\zeta_2, s, \theta, \varphi)]^{2(1-4/\alpha)}$$

$$\times [\tfrac{1}{2}\{\gamma(\zeta_2 s)\cos^\alpha\theta + \gamma(\zeta_2)\sin^\alpha\theta + \gamma(\zeta_2, s, \theta, \varphi)\}]^{2(1-4/\alpha)}, \quad \phi \gg 1, \tag{C16}$$

where $\gamma(\zeta)$ is given by (B15) and

$$\gamma(\zeta, s, \theta, \varphi) = (\zeta s)^{\alpha+1}(1+\alpha)^{-1}\,|1 - 2\sin\theta\cos\theta\cos\varphi|^{\alpha/2}$$

$$+ \zeta^{\alpha+1}\int_s^1 dx\,|s^2\cos^2\theta + x^2\sin^2\theta - 2sx\cos\theta\sin\theta\cos\varphi|^{\alpha/2}$$

$$+ (1-\zeta)\zeta^\alpha\,|s^2\cos^2\theta + \sin^2\theta - 2s\cos\theta\sin\theta\cos\varphi|^{\alpha/2}.$$

Table II was obtained by use of (C5) and (C16) and the integrations were made by a numerical method.

[1]K. Furutsu, J. Opt. Soc. Am. 62, 240 (1972).
[2]V. I. Klyatskin and V. I. Tatarskii, Izv. Vyssh. Uchebn. Zaved. Radiofiz. 15, 1433 (1972).
[3]V. I. Shishov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. 11, 866 (1968).
[4]V. I. Tatarskii, Zh. Eksp. Teor. Fiz. 56, 2106 (1969) [Sov. Phys. JETP 29, 1133 (1969)].
[5]J. E. Molyneux, J. Opt. Soc. Am. 61, 248 (1971).
[6]M. J. Beran and T. L. Ho, J. Opt. Soc. Am. 59, 1134 (1969).
[7]W. P. Brown, Jr., J. Opt. Soc. Am. 62, 45 (1972).
[8]V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).
[9]J. W. Strohbehn, Proc. IEEE 56, 1301 (1968).
[10]V. I. Klyatskin and V. I. Tatarskii, Izv. Vyssh. Ucheben. Zaved. Radiofiz. 13, 1061 (1970).
[11]A. M. Whitman and M. J. Beran, J. Opt. Soc. Am. 60, 1595 (1970).
[12]K. S. Gochelashvili and V. I. Shishov, Zh. Eksp. Teor. Fiz. 66, 1237 (1974) [Sov. Phys. JETP 39, 605 (1974)].
[13]A. M. Prokhorov, F. V. Bunkin, K. S. Gochelashvili, and V. I. Shishov, Proc. IEEE 63, 790 (1975).
[14]R. L. Fante, Radio Sci. 10, 77 (1975).
[15]I. M. Dazkesamanskaya and V. I. Shishov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. 13, 16 (1970).
[16]W. P. Brown, Jr., J. Opt. Soc. Am. 62, 966 (1972).
[17]In the derivation in Ref. 1, the condition $|\arg(it)| < \pi$ in Eq. (108) should read $|\arg(it)| \leq \pi/2$. The integral representation (1.12) can also be derived from a different viewpoint: H. Bremmer, Radio Sci. 8, 511 (1973).
[18]The condition (1.11) satisfies the Carleman condition for the existence of a unique distribution function: J. W. Strohbehn, Ting-i Wang, and J. P. Speck, Radio Sci. 10, 59 (1975).
[19]K. Furutsu and Y. Furuhama, Opt. Acta 20, 707 (1972).
[20]D. L. Fried, Appl. Opt. 12, 422 (1972).
[21]P. J. Titterton, Appl. Opt. 12, 423 (1972).
[22]M. Tateiba, IEEE Trans. AP-23, 493 (1975).
[23]A detailed equation formulation was given in Ref. 1.
[24]R. Zwanzig, Phys. Rev. 124, 983 (1961).
[25]H. Mori, Prog. Theor. Phys. (Kyoto) 33, 423 (1965).
[26]J. Weinstock, Phys. Fluids 12, 1045 (1969); 15, 454 (1972).
[27]J. H. Misguich and R. Balescu, J. Plasma Phys. 13, 385 (1975); Physica 79C, 373 (1975).
[28]A detailed theory on the cluster expansion was given by R. Kubo, J. Phys. Soc. Jpn. 17, 1100 (1962).
[29]M. E. Gracheva, A. S. Gurvich, S. S. Kashkarov, and V. V. Pokasov, Preprint, Dept. of Oceanology, Physics, Atmospheres, and Geography, USSR Academy of Sciences.
[30]G. N. Watson, *Theory of Bessel Function* (Cambridge U. P., Cambridge, 1952), 2nd ed., p. 392.

On generation of solutions of Einstein's equations

C. Hoenselaers

Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan
(Received 19 January 1976)

The equations for two-Killing-vector solutions of Einstein's equations are looked at from the point of view of the Lagrangian formalism. The original Lagrangian, written in terms of the metric, will undergo a Legendre transformation leading to the Ernst Lagrangian and from there by the same procedure to others. So one gets a sequence of Lagrangians, and by performing an invariance transformation at each step one obtains new solutions of Einstein's equations. A convenient method for dealing with Lagrangians quadratic in the velocities is outlined.

1. INTRODUCTION

Considerable effort has been made to generate solutions of Einstein's equations[1,2] or even the Einstein–Maxwell equations,[3,4] but up to now no one has successed in giving a transformation generating the Tomimatsu–Sato metrics from the Weyl metrics.

Moreover, it has been shown[5] that some of the known methods for generating solutions give rise to asymptotically nonflat solutions.

As the methods up to now either do not use the Lagrangian formalism (Ref. 2) or do not use the complete Lagrangian for the gravitational field (Ref. 4), the question arises whether the analysis of the complete Einstein Lagrangian reveals any new possibility of obtaining new solutions.

A method for dealing with Lagrangians quadratic in the derivatives of the fields, just the type needed for our problem, is outlined in Appendix A.

Appendix B contains a special treatment of the Lagrangians used, which may be useful for further studies.

A sketch of the applications of our method is given in Appendix C.

Roughly speaking, the result is that, by a chain of Legendre transformations of the original Lagrangian, the second one will turn out to be the Ernst Lagrangian and the third one will also be given, and applying an invariance-transformation at every intermediate step, one can obtain an infinite-parameter solution of the field equations. The question, whether the method reveals all 2-Killing-vector solutions of Einstein's equations has unfortunately to remain unanswered. Although we were not able to give precisely the Schwarzschild → Kerr transformation, from looking at the formulas it is very likely to be built up from a chain of infinitesimal transformations of the discussed type.

2. GENERATING SOLUTIONS

Let $(M, g_{\alpha\beta})$ be a 4-manifold with a metric of signature $(+++-)$ admitting two Killing-vectors ξ^α, η^α and let

$$\xi^\alpha \xi_\alpha =: m, \quad \xi^\alpha \eta_\alpha =: l, \quad \eta_\alpha \eta^\alpha =: n. \tag{1}$$

Then the line element can be written[6,7]

$$ds^2 = \exp(2k)(dx_1^2 + dx_2^2) + n\, d\phi^2 + 2l\, d\phi\, dt + m\, dt^2. \tag{2}$$

If we take ξ^α to be a timelike Killing vector, an assumption made only for convenience (taking it spacelike would change some signs but not the essentials of our analysis), we can set

$$r^2 := l^2 - mn > 0.$$

The Einstein equations are to be derived from the Lagrangian density

$$\sqrt{|g|}R = 2r\nabla^2 k + 2\nabla^2 r - (1/2r)(\nabla l^2 - \nabla m \nabla n)$$

where ∇ denotes the two-dimensional derivative operator $\binom{\partial_1}{\partial_2}$. After we write the second derivatives as divergencies, omit the divergence terms, and change the total sign, the Lagrangian reveals the form

$$L = 2\nabla k \nabla r + (1/2r)(\nabla l^2 - \nabla m \nabla n) \tag{3}$$

or equivalently

$$L = \tfrac{1}{2}r\left\{\nabla(\tfrac{5}{4}\ln r + 2k)^2 + \nabla\left[ar\sinh\left(\frac{m+n}{2r}\right)\right]^2 + \left[1 + \left(\frac{m+n}{2r}\right)^2\right]\nabla\left[\operatorname{arccot}\left(\frac{m-n}{2l}\right)\right]^2 - \nabla(-\tfrac{3}{4}\ln r - 2k)^2\right\}. \tag{4}$$

According to Appendix A, we now have to solve

$$X_{A;BC} = R_{ABCD}X^D$$

in one of the metrics determined by (3) or (4).[8] However, the solutions of this equation turn out to be [in the metric given by (3)] ($\alpha \cdots \zeta$ constants)

$$X^A = \begin{pmatrix} \alpha + \beta r \\ \gamma \cdot m + \delta l + \epsilon m \\ \gamma \cdot n + \zeta l - \epsilon n \\ \gamma \cdot l + \tfrac{1}{2}(\zeta m + \delta n) \end{pmatrix} \tag{5}$$

being Killing vectors for $\beta = \gamma = 0$.

As far as the corresponding finite transformations are concerned, α and β represent an addition of a constant and a multiple of r respectively to k, γ gives a conformal transformation of the Lagrangian by e^γ, to be absorbed by a trivial coordinate transformation, and the transformations belonging to δ, ϵ, ζ are already known[9]; moreover, they also can be absorbed by a coordinate transformation.

We have, anyway, three nonnull hypersurface-orthogonal Killing congruences and the gradients of the functions to be chosen as adjusted to them have to be proportional to

$$\begin{pmatrix}0\\-l\\0\\m\end{pmatrix}\begin{pmatrix}0\\-m\\n\\0\end{pmatrix}\begin{pmatrix}0\\0\\-l\\n\end{pmatrix}.$$

One can see that functions of the arguments

$$l/m,\quad n/m,\quad l/n \tag{6}$$

respectively fulfill the requirements. As m and n are treated completely symmetrically in the original Lagrangian, we do not expect to get different results from the first and third, and so we will deal with the first and second only.

As far as the second, the less interesting, is concerned, one arrives at

$$L=2\nabla(k+\tfrac{1}{4}\ln r)\nabla r-(r/2)\{[\nabla(\text{arccosh})(l/r)]^2 - [(l/r)^2-1]\,\nabla[\tfrac{1}{2}\ln(m/n)]^2\} \tag{7}$$

and after the Legendre transformation (A10) at

$$L'=2\nabla(k+\tfrac{1}{4}\ln r)\nabla r-(r/2)\nabla[\text{arccosh}(l/r)]^2 - (2/r)[(l/r)^2-1]^{-1}\nabla\psi^2, \tag{8}$$

which can easily be shown to admit only trivial and not to new solutions leading vectors X^A from (A4).

More interesting is the first of the expressions (6). After the following choice of variables (in order to get the same variables as those used in Ref. 10),

$$m=-f,\quad n=r^2/f-f\omega^2,\qquad l=\omega f,\quad k=\gamma-\tfrac{1}{2}\ln f \tag{9}$$

one arrives at

$$L=2\nabla\gamma\nabla r+\tfrac{1}{2}[(f^2/r)\nabla\omega^2-(r/f^2)\nabla f^2] \tag{10}$$

and with

$$\nabla^*\psi=(f^2/r)\nabla\omega \tag{11}$$

one gets the familiar form

$$L'=2\nabla\gamma\nabla r-(r/2f^2)(\nabla f^2+\nabla\psi^2). \tag{12}$$

Here now (A4) has the solutions

$$X^A=\begin{pmatrix}\alpha+\beta r\\ \gamma r\\ f(\delta+\epsilon\psi)\\ (\epsilon/2)(\psi^2-f^2)+\delta\psi+\zeta\end{pmatrix} \tag{13}$$

(Killing vectors for $\beta=\gamma=0$).

α, β, γ again describe rather trivial transformations while δ, ϵ, ζ belong to the transformations given in Ref. 1.

However, we are not restricted, performing again Legendre transformations with respect to the δ, ϵ Killing congruences.[11]

Adjusting the variables to the Killing vector

$$\begin{pmatrix}0\\0\\f\\\psi\end{pmatrix},$$

one sees that an appropriate choice is arctan (f/ψ) and $\frac{1}{2}\ln(f^2+\psi^2)$, but one finds that the Legendre-transformed Lagrangian is of type (8).

By inversion on the unit circle in the f—ψ plane the Lagrangian (12) can be shown to be equivalent to

$$L'=2\nabla\gamma\nabla r-\frac{r}{2}\left(\frac{f^2+\psi^2}{f}\right)^2\left[\nabla\left(\frac{f}{f^2+\psi^2}\right)^2+\nabla\left(\frac{\psi}{f^2+\psi^2}\right)^2\right] \tag{14}$$

and, going backwards from (12) via (11), (10), (9) to (3), one has for L''

$$L''=2\nabla k''\nabla r+(1/2r)(\nabla l''^2-\nabla m''\nabla n''), \tag{15}$$

and the connections between the new and old functions are

$$\begin{aligned}
\nabla^*\chi&=r\cdot\left(\frac{f^2+\psi^2}{f}\right)^2\nabla\left(\frac{\psi}{f^2+\psi^2}\right),\\
m''&=\frac{m}{m^2+\psi^2},\\
n''&=r^2\cdot\frac{f^2+\psi^2}{f}-\frac{f\chi^2}{f^2+\psi^2},\\
l''&=\frac{\chi f}{f^2+\psi^2},\\
k''&=k+\tfrac{1}{2}\ln(f^2+\psi^2).
\end{aligned} \tag{16}$$

Now one has the possibility of Legendre-transforming this Lagrangian with respect to the third of the expressions (6), coming to L''' of the type (12), and so on. Thus, performing at every intermediate step a transformation according to (13) or (5), one can generate an infinite-parameter solution of Einstein's equations.

Some interesting questions arise and, at the present state of affairs, have to remain unanswered. For example, besides the problem already mentioned in the introduction, we have the question in which way type-(5) transformations of L'' affect the multipole moments. We know already[5] that a transformation of L' mixes the mass- and angular-momentum moments.

ACKNOWLEDGMENTS

The author wishes to thank the Research Institute for Fundamental Physics for its hospitality and specially express his gratitude to Professor H. Sato for encouragement and many valuable discussions.

APPENDIX A

Suppose that we are given N fields, denoted by F^A, and the field equations to be derived from the Lagrangian

$$L=G_{AB}\nabla F^A\nabla F^B, \tag{A1}$$

where G_{AB} is a nonsingular $N\times N$ matrix, whose components are functions of the F^A, and ∇ denotes a flat-space derivative operator. We will regard G_{AB} as the metric in an abstract N-dimensional Riemann space, and all the symbols used will have their usual meanings.[12] By varying (A1) with respect to F^A, the field equations, not very surprisingly, turn out to be

$$\nabla^2 F^A + \Gamma^A_{BC}\nabla F^B \nabla F^C = 0. \tag{A2}$$

By performing now an infinitesimal transformation of the form

$$E^A = F^A + \epsilon X^A(F), \tag{A3}$$

the new fields E^A will satisfy (A2), provided that F^A satisfies it and

$$X_{A;BC} = R_{ABCD}X^D. \tag{A4}$$

If (A2) is analogous to the equations of geodesics, this may be thought of as an analogy to the equation of geodesic deviation. As this equation implies the existence of a symmetric covariantly constant second rank tensor

$$Y_{(AB);C} = 0, \tag{A5}$$

$$Y_{(AB)} = X_{(A;B)} \tag{A6}$$

[in fact, (A4) is equivalent to (A5) and (A6)] their solutions contain not only the Killing vectors of G_{AB} [$Y_{(AB)} \equiv 0$, leaving (A1) invariant], the conformal Killing vectors with constant divergence [$Y_{(AB)} = \text{const} \circ G_{AB}$, multiplying the right-hand side of (A1) by a constant factor], but also other solutions iff G_{AB} is decomposible.[13]

If we assume that X^A is a Killing vector, i.e., a generator of an invariance of the Lagrangian (A1), then by Noether's theorem this fact gives rise to the existence of a divergence-free vector field, which can be written easily as

$$\eta_\alpha = X_A \nabla_\alpha F^A. \tag{A7}$$

Note added in proof: Let X_A be a conformal Killing vector with constant divergence

$$X_{A;B} = G_{AB} \circ \text{const};$$

then

$$\nabla(X_A \nabla F^A) = (X_{A,B} - \Gamma^C_{AB}X_C)\nabla F^A \nabla F^B$$
$$= \text{const} \circ G_{AB}\nabla F^A \nabla F^B,$$

which means that the divergence of $X_A \nabla F^A$ will vanish iff the Lagrangian vanishes for the actual fields, i.e., for the solutions of the field equations (A2) denoted by F^A_s. The Einstein Lagrangian can be written

$$L_{\text{Ein}} = \bar{G}_{AB}\nabla F^A \nabla F^B + H_A \nabla^2 F^A$$

or

$$L_{\text{Ein}} = (\bar{G}_{AB} - H_{(A,B)})\nabla F^A \nabla F^B + \nabla(H_A \nabla F^A),$$

which leads to (A1) by setting $G_{AB} = \bar{G}_{AB} - H_{(A,B)}$ and dropping the divergence term. However, it is known that $L_{\text{Ein}}(F^A_s) \equiv 0$, and hence we have

$$\nabla[(X_A + \text{const} \circ H_A)\nabla F^A]_s \equiv 0.$$

Let the F^A be functionals of a function λ which satisfies

$$\nabla^2\lambda = 0, \quad (\nabla\lambda)^2 \neq 0$$

and denote $\mathring{F}^A = dF^A/d\lambda$. Then the field equations (A2) become

$$\mathring{F}^A{}_{;B}\mathring{F}^B = 0, \tag{A8}$$

which can be solved using the well-known methods for the geodesic equation. The general solution involves $2n$ constants, thus giving a $2n$-parameter solution of the field equations.[14]

A very nice special case occurs if ∇ is the two-dimensional derivative operator $\binom{\partial_1}{\partial_2}$. Suppose that this is the case and G_{AB} admits a hypersurface-orthogonal nonnull Killing vector

$$X_{[A;BC]} = 0 = \left(\frac{1}{X_C X^C} X_{[A}\right)_{;B]}.$$

Then, by a proper choice of the F^A, the Lagrangian can be written

$$L = H_{AB}\nabla F^A \nabla F^B + V(\nabla K)^2 \tag{A9}$$

and the field equations for K read

$$\nabla(V\nabla K) = 0$$

implying the existence of a function M, defined by

$$\nabla^* M = V\nabla K. \tag{A10}$$

(∇^* denotes the conjugated derivative operator: $\nabla^{*\alpha} = \epsilon^\alpha_\beta \nabla^\beta$.)

The new Lagrangian as a function of ∇M is obtained by performing a Legendre transformation

$$L' = L - \frac{\partial L}{\partial \nabla K}\nabla K. \tag{A11}$$

It may easily be checked that this new Lagrangian gives the correct field-equations in terms of M.[15]

As the transformation from K to M via (A10) is by no means contained in (A3), the new Lagrangian leads to new (A3)-type transformations, which are quite distinct from the old ones of the old Lagrangian.

The method may be generalized to Lagrangians of the type

$$L = G_{AB}\nabla F^A \nabla F^B + S(F^C).$$

This is invariant under an (A3) transformation iff

$$X_{(A;B)} = S_{,A}X^A = 0$$

and the field equations are

$$\nabla^2 F^A + \Gamma^A_{BC}\nabla F^B \nabla F^C = G^{AB}S_{,B}.$$

Note added in proof: Another generalization is Lagrangians of the type

$$L = \xi_A \nabla F^A,$$

where ξ_A is a vector, whose components are functions of the F^A, in the coordinate space (coordinate indices suppressed) as well as in the manifold of the F^A. Note, that in this case no metric structure for the F^A is required. The Lagrangian is invariant under (A3) if

$$D_{L(X)}\xi = 0$$

and the field equations are

$$\xi_{[A,B]}\nabla F^A = 0,$$

where D_L denotes the Lie derivative in the F^A-manifold.

APPENDIX B

The Lagrangians (3) or (4) can be written as the metric on an hypersurface S in a flat 5-space.

By the transformation

$$\zeta=\tfrac{1}{2}k+\tfrac{3}{4}\ln r+r,\quad \iota=r^{1/2}l,$$
$$\kappa=\tfrac{1}{2}r^{1/2}(m-n),\quad \eta=\tfrac{1}{2}r^{1/2}(m+n) \tag{B1}$$
$$\lambda=-\tfrac{1}{2}k-\tfrac{3}{4}\ln r+r$$

we get

$$2L=\nabla\zeta^2+\nabla\iota^2+\nabla\kappa^2-\nabla\eta^2-\nabla\lambda^2,$$
$$S:\ \zeta+\lambda=2(\iota^2+\kappa^2-\eta^2) \tag{B2}$$

as expressions for the Lagrangian and the hypersurface. This hypersurface can be shown to be not lightlike.

Some solutions of Eq. (A4) may now be obtained by the following theorem:

Let there be given in an n-dimensional Riemann space a nonnull hypersurface with unit normal vector n_α, Riemann tensor $\bar{R}^\alpha_{\beta\gamma\delta}$, and covariant derivative formed with respect to the induced metric on the hypersurface, and let ξ_α satisfy

$$\text{(i) } \xi_\alpha n^\alpha=0,$$
$$\text{(ii) } D_{L(n)}\xi=0, \tag{B3}$$
$$\text{(iii) } \xi_{\alpha;\beta\gamma}=R_{\alpha\beta\gamma\delta}\xi^\delta;$$

then it also satisfies

$$\xi_{\alpha\|\beta\gamma}=\bar{R}_{\alpha\beta\gamma\delta}\xi^\delta. \tag{B4}$$

As the proof of this theorem is easy using standard books,[16] we will not give it here.[17]

However, as the 5-space is flat, the solutions of (B3iii) are easily to be found as

$$\xi^A=a^A_B\xi^B+C^A,$$

involving 30 constants a^A_B, C^A, which one has to choose according to (B3i, ii), finding the surviving solution to be given by (5).

Whether the form (B2) of the Lagrangian (3) [of course, a similar treatment is possible with (12)] is useful for finding new solutions will be the subject of a forthcoming paper.

We note that the full Einstein Lagrangian, involving ten functions, may be imbedded in a flat 55-space.

APPENDIX C

As a sketch of the application of the methods described in this paper, we want to discuss one example.

We examine the behavior of the Schwarzschild-solution under an infinitesimal transformation of L'' [Eq. (15)].

In prolate spheroidal coordinates the Schwarzschild metric is for unit mass

$$ds^2=(\cosh u+1)^2(du^2+d\theta^2+\sin^2\theta\, d\phi^2)-\frac{\cosh u-1}{\cosh u+1}dt^2$$

or

$$k=\ln(\cosh u+1),\quad m=-\frac{\cosh u-1}{\cosh u+1},$$
$$n=(\cosh u+1)^2\sin^2\theta,\quad l=0. \tag{C1}$$

This leads by direct calculation via Eqs. (9), (11), (14), (16) to the variables

$$k''=\ln(\cosh u-1),\quad m''=-\frac{\cosh u+1}{\cosh u-1},$$
$$n''=(\cosh u-1)^2\sin^2\theta,\quad l''=0,$$

used in Eqs. (15), (16). Applying now the infinitesimal transformation [Eq. (5)]

$$m''^*=m''+\delta l',\quad l''^*=l''+\tfrac{1}{2}\delta n'',$$
$$n''^*=n'',\quad k''^*=k''$$

gives

$$l''^*=\tfrac{1}{2}\delta(\cosh u-1)^2\sin^2\theta$$

while the other variables remain unchanged. From Eq. (16) it can easily be calculated that

$$\left(\frac{\psi}{f^2+\psi^2}\right)^*=\delta(\cosh u+2)\cos\theta$$

and one obtains

$$\omega^*=-\tfrac{1}{2}\delta(\cosh^2 u+4\cosh u+7)\sin^2\theta$$

and finally

$$l^*=-\tfrac{1}{2}\delta\sin^2\theta(\cosh^2 u+4\cosh u+7)\frac{\cosh u-1}{\cosh u+1}. \tag{C2}$$

m, n, k are unchanged due to the fact that only infinitesimal δ was considered.

[1]R. Geroch, J. Math. Phys. 12, 918 (1971).
[2]R. Geroch, J. Math. Phys. 13, 394 (1972).
[3]W. Kinnersley, J. Math. Phys. 14, 651 (1973).
[4]C. Hoenselaers, Univ. Bielefeld Preprint, Bi 5/75.
[5]R. O. Hansen, J. Math. Phys. 15, 46 (1974).
[6]A. Papapetrou, Ann. Inst. H. Poincaré 4, 83 (1966).

[7]This form of the line element can be obtained without using Einstein's equations. One of the equations will turn out to be $\nabla^2 r=0$, which enables us to choose r as one coordinate and z, defined by $\nabla^* z=\nabla r$, as the other. But as then the Lagrangian will become explicitly coordinate dependent, we do not intend to use this possibility.
[8]For actual computations (4) is the more convenient.

[9]R. Matzner and C. Misner, Phys. Rev. 154, 1229 (1967).
[10]F. Ernst, Phys. Rev. 167, 1175 (1968).

[11]We omit the ξ-Killing-vector here, because a Legendre transformation would just lead backwards to (10).

[12]G. Neugebauer and D. Kramer, Ann. Physik 24, 62 (1969).
[13]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1949), p. 142.

[14]It may be noted that solutions of this type lead either for the Lagrangian (3) or (12) to asymptotically nonflat metrics.
[15]A Legendre transformation of this kind is known in the theory of the gyroscope, when passing from the Lagrange function to the Routh function.

[16]S. Hawking and G. Ellis, *The Large Scale Structure of Space—Time* (Cambridge U. P., Cambridge, 1973), pp. 44ff.

[17]We only mention that some of the Killing vectors of the hypersurface may be found by

$$\xi_{(\alpha;\beta)}=0,\quad \xi_\alpha n^\alpha=0.$$

The fact, that no condition like (B3ii) has to be imposed arises from the Killing equation being a first-order equation, while (B3iii) is a second-order equation.

Homothetic and conformal motions in spacelike slices of solutions of Einstein's equations*

Beverly K. Berger

Physics Department, Yale University, New Haven, Connecticut 06520
(Received 1 March 1976)

Components of Killing's equation are used to obtain constraints satisfied in a spacelike hypersurface by the intrinsic metric and extrinsic curvature in the presence of a spacetime conformal motion for a solution of Einstein's equations. If the conformal motion is either a homothetic motion or a motion, it is shown that these Killing constraints are preserved by the Einstein evolution equations. It is then shown that the generator of the homothetic motion (homothetic Killing vector) can be constructed if the Killing constraints are satisfied by a set of initial data. It is shown that a homothetic motion in the intrinsic metric is a spacetime homothetic motion if the extrinsic curvature is transformed correctly under the spatial homothetic motion. Further restrictions on a proper conformal motion due to the fact that it is not identically a curvature collineation are obtained. Restrictions on the matter–stress–energy tensor are discussed. Examples are presented.

I. INTRODUCTION

To consider a solution of Einstein's equations as the time evolution of an initial spacelike hypersurface has proven useful in, for example, quantization of gravity,[1] the gravitational initial value problem,[2] and the computer evolution of colliding black holes and collapsing stars.[3,4] It is interesting to derive the constraints imposed on the initial data in the hypersurface due to the presence in the spacetime of some generalized symmetry.[5] Here we discuss the effect in the spacelike hypersurface of a spacetime conformal motion (conformal symmetry) which has as two special cases a homothetic motion (self-similarity) and a motion (isometry).[6–8]

We first consider the results for a solution of Einstein's equations *in vacuo* in the presence of a homothetic motion or motion: It is found that, in addition to the Hamiltonian and momentum constraints, the spatial metric and extrinsic curvature must satisfy additional Killing constraints which depend on the components of the generator of the homothetic motion (homothetic Killing vector) normal and tangent to the hypersurface. These constraints are obtained easily using the components of the spacetime Killing's equation and its Lie derivative normal to the hypersurface. It is shown that these constraints are preserved by Einstein's equations. These results have previously been obtained in the special case of an isometry by Moncrief,[9] although the methods used here (due to Geroch[10] as elaborated by Smarr[3]) are much more straightforward. A great simplification is achieved by using the conformal Killing vector field as the coordinate congruence.[11]

Conversely, it is shown by constructing the homothetic Killing vector that a set of initial data on a spacelike hypersurface which satisfies these Killing constraints in addition to the Einstein constraints also satisfies the spacetime Killing's equation. It is also shown that a spatial homothetic Killing vector which satisfies the hypersurface constraints for a homothetic Killing vector with no normal component is a spacetime homothetic Killing vector.[12]

Next, the more general case of a conformal motion which is neither a homothetic motion nor a motion is considered: In this case, also, the spatial projection of Killing's equation and its time evolution yield constraints in a spacelike hypersurface. These constraints are *not* preserved by Einstein's evolution equations which in fact restrict the conformal motions compatible with solutions of Einstein's equations. Additional restrictions in the hypersurface arise from requiring the Lie derivative along the conformal Killing vector of the Einstein constraints to be zero. These restrictions and those formed by their time derivatives are so severe that, as shown by Collinson and French[6] using Newman–Penrose techniques, only type N spacetimes with twist-free geodesic rays can have a proper conformal motion. These restrictions serve, in general, to prevent a conformal motion in spacelike initial data from being a spacetime conformal motion.

In Sec. II, the basic notation and formalism[3,10] is presented. Section III contains the projections of Killing's equation and the derivation of the Killing constraints for a spacetime possessing a conformal motion. In Sec. IV, the constraints are presented for the special case of a homothetic motion and are shown to be conserved (using Appendix A). It is then shown possible to construct a homothetic Killing vector from the constraints.

In Sec. V, it is shown using Appendix A that for a proper conformal motion the Killing constraints are not preserved by Einstein's equations. Further restrictions are obtained using the Einstein constraints. In Sec. VI, the modifications required in the presence of matter are discussed. The results are summarized in Sec. VII.

In Appendix B three examples are presented. The first, the dust-filled Friedmann universe with flat spacelike slices,[13] is shown to contain a proper conformal motion. The second, the empty Kasner universe, is shown to contain a homothetic Killing vector.[12] For the third, Schwarzschild in the usual coordinates,[13] the restriction is obtained which prevents a spatial conformal Killing vector from being a spacetime conformal Killing vector.

II. DEFINITIONS AND NOTATION

Assume there exists a spacetime (ST) with metric

g_{ab} which satisfies Einstein's equations *in vacuo*

$$R_{ab} - \tfrac{1}{2} g_{ab} R = 0, \quad (1)$$

where $a, b = 0, 1, 2, 3$. The ST is assumed to possess a conformal motion generated by the vector field ξ^a such that

$$\pounds_\xi g_{ab} = \sigma g_{ab}, \quad (2)$$

where $\pounds$ denotes Lie derivative defined by $\pounds_\xi g_{ab} = \nabla_a \xi_b + \nabla_b \xi_a$ (for ∇ the covariant derivative with respect to g_{ab}) and σ is an arbitrary function. The vector ξ^a is said to be a conformal Killing vector (CKV). If $\sigma = c$, a constant, the ST possesses a homothetic motion (self-similarity) and ξ^a is a homothetic Killing vector (HKV). If $c = 0$, the ST possesses a motion (isometry) and ξ^a is a Killing vector (KV).[5]

Consider a spacelike hypersurface S_t, defined as having constant coordinate time t, embedded in the ST.[14] The vector

$$n^a = -\lambda \nabla_a t \quad (3)$$

(where $\lambda\, dt$ is the proper time along n between S_t and S_{t+dt}) is the future pointing unit timelike normal to S_t. Thus S_t has the intrinsic metric

$$h_{ab} = g_{ab} + n_a n_b \quad (4)$$

and extrinsic curvature

$$K_{ab} = -h^c{}_a h^d{}_b \nabla_c n_d. \quad (5)$$

Einstein's equations (1) become four constraint equations

$$K_{ab} K^{ab} - K^2 - \mathcal{R} = 0 \quad (6)$$

and

$$D_a K^a{}_b - D_b K = 0 \quad (7)$$

(where $K \equiv h^{ab} K_{ab}$, indices on spatial tensors are raised and lowered with h_{ab} and h^{ab}, and $\mathcal{R}$ is the curvature scalar formed from h_{ab} using the intrinsic covariant derivative D defined[10] by $D_a A_{bc} = h^e{}_a h^f{}_b h^d{}_c \nabla_e A_{fd}$ such that $D_a h_{bc} = 0$) and twelve evolution equations

$$(\pounds_v - \pounds_\mu) h_{ab} = -2\lambda K_{ab} \quad (8)$$

and[15]

$$(\pounds_v - \pounds_\mu) K_{ab} = -2\lambda K_{ac} K^c{}_b + \lambda K K_{ab} + \lambda \mathcal{R}_{ab} - D_a D_b \lambda \quad (9)$$

(where $\mathcal{R}_{ab}$ is the Ricci tensor formed from D and h_{ab}). The vector field $v^a = \lambda n^a + \mu^a$ has been chosen as the coordinate congruence so that $\pounds_v \to \partial/\partial t$ and the shift vector μ^a joins points on S_t and S_{t+dt} with the same spatial coordinates. It will simplify subsequent calculations to remark that Eq. (9) may be rewritten[3] as

$$(\pounds_v - \pounds_\mu) K_{ab} = -\lambda K_{ac} K^c{}_b - \lambda P_{ab} - D_a D_b \lambda, \quad (10)$$

where

$$P_{ab} = -h_m{}^r h_a{}^n h_b{}^p R^m{}_{npr} \quad (11)$$

for R_{mnpr} a component of the ST Riemann tensor. Comparing Eqs. (9) and (10) gives[3]

$$P_{ab} = -\mathcal{R}_{ab} - K K_{ab} + K_{ca} K^c{}_b. \quad (12)$$

III. PROJECTIONS OF KILLING'S EQUATION AND THE KILLING CONSTRAINTS

It is now possible to obtain the components of Killing's equation (2) with respect to the hypersurface S_t. If the CKV is written as $\xi^a = \alpha n^a + \beta^a$ (where $\beta^a n_a = 0$), the normal component of Eq. (2) is, for an arbitrary coordinate congruence $v^a = \lambda n^a + \mu^a$,

$$\pounds_{\lambda n} \alpha = -\pounds_\beta \lambda + \tfrac{1}{2}\sigma\lambda, \quad (13)$$

where we have used the relation[3]

$$n^a \nabla_a n_b = D_b \ln\lambda. \quad (14)$$

The mixed component of Eq. (2) yields

$$\pounds_{\lambda n} \beta^a = \lambda D^a \alpha - \alpha D^a \lambda, \quad (15)$$

as is seen from Eqs. (5), (13), and (14) and the fact that K_{ab} is a symmetric tensor.

The spatial projection and its time evolution are found most simply by taking as the coordinate congruence v the conformal Killing vector field $\xi^a = \alpha n^a + \beta^a$—i.e., the lapse function is taken to be the negative of the normal component of the CKV and the shift vector to be its spatial projection.[11]

With this choice of coordinates, the normal component of Eq. (2) may be written

$$\pounds_\xi \alpha = \tfrac{1}{2}\sigma\alpha \quad (16)$$

while Eq. (15) becomes

$$\pounds_{\alpha n} \beta^a = 0 \quad (17)$$

or equivalently

$$\pounds_\xi n^a = -\tfrac{1}{2}\sigma n^a. \quad (18)$$

The spatial component of Eq. (2) yields a constraint on S_t:

$$D_a \beta_b + D_b \beta_a - 2\alpha K_{ab} = \sigma h_{ab}. \quad (19)$$

Since Eq. (18) is sufficient to show that

$$D_a \beta_b + D_b \beta_a = \pounds_\beta h_{ab} \quad (20)$$

(where it is recalled that $\pounds$ signifies a ST Lie derivative), Eq. (8) may be used to rewrite Eq. (19) as

$$\pounds_\xi h_{ab} = \sigma h_{ab}. \quad (21)$$

To obtain an additional constraint on S_t, operate on Eq. (21) with $\pounds_\xi - \pounds_\beta = \pounds_{\alpha n}$ and use the relation[16]

$$\pounds_v \pounds_w h_{ab} = \pounds_w \pounds_v h_{ab} + \pounds_{\pounds_v w} h_{ab}. \quad (22)$$

From Eqs. (8), (9), (16), and (18) and the relation

$$\pounds_\beta K_{ab} = {}^3\pounds_\beta K_{ab} \quad (23)$$

for ${}^3\pounds_\beta A_{ab} \equiv \beta^c D_c A_{ab} + A_{cb} D_a \beta^c + A_{ac} D_b \beta^c$ it is easily shown that

$$\pounds_\xi K_{ab} = \tfrac{1}{2}\sigma K_{ab} - \tfrac{1}{2} h_{ab} \pounds_n \sigma \quad (24)$$

or

$${}^3\pounds_\beta K_{ab} - 2\alpha K_{ac} K^c{}_b + \alpha K K_{ab} + \alpha \mathcal{R}_{ab} - D_a D_b \alpha = \tfrac{1}{2}\sigma K_{ab} - \tfrac{1}{2} h_{ab} \pounds_n \sigma. \quad (25)$$

It is to be noted that Eqs. (19) and (25) are constraints which must be satisfied by the intrinsic metric and extrinsic curvature on S_t in the presence of a CKV. These equations are independent of coordinate choice where α, β^a are components of the CKV and σ and $\pounds_n\sigma$ are arbitrary. This may be seen directly by rederiving the constraints using arbitrary coordinates or by noting that Eq. (25) is the spatial projection of the invariant equation[17]

$$\pounds_\xi n_a\Gamma^a{}_{bc} = \tfrac{1}{2}\sigma n_a\Gamma^a{}_{bc} - \tfrac{1}{2}g_{bc}\pounds_n\sigma + n_{(b}\nabla_{c)}\sigma \tag{26}$$

since $K_{ab} = h^c_a h^d_b n_e \Gamma^e{}_{cd}$ where the relation for the Lie derivative of the affine connection has been used. [The convenient notation of Eqs. (19) and (24) is not coordinate independent since it requires $v = \xi$ in Eqs. (8) and (9).]

IV. RESULTS FOR A HOMOTHETIC MOTION

Under a homothetic motion any geometric object with dimension (length)^q transforms with a factor $\exp(qc/2)$ if the metric transforms as $g_{ab} \to e^c g_{ab}$.[7] Thus for any object φ (with arbitrary indices),

$$\pounds_\xi \varphi = \tfrac{1}{2}qc\varphi. \tag{27}$$

For $\sigma = c$, the components of Killing's equation, Eqs. (16), (18), (21) and the constraint Eq. (24) become

$$\pounds_\xi \alpha = \tfrac{1}{2}c\alpha, \tag{28}$$

$$\pounds_\xi n^a = -\tfrac{1}{2}cn^a, \tag{29}$$

$$\pounds_\xi h_{ab} = ch_{ab}, \tag{30}$$

and

$$\pounds_\xi K_{ab} = \tfrac{1}{2}cK_{ab}. \tag{31}$$

From the definitions of α, n^a, h_{ab}, and K_{ab}, it is clear that the values of q in Eqs. (28)–(31) are those expected from dimensional analysis where coordinates are assumed to be dimensionless.

For further analysis it is useful to rewrite Eqs. (30) and (31) in the coordinate independent form as Proposition 1.

Proposition 1: If a vacuum spacetime possesses a homothetic motion $\pounds_\xi g_{ab} = cg_{ab}$, then in any spacelike hypersurface of the spacetime the intrinsic metric, h_{ab}, and extrinsic curvature, K_{ab}, must satisfy

$$D_a\beta_b + D_b\beta_a = 2\alpha K_{ab} + ch_{ab} \tag{32}$$

and

$$\beta^c D_c K_{ab} + K_{ac}D_b\beta^c + K_{cb}D_a\beta^c = 2\alpha K_{ac}K^c{}_b - \alpha K K_{ab} - \alpha \mathcal{R}_{ab} + D_a D_b \alpha + \tfrac{1}{2}cK_{ab}, \tag{33}$$

where $\xi^a = \alpha n^a + \beta^a$ is the homothetic Killing vector.

For $c = 0$, the spacetime possesses an isometry and we have the result found by Moncrief[9] and Berezdivin[18]:

Corollary 1: If a vacuum spacetime possesses an isometry with generator ξ, then in any spacelike hypersurface

$$D_a\beta_b + D_b\beta_a = 2\alpha K_{ab} \tag{34}$$

and

$$\beta^c D_c K_{ab} + K_{ac}D_b\beta^c + K_{cb}D_a\beta^c = 2\alpha K_{ac}K^c{}_b - \alpha K K_{ab} - \alpha \mathcal{R}_{ab} + D_a D_b \alpha \tag{35}$$

for $\xi^a = \alpha n^a + \beta^a$.

Proposition 2: The Killing equation constraints (32) and (33) are conserved by the Einstein evolution equations.

Proof: Since the evolution of Eq. (32) yields Eq. (33) (Sec. III), we need only show that Eq. (33) is conserved. Evolve with $(\pounds_\xi - \pounds_\beta)$ using Eqs. (8) and (9) for $v^a = \xi^a$. Equations (29), (17), and (22) (for K_{ab} rather than h_{ab}) may be used to give in these coordinates

$$\pounds_\xi[(\pounds_\xi - \pounds_\beta)K_{ab}] = \tfrac{1}{2}c(\pounds_\xi - \pounds_\beta)K_{ab}. \tag{36}$$

The left-hand side of Eq. (36) is calculated explicitly by specializing the calculation in Appendix A to the case $\sigma = c$, a constant, which eliminates all terms in derivatives of σ. It is immediately seen from Eq. (A19) that an identity results.

Proposition 2 has been proved for isometries by Moncrief.[9] It is now possible to prove

Proposition 3: If in a spacelike hypersurface, S_t, a set of initial data satisfy, in addition to the Einstein constraints,

$$D_a\beta_b + D_b\beta_a = 2\alpha K_{ab} + ch_{ab} \tag{37}$$

and

$$\beta^c D_c K_{ab} + K_{ac}D_b\beta^c + K_{cb}D_a\beta^c = 2\alpha K_{ac}K^c{}_b - \alpha K K_{ab} - \alpha \mathcal{R}_{ab} + D_a D_b \alpha + \tfrac{1}{2}cK_{ab} \tag{38}$$

for some scalar α and spatial vector β^a, then the spacetime development of S_t contains a homothetic motion with generator $\xi^a = \alpha n^a + \beta^a$.

Proof: To evolve S_t, choose the coordinate congruence $\xi^a = \alpha n^a + \beta^a$. It is easy to show using calculations in Appendix A and

$$\pounds_{\alpha n} h^b{}_a = 0 \tag{39}$$

[which follows from Eqs. (4) and (14)] that $(\pounds_\xi - \pounds_\beta)$ operating on Eq. (37) yields Eq. (38) and $(\pounds_\xi - \pounds_\beta)$ operating on Eq. (38) yields an identity if and only if Eqs. (17) and (28) are true. But Eqs. (37), (38), (17), and (28) are equivalent to Killing's equation (2) for $\sigma = c$. Thus ξ^a is an HKV. Therefore, Eqs. (37) and (38) may be used to construct an HKV if h_{ab}, K_{ab} are known (in any coordinates).

Examples are given in Appendix B. Proposition 3 has been proved for isometries by Moncrief[9] and Berezdivin.[18]

Corollary 3: A homothetic motion in initial data such that

$$^3\pounds_\beta h_{ab} = ch_{ab} \tag{40}$$

and

$$^3\pounds_\beta K_{ab} = \frac{c}{2}K_{ab} \tag{41}$$

is a homothetic motion of the spacetime.

Proof: For a coordinate congruence $v^a = \lambda n^a$ repeat the proof of Proposition 3 since Eqs. (40) and (41) are just Eqs. (37) and (38) with $\alpha=0$. We require Eqs. (13) and (15) with $\alpha=0$ and thus have Killing's equation (2) for $\sigma=c$ with $\xi^a=\beta^a$.

Corollary 3 has previously been obtained by Eardley[7] for self-similar cosmologies and as restricted to isometries by Berezdivin.[18]

As an aside, we mention that it is easy to show that for any HKV ξ

$$\pounds_\xi(K_{ab}K^{ab}-K^2-\mathcal{R})=-c(K_{ab}K^{ab}-K^2-\mathcal{R}) \tag{42}$$

and

$$\pounds_\xi(D_aK^{ab}-D^bK)=-\tfrac{1}{2}c(D_aK^{ab}-D^bK) \tag{43}$$

so that the Einstein constraints cannot spoil the compatibility of Einstein's equations and a homothetic motion. (This has been pointed out for isometries by Berezdivin.[18])

V. PROPER CONFORMAL MOTIONS ($\sigma \neq$ CONST)

It is now possible to show that the constraints (19) and (25) are not preserved by Einstein's equations for a proper conformal motion by operating on Eq. (25) with $\pounds_{\alpha n}$ and showing that an identity does not result. Using Eq. (22) for K_{ab} rather than h_{ab} to commute Lie derivatives yields

$$\pounds_\xi(\pounds_\xi-\pounds_\beta)K_{ab}=\tfrac{1}{2}\sigma(\pounds_\xi-\pounds_\beta)K_{ab}+\tfrac{3}{2}\alpha K_{ab}\pounds_n\sigma - \tfrac{1}{2}\alpha h_{ab}\pounds_n\pounds_n\sigma. \tag{44}$$

The term by term calculation of the left hand side of Eq. (44) using Eq. (10) is given in Appendix A. The result of the direct calculation is

$$\pounds_\xi(\pounds_\xi-\pounds_\beta)K_{ab}=\tfrac{1}{2}\sigma(\pounds_\xi-\pounds_\beta)K_{ab}+\tfrac{1}{2}\alpha K_{ab}\pounds_n\sigma - \alpha D_aD_b\sigma-\tfrac{1}{2}\alpha h_{ab}D_cD^c\sigma-\tfrac{1}{2}\alpha h_{ab}K\pounds_n\sigma - \tfrac{1}{2}h_{ab}(D_c\alpha)(D^c\sigma), \tag{45}$$

which when compared with Eq. (44) yields the relation (for an arbitrary coordinate congruence $v^a=\lambda n^a+\mu^a$)

$$-D_aD_b\sigma-K_{ab}\pounds_n\sigma+\tfrac{1}{2}h_{ab}[\pounds_n\pounds_n\sigma-\lambda^{-1}(D^c\lambda)(D_c\sigma) - D_cD^c\sigma-K\pounds_n\sigma]=0. \tag{46}$$

Equation (46) is a restriction on σ.

In addition, the Lie derivative of the Einstein constraint Eqs. (6) and (7) require

$$\pounds_\xi(K_{ab}K^{ab}-K^2-\mathcal{R})=0 \tag{47}$$

and

$$\pounds_\xi(D_aK^a{}_b-D_bK)=0. \tag{48}$$

Direct calculation of Eq. (47) using Eq. (6), the trace of Eq. (12), and Eq. (A11) yields the restriction

$$D_cD^c\sigma+K\pounds_n\sigma=0. \tag{49}$$

Direct calculation of Eq. (48) using the trace of[3]

$$-D_aK_{bc}+D_bK_{ac}=h_a{}^mh_b{}^nh_c{}^pn^rR_{mnpr} \tag{50}$$

and Eq. (A6) yields

$$D_b\pounds_n\sigma+K_{ab}D^a\sigma=0. \tag{51}$$

It is immediately seen that Eqs. (46), (49), and (51) are not automatically satisfied for σ nonconstant. The restrictions arise because a conformal motion must be a conformal collineation[5]—i.e., Eq. (2) implies

$$\pounds_\xi C^a{}_{bcd}=0, \tag{52}$$

where $C^a{}_{bcd}$ is the Weyl tensor. In a vacuum ST Eq. (52) is equivalent to

$$\pounds_\xi R^a{}_{bcd}=0. \tag{53}$$

Although Eq. (52) is an identity, Eq. (53) is not and becomes a restriction on σ. Various components and contractions of Eq. (53) have been used by Collison and French[6] to show that only type N vacuum ST's with twist-free geodesic rays can possess conformal motions with σ nonconstant.

The additional restrictions imposed on proper conformal motions by Eqs. (46), (49), (51) and their time derivatives allow the statement of

Proposition 4: A proper conformal motion in initial data such that

$$^3\pounds_\beta h_{ab}=\sigma h_{ab} \tag{54}$$

and

$$^3\pounds_\beta K_{ab}=\tfrac{1}{2}\sigma K_{ab}-\tfrac{1}{2}h_{ab}\pounds_n\sigma \tag{55}$$

is, in general, *not* a conformal motion of the spacetime.

The example of the Schwarzschild solution is presented in Appendix B.

VI. NONVACUUM SPACETIMES

Here we briefly outline the modifications which must be made in the presence of an arbitrary stress–energy tensor (with all constants absorbed)[3]

$$T_{ab}=\rho n_an_b-2n_{(a}t_{b)}+S_{ab}, \tag{56}$$

where t_a and S_{ab} lie in S_t. The Einstein constraints (6) and (7) become

$$K_{ab}K^{ab}-K^2-\mathcal{R}=-2\rho \tag{57}$$

and

$$D_aK^a{}_b-D_bK=-t_b \tag{58}$$

while the evolution equation (8) for h_{ab} is unchanged and that for K_{ab} becomes[3]

$$(\pounds_v-\pounds_\mu)K_{ab}=-2\lambda K_{ac}K^c{}_b+\lambda KK_{ab}+\lambda\mathcal{R}_{ab} - D_aD_b\lambda-\lambda S_{ab}+\tfrac{1}{2}\lambda h_{ab}T, \tag{59}$$

where $T\equiv h^{ab}T_{ab}$. The Killing constraint equation (19) is unchanged while Killing constraint equation (25) picks up a term $\lambda S_{ab}-\frac{1}{2}\lambda h_{ab}T$ on the right-hand side. This is also true in all other rewritings of Eq. (25) such as Eqs. (35), (38), etc. For a homothetic motion, the constraints are conserved if

$$\pounds_\xi T_{ab}=0, \tag{60}$$

a result expected from dimensional analysis.

For a proper conformal motion, the restrictions (46), (49), and (51) are obtained by replacing the zero on the right-hand sides of Eqs. (46), (49), and (51) with $\pounds_\xi S_{ab}-\frac{1}{2}h_{ab}\pounds_\xi T$, $-\pounds_\xi\rho$, and $-\pounds_\xi t_b$ respectively. The complexity of these constraints suggests that solutions of the nonvacuum Einstein equations possessing proper conformal motions must be rare. An example, however, is given in Appendix B.

VII. DISCUSSION

The compatibility between homothetic motions and Einstein's equations has been shown by obtaining the constraints satisfied by spacelike initial data in the presence of a spacetime homothetic Killing vector and showing that these constraints are conserved by Einstein equations. It is also shown that a spacetime homothetic Killing vector can be constructed from a set of initial data which satisfy the Killing constraints.

The compatibility cannot be generalized to a proper conformal motion since the constraints obtained in this case are not in general preserved by Einstein's equations. Thus a proper conformal motion in spacelike initial data is in general not a spacetime proper conformal motion.

ACKNOWLEDGMENTS

I wish to thank Douglas M. Eardley and Larry Smarr for many useful discussions concerning this work.

APPENDIX A: COMPUTATION OF $\pounds_\xi(\pounds_\xi - \pounds_\beta)K_{ab}$

From Eq. (10),

$$(\pounds_\xi - \pounds_\beta)K_{ab} = -\alpha K_{ac}K^c{}_b - \alpha P_{ab} - D_aD_b\alpha. \tag{A1}$$

Consider the Lie derivative of each term on the right-hand side of Eq. (A1): From Eqs. (24), (16), and

$$\pounds_\xi h^{ab} = -\sigma h^{ab} \tag{A2}$$

direct substitution yields

$$\pounds_\xi(-\alpha K_{ac}K^c{}_b) = -\tfrac{1}{2}\sigma\alpha K_{ac}K^c{}_b + \alpha K_{ab}\pounds_n\sigma. \tag{A3}$$

From the definition of P_{ab}, Eq. (11), and Eqs. (16), (18), (A2), and

$$\pounds_\xi n_a = \tfrac{1}{2}\sigma n_a, \tag{A4}$$

it is easy to show that

$$-\pounds_\xi \alpha P_{ab} = -\tfrac{1}{2}\sigma\alpha P_{ab} - h^r{}_m h^n{}_a h^p{}_b \pounds_\xi R^m{}_{npr}. \tag{A5}$$

It is well known that[5,16]

$$\begin{aligned}\pounds_\xi R^m{}_{npr} = \tfrac{1}{2}g^{ms}\{&\nabla_p[\nabla_n(\pounds_\xi g_{rs}) + \nabla_r(\pounds_\xi g_{ns}) \\ &- \nabla_s(\pounds_\xi g_{nr})] - \nabla_r[\nabla_n(\pounds_\xi g_{ps}) \\ &+ \nabla_p(\pounds_\xi g_{sn}) - \nabla_s(\pounds_\xi g_{pn})]\}.\end{aligned} \tag{A6}$$

Substituting Eq. (2) in (A6) yields

$$\begin{aligned}\pounds_\xi R^m{}_{npr} = \tfrac{1}{2}(&\delta^m{}_r\nabla_n\nabla_p\sigma - g_{rn}\nabla^m\nabla_p\sigma \\ &- \delta^m{}_p\nabla_n\nabla_r\sigma + g_{pn}\nabla^m\nabla_r\sigma).\end{aligned} \tag{A7}$$

Substitution of Eq. (A7) in Eq. (A5) gives

$$\begin{aligned}-\pounds_\xi\alpha P_{ab} = &-\tfrac{1}{2}\sigma\alpha P_{ab} - \tfrac{1}{2}\alpha h^n{}_a h^p{}_b\nabla_n\nabla_p\sigma \\ &- \tfrac{1}{2}\alpha h_{ab}h^r{}_m\nabla^m\nabla_r\sigma.\end{aligned} \tag{A8}$$

The second term on the right-hand side of Eq. (A8) becomes

$$\begin{aligned}-\tfrac{1}{2}\alpha h^n{}_a h^p{}_b\nabla_n\nabla_p\sigma &= -\tfrac{1}{2}\alpha h^n{}_a h^p{}_b\nabla_n g_p{}^s\nabla_s\sigma \\ &= -\tfrac{1}{2}\alpha D_aD_b\sigma - \tfrac{1}{2}\alpha K_{ab}\pounds_n\sigma\end{aligned} \tag{A9}$$

while the third term becomes

$$\begin{aligned}&-\tfrac{1}{2}\alpha h_{ab}h^r{}_m\nabla^m\nabla_r\sigma \\ &\quad= -\tfrac{1}{2}\alpha h_{ab}D_cD^c\sigma - \tfrac{1}{2}\alpha h_{ab}K\pounds_n\sigma.\end{aligned} \tag{A10}$$

Combining Eqs. (A8)—(A10) yields

$$\begin{aligned}\pounds_\xi(-\alpha P_{ab}) = &-\tfrac{1}{2}\sigma\alpha P_{ab} - \tfrac{1}{2}\alpha D_aD_b\sigma - \tfrac{1}{2}\alpha K_{ab}\pounds_n\sigma \\ &- \tfrac{1}{2}\alpha h_{ab}D_cD^c\sigma - \tfrac{1}{2}\alpha h_{ab}K\pounds_n\sigma.\end{aligned} \tag{A11}$$

Furthermore,

$$\begin{aligned}\pounds_\xi(-D_aD_b\alpha) &= -\pounds_\xi(h^c_a h^d_b\nabla_c h^e_d\nabla_e\alpha) \\ &= -h^c_a h^d_b\pounds_\xi\nabla_c\nabla_d\alpha + \pounds_\xi(K_{ab}\pounds_n\alpha).\end{aligned} \tag{A12}$$

But

$$\begin{aligned}\pounds_\xi\nabla_c\nabla_d\alpha &= \tfrac{1}{2}\pounds_\xi\pounds_{\nabla\alpha}g_{ab} \\ &= \tfrac{1}{2}\pounds_{\nabla\alpha}(\sigma g_{ab}) + \tfrac{1}{2}\pounds_{\pounds_\xi\nabla\alpha}g_{ab}.\end{aligned} \tag{A13}$$

Now

$$\pounds_\xi\nabla^a\alpha - \nabla^a\pounds_\xi\alpha = -(\nabla^c\alpha)\sigma g^a_c. \tag{A14}$$

Thus Eq. (A13) becomes

$$\pounds_\xi\nabla_c\nabla_d\alpha = \tfrac{1}{2}\sigma\nabla_c\nabla_d\alpha + \tfrac{1}{2}\alpha\nabla_c\nabla_d\sigma + \tfrac{1}{2}g_{ab}(\nabla^c\alpha)(\nabla_c\sigma), \tag{A15}$$

and

$$\begin{aligned}\pounds_\xi(K_{ab}\pounds_n\alpha) = &\tfrac{1}{2}\sigma K_{ab}\pounds_n\alpha - \tfrac{1}{2}h_{ab}(\pounds_n\alpha)(\pounds_n\sigma) \\ &+ \tfrac{1}{2}\alpha K_{ab}\pounds_n\sigma.\end{aligned} \tag{A16}$$

Using

$$(\nabla^c\alpha)(\nabla_c\sigma) = (D^c\alpha)(D_c\sigma) - (\pounds_n\alpha)(\pounds_n\sigma) \tag{A17}$$

and combining Eqs. (A12)—(A17) gives

$$\pounds_\xi(-D_aD_b\alpha) = -\tfrac{1}{2}\sigma D_aD_b\alpha - \tfrac{1}{2}\alpha D_aD_b\sigma - \tfrac{1}{2}h_{ab}(D_c\alpha)(D^c\sigma). \tag{A18}$$

Combining Eqs. (A3), (A11), and (A18) gives

$$\begin{aligned}\pounds_\xi(\pounds_\xi - \pounds_\beta)K_{ab} = &\tfrac{1}{2}\sigma(\pounds_\xi - \pounds_\beta)K_{ab} + \tfrac{1}{2}\alpha K_{ab}\pounds_n\sigma \\ &- \alpha D_aD_b\sigma - \tfrac{1}{2}\alpha h_{ab}D_cD^c\sigma \\ &- \tfrac{1}{2}\alpha h_{ab}K\pounds_n\sigma - \tfrac{1}{2}h_{ab}(D_c\alpha)(D^c\sigma).\end{aligned} \tag{A19}$$

APPENDIX B: ILLUSTRATIVE EXAMPLES OF ST's POSSESSING GENERALIZED SYMMETRIES[19]

1. ST with CKV

Consider the dust-filled Friedmann universe with zero spatial curvature:

$$ds^2 = -dt^2 + a^2(t)\delta_{ij}dx^i dx^j \quad (i, j = 1, 2, 3), \tag{B1}$$

where

$$T_{ab} = \rho\delta_{a0}\delta_{b0} \tag{B2}$$

for

$$\rho = \rho_0(a/a_0)^{-3} \tag{B3}$$

and

$$a = a_0 t^{2/3}. \tag{B4}$$

For the coordinates of Eq. (B1),

$$n^a = (1, 0, 0, 0) \quad \text{(B5)}$$

so that $\pounds_n = \partial/\partial t$. From Eq. (8) in the coordinates of (B1), it is seen that

$$K_{ij} = -a\dot{a}\delta_{ij} \quad \text{(B6)}$$

where $\cdot \equiv d/dt$. The simplest case is to assume that there exists a CKV parallel to n^a. Then Eqs. (19) and (25) yield

$$\alpha = a, \quad \sigma = 2\dot{a}. \quad \text{(B7)}$$

Thus the CKV is $\xi^a = (a, 0, 0, 0)$ or

$$\pounds_\xi = \frac{\partial}{\partial\eta} \quad \text{for } a\,d\eta = dt, \quad \text{(B8)}$$

as is expected. Direct substitution using Eqs. (B7), (B2)–(B4), and (B8) shows that Eqs. (46), (49), and (51) as amended in Sec. VI for $T_{ab} \neq 0$ are satisfied in this case.

2. ST with an HKV[20]

Consider the empty Kasner universe with

$$ds^2 = -dt^2 + t^{2p_1}\,dx^2 + t^{2p_2}\,dy^2 + t^{2p_3}\,dz^2, \quad \text{(B9)}$$

where the p_i are constants such that

$$\sum p_i = 1 = \sum p_i^2. \quad \text{(B10)}$$

For the coordinates of Eq. (B9), Eq. (8) yields

$$K_{ij} = -p_i\, t^{2p_i - 1}\delta_{ij}. \quad \text{(B11)}$$

Substitution in Eqs. (32) and (33) yields [assuming $\alpha = \alpha(t)$]

$$\alpha = \tfrac{1}{2}ct, \quad \beta^i = \tfrac{1}{2}c(1-p_i)x^i \quad \text{(B12)}$$

so that

$$\xi = (t; (1-p_i)x^i) \quad \text{(B13)}$$

is an HKV. The HKV of Eq. (B13) is well known[7] and may be found directly from Eq. (2).

3. A spatial CKV in Schwarzschild

The spherically symmetry spacelike hypersurfaces S_t in the Schwarzschild solution are conformally flat.[3] The metric (exterior) is

$$ds^2 = -(1-2m/r)\,dt^2 + (1-2m/r)^{-1}\,dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \quad \text{(B14)}$$

From Eq. (8)

$$K_{ij} = 0. \quad \text{(B15)}$$

Conformal flatness of the S_t implies that there exists β^i in S_t such that

$${}^3\pounds_\beta h_{ij} = \sigma h_{ij} \quad \text{(B16)}$$

for nonconstant σ. To satisfy the constraint Eq. (55) and Eq. (B15), it is necessary to choose

$$\pounds_n \sigma = 0. \quad \text{(B17)}$$

But

$$n^a = (1-2m/r)^{-1/2}(1, 0, 0, 0), \quad \text{(B18)}$$

so that Eq. (B17) is satisfied by choosing

$$\frac{\partial\sigma}{\partial t} = 0. \quad \text{(B19)}$$

For the coordinate system of Eq. (B14), Eq. (46), becomes [using (B19) and (B15)]

$$\sigma_{|ij} - \tfrac{1}{2}h_{ij}\sigma_{|l}\alpha^{|l} + \tfrac{1}{2}h_{ij}\sigma_{|l}{}^{l} = 0, \quad \text{(B20)}$$

where α is the lapse function in the metric (B14). But Eq. (49) becomes

$$\sigma_{|l}{}^{l} = 0, \quad \text{(B21)}$$

so that Eq. (B21) and the trace of Eq. (B20) force

$$\sigma_{|l}\alpha^{|l} = 0. \quad \text{(B22)}$$

Since $\alpha_{|l} \neq 0$, Eq. (B22) and spherical symmetry imply $\sigma =$ const. Thus β^i cannot be a proper CKV of the ST.

*Research supported in part by the National Science Foundation under Grant No. GP-36317.

[1]See, for example, M. Ryan, *Hamiltonian Cosmology* (Springer-Verlag, Berlin, 1972).

[2]N. O'Murchadha and J.W. York, Jr., Phys. Rev. D **10**, 428 (1974); J.A. Wheeler, in *Battelle Rencontres*, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, 1968), Chap. IX.

[3]L. Smarr, Ph.D. thesis (University of Texas, 1974).

[4]D.M. Eardley, Phys. Rev. D 12, 3072 (1975); D.M. Eardley and L. Smarr, to be published; D.R. Brill, to be published.

[5]For a discussion of the relationships among various general symmetries see G.H. Katzin, J. Levin, and W.R. Davis, J. Math. Phys. 10, 617 (1969).

[6]A discussion of vacuum spacetimes which possess conformal motions is given in C.D. Collinson and D.C. French, J. Math. Phys. 8, 701 (1967).

[7]An extensive discussion of self-similar cosmological models is found in D.M. Eardley, Commun. Math. Phys. 37, 287 (1974).

[8]See also B.B. Godfrey, General Relativity Gravitation 3, 3 (1972); C.B.G. McIntosh, to be published; R. Sigal, to be published.

[9]V. Moncrief, J. Math. Phys. 16, 493 (1975).

[10]R. Geroch, J. Math. Phys. 13, 956 (1972).

[11]This is due to a suggestion by J.M. Bardeen (private communication).

[12]This has also been shown by Eardley (Ref. 7).

[13]See, for example, C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[14]The notation used here is that of Refs. 3 and 10. To obtain the more standard notation of Ref. 13, for a coordinate congruence $v^a = \alpha n^a + \beta^a$ set $v^a = (1,0,0,0)$; if A_{ab} is spatial $A_{ab} \to A_{ij}$, $i,j = 1,2,3$, $D_a\beta_b \to \beta_{i|j}$, $\pounds_v \to \partial/\partial t$.

[15]Equation (9) differs from that given in Refs. 13 and 9 by a factor of $\tfrac{1}{2}\times$ Eq. (6) (which is, of course, zero).

[16]K. Yano, *Theory of Lie Derivatives* (North-Holland, Amsterdam, 1955).

[17]Here the relation

$$\pounds_\xi(\Gamma^a{}_{bc}) = \tfrac{1}{2}g^{ad}[\nabla_c(\pounds_\xi g_{db}) + \nabla_b(\pounds_\xi g_{dc}) - \nabla_d(\pounds_\xi g_{bc})]$$

has been used. See Refs. 5 and 16.

[18]R. Berezdivin, J. Math. Phys. 15, 1963 (1974).

[19]For the purposes of these calculations, 3+1 notation rather than projective notation is most useful. See Ref. 14 for the relation between the two. These relations are proven in Ref. 3.

[20]This example is given by Eardley (Ref. 7).

A supereigenvalue problem for the solution of the generator coordinate integral equation*

L. Lathouwers†

Faculty of Science, Rijksuniversitair Centrum, Antwerp, Belgium
(Received 16 March 1976)

The unconventional Griffin–Hill–Wheeler integral equation is replaced by a classical Fredholm eigenvalue problem which is the continuous analog of Löwdin's supersecular equation.

The generator coordinate method is a variational procedure based on trial functions of the form[1]

$$\psi(x)=\int f(\alpha)\,\phi(x\,|\,\alpha)\,d\alpha. \tag{1}$$

The many particle functions $\phi(x|\alpha)$ are termed intrinsic states and can conveniently be chosen for the problem under consideration. In order to represent bound states they are assumed to be square integrable (L^2) in the dynamical variables x for all values of the parameters α. The folding of $\phi(x|\alpha)$ with superposition amplitudes $f(\alpha)$ through the integration over the α range generates a class of trial functions $\psi(x)$ depending solely on the particle degrees of freedom. The parameters α are generally referred to as generator coordinates (GC). The amplitudes $f(\alpha)$ which make the expectation value of the Hamiltonian H with respect to (1) stationary satisfy the, Griffin–Hill–Wheeler (GHW) integral equation

$$\int[H(\alpha,\beta)-E\Delta(\alpha,\beta)]f(\beta)\,d\beta=0, \tag{2}$$

where

$$H(\alpha,\beta)=\int \phi^*(x\,|\,\alpha)\,H\phi(x\,|\,\beta)\,dx, \tag{3}$$

$$\Delta(\alpha,\beta)=\int \phi^*(x\,|\,\alpha)\,\phi(x\,|\,\beta)\,dx, \tag{4}$$

are the Hermitian Hamiltonian and metric kernels. For bound state problems this equation has to be solved under the restriction that $\psi(x)$ is L^2,

$$\|\psi\|=\left[\int\int f^*(\alpha)\,\Delta(\alpha,\beta)f(\beta)\,d\alpha\,d\beta\right]^{1/2}=\|f\|_\Delta, \tag{5}$$

i.e., the amplitudes $f(\alpha)$ the amplitudes $f(\alpha)$ should have a finite norm in the nondiagonal $\Delta(\alpha,\beta)$ metric. The eigenvalues E_i of the GC eigenvalue problem (2) and (5) are upper bounds to the exact eigenvalues of H. The properties of the corresponding eigenamplitudes $f_i(\alpha)$ have recently been studied.[2] These quantities can behave in a variety of ways, ranging from L^2 functions over tempered distributions to singular non-L^2 functions, depending on a balance between dynamical and geometrical factors. Numerical methods, such as discretization techniques, for solving the GHW equation should therefore be handled with extreme care.

Both theoretically and practically it would be convenient to replace the unconventional GHW integral equations by classical Fredholm equations.[3] One way of doing this has been proposed in Ref. 2. In the following, I will present an alternative procedure which is the integral equation analog of Löwdin's super-secular equation[4] for treating the diagonalization of H in a finite nonorthogonal basis. The first step is to renormalize the intrinsic function $\phi(x|\alpha)$, i.e., to multiply it with an L^2 function in α, nonzero almost everywhere, such that it is L^2 both in x and α. If $\phi(x|\alpha)$ is in the domain of H, i.e., when $H\phi(x|\alpha)$ is L^2 in x for all α, the renormalization function can be chosen such that the metric kernel and the Hamiltonian kernel are L^2 kernels. Thus there exist two positive constants C_Δ and C_H for which

$$\int\int|\Delta(\alpha,\beta)|^2\,d\alpha\,d\beta=C_\Delta<+\infty, \tag{6}$$

$$\int\int|H(\alpha,\beta)|^2\,d\alpha\,d\beta=C_H<+\infty, \tag{7}$$

Instead of solving the GHW equation one can consider the associated equation

$$\int K(\alpha,\beta\,|\,E)\,g(\beta\,|\,E)\,d\beta=\lambda(E)\,g(\alpha\,|\,E), \tag{8}$$

i.e., the formal eigenvalue problem for the Hermitian kernel $K(\alpha,\beta|E)=H(\alpha,\beta)-E\Delta(\alpha,\beta)$. Taking over the terminology of Löwdin, (8) will be referred to as the super-GHW equation. If the kernels $H(\alpha,\beta)$ and $\Delta(\alpha,\beta)$ are L^2 so is $K(\alpha,\beta|E)$. Indeed one readily verifies that

$$\int|K(\alpha,\beta\,|\,E)|^2\,d\alpha\,d\beta\leq C_H+E^2C_\Delta+2|E|C_H C_\Delta \tag{9}$$

showing that $K(\alpha,\beta|E)$ is L^2 for all E. Hence the super-GHW equation is a classical Hermitian Fredholm equation of the second kind. Its eigenvalues $\lambda_i(E)$ and corresponding L^2 eigenfunctions $g_i(\alpha|E)$ will depend upon the energy as a parameter. Taking the partial derivative of (8) with respect to E, multiplying to the left with $g^*(\alpha|E)$, and integrating over α, one obtains a closed expression for the derivative of the supereigenvalues

$$\lambda_i'(E)=\frac{(-)\int\int g_i^*(\alpha\,|\,E)\,\Delta(\alpha,\beta)\,g_i(\beta\,|\,E)\,d\alpha\,d\beta}{\int\int g_i^*(\alpha\,|\,E)\,g_i(\alpha\,|\,E)\,d\alpha}. \tag{10}$$

Thus the supereigenvalues are monotonically decreasing functions of energy. Each of them will cut the E axis in a certain point E_i at which the super-GHW equation reduces to

$$\int[H(\alpha,\beta)-E_i\Delta(\alpha,\beta)]\,g_i(\beta\,|\,E_i)\,d\beta=0. \tag{11}$$

Identifying this expression with the original GHW equation, one concludes that the zeros E_i of the supereigenvalues $\lambda_i(E)$ equal the GHW eigenvalues while the eigenamplitudes $f_i(\alpha)$ can be identified with $g_i(\alpha|E_i)$. It follows from the Fredholm theory that $f_i(\alpha)$, if L^2, is orthogonal to all eigenfunctions $g_j(\alpha|E_i)$, with $i\neq j$, of the kernel $K(\alpha,\beta|E_i)$. Rewriting (10) at $E=E_i$ yields

$$\lambda_i'(E_i) = (-)\left(\frac{\|f_i\|_\Delta}{\|f_i\|}\right)^2 . \qquad (12)$$

Therefore, if a solution of the GHW equation is non-L^2 this will be marked by a zero slope crossing of the associated supereigenvalue and the energy axis. The shape of the supereigenvalues betrays non-L^2 solutions of the GHW equation. In order to find E_i one has to locate the zero point of $\lambda_i(E)$. This can be done, e.g., by the Newton—Raphson method[5] based on the iteration of the equation

$$E_i^{(n+1)} = E_i^{(n)} - \left(\frac{\lambda_i(E_i^{(n)})}{\lambda_i'(E_i^{(n)})}\right), \qquad (13)$$

where $E_i^{(0)}$ is an initial guess for E_i. Inserting the explicit expressions for $\lambda_i(E)$ and $\lambda_i'(E)$ one obtains

$$E_i^{(n+1)} = \frac{\iint g_i^*(\alpha \mid E_i^{(n)}) H(\alpha,\beta)\, g_i(\beta \mid E_i^{(n)})\, d\alpha\, d\beta}{\iint g_i^*(\alpha \mid E_i^{(n)}) \Delta(\alpha,\beta)\, g_i(\beta \mid E_i^{(n)})\, d\alpha\, d\beta} . \qquad (14)$$

Hence the Newton—Raphson method is equivalent to the iteration of the GC energy expectation value of the solution of the super-GHW equation. This procedure might not be the best one to determine the zeros of the supereigenvalues but it has the advantage of being closely related to the GC variational principle. From the numerical point of view the replacement of the GHW-equation by the super-GHW equation has the advantage that for the latter one, being an Hermitian Fredholm equation, accurate quadrature methods exist. Under quite general conditions[6] on the kernel $K(\alpha,\beta \mid E)$, error bounds for the supereigenvalues can be given. The problem of approximate linear dependence, occurring in the discretization of the GHW equation due to the nondiagonal metric $\Delta(\alpha,\beta)$, does not arise at all.

*Work supported by the Interuniversitair Instituut voor Kernwetenschappen, Belgium and the Swedish Institute, Sweden.
†Presently at the Department of Quantum Chemistry, Uppsala University, P.O. Box 518, S-751 20 Uppsala, Sweden.
[1]J. J. Griffin and J. A. Wheeler, Phys. Rev. **108**, 311 (1957).
[2]L. Lathouwers, submitted for publication in Ann. Phys.
[3]F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[4]P. O. Löwdin, Int. J. Quantum Chem. **S1**, 811 (1967).
[5]A. Ralston, *A First Course in Numerical Analysis* (McGraw-Hill, New York, 1965).
[6]H. Wielandt, *Proceedings of the Sixth Symposium in Applied Mathematics of the American Mathematical Society* (McGraw-Hill, New York, 1956).

Measure-theoretic representations of the SU(2) ⊗ SU(2) current algebra with PCAC

Zbigniew Haba

Institute of Theoretical Physics, University of Wrocław, Wrocław, Poland
(Received 27 June 1975; revised manuscript received 15 October 1975)

The SU(2)⊗SU(2) current algebra, with scalar fields φ^a σ describing PCAC, is considered. The algebra is exponentiated in order to get a group C. These exponentiated currents are represented in the space L^2_μ of functions defined on S' (Gel'fand–Vilenkin formalism). The unitary representation is described by a quasi-invariant measure μ on S'. Then, the representation of the SU(2)⊗SU(2) current algebra extended by the fields can be obtained, with currents and fields defined as self-adjoint operators with a dense invariant domain in Hilbert space. The Gaussian measure μ_B gives an example of a measure describing representations of the SU(2)⊗SU(2) current algebra with conserved currents. Such a theory is nonrelativistic (ultralocal). In order to obtain chiral symmetry breaking and PCAC, the Gaussian measure is not sufficient and should be modified by the singular interaction factor giving a new measure $d\mu_1 = \lim_{h\to 1}\{\int \exp[-H_I(h)]\,d\mu_B\}^{-1}\exp[-H_I(h)]\,d\mu_B$. This model is discussed, and the existence of the charges is shown.

I. INTRODUCTION

The success of the Gell-Mann–Feynman current algebra[1] may be treated as a justification of some basic principles of the theory. Because of this success there is the conviction that currents play a fundamental role in hadron physics. Furthermore, it has been assumed that currents can be treated as the basic canonical variables, "coordinates" of hadrons.[2,3] In such a case our starting point should be the commutation relations among the currents, which must be represented in Hilbert space. The representation theory of the nonrelativistic current algebra has been developed by G. A. Goldin[4] (see also earlier works[5]). These representations have found fruitful applications in the statistical physics.[6] It is interesting to investigate current algebras, which are relevant in the relativistic theory, although some difficulties may appear in this case.[7]

In our earlier paper[8] we have considered representations of the SU(2) current algebra with a c-number Schwinger term. The present work deals with a much larger algebra admitting the partial conservation of the axial current (PCAC), provided an appropriate Hamiltonian is chosen. We are using exponentiated currents and fields $\varphi^a\sigma$. From the commutation relations of the SU(2)⊗SU(2) current algebra extended by fields, through the process of formal exponentiation we get the commutation relations among the exponentiated currents and fields (Sec. III) which form a group C. These commutation relations are our starting point for further considerations. Unitary representations are described by a quasi-invariant measure μ defined on S' (Sec. IV). If some additional assumptions are fulfilled, we can obtain representations of the current algebra (Sec. VI) with currents and fields defined as self-adjoint operators with a dense invariant domain in the Hilbert space L^2_μ. As an example of a measure, which fulfils all the above conditions, we consider the Gaussian measure (Sec. V). As is well known, this measure gives also representations of the canonical commutation relations for fields. If we want the currents to be constructed from fields in a local way and to be conserved, we have to restrict ourselves to nonrelativistic theory (we obtain so-called ultralocal theory[9,10] In order to obtain PCAC (Sec. VII), the SU(2)⊗SU(2) symmetry should be broken. We can do it similarly as in the σ-model[1,11] (the linear term is sufficient). Then we have to modify the Gaussian measure by the factor $\exp[\int f_\pi \mu^2 h(\mathbf{x})\sigma(\mathbf{x})\,d\mathbf{x}] = \exp[-H_I(h)]$[12] ($h$ denotes the cutoff). The cut-off can be removed giving a measure μ_1. In the Hilbert space $L^2_{\mu_1}$ we can define the self-adjoint Hamiltonian, which ensures PCAC. We show further that the formal charge $Q^a = \int V_0^a(\mathbf{x})\,d\mathbf{x}$ exists and is the generator of SU(2) symmetry. The formal charge $Q_5^a = \int A_0^a(\mathbf{x})\,d\mathbf{x}$ does not exist, but it defines a bilinear form, which is the form of an operator Q_5^a.[13] The vacuum is not annihilated by Q_5^a.[14]

II. THE CURRENT ALGEBRA

We consider the usual commutation relations (with a c-number Schwinger term) among the vector and axial SU(2)⊗SU(2) currents[1,3,15]:

$$\begin{aligned}
&[V_0^a(\mathbf{x}), V_0^b(\mathbf{y})] = i\epsilon^{abc}V_0^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[A_0^a(\mathbf{x}), A_0^b(\mathbf{y})] = i\epsilon^{abc}V_0^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[V_0^a(\mathbf{x}), A_0^b(\mathbf{y})] = i\epsilon^{abc}A_0^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[V_0^a(\mathbf{x}), V_k^b(\mathbf{y})] = i\epsilon^{abc}V_k^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}) - iC\delta^{ab}\partial_k\delta(\mathbf{x}-\mathbf{y}),\\
&[A_0^a(\mathbf{x}), A_k^b(\mathbf{y})] = i\epsilon^{abc}A_k^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}) - iC\delta^{ab}\partial_k\delta(\mathbf{x}-\mathbf{y}),\\
&[V_0^a(\mathbf{x}), A_k^b(\mathbf{y})] = i\epsilon^{abc}A_k^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[A_0^a(\mathbf{x}), V_k^b(\mathbf{y})] = i\epsilon^{abc}A_k^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}).
\end{aligned}\tag{II.1}$$

The remaining commutators vanish

$$[A_k^a(\mathbf{x}), A_l^b(\mathbf{y})] = [V_k^a(\mathbf{x}), V_l^b(\mathbf{y})] = [V_k^a(\mathbf{x}), A_l^b(\mathbf{y})] = 0. \tag{II.2}$$

If we wish to obtain PCAC, we must supplement to the currents $V_\mu^a A_\mu^a$ scalar fields $\varphi^a\sigma$. Here, φ^a is the pseudoscalar, isovector pion field and σ is a scalar, isoscalar field. The natural choice of the commutation relations among the fields and the currents is the folfollowing[16,29]:

$$\begin{aligned}
&[V_0^a(\mathbf{x}), \varphi^b(\mathbf{y})] = i\epsilon^{abc}\varphi^c(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[A_0^a(\mathbf{x}), \varphi^b(\mathbf{y})] = i\delta^{ab}\sigma(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&[A_0^a(\mathbf{x}), \sigma(\mathbf{y})] = -i\varphi^a(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),
\end{aligned}\tag{II.3}$$

$$[V^a_\mu(\mathbf{x}),\sigma(\mathbf{y})]=[A^a_k(\mathbf{x}),\sigma(\mathbf{y})]=$$
$$=[V^a_k(\mathbf{x}),\varphi^b(\mathbf{y})]=[A^a_k(\mathbf{x}),\varphi^b(\mathbf{y})]=0.$$

The currents and the fields are operator valued distributions.[17] So, they should be smeared out with test functions from $\mathcal{S}$. We shall use the following notation[8]:

$$\mathbf{V}_0(\mathbf{f})=\sum_{a=1}^{3} V^a_0(f^a),\quad \vec{\mathbf{V}}(\vec{\mathbf{g}})=\sum_{a=1}^{3}\sum_{k=1}^{3} V^a_k(g^a_k),\quad \mathbf{A}_0(\mathbf{h})=\sum_{a=1}^{3} A^a_0(h^a)$$

$$\vec{\mathbf{A}}(\vec{\mathbf{q}})=\sum_{a=1}^{3}\sum_{k=1}^{3} A^a_k(q^a_k),\quad \boldsymbol{\varphi}(\mathbf{l})=\sum_{a=1}^{3}\varphi^a(l^a),\quad \text{and}$$

$$\sigma(b)=\int\sigma(\mathbf{x})b(\mathbf{x})\,d\mathbf{x} \tag{II.4}$$

After smearing out with test functions, the current algebra (II.1), (II.2), (II.3) takes the form

$$\begin{aligned}
&[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]=i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}'),\\
&[\mathbf{A}_0(\mathbf{h}),\mathbf{A}_0(\mathbf{h}')]=i\mathbf{V}_0(\mathbf{h}\times\mathbf{h}'),\\
&[\mathbf{V}_0(\mathbf{f}),\mathbf{A}_0(\mathbf{h})]=i\mathbf{A}_0(\mathbf{f}\times\mathbf{h}),\\
&[\mathbf{V}_0(\mathbf{f}),\vec{\mathbf{V}}(\vec{\mathbf{g}})]=i\vec{\mathbf{V}}(\mathbf{f}\times\vec{\mathbf{g}})-iC\int\vec{\partial}\mathbf{f}\vec{\mathbf{g}}\,d\mathbf{x},\\
&[\mathbf{A}_0(\mathbf{h}),\vec{\mathbf{A}}(\vec{\mathbf{q}})]=i\vec{\mathbf{A}}(\mathbf{h}\times\vec{\mathbf{q}})-iC\int\vec{\partial}\mathbf{h}\vec{\mathbf{q}}\,d\mathbf{x},\\
&[\mathbf{A}_0(\mathbf{h}),\vec{\mathbf{V}}(\vec{\mathbf{g}})]=i\vec{\mathbf{A}}(\mathbf{h}\times\vec{\mathbf{g}}),\\
&[\mathbf{V}_0(\mathbf{f}),\boldsymbol{\varphi}(\mathbf{l})]=i\boldsymbol{\varphi}(\mathbf{f}\times\mathbf{l}),\\
&[\mathbf{A}_0(\mathbf{h}),\boldsymbol{\varphi}(\mathbf{l})]=i\sigma(\mathbf{h}\cdot\mathbf{l}),\\
&[\mathbf{A}_0(\mathbf{h}),\sigma(b)]=-i\boldsymbol{\varphi}(b\mathbf{h}).
\end{aligned} \tag{II.5}$$

Here $(\vec{\partial}\mathbf{f})^a_k=\partial_k f^a$ $(\mathbf{f}\times\vec{\mathbf{q}})^a_k=\epsilon^{abc}f^b q^c_k$. The remaining commutators are equal to zero.

III. THE EXPONENTIATED COMMUTATION RELATIONS

In order to work with bounded operators, we shall consider exponentiated currents. Using the formula

$$(\exp A)B[\exp(-A)]=\sum_{n=0}^{\infty}\frac{1}{n!}\underbrace{[A,\ldots,[A,B]]}_{n\text{ times}} \tag{III.1}$$

and the commutation relations (II.5), we get the formulas

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\mathbf{V}_0(\mathbf{f}')]\}\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}=\exp[i\mathbf{V}_0(R(\mathbf{f},-|\mathbf{f}|)\mathbf{f}')]; \tag{III.2}$$

here $R(\mathbf{f},-|\mathbf{f}|)\mathbf{f}'=\mathbf{f}'-(\underline{\mathbf{f}}\times\mathbf{f}')\sin|\mathbf{f}|-\underline{\mathbf{f}}\times(\underline{\mathbf{f}}\times\mathbf{f}')(\cos|\mathbf{f}|-1)$ is the rotation of the vector $\mathbf{f}'(\mathbf{x})$ around the vector $\mathbf{f}(\mathbf{x})$ $(|\mathbf{f}(\mathbf{x})|)^{-1}=\underline{\mathbf{f}}(\mathbf{x})$ by the angle $-|\mathbf{f}(\mathbf{x})|$:

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\cdot\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\}\cdot\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}=\exp\{i[R(\mathbf{f},-|\mathbf{f}|)\mathbf{g}+C\int_0^1 ds\int\vec{\partial}\mathbf{f}\,R(\mathbf{f},-s|\mathbf{f}|)\vec{\mathbf{g}}\,d\mathbf{x}]\}, \tag{III.3}$$

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\cdot\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\cdot\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}=\exp[i\mathbf{A}_0(R(\mathbf{f},-|\mathbf{f}|)\mathbf{h})], \tag{III.4}$$

$$\begin{aligned}
&\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\circ\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\}\cdot\{\exp[-i\mathbf{A}_0(\mathbf{h})]\}\\
&=\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}-\underline{\mathbf{h}}\times(\underline{\mathbf{h}}\times\vec{\mathbf{q}})(\cos|\mathbf{h}|-1))]\}\\
&\quad\cdot\{\exp[-i\mathbf{V}(\underline{\mathbf{h}}\times\vec{\mathbf{q}}\sin|\underline{\mathbf{h}}|)]\}\\
&\quad\cdot\{\exp[iC\int_0^1 ds\int\vec{\partial}\mathbf{h}\circ(\vec{\mathbf{q}}-\underline{\mathbf{h}}\times(\underline{\mathbf{h}}\times\vec{\mathbf{q}})(\cos|s\underline{\mathbf{h}}|-1)\,d\mathbf{x}]\},
\end{aligned} \tag{III.5}$$

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\circ\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\}\cdot\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}=\exp[i\vec{\mathbf{A}}(R(\mathbf{f},-|\mathbf{f}|\vec{\mathbf{q}})], \tag{III.6}$$

$$\begin{aligned}
&\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\cdot\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\}\cdot\{\exp[-i\mathbf{A}_0(\mathbf{h})]\}\\
&=\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\cdot\{\exp[i\sum_k\mathbf{V}_k(\underline{\mathbf{h}}\times(\underline{\mathbf{h}}\times\mathbf{g}_k)(\cos|\mathbf{h}|-1))]\}\\
&\quad\cdot\{\exp[-i\sum_k\mathbf{A}_k(\underline{\mathbf{h}}\times\mathbf{g}_k\sin|\mathbf{h}|)]\}\\
&\quad\cdot\{\exp[iC\sum_k\int\partial_k\mathbf{h}\int_0^1 ds\,\underline{\mathbf{h}}\times\mathbf{g}_k\sin|s\mathbf{h}|]\},
\end{aligned} \tag{III.7}$$

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\cdot\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}\cdot\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}=\exp[i\boldsymbol{\varphi}(R(\mathbf{f},-|\mathbf{f}|)\mathbf{l})], \tag{III.8}$$

$$\begin{aligned}
&\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\circ\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}\cdot\{\exp[-i\mathbf{A}_0(\mathbf{h})]\}\\
&=\{\exp[i\boldsymbol{\varphi}(\mathbf{l}+\underline{\mathbf{h}}(\underline{\mathbf{h}}\cdot\mathbf{l})(\cos|\mathbf{h}|-1)]\}\\
&\quad\cdot\{\exp[-i\sigma(\underline{\mathbf{h}}\cdot\mathbf{l}\sin|\mathbf{h}|)]\},
\end{aligned} \tag{III.9}$$

$$\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\circ\{\exp[i\sigma(b)]\}\cdot\{\exp[-\mathbf{A}_0(\mathbf{h})]\}=\exp[i\sigma(b\cos|\mathbf{h}|)]\exp[i\boldsymbol{\varphi}(\underline{\mathbf{h}}b\sin|\mathbf{h}|)]. \tag{III.10}$$

Our formal exponentiation procedure can be justified if there exists a dense invariant domain of analytic vectors.[18] We can treat, however, Eqs. (III.2)–(III.10) as our starting point for representation of the commutation relations (II.5). Only in our final stage (Sec. VI) it is shown that such a domain exists if we restrict ourselves to measures fulfilling some additional assumptions. The group of all such exponents (III.2)–(III.10) [if the algebra (II.5) is integrable to a group[18]] we shall call $\mathcal{C}$, and its elements will be denoted

$$\begin{aligned}
&(\mathcal{G}:\mathbf{f},\vec{\mathbf{g}},\mathbf{h},\vec{\mathbf{q}}:\,\mathbf{1},b)\\
&=\{\exp(i\mathcal{G})\}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\\
&\quad\times\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\}\{\exp[i\mathbf{A}_0(\mathbf{h})]\}\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\}\\
&\quad\times\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}\{\exp[i\sigma(b)]\}.
\end{aligned} \tag{III.11}$$

IV. REPRESENTATION THEORY

It is natural to assume in the relativistic theory[7] that the representation is cyclic under all currents and the fields $\varphi^a\sigma$. From this assumption it follows (see, e.g., Ref. 19) the cyclicity under the maximal commutative subalgebra $\mathcal{K}$ generated by $\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma$. So, we shall further assume the cyclicity,[20] and we denote the cyclic vector by Ω. Then, the functional[21]

$$L(\vec{\mathbf{g}},\vec{\mathbf{q}},\mathbf{l},b)=(\Omega,\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})\}\cdot\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\}$$

$$\cdot\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}\cdot\{\exp[i\sigma(b)]\},\Omega) \qquad (\text{IV}.1)$$

is positively definite, i.e.,

$$\sum_{i,j=1}^{n} \xi_i\bar{\xi}_j L(\vec{\mathbf{g}}_i-\vec{\mathbf{g}}_j,\vec{\mathbf{q}}_i,\vec{\mathbf{q}}_j,\mathbf{l}_i-\mathbf{l}_j,b_i-b_j)\geq 0 \qquad (\text{IV}.2)$$

for arbitrary $\xi_i,\vec{\mathbf{g}}_i,\vec{\mathbf{q}}_i,\mathbf{l}_i,b_i$.

It is the Fourier transform [22] of a measure μ defined on the direct sum of 22 dual spaces $\mathcal{S}'$

$$L(\vec{\mathbf{g}},\vec{\mathbf{q}},\mathbf{l},b) = \int\{\exp[i(\vec{\mathbf{V}},\vec{\mathbf{g}})]\}\cdot\{\exp[i(\vec{\mathbf{A}},\vec{\mathbf{q}})]\} \cdot\{\exp[i(\boldsymbol{\varphi},\mathbf{l})]\}\cdot\{\exp[i(\sigma,b)]\}\,d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma);$$

here

$$(\vec{\mathbf{V}},\vec{\mathbf{g}})=\sum_{a=1}^{3}\sum_{k=1}^{3}(V_k^a,g_k^a)$$

and similarly for $\vec{\mathbf{A}},\boldsymbol{\varphi}$, and we denote elements of the direct sum of dual spaces (distributions) by $\vec{\mathbf{V}}$, $\vec{\mathbf{A}}$, $\boldsymbol{\varphi}$, σ, correspondingly. So, to each representation of the commutative subgroup $\mathcal{K}$ of $\mathcal{C}$ there corresponds a measure on $\mathcal{S}'$ and vice versa; having the measure μ, we can construct the representation of $\mathcal{K}$ in the space $\mathcal{L}^2_\mu$ of square integrable (in μ) functions. Namely, Ω is represented by the function $1(\cdots)$ identically equal to 1 and[22]

$$\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) = \{\exp[i(\vec{\mathbf{V}},\vec{\mathbf{g}})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma),$$

$$\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) = \{\exp[i(\vec{\mathbf{A}},\vec{\mathbf{q}})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma),$$

$$\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) = \{\exp[i(\boldsymbol{\varphi},\mathbf{l})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma), \qquad (\text{IV}.3)$$

$$\{\exp[i\sigma(b)]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) = \{\exp[i(\sigma,b)]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma).$$

Now, we would like to solve the problem of whether the space $\mathcal{L}^2_\mu$ carries a unitary representation of the whole group $\mathcal{C}$. To answer this question we assume first that there exists a unitary representation of $\mathcal{C}$ in the space generated from Ω by the commutative subgroup $\mathcal{K}$. Then, we can find the action of this representation on the set of vectors

$$\sum_{ijkn} c_{ijkn}\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_i)]\}\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_j)]\} \times\{\exp[i\boldsymbol{\varphi}(\mathbf{l}_k)]\}\{\exp[i\sigma(b_n)]\}\Omega.$$

Namely {for example, taking $\exp[i\mathbf{V}_0(\mathbf{f})]$},

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\sum c_{ijkn}\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_i)]\}\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_j)]\} \times\{\exp[i\boldsymbol{\varphi}(\mathbf{l}_k)]\}\{\exp[i\sigma(b_n)]\}\Omega$$

$$=\sum c_{ijkn}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_i)]\}$$
$$\times\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_j)]\}$$
$$\times\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\boldsymbol{\varphi}(\mathbf{l})]\}$$
$$\times\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\sigma(b)]\}$$
$$\times\{\exp[-i\mathbf{V}_0(\mathbf{f})]\}\{\exp[i\mathbf{V}_0(\mathbf{f})]\}\Omega.$$

We can now use the commutation relations (III.2)–(III.11) to obtain the transformation of $\vec{\mathbf{g}},\vec{\mathbf{q}},\mathbf{l}$, and b. The action of $\exp[i\mathbf{V}_0(\mathbf{f})]$ on Ω is undetermined, but this unknown function in $\mathcal{L}^2_\mu$ can be fixed uniquely (up to the phase) from the assumption of unitarity. So, we reach the following (compare with Refs. 4, 22, 8):

Theorem I: There exists a unitary representation of the group $\mathcal{C}$ of exponentiated currents only if there exists a measure μ which is quasi-invariant under the action of $\mathcal{C}$ [Eqs. (III.2)–(III.10)]. In such a case the representation is given by the formulas (IV.3) and

$$\{\exp[i\mathbf{V}_0(\mathbf{f})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) =\left(\frac{d\mu(\mathbf{f}^*\vec{\mathbf{V}},\mathbf{f}^*\vec{\mathbf{A}},\mathbf{f}^*\boldsymbol{\varphi};\sigma)}{d\mu(\mathbf{V},\mathbf{A},\boldsymbol{\varphi},\sigma)}\right)^{1/2}F(\mathbf{f}^*\vec{\mathbf{V}},\mathbf{f}^*\vec{\mathbf{A}},\mathbf{f}^*\boldsymbol{\varphi},\sigma), \qquad (\text{IV}.4)$$

$$\{\exp[i\mathbf{A}_0(\mathbf{h})]\}F(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma) =\left(\frac{d\mu(\mathbf{h}^*\vec{\mathbf{V}},\mathbf{h}^*\vec{\mathbf{A}},\mathbf{h}^*\boldsymbol{\varphi},\mathbf{h}^*\sigma)}{d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma)}\right)^{1/2}F(\mathbf{h}^*\vec{\mathbf{V}},\mathbf{h}^*\vec{\mathbf{A}},\mathbf{h}^*\boldsymbol{\varphi},\mathbf{h}^*\sigma), \qquad (\text{IV}.5)$$

where $d\mu(\mathbf{h}^*\vec{\mathbf{V}},\mathbf{h}^*\vec{\mathbf{A}},\mathbf{h}^*\boldsymbol{\varphi},\mathbf{h}^*\sigma)/d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\boldsymbol{\varphi},\sigma)$ is the Radon–Nikodym derivative and the action of $\mathbf{f}^*$ and $\mathbf{h}^*$ is defined as follows

$$(\mathbf{f}^*\vec{\mathbf{V}},\vec{\mathbf{g}})=(\vec{\mathbf{V}},R(\mathbf{f},-|\mathbf{f}|)\vec{\mathbf{g}}) +C\int_0^1 ds\int\vec{\partial}\mathbf{f}R(\mathbf{f},-s|\mathbf{f}|)\vec{\mathbf{g}}\,d\mathbf{x}, \qquad (\text{IV}.6)$$

$$(\mathbf{f}^*\vec{\mathbf{A}},\vec{\mathbf{q}})=(\vec{\mathbf{A}},R(\mathbf{f},-|\mathbf{f}|)\vec{\mathbf{q}}), \qquad (\text{IV}.7)$$

$$(\mathbf{f}^*\boldsymbol{\varphi},\mathbf{l})=(\boldsymbol{\varphi},R(\mathbf{f},-|\mathbf{f}|)\mathbf{l}), \qquad (\text{IV}.8)$$

and

$$(\mathbf{h}^*\vec{\mathbf{V}},\vec{\mathbf{g}})=(\vec{\mathbf{V}},\vec{\mathbf{g}}-\underset{\sim}{\mathbf{h}}\times(\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{g}})(\cos|\mathbf{h}|-1)) +C\int_0^1 ds\int\vec{\partial}\mathbf{h}(\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{g}})(\sin s|\mathbf{h}|)\,d\mathbf{x} -(\vec{\mathbf{A}},\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{g}}\sin|\mathbf{h}|), \qquad (\text{IV}.9)$$

$$(\mathbf{h}^*\vec{\mathbf{A}},\vec{\mathbf{q}})=(\vec{\mathbf{A}},\vec{\mathbf{q}}-\underset{\sim}{\mathbf{h}}\times(\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{q}})(\cos|\mathbf{h}|-1)) -(\vec{\mathbf{V}},\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{q}}\sin|\mathbf{h}|) +C\int_0^1 ds\int\vec{\partial}\mathbf{h}((\vec{\mathbf{q}}-\underset{\sim}{\mathbf{h}}\times(\underset{\sim}{\mathbf{h}}\times\vec{\mathbf{q}})(\cos|s\underset{\sim}{\mathbf{h}}|-1)), \qquad (\text{IV}.10)$$

$$(\mathbf{h}^*\boldsymbol{\varphi},\mathbf{l})=(\boldsymbol{\varphi},\mathbf{l}+\underset{\sim}{\mathbf{h}}(\underset{\sim}{\mathbf{h}}\cdot\mathbf{l})(\cos|\underset{\sim}{\mathbf{h}}|-1))-(\sigma,\underset{\sim}{\mathbf{h}}\mathbf{l}\sin|\mathbf{h}|), \qquad (\text{IV}.11)$$

$$(\mathbf{h}^*\sigma,b)=(\sigma,b\cos|\mathbf{h}|)+(\boldsymbol{\varphi},\underset{\sim}{\mathbf{h}}b\sin|\mathbf{h}|). \qquad (\text{IV}.12)$$

From these formulas we can see that the chiral transformation generated by $\mathbf{h}^*$ mixes vector and axial currents, and scalar and pseudoscalar fields. It is so, because it is the $O(4)\sim SU(2)\otimes SU(2)$ rotation in the six-dimensional and four-dimensional spaces, correspondingly. Moreover, the transformation includes a translation by a function similarly to that for vector fields.

V. AN EXAMPLE OF A QUASIINVARIANT MEASURE

We shall consider the Gaussian measure.[22] Its Fourier transform is given by

$$(\int \exp[i(\varphi,f)]\,d\mu_B)=\exp[-B(f,f)/4], \tag{V.1}$$

where B is a bilinear form. On the cylinder sets Z,

$$Z=\{\varphi\in\mathcal{S}';((\varphi,f_1),(\varphi,f_2),\ldots,(\varphi,f_n))\subset R^n, f_i\in\mathcal{S}\}, \tag{V.2}$$

the measure $d\mu_B$ is given by the finite-dimensional measure

$$\mu_B(Z)=\int_A dx_1\cdots dx_n\{\exp(-\sum_{i,j=1}^{n} x_i(B^{-1})_{ij}x_j)\}\times\pi^{-n/2}(\det B)^{-1/2}, \tag{V.3}$$

where $B_{ij}=B(f_i,f_j)$. The Gaussian measure is quasi-invariant under translations,[22] and its Radon–Nikodym derivative is given by[22,23,12]

$$\frac{d\mu_B(\varphi+f)}{d\mu_B(\varphi)}=\exp[-(2(\varphi,K_x^{1/2}f)+(K_x^{1/2}f,f))], \tag{V.4}$$

where we assume[24] that

$$B(f,g)=(K_x^{-1/2}f,g)\quad\text{with}\quad(f,g)=\int f(\mathbf{x})g(\mathbf{x})\,d\mathbf{x}$$

and $K_x^{-1/2}$ is an operator in $L^2(R^3)$.

We would like to define now a quasi-invariant measure on the space of $\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma$. Because the test functions $\vec{\mathbf{g}},\vec{\mathbf{q}}$ form a basis of the $j_1=1$, $j_2=1$ representation of the group $O(4)\sim SU(2)\otimes SU(2)$ [25] and the functions 1, b a basis of the $j_1=1$, $j_2=0$ representation, we introduce the notation

$$S_i^v(\mathbf{x})=\begin{cases} g_i^a(\mathbf{x}) & \text{for } v=1,2,3,\ i=1,2,3,\\ q_i^b(\mathbf{x}) & \text{for } v=4,5,6,\ i=1,2,3,\end{cases}$$

$$T^u(\mathbf{x})=\begin{cases} l^a(\mathbf{x}) & \text{for } u=1,2,3,\\ b(\mathbf{x}) & \text{for } u=4.\end{cases}$$

As follows from (IV.6)–(IV.12) the transformation of the functions S,T is the composition of an $O(4)$ rotation in the isotopic space and a translation by a rotated function. In order to construct a quasi-invariant (under the action of $\mathbf{f}^*,\mathbf{h}^*$) measure, it is sufficient to use invariant under $O(4)$ rotations bilinear forms $B(S,S')$ and $B(T,T')$. They should form a scalar product in the $O(4)$ representation space. So, we get

$$B(S,S';T,T')=B(S,S')+B(T,T')$$
$$=\sum_{i,j=1}^{3}\sum_{v=1}^{6}\int S_i^v(\mathbf{x})\Delta_{ij}^S(\mathbf{x}-\mathbf{y})S_j^v(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}$$
$$+\sum_{u=1}^{4}\int T^u(\mathbf{x})\Delta^T(\mathbf{x}-\mathbf{y})T^u(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}. \tag{V.5}$$

The Δ functions should have such a form that the space rotations

$$T_R g_i^a(\mathbf{x})=R_{ik}g_k^a(R\mathbf{x}),\quad T_R q_i^a(\mathbf{x})=R_{ik}q_k^a(R\mathbf{x}),$$
$$T_R\varphi^a(\mathbf{x})=\varphi^a(R\mathbf{x}),\quad\text{and}\quad T_R\sigma(\mathbf{x})=\sigma(R\mathbf{x}) \tag{V.6}$$

leave the form B invariant. We could also introduce the chiral test functions

$$\vec{\mathbf{g}}_R=(\vec{\mathbf{g}}+\vec{\mathbf{q}})/2\quad\text{and}\quad g_L(\vec{\mathbf{g}}-\vec{\mathbf{q}})/2.$$

Then the rotations (IV.6)–(IV.12) would be simple $SU(2)$ rotations of $\vec{\mathbf{g}}_R$ and $\vec{\mathbf{g}}_L$ separately. Finally, let us underline that in such a space $\mathcal{L}^2_{\mu_B}$ we have a representation of the canonical commutation relations for vector fields.

VI. RECOVERING OF THE CURRENT ALGEBRA

We were begining with the current algebra (II.1)–(II.3) [or (II.5)]. But instead of considering its representations directly, we preferred to deal with the commutation relations (III.2)–(III.10). In such a way we omit some difficult problems concerning the domains of definition of the operators forming the algebra. It is now much easier to get representations of the algebra from the representations of the group $\mathcal{C}$. The currents are defined as generators of the corresponding one-parameter subgroups. They are self-adjoint on the set of vectors (in general different for different currents) on which the following strong limits exist (Stone's theorem)

$$\vec{\mathbf{V}}(\vec{\mathbf{g}})F=\lim_{t\to 0}\left(\frac{\exp[it\vec{\mathbf{V}}(\vec{\mathbf{g}})]-1}{it}\right)F, \tag{VI.1}$$

$$\vec{\mathbf{A}}(\vec{\mathbf{q}})F=\lim_{t\to 0}\left(\frac{\exp[it\vec{\mathbf{A}}(\vec{\mathbf{q}})]-1}{it}\right)F, \tag{VI.2}$$

$$\varphi(\mathbf{l})F=\lim_{t\to 0}\left(\frac{\exp[it\varphi(\mathbf{l})]-1}{it}\right)F, \tag{VI.3}$$

$$\sigma(b)F=\lim_{t\to 0}\left(\frac{\exp[it\sigma(\mathbf{l})]-1}{it}\right)F, \tag{VI.4}$$

$$\mathbf{V}_0(\mathbf{f})F=\lim_{t\to 0}\left(\frac{\exp[it\mathbf{V}_0(\mathbf{f})]-1}{it}\right)F, \tag{VI.5}$$

$$\mathbf{A}_0(\mathbf{h})F=\lim_{t\to 0}\left(\frac{\exp[it\mathbf{A}_0(\mathbf{h})]-1}{it}\right)F, \tag{VI.6}$$

where the action of the subgroups $\exp(itA)$ is given by Eqs. (IV.4)–(IV.12). Now, our treatment is quite similar to that in Ref. 4. So, we shall only outline the proofs. First of all we state

Theorem II:

(a)

$$[\vec{\mathbf{V}}(\vec{\mathbf{g}}_1),\vec{\mathbf{V}}(\vec{\mathbf{g}}_2)]=[\vec{\mathbf{A}}(\vec{\mathbf{q}}_1),\vec{\mathbf{A}}(\vec{\mathbf{q}}_2)]=[\vec{\mathbf{V}}(\vec{\mathbf{g}}),\vec{\mathbf{A}}(\vec{\mathbf{q}})]$$
$$=[\vec{\mathbf{V}}(\vec{\mathbf{g}}),\varphi(\mathbf{l})]=[\vec{\mathbf{A}}(\vec{\mathbf{q}}),\varphi(\mathbf{l})]=[\varphi(\mathbf{l}),\sigma(b)]$$
$$=[\vec{\mathbf{V}}(\vec{\mathbf{g}}),\sigma(b)]=[\vec{\mathbf{A}}(\vec{\mathbf{q}}),\sigma(b)]=0; \tag{VI.7}$$

(b)

$$[\mathbf{V}_0(\mathbf{f}),\vec{\mathbf{V}}(\vec{\mathbf{g}})]=i\vec{\mathbf{V}}(\mathbf{f}\times\vec{\mathbf{g}})-iC\int\partial\vec{\mathbf{f}}\circ\vec{\mathbf{g}}\,d\mathbf{x},$$

$$[\mathbf{V}_0(\mathbf{f}),\vec{\mathbf{A}}(\vec{\mathbf{q}})]=i\vec{\mathbf{A}}(\mathbf{f}\times\vec{\mathbf{q}}),$$

$$[\mathbf{A}_0(\mathbf{h}),\vec{\mathbf{V}}(\vec{\mathbf{g}})]=i\vec{\mathbf{A}}(\mathbf{h}\times\vec{\mathbf{g}}),$$

$$[\mathbf{A}_0(\mathbf{h}),\vec{\mathbf{A}}(\vec{\mathbf{q}})]=i\vec{\mathbf{A}}(\mathbf{h}\times\vec{\mathbf{q}})-iC\int\partial\vec{\mathbf{h}}\vec{\mathbf{q}}\,d\mathbf{x},$$

$$[\mathbf{V}_0(\mathbf{f}),\varphi(\mathbf{l})]=i\varphi(\mathbf{f}\times\mathbf{l}),\tag{VI. 8}$$

$$[\mathbf{V}_0(\mathbf{f}),\sigma(b)]=0,$$

$$[\mathbf{A}_0(\mathbf{h}),\varphi(\mathbf{l})]=i\sigma(\mathbf{h}\circ\mathbf{l}),$$

$$[\mathbf{A}_0(\mathbf{h}),\sigma(b)]=-i\varphi(\mathbf{h}\circ b),$$

The formulas are true on each vector, on which they have sense.

Proof: Part (a) of the theorem is obvious. The first formula in (VI. 8) has been proved in Ref. 8 (Appendix A). It is clear that the next three formulas can be proved proved in a quite similar way. So, we only give the proof of the seventh formula. Let us consider (the lhs is well defined if and only if the strong limits on the rhs exist)

$$\mathbf{A}_0(\mathbf{h})\varphi(\mathbf{l})F=\lim_{t\to0}\lim_{p\to0}\frac{\exp[it\mathbf{A}_0(\mathbf{h})]-1}{it}\times\left(\frac{\exp[pi\varphi(\mathbf{l})]-1}{ip}\right)F.\tag{VI. 9}$$

Because we know the action of $\exp[it\mathbf{A}_0(\mathbf{h})]$ and $\exp[ip\varphi(\mathbf{l})]$ from Eqs. (IV. 5) and (IV. 9)–(IV. 12) we can commute the operators on the rhs of Eq. (VI. 9). We obtain then

$$=\lim_{t\to0}\lim_{p\to0}\left(\frac{\exp[ip\varphi(\mathbf{l})]-1}{ip}\right)\times\left(\frac{\exp[it\mathbf{A}_0(\mathbf{h})]-1}{it}\right)F+\lim_{t\to0}\lim_{p\to0}\exp[ip\varphi(\mathbf{l})]\times\frac{1}{it}\cdot\frac{1}{ip}\{(\exp[ip\varphi(\underset{\sim}{\mathbf{h}}(\underset{\sim}{\mathbf{h}}\cdot\mathbf{l}))(\cos|t\mathbf{h}|-1)]\}\times\{\exp[pi\sigma(\underset{\sim}{\mathbf{h}}\mathbf{l}\sin|t\mathbf{h}|)]-1)\}\{\exp[it\mathbf{A}_0(\mathbf{h})]\}F.$$

The existence of the limit $p\to0$ follows from the requirement that F belongs to the domain of φ and σ. It is equal

$$\frac{1}{it}[\varphi(\underset{\sim}{\mathbf{h}}(\underset{\sim}{\mathbf{h}}\cdot\mathbf{l})(\cos|t\mathbf{h}|-1)+\sigma(\underset{\sim}{\mathbf{h}}\cdot\mathbf{l}\sin|t\mathbf{h}|)]$$

Now, taking the limit $t\to0$ we get the result. QED

From the formulas (VI. 8), using the Jacobi identities, we obtain immediately the remaining commutators (II. 5) in the form of a commutator with the operators from the commutative subalgebra generated by $\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma$. So, commuting $\mathbf{V}_0(\mathbf{f}')$ with (VI. 8), we get

$$[\vec{\mathbf{V}}(\vec{\mathbf{g}}),[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]]=[\vec{\mathbf{V}}(\vec{\mathbf{g}}),i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}')],$$

$$[\vec{\mathbf{A}}(\vec{\mathbf{q}}),[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]]=[\vec{\mathbf{A}}(\vec{\mathbf{q}}),i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}')],$$

$$[\varphi(\mathbf{l}),[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]]=[\varphi(\mathbf{l}),i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}')],$$

$$[[\sigma(b),[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]]=[\sigma(b),i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}')].\tag{VI. 10}$$

In other words $[\mathbf{V}_0(\mathbf{f}),\mathbf{V}_0(\mathbf{f}')]-i\mathbf{V}_0(\mathbf{f}\times\mathbf{f}')$ commutes with the maximal commutative subalgebra. The same conclusion is true for $\mathbf{V}_0(\alpha\mathbf{f}+\alpha'\mathbf{f}')-\alpha\mathbf{V}_0(\mathbf{f})-\alpha'\mathbf{V}_0(\mathbf{f}')$. Similar statements can be obtained for the axial current $\mathbf{A}_0(\mathbf{h})$.

We would like now to construct a common, dense, invariant domain D, on which the formulas (VI. 7), (VI. 8), (VI. 10) will be fulfilled. Then we can show that from the expressions of the form (VI. 10) it follows that all the commutation relations (II. 5) are fulfilled on D. To construct such a domain D, we must assume the following.

1. $$\int P(s_1,\ldots,s_\alpha:t_1,\ldots,t_\beta:v_1,\ldots,v_\gamma:w_1,\ldots,w_\delta)\times d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)<\infty,\tag{VI. 11}$$

for all polynomials P. Here, we denote

$$s_i=(\vec{\mathbf{V}},\vec{\mathbf{g}}_i),\quad t_i=(\vec{\mathbf{A}},\vec{\mathbf{q}}_i),\quad v_i=(\varphi,\mathbf{l}_i),\quad w_i=(\sigma,b_i).$$

If we want the cyclic vector $\Omega=1$ to belong to this domain, we must assume the following.

2. The strong limits

$$\lim_{t\to0}\frac{1}{it}\left[\left(\frac{d\mu(t\mathbf{f}^*\vec{\mathbf{V}},t\mathbf{f}^*\vec{\mathbf{A}},t\mathbf{f}^*\varphi,\sigma)}{d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)}\right)^{1/2}-1\right]$$

and

$$\lim_{t\to0}\frac{1}{it}\left[\left(\frac{d\mu(t\mathbf{h}^*\vec{\mathbf{V}},t\mathbf{h}^*\vec{\mathbf{A}},t\mathbf{h}^*\varphi,t\mathbf{h}^*\sigma)}{d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)}\right)^{1/2}-1\right]$$

exist.

We take[4] as the domain D the set of all functions $\beta(s_1,\ldots,s_\alpha:t_1,\ldots,t_\beta:v_1,\ldots,v_\gamma:w_1,\ldots,w_\delta)$, which are infinitely differentiable and polynomially bounded. It can be shown[4] that the functions β form a dense set in $\mathcal{L}^2_\mu$ and belong to the domain of definition of $\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma$. In order to check that the β's are in the domain of $\mathbf{V}_0(\mathbf{f})$ and $\mathbf{A}_0(\mathbf{h})$ we have to show that the limits (VI. 5) and (VI. 6) exist on β. We denote the action of the subgroups $\exp[it\mathbf{V}_0(\mathbf{f})]$ [(IV. 4)] and $\exp[it\mathbf{A}_0(\mathbf{h})]$ [(IV. 5)] on the arguments of β by β_{tf} and $\beta_{t\mathbf{h}}$ corresp. [i.e., $\beta_{t\mathbf{h}}(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)=\beta(t\mathbf{h}^*\vec{\mathbf{V}},t\mathbf{h}^*\vec{\mathbf{A}},t\mathbf{h}^*\varphi,t\mathbf{h}^*\sigma)$]. Then the limits (VI. 5) and (VI. 6) can be divided into three parts, e.g.,

$$\lim_{t\to0}\frac{\exp[it\mathbf{V}_0(\mathbf{f})]-1}{it}\beta=\lim_{t\to0}\frac{\beta_{tf}-\beta}{it}+\lim_{t\to0}\frac{1}{it}\left[\left(\frac{d\mu(t\mathbf{f}^*\vec{\mathbf{V}},t\mathbf{f}^*\vec{\mathbf{A}},t\mathbf{f}^*\varphi,t\mathbf{f}^*\sigma)}{d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)}\right)^{1/2}-1\right]\beta-\lim\frac{1}{it}\left[\left(\frac{d\mu(t\mathbf{f}^*\vec{\mathbf{V}},t\mathbf{f}^*\vec{\mathbf{A}},t\mathbf{f}^*\varphi,t\mathbf{f}^*\sigma)}{d\mu(\vec{\mathbf{V}},\vec{\mathbf{A}},\varphi,\sigma)}\right)^{1/2}-1\right](\beta_{tf}-\beta).\tag{VI. 13}$$

The limit of the first term exists because the functions β are differentiable. The limit of the second term exists by the assumption 2. The last limit equals zero. The explicit computation of (VI. 13) gives the result

$$\mathbf{V}_0(\mathbf{f})\beta(s_1,\ldots,s_\alpha:t_1,\ldots,t_\beta:v_1,\ldots,v_\gamma:w_1,\ldots,w_\delta)$$

$$= \sum_{j=1}^{\alpha} \frac{1}{i} \frac{\partial\beta}{\partial s_j} (\vec{\mathbf{V}}, \mathbf{f}\times\vec{\mathbf{g}}_j) + \sum_{j=1}^{\beta} \frac{1}{i} \frac{\partial\beta}{\partial t_j} (\vec{\mathbf{A}}, \mathbf{f}\times\vec{\mathbf{q}}_j) + \sum_{j=1}^{\gamma} \frac{1}{i} \frac{\partial\beta}{\partial v_j} (\varphi, \mathbf{f}\times\mathbf{l}_j) + C\frac{1}{i} \sum_{j=1}^{\alpha} \int \vec{\partial}\mathbf{f}\vec{\mathbf{g}}_j \, d\mathbf{x} \frac{\partial\beta}{\partial s_j} + \frac{1}{i} M^{\mathbf{f}}(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma)\beta, \qquad \text{(VI. 14)}$$

where

$$M^{\mathbf{f}}(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma) = \lim_{t\to 0} \frac{1}{t} \left[\left(\frac{d\mu(t\mathbf{f}^*\vec{\mathbf{V}}, t\mathbf{f}^*\vec{\mathbf{A}}, t\mathbf{f}^*\varphi, \sigma)}{d\mu(\vec{\mathbf{V}}), \vec{\mathbf{A}}, \varphi, \sigma)} \right)^{1/2} - 1 \right], \qquad \text{(IV. 15)}$$

and similarly

$$\mathbf{A}_0(\mathbf{h})\beta(s_1, \ldots, s_\alpha : t_1, \ldots, t_\beta : v_1, \ldots, v_\gamma : w_1, \ldots, w_\delta) = \sum_{j=1}^{\alpha} \frac{1}{i} \frac{\partial\beta}{\partial s_j} (\vec{\mathbf{A}}, \mathbf{h}\times\vec{\mathbf{g}}_j) + \sum_{j=1}^{\beta} \frac{1}{i} \frac{\partial\beta}{\partial t_j} (\vec{\mathbf{V}}, \mathbf{h}\times\vec{\mathbf{q}}_j) - \sum_{j=1}^{\gamma} \frac{1}{i} \frac{\partial\beta}{\partial v_j} (\sigma, \mathbf{h}\cdot\mathbf{l}_j) + \sum_{j=1}^{\delta} \frac{1}{i} \frac{\partial\beta}{\partial w_j} (\varphi, b_j \mathbf{h}) \times C\frac{1}{i} \sum_{j=1}^{\beta} \int \vec{\partial}\mathbf{h}\, \mathbf{q}_j \, d\mathbf{x} \frac{\partial\beta}{\partial t_j} + \frac{1}{i} M^{\mathbf{h}}(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma)\beta, \qquad \text{(VI. 16)}$$

where

$$M^{\mathbf{h}}(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma) = \lim_{t\to 0} \frac{1}{t} \left[\left(\frac{d\mu(t\mathbf{h}^*\vec{\mathbf{V}}, t\mathbf{h}^*\vec{\mathbf{A}}, t\mathbf{h}^*\varphi, t\mathbf{h}^*\sigma)}{d\mu(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma)} \right)^{1/2} - 1 \right]. \qquad \text{(IV. 17)}$$

Now, we come back to Eqs. (VI. 10) and we can state

Theorem III: The commutation relations (II. 5) are fulfilled on D.

Proof: Some of the commutation relations (II. 5) were proven already in Theorem II. It remains to show that from the relations of the form (VI. 10) there follow the corresponding relations (II. 5). The proof goes quite similarly as in Ref. 4. Let us take for example the operator $\mathbf{R} = [\mathbf{A}_0(\mathbf{h}), \mathbf{A}_0(\mathbf{h}')] - i\mathbf{V}_0(\mathbf{h}\times\mathbf{h}')$. Because of Eqs. (VI. 10) and the assumptions of cyclicity it is sufficient to prove that $\mathbf{R}\Omega = 0$. Using Eqs. (VI. 14)–(VI. 17), we can check that $\mathbf{R}\Omega$ is a real function. But $\mathbf{R}^+ = -\mathbf{R}$, so $\mathbf{R}\Omega$ should be purely imaginary, which implies $\mathbf{R}\Omega = 0$. In the similar way we prove the remaining relations. QED

The generators of space rotations and space translations are also defined on the functions β and they are given by the formulas

$$P_i\beta(s_1, \ldots, s_\alpha : t_1, \ldots, t_\beta : v_1, \ldots, v_\gamma : w_1, \ldots, w_\delta) = -i\sum \frac{\partial\beta}{\partial s_j} (\vec{\mathbf{V}}, \partial_i\vec{\mathbf{g}}_j) - i\sum_{j=1}^{\beta} \frac{\partial\beta}{\partial t_j} (\vec{\mathbf{A}}, \partial_i\vec{\mathbf{q}}_j) - i\sum_{j=1}^{\gamma} \frac{\partial\beta}{\partial v_j} (\varphi, \partial_i\mathbf{l}_j) - i\sum_{j=1}^{\delta} \frac{\partial\beta}{\partial w_j} (\sigma, \partial_i b_j) \qquad \text{(VI. 18)}$$

$$M_{ik}\beta(s_1, \ldots, s_\alpha : t_1, \ldots, t_\beta : v_1, \ldots, v_\gamma : w_1, \ldots, w_\delta) = i\sum_{L=1}^{\alpha} (\vec{\mathbf{V}}, x_k\partial_i\vec{\mathbf{g}}_L(\mathbf{x}) - x_i\partial_k\vec{\mathbf{g}}_L(\mathbf{x})) \frac{\partial\beta}{\partial s_L} - i\sum_{L=1}^{\alpha}\sum_{r=1}^{3} [(\vec{\mathbf{V}}_i, \delta_{kr}\mathbf{g}_{Lr}) - (\mathbf{V}_k, \delta_{ir}\mathbf{g}_{Lr})] \frac{\partial\beta}{\partial s_L} + i\sum_{L=1}^{\beta} (\vec{\mathbf{A}}, x_k\partial_i\vec{\mathbf{q}}_L(\mathbf{x}) - x_i\partial_k\vec{\mathbf{q}}_L(\mathbf{x})) \frac{\partial\beta}{\partial t_L} - i\sum_{L=1}^{\beta}\sum_{r=1}^{3} [(\mathbf{A}_i, \delta_{kr}\mathbf{q}_{Lr}) - (\mathbf{A}_k, \delta_{ir}\mathbf{q}_{Lr})] \frac{\partial\beta}{\partial t_L} + i\sum_{L=1}^{\gamma} (\varphi, x_k\partial_i\mathbf{l}_L(\mathbf{x}) - x_i\partial_k\mathbf{l}_L(\mathbf{x})) \frac{\partial\beta}{\partial v_L} + i\sum_{L=1}^{\delta} (\sigma, x_k\partial_i b_L(\mathbf{x}) - x_i\partial_k b_L(\mathbf{x})) \frac{\partial\beta}{\partial w_L}. \qquad \text{(VI. 19)}$$

They are self-adjoint if and only if the measure μ is invariant under space translations $f(\mathbf{x}) \to f(\mathbf{x}+\mathbf{a})$ and space rotations (V. 6).

VII. ULTRALOCAL MODEL OF THE SU(2) ⊗ SU (2) CURRENT ALGEBRA

Up to here we were dealing only with general features of the representations of the SU(2)⊗SU(2) current algebra (II. 5). For more concrete considerations we should choose a measure fulfilling all our requirements. Then we have to construct Hamiltonian in order to introduce time into the theory. We expect that similarly as in field theory[26] this Hamiltonain is uniquely determined by the measure μ. Also the conservation or non-conservation of the currents V^a_μ, A^a_μ depends on the choice of Hamiltonian, and so on the choice of the measure μ. Because of the physical sense of the current algebra (II. 5) describing the SU(2)⊗SU(2) symmetry, we require the conservation of the vector current V^a_μ and the partial conservation of the axial current A^a_μ (PCAC). The axial current should be conserved in the limit when symmetry breaking interaction disappears. It is clear that the programme we have outlined is difficult to perform. The difficulties, which arise here, are the same as in constructive field theory. In Sec. V. we have shown that the Gaussian measure fulfills the requirements of describing a representation of the SU(2)⊗SU(2) current algebra (plus the fields $\varphi^a\sigma$). So, we will start here, similarly as in the constructive field theory[24,27] with the Gaussian measure.[28] In the space $\mathcal{L}^2_{\mu_B}$ the representation of the canonical pair of the field variables can be constructed

$$\exp[i\varphi(f)]F(\varphi) = \exp[i(\varphi, f)]F(\varphi),$$

$$\exp[i\pi(f)]F(\varphi) = \left(\frac{d\mu_B(\varphi+f)}{d\mu_B(\varphi)} \right)^{1/2} F(\varphi+f). \qquad \text{(VII. 1)}$$

Using (V. 4), we obtain in our case

$$V^a_k(g^{a\prime}_k)\beta(s,t,v,w) = (V^a_k, g^{a\prime}_k)\beta(s,t,v,w),$$

$$\Pi^a_{Vk}(g^{a\prime}_k)\beta(s,t,v,w) = \frac{1}{i} \sum_{L=1}^{\alpha} \frac{\partial\beta}{\partial s_L} (g^a_{Lk}, g^{a\prime}_k) - \frac{1}{i} (V^a_k, K_x^{1/2} g^{a\prime}_k)\beta(s,t,v,w) \qquad \text{(VII. 2)}$$

(no summing over a, k),

$$A^a_k(q^{a\prime}_k)\beta(s,t,v,w) = (A^a_k, q^{a\prime}_k)\beta(s,t,v,w),$$

$$\Pi^a_{Ak}(q^{a\prime}_k)\beta(s,t,v,w) = \frac{1}{i} \sum_{L=1}^{\beta} \frac{\partial\beta}{\partial t_L} (q^a_{Lk}, q^{a\prime}_k) - \frac{1}{i} (A^a_k, K_x^{1/2} q^{a\prime}_k)\beta(s,t,v,w), \qquad \text{(VII. 3)}$$

$$\varphi^a(l^{a\prime})\beta(s,t,v,w) = (\varphi^a, l^{a\prime})\beta(s,t,v,w),$$

$$\Pi^a(l^{a\prime})\beta(s,t,v,w) = \frac{1}{i}\sum_{L=1}^{\gamma}\frac{\partial\beta}{\partial v_L}(l_L^a, l^{a\prime}) - \frac{1}{i}(\varphi^a, \mu_x^{1/2} l^{a\prime})\beta(s,t,v,w), \tag{VII.4}$$

and

$$\sigma(b')\beta(s,t,v,w) = (\sigma, b')\beta(s,t,v,w),$$

$$\Pi(b')\beta(s,t,v,w) = \frac{1}{i}\sum_{L=1}^{6}\frac{\partial\beta}{\partial w_L}(b_L, b') - \frac{1}{i}(\sigma, \mu_x^{1/2} b')\beta(s,t,v,w). \tag{VII.5}$$

For relativistic vector fields

$$B(\vec{\mathbf{g}}_1, \vec{\mathbf{g}}_2) = \sum_{a,i,j=1}^{3} \int g_i^a(\mathbf{k})_1 \left(\frac{\delta_{ij} - k_i k_j/m^2}{(|\mathbf{k}|^2 + m^2)^{1/2}}\right) g_j^a(\mathbf{k})_2 \, d\mathbf{k}. \tag{VII.6}$$

This form can be diagonalized and we can find K_{V_x}, Eq.(V.4) (or $K_V(\mathbf{k})$ in the momentum space). However, if $K_V(\mathbf{k})$ depends on $\mathbf{k}$, K_{V_x} is a nonlocal operator and the currents (VI.14) and (VI.16) are nonlocal functions of the fields. In such a case we are not able to construct the Hamiltonian, which will ensure the conservation of currents. So, we omit the "gradients" k_i in (VII.6) getting an ultralocal theory[9,10] with the bilinear form

$$B(\vec{\mathbf{g}}, \vec{\mathbf{q}}, \mathbf{l}, b; \vec{\mathbf{g}}', \vec{\mathbf{q}}', \mathbf{l}', b') = \sqrt{C}\int \vec{\mathbf{g}}(\mathbf{x})\cdot\vec{\mathbf{g}}'(\mathbf{x})\,d\mathbf{x} + \sqrt{C}\int \vec{\mathbf{q}}(\mathbf{x})\vec{\mathbf{q}}(\mathbf{x})\,d\mathbf{x} + \frac{1}{\mu_0}\int \mathbf{l}(\mathbf{x})\cdot\mathbf{l}'(\mathbf{x})\,d\mathbf{x} + \frac{1}{\mu_0}\int b(\mathbf{x})b'(\mathbf{x})\,d\mathbf{x}. \tag{VII.7}$$

In such a case we find quite similarly as in Ref. 8 that the currents $\vec{\mathbf{V}}, \vec{\mathbf{A}}$ are expressed in terms of fields in the following way:

$$\mathbf{V}_0(\mathbf{f}) = (\vec{\Pi}_V \times \vec{\mathbf{V}})(\mathbf{f}) + (\vec{\Pi}_A \times \vec{\mathbf{A}})(\mathbf{f}) + (\Pi \times \boldsymbol{\varphi})(\mathbf{f}) + C(\partial\vec{\Pi}_V)(\mathbf{f}),$$

where

$$\vec{\Pi} \times \vec{\mathbf{V}} = \sum_{k=1}^{3} \Pi_k \times \mathbf{V}_k, \quad \partial\cdot\vec{\Pi} = \partial_k \cdot \Pi_{Vk}, \tag{VII.8}$$

$$\mathbf{A}_0(\mathbf{h}) = (\vec{\Pi}_V \times \vec{\mathbf{A}})(\mathbf{h}) + (\vec{\Pi}_A \times \vec{\mathbf{V}})(\mathbf{h}) - (\sigma\boldsymbol{\Pi})(\mathbf{h}) + (\Pi\boldsymbol{\varphi})(\mathbf{h}) + C(\partial\cdot\vec{\Pi}_A)(\mathbf{h}). \tag{VII.9}$$

It is known that the current algebra with c-number Schwinger term [Eqs. (II.1) and (II.2)] is closely connected with the Yang–Mills theory.[1,29,30] It is a basis of the vector mesons dominance in Ref. 30. In this case the currents are expressed by the canonical Yang–Mills field variables. In another approach[29,3] a noncanonical current theory is obtained as a limit of the Yang–Mills theory. In Ref. 8 we have shown that (when the fields φ^a, σ are absent) we get our currents expressed by fields in the same way as in the Yang–Mills theory. We have now a similar situation. Let us take the Yang–Mills fields and φ^a, σ fields with the Lagrangian[16] (which combines the features of the σ-model[11] and Yang–Mills theory)

$$\begin{aligned} L = &\sum_{\mu,\nu=0}^{3} -\tfrac{1}{4}(\partial_\mu \mathbf{v}_\nu - \partial_\nu \mathbf{v}_\mu - g_0\mathbf{v}_\mu\times\mathbf{v}_\nu - g_0\mathbf{a}_\mu\times\mathbf{a}_\nu)^2 \\ &\times \sum_{\mu,\nu=0}^{3} -\tfrac{1}{4}(\partial_\mu\mathbf{a}_\nu - \partial_\nu\mathbf{a}_\mu - g_0\mathbf{v}_\mu\times\mathbf{a}_\nu + g_0\mathbf{v}_\nu\times\mathbf{a}_\mu)^2 \\ &+ \tfrac{1}{2}m^2\sum_{\mu=0}^{2}(\mathbf{v}_\mu{}^2 + \mathbf{a}_\mu{}^2) + \tfrac{1}{2}\sum_{\mu=0}^{3}(\partial_\mu\sigma - g_0\boldsymbol{\varphi}\cdot\mathbf{a}_\mu)^2 \\ &+ \tfrac{1}{2}\sum_{\mu=0}^{3}(\partial_\mu\boldsymbol{\varphi} - g_0\mathbf{v}_\mu\times\boldsymbol{\varphi} + g_0 f_\pi\sigma\cdot\mathbf{a}_\mu)^2 \\ &+ \tfrac{1}{2}\mu_0^2(\boldsymbol{\varphi}^2 + \sigma^2) + f_\pi\mu_0{}^2\sigma. \end{aligned} \tag{VII.10}$$

Writing the Lagrange equations and omitting terms proportional to $g_0{}^2$, we get from these equations the formulas (VII.8) and (VII.9) for $\mathbf{V}_0(\mathbf{f})$ and $\mathbf{A}_0(\mathbf{h})$. These considerations suggest the form of the ultralocal Hamiltonian. We obtain it omitting some terms in the Yang–Mills Hamiltonian. Namely, we omit self-interaction of the Yang–Mills fields and terms with spatial derivatives. We get

$$\begin{aligned} H(\boldsymbol{\alpha}) = &\tfrac{1}{2}\int\{:\vec{\Pi}_V\cdot\vec{\Pi}_V:(\mathbf{x}) + :\vec{\Pi}_A\cdot\vec{\Pi}_A:(\mathbf{x}) \\ &+ (1/C):\vec{\mathbf{V}}\cdot\vec{\mathbf{V}}:(\mathbf{x}) + (1/C):\vec{\mathbf{A}}\cdot\vec{\mathbf{A}}:(\mathbf{x})\}d\mathbf{x} \\ &+ \tfrac{1}{2}\int\{:\boldsymbol{\Pi}\cdot\boldsymbol{\Pi}:(\mathbf{x}) + :\Pi\cdot\Pi:(\mathbf{x}) \\ &+ \mu_0{}^2:\boldsymbol{\varphi}\cdot\boldsymbol{\varphi}:(\mathbf{x}) + \mu_0{}^2:\sigma^2:(\mathbf{x})\}d\mathbf{x} \\ &- f_\pi\mu_0{}^2\int\boldsymbol{\alpha}(\mathbf{x})\sigma(\mathbf{x})\,d\mathbf{x}. \end{aligned} \tag{VII.11}$$

The operator $H(\alpha)$ written above needs some explanations. The Gaussian measure describes the Fock representation of the canonical commutation relations.[31] It is known[32] that the operator $H(\alpha)$ can be defined in the Fock space and is self-adjoint if and only if $\int\alpha^2(\mathbf{x})\,d\mathbf{x} < \infty$. For this purpose we have introduced the cutoff, which must be removed in order to obtain the partial conservation of the axial current (PCAC). Just the last term breaks the chiral SU(2)⊗SU(2) symmetry and gives PCAC. Using the canonical commutation relations and taking the formal limit $\alpha \to 1$, we get

$$\begin{aligned} &[H, V_0^a(\mathbf{x})] = -i\partial_k V_k^a(\mathbf{x}), \\ &[H, A_0^a(\mathbf{x})] = -i\partial_k A_k^a(\mathbf{x}) - if_\pi\mu_0^2\varphi^a(\mathbf{x}). \end{aligned} \tag{VII.12}$$

So, after introducing

$$V_\mu^a(\mathbf{x},t) = [\exp(iHt)]\cdot V_\mu^a(\mathbf{x})\cdot[\exp(-iHt)],$$

$$A_\mu^a(\mathbf{x},t) = [\exp(iHt)]\cdot A_\mu^a(\mathbf{x})\cdot[\exp(-iHt)],$$

$$\varphi^a(\mathbf{x},t) = [\exp(iHt)]\varphi^a(\mathbf{x})[\exp(-iHt)],$$

$$\sigma(\mathbf{x},t) = [\exp(iHt)]\cdot\sigma(\mathbf{x})\cdot[\exp(-iHt)],$$

we get

$$\partial^\mu V_\mu^a(\mathbf{x},t) = 0, \quad \partial^\mu A_\mu^a(\mathbf{x},t) = f_\pi\mu_0^2\varphi^a(\mathbf{x},t). \tag{VII.13}$$

In order to represent H $[H = \lim_{\alpha\to 1} H(\alpha)]$ and the relations (VII.12), (VII.13) in Hilbert space we must have non-Gaussian measures. We apply here the standard methods of constructive field theory.[24,12,27] First, let us notice that $H(\alpha)\Omega = H(\alpha)1 \neq 0$ and we have to construct the physical vacuum state Ω_α which equals[17] $[\int \alpha^2(\mathbf{x})d\mathbf{x} < \infty]$

$$\Omega_\alpha = \exp[i\Pi(f_\pi \alpha)\Omega] = \exp(-f_\pi \mu_0(\sigma, \alpha) - (f_\pi/2\mu_0)(\alpha, \alpha)), \quad \text{(VII.14)}$$

then

$$H_\alpha \Omega_\alpha = 0. \quad \text{(VII.15)}$$

The functional

$$\begin{aligned} L(\vec{\mathbf{g}}, \vec{\mathbf{q}}, \mathbf{l}, b) &= (\Omega_\alpha, \{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}})]\} \cdot \{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}})]\} \\ &\quad \times \{\exp[i\varphi(\mathbf{l})]\}\{\exp[i\sigma(b)]\}\Omega_\alpha) \\ &= \{\exp[-(\sqrt{C}/4)(\vec{\mathbf{g}}, \vec{\mathbf{g}})]\} \cdot \{\exp[-(\sqrt{C}/4)(\vec{\mathbf{q}}, \vec{\mathbf{q}})]\} \\ &\quad \cdot \{\exp[-(1/4\mu_0)(\mathbf{l}, \mathbf{l})]\} \\ &\quad \times \{\exp[-(1/4\mu_0)(b, b)]\} \cdot \{\exp[-if_\pi(b, \alpha)]\} \end{aligned} \quad \text{(VII.16)}$$

defines a new measure, which equals

$$d\mu_\alpha(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma) = \frac{\exp[-2f_\pi \mu_0(\sigma, \alpha)]}{\exp[(f_\pi/\mu_0)(\alpha, \alpha)]} d\mu_B(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma). \quad \text{(VII.17)}$$

Now, the limit[12] $\alpha \to 1$ of the functional L_α [(VII.16)] exists

$$\begin{aligned} L_1(\vec{\mathbf{g}}, \vec{\mathbf{q}}, \mathbf{l}, b) &= \{\exp[-(\sqrt{C}/4)(\vec{\mathbf{g}}, \vec{\mathbf{g}})]\}\{\exp[-(\sqrt{C}/4)(\vec{\mathbf{q}}, \vec{\mathbf{q}})]\} \\ &\quad \times \{\exp[-(1/4\mu_0)(\mathbf{l}, \mathbf{l})]\} \circ \{\exp[-(1/4\mu_0)(b, b)]\} \\ &\quad \times \{\exp[-if_\pi \textstyle\int b(\mathbf{x})\, d\mathbf{x}]\} \end{aligned} \quad \text{(VII.18)}$$

and defines a new measure, which can be considered as the limit (in the weak sense)

$$d\mu_1(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma) = = \lim_{\alpha \to 1} \frac{\exp[-2f_\pi \mu_0(\sigma, \alpha)]}{\exp[(f_\pi/\mu_0)(\alpha, \alpha)]} d\mu_B(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma). \quad \text{(VII.19)}$$

It can be shown that the measure $d\mu_1$ is also quasi-invariant under the action of $\mathbf{f}^*\mathbf{h}^*$ [(IV.6)–(IV.12)] and its Radon–Nikodym derivatives are (see the proof in $P(\varphi)_2$[12])

$$\frac{d\mu_1(\mathbf{f}^*\vec{\mathbf{V}}, \mathbf{f}^*\vec{\mathbf{A}}, \mathbf{f}^*\varphi, \sigma)}{d\mu_1(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma)} = \{\exp[-2\sqrt{C}(\vec{\mathbf{V}}, \partial\vec{\mathbf{f}})]\}\{\exp[-C^{3/2}(\partial\vec{\mathbf{f}}, \partial\vec{\mathbf{f}})]\}, \quad \text{(VII.20)}$$

$$\begin{aligned} \frac{d\mu_1(\mathbf{h}^*\mathbf{V}, \mathbf{h}^*\mathbf{A}, \mathbf{h}^*\varphi, \mathbf{h}^*\sigma)}{d\mu_1(\vec{\mathbf{V}}, \vec{\mathbf{A}}, \varphi, \sigma)} &= \{\exp[-2\sqrt{C}(\vec{\mathbf{A}}, \partial\vec{\mathbf{h}})]\} \cdot \{\exp[-C^{3/2}(\partial\vec{\mathbf{h}}, \partial\vec{\mathbf{h}})]\} \\ &\quad \times \{\exp[-2f_\pi \mu_0(\sigma, \cos|\mathbf{h}| - 1)]\} \\ &\quad \cdot \{\exp[-2f_\pi \mu_0(\varphi, \underset{\sim}{\mathbf{h}} \sin|\mathbf{h}|)]\}. \end{aligned} \quad \text{(VII.21)}$$

We can take as the Hamiltonian the operator $H = = \int a_{in}^+(\mathbf{k}) a_{in}(\mathbf{k})\, d\mathbf{k}$, where the asymptotic fields are defined in Ref. 32. It can be considered as a limit[24] $\alpha \to 1$ of $H(\alpha) - (\Omega, H(\alpha)\Omega)$. This shows that in $L^2_{\mu_1}$ the operator $H = \lim_{\alpha \to 1} H(\alpha)$ is well defined and self-adjoint (and $H\Omega_1 = 0$). So, the relations (VII.12) and (VII.13) are fulfilled (H is defined on D).

On the basis of this simple model we are going to investigate the chiral symmetry breaking. As the first consequence we get a nonzero vacuum expectation value of the field σ

$$(\Omega_1, \sigma(b)\Omega_1) = \frac{1}{i}\left\{\frac{d}{ds}(\Omega_1, \exp[is\sigma(b)]\Omega_1)\right\}_{s=0} = -f_\pi \int b(\mathbf{x})\, d\mathbf{x} \quad \text{(VII.22)}$$

So,

$$(\Omega_1, b(\mathbf{x}, t)\Omega_1) = -f_\pi \quad \text{(VII.23)}$$

In the language of Feynman diagrams this effect gives the tadpole graphs (Ref. 1). We would have the "spontaneous" breaking of chiral symmetry if we had introduced the interaction $\lambda(\varphi^2 + \sigma^2)^2$ (Ref. 1) into the Hamiltonian (VII.11). Let us notice that because of the change of the Radon–Nikodym derivative (VII.21) the form of the current $\mathbf{A}_0(\mathbf{h})$ [(VI.16)] is changed. However, the expressions of the currents in terms of fields (VII.8) and (VII.9) remain unchanged, because the changes of A_0^a and σ compensate one another. Let us consider the charges

$$Q^a = \int V_0^a(\mathbf{x})\, d\mathbf{x} \quad \text{and} \quad Q_5^a = \int A_0^a(\mathbf{x})\, d\mathbf{x}. \quad \text{(VII.24)}$$

Their existence depends in a crucial way on the Radon–Nikodym derivatives of the measure in Eqs. (IV.4) and (IV.5). If there exist the pointwise limits $h^a \to 1$, $f^a \to 1$ of the Radon–Nikodym derivatives, then the charges (VII.24) exist. This condition is fulfilled for the Gaussian measure and even for every measure, which is invariant under $SU(2) \otimes SU(2)$ rotations (because $\partial\vec{\mathbf{h}}, \partial\vec{\mathbf{f}} \to 0$). If the measure μ is not invariant under $SU(2) \otimes SU(2)$ rotations, then the cyclic vector (vacuum) $\Omega_1 = 1$ is not invariant under chiral $SU(2) \otimes SU(2)$ transformations. We cannot change the vacuum $\Omega_1 = 1$ because it is fixed by the condition $H\Omega_1 = 0$. In our case, as can be seen from Eqs. (IV.4) and (VII.20), the limit $f^a \to 1$ of the Radon–Nikodym derivative (VII.20) exists and $\lim_{f^a \to 1} Q^a(f^a) = Q^a$ defines the self-adjoint charge, which is the generator of SU(2) symmetry. The vacuum is invariant under SU(2) transformations

$$\exp(iq^a Q^a)\Omega_1 = \Omega_1. \quad \text{(VII.25)}$$

When cutoff is present, the pointwise limit $h^a \to 1$ of the Radon–Nikodym derivative (VII.21) also exists and defines the self-adjoint charge Q_5^a.

We have quite different situation after removing of the cutoff. Let us compute the matrix elements of $\exp[i\mathbf{A}_0(\mathbf{h})]$:

$$\begin{aligned} (F_1, \exp[i\vec{\mathbf{A}}_0(s\mathbf{h})]F_2) &= (\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_1)]\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_1)]\exp[i\varphi(\mathbf{l}_1)] \\ &\quad \times \exp[i\sigma(b_1)]\Omega_1, \exp[i\mathbf{A}_0(s\mathbf{h})]\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_2)] \\ &\quad \times \exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_2)]\exp[i\varphi(\mathbf{l}_2)]\exp[i\sigma(b_2)]\Omega_1) \end{aligned}$$

$$= (\{\exp[i\vec{\mathbf{A}}(\vec{\mathbf{q}}_1 - s\mathbf{h}^*\vec{\mathbf{q}}_2)]\}\{\exp[i\vec{\mathbf{V}}(\vec{\mathbf{g}}_1 - s\mathbf{h}^*\vec{\mathbf{g}}_2)]\}$$

$$\times\{\exp[i\varphi(\mathbf{l}_1 - s\mathbf{h}^*\mathbf{l}_2)]\}\{\exp[i\sigma(b_1 - s\mathbf{h}^* b_2)]\}\Omega_1,$$

$$\{\exp[i\mathbf{A}_0(s\mathbf{h})]\}\Omega_1)$$

$$= (F^s_{1,2}, \exp[i\mathbf{A}_0(s\mathbf{h})]\Omega_1). \qquad \text{(VII. 26)}$$

The matrix elements (VII. 26) can be computed, and we obtain

$$(F_1, \exp[i\mathbf{A}_0(s\mathbf{h})]F_2)$$

$$= \{\exp[-(\sqrt{C}/4)(\vec{\mathbf{g}}^s_{1,2}, \vec{\mathbf{g}}^s_{1,2})]\}\{\exp[-(\sqrt{C}/4)(\vec{\mathbf{q}}^s_{1,2}, \vec{\mathbf{q}}^s_{1,2})]\}$$

$$\times\{\exp[-(1/4\mu_0)(\mathbf{l}^s_{1,2}, \mathbf{l}^s_{1,2})]\}\{\exp[-(1/4\mu_0)(b^s_{1,2}, b^s_{1,2})]\}$$

$$\times\{\exp(-if_\pi \int b^s_{1,2}(\mathbf{x})(\tfrac{1}{2} + \tfrac{1}{2}\cos|s\mathbf{h}|)\,d\mathbf{x})\}$$

$$\times\{\exp(-if_\pi \int \mathbf{l}^s_{1,2}(\mathbf{x})\,\underline{\mathbf{h}}(\mathbf{x})\tfrac{1}{2}\sin|s\mathbf{h}|\,d\mathbf{x})\}$$

$$\times\{\exp(-(f_\pi|/2\mu_0)\int(1-\cos|s\mathbf{h}|)\,d\mathbf{x})\}, \qquad \text{(VII. 27)}$$

$$\vec{\mathbf{g}}_{1,2} = \vec{\mathbf{g}}_1 - s\mathbf{h}^*\vec{\mathbf{g}}_2.$$

Let us formulate our results in the form of a theorem:

Theorem IV: (a) The strong limit $\exp(isQ^a) = \lim_{f^a\to 1} \exp[is V^a_0(f^a)]$ exists and defines the self-adjoint charge Q^a, which is the generator of SU(2) symmetry leaving the vacuum invariant.

(b) The strong limit of $\exp[is\mathbf{A}_0(\mathbf{h})]$ does not exist. The weak limit is equal to zero.

(c) The weak limit of $\mathbf{A}_0(\mathbf{h})$ does not exist, but the bilinear form

$$\lim_{h\to 1}(F, A^a_0(h)F') = Q^a_5(F, F'), \quad F, F' \in D,$$

defines a symmetric operator $Q^a_5(F,F') = (F, Q^a_5 F')$.

(d) $Q^a_5\Omega_1 \neq 0$.

Proof: (a) This statement follows from our previous considerations.

(b) From Eq. (VII. 27) we can see that when $\mathbf{h}\to 1$ $(F, \exp[i\mathbf{A}_0(\mathbf{h})]F') \to 0$ for F, $F' \in D$ and the sequence $\exp[is\mathbf{A}_0(\mathbf{h})]F'$ is bounded. So, $\exp[is\mathbf{A}_0(\mathbf{h})]$ tends weakly to zero.[33] Of course, the strong limit of $\exp[is\mathbf{A}_0(\mathbf{h})]$ cannot be zero; therefore, it does not exist.

(c) Let us take

$$F = \sum a_i F^i_1, \quad F' = \sum c^i F^i_2,$$

where F^i_1, F^i_2 have the form of exponents similar as in the expression (VII. 26). We shall use a technical assumption $\int l^a_i(\mathbf{x})\,d\mathbf{x} = 1$, it does not reduce the set of vectors F because $\varphi(\mathbf{l})$ is linear in $\mathbf{l}$. Equation (VII. 27) can be written in the form

$$(F, \exp[is\mathbf{A}_0(\mathbf{h})]F') = \sum_{i,j} a^i c^j (F^i_1, \exp[is\mathbf{A}_0(\mathbf{h})]F^j_2)$$

$$= \sum_{i,j} a^i c^j \{\exp[(-if_\pi/2)\int b^s_{i,j}\,(1-\cos|s\mathbf{h}|)\,d\mathbf{x}]\}$$

$$\times\{\exp[(-f_\pi/2\mu_0)\int(1-\cos|s\mathbf{h}|)\,d\mathbf{x}]\}$$

$$\times\{\exp(-if_\pi\tfrac{1}{2}\int \mathbf{l}^s_{i,j}\underline{\mathbf{h}}\sin|s\mathbf{h}|\,d\mathbf{x})\}$$

$$\times(F, \exp[is\mathbf{A}_0(\mathbf{h})]_{\text{free}}F'), \qquad \text{(VII. 28)}$$

where $\exp[is\mathbf{A}_0(\mathbf{h})]_{\text{free}}$ means that the action of $\exp[is\mathbf{A}_0(\mathbf{h})]$ is the same as in the case without interaction, i.e., we use the Radon–Nikodym derivative of the Gaussian measure. We denote by $\exp(ir_j)$ the phase factors depending on j. Further we introduce the vector $F^s_{\underline{\mathbf{h}}} = \sum_j c_j \exp(ir_j)F^j_2$, differentiate both sides of (VII. 28) over s, and put $s=0$. We then get

$$(F, \mathbf{A}_0(\mathbf{h})F') = -\sum_{i,j}\tfrac{1}{2}f_\pi(\int \mathbf{l}_j\mathbf{h}\,d\mathbf{x})a^i F^i_1, F'$$

$$+\frac{d}{ds}\Big(\big[(F, \exp[is\mathbf{A}_0(\mathbf{h})]_{\text{free}}F^s_{\mathbf{h}})\big]\Big)_{s=0} \qquad \text{(VII. 29)}$$

Now, the limits $\mathbf{h}\to 1$ of both terms of Eq. (VII. 29) exist [the existence of the limit of the second term can be shown similarly as in (a)]. Using our assumption $\int l^a_j(\mathbf{x})\,d\mathbf{x} = 1$ and denoting

$$\mathbf{Q}_5(F,F') = \lim_{h\to 1}(F, \mathbf{A}_0(\mathrm{h})F'),$$

we get

$$|\mathbf{Q}_5(F,F')| \le K_{F'}\|F\|. \qquad \text{(VII. 30)}$$

The condition (VII. 30) is necessary and sufficient[13,33] in order to define the operator Q^α_5

$$(F; Q^a_5, F') = Q^a_5(F, F')$$

(d) follows immediately from Eq. (VII. 29) QED

The part (d) of the theorem gives an illustration of the Coleman theorem.[14] We are not able to prove that Q^a_5 is self-adjoint. This is a difficult problem.[13] So, we do not know whether Q^a_5 is the generator of the chiral symmetry or not. Comparing our results with Ref. 13, we can see that the situation (concerning the charges), which we have is much better than in general relativistic *QFT*. This is due to the ultralocality and simple form of the interaction.

VIII. FINAL REMARKS

We were considering the problem of representations of the SU(2)⊗SU(2) current algebra with c-number Schwinger term and with additional fields needed for PCAC. All the representations are described by a quasi-invariant measure. So, a difficult problem appears to be finding such measures and answering the question, which current theories admit the c-number Schwinger term. Each such a representation should give important information about properties of the theory of currents. Finally, we would like to emphasize that the ultralocal theory has arisen in our case because of the use of the Gaussian measure. There should exist measures giving nontrivial relativistic theories e.g., the measure describing the Yang–Mills fields.[29]

[1]V. De Alfaro, S. Fubini, G. Furlan, and C. Rossetti, *Currents in Hadron Physics* (North-Holland, Amsterdam, 1973).
[2]R.F. Dashen and D.H. Sharp, Phys. Rev. **165**, 1857 (1968); and D.H. Sharp, *ibid.* 165, 1867 (1968).

[3]H. Sugawara, Phys. Rev. **170**, 1659 (1968).
[4]G. A. Goldin, J. Math. Phys. **12**, 462 (1971).
[5]J. Grodnik and D.H. Sharp, Phys. Rev. D **1**, 1531 (1970).
[6]G. A. Goldin, J. Grodnik, R.T. Powers, and D.H. Sharp, J. Math. Phys. **15**, 88 (1974).
[7]G. A. Goldin and D.H. Sharp, in 1969 *Battelle Rencontres: Group Representations*, edited by V. Bargmann (Springer-Verlag, Berlin, 1970).
[8]Z. Haba, Rep. Math. Phys. **8**, 61 (1975).
[9]J.R. Klauder, Comm. Math. Phys. **18**, 307 (1970).
[10]C.M. Newman, Comm. Math. Phys. **26**, 169 (1972).
[11]M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960).
[12]J. Fröhlich, Ann. Inst. H. Poincaré **21**, 271 (1974).
[13]C.A. Orzalesi, Rev. Mod. Phys. **42**, 405 (1970).
[14]S. Coleman, J. Math. Phys. **7**, 787 (1966).
[15]T.D. Lee, S. Weinberg, and B. Zumino, Phys. Rev. Lett. **18**, 1029 (1967).
[16]B.W. Lee and H.T. Nieh, Phys. Rev. **166**, 1507 (1968).
[17]We call "current" a real, Wightman (pseduo-) vector field [J.T. Łopuszański and J. Lukierski, Rep. Math. Phys. **1**, 265 (1971)].
[18]E. Nelson, Ann. Math. **70**, 572 (1959).
[19]K. Maurin, *Methods of Hilbert Spaces* (PWN, Warsaw, 1972).
[20]In the noncyclic case we would have direct integrals of cyclic representations.
[21]We shall use the straightforward, easy notation denoting the representation of the subgroup in the same way as the formal exponents in Sec. III
[22]I.M. Gel'fand and N.Y. Vilenkin, *Generalized Functions, Vol. IV: Some Applications of the Harmonic Analysis* (Academic, New York, 1964).
[23]Y. Umemura, Publ. Res. Inst. Math. Sci. (Kyoto) A **1**, 1 (1966).
[24]J. Glimm and A. Jaffe, "Boson quantum field models," in *Mathematics of Contemporary Physics* (Academic, New York, 1972).
[25]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row and Peterson, Evanston, Ill., 1961).
[26]H. Araki, J. Math. Phys. **1**, 492 (1960).
[27]J. Dimock and J. Glimm, Advan. Math. **12**, 58 (1974).
[28]Unfortunately, such theories do not have a pure current character.[2,3]
[29]K. Bardakci, Y. Frishman, and M.B. Halpern, Phys. Rev. **170**, 1353 (1968).
[30]J. J. Sakurai, *Currents and Mesons* (University of Chicago Press, Chicago, 1969).
[31]S. Nagamachi and N. Mugibayashi, Progr. Theor. Phys. (Kyoto) **50**, 1066 (1973).
[32]G.G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Interscience, New York, 1972).
[33]N.I. Akhiezer and I.M. Glasman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961).

Pseudo white noises: δ-correlated processes with finite memory

A. Brissaud

Ecole Nationale Supérieure de Techniques Avancées and Ecole Polytechnique, Centre de Physique Théorique, 91120 Palaiseau, France

J. Léorat and A. Mazure

Université Paris VII and Observatoire de Meudon, 92190 Meudon, France
(Received 29 December 1975)

Stochastic processes are constructed from the trajectories of mutually independent particles subject to a central potential, and their characteristic functional is given. For a wide class of such processes, time dependence of the correlation function is found to follow a power law. Particular attention is paid to finite memory processes whose distribution function is a Dirac δ distribution, and an example of δ′correlated processes is also given.

1. INTRODUCTION

We introduce a class of stochastic processes constructed from the trajectories of mutually independent particles subject to central forces. These processes generalize the Holtsmark process[1,2] stochastic model for the microfield in an (electric or gravitational) plasma.

These processes are defined in Sec. 2; as for Gaussian processes, the analytical expression of the characteristic functional (and, consequently, of the joint probabilities) can be written explicitly.

Considering the similarity properties of certain trajectories, Sec. 3 shows that the time dependence of the correlation function of some processes is given by a power law, or, in particular cases, by a Dirac distribution ("pseudo white noises").

One of these white noises is more thoroughly studied in Sec. 4. In the contrary of Gaussian white noises, the time derivative of its samples is almost surely finite in any finite interval. It is also interesting to note that a stationary particle distribution may lead to nonstationary processes.

Finally, Sec. 5 studies a family of processes constructed from the trajectories of independent charged particles in a Coulomb potential. A vector valued pseudo white noise is obtained when $\mathcal{E}_0$, total energy per particle, is zero. In small or large $\mathcal{E}_0$ limit, it is shown that the statistical properties of these processes do not converge uniformly towards the statistical properties of the limiting processes, i.e., pseudo white noise and Holstmark process. Several results of this section are useful for the treatment of ionic lines emitted by a plasma.

2. DEFINITIONS

In the n-dimensional Euclidean space $\mathbb{R}^n$, we consider an infinite set $\{i\}$ of identical point particles without mutual interactions subject to a central potential. The trajectories of the particles are determined by the initial positions and velocities. This system is randomized by distributing identically and independently the initial positions and velocities. If only stationary and isotropic systems are considered, it is well known[3] that the distribution function is an arbitrary function of the energy per particle

$$\mathcal{E}=\frac{v^2}{2}+V(r), \tag{2.1}$$

where $\mathbf{r}$ (modulus r) and $\mathbf{v}$ (modulus v) are the position and velocity of the particle and V is the central potential. Hereafter, we always consider particles with the same energy $\mathcal{E}_0$; thus, the stationary distribution function is of the form

$$f(\mathbf{r},\mathbf{v})=\frac{C}{4\pi}\,\delta(\mathcal{E}-\mathcal{E}_0), \tag{2.2}$$

where C is an arbitrary constant and δ the Dirac distribution. With $\mathbf{r}$ and $\mathbf{v}$ variables, Eq. (2.2) reads

$$\left.\begin{aligned} f(\mathbf{r},\mathbf{v})&=\frac{C}{4\pi w(r)}\,\delta(v-w(r))\\ w(r)&=\{2[\mathcal{E}_0-V(r)]\}^{1/2}\end{aligned}\right\}\quad \text{if } \mathcal{E}_0-V(r)>0, \tag{2.3}$$

$$f(\mathbf{r},\mathbf{v})=0 \qquad \text{if } \mathcal{E}_0-V(r)<0. \tag{2.4}$$

On the above-defined random set, we may construct the stochastic processes

$$M(t)=\sum_i m(\mathbf{r}_i(t),\mathbf{v}_i(t)), \tag{2.5}$$

where the summation is over the particles of $\{i\}$ and m is a given real function of $\mathbf{r}$ and $\mathbf{v}$. These processes are indefinitely divisible[4] because the random functions

$$m_i(t)=m(\mathbf{r}_i(t),\mathbf{v}_i(t)) \tag{2.6}$$

are mutually independent and identically distributed.

More precisely, we define $M(t)$ as the limit of a process $M_R(t)$ constructed as follows. Let (S_n) be the hypersphere of radius R centered at the origin. The process $M_R(t)$ is obtained by restraining the summation in (2.5) to the only particles present at a given time $t=0$ in the sphere (S_n).

At this initial time, the number N of particles in the sphere is distributed according to the Poisson law,

$$P_R(N) = (\bar{N}_R^N/N!)\exp(-\bar{N}_R), \tag{2.7}$$

where $\bar{N}_R$ is the average number of particles in (S_n), given by

$$\bar{N}_R = \int_{(S_n)} d^n r \int d^n v f(\mathbf{r}, \mathbf{v}). \tag{2.8}$$

Note that the probability distribution of a given particle in the sphere (S_n) is

$$\mathrm{P}_R(\mathbf{r}, \mathbf{v}) = f(\mathbf{r}, \mathbf{v})/\bar{N}_R. \tag{2.9}$$

The process $M_R(t)$ may be characterized by it characteristic functional[5]

$$K_R(Z(s)) = \langle \exp[i\int_{-\infty}^{+\infty} Z(s) M_R(s)\, ds]\rangle, \tag{2.10}$$

where $Z(t)$ is a real distribution and the average is taken over the initial distribution of particles in (S_n). By definition of the process $M_R(t)$, we get

$$K_R(Z(s)) = \sum_{N=0}^{\infty} P_R(N)\{\langle \exp[i\int_{-\infty}^{+\infty} Z(s) m(s)\, ds]\rangle_R\}^N, \tag{2.11}$$

where $\langle\circ\rangle_R$ denotes an average over the distribution P_R [Eq. (2.9)]. Then, using Eqs. (2.7)–(2.9), we obtain

$$\log K_R(Z(s)) = -\int_{(S_n)} d^n r \int d^n v\; f(\mathbf{r}, \mathbf{v})\{1 - \exp[i\int_{-\infty}^{+\infty} Z(s) m(s)\, ds]\}. \tag{2.12}$$

Finally, the characteristic functional $K(Z(s))$ of the limiting process $M(t)$, limit of $K_R(Z(s))$ for $R \to \infty$, is given by

$$\log K(Z(s)) = -\int d^n r \int d^n v f(\mathbf{r}, \mathbf{v})\{1 - \exp[i\int_{-\infty}^{+\infty} Z(s) m(s)\, ds]\}. \tag{2.13}$$

Note that (2.13) includes the multivariate probability distribution $P(M_1, t_1; \cdots; M_k, t_k)$, Fourier transform of the characteristic functional of $Z(t) = \sum_{j=1}^{k} z_j \delta(t - t_j)$.

3. SIMILAR TRAJECTORIES AND CORRELATION FUNCTIONS

The aim of this section is to analyze the consequences of certain similarity properties of trajectories. First, let us look at the simple case of particles describing straight lines in $\mathbb{R}^n$ with constant speed v. The motion can be written in the suitable basis of $\mathbb{R}^2$

$$\frac{d^2\rho}{d\tau^2} = 0, \quad \rho(0) = \begin{pmatrix}1\\0\end{pmatrix}, \quad \left.\frac{d\rho}{d\tau}\right|_{\tau=0} = \begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix}, \tag{3.1}$$

where $\tau = vt/r(0)$ and $\rho(\tau) = \mathbf{r}(t)/r(0)$. Equation (3.1) clearly shows that ρ depends on t and $r(0)$ through the only ratio $t/r(0)$, and this result is also valid for any quantity constructed from ρ and $d\rho/d\tau$, particularly for any angle related to the trajectory.

Now consider two real functions, m and m', of the form

$$m(t) = r^p(0)\varphi(\rho, \frac{d\rho}{d\tau}), \quad m'(t) = r^q(0)\varphi'(\rho, \frac{d\rho}{d\tau}), \tag{3.2}$$

where p and q are real numbers and where φ and φ' only depend on t and $r(0)$ through $\rho(\tau)$ and its derivatives; let us calculate

$$\Gamma(t) = \sum_{i,j} \langle m_i(0) m'_j(t)\rangle, \tag{3.3}$$

where the average is taken over the initial (homogeneous and isotropic) distribution of particles; assuming $\langle m_i(0)\rangle$ is zero and using the particle independence, we obtain an expression of the form

$$\Gamma(t) = \int_0^{\pi} d\alpha \int_0^{\infty} dr\, r^{n-1} r^p r^q \Theta(t/r, \alpha), \tag{3.4}$$

where the integration variable r stands for $r(0)$; then a simple change of variable, and the use of time and angle symmetries, lead to

$$\Gamma(t) = \epsilon A t^{p+q+n}, \tag{3.5}$$

where ϵ may be either 1 or $\mathrm{sgn}(t)$. Thus, if the constant A is neither zero nor infinite, the time dependence of $\Gamma(t)$ is given by a power law.

For example, let us take for m and m' the components of the Coulomb field $\mathbf{r}(t)/r^3(t)$ in $\mathbb{R}^3$ (Holtsmark process[1,2]), corresponding to $p = q = -2$ and $n = 3$; from Eq. (3.5) we immediately recover the Taylor's result[6]

$$\Gamma(t) = A|t|^{-1}. \tag{3.6}$$

We can also look for a "pseudo white noise" $M(t)$ in $\mathbb{R}^3$ such as

$$\int_0^t ds \langle M(0) M(s)\rangle = \langle M(0) \int_0^t M(s)\, ds\rangle = A_1\, \mathrm{sgn}(t). \tag{3.7}$$

If $\int_0^t m(s)\, ds$ is of the form $m'(t)$ of (3.2), p and q must necessarily satisfy $q = p + 1$ and $p + q + 3 = 0$, which gives $p = -2$; but, using (3.5), we find that the autocorrelation function of processes corresponding to $p = -2$ is proportional to $|t|^{-1}$; in fact, if A_1 is finite in (3.7), A vanishes in (3.5) and a more complete analysis of (3.4) shows that the autocorrelation function of $M(t)$ is really a Dirac distribution. Section 4 studies an example of this case, the scalar pseudo white noise $M(t) = (d/dt)(\sum_i [r_i(t)]^{-1})$.

Considering again particles in $\mathbb{R}^n$, let us look now at the more general problem of trajectories in the central attractive potential

$$V(r) = -r^{-2a}, \tag{3.8}$$

where a is a positive number. Among them, it is easy to see that trajectories that correspond to zero energy show similarity properties analogous to the preceding ones. Indeed, since the speed is determined at each point, $v(\mathbf{r}) = 2^{1/2} r^{-a}$, the equation of motion can be written in $\mathbb{R}^2$

$$\frac{d^2\rho}{d\tau^2} = -\frac{a\rho(\tau)}{[\rho(\tau)]^{2(a+1)}}, \quad \rho(0) = \begin{pmatrix}1\\0\end{pmatrix}, \quad \left.\frac{d\rho}{d\tau}\right|_{\tau=0} = \begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix}, \tag{3.9}$$

where $\tau = 2^{1/2} t/[r(0)^{a+1}]$ and $\rho(\tau) = \mathbf{r}(t)/r(0)$. Proceeding in the same way as for the straight line trajectories, and using expression (2.3) for the initial distribution of particles, we get an expression of the form

$$\Gamma(t) = \int_0^{\pi} d\alpha \int_0^{\infty} dr\, r^{n-1} r^p r^q r^{-a} \Theta(t/r^{a+1}, \alpha) \tag{3.10}$$

for the correlation function $\Gamma(t)$ defined by (3.3); integrating, we obtain a power law again,

$$\Gamma(t) = \epsilon A t^{(p+q+n-a)/(a+1)}. \qquad (3.11)$$

As for the straight line case, let us look for a pseudo white noise constructed from particles in $\mathbb{R}^3$. In order to satisfy (3.7) and (3.11), we must have $q = p + a + 1$ and $p + q + 3 - a = 0$. So, for any a, we find the same result as for straight line trajectories, i.e., $p = -2$. Section (5) will analyze a vector valued pseudo white noise of this type, the random field $\mathbf{M}(t) = \sum_i \mathbf{r}_i(t)/r_i^3(t)$, where the $\mathbf{r}_i(t)$ are parabolic trajectories.

Concluding this section, let us make some remarks about the correlation functions (3.5) and (3.11). First we note that this power law time dependence is associated with a quite large class of processes since it involves four free parameters characterizing the correlated functions, the dimension of space, and the central potential. Second, it is easy to see that the variance of these processes is infinite. Third, it may be interesting to consider linear stochastic equations whose random coefficients are of the form $\sum_i m_i(t)$. Indeed, in the small Kubo number[7,8] limit, we know that the correlation function of the coefficients plays a main role in the average solution of these equations; if $\Gamma(t)$ is not integrable for short times (i.e., if $p + q + n + 1 \leqslant 0$), the Bourret equation[9] cannot be used, but one may look for an approximate average solution by modelizing the coefficients with suitably correlated Kangaroo Processes. The justification of this method is given in a particular case in Refs. 10,11.

4. A SCALAR PSEUDO WHITE NOISE

In this section, we exhibit a first example of those δ-correlated stochastic processes foreseen in the preceding section.

Let $\{i\}$ be a set of free particles in $\mathbb{R}^3$, describing straight line trajectories

$$\mathbf{r}_i(t) = \mathbf{r}_i(0) + \mathbf{v}_i t, \qquad (4.1)$$

and initially distributed according to

$$f(\mathbf{r}, \mathbf{v}) = d\delta(v - w)/4\pi w^2, \qquad (4.2)$$

where d is the uniform density, w the speed (common to all particles) and δ the Dirac distribution; in Sec. 2 we recalled that such a distribution is stationary.

The "scalar pseudo-noise" $M(t)$ is defined by

$$M(t) = \frac{dU}{dt}, \qquad (4.3)$$

where

$$U(t) = \sum_i r_i^{-1}(t) \qquad (4.4)$$

is the Coulomb potential of the particles at the origin.

Let us look at some statistical properties of this process, and first to its mean. Integrating $M(t)$, we obtain

$$\left\langle \int_0^t M(s)\, ds \right\rangle = \langle U(t) - U(0) \rangle, \qquad (4.5)$$

where $\langle \circ \rangle$ denotes an average over the distribution of particles at time $t = 0$. In the rhs of Eq. (4.5), the angular average may be done as follows: One particle situated in $\mathbf{r}$ with a random isotropic speed direction at time $t = 0$ is, at time t, uniformly distributed upon the sphere whose center is $\mathbf{r}$ and radius wt: thus, the angular average of the particle potential at time t is either $(wt)^{-1}$ (if the origin is inside the sphere) or r^{-1} (if the origin is outside); applied to Eq. (4.5), this result gives

$$\left\langle \int_0^t M(s)\, ds \right\rangle = 4\pi d \int_0^{wt} dr\, r^2 [(wt)^{-1} - r^{-1}] = -(2\pi d/3) w^2 t^2. \qquad (4.6)$$

$M(t)$ being defined as a limiting process (see Sec. 2) we can commute averaging and time derivation in the lhs of (4.6), obtaining

$$\langle M(t) \rangle = -(4\pi d/3) w^2 t. \qquad (4.7)$$

Although the particle distribution is stationary, we see that $M(t)$ is not stationary and that its definition must include the "initial" time when the average over the particle distribution is taken. This result must be interpreted in terms of processes $M_R(t)$ of Sec. 2: clearly, the average potential of particles initially in a sphere (S_n) is a decreasing function of time because these particles are more and more distant; Eqs. (4.5)–(4.6) show that this decrease does not reduce to zero in the limit $R \to \infty$, because of the dominant effect of distant particles.

Now look at the moments of the zero mean process

$$M'_R(t) = M_R(t) - \langle M_R(t) \rangle. \qquad (4.8)$$

It is easy to see that these moments can be expressed in terms of quantities of the form

$$Q_R^k(t_1, \ldots, t_k) = \int_0^R r^2\, dr \int_0^\pi d\alpha \sin\alpha\, m(t_1) \times \cdots \times m(t_k), \qquad (4.9)$$

where

$$m(t_j) = \left[\frac{d}{dt}\, r^{-1}(t) \right]_{t=t_j} = \frac{w(r\cos\alpha + wt_j)}{(r^2 + w^2 t_j^2 + 2rwt_j \cos\alpha)^{3/2}}, \qquad (4.10)$$

and where k *is at least two*. In the limit $R \to \infty$, the Q_R^k are stationary because, for any h, the change of variable $\mathbf{r}' = \mathbf{r} + \mathbf{v}h$ only modifies the integration of (4.9) in a shell of radius R and thickness $2wh$, whose contribution gives zero in the large R limit, for $k \geqslant 2$. Thus $M'(t)$ is stationary (in the sense that all its moments are invariable in a time translation); $M(t)$ appears as the sum of a drift term $-(4\pi d/3)w^2 t$ and of a stationary stochastic process $M'(t)$ whose main statistical properties will be studied now.

First, consider the covariance

$$\Gamma(t) = \langle M'(t + t') M'(t') \rangle = \langle M(t) M(0) \rangle, \qquad (4.11)$$

and calculate its integral

$$\Gamma_1(t) = \int_0^t \langle M(s) M(0) \rangle\, ds. \qquad (4.12)$$

From equations (4.1)–(4.4), Γ_1 can be written explicitly

$$\Gamma_1(t) = -2\pi w\, d \int_0^\infty ds \int_0^\pi d\alpha \sin\alpha \cos\alpha (s^2 + t^2 + 2st\cos\alpha)^{-1/2}. \tag{4.13}$$

Integrating, we obtain

$$\Gamma_1(t) = 2\pi w\, d\, \mathrm{sgn}(t), \tag{4.14}$$

or, equivalently

$$\Gamma(t) = 4\pi w\, d\delta(t). \tag{4.15}$$

Now, let us stress two interesting points.

First, the covariance of $M(t)$ is a Dirac distribution; this result may seem strange because the memory of the process is clearly finite (the time derivative of the process is almost surely finite); in fact, this is a purely statistical effect.

Second, we obtained an "analytical" expression of the Dirac distribution

$$\delta(t - t') = \tfrac{1}{2} \int_0^\infty ds \int_0^\pi d\alpha\, s^2 \sin\alpha \times \frac{t + s\cos\alpha}{(t^2 + s^2 + 2st\cos\alpha)^{3/2}} \frac{t' + s\cos\alpha}{(t'^2 + s^2 + 2st'\cos\alpha)^{3/2}}, \tag{4.16}$$

expression which may be simplified by setting $t' = 0$.

When solving stochastic equations, two other statistical properties are important, the characteristic function

$$K(z) = \langle \exp(izM) \rangle, \tag{4.17}$$

and the instantaneous probability distribution

$$P(M) = (2\pi)^{-1} \int dz \exp(-izM) K(z). \tag{4.18}$$

Since the zero mean process $M'(t) = M(t) - \langle M(t) \rangle$ is stationary, it is sufficient to calculate $K(z)$ and $P(M)$ at time $t = 0$; using the general expression of the characteristic functional (2.13), we first recover the Holstmark-like result

$$\log[K(z)] = -(4d/15)(2\pi w)^{3/2} |z|^{3/2}, \tag{4.19}$$

which leads to the following asymptotic expansions of $P(M)$ for large and small M

$$P(M) = 2^{47/12} 3^{-5/6} 5^{-11/6} \pi^{9/4} d w^{3/2} M^{-5/2} + O\{d^2 w^3 M^{-4}\},$$
$$P(M) = \left(\frac{15}{4d}\right)^{2/3} \frac{\Gamma(2/3)}{3\pi^2 w} + O\{d^{-2} w^{-3} M^2\}. \tag{4.20}$$

Note that $K(z)$ and $P(M)$ are not Gaussian.

Now look at the properties of the time average of $M(t)$,

$$N(t) = t^{-1} \int_0^t M(s)\, ds, \tag{4.21}$$

where $t = 0$ is again the time when the average over the particle distribution is taken.

The characteristic function of $N(t)$ may be written

$$K_1(z, t) = \langle \exp(izt^{-1} \int_0^t M(s)\, ds) \rangle; \tag{4.22}$$

using (2.13), we have

$$\log[K_1(z, t)] = -2\pi d \int_0^\infty dr \int_0^\pi d\alpha\, r^2 \sin\alpha \times (1 - \exp\{izt^{-1}[(r^2 + w^2 t^2 + 2wt\cos\alpha)^{-1/2} - r^{-1}]\}), \tag{4.23}$$

which can be rewritten in the form

$$\log K_1 = -2\pi\, d w^3 t^3 \psi(z w^{-1} t^{-2}),$$
$$\psi(x) = \int_0^\infty d\rho \int_0^\pi d\alpha\, \rho^2 \sin\alpha \times (1 - \exp\{ix[(\rho^2 + 1 + 2\rho\cos\alpha)^{-1/2} - \rho^{-1}]\}). \tag{4.24}$$

Expanding $\psi(x)$ for small x, we get

$$\psi(x) = ix/3 + x^2 + O(x^3 \log x); \tag{4.25}$$

so, in the long time limit, we obtain the Gaussian result

$$\log[K_1(z, t)] = iz\langle N(t) \rangle - \tfrac{1}{2} z^2 \langle N'(t) N'(t) \rangle, \tag{4.26}$$

where we have used the expressions

$$\langle N(t) \rangle = \tfrac{1}{2} \langle M(t) \rangle = -(2\pi d/3) w^2 t,$$
$$N'(t) = N(t) - \langle N(t) \rangle, \tag{4.27}$$
$$\langle N'(t) N'(t) \rangle = 2\pi\, dw/t.$$

We conclude that $N(t)$ is distributed as the time average of a Gaussian white noise in the long time limit. But we shall see now that its distribution is completely different in the opposite limit. In order to avoid a detailed proof (see Ref. 12 in a similar case), let us remark that only large ρ's play a role in the expression $\exp\{ix[(\rho^2 + 1 + 2\rho\cos\alpha)^{-1/2} - \rho^{-1}]\}$ when calculating $\psi(x)$ for large x. Expanding the argument to the first order in ρ^{-1}, we exactly obtain the corresponding term of the characteristic function (4.17) of M; we conclude that $N(t)$ and M are identically distributed in the short time limit. This result, which would have no significance for a Gaussian white noise, can be interpreted in terms of the average solution of the pseudo harmonic oscillator,

$$\frac{d}{dt} q(t) = i\lambda M(t) q(t), \quad q(0) = 1; \tag{4.28}$$

indeed, this average solution reads

$$\langle q(t) \rangle = \langle \exp(i\lambda \int_0^t M(s)\, ds) \rangle = K_1(\lambda t, t); \tag{4.29}$$

thus, from the preceding result, and in the short time limit,

$$\langle q(t) \rangle \approx K(\lambda t) = \langle \exp[i\lambda t M(0)] \rangle. \tag{4.30}$$

Equation (4.30) means that, for large λ, the complete damping of $\langle q(t) \rangle$ is given by the static approximation,[8] i.e., by treating $M(t)$ as a random *variable* distributed according to $P(M)$.

5. COULOMB FIELD AT A CHARGED POINT

Stochastic Stark broadening of atomic and ionic spectral lines emitted by a plasma is an interesting field for applying stochastic equation methods,[13,10,11] because line profiles are expressed in terms of the average solution of a linear equation in which the "stochastic" electric microfield appears in the coefficients. Being the sum of individual fields due to the plasma electrons and ions, whose trajectories in the vicinity of the emitter are approximately either straight lines (atomic emitter) or hyperboles (ionic emitter), this microfield gives a physical background to the following process.

Let $\{i\}$ be a set of equienergetic, mutually indepen-

TABLE I. Characteristic function $K(z)$ and probability distribution P, (M) of straight line, parabolic and hyperbolic processes. $\alpha_1 = 4/15\,(2\pi)^{3/2}$; $\alpha_2 = (128/45)\pi[(1+\sqrt{2})/\sqrt{2}]^{1/2}\Gamma(3/4)$; $\beta_1 = (75/32\pi^{-4}$; $\beta_2 = (8/5\pi)(\alpha_2)^{-12/5}\Gamma(12/5)$; $\gamma_2 = (128/45)(1+\sqrt{2})\Gamma(3/4)$.

Trajectories	$\log[K(z)]$	$P_1(M)$ (small M)	$P_1(M)$ (large M)
Straight lines	$-\alpha_1 C(2\mathcal{E}_0)^{1/2}\|z\|^{3/2}$	$\beta_1 C^{-2}(2\mathcal{E}_0)^{-1}M^2 + \mathcal{O}(C^{10/3}\mathcal{E}_0^{5/3}M^4)$	$2\pi C(2\mathcal{E}_0)^{1/2}M^{-5/2} + \mathcal{O}(C^2\mathcal{E}_0 M^{-4})$
Parabolas	$-\alpha_2 C\|z\|^{5/4}$	$\beta_2 C^{-12/5}M^2 + \mathcal{O}(C^{-4}M^4)$	$\gamma_2 C M^{-9/4} + \mathcal{O}(C^2 M^{-7/2})$
Hyperbolas (to the first nonvanishing order)	small $\|z\|\mathcal{E}_0^2$: as for parabolas large $\|z\|\mathcal{E}_0^2$: as for straight lines	proportional to M^2	as for parabolas

dent, particles in $\mathbb{R}^3$, submitted to the central attractive potential

$$V(r) = -r^{-1}, \tag{5.1}$$

and randomly distributed at the initial time (see Sec. 2). According to the common positive energy $\mathcal{E}_0$, three types of trajectories may occur: in the general case, the motion is hyperbolic and the stationary density [see Eq. (2.3)] reads

$$d(r) = C[2(\mathcal{E}_0 + r^{-1})]^{1/2}. \tag{5.2}$$

For zero $\mathcal{E}_0$, the motion is parabolic and the density is $2^{1/2}Cr^{-1/2}$. In the large $\mathcal{E}_0$ limit, we recover straight line trajectories with uniform density $d = 2^{1/2}C\mathcal{E}_0^{1/2}$. Processes constructed from bound trajectories ($\mathcal{E}_0 < 0$) will not be studied here.

Now consider the vector valued process

$$\mathbf{M}(t) = \sum_i \mathbf{r}_i(t)/[r_i(t)]^3, \tag{5.3}$$

where the summation is extended to the above set $\{i\}$. As seen in Secs. 2 and 4, such a process must be defined more precisely by restraining the sum in Eq. (5.3) to the only particles that are inside a sphere a radius R at a given time, and by taking the limit $R \to \infty$. In this limit, it is clear that the mean $\langle \mathbf{M}(t)\rangle$ is zero (for symmetry reasons) and that higher order moments are stationary (see the scalar pseudo-white-noise of Sec. 4). Now look at some main statistical properties of $\mathbf{M}(t)$.

A. Correlation function

For symmetry reasons, the correlation tensor

$$\Gamma_{\alpha\beta}(t) = \langle M_\alpha(t) M_\beta(0)\rangle \tag{5.4}$$

is given by its trace

$$\Gamma(t) = \langle \mathbf{M}(t)\cdot\mathbf{M}(0)\rangle, \tag{5.5}$$

where the point denotes a scalar product and where $\langle\cdot\rangle$ indicates an average over the initial distribution of particles.

In the no-potential case, the well-known result[6]

$$\Gamma(t) = 4\pi\, d/w\,|t| = 4\pi C/|t| \tag{5.6}$$

is not integrable for short time; on the contrary, in the case of particles subject to a Coulomb potential, there always exists the function

$$\Gamma_1(t) = \int_0^t \Gamma(s)\,ds; \tag{5.7}$$

after angular average, the Appendix shows that this function reads

$$\Gamma_1(t) = 2\pi C \int_0^\infty [\{r_-(t)\}^{-1} - \{r_+(t)\}^{-1}]\,dr, \tag{5.8}$$

where $r_-(t)$ [resp. $r_+(t)$] is the position at time t of a particle whose initial position is r and whose initial velocity is directed towards the center (resp. opposite). Since $r_-(t) = r_+(-t)$, $\Gamma_1(t)$ is an odd function of t.

First look at zero energy particles, following parabolic trajectories; in this case, the axial motion is given by $r_\pm(t) = (|r^{3/2} \pm 3t/2^{1/2}|)^{2/3}$; setting $2u = r(3t/4)^{-2/3}$, Eq. (5.8) reads

$$\Gamma_1(t) = 2\pi C\,\mathrm{sgn}(t)\int_0^\infty du[|u^{3/2} - 1|)^{-2/3} - (|u^{3/2} + 1|)^{-2/3}], \tag{5.9}$$

or

$$\Gamma_1 = A\,\mathrm{sgn}(t), \tag{5.10}$$

where A is a numerical constant ($\approx 2\pi C \times 1.81$). Since Γ_1 is a step function, $\Gamma(t)$ is the Dirac distribution

$$\Gamma(t) = 2A\delta(t). \tag{5.11}$$

Therefore, in the case of parabolic trajectories, the Coulomb field $\mathbf{M}(t)$ is a vector valued "pseudo-white-noise" in the sense of Sec. 4.

Now consider positive energy particles, following hyperbolic trajectories; in this case the Appendix shows that $\Gamma_1(t)$ can be written $4\pi C\gamma[(2\mathcal{E}_0)^{3/2}t]$, where γ is an universal function. From this expression, straightforward calculations show that the hyperbolic covariance $\Gamma(t)$ is the sum of the Dirac distribution (5.11) (parabolic covariance) and of a function, integrable for finite times, and behaving for long times as (5.6) (straight line covariance).

B. Characteristic function and probability distribution

The characteristic function $K(\mathbf{z})$ and the instantaneous probability distribution $P(\mathbf{M})$ are defined by

$$K(\mathbf{z}) = \langle \exp(i\mathbf{z}\cdot\mathbf{M})\rangle, \tag{5.12}$$
$$P(\mathbf{M}) = (2\pi)^{-3}\int d^3z\,\exp(-i\mathbf{z}\cdot\mathbf{M})K(z).$$

Since $\mathbf{M}$ is isotropic, we shall only consider here

$$P_1(M) = 4\pi M^2 P(\mathbf{M}). \tag{5.13}$$

Using Eq. (2.13) and isotropy, we obtain

$$\log[K(\mathbf{z})] = -4\pi C \int_0^\infty dr\, r^2 [2(\mathcal{E}_0 + r^{-1})]^{1/2} \left[1 - \frac{\sin(z/r^2)}{z/r^2}\right], \tag{5.14}$$

where $z = |\mathbf{z}|$. According to the various trajectories, the behavior of $K(\mathbf{z})$ and $P_1(M)$ is shown in Table I.

C. Time average of the process

Now consider the time average of the process,

$$\mathbf{N}(t) = t^{-1} \int_0^t \mathbf{M}(s)\, ds, \tag{5.15}$$

and, especially, its characteristic function

$$K_1(\mathbf{z}, t) = \langle \exp[it^{-1}\mathbf{z} \cdot \int_0^t \mathbf{M}(s)\, ds]\rangle. \tag{5.16}$$

For straight line trajectories, such an expression was often calculated[14-16]; in Ref. 12, it is shown in details that $K_1(\mathbf{z},t)$ behaves as $K(\mathbf{z})$ (given by Table I) in the short time limit and behaves as $\exp[-(4\pi C/3) z^2 t^{-1} \times \log t]$ in the long time limit. If trajectories are hyperbolic or parabolic, it can be shown in the same way that the short time limit of $K_1(\mathbf{z},t)$ is again $K(\mathbf{z})$ and that the long time limit is the Gaussian limit

$$\log[K_1(\mathbf{z},t)] = -(z^2/6)\langle \mathbf{N}(t) \cdot \mathbf{N}(t)\rangle. \tag{5.17}$$

D. Limits of the statistical properties of the hyperbolic process for large or small $\mathcal{E}_0$

By comparing the preceding results, it is clear that we do not obtain a uniform convergence of the hyperbolic statistical properties towards the straight line or parabolic properties, in the large or small $\mathcal{E}_0$ limit. A complete discussion might be tedious, but it is interesting to consider at least an example, for instance the long time behavior of $K_1(\mathbf{z},t)$.

In the parabolic case, using (5.11) in (5.17), we get the asymptotic behavior

$$\log[K_1(\mathbf{z},t)] = -Az^2(3t)^{-1}, \quad \text{with } A \approx 2\pi C \times 1.81. \tag{5.18}$$

In the hyperbolic case, using now the long time behavior (5.6) of the covariance we obtain

$$\log[K_1(\mathbf{z},t)] = -4\pi C z^2 (3t)^{-1} \log t. \tag{5.19}$$

Since this last expression does not depend upon $\mathcal{E}_0$, it certainly does not lead to (5.18) in the small $\mathcal{E}_0$ limit; in fact we should have written $\log(\mathcal{E}_0^{3/2} t)$ instead of $\log t$ in (5.19), because this asymptotic expression is only valid for times larger than $\mathcal{E}_0^{-3/2}$; this explains the apparent paradox, and also shows that the convergence of the hyperbolic K_1 towards the parabolic K_1 is not uniform in the small $\mathcal{E}_0$ limit. Similar conclusions would be obtained for the large $\mathcal{E}_0$ limit and for other statistical properties.

6. CONCLUDING REMARKS

This paper has defined a rather large class of stochastic processes with infinite variance. Some examples of these have been studied in Secs. 4–5. Among other potentially interesting processes of this class, we note the δ'-correlated processes, which correspond to $2p = -(n + a + 2)$ in the notation of Sec. 3; the simplest example is probably constructed in $\mathbb{R}^2$ from straight line trajectories and $m(t) = (d^2/dt^2)\log[r(t)]$. We also note the quite simple processes constructed from trajectories in the one-dimensional space $\mathbb{R}^1$, whose possible discontinuities at each passage of a particle could lead to interesting problems.

ACKNOWLEDGMENTS

We are indebted to U. Frisch and Hénon for very helpful comments.

APPENDIX: HYPERBOLIC TRAJECTORIES

First, we recall the expression of hyperbolic trajectories.[17] Let $\xi(t)$ be the motion parameter, satisfying

$$e[\sinh\xi(t) - \sinh\xi(0)] - [\xi(t) - \xi(0)] = (2\mathcal{E}_0)^{3/2} t, \tag{A1}$$

where e is the eccentricity and $\xi(0)$ the initial parameter,

$$\begin{aligned} e &= [1 + 4\mathcal{E}_0 r(1 + \mathcal{E}_0 r)\sin^2\alpha]^{1/2}, \\ e\sinh\xi(0) &= -2[\mathcal{E}_0 r(1 + \mathcal{E}_0 r)]^{1/2}\cos\alpha, \end{aligned} \tag{A2}$$

r is the initial distance of the particle, and α the angle between the initial position and velocity.

The trajectory $\mathbf{r}(t)$ can be expressed in terms of $\xi(t)$; in particular we have

$$r(t) = [e\cosh\xi(t) - 1]/2\mathcal{E}_0, \tag{A3}$$

and

$$\left|\int_0^t \frac{\mathbf{r}(s)}{[r(s)]^3} ds\right|^2 = \frac{2}{r} \frac{\cosh[\xi(t) - \xi(0)] - 1}{e\cosh\xi(t) - 1}. \tag{A4}$$

Now calculate the covariance of the process $\mathbf{M}(t)$ of Sec. 5. We have first

$$\int_0^t ds \int_0^s ds' \langle \mathbf{M}(s') \cdot \mathbf{M}(0)\rangle = \tfrac{1}{2} \int d^3r \int d^3v\, f(\mathbf{r},\mathbf{v}) \times \left| \int_0^t \frac{\mathbf{r}(s)}{[r(s)]^3} ds\right|^2. \tag{A5}$$

Using Eqs. (2.3), (A4), and (A5), we obtain

$$\int_0^t ds \int_0^s ds' \langle \mathbf{M}(s') \cdot \mathbf{M}(0)\rangle = 4\pi c (2\mathcal{E}_0)^{-1/2} \int_0^\infty dr \int_0^\pi d\alpha \sin\alpha \times [\mathcal{E}_0 r(1 + \mathcal{E}_0 r)]^{1/2} \{\cosh[\xi(t) - \xi(0)] - 1\}/[e\cosh\xi(t) - 1]. \tag{A6}$$

Then, from the identities

$$\frac{d}{d\alpha}[\xi(t) - \xi(0)] = -2\sin\alpha[\mathcal{E}_0 r(1 + \mathcal{E}_0 r)]^{1/2} \times \{\cosh[\xi(t) - \xi(0)] - 1\}/[e\cosh\xi(t) - 1], \tag{A7}$$

$$\frac{d}{dt}[\xi(t) - \xi(0)] = (2\mathcal{E}_0)^{1/2}[r(t)]^{-1},$$

we get

$$\Gamma_1(t) = \int_0^t ds \langle \mathbf{M}(s) \cdot \mathbf{M}(0)\rangle = 2\pi C \int_0^\infty [\{r_-(t)\}^{-1} - \{r_+(t)\}^{-1}]\, dr, \tag{A8}$$

where $r_-(t)$ and $r_+(t)$ are the distance of a particle, initially in r, whose velocity is initially directed either towards the center or in the opposite direction.

By using (A1)–(A3) and (A8), $\Gamma_1(t)$ can be written in the form

$$\Gamma_1(t) = 4\pi C\gamma[(2\mathcal{E}_0)^{3/2}t], \tag{A9}$$

where $\gamma(s)$ is given by

$$\gamma(s) = \int_0^\infty d\rho[\{\cosh\xi^-(s) - 1\}^{-1} - \{\cosh\xi^+(s) - 1\}^{-1}],$$

$$\sinh\xi^\pm(s) - \xi^\pm(s) = s + \sinh\xi^\pm(0) - \xi^\pm(0), \tag{A10}$$

$$\sinh\xi^\pm(0) = \pm 2[\rho(1+\rho)]^{1/2}.$$

[1]J. Holtsmark, Ann. Phys. (Leipzig) 58, 577 (1919).
[2]S. Chandrasekhar, Rev. Mod. Phys. 15, 1 (1943).
[3]J.H. Jeans, Monthly Notices 76, 71 (1915).
[4]P. Levy, *Processus stochastiques et mouvement brownien* (Gauthier-Villars, Paris, 1965).
[5]E. Hopf, Proc. Symp. Appl. Math. 7 (1957).
[6]J.B. Taylor, Phys. Fluids 3, 792 (1960).
[7]R. Kubo, J. Math. Phys. 4, 174 (1963).
[8]A. Brissaud and U. Frisch, J. Math. Phys. 15, 524 (1974).
[9]R. Bourret, Nuovo Cimento 26, 1 (1962).
[10]A. Brissaud and U. Frisch, J. Quant. Spectrosc. Radiat. Transfer 11, 1767 (1971).
[11]A. Brissaud, C. Goldbach, J. Léorat, A. Mazure, and G. Nollez, "On the validity of the model microfield method as applied to stark broadening of neutral lines," to be published in J. Phys. B.
[12]U. Frisch and A. Brissaud, J. Quant. Spectrosc. Radiat. Transfer 11, 1753 (1971).
[13]H.R. Griem, *Spectral line broadening by plasmas* (Academic, New York, 1974).
[14]P.W. Anderson and J.D. Talman, Proc. Conf. on Broadening of Spectral Lines, Univ. of Pittsburgh, Bell Telephone Syst. Monogr. 3117 (1956).
[15]L. J. Roszman and C.F. Hooper, Phys. Rev. A 7, 2121 (1973).
[16]E.R. Segre and D. Voslamber, Phys. Lett. A 46, 397 (1974).
[17]L. Landau and E. Lifschitz, *Mechanics* (Mir ed., Moscow, 1969).

Prolongation structures of nonlinear evolution equations. II

F. B. Estabrook and H. D. Wahlquist

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103
(Received 21 October 1975)

The prolongation structure of a closed ideal of exterior differential forms is further discussed, and its use illustrated by application to an ideal (in six dimensions) representing the cubically nonlinear Schrödinger equation. The prolongation structure in this case is explicitly given, and recurrence relations derived which support the conjecture that the structure is open—i.e., does not terminate as a set of structure relations of a finite-dimensional Lie group. We introduce the use of multiple pseudopotentials to generate multiple Bäcklund transformation, and derive the double Bäcklund transformation. This symmetric transformation concisely expresses the (usually conjectured) theorem of permutability, which must consequently apply to all solutions irrespective of asymptotic constraints.

I. INTRODUCTION

In the first of these papers[1] we introduced a geometric method for finding a hierarchy of potentials and pseudopotentials (denoted y^i) for sets of nonlinear partial differential equations with two independent variables. By "geometric" is meant the systematic use of the formalism of differential geometry, in particular the representation of the partial differential equations as a closed ideal of differential forms. To find the potentials and pseudopotentials, it turns out that one must solve—find representation of—a Lie structure of auxiliary vector fields. We have denoted this the "prolongation structure," or PS. The vector fields, say $X_1, X_2, \cdots$, and their commutation relations are defined in the space of "prolongation" variables y^i, while the defining equations for the potentials, or pseudopotentials, appear as 1-forms in the space of *all* variables—the original, or primitive variables (independent variables x and t, say, and dependent variables z^A) plus the y^i,

$$\omega^i \equiv dy^i + F^i(z^A, y^i)\, dx + G^i(z^A, y^i)\, dt.$$

The F^i and G^i depend on the y^i by being linear in the components $X^i = X \cdot dy^i$ of the X's. These 1-forms each lead to a conservation law, since $d\omega^i$ is required to be in the *prolonged* ideal: the ideal representing the original partial differential set, to which is also adjoined the ω^i. Consequently, for any solution manifold of the ideal, the ω^i are exact and Stokes' theorem says that $\oint \omega^i$, confined to that manifold, vanishes. This hierarchy of higher conservation laws is essentially different from those previously recognized [for equations such as the Korteweg–de Vries (KdV) and nonlinear Schrödinger], which involved repeated partial derivatives.

Coordinate transformation in the space of the y^i leaves the abstract algebraic relations—the PS—among the X's unchanged: It is equivalent to changing the X's individually by local similarity transformation, while forming trivial superpositions of the pseudopotentials and associated conservation laws. It is important, however, that coordinates must exist for which the X's are homogeneous first degree in the y^i—the finding of these is equivalent to finding matrix (or linear) representations of the X's, a standard procedure in the discussion of Lie groups. It is in such coordinate frames that one finds the linear auxiliary partial differential equations for pseudopotentials already known in application to the boundary value problem as the method of inverse scattering. More general linear representations surely can be found, and their application remains to be investigated.

The PS is further invariant under less trivial transformations in the space of the y^i: Linear superpositions of the X's can be found which keep their commutation relations invariant. This is a group of automorphisms which, in the KdV case, at least, turns out to be a 2-parameter group isomorphic to an invariance group of the initial KdV equation, before prolongation. These automorphisms imply that some of the pseudopotentials y^i can in fact constitute a continuous family. For the case of a two-dimensional representation of the KdV, this degenerates to a 1-parameter family of pseudopotentials, the automorphism parameter appearing as the so-called eigenvalue in the associated linear problem.

The existence of prolongation structures also seems to be closely related to the possibility of solution methods by Bäcklund transformation. We have found these to be derivable as discrete invariance transformations of the prolonged differential ideal, when a true pseudopotential exists. They again involve the automorphism parameters explicitly.

We speculate that the existence of a nontrivial PS may be a useful defining algebraic characteristic for the entire class of nonlinear equations now under intensive study in many contexts, roughly, those equations having solutions with nonlinear superposition properties, such as "solitons." By nontrivial we mean that the PS is non-Abelian, for then true pseudopotentials exist—viz., those y^i whose F^i and G^i cannot by coordinate transformation be made independent of y^i. Three now-classic soliton equations are the Korteweg–de Vries, sine-Gordon and nonlinear Schrödinger. We have given a PS for the first,[1] involving seven vectors $\mathbf{X}_i$ and an apparently open set derived from their commutators. A nontrivial PS also is readily derivable for the sine-Gordon equation. In the present paper we consider the PS of the third, which again is nontrivial and seems almost certainly to be open. As with the first two equations, this PS has useful low dimensional representations, and allows Bäcklund transformations. We introduce the use of multiple pseudopotentials, belonging to different values of the automorphism parameter, for the derivation of the multiple Bäcklund transformation.

The existence of simple pseudopotentials has been carefully discussed by Corones[2]; he also derives a prolongation structure for the Hirota equation. Prolongation structures for the Boussinesq equation and the nonlinear wave-envelope equations have been obtained by Morris.[3] Most recently Morris has devised an algorithm for extending an equation of evolution in one spatial dimension, which has a PS, to derive a related evolution equation in two spatial dimensions also having a PS.[4] In this way he has systematically discussed both the extension of the Boussinesq equation to the Kadomtsev–Petviashvili–Dryuma equation, and the extension of an interesting new nonlinear system having a PS to the generalization of the nonlinear Schrödinger equation of Ablowitz and Haberman.

The nonexistence of a non-Abelian PS does not of course mean that the systematic methods of differential geometry cannot then be useful for treating other classes of nonlinear differential systems. Many properties such as invariance operations, variational formulations, characteristics, and special solution sets such as similarity solutions, can in our opinion best be understood in this way. But the search for a PS, and the consequent conservation laws and potentials, does seem to be a first operation to try on a "well-formulated" set of partial differential equations—i.e., a set belonging to the regular integral manifolds (of maximum dimension) of a closed differential ideal. In the Appendix it is shown how such a search differentiates the KdV and modified KdV equations from those with higher order nonlinearities—the latter having only an Abelian PS.

II. THE PS OF THE NONLINEAR SCHRÖDINGER EQUATION

The cubically nonlinear Schrödinger equation has been treated extensively in the recent literature of nonlinear wave equations, cf. Whitham[5] and Scott *et al.*[6] With one sign ($\epsilon = -1$ in the following) of the nonlinear term, the equation describes stationary two-dimensional self-focusing of plane wave trains in nonlinear media (cf. Zakharov and Shabat,[7] Hirota,[8] and references therein), or the time dependent phenomenon of self-modulation (leading to the so-called "envelope" solitons). With the other sign ($\epsilon = +1$) the solutions have much greater stability; nevertheless, so-called envelope-hole solutions, etc., have been studied.[9,10] The local analytical method expounded in the present paper is applicable in either case. We present the prolongation structure, and construct some of the resulting new potential and pseudopotential conservation laws, and the inverse scattering equations. We use the pseudopotential to find the single Bäcklund transformation (independently given by Lamb[11]) and are able also to find a double Bäcklund transformation, or theorem of permutability.

The equation

$$i\psi_t + \psi_{xx} - \tfrac{1}{2}\epsilon\bar{\psi}\psi^2 = 0, \tag{1}$$

where $\epsilon = \pm 1$ and the bar denotes complex conjugate, can be expressed as the set of differential forms

$$\begin{aligned} \alpha_1 &= d\psi\wedge dt + \xi\, dx\wedge dt \\ \alpha_2 &= i d\psi\wedge dx + d\xi\wedge dt + \tfrac{1}{2}\epsilon\bar{\psi}\psi^2\, dx\wedge dt, \end{aligned} \tag{2}$$

together with the complex conjugates $\bar{\alpha}_1$ and $\bar{\alpha}_2$. These four 2-forms live in a six-dimensional space of primitive variables ψ, $\bar{\psi}$, ξ, $\bar{\xi}$, x, t, and have two-dimensional integral manifolds which are the solutions of (1). In the prolonged space of variables ψ, $\bar{\psi}$, ξ, $\bar{\xi}$, x, t, y^k, we search for Pfaffians of the form

$$\omega^k = dy^k + F^k(\psi, \bar{\psi}, \xi, \bar{\xi}, y^i)\, dx - iG^k(\psi, \bar{\psi}, \xi, \bar{\xi}, y^i)\, dt \tag{3}$$

(the factor $-i$ in the dt term proves convenient later), which are such that $d\omega^k$ are in the prolonged ideal α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$, ω^k. Following the same procedures as in Sec. I, we find overdetermined partial differential equations requiring a decomposition into polynomials in the primitive variables

$$\begin{aligned} F^k &= \tfrac{1}{2}[X_1^k + \psi\bar{\psi}X_2^k - 2\psi Z_1^k - 2\bar{\psi}\bar{Z}_1^k], \\ G^k &= \tfrac{1}{2}[(\xi\bar{\psi} - \bar{\xi}\psi)X_2^k - 2\xi Z_1^k + 2\bar{\xi}\bar{Z}_1^k \\ &\qquad + Y_2^k + \psi\bar{\psi}Y_2^k + \psi Z_2^k - \bar{\psi}\bar{Z}_2^k], \end{aligned} \tag{4}$$

where the vectors $Z_m(y^i)$ are complex, while $X_m(y^i)$ are real and $Y_m(y^i)$ pure imaginary,

$$\bar{X}_m = X_m, \quad \bar{Y}_m = -Y_m. \tag{5}$$

The remaining partial differential equations involve only dependency on the y^i and are all of commutator form, and so they define the prolongation structure:

$$\begin{aligned} &[X_1, X_2] = [X_1, Y_2] = [X_2, Y_1] = [X_2, Z_1] = [Z_1, Z_2] = 0, \\ &[X_1, Z_1] = Z_2, \quad [Z_1, \bar{Z}_1] = \tfrac{1}{2}Y_1, \\ &\tfrac{1}{2}[X_2, Z_2] + [Y_1, Z_1] - \epsilon Z_1 = 0, \quad [X_1, Z_2] + 2[Y_2, Z_1] = 0, \\ &[X_1, Y_1] + [X_2, Y_2] + 2[Z_1, \bar{Z}_2] - 2[\bar{Z}_1, Z_2] = 0, \end{aligned} \tag{6}$$

together with complex conjugates.

A number of further relations are derivable using the Jacobi identity. We find that X_2 must commute with all these vectors, and also that

$$[X_1, Y_1] = [Y_1, Y_2] = 0, \tag{7}$$

from which follow

$$[Y_1, Z_1] = \epsilon Z_1, \quad [Y_1, Z_2] = \epsilon Z_2. \tag{8}$$

At this point we have all the noncomplex vectors $\{X_m, Y_m\}$ constituting an Abelian subalgebra. Also, the set $\{Y_1, Z_1, \bar{Z}_1\}$ satisfies

$$[Y_1, Z_1] = \epsilon Z_1, \quad [Z_1, \bar{Z}_1] = \tfrac{1}{2}Y_1, \tag{9}$$

which is the algebra of the 3-parameter rotation group in complex notation. If we define a new complex generator

$$Z_3 \equiv [X_1, Z_2], \tag{10}$$

we can split up one of the remaining relations in Eq. (6) to get

$$[Y_2, Z_1] = -\tfrac{1}{2}Z_3, \tag{11}$$

as well as

$$[Z_1, Z_3] = [Z_2, Z_3] = 0, \quad [Y_1, Z_3] = \epsilon Z_3. \tag{12}$$

As was the case for the Korteweg–de Vries equation, this structure does not appear to terminate; in fact, it looks quite convincing that this case must be open. Noting in Eqs. (6) and (10) that X_1 has the property of generating higher complex vectors by the recursion relation

$$[X_1, Z_n] \equiv Z_{n+1}, \tag{13}$$

we consider the relations

$$[X_2, Z_n] = 0, \quad [Y_1, Z_n] = \epsilon Z_n,$$
$$[Y_2, Z_n] = -\tfrac{1}{2}Z_{n+2}, \ [Z_m, Z_n] = 0, \tag{14}$$

all of which are valid for the lowest values of n and m. These can be shown to be self-consistent, and to hold in general for arbitrary (n, m) by an argument of mathematical induction. There are two remaining relations to consider,

$$[Z_1, \bar{Z}_1] = \tfrac{1}{2}Y_1, \quad [Z_1, \bar{Z}_2] = [\bar{Z}_1, Z_2], \tag{15}$$

so that we have not as yet quite *proved* that this system does not terminate.

III. THE PSEUDOPOTENTIAL

To obtain solutions of the prolongation structure we shall use an approach based on the sub-algebras which have appeared. First, since the vector X_2 must commute with the entire structure, it generates a single independent conservation law, whose potential, r, need never appear in any other Pfaffians. Taking $X_2 = \partial/\partial r$, we obtain from Eqs. (3) and (4)

$$\omega_0 = dr + \tfrac{1}{2}\psi\bar{\psi}\,dx - (i/2)(\xi\bar{\psi} - \bar{\xi}\psi)\,dt, \tag{16}$$

which immediately gives the usual conservation law for probability density. Henceforth we may set $X_2 = 0$.

Additional solutions are easily obtained by exploiting the well-known representations for the rotation algebra of Eq. (9). It will be useful to write these in terms of two complex coordinates, y and z with associated basis vectors, $b_1 = \partial/\partial y$, $b_2 = \partial/\partial z$. A simple representation satisfying the structure equations then is

$$X_1 = 2(kyb_1 + \bar{k}\bar{y}\bar{b}_1),$$
$$Y_1 = \epsilon(yb_1 - \bar{y}\bar{b}_1 - \tfrac{1}{2}b_2 + \tfrac{1}{2}\bar{b}_2),$$
$$Y_2 = -2(k^2yb_1 - \bar{k}^2\bar{y}\bar{b}_1), \tag{17}$$
$$Z_1 = -\tfrac{1}{2}(y^2b_1 - \epsilon\bar{b}_1 - y_1b_2),$$
$$Z_2 = -(ky^2b_1 + \epsilon\bar{k}\bar{b}_1 - kyb_2),$$

where k is an arbitrary complex constant. So for each choice of k we have a solution.

The corresponding Pfaffians are

$$\omega^1 = dy + \tfrac{1}{2}[y^2\psi - \epsilon\bar{\psi} + 2ky](dx + ik\,dt)$$
$$- (i/2)[\xi y^2 + \epsilon\bar{\xi} + \epsilon\psi\bar{\psi}y]\,dt, \tag{18a}$$

$$\omega^2 = dz - \tfrac{1}{2}\psi y(dx + ik\,dt) + (i/2)(\xi y + \tfrac{1}{2}\epsilon\psi\bar{\psi})\,dt, \tag{18b}$$

together with their complex conjugates. We see from ω^1 that y is a pseudopotential depending only on the primitive variables, and from ω^2 that z is a potential (depending, however, on y). The equation

$$y_{,x} = -\tfrac{1}{2}(y^2\psi - \epsilon\bar{\psi} + 2ky) \tag{19}$$

is linearizable with the substitution

$$y = (2/\psi)(u_x/u) \tag{20}$$

giving

$$u_{xx} - (\psi_x/\psi - k)u_x - \tfrac{1}{4}\epsilon\psi\bar{\psi}u = 0, \tag{21}$$

and the equation

$$z_{,x} = \tfrac{1}{2}\psi y = u_x/u \tag{22}$$

suggests the new variables

$$z = \ln u, \quad y = v/u. \tag{23}$$

The well-known linear inverse scattering forms result:

$$\omega^3 \equiv u\omega^2 = du - \tfrac{1}{2}\psi v(dx + ik\,dt) + (i/2)(\xi v + \tfrac{1}{2}\epsilon\psi\bar{\psi}u)\,dt, \tag{24a}$$

$$\omega^4 \equiv u\omega^1 + v\omega^2 = dv - \tfrac{1}{2}[\epsilon\bar{\psi}u - 2kv](dx + ik\,dt)$$
$$- i\epsilon/2[\bar{\xi}u + \tfrac{1}{2}\psi\bar{\psi}v]\,dt. \tag{24b}$$

These could, of course, have been found directly from the prolongation structure by searching for a 2×2 matrix representation. The 1-parameter family of nonlinear pseudopotentials y, each of which occurs by itself in a 1-form (18a), seems, however, to be more closely tied to other methods of solution.

IV. TRANSFORMATIONS OF THE PROLONGED IDEAL

We will search for Bäcklund transformations as discrete transformations of the ideal (2), as prolonged with a set of pseudopotentials y_1, y_2, etc., characterized, respectively, by parameters k_1, k_2, etc., and defined by Pfaffians as in (18a) and its conjugate (we lower the labels on y_i, k_i to remind us of this specialization),

$$\omega_i \equiv dy_i + \tfrac{1}{2}[y_i^2\psi - \epsilon\bar{\psi} + 2k_iy_i](dx + ik_i\,dt)$$
$$- (i/2)[\xi y_i^2 + \epsilon\bar{\xi} + \epsilon\psi\bar{\psi}y_i]\,dt. \tag{25}$$

That is, we write

$$\alpha_1' = d\psi' \wedge dt + \xi'\,dx \wedge dt \tag{26}$$

$$\alpha_2' = id\psi' \wedge dx + d\xi' \wedge dt + \tfrac{1}{2}\epsilon\bar{\psi}'\psi'^2\,dx \wedge dt, \tag{27}$$

take $\psi' = \psi'(\psi, \bar{\psi}, \xi, \bar{\xi}, y_i, \bar{y}_i)$, and require α_1', α_2' and their conjugates to belong to the ideal of α_1, α_2, ω_i and their conjugates.

From (26) one finds ξ'; entering (27) with all the information from the ideal, one with some labor derives a single equation with unknown functions of the y_i as coefficients of polynomials in ξ, $\bar{\xi}$, $\psi^2\bar{\psi}$, ψ^2, $\bar{\psi}^2$, $\psi\bar{\psi}$, ψ, $\bar{\psi}$, and 1. These must independently vanish, and so present an overdetermined set to which we cannot be sure a nontrivial solution exists; i.e., we are left with what we might call a nonlinear Lie problem of integration in the space y_i. To be completely explicit, we find that

$$\psi' = \psi + S(y_i, \bar{y}_i), \tag{28}$$

where S must satisfy six derivation equations:

$$\underset{\bar{Z}_1}{\pounds} S = 0, \tag{29}$$

$$S + \epsilon \underset{Y_1}{\pounds} S = 0, \tag{30}$$

$$\epsilon S^2 - 2\underset{\bar{Z}_2}{\pounds} S = 0, \tag{31}$$

$$\epsilon S\bar{S} + \mathcal{L}_{X_1}\mathcal{L}_{Z_1} S = 0, \tag{32}$$

$$\epsilon\bar{S} - 2\mathcal{L}_{Z_1}\mathcal{L}_{Z_1} S = 0, \tag{33}$$

$$-\tfrac{1}{2}\mathcal{L}_{X_1}\mathcal{L}_{X_1} S + \mathcal{L}_{Y_2} S + \epsilon S^2\bar{S} = 0. \tag{34}$$

The vectors X_1, Y_1, Z_1, Y_2, Z_2 are those previously introduced, specialized in that only pseudopotentials y_i are now present. From ω_i, Eq. (25), we can read off their $(y_1, \bar{y}_1, y_2, \bar{y}_2, \cdots)$ components:

$$\begin{aligned}
X_1 &= (2k_1 y_1, 2\bar{k}_1\bar{y}_1, \cdots) = \bar{X}_1,\\
Y_1 &= (\epsilon y_1, -\epsilon\bar{y}_1, \cdots) = -\bar{Y}_1,\\
Z_1 &= (-\tfrac{1}{2}y_1^2, \tfrac{1}{2}\epsilon, \cdots),\\
\bar{Z}_1 &= (\tfrac{1}{2}\epsilon, -\tfrac{1}{2}\bar{y}_1^2, \cdots),\\
Y_2 &= (-2k_1^2 y_1, 2\bar{k}_1^2\bar{y}_1, \cdots) = -\bar{Y}_2\\
Z_2 &= (-k_1 y_1^2, -\epsilon\bar{k}_1, \cdots),\\
\bar{Z}_2 &= (-\epsilon k_1, -\bar{k}_1\bar{y}_1^2, \cdots).
\end{aligned} \tag{35}$$

Then

$$\bar{Z}_3 = (+2\epsilon k_1^2, -2\bar{k}_1^2\bar{y}_1^2, \cdots), \tag{36}$$

etc.

Other relations in such a Lie problem are deduced by repeated derivation and the use of the Jacobi identity. For example, operating on (33) with $\bar{Z}_1$, using the properties of the Lie derivative and (9), we get

$$\begin{aligned}
0 &= \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} - 2\mathcal{L}_{\bar{Z}_1}\mathcal{L}_{Z_1}\mathcal{L}_{Z_1} S\\
&= \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} - 2\mathcal{L}_{[\bar{Z}_1, Z_1]}\mathcal{L}_{Z_1} S - 2\mathcal{L}_{Z_1}\mathcal{L}_{\bar{Z}_1}\mathcal{L}_{Z_1} S\\
&= \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} + \mathcal{L}_{Y_1}\mathcal{L}_{Z_1} S - 2\mathcal{L}_{Z_1}\mathcal{L}_{[\bar{Z}_1, Z_1]} S - 2\mathcal{L}_{Z_1}\mathcal{L}_{Z_1}\mathcal{L}_{\bar{Z}_1} S\\
&= \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} + \mathcal{L}_{[Y_1, Z_1]} S + 2\mathcal{L}_{Z_1}\mathcal{L}_{Y_1} S - 2\mathcal{L}_{Z_1}\mathcal{L}_{Z_1}\mathcal{L}_{\bar{Z}_1} S\\
&= \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} + \epsilon\mathcal{L}_{Z_1} S + 2\mathcal{L}_{Z_1}\mathcal{L}_{Y_1} S - 2\mathcal{L}_{Z_1}\mathcal{L}_{Z_1}\mathcal{L}_{\bar{Z}_1} S,
\end{aligned}$$

then using (29) and (30),

$$0 = \epsilon\mathcal{L}_{\bar{Z}_1}\bar{S} + \epsilon\mathcal{L}_{Z_1} S - 2\epsilon\mathcal{L}_{Z_1} S$$

so finally

$$\mathcal{L}_{\bar{Z}_1}\bar{S} = \mathcal{L}_{Z_1} S. \tag{37}$$

This in turn implies (33), so (33) can be dropped.

By operating on (34) with $\mathcal{L}_{\bar{Z}_1}$ and using (31), one similarly shows that

$$-\mathcal{L}_{\bar{Z}_3} S + \epsilon S\mathcal{L}_{X_1} S + \epsilon S^2\mathcal{L}_{Z_1} S = 0. \tag{38}$$

We have been able to proceed farther only in the cases of one pseudopotential and two. Integration of these, however, is sufficient to give respectively the Bäcklund transformation and the theorem of permutability. The Bäcklund transformation acts, at least in the case $\epsilon = -1$, to add a soliton to any given solution; the theorem of permutability can be used to generate algebraically a ladder of solutions with added solitons, once the Bäcklund transformation of the originally given solution has been integrated to achieve the first step up the ladder.

V. BÄCKLUND TRANSFORMATIONS

Using only one pseudopotential y, from (29), (30), and (31), one readily integrates for $S(y, \bar{y})$, and can check that (32), etc., are satisfied. The result is[11,2]

$$\psi' = \psi - 2(k + \bar{k})\bar{y}/(y\bar{y} - \epsilon). \tag{39}$$

The simplest Bäcklund transformations of a given solution $\psi(x, t)$, and $\xi = -\psi_x$, thus is accomplished by integrating the (guaranteed integrable!) partial differential equations resulting from one Pfaffian ω,

$$\begin{aligned}
y_x &= -\tfrac{1}{2}(y^2\psi - \epsilon\bar{\psi} + 2ky),\\
y_t &= -ik\tfrac{1}{2}(y^2\psi - \epsilon\bar{\psi} + 2ky) + (i/2)(\xi y^2 + \epsilon\bar{\xi} + \epsilon\psi\bar{\psi}y),
\end{aligned} \tag{40}$$

for the pseudopotential $y(x, t)$, and then substituting into (39).

VI. THE THEOREM OF PERMUTABILITY

To integrate the case of two pseudopotentials, y_1 and y_2, we first introduce variables $u_1 = \bar{y}_1^{-1} - \epsilon y_1$, $u_2 = \bar{y}_2^{-1} - \epsilon y_2$, and $u_3 = y_1 - y_2$. From (35),

$$\mathcal{L}_{\bar{Z}_1} u_1 = \mathcal{L}_{\bar{Z}_1} u_2 = \mathcal{L}_{\bar{Z}_1} u_3 = 0,$$

so from (29) S must be a function of only these three. Moreover,

$$\mathcal{L}_{Y_1} u_1 = \epsilon u_1, \quad \mathcal{L}_{Y_1} u_2 = \epsilon u_2, \quad \mathcal{L}_{Y_1} u_3 = \epsilon u_3,$$

so from (30) S^{-1} must be homogeneous of first order in them.

Now

$$\mathcal{L}_{\bar{Z}_2} u_1 = k_1 + \bar{k}_1, \quad \mathcal{L}_{\bar{Z}_2} u_2 = k_2 + \bar{k}_2, \quad \mathcal{L}_{\bar{Z}_2} u_3 = -\epsilon(k_1 - k_2),$$

and it can next be shown that (31) is satisfied if

$$S^{-1} = u_3/2(k_1 - k_2) + (u_1 + \gamma_1 u_3)^{1/2}(u_2 + \gamma_2 u_3)^{1/2} G(x), \tag{41}$$

where

$$\gamma_1 = \epsilon(k_1 + \bar{k}_1)/(k_1 - k_2), \quad \gamma_2 = \epsilon(k_2 + \bar{k}_2)/(k_1 - k_2),$$

and we take

$$x^2 = (u_1 + \gamma_1 u_3)/(u_2 + \gamma_2 u_3), \quad \mathcal{L}_{\bar{Z}_2} x = 0.$$

The only remaining first order Lie equation we have found, (38), is then an ordinary differential equation for $G(x)$; after considerable manipulation it reduces to

$$\begin{aligned}
&xG'(x)[-\{(k_1 + \bar{k}_2)(k_2 + \bar{k}_2)x - (k_1 + \bar{k}_1)(k_2 + \bar{k}_1)x^{-1}\}G^2\\
&\quad + \epsilon(\bar{k}_1 - \bar{k}_2)G + \tfrac{1}{4}(x - x^{-1})]\\
&\quad + G^3[(k_2 + \bar{k}_2)(k_1 + \bar{k}_2)x + (k_1 + \bar{k}_1)(\bar{k}_1 + k_2)x^{-1}]\\
&\quad + \epsilon G^2(k_1 + \bar{k}_1 + k_2 + \bar{k}_2) + \tfrac{1}{4}G(x + x^{-1}) = 0,
\end{aligned} \tag{42}$$

where $G'(x)$ is dG/dx. The general solution of this is

$$(G-\alpha_1 x)(G-\alpha_2 x^{-1})=\lambda(G-\beta_1 x)(G-\beta_2 x^{-1}), \tag{43}$$

where

$$\begin{aligned}4\alpha_1&=(1-\epsilon)/(k_2+\bar{k}_1)-(1+\epsilon)/(k_1+\bar{k}_1),\\4\beta_1&=(1-\epsilon)/(k_1+\bar{k}_1)-(1+\epsilon)/(k_2+\bar{k}_1),\\4\alpha_2&=(1-\epsilon)/(k_1+\bar{k}_2)-(1+\epsilon)/(k_2+\bar{k}_2),\\4\beta_2&=(1-\epsilon)/(k_2+\bar{k}_2)-(1+\epsilon)/(k_1+\bar{k}_2),\end{aligned} \tag{44}$$

and λ is the constant of integration.

We have verified by direct substitution, using the very convenient MACSYMA symbolic manipulation program of the MIT Mathlab Group,[12] that the remaining equations (32), (33), and (34) are all satisfied by the single choice $\lambda=1$, which reduces (43) to a linear result for G,

$$2G=\epsilon(\bar{k}_1-\bar{k}_2)/[(k_2+\bar{k}_2)(k_1+\bar{k}_2)x-(k_1+\bar{k}_1)(k_2+\bar{k}_1)x^{-1}]. \tag{45}$$

Substituting into (41), and converting to the original pseudopotential variables gives finally a reasonable form for the double Bäcklund transformation

$$\psi'=\psi+2Q/R, \tag{46}$$

where

$$\begin{aligned}Q=\;&\epsilon(k_1+\bar{k}_1)(k_2+\bar{k}_2)(k_1-k_2-\bar{k}_1+\bar{k}_2)(y_1-y_2)\bar{y}_1\bar{y}_2\\&+(k_2+\bar{k}_2)(k_1+\bar{k}_2)(k_1-k_2)(1-\epsilon y_1\bar{y}_1)\bar{y}_2\\&-(k_1+\bar{k}_1)(k_2+\bar{k}_1)(k_1-k_2)(1-\epsilon y_2\bar{y}_2)\bar{y}_1,\end{aligned} \tag{47}$$

and

$$\begin{aligned}R=&-(k_1+\bar{k}_1)(k_2+\bar{k}_2)(y_1-y_2)(\bar{y}_1-\bar{y}_2)\\&+\epsilon(\bar{k}_1-\bar{k}_2)(k_1-k_2)(1-\epsilon y_1\bar{y}_1)(1-\epsilon y_2\bar{y}_2).\end{aligned} \tag{48}$$

If k_1, say, $\rightarrow 0$, this becomes (39).

APPENDIX

To investigate the question of the uniqueness of the KdV equation with respect to the soliton phenomenon, we may consider the prolongation structures of generalized KdV-like equations

$$u_t+u_{xxx}+f(u)u_x=0, \tag{A1}$$

where $f(u)$ is an arbitrary function of u. The result for this equation, analogous to Eq. (32) of Ref. (1), is

$$\begin{aligned}F^k&=2X_1^k+2uX_2^k+3u^2X_3^k,\\G^k&=-2(p+g')X_2^k+3[z^2-2u(p+g')+2g]X_3^k\\&\quad+8X_4^k+8uX_5^k+4u^2X_6^k+4zX_7^k,\end{aligned} \tag{A2}$$

where the function $g(u)$ is defined by

$$g''=\frac{d^2g(u)}{du^2}\equiv f(u). \tag{A3}$$

Most of the structure equations of Eq. (37) of Ref. (1) for the X_m are unchanged. In the course of the derivation, however, one encounters the equation

$$\left(\frac{d^5g}{du^5}\right)X_7=0. \tag{A4}$$

There is no constraint on $g(u)$ [or $f(u)$] if we take $X_7=0$. However, again from Eq. (37), this implies $X_5=X_6=0$, and the algebra of the remaining four generators is complete and Abelian. No pseudopotentials therefore can exist in this case, and only the well-known conservation laws result.

If $X_7\neq 0$, $g(u)$ must be a quartic, but it is easily shown that the constant and linear terms [which do not affect $f(u)$] can be dropped without loss of generality. Thus, taking

$$g(u)=\alpha u^2+\beta u^3+\gamma u^4, \tag{A5}$$

we have

$$f(u)=2\alpha+6\beta u+12\gamma u^2, \tag{A6}$$

demonstrating that the combined KdV-modified KdV equation is the only example of equations of type (A1) admitting pseudopotentials. For general coefficients $\{\alpha,\beta,\gamma\}$ we find three modified structure equations

$$\begin{aligned}&[X_1,X_5]+[X_2,X_4]+\tfrac{1}{2}\alpha X_7=0,\\&[X_3,X_4]+[X_1,X_6]+\tfrac{1}{2}\beta X_7=0,\\&[X_2,X_6]=-2\gamma X_7,\end{aligned} \tag{A7}$$

and it follows from these that

$$[X_6,X_7]=\tfrac{1}{2}\beta X_6-2\gamma X_5. \tag{A8}$$

All other relations in Eq. (37) and Eq. (38) of Ref. (1) are unchanged.

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS7-100, sponsored by the National Aeronautics and Space Administration.

[1]H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1–7 (1975).
[2]J. Corones, J. Math. Phys. **17**, 756–59 (1976).
[3]H.C. Morris, "Prolongation Structures and a Generalized Inverse Scattering Problem," preprint TCD 1975-6, Trinity College, Dublin, J. Math. Phys. **17** (1976), to be published.
[4]H.C. Morris, "Prolongation Structures and Nonlinear Evolution Equations in Two Spatial Dimensions," preprint TCD 1975-10, Trinity College, Dublin, J. Math. Phys. **17** (1976), to be published; "Prolongation Structures and Nonlinear Evolution Equations in Two Spatial Dimensions II: A Generalized Nonlinear Schrödinger Equation," report TCD 1976-3, Trinity College, Dublin.
[5]G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[6]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[7]V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP, **34**, 62 (1972)].
[8]R. Hirota, J. Math. Phys. **14**, 805 (1973).
[9]A. Hasegawa and F. Tappert, Appl. Phys. Lett. **23**, 171 (1973).
[10]R. Hirota, "Direct Method of Finding Exact Solutions of Nonlinear Evolution Equations," (preprint).
[11]G.L. Lamb, Jr., J. Math. Phys. **15**, 2157 (1974).
[12]The Mathlab Group is supported by the Advanced Research Projects Agency (ARPA), under Office of Naval Research Contract N00014-70-A-0362-0006.

N-level system in contact with a singular reservoir*

Vittorio Gorini

Istituto di Fisica dell'Università, 20133 Milano, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Italy

Andrzej Kossakowski

Institute of Physics, N. Copernicus University, 87100 Toruń, Poland
(Received 19 March 1975)

We study a model of an N-level atom coupled linearly to an infinite free boson bath whose time correlation functions are Gaussian. We prove that, in the limit when the decay time of the correlations of the bath goes to zero, the reduced dynamics of the atom is given by a completely positive dynamical semigroup. By varying the interaction parameters, any such semigroup can be obtained in the limit. We also discuss the formally analogous situation of an N-level atom whose Hamiltonian contains an external fluctuating Gaussian stationary contribution.

I. INTRODUCTION

In a preceding paper,[1] hereafter referred to as I, we have proved that a linear operator $L:M(N)\to M(N)$ is the generator of a completely positive dynamical semigroup of $M(N)$ in the Schrödinger picture iff it can be written in the form

$$L:\rho\to L\rho=-i[H,\rho]+\frac{1}{2}\sum_{i,j=1}^{N^2-1}c_{ij}\{[F_i,\rho F_j^*]+[F_i\rho,F_j^*]\},\quad \rho\in M(N), \tag{1.1}$$

where $H=H^*$; $\mathrm{Tr}(H)=0$, $\mathrm{Tr}(F_i)=0$, $\mathrm{Tr}(F_i^*F_j)=\delta_{ij}$ $(i,j=1,2,\ldots,N^2-1)$ and $\{c_{ij}\}$ is a complex positive matrix. For a given L, H is uniquely determined by the condition $\mathrm{Tr}(H)=0$ and $\{c_{ij}\}$ is uniquely determined by the choice of the F_i's.[2] By the duality relation $\mathrm{Tr}[(L\rho)A]=\mathrm{Tr}[\rho(L^*A)]$, the generator of the corresponding Heisenberg dynamical semigroup has the form

$$L^*:A\to L^*A=i[H,A]+\frac{1}{2}\sum_{i,j=1}^{N^2-1}c_{ij}\{[F_j^*,A]F_i+F_j^*[A,F_i]\},\quad A\in M(N). \tag{1.2}$$

The condition of positivity of the matrix $\{c_{ij}\}$ expresses implicitly N^2-1 inequalities which have to be satisfied by the physical parameters which characterize the dynamical evolution (such as f.i. relaxation times and components of equilibrium states) if the latter is to be completely positive. As an application, these inequalities were explicitly worked out in I for the case of a two-level system. In this paper, we study a model of an N-level atom S coupled to a boson reservoir R initially in the vacuum state and whose time correlation functions are Gaussian (which in particular implies the one-particle energy operator to be unbounded from below as well as from above). We derive an explicit formula for the reduced Heisenberg dynamics $t\to\mu_t$ of the system and then show that in the limit interaction when the two-point time correlation functions of R become δ functions, $t\to\mu_t$ goes over into a completely positive Heisenberg dynamical semigroup. In this limit, the meaning of the coefficients $\{c_{ij}\}$ becomes clear: they are directly related to the strength of the interaction and appear as coefficients of the two-point time correlation functions of the reservoir operators occurring in the expansion of the interaction Hamiltonian over the matrices F_i. In this connection, any completely positive dynamical semigroup can be obtained by suitably changing the interaction parameters, since the matrix $\{c_{ij}\}$ need not satisfy any restriction beyond positivity. Our model is similar to one considered by Hepp and Lieb.[3] These authors however make use of fermion reservoirs. We develop the model in Sec. II. In Sec. III we briefly discuss the case, which can be treated formally in an analogous way, of an N-level system whose Hamiltonian contains a stationary stochastic contribution. An example is provided by the motion of a spin magnetic moment in an external fluctuating magnetic field.[4] Without restriction to the Hilbert space being finite dimensional, the general problem has also been recently studied in some detail by Fox.[5] For earlier papers on systems with stochastic Hamiltonians see Refs. 6 and 7. For an extension of our model to an N-level system coupled to an infinite quasifree Fermi or Bose reservoir at an arbitrary temperature see Ref. 8.

II. MODEL

We describe our model reservoir by the Fock representation on Fock space $\mathcal{H}'$ of N^2 independent Bose fields over the test function space $\mathcal{S}(\mathbf{R})$ of complex functions of rapid decrease on the real line,

$$[a_\alpha(f),a_\beta(g)]=0,\ [a_\alpha(f),a_\beta^*(g)]=\delta_{\alpha\beta}(f,g),\quad \alpha,\beta=1,2,\ldots,N^2;\ (f,g)=\int_{-\infty}^{\infty}\overline{f(\omega)}g(\omega)\,d\omega, \tag{2.1}$$

with the free Heisenberg evolution defined by

$$t\to\alpha_t^R:a_\alpha(f)\to a_\alpha(f_{-t}),\quad t\geq 0, \tag{2.2}$$

where

$$f_t(\omega)=\exp(-i\omega t)f(\omega) \tag{2.3}$$

The infinitesimal generator of $t\to\alpha_t^R$ is formally given by $i[H_R,\cdot\,]$, where

$$H_R=\sum_{\alpha=1}^{N^2}\int_{-\infty}^{\infty}\omega a_\alpha^*(\omega)a_\alpha(\omega)\,d\omega, \tag{2.4}$$

$a(\omega)$ being the improper fields, $a(f)=\int_{-\infty}^{\infty}\overline{f(\omega)}a(\omega)\,d\omega$.

The initial state of the reservoir is assumed to be the Fock vacuum Ω, $a(f)\Omega = 0\ \forall f \in \mathcal{S}(\mathbb{R})$. The Hilbert space of the system S is the N-dimensional unitary space $\mathcal{H} = \mathbb{C}^N$. Let H_S denote the system's Hamiltonian and let $\{F_\alpha\}_{\alpha=1,2,\ldots,N^2}$ be a complete orthonormal set (c.o.s.) of $N\times N$ self-adjoint matrices, $F_\alpha = F_\alpha^*$, $(F_\alpha, F_\beta) = \delta_{\alpha\beta}$ $(\alpha, \beta = 1, 2, \ldots, N^2)$, with $F_{N^2} = (1/\sqrt{N})\mathbb{1}$. We assume R to be linearly coupled to S by the following interaction Hamiltonian

$$V = \sum_{\alpha=1}^{N^2} F_\alpha \otimes \varphi_\alpha^\epsilon, \tag{2.5}$$

with

$$\varphi_\alpha^\epsilon = \frac{1}{(2\pi)^{1/2}} \sum_{\beta=1}^{N^2} [\bar{\mu}_\beta \bar{f}_\alpha^\beta a_\beta(f^\epsilon) + \mu_\beta f_\alpha^\beta a_\beta^*(f^\epsilon)], \tag{2.6}$$

where

$$f^\epsilon(\omega) = \exp(-\epsilon^2\omega^2/8), \quad \epsilon > 0 \tag{2.7}$$

and

$$\sum_{\alpha=1}^{N^2} \bar{f}_\alpha^\mu f_\alpha^\nu = \delta_{\mu\nu}.$$

Set $\varphi_\alpha^\epsilon(t) = \alpha_{-t}^R \varphi_\alpha^\epsilon$. Then we have

$$(\Omega, \varphi_{\alpha_1}^\epsilon(t_1) \cdots \varphi_{\alpha_{2n+1}}^\epsilon(t_{2n+1})\Omega) = 0, \quad n = 0, 1, 2, \ldots \tag{2.8}$$

and

$$\begin{aligned}(\Omega, \varphi_{\alpha_1}^\epsilon(t_1) \cdots \varphi_{\alpha_{2n}}^\epsilon(t_{2n})\Omega) &= \sum_{p\in\mathcal{P}_n} \prod_{r=1}^{n} (\Omega, \varphi_{\alpha_{p(2r-1)}}^\epsilon(t_{p(2r-1)}) \\ &\quad \times \varphi_{\alpha_{p(2r)}}^\epsilon(t_{p(2r)})\,\Omega), \quad n = 1, 2, \ldots,\end{aligned} \tag{2.9}$$

where the summation is extended to the set $\mathcal{P}_n$ of all permutations p of $(1, 2, \ldots, 2n)$ such that $p(2r-1) < p(2r)$ and $p(2r-1) < p(2r+1)$. The two-point time correlation function is given by

$$\begin{aligned}(\Omega, \varphi_\alpha^\epsilon(t)\varphi_\beta^\epsilon(s)\Omega) &= \frac{1}{2\pi} \sum_{\gamma,\delta=1}^{N^2} \bar{\mu}_\gamma \mu_\delta \bar{f}_\alpha^\gamma f_\beta^\delta (\Omega, a_\gamma(f_t^\epsilon) a_\delta^*(f_s^\epsilon)\Omega) \\ &= \frac{1}{2\pi}\left(\sum_{\gamma=1}^{N^2} |\mu_\gamma|^2 \bar{f}_\alpha^\gamma f_\beta^\gamma\right)(f_t^\epsilon, f_s^\epsilon) = c_{\beta\alpha}\,\delta_\epsilon(t-s),\end{aligned} \tag{2.10}$$

where

$$c_{\beta\alpha} = \sum_{\gamma=1}^{N^2} |\mu_\gamma|^2 \bar{f}_\alpha^\gamma f_\beta^\gamma \tag{2.11}$$

and

$$\delta_\epsilon(t-s) = (1/\epsilon\sqrt{\pi}) \exp[-(t-s)^2/\epsilon^2]. \tag{2.12}$$

Let $t \to \alpha_t^\epsilon$, $t \geq 0$ be the Heisenberg evolution of the total system $S+R$ in the space $\mathcal{H}\otimes\mathcal{H}'$. Denote by $t \to \alpha_t^S$, $t \geq 0$, the free Heisenberg evolution of S in $\mathcal{H}$ (whose generator is $i[H_S, \circ])$ and define $\beta_t^\epsilon = (\alpha_{-t}^S \otimes \mathbb{1})(\mathbb{1} \otimes \alpha_{-t}^R)\alpha_t^\epsilon$. It is convenient to study the reduced dynamics of the system S in an interaction picture corresponding to the free evolution of S. If we take into account the invariance of Ω under the free evolution of R, this is defined by

$$(x_t, (\gamma_t^\epsilon A)y_t) = (x_t \otimes \Omega, [\beta_t^\epsilon(A\otimes\mathbb{1})]y_t \otimes \Omega), \quad A \in M(N), \quad t \geq 0, \tag{2.13}$$

where

$$x_t = \exp(-iH_S t)x, \quad x \in \mathcal{H} \quad \text{(same for } y_t) \tag{2.14}$$

and $\beta_t^\epsilon(A\otimes\mathbb{1})$ satisfies the following differential equation:

$$\begin{aligned}\frac{d\beta_t^\epsilon(A\otimes\mathbb{1})}{dt} &= i[(\alpha_{-t}^S \otimes \mathbb{1})(\mathbb{1}\otimes\alpha_{-t}^R)V, \beta_t^\epsilon(A\otimes\mathbb{1})] \\ &= i\sum_{\alpha=1}^{N^2} [F_\alpha(t)\otimes\varphi_\alpha^\epsilon(t), \beta_t^\epsilon(A\otimes\mathbb{1})],\end{aligned} \tag{2.15}$$

where

$$F_\alpha(t) = \exp(-iH_S t) F_\alpha \exp(iH_S t). \tag{2.16}$$

Introducing the Liouvillian

$$M^\epsilon(t) = i\sum_{\alpha=1}^{N^2} [F_\alpha(t)\otimes\varphi_\alpha^\epsilon(t), \circ], \tag{2.17}$$

β_t^ϵ is formally given by the Dyson series

$$\begin{aligned}\beta_t^\epsilon &= T\exp\left(\int_0^t ds\, M^\epsilon(s)\right) \\ &= \mathbb{1} + \sum_{n=1}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n M^\epsilon(t_1)M^\epsilon(t_2)\cdots M^\epsilon(t_n), \\ &\qquad t \geq 0.\end{aligned} \tag{2.18}$$

More generally, we define

$$\beta_{t,s}^\epsilon = T\exp\left(\int_s^t du\, M^\epsilon(u)\right), \quad t \geq s \geq 0. \tag{2.19}$$

By (2.13), the reduced dynamics γ_t^ϵ is obtained by averaging (2.18) over Ω,

$$\begin{aligned}\gamma_t^\epsilon A &= \langle\beta_t^\epsilon(A\otimes\mathbb{1})\rangle \\ &= A + \sum_{n=1}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n} \\ &\quad \times \langle M^\epsilon(t_1)M^\epsilon(t_2)\cdots M^\epsilon(t_{2n})A\otimes\mathbb{1}\rangle, \quad t \geq 0,\end{aligned} \tag{2.20}$$

where we have employed the notation $\langle\cdots\rangle = (\Omega, \ldots\Omega)$. Owing to (2.8), there appear in (2.20) only terms containing products of an even number of Liouvillians.

We now proceed to obtain a simple formal derivation of a compact formula for the time derivative of (2.20), from which we will obtain a description of the reduced dynamics in terms of a dynamical semigroups in the limit $\epsilon \downarrow 0$. A rigorous but lengthy proof of this result is given in Appendices A and B, where we show that for any given $t \geq 0$ the averaged Dyson series (2.20) is uniformly convergent on any finite interval $0 \leq \epsilon \leq \epsilon_0$ (Appendix A) and that the semigroup dynamics is rigorously attained as $\epsilon \downarrow 0$ (Appendix B). We need the following formula, which is proved in Appendix C:

$$\begin{aligned}\langle M^\epsilon(t)M^\epsilon(t_1)\cdots M^\epsilon(t_{2n-1})(A\otimes\mathbb{1})\rangle &= i\sum_{j=1}^{2n-1} \{\langle\varphi_\alpha^\epsilon(t)\varphi_{\alpha_j}^\epsilon(t_j)\rangle F_\alpha(t)\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j)M^\epsilon(t_{j+1})\cdots M^\epsilon(t_{2n-1})(A\otimes\mathbb{1})\rangle \\ &\quad - \langle\varphi_{\alpha_j}^\epsilon(t_j)\varphi_\alpha^\epsilon(t)\rangle\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j)M^\epsilon(t_{j+1})\cdots M^\epsilon(t_{2n-1})(A\otimes\mathbb{1})\rangle F_\alpha(t)\}, \quad t \geq t_1 \geq \cdots \geq t_{2n-1},\end{aligned} \tag{2.21}$$

where summation from 1 to N^2 over repeated indices α_j is understood and where

$$\mathcal{L}_\alpha(t) := i[F_\alpha(t)\otimes \mathbb{1}, \cdot]. \tag{2.22}$$

From (2.20) we get

$$\frac{d(\gamma_t^\epsilon A)}{dt} = \sum_{n=1}^{\infty}\int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-2}} dt_{2n-1} \langle M^\epsilon(t) M^\epsilon(t_1)\cdots M^\epsilon(t_{2n-1})(A\otimes\mathbb{1})\rangle, \tag{2.23}$$

Then, introducing the notation

$$I_n^\epsilon(t,t_0) = \int_{t_0}^t dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n-1}} dt_n M^\epsilon(t_1) M^\epsilon(t_2)\cdots M^\epsilon(t_n), \tag{2.24}$$

and substituting (2.21) into (2.23), we obtain

$$\frac{d(\gamma_t^\epsilon A)}{dt} = i\sum_{n=1}^{\infty}\sum_{j=1}^{2n-1}\int_0^t dt_1 \cdots \int_0^{t_{j-1}} dt_j \{\langle \varphi_\alpha^\epsilon(t)\varphi_{\alpha_j}^\epsilon(t_j)\rangle F_\alpha(t)\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle$$

$$- \langle \varphi_{\alpha_j}^\epsilon(t_j)\varphi_\alpha^\epsilon(t)\rangle\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle F_\alpha(t)\}$$

$$= i\sum_{n=1}^{\infty}\sum_{j=1}^{2n-1}\int_0^t dt_j \int_{t_j}^t dt_1 \int_{t_j}^{t_1} dt_2 \cdots \int_{t_j}^{t_{j-2}} dt_{j-1} \{\langle \varphi_\alpha^\epsilon(t)\varphi_{\alpha_j}^\epsilon(t_j)\rangle F_\alpha(t)\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle$$

$$- \langle \varphi_{\alpha_j}^\epsilon(t_j)\varphi_\alpha^\epsilon(t)\rangle\langle M^\epsilon(t_1)\cdots M^\epsilon(t_{j-1})\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle F_\alpha(t)$$

$$= i\sum_{n=1}^{\infty}\sum_{j=1}^{2n-1}\int_0^t dt_j \{\langle \varphi_\alpha^\epsilon(t)\varphi_{\alpha_j}^\epsilon(t_j)\rangle F_\alpha(t)\langle I_{j-1}^\epsilon(t,t_j)\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle$$

$$- \langle \varphi_{\alpha_j}^\epsilon(t_j)\varphi_\alpha^\epsilon(t)\rangle\langle I_{j-1}^\epsilon(t,t_j)\mathcal{L}_{\alpha_j}(t_j) I_{2n-1-j}^\epsilon(t_j,0)(A\otimes\mathbb{1})\rangle F_\alpha(t)\} = i\sum_{n=0}^{\infty}\sum_{j=0}^{2n}\int_0^t ds\, \{\langle \varphi_\alpha^\epsilon(t)\varphi_\beta^\epsilon(s)\rangle F_\alpha(t)$$

$$\times \langle I_j^\epsilon(t,s)\mathcal{L}_\beta(s) I_{2n-j}^\epsilon(s,0)(A\otimes\mathbb{1})\rangle - \langle \varphi_\beta^\epsilon(s)\varphi_\alpha^\epsilon(t)\rangle\langle I_j^\epsilon(t,s)\mathcal{L}_\beta(s) I_{2n-j}^\epsilon(s,0)(A\otimes\mathbb{1})\rangle F_\alpha(t)\}$$

$$= i\sum_{m,n=0}^{\infty}\int_0^t ds \{\langle \varphi_\alpha^\epsilon(t)\varphi_\beta^\epsilon(s)\rangle F_\alpha(t)\langle I_n^\epsilon(t,s)\mathcal{L}_\beta(s) I_m^\epsilon(s,0)(A\otimes\mathbb{1})\rangle$$

$$- \langle \varphi_\beta^\epsilon(s)\varphi_\alpha^\epsilon(t)\rangle\langle I_n^\epsilon(t,s)\mathcal{L}_\beta(s) I_m^\epsilon(s,0)(A\otimes\mathbb{1})F_\alpha(t)\}.$$

Hence, from (2.10), (2.18), (2.19), and (2.29) we get (now writing explicitly the summation over α and β)

$$\frac{d(\gamma_t^\epsilon A)}{dt} = i\sum_{\alpha,\beta=1}^{N^2} c_{\alpha\beta}\int_0^t ds\, \delta_\epsilon(t-s)\{F_\beta(t)\langle \beta_{t,s}^\epsilon \mathcal{L}_\alpha(s)\beta_s^\epsilon(A\otimes\mathbb{1})\rangle - \langle \beta_{t,s}^\epsilon \mathcal{L}_\beta(s)\beta_s^\epsilon(A\otimes\mathbb{1})\rangle F_\alpha(t)\} \tag{2.25}$$

Since $\lim_{\epsilon\downarrow 0}\delta_\epsilon(u) = \delta(u)$ by formally taking the limit as $\epsilon\downarrow 0$ and setting $\lim_{\epsilon\downarrow 0}\gamma_t^\epsilon A = \gamma_t A$, we finally obtain $d(\gamma_t A)/dt = S(t)(\gamma_t A)$, where

$$S(t) = \frac{1}{2}\sum_{\alpha,\beta=1}^{N^2} c_{\alpha\beta}\{[F_\beta(t),\cdot\,]F_\alpha(t) + F_\beta(t)[\,\cdot\,, F_\alpha(t)]\}. \tag{2.26}$$

Hence

$$\gamma_t A = A + \sum_{n=1}^{\infty}\int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \times S(t_1)S(t_2)\cdots S(t_n)A, \quad t \geq 0. \tag{2.27}$$

As stated above, rigorous proofs of the convergence of the series (2.20) and of the convergence of $\gamma_t^\epsilon A$ to (2.27) as $\epsilon\downarrow 0$ are given in Appendices A and B, respectively. The Heisenberg reduced dynamics is defined by $(x, (\mu_t^\epsilon A)y) = (x_t, (\gamma_t^\epsilon A)y_t)$. Hence, by (2.16), (2.27), and (2.26), the limit reduced dynamics $t \to \mu_t = \lim_{\epsilon\downarrow 0}\mu_t^\epsilon$, $t \geq 0$, is a completely positive dynamical semigroup $t \to \mu_t = \exp(Lt)$, whose generator is given by

$$L: A \to LA = i[H_S + \hat{H}, A] + \frac{1}{2}\sum_{i,j=1}^{N^2-1} c_{ij}\{[F_j, A]F_i + F_j[A, F_i]\}, \quad A\in M(N), \tag{2.28}$$

where

$$\hat{H} = -\frac{1}{\sqrt{N}}\sum_{l=1}^{N^2-1}(Im\, c_{lN^2})F_l. \tag{2.29}$$

Equation (2.28) has precisely the form (1.6) (recall that the F_i's have been chosen to be self-adjoint). The term $\hat{H}$ describes a shift of the energy levels produced by the coupling of the system to the reservoir. By varying the coefficients of the expansion (2.6) one can obtain, in the limit $\epsilon\downarrow 0$, any given completely positive dynamical semigroup, since one can fix at will the eigenvalues $|\mu_\gamma|^2$ and the corresponding eigenvectors f_α^ν of the matrix $\{c_{\alpha\beta}\}$. In our model, the interaction of the system with the reservoir takes place through a process of emission and absorption of Bose quanta and the eigenvalues of the matrix $\{c_{\alpha\beta}\}$ play the role of coupling constants.

III. N-LEVEL SYSTEM WITH FLUCTUATING HAMILTONIAN

In a way formally analogous to the above one can study the behavior of an N-level system on which the effect of the surroundings can be represented by the addition to the Hamiltonian of a stationary stochastic term. An example thereof which has been recently studied in detail is provided by the motion of a spin magnetic moment in an external fluctuating magnetic field.[4] We refer to other existing work[5-7] for general treatments and for the discussion of other physical applications, and confine ourselves here to a few remarks concerning the derivation of the average dynamics of the system. The Hamiltonian of the system is given by

$$H(t)=H_S+\tilde{H}(t), \tag{3.1}$$

where $\tilde{H}(t)$ is the external stochastic term. Expand $\tilde{H}(t)$ over a c.o.s. of self-adjoint matrices $\{F_\alpha\}_{\alpha=1,2,\ldots,N^2}$ with $F_{N^2}=(1/\sqrt{N})\mathbb{1}$:

$$\tilde{H}(t)=\sum_{\alpha=1}^{N^2} F_\alpha\varphi_\alpha(t). \tag{3.2}$$

We consider the situation when the external "classical fields" are Gaussian and stationary. In other words, their time correlation functions $\langle\varphi_{\alpha_1}(t_1)\cdots\varphi_{\alpha_n}(t_n)\rangle$ are specified by the generating functional

$$F(\chi)=\left\langle\exp\left(i\sum_{\alpha=1}^{N^2}\int_{-\infty}^{\infty}dt\,\chi_\alpha(t)\varphi_\alpha(t)\right)\right\rangle$$
$$=\left(-\frac{1}{2}\sum_{\alpha,\beta=1}^{N^2}c_{\alpha\beta}\int_{-\infty}^{\infty}dt\int_{-\infty}^{\infty}ds\,\chi_\alpha(t)\chi_\beta(s)\delta_\epsilon(t-s)\right), \tag{3.3}$$

where $\delta_\epsilon(t-s)$ is given by (2.12). By successively computing the functional derivatives of (3.3) at $t=0$ we obtain that the odd-point correlation functions vanish, whereas

$$\langle\varphi_\alpha(t)\varphi_\beta(s)\rangle=c_{\alpha\beta}\delta_\epsilon(t-s) \tag{3.4}$$

and the higher-order even-point correlation functions are expressed in terms of (3.4) by Wick's formula (2.9). The dynamics of the system is obtained by averaging (3.2) and can be computed exactly in the same way as in Sec. II. One arrives once more at formula (2.25) for the average dynamics in the interaction picture relative to the free evolution and, in the limit $\epsilon\downarrow 0$ (white noise), the average Heisenberg dynamics is a completely positive dynamical semigroup whose generator is given by (2.28). However, in the present case the positive matrix $\{c_{\alpha\beta}\}$ is symmetric. Therefore, $\hat{H}=0$ (no level shift) and the generator L is written

$$L: A\to LA=i[H_S,A]$$
$$-\frac{1}{2}\sum_{i,j=1}^{N^2-1}c_{ij}[F_i,[F_j,A]],\quad A\in M(N). \tag{3.5}$$

The corresponding generator of the Schrödinger dynamics is

$$L_*:\rho\to L_*\rho=-i[H_S,\rho]-\frac{1}{2}\sum_{i,j=1}^{N^2-1}c_{ij}[F_i,[F_j,\rho]], \tag{3.6}$$

As $L_*\mathbb{1}=0$, the central state $(1/N)\mathbb{1}$ is invariant under the evolution. This was to be expected, since the dynamics is an average of Hamiltonian dynamics each of which is centre preserving. In the case $N=2$, the commutation relations $[F_i,F_j]=i\sqrt{2}\sum_{l=1}^{3}\epsilon_{ijl}F_l$ $(i,j=1,2,3$; ϵ_{ijl} is the Levi-Civita symbol) show that the condition that the matrix $\{c_{ij}\}_{i,j=1,2,\ldots,N^2-1}$ be symmetric is also necessary for the central state to be an invariant state of the dynamical semigroup. On the other hand, the condition is only sufficient if $N>2$ and not all center-preserving dynamical semigroups can be obtained by the procedure above. This fact seems to indicate that, unless $N=2$, not all center-preserving completely positive dynamical maps[9] of an N-level system can be obtained as convex combinations of unitary transformations $\rho\to U\rho U^*$, $U^*U=\mathbb{1}$.

IV. CONCLUDING REMARKS

In this paper, we have constructed a simplified model of an N-level atom S in contact with an infinite free boson reservoir R, which allows for the derivation of a formally simple formula [Eq. (2.25)] for the reduced dynamics of S and shows that the latter becomes rigorously Markovian in the limit of infinitely short correlation times of the reservoir operators which occur in the interaction. The model displays as well the completely positive character of the reduced dynamics of S as stemming from the automorphic nature of the global dynamics of $S+R$ (as discussed in I) and is sufficiently general to allow for the derivation, in the Markovian limit, of any completely positive dynamical semigroup for S by suitably varying the interaction parameters.

ACKNOWLEDGMENTS

The authors acknowledge several stimulating and enlightening discussion with E.C.G. Sudarshan, whom they also wish to thank warmly for the invitation to visit the Center for Particle Theory of the Physics Department of the University of Texas at Austin, where part of this work was accomplished. One of use (V.G.) is also grateful to R.S. Ingarden for the hospitality at the Physics Department of the Nicholas Copernicus University, Toruń, Poland, where the first draft of the paper was written.

APPENDIX A

In this appendix we show that for given $t\geq 0$, $A\in M(N)$, and $\epsilon_0>0$ the series (2.20) is uniformly convergent in $M(N)$ on the interval $0\leq\epsilon\leq\epsilon_0$. First observe that

$$M^\epsilon(t_1)\cdots M^\epsilon(t_n)(A\otimes\mathbb{1})$$
$$=i^n\sum_{[i,k,n]}(-1)^{n-k}V^\epsilon(t_{i_1})\cdots V^\epsilon(t_{i_k})(A\otimes\mathbb{1})$$
$$\times V^\epsilon(t_{i_{k+1}})\cdots V^\epsilon(t_{i_n}), \tag{A1}$$

where $V^\epsilon(t)=\sum_{\alpha=1}^{N^2}F_\alpha(t)\otimes\varphi_\alpha^\epsilon(t)$ and where the summation is extended over all partitions $[i,k,n]=(i_1,\ldots,i_k)$, $(i_{k+1},\ldots,i_n)$, $k=0,\ldots,n$, of $(1,\ldots,n)$ such that $i_1<i_2<\cdots<i_k$ and $i_{k+1}>i_{k+2}>\cdots>i_n$. There are 2^n such partitions. We also introduce the notations $F=\sup_\alpha\|F_\alpha\|$ and $C^2=\sum_{\alpha,\beta=1}^{N^2}|c_{\alpha\beta}|$. Since the norm of a matrix is invariant under a unitary transformation, we have $F=\sup_{\alpha,t}\|F_\alpha(t)\|$. Then, if $(i_1,\ldots,i_k)$, $(i_{k+1},\ldots,i_{2n})$ is a partition as above and $x,y\in\mathcal{H}=\mathbb{C}^N$ we have, using (2.9) and (2.10),

$$|(x\otimes\Omega, V^\epsilon(t_{i_1})\cdots V^\epsilon(t_{i_k})(A\otimes\mathbb{1})V^\epsilon(t_{i_{k+1}})\cdots V^\epsilon(t_{i_{2n}})y\otimes\Omega)|$$

$$=\Big|\sum_{\alpha_1,\ldots,\alpha_{2n}}(\Omega,\varphi^\epsilon_{\alpha_1}(t_{i_1})\cdots\varphi^\epsilon_{\alpha_{2n}}(t_{i_{2n}})\Omega)(x,F_{\alpha_1}(t_{i_1})\cdots F_{\alpha_k}(t_{i_k})AF_{\alpha_{k+1}}(t_{i_{k+1}})\cdots F_{\alpha_{2n}}(t_{i_{2n}})y)\Big|$$

$$\leqslant\sum_{\alpha_1,\ldots,\alpha_{2n}}|(\Omega,\varphi^\epsilon_{\alpha_1}(t_{i_1})\cdots\varphi^\epsilon_{\alpha_{2n}}(t_{i_{2n}})\Omega)|\;\|x\|\;\|y\|\;\|A\|F^{2n}$$

$$\leqslant\sum_{\alpha_1,\ldots,\alpha_{2n}}\sum_{p\in\mathcal{P}_n}\prod_{r=1}^{n}|c_{\alpha_{p(2r)}\alpha_{p(2r-1)}}|\,|\delta_\epsilon(t_{i_{p(2r-1)}}-t_{i_{p(2r)}})|\,\|x\|\;\|y\|\;\|A\|F^{2n}$$

$$=\|x\|\;\|y\|\;\|A\|(CF)^{2n}\sum_{p\in\mathcal{P}_n}\prod_{r=1}^{n}\delta_\epsilon(t_{i_{p(2r-1)}}-t_{i_{p(2r)}}).$$

Hence, since

$$\sum_{p\in\mathcal{P}_n}\prod_{r=1}^{n}\delta_\epsilon(t_{i_{p(2r-1)}}-t_{i_{p(2r)}})$$

is a symmetric function of $t_1,\ldots,t_{2n}$,

$$|(x\otimes\Omega,M^\epsilon(t_1)\cdots M^\epsilon(t_{2n})(A\otimes\mathbb{1})y\otimes\Omega)|\leqslant(2CF)^{2n}\|x\|\;\|y\|\;\|A\|\sum_{p\in\mathcal{P}_n}\prod_{r=1}^{n}\delta_\epsilon(t_{p(2r-1)}-t_{p(2r)}).\tag{A2}$$

From (2.20) and (2.24) we have

$$(x,(\gamma^\epsilon_tA)y)=(x,Ay)+\sum_{n=1}^{\infty}(x\otimes\Omega,I^\epsilon_{2n}(t,0)(A\otimes\mathbb{1})y\otimes\Omega),\quad t\geqslant0,\tag{A3}$$

and, from (A2), noting that the summation at the right-hand side (2.9) contains $(2n)!/2^nn!$ terms,

$$|(x\otimes\Omega,I^\epsilon_{2n}(t,0)(A\otimes\mathbb{1})y\otimes\Omega)|$$

$$\leqslant\int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{2n-1}}dt_{2n}|(x\otimes\Omega,M^\epsilon(t_1)M^\epsilon(t_2)\cdots M^\epsilon(t_{2n})(A\otimes\mathbb{1})y\otimes\Omega)|$$

$$\leqslant(2CF)^{2n}\|x\|\;\|y\|\;\|A\|\frac{1}{(2n)!}\int_0^t dt_1\int_0^t dt_2\cdots\int_0^t dt_{2n}\sum_{p\in\mathcal{P}_n}\prod_{r=1}^{n}\delta_\epsilon(t_{p(2r-1)}-t_{p(2r)})$$

$$=(2CF)^{2n}\|x\|\;\|y\|\;\|A\|\frac{1}{2^nn!}\Big(\int_0^t dt_1\int_0^t dt_2\,\delta_\epsilon(t_1-t_2)\Big)^n,\quad t\geqslant0.\tag{A4}$$

Now if $t\geqslant0$ we have

$$\int_0^t dt_1\int_0^t dt_2\,\delta_\epsilon(t_1-t_2)=\int_0^t dt_1\int_{-\epsilon^{-1}(t-t_1)}^{\epsilon^{-1}t_1}dx\frac{\exp(-x^2)}{\sqrt{\pi}}\quad\text{and}\quad\lim_{\epsilon\downarrow0}\int_0^t dt_1\int_0^t dt_2\,\delta_\epsilon(t_1-t_2)=t.$$

Therefore

$$|(x\otimes\Omega,I^\epsilon_{2n}(t,0)(A\otimes\mathbb{1})y\otimes\Omega)|\leqslant\|x\|\;\|y\|\;\|A\|[(2F^2C^2t)^n/n!],\quad t\geqslant0,\quad\epsilon\geqslant0.\tag{A5}$$

From (A5) there follows the uniform convergence of (A3) on any finite interval $0\leqslant\epsilon\leqslant\epsilon_0$.[10]

APPENDIX B

From the uniform convergence of (2.20) on every finite interval $0\leqslant\epsilon\leqslant\epsilon_0$ (see Appendix A) it follows that, in order to prove that $\lim_{\epsilon\downarrow0}\gamma^\epsilon_tA=\gamma_tA$, where γ_tA is given by (2.27), it is sufficient to show that (2.20) tends to (2.27) termwise.[11] In other words, in terms of matrix elements, we must show that

$$\lim_{\epsilon\downarrow0}(x\otimes\Omega,I^\epsilon_{2n}(t,0)(A\otimes\mathbb{1})y\otimes\Omega)$$
$$=\int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n\times(x,[S(t_1)S(t_2)\cdots S(t_n)A]y),\quad t\geqslant0,\tag{B1}$$

The statement is trivial for $t=0$, hence we assume throughout $t>0$. We need some preliminary lemmas.

Lemma 1. Define the operator $\phi(t,s)$: $M(N)\to M(N)$, $t\geqslant s\geqslant0$, as

$$\phi(t,s)B=\frac{1}{2}\sum_{\alpha,\beta=1}^{N^2}c_{\alpha\beta}\times\{[F_\beta(s),B]F_\alpha(t)+F_\beta(t)[B,F_\alpha(s)]\},\quad B\in M(N).\tag{B2}$$

Then, with the notations of Appendix A, we have

$$\|\phi(t_1,t_2)\phi(t_3,t_4)\cdots\phi(t_{2n-1},t_{2n})B\|\leqslant(2C^2F^2)^n\|B\|.\tag{B3}$$

Proof. We have

$$\|\phi(t,s)B\|\leqslant\frac{1}{2}\sum_{\alpha,\beta=1}^{N^2}|c_{\alpha\beta}|\;\|F_\beta(s)BF_\alpha(t)-BF_\beta(s)F_\alpha(t)+F_\beta(t)BF_\alpha(s)-F_\beta(t)F_\alpha(s)B\|$$
$$\leqslant\frac{1}{2}\sum_{\alpha,\beta=1}^{N^2}|c_{\alpha\beta}|4F^2\|B\|=2\|B\|C^2F^2.$$

By iteration, the statement follows.

Lemma 2. With the notation $U_t = \exp(-iH_S t)$ [see Eq. (2.16)] and those of Appendix A we have

$$\|F_\alpha(t) - F_\alpha(s)\| \le 2F\|U_{t-s} - \mathbb{1}\|, \quad t \ge s > 0. \tag{B4}$$

Proof.

$$\|F_\alpha(t) - F_\alpha(s)\| = \|U_{t-s}F_\alpha(s)U^*_{t-s} - F_\alpha(s)\| = \|(U_{t-s} - \mathbb{1})F_\alpha(s)U^*_{t-s} + F_\alpha(s)(U^*_{t-s} - \mathbb{1})\|$$

$$= \|(U_{t-s} - \mathbb{1})F_\alpha(s)U^*_{t-s} + F_\alpha(s)U^*_{t-s}(\mathbb{1} - U_{t-s})\| \le \|U_{t-s} - \mathbb{1}\| 2\|F_\alpha(s)U^*_{t-s}\| \le 2\|F_\alpha(s)\| \|U_{t-s} - \mathbb{1}\| \le 2F\|U_{t-s} - \mathbb{1}\|.$$

Lemma 3. With the notations of Appendix A and of Lemmas 1 and 2 we have

$$\|\phi(t,s)B - \phi(s,s)B\| \le 4(CF)^2\|B\| \|U_{t-s} - \mathbb{1}\|, \quad t \ge s > 0, \quad B \in M(N). \tag{B5}$$

Proof. Using the result of Lemma 2 we obtain

$$\|\phi(t,s)B - \phi(s,s)B\| \le \frac{1}{2}\sum_{\alpha,\beta=1}^{N^2} |c_{\alpha\beta}| \, \|F_\beta(s)B(F_\alpha(t) - F_\alpha(s))$$

$$+ (F_\beta(t) - F_\beta(s))BF_\alpha(s) - BF_\beta(s)(F_\alpha(t) - F_\alpha(s)) - (F_\beta(t) - F_\beta(s))F_\alpha(s)B\|$$

$$\le \sum_{\alpha,\beta=1}^{N^2} |c_{\alpha\beta}|\{\|F_\alpha(t) - F_\alpha(s)\| \, \|F_\beta(s)\| \, \|B\| + \|F_\beta(t) - F_\beta(s)\| \, \|F_\alpha(s)\| \, \|B\|\} \le 4(CF)^2\|B\| \, \|U_{t-s} - \mathbb{1}\|.$$

Lemma 4. Given an integrable function $\lambda(t_1, t_2, \ldots, t_{2n-1}, t_{2n})$ we have

$$\int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}\, \lambda(t_1, t_2, \ldots, t_{2n-1}, t_{2n})$$

$$= \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \int_0^{t-s_1} dx_1 \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n\, \lambda(s_1 + x_1, s_1, \ldots, s_n + x_n, s_n), \quad t > 0. \tag{B6}$$

Proof. For a function of two variables $\mu(t_1, t_2)$ is easily seen that

$$\int_0^{t} dt_1 \int_0^{t_1} dt_2\, \mu(t_1, t_2) = \int_0^t ds \int_0^{t-s} dx\, \mu(s + x, s).$$

By applying repeatedly this formula the statement follows.

In the following, we also use the notation

$$\Delta(x) = \frac{2}{\sqrt{\pi}} \exp(-x^2) \tag{B7}$$

Proof of (B1). From (B1), (2.9), (2.10), and (B2), we have

$$(x \otimes \Omega, M^\epsilon(t_1) \cdots M^\epsilon(t_{2n})(A \otimes \mathbb{1})y \otimes \Omega) = i^{2n} \sum_{\alpha_1, \ldots, \alpha_{2n}} \sum_{\{i, l, 2n\}} (-1)^{2n-l}(\Omega, \varphi^\epsilon_{\alpha_1}(t_{i_1}) \cdots \varphi^\epsilon_{\alpha_{2n}}(t_{i_{2n}})\Omega)$$

$$\times (x, F_{\alpha_1}(t_{i_1}) \cdots F_{\alpha_l}(t_{i_l})AF_{\alpha_{l+1}}(t_{i_{l+1}}) \cdots F_{\alpha_{2n}}(t_{i_{2n}})y) = 2^n\delta_\epsilon(t_1 - t_2)\delta_\epsilon(t_3 - t_4) \cdots \delta_\epsilon(t_{2n-1} - t_{2n})$$

$$\times (x, [\phi(t_1, t_2)\phi(t_3, t_4) \cdots \phi(t_{2n-1}, t_{2n})]y) + (x, [\mathcal{J}(t_1, \ldots, t_{2n})A]y), \tag{B8}$$

where in the operator $\mathcal{J}(t_1, \ldots, t_{2n})$ only terms occur which contain products $\prod_{r=1}^n \delta_\epsilon(t_{p(2r-1)} - t_{p(2r)})$ in which the permutation p is such that $|p(2r-1) - p(2r)| > 1$ for some r. Hence we obtain, analogously to (B2),

$$|(x, [\mathcal{J}(t_1, \ldots, t_{2n})A]y)| \le (2CF)^{2n}\|x\| \, \|y\| \, \|A\| \sum_{p \in \mathcal{P}_n} \prod_{r=1}^n \delta_\epsilon(t_{p(2r-1)} - t_{p(2r)}) - \delta_\epsilon(t_1 - t_2)\delta_\epsilon(t_3 - t_4) \cdots \delta_\epsilon(t_{2n-1} - t_{2n}) \ . \tag{B9}$$

From (2.26) and (B2) we have $S(t) = \phi(t,t)$, hence, by (2.24) and (B8),

$$|(x \otimes \Omega, I^\epsilon_{2n}(t, 0)(A \otimes \mathbb{1})y \otimes \Omega) - \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n\, (x, [S(t_1)S(t_2) \cdots S(t_n)A]y)|$$

$$\le \Big| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}\, 2^n\delta_\epsilon(t_1 - t_2)\delta_\epsilon(t_3 - t_4) \cdots \delta_\epsilon(t_{2n-1} - t_{2n})(x, [\phi(t_1, t_2)\phi(t_3, t_4) \cdots \phi(t_{2n-1}, t_{2n})A]y)$$

$$- \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n\, (x, [\phi(t_1, t_1)\phi(t_2, t_2) \cdots \phi(t_n, t_n)A]y)\Big|$$

$$+ \Big| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}(x, [\mathcal{J}(t_1, \ldots, t_{2n})A]y)\Big|, \quad t > 0. \tag{B10}$$

Let us denote by $\alpha^\epsilon_n(t)$ and $\beta^\epsilon_n(t)$ the first and second term at the right-hand side of (B10), respectively. From (B9), Lemma 4., (B7), and the symmetry of the function $\sum_{p \in \mathcal{P}_n} \prod_{r=1}^n \delta_\epsilon(t_{p(2r-1)} - t_{p(2r)})$, we have

$$\beta_n^\epsilon(t) \le (2CF)^{2n}\|x\|\,\|y\|\,\|A\| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n} \Big(\sum_{p\in P_n} \prod_{r=1}^{n} \delta_\epsilon(t_{p(2r-1)} - t_{p(2r)}) - \delta_\epsilon(t_1-t_2)\delta_\epsilon(t_3-t_4)\cdots\delta_\epsilon(t_{2n-1}-t_{2n})\Big)$$

$$= (2CF)^{2n}\|x\|\,\|y\|\,\|A\| \Big(\frac{1}{(2n)!}\int_0^t dt_1 \int_0^t dt_2 \cdots \int_0^t dt_{2n} \sum_{p\in P_n} \prod_{r=1}^{n} \delta_\epsilon(t_{p(2r-1)} - t_{p(2r)})$$

$$- \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \int_0^{t-s_1} dx_1 \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n\, \delta_\epsilon(x_1)\delta_\epsilon(x_2)\cdots\delta_\epsilon(x_n)\Big)$$

$$= (2CF)^{2n}\|x\|\,\|y\|\,\|A\| \Big[\frac{1}{2^n n!}\Big(\int_0^t dt_1 \int_0^t dt_2\, \delta_\epsilon(t_1-t_2)\Big)^n - \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \int_0^{t-s_1} dx_1$$

$$\times \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n\, \delta_\epsilon(x_1)\delta_\epsilon(x_2)\cdots\delta_\epsilon(x_n)\Big] = (2C^2F^2)^n \|x\|\,\|y\|\,\|A\| \Big[\frac{1}{n!}\Big(\frac{1}{2}\int_0^t |\, dx_2 \int_{-\epsilon^{-1}(t-t_1)}^{\epsilon^{-1}t_1} dx\, \Delta(x)\Big)^n$$

$$- \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \int_0^{\epsilon^{-1}(t-s_1)} dx_1 \int_0^{\epsilon^{-1}(s_1-s_2)} dx_2 \cdots \int_0^{\epsilon^{-1}(s_{n-1}-s_n)} dx_n\, \Delta(x_1)\Delta(x_2)\cdots\Delta(x_n)\Big].$$

Since $\frac{1}{2}\int_{-\infty}^{\infty}\Delta(x)\,dx = 1$ and $t>0$, we have

$$\lim_{\epsilon\downarrow 0} \frac{1}{n!}\Big(\frac{1}{2}\int_0^t dt_1 \int_{-\epsilon^{-1}(t-t_1)}^{\epsilon^{-1}t_1} dx\,\Delta(x)\Big)^n = \frac{t^n}{n!};\ \text{also,}\ \lim_{\epsilon\downarrow 0}\int_0^{\epsilon^{-1}(s_{l-1}-s_l)} \Delta(x)\,dx = 1 \quad \text{for } s_{l-1} > s_l.$$

Therefore, since $\int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n = t^n/n!$, we see that

$$\lim_{\epsilon\downarrow 0}\int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \int_0^{\epsilon^{-1}(t-s_1)} dx_1 \int_0^{\epsilon^{-1}(s_1-s_2)} dx_2 \cdots \int_0^{\epsilon^{-1}(s_{n-1}-s_n)} dx_n\, \Delta(x_1)\Delta(x_2)\cdots\Delta(x_n) = \frac{t^n}{n!}.$$

Hence $\lim_{\epsilon\downarrow 0}\beta_n^\epsilon(t) = 0$. We now turn to the consideration of $\alpha_n^\epsilon(t)$. Using lemmas 1 and 3, we have

$$\alpha_n^\epsilon(t) = \Big| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}\, 2^n \delta_\epsilon(t_1-t_2)\delta_\epsilon - t_4)\cdots\delta_\epsilon(t_{2n-1}-t_{2n})$$

$$\times\{(x, [(\phi(t_1,t_2) - \phi(t_2,t_2))\phi(t_3,t_4)\cdots\phi(t_{2n-1},t_{2n})A]y)$$

$$+ (x, [\phi(t_2,t_2)(\phi(t_3,t_4) - \phi(t_4,t_4))\phi(t_5,t_6)\cdots\phi(t_{2n-1},t_{2n})A]y)$$

$$+\cdots+ (x, [\phi(t_2,t_2)\phi(t_4,t_4)\cdots\phi(t_{2n-2},t_{2n-2})(\phi(t_{2n-1},t_{2n}) - \phi(t_{2n},t_{2n}))A]y) + (x, [\phi(t_2,t_2)\phi(t_4,t_4)\cdots\phi(t_{2n},t_{2n})A]y)\}$$

$$- \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n (x, [\phi(t_1,t_1)\phi(t_2,t_2)\cdots\phi(t_n,t_n)A]y)\Big|$$

$$\le \|x\|\,\|y\| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}\, 2^n\delta_\epsilon(t_1-t_2)\cdots\delta_\epsilon(t_{2n-1}-t_{2n})$$

$$\times\{\|(\phi(t_1,t_2) - \phi(t_2,t_2))\phi(t_3,t_4)\cdots\phi(t_{2n-1},t_{2n})A\| + \cdots + \|\phi(t_2,t_2)\cdots\phi(t_{2n-2},t_{2n-2})(\phi(t_{2n-1},t_{2n}) - \phi(t_{2n,2n}))A\|\}$$

$$+ \Big|\int_0^t dt_1 \cdots \int_0^{t_{2n-1}} dt_{2n}\, 2^n\delta_\epsilon(t_1-t_2)\cdots\delta_\epsilon(t_{2n-1}-t_{2n})(x, [\phi(t_2,t_2)\cdots\phi(t_{2n},t_{2n})A]y)$$

$$- \int_0^t dt_1 \cdots \int_0^{t_{n-1}} dt_n (x, [\phi(t_1,t_1)\cdots\phi(t_n,t_n)A]y)\Big| \le 2(2C^2F^2)^n\|x\|\,\|y\|\,\|A\| \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{2n-1}} dt_{2n}\, 2^n\delta_\epsilon(t_1-t_2)\cdots$$

$$\delta_\epsilon(t_{2n-1}-t_{2n})$$

$$\times(\|U_{t_1-t_2} - \mathbb{1}\| + \cdots + \|U_{t_{2n-1}-t_{2n}} - \mathbb{1}\|) + \Big|\int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \Big(\int_0^{t-s_1} dx_1 \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n$$

$$\times 2^n\delta_\epsilon(x_1)\delta_\epsilon(x_2)\cdots\delta_\epsilon(x_n) - 1\Big)(x, [\phi(s_1,s_1)\cdots\phi(s_n,s_n)A]y)\Big| \le (2C^2F^2)^n\|x\|\,\|y\|\,\|A\|\Big\{\int_0^t ds_1 \cdots \int_0^{t_{n-1}} ds_n$$

$$\times\Big[2\int_0^{t-s_1} dx_1 \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n 2^n\delta_\epsilon(x_1)\cdots\delta_\epsilon(x_n)(\|U_{x_1} - \mathbb{1}\| + \cdots + \|U_{x_n} - \mathbb{1}\|)$$

$$+ \Big(\int_0^{t-s_1} dx_1 \int_0^{s_1-s_2} dx_2 \cdots \int_0^{s_{n-1}-s_n} dx_n\, 2^n\delta_\epsilon(x_1)\cdots\delta_\epsilon(x_n) - 1\Big)\Big]\Big\}$$

$$= (2C^2F^2)^n\|x\|\,\|y\|\,\|A\|\Big\{\int_0^t ds_1 \cdots \int_0^{s_{n-1}} ds_n\Big[2\int_0^{\epsilon^{-1}(t-s_1)} dx_1 \cdots \int_0^{\epsilon^{-1}(s_{n-1}-s_n)} dx_n\, \Delta(x_1)\cdots\Delta(x_n).$$

$$\times(\|U_{\epsilon x_1} - \mathbb{1}\| + \cdots + \|U_{\epsilon x_n} - \mathbb{1}\|) + \Big(\int_0^{\epsilon^{-1}(t-s_1)} dx_1 \cdots \int_0^{\epsilon^{-1}(s_{n-1}-s_n)} dx_n\, \Delta(x_1)\cdots\Delta(x_n) - 1\Big)\Big]\Big\}.$$

Since $t>0$ and since $\lim_{\epsilon\downarrow 0}\|U_{\epsilon x_l} - \mathbb{1}\| = 0$ $(l = 1, \ldots, n)$ and $\int_0^\infty \Delta(x)\,dx = 1$, this expression tends to zero as $\epsilon\downarrow 0$. This completes the proof of (B1).

APPENDIX C

We note that (2.9) in equivalent to the following formula (for convenience, we drop the ϵ in the sequel):

$$\langle\varphi_{\alpha_1}(t_1)\varphi_{\alpha_2}(t_2)\cdots\varphi_{\alpha_{2n}}(t_{2n})\rangle=\sum_{j=2}^{2n}\langle\varphi_{\alpha_1}(t_1)\varphi_{\alpha_j}(t_j)\rangle\langle\varphi_{\alpha_2}(t_2)\cdots\varphi_{\alpha_{j-1}}(t_{j-1})\varphi_{\alpha_{j+1}}(t_{j+1})\cdots\varphi_{\alpha_{2n}}(t_{2n})\rangle$$
$$=\sum_{j=1}^{2n-1}\langle\varphi_{\alpha_j}(t_j)\varphi_{\alpha_{2n}}(t_{2n})\rangle\langle\varphi_{\alpha_1}(t_1)\cdots\varphi_{\alpha_{j-1}}(t_{j-1})\varphi_{\alpha_{j+1}}(t_{j+1})\cdots\varphi_{\alpha_{2n-1}}(t_{2n-1})\rangle,\quad n=1,2,\ldots. \tag{C1}$$

We have (summation over repeated Greek indices is understood throughout in the following)

$$\langle M(t)M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})\rangle=iF_\alpha(t)\langle(\mathbb{1}_N\otimes\varphi_\alpha(t))[M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})]\rangle$$
$$-i\langle[M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})](\mathbb{1}_N\otimes\varphi_\alpha(t))\rangle F_\alpha(t), \tag{C2}$$

Using (A1) and (C1), the first term at the right-hand side of (C2) can be expressed as follows:

$$iF_\alpha(t)\langle(\mathbb{1}_N\otimes\varphi_\alpha(t))[M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})]\rangle=iF_\alpha(t)i^{(2n-1)}\sum_{\{i,l,2n-1\}}(-1)^{2n-1-l}$$
$$\times\langle\varphi_\alpha(t)\varphi_{\alpha_1}(t_{i_1})\cdots\varphi_{\alpha_{2n-1}}(t_{i_{2n-1}})\rangle F_{\alpha_1}(t_1)\cdots F_{\alpha_l}(t_{i_l})AF_{\alpha_{l+1}}(t_{i_{l+1}})\cdots F_{\alpha_{2n-1}}(t_{i_{2n-1}})$$
$$\times iF_\alpha(t)i^{(2n-1)}\sum_{\{i,l,2n-1\}}\sum_{j=1}^{2n-1}(-1)^{2n-1-l}\langle\varphi_\alpha(t)\varphi_{\alpha_j}(t_{i_j})\rangle\langle\varphi_{\alpha_1}(t_{i_1})\cdots\varphi_{\alpha_{j-1}}(t_{i_{j-1}})\varphi_{\alpha_{j+1}}(t_{i_{j+1}})\cdots\varphi_{\alpha_{2n-1}}(t_{i_{2n-1}})\rangle$$
$$\times F_{\alpha_1}(t_{i_1})\cdots F_{\alpha_l}(t_{i_l})AF_{\alpha_{l+1}}(t_{i_{l+1}})\cdots F_{\alpha_{2n-1}}(t_{i_{2n-1}}).$$

We now select all the terms in the double summation for which i_j is equal to a given integer k $(1\leq k\leq 2n-1)$. Clearly, the sum of all these terms is equal to

$$iF_\alpha(t)\langle\varphi_\alpha(t)\varphi_\beta(t_k)\rangle\langle M(t_1)\cdots M(t_{k-1})\mathcal{L}_\beta(t_k)M(t_{k+1})\cdots M(t_{2n-1})(A\otimes\mathbb{1})\rangle.$$

Hence we have

$$iF_\alpha(t)\langle(\mathbb{1}_N\otimes\varphi_\alpha(t))[M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})]\rangle=\sum_{k=1}^{2n-1}\langle\varphi_\alpha(t)\varphi_{\alpha_k}(t_k)\rangle F_\alpha(t)\langle M(t_1)\cdots M(t_{k-1})\mathcal{L}_{\alpha_k}(t_k)M(t_{k+1})\cdots M(t_{2n-1})(A\otimes\mathbb{1})\rangle. \tag{C3}$$

In the same way, we get that

$$i\langle[M(t_1)\cdots M(t_{2n-1})(A\otimes\mathbb{1})](\mathbb{1}_N\otimes\varphi_\alpha(t))\rangle F_\alpha(t)=i\sum_{k=1}^{2n-1}\langle\varphi_{\alpha_k}(t_k)\varphi_\alpha(t)\rangle\langle M(t_1)\cdots M(t_{k-1})\mathcal{L}_{\alpha_k}(t_k)M(t_{k+1})\cdots M(t_{2n-1})(A\otimes\mathbb{1})\rangle F_\alpha(t). \tag{C4}$$

Substituting (C3) and (C4) into (C2), we obtain (2.21).

*Supported in party by USAEC, Contract No. AT(40-1)3992, by INFN, Sezione di Milano and by the Institute of Mathematics, Polish Academy of Sciences, Warsaw.

[1]V. Gorini, A Kossakowski and E.C.G. Sudarshan, J. Math. Phys. 17, 821 (1976). We refer to this paper for notations and terminology.

[2]During the preparation of the present paper, we have learned that a more general result has been recently obtained independently by G. Lindblad ["On the Generators of Quantum Dynamical Semigroups," preprint Royal Institute of Technology, Stockholm, TRITA-TRY-75-1, 1975]. In this paper, the author was able to show, by means of the introduction of a suitable dissipation function, that the generator L of a norm continuous completely positive Schrödinger dynamical semigroup $t\to\exp(Lt):\mathcal{T}(\mathcal{H})\to\mathcal{T}(\mathcal{H})$, $t\geq 0$, $\mathcal{H}$ separable, can be written in the form

$$L:\rho\to L\rho=-i[H,\rho]+\tfrac{1}{2}\sum_i\{[V_i,\rho V_i^*]+[V_i\rho,V_i^*]\},$$

where H is a bounded self-adjoint operator and $\{V_i\}$ is a (possibly finite) sequence of bounded operators such that $\sum_i V_i^*V_i$ converges ultraweakly in $\mathcal{B}(\mathcal{H})$. The series in the expression of $L\rho$ converges in the trace norm, for all $\rho\in\mathcal{T}(\mathcal{H})$. The restriction of norm continuity is a strong one, since it is equivalent to the boundedness of L, a property which is not satisfied in general in physical applications, when $\mathcal{H}$ is infinite dimensional. However, there is some hope that the foregoing result can be extended to the general case with unbounded operators H and V_i [G. Lindblad (private communication)].

[3]K. Hepp and E.H. Lieb, Helv. Phys. Acta **46**, 573 (1973).

[4]C. Itzykson, Commun. Math. Phys. **36**, 19 (1974).

[5]R.F. Fox, J. Math. Phys. **13**, 1196 (1972); **13**, 1726 (1972); **14**, 20 (1973); **14**, 1187 (1973); **15**, 217 (1974); **15**, 1479 (1974); **15**, 1918 (1974); **16**, 289 (1975).

[6]H. Primas, Helv. Phys. Acta **34**, 36 (1961).

[7]R. Kubo, J. Math. Phys. **4**, 174 (1963).

[8]A. Frigerio and V. Gorini, J. Math. Phys. (to be published).

[9]By a *dynamical map* we mean a positive trace preserving linear map $M(N)\to M(N)$.

[10]J.M. Hyslop, *Infinite Series* (Oliver and Boyd, London, 1965), theorem 36.

[11]Reference 10, theorem 39.

On space–time Killing tensors with a Segre characteristic [(11)(11)]

I. Hauser and R. J. Malhiot

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 24 March 1975)

We consider the set of all space–times which admit a Killing tensor $K_{\alpha\beta}$ whose Segre characteristic is [(11)(11)] and whose eigenvalues λ^1 and λ^2 satisfy the condition that $d\lambda^1 \wedge d\lambda^2$ is not null. We prove that there locally exist scalar fields ϕ^3, ϕ^4 such that λ^1, λ^2, ϕ^3, ϕ^4 constitute a chart, and $d\lambda^1 \cdot d\lambda^2 = d\lambda^i \cdot d\phi^r = 0$ for $i = 1, 2$ and $r = 3, 4$. Also, relative to this coordinate system, the metric has the same general form as Carter's Hamilton–Jacobi separable metric except that his arbitrary functions of λ^i are replaced by functions of λ^i, ϕ^3, ϕ^4. If $R^{\alpha\beta}(\nabla_\alpha\lambda^1)(\nabla_\beta\lambda^2) = 0$, there exists a two parameter Abelian group of isometries which commute with $K_{\alpha\beta}$; also (ϕ^3,ϕ^4) can be chosen so that $\partial/\partial\phi^3$ and $\partial/\partial\phi^4$ are the Killing vectors. $R^{\alpha\beta}(\nabla_\alpha\lambda^1)(\nabla_\beta\lambda^2) = 0$ is necessary and sufficient for Schrödinger equation separability in case (a) of Carter. A Newtonian analog of our results is discussed.

1. INTRODUCTION

Amongst the space—times with two parameter Abelian isometry groups, the Kerr metric and the other Hamilton—Jacobi separable space—times studied by Carter[1] each admits a Killing tensor $K_{\alpha\beta}$ whose Segre characteristic is [(11)(11)]. We recall that a Segre characteristic of [(11)(11)] means that there exists a null tetrad[2] consisting of two real vectors k_α, m_α and two complex vectors t_α, t^*_α such that $k_\alpha m^\alpha = t_\alpha t^{\alpha *} = 1$ and such that

$$K_{\alpha\beta} = \lambda^1(k_\alpha m_\beta + k_\beta m_\alpha) + \lambda^2(t_\alpha t^*_\beta + t_\beta t^*_\alpha). \tag{1}$$

λ^1 and λ^2 are the eigenvalues of the tensor.

We are going to consider those space—times which admit a [(11)(11)] Killing tensor subject to the condition that the 2-surfaces $\lambda^i = \text{const}$ $(i = 1, 2)$ not be null almost everywhere; i.e.,

$$(d\lambda^1 \wedge d\lambda^2) \cdot (d\lambda^1 \wedge d\lambda^2) \neq 0 \tag{2}$$

almost everywhere.

However, we are not going to postulate the existence of any Killing vectors as an independent assumption. In fact, it is our objective to demonstrate that, for a broad class of matter tensors which include the vacuum as a special case, Eqs. (1) *and* (2) *are sufficient to imply the existence of a two parameter Abelian isometry group* which automatically commutes with the given Killing tensor and whose surfaces of transitivity are the surfaces of constant λ^i.

Specifically, we will prove the following theorems concerning any space—time which admits a [(11)(11)] Killing tensor:

(I) The linearly independent eigenvectors k_α and m_α of the Killing tensor are shear-free geodesic fields and are, therefore, principal null vectors of the conform tensor.[3]

(II) For any given point p_0 in the support of $(d\lambda^1 \wedge d\lambda^2) \cdot (d\lambda^1 \wedge d\lambda^2)$, there exist scalar fields ϕ^3 and ϕ^4 such that $x^1 = \lambda^1$, $x^2 = \lambda^2$, $x^3 = \phi^3$, $x^4 = \phi^4$ constitute a chart which covers p_0, and

$$d\lambda^i \cdot d\phi^r = 0, \quad i = 1, 2, \quad r = 3, 4,$$

$$d\lambda^1 \cdot d\lambda^2 = 0.$$

(III) The metric tensor components relative to the above chart (Theorem II) have the same general form as Carter's[1] Hamilton—Jacobi separable metric except that his arbitrary functions of λ^i are to be replaced by functions of λ^i, ϕ^3, and ϕ^4. As we will see, the dependences on ϕ^3 and ϕ^4 are not arbitrary. (A complete wording of this theorem will appear in Sec. 3.)

(IV) Suppose that, in addition to Eqs. (1) and (2), there exists a two parameter Abelian isometry group which commutes with $K_{\alpha\beta}$. Then,

(a) the 2-surfaces of transitivity are not null and are the surfaces of constant λ^1 and λ^2,

(b) ϕ^3 and ϕ^4 (see Theorem II) can be chosen so that $\partial/\partial\phi^3$ and $\partial/\partial\phi^4$ are Killing vectors for the group, and

(c) as a corollary, the existence of this kind of symmetry group taken together with Eqs. (1) and (2) are collectively equivalent to Carter's[1] Hamilton—Jacobi separability criteria for the case where his U_1 and U_{-1} are not constants.

(V) Equations (1) and (2) and the additional assumption

$$R^{\alpha\beta}(\nabla_\alpha\lambda^1)(\nabla_\beta\lambda^2) = 0 \tag{3}$$

are sufficient to imply the existence of a two parameter Abelian isometry group which commutes with $K_{\alpha\beta}$.

(VI) Equations (1), (2), and (3) are collectively equivalent to Carter's[1] Schrödinger equation separability criteria for the case where his U_1 and U_{-1} are not constants [his Case (a)].

The above theorems are in the order in which they will appear during our presentation of the subject. Theorems (I) and (IV) have been independently proven by Dietz.[4] Theorems II, III, V, and VI constitute new results.

It should be noted that Theorem I follows directly from the assumption that the Killing tensor has a Segre characteristic [(11)(11)]; i.e., Eq. (2) need not be assumed.

Theorem II is important, because it provides a canonical coordinate system which can be used for the analysis of all space—times which admit $[(11)(11)]$ Killing tensors subject to Eq. (2). The specifics of this coordinate system and the precise form of the corresponding metric are contained in our proof of Theorem III. As regards Theorem IV, note that an immediate consequence of this theorem is the invariance of the space—time with respect to simultaneous inversion of both Killing vectors.[4] Theorem V is especially striking, since the premises say *nothing* about the algebraic classification of the conform tensor; nothing is assumed except Eqs. (1), (2), and (3). The result is consistent with and throws new light on the observations of Hughston, Penrose, Sommers, and Walker[5] on Killing tensors of type D vacuums. Theorem VI provides us with a physical criterion (vanishing of a given matter tensor component) for Schrödinger equation separability.

In Sec. 2, we will substitute Eqs. (1) and (2) into the defining equation $\nabla_{(\alpha}K_{\beta\gamma)}=0$ for a Killing tensor, and we will derive the well-known necessary and sufficient set of conditions[6] which the coefficients of rotation must satisfy if the space—time is to admit a $[(11)(11)]$ Killing tensor. In Sec. 3, we will prove Theorems I—IV. In Sec. 4, we will prove Theorems V and VI. In Sec. 5, we will describe an interesting Newtonian counterpart of our Theorems II, III, and V, and we will briefly discuss the bearing of our results on the search for nonstationary space—times which admit nonreducible Killing tensors. Also, we will discuss some of the work which remains to be done on $[(11)(11)]$ Killing tensors.

2. KILLING TENSOR EQUATIONS IN NULL TETRAD FORM

We let k, m, t, t^* denote any null tetrad[2] consisting of real 1-forms k, m and complex 1-forms t, t^* such that t^* is the complex conjugate of t, and $k\cdot m = t\cdot t^* = 1$.

The script variables which we use for coordinate tetrad components of tensors are

$\alpha, \beta, \cdots$ with values 1, 2, 3, 4,

i, j with values 1, 2,

r, s with values 3, 4.

For null tetrad[7] components, we use script variables

$a, b, \cdots$ with values k, m, t, t^*.

The metric components $g_{ab}=g_{(ab)}$ all vanish except for $g_{km}=g_{tt^*}=1$; also, $g^{ab}=g_{ab}$. Therefore, for any 1-form u, we have $u_m = u^k$, $u_{t^*}=u^t$, etc.

The null tetrad components of the differential of any scalar field ϕ are denoted by $d_a\phi$. To compute the null tetrad components of covariant derivatives, we simply adopt the equations of the classical theory[6] of orthonormal components with the only changes being in the specific values of the metric components. For example, the null tetrad components of the covariant derivative of any second order tensor are given by

$$\nabla_c K_{ab} = d_c K_{ab} - \Omega_{cad}K^d{}_b - \Omega_{cbd}K_a{}^d. \tag{4}$$

$\Omega_{cab}=-\Omega_{cba}$ are the *coefficient of rotation*,[6] which are defined to be the null tetrad components of the covariant derivatives of the null tetrad members. For example,

$$\Omega_{tkm}=t^\alpha(\nabla_\alpha k_\beta)m^\beta.$$

It is convenient to introduce the special notations

$$P_a=\Omega_{akm},\quad iQ_a=\Omega_{att^*},$$
$$v_a=\Omega_{akt},\quad w_a=\Omega_{amt^*}. \tag{5}$$

P_a, Q_a, v_a and w_a may be complex. $v_k, v_t, \mathrm{Re}v_{t^*}$, and $\mathrm{Im}v_{t^*}$ are the geodesy, shear, divergence, and twist of the null congruence defined by k; the corresponding optical parameters for the null congruence defined by m are w_m, w_{t^*}, $\mathrm{Re}w_t$, and $\mathrm{Im}w_t$. All of the coefficients of rotation are simple multiples of P_a, Q_a, v_a, w_a, v_a^*, w_a^*. As a note of caution, we always use the convention $v_a^* =(v_a)^*$.

We select our null tetrad so that the $[(11)(11)]$ Killing tensor $K_{\alpha\beta}$ has the canonical form given by Eq. (1), i.e., all $K_{ab}=K_{(ab)}$ vanish except for

$$K_{km}=\lambda^1,\quad K_{tt^*}=\lambda^2.$$

We proceed to substitute this canonical form into the defining equation for a Killing tensor, viz.,

$$\nabla_{(c}K_{ab)}=0.$$

Equations (4) and (5) are used in the calculation, which is straightforward and yields the following necessary and sufficient set of conditions for K_{ab} to be a Killing tensor:

$$v_k=v_t=w_m=w_{t^*}=0, \tag{6}$$

$$d\lambda^1=\rho^2[(w_k+v_m^*)t+(w_k^*+v_m)t^*],$$
$$d\lambda^2=\rho^2[(w_t+w_t^*)k+(v_{t^*}+v_{t^*}^*)m], \tag{7}$$

where we let[8]

$$\rho^2=\lambda^2-\lambda^1. \tag{8}$$

From Eqs. (6), we immediately see that *both k and m are shear-free null geodesic fields and are, therefore, principal null vectors of the conform tensor.*[3] This is Theorem I of the Introduction.

So far, we have not used the nondegeneracy condition of Eq. (2). From Eq. (7), we see that Eq. (2) is equivalent to the condition

$$\rho^2(w_k+v_m^*)(w_t+w_t^*)(v_{t^*}+v_{t^*}^*)\neq 0. \tag{9}$$

If the space—time is such that ρ^2 is positive definite or is negative definite in the support of $(d\lambda^1\wedge d\lambda^2)$, $(d\lambda^1\wedge d\lambda^2)$, then we will select the arbitrary sign in $K_{\alpha\beta}$ so that $\rho^2>0$ in the support. Otherwise, we will restrict our subsequent calculations to those events at which $\rho^2>0$. However, this involves no loss of generality, because the results for negative ρ^2 can be obtained from the results for positive ρ^2 by simply replacing ρ with $|\rho|$.[8]

The next step is to apply the integrability conditions $d^2\lambda^i=0$ to Eqs. (7). Here, we adopt the convention of suppressing the symbol $\wedge$ in exterior products of differential forms.[7] For example, $d^2\lambda^i=d\wedge d\lambda^i$, and $uv=u\wedge v$. The exterior derivatives of the null tetrad members in Eqs. (7) are computed from the equations[7]

$$dk=Pk+v^*t+vt^*,$$

$$dm = -Pm + wt + w^*t^*,\qquad(10)$$
$$dt = -w^*k - vm + iQt,$$

where P, Q, v, w are the 1-form whose components are P_a, Q_a, v_a, w_a respectively. The results of our straightforward efforts are expressible in terms of components relative to the 2-form basis km, tt^*, kt, kt^*, mt^*, mt as follows:

$$|v_m| = |w_k|,\quad v_{t^*}w_t = (v_{t^*}w_t)^*,\qquad(11)$$
$$d_t(w_k + v_m^*) - d_{t^*}(w_k^* + v_m) = -iQ_t(w_k + v_m^*) - iQ_t^*(w_k^* + v_m),\qquad(12)$$
$$d_k(w_k + v_m^*) = -(2v_{t^*} + v_t^{**} + iQ_k)(w_k + v_m^*),\qquad(13)$$
$$d_m(w_k + v_m^*) = -(w_t + 2w_t^* + iQ_m)(w_k + v_m^*),\qquad(14)$$
$$d_k(w_t + w_t^*) - d_m(v_{t^*} + v_t^{**}) = -P_k(w_t + w_t^*) - P_m(v_{t^*} + v_t^{**}),\qquad(15)$$
$$d_t(v_{t^*} + v_t^{**}) = (v_m + 2w_k^* + P_t)(v_{t^*} + v_t^{**}),\qquad(16)$$
$$d_t(w_t + w_t^*) = (2v_m + w_k^* - P_t)(w_t + w_t^*).\qquad(17)$$

Equations (9) and (11) now give us the opportunity to achieve a considerable simplification by taking advantage of the fact that the canonical form of Eq. (1) is invariant under the transformation

$$k \to e^{\chi}k,\quad m \to e^{-\chi}m,\quad t \to e^{-i\phi}t,$$

where χ and ϕ are any real fields. Under this transformation, as can be seen from Eqs. (5),

$$v_m \to e^{-i\phi}v_m,\quad w_k \to e^{i\phi}w_k,$$
$$v_{t^*} \to e^{\chi}v_{t^*},\quad w_t \to e^{-\chi}w_t.$$

Therefore, we can and do select our null tetrad (which is then uniquely determined except for a discrete group) so that

$$v_m = w_k \equiv \gamma_1 + i\delta_1,$$
$$v_{t^*} = \epsilon w_t \equiv \gamma_2 + i\delta_2,$$
$$\epsilon = \pm 1,\quad \gamma_i,\ \delta_i \text{ are real.}\qquad(18)$$

The integrability conditions of Eqs. (11) are now automatically satisfied, and we have selected a null tetrad for which Eqs. (7) and (9) reduce to the simple forms

$$d\lambda^1 = 2\rho^2\gamma_1(t + t^*),\quad d\lambda^2 = 2\rho^2\gamma_2\epsilon(k + \epsilon m),\qquad(19)$$
$$\gamma_1\gamma_2 \neq 0.\qquad(20)$$

Note that $d\lambda^1 \cdot d\lambda^2 = 0$, that $d\lambda^1$ is spacelike, and that $d\lambda^2$ is spacelike or timelike according as $\epsilon = 1$ or $\epsilon = -1$ respectively.

In addition to γ_i and δ_i, the following scalar fields occur with sufficient frequency in calculations to merit special notations:

$$\alpha_1 \equiv \tfrac{1}{2}(Q_t + Q_t^*),\qquad i\beta_1 \equiv \tfrac{1}{2}(Q_t - Q_t^*)$$
$$\alpha_2 \equiv \tfrac{1}{2}(P_k + \epsilon P_m),\quad \beta_2 \equiv \tfrac{1}{2}(P_k - \epsilon P_m).\qquad(21)$$

With the aid of Eqs. (18) and (21), the remaining integrability conditions, Eqs. (12) to (17), become

$$P_t = i\delta_1,\quad -Q_k = \epsilon Q_m = \delta_2,\qquad(22)$$
$$d_k\gamma_1 = \epsilon d_m\gamma_1 = -d_t\gamma_2 = -3\gamma_1\gamma_2,\qquad(23)$$
$$(d_t - d_{t^*})\gamma_1 = -2i\alpha_1\gamma_1,$$
$$(d_k - \epsilon d_m)\gamma_2 = -2\alpha_2\gamma_2.\qquad(24)$$

With the results of Eqs. (22), note that all of the coefficients of rotation are now simple numerical multiples or linear combinations of the eight real fields α_i, β_i, γ_i, δ_i. As regards the rest of the integrability conditions, once we have introduced a coordinate system in Sec. 3, Eqs. (23) will become tractable to integration, and Eqs. (24) will enable us to compute α_1 and α_2 in the terms of derivatives of γ_1 and γ_2.

Before we introduce any coordinate system, however, we can extract more flavor from the above integrability conditions and from Eqs. (6) and (18) by exploiting the expressions[7] for the null tetrad components of the Riemann tensor in terms of the coefficients of rotation and the differentials of these coefficients. These relations are most conveniently expressed in terms of the components

$$S_{ab} = R_{ab} - \tfrac{1}{4}g_{ab}R\qquad(25)$$

of the traceless Ricci tensor, null tetrad components of the conform tensor, and the curvature scalar R. The only components which are pertinent to the proofs of our theorems or to the discussion of these theorems are, after setting $v_k = v_t = w_m = w_{t^*} = 0$, as follows[7,9]:

$$R_{ktkt} = R_{mtmt} = 0\qquad(26a)$$
$$R_{mt^*kt} = d_m v_{t^*} - d_{t^*}v_m + v_m v_m^* + v_{t^*}w_t^* + iQ_t^* v_m - P_m v_{t^*},\qquad(26b)$$
$$R_{ktmt^*} = d_k w_t - d_t w_k + w_k w_k^* + w_t v_t^{**} - iQ_t w_k + P_k w_t,\qquad(26c)$$
$$\tfrac{1}{2}S_{tt} = -d_t v_m + (v_m)^2 + iQ_t v_m,\qquad(26d)$$
$$\tfrac{1}{2}S_{tt} = -d_t w_k^* + (w_k^*)^2 + iQ_t w_k^*,\qquad(26e)$$
$$\tfrac{1}{2}S_{kk} = d_k v_{t^*} + (v_{t^*})^2 - P_k v_{t^*},\qquad(26f)$$
$$\tfrac{1}{2}S_{mm} = d_m w_t + (w_t)^2 + P_m w_t,\qquad(26g)$$
$$S_{kt} = -d_k v_m + d_t v_{t^*} - v_m(v_{t^*} - v_t^{**} - iQ_k) - v_{t^*}(v_m - w_k^* + P_t),\qquad(26h)$$
$$S_{mt^*} = -d_m w_k + d_{t^*}w_t - w_k(w_t - w_t^* - iQ_m) + w_t(v_m^* - w_k + P_t^*).\qquad(26i)$$

The expression for $\frac{1}{2}S_{tt}$ in Eq. (26d) derives from a computation of R_{mtkt}, whereas the expression for $\frac{1}{2}S_{tt}$ in Eq. (26e) derives from a computation of R_{ktmt}. The zero values of the conform tensor components in Eqs. (26a) are equivalent to the statement that k and m are principal null directions, a fact which we have already noted. The equal components in Eqs. (26b) and (26c) are each equal to $c_0 + (R/12)$, where c_0 is a conform tensor component.[7]

After substituting from Eqs. (6), (18), and (22)–(24) into the above Eqs. (26), we derive the following key results:

$$\alpha_1\delta_1 = \alpha_2\delta_2 = 0,\qquad(27)$$
$$R_{tt} - R_{tt}^* = 8i\alpha_1\gamma_1,\qquad(28)$$
$$R_{kk} - R_{mm} = -8\alpha_2\gamma_2,\qquad(29)$$

$$R^{\alpha\beta}(\nabla_\alpha \lambda^1)(\nabla_\beta \lambda^2) = 96\epsilon\rho^4\gamma_1\gamma_2(\gamma_1\gamma_2 + \delta_1\delta_2). \quad (30)$$

To derive the above equations, we start by using the fact that the Riemann tensor components in Eqs. (26b) and (26c) are identically equal. Computation of the real and imaginary parts of the difference of these equal components nets

$$-i(d_t - d_{t^*})\delta_1 + 2\alpha_1\delta_1 = 0,$$
$$(d_k - \epsilon d_m)\delta_2 + 2\alpha_2\delta_2 = 0. \quad (31)$$

From the real part of the difference of the two equal components in Eqs. (26d) and (26e),

$$-i(d_t - d_{t^*})\delta_1 - 2\alpha_1\delta_1 = 0. \quad (32a)$$

From the vanishing of the imaginary part of the difference of the two real components in Eqs. (26f) and (26g),

$$(d_k - \epsilon d_m)\delta_2 - 2\alpha_2\delta_2 = 0. \quad (32b)$$

Comparison of Eqs. (32) with (31) then yields Eqs. (27).

Next, Eq. (28) is derived from the imaginary part of the sum of the equal components in Eqs. (26d) and (26e). Equation (29) is derived from the difference of the components in Eqs. (26f) and (26g). Finally, Eq. (30) is derived by using Eqs. (19) to show that

$$R^{\alpha\beta}(\nabla_\alpha \lambda^1)(\nabla_\beta \lambda^2) = 4\epsilon\rho^4\gamma_1\gamma_2(S_{kt} + S^*_{kt} + \epsilon S_{mt} + \epsilon S_{mt^*}),$$

whereupon Eqs. (26h) and (26i) and the reality condition $S^*_{mt} = S_{mt^*}$ do the trick.

The results in Eqs. (27) constitute a remarkable simplification, which will be exploited after we introduce a natural choice of coordinates in the next section.

3. CHOICE OF COORDINATES

It is desirable to give the reader some perspective by first discussing the main results which we are going to derive in this section. After this discussion, the derivations will be outlined.

The 1-forms

$$\omega_x = t + t^*, \qquad \omega_+ = k + \epsilon m,$$
$$\omega_y = -i(t - t^*), \quad \omega_- = k - \epsilon m \quad (33)$$

constitute an orthogonal tetrad, which can be normalized[10] by dividing each tetrad member by $\sqrt{2}$. We will establish the local existence of real valued fields ϕ^r, $E^{(i)}$, $F^{(i)}$, $G_r^{(i)}$ ($i = 1, 2$ and $r = 3, 4$) such that[8]

$$\omega_x = (\rho/2E^{(1)})\,d\lambda^1, \qquad \omega_+ = (\rho/2E^{(2)})\,d\lambda^2, \quad (34a)$$
$$\omega_y = A_r\,d\phi^r, \qquad \omega_- = B_r\,d\phi^r, \quad (34b)$$
$$A_r = \rho\Delta^{-1}F^{(1)}G_r^{(2)}, \qquad B_r = \rho\Delta^{-1}F^{(2)}G_r^{(1)}, \quad (34c)$$
$$\Delta = G_3^{(2)}G_4^{(1)} - G_4^{(2)}G_3^{(1)} > 0, \quad (34d)$$

where summation over r is understood in Eq. (34b), and a parenthesized superscript (i) indicates that the field is expressible as a function of λ^1, λ^2, ϕ^3, ϕ^4 such that

$$\frac{\partial E^{(i)}}{\partial\lambda^j} = \frac{\partial F^{(i)}}{\partial\lambda^j} = \frac{\partial G_r^{(i)}}{\partial\lambda^j} = 0 \quad \text{if } i \neq j.$$

The metric involves only six independent unspecified fields, since it is invariant under the group of substitutions

$$F^{(i)} \to \Psi^{(i)}F^{(i)}, \quad G_r^{(i)} \to \Psi^{(i)}G_r^{(i)},$$

where $\Psi^{(1)}$ and $\Psi^{(2)}$ are any scalar fields which have no zeros in the domain of $(\lambda^1, \lambda^2, \phi^3, \phi^4)$ and which satisfy $\partial\Psi^{(i)}/\partial\lambda^j = 0$ if $i \neq j$.

ϕ^3 and ϕ^4 are arbitrary up to any coordinate transformation which does not involve λ^1 and λ^2. Therefore, for any given point p_0 in the domain of our chart, we can always select (ϕ^3, ϕ^4) such that $G_3^{(2)}G_4^{(1)} > 0$ in some neighborhood of p_0. Then, in that neighborhood of p_0, Eqs. (34b) are expressible in the relatively simple forms

$$\omega_y = \rho(1 - J^{(1)}J^{(2)})^{-1}H^{(1)}(d\phi^3 + J^{(2)}\,d\phi^4),$$
$$\omega_- = \rho(1 - J^{(1)}J^{(2)})^{-1}H^{(2)}(J^{(1)}\,d\phi^3 + d\phi^4), \quad (35)$$

where $1 - J^{(1)}J^{(2)} > 0$, and

$$H^{(1)} = F^{(1)}/G_4^{(1)}, \quad H^{(2)} = F^{(2)}/G_3^{(2)},$$
$$J^{(1)} = G_3^{(1)}/G_4^{(1)}, \quad J^{(2)} = G_4^{(2)}/G_3^{(2)}. \quad (36)$$

The fields $E^{(i)}$, $F^{(i)}$, $G_r^{(i)}$ are not arbitrary. We will derive the following relations which they may must satisfy for $i = 1$, $j = 2$ and for $i = 2$, $j = 1$:

$$G_4^{(i)}\frac{\partial E^{(j)}}{\partial\phi^3} - G_3^{(i)}\frac{\partial E^{(j)}}{\partial\phi^4} = 0, \quad (37)$$

also,

$$\frac{\partial A_3}{\partial\phi^4} - \frac{\partial A_4}{\partial\phi^3} = 0, \quad \frac{\partial B_3}{\partial\phi^4} - \frac{\partial B_4}{\partial\phi^3} = 0. \quad (38)$$

For the general class of metrics defined by Eqs. (1) and (2), Eqs. (37) and (38) are the only constraints on the fields $E^{(i)}$, $F^{(i)}$, $G_r^{(i)}$. The general solution of these equations is not pertinent to our current objective of proving certain theorems. However, one solution will be obtained in the process of constructing our proofs, and our work in progress on the general solution will be discussed in Sec. 5.

Finally, we will derive the following expressions for the coefficients of rotation, where $j = 1, 2$:

$$\gamma_j = \epsilon^{j-1}\rho^{-3}E^{(j)}, \quad (39)$$

$$\alpha_j = \frac{-\epsilon^{j-1}}{\rho E^{(j)}F^{(j)}}\left(G_4^{(j)}\frac{\partial E^{(j)}}{\partial\phi^3} - G_3^{(j)}\frac{\partial E^{(j)}}{\partial\phi^4}\right), \quad (40)$$

$$\beta_j = 2\epsilon^{j-1}(-1)^j E^{(j)}\left(\frac{\Delta}{\rho^2F^{(j)}}\right)\frac{\partial}{\partial\lambda^j}(\Delta^{-1}\rho F^{(j)}); \quad (41)$$

for $i = 1$, $j = 2$ and for $i = 2$, $j = 1$:

$$\delta_i = \frac{-(-\epsilon)^{j-1}E^{(j)}}{\rho\Delta F^{(j)}}\left(G_3^{(j)}\frac{\partial G_4^{(j)}}{\partial\lambda^j} - G_4^{(j)}\frac{\partial G_3^{(j)}}{\partial\lambda^j}\right)F^{(i)}. \quad (42)$$

Also, the following useful relation derives from Eqs. (34d) and (42):

$$\gamma_1\gamma_2 + \delta_1\delta_2 = \rho^4\gamma_1\gamma_2\frac{\partial^2}{\partial\lambda^1\partial\lambda^2}\ln\left(\frac{\rho^2}{\Delta}\right). \quad (43)$$

We start our derivations by computing the following 2-forms from Eqs. (6), (10), (18), (21), and (22):

$$d\omega_x = (\alpha_1\omega_y + \epsilon\gamma_2\omega_+)\omega_x, \quad (44a)$$
$$d\omega_+ = (-\epsilon\alpha_2\omega_- - \gamma_1\omega_x)\omega_+, \quad (44b)$$
$$d\omega_y = (-\beta_1\omega_x + \epsilon\gamma_2\omega_+)\omega_y + (2\epsilon\delta_1\omega_+)\omega_-, \quad (44c)$$
$$d\omega_- = (2\delta_2\omega_x)\omega_y + (-\gamma_1\omega_x + \epsilon\beta_2\omega_+)\omega_-. \quad (44d)$$

Equations (44c) and (44d) and the theorem of Frobenius imply the existence of scalar fields ϕ^r, A_r, B_r ($r=3,4$) such that Eqs. (34b) hold. This is the key idea of our derivation, and this completes the proof of Theorem II as stated in the Introduction.

The rest of the derivation uses the integrability conditions for Eqs. (19) and (34b). Specifically, $d\omega_x$, $d\omega_+$, $d\omega_y$, $d\omega_-$ are computed from Eqs. (19) and (34b) in terms which involve partial derivatives of the rotation coefficients with respect to λ^i and ϕ^r. The results of this computation are equated to their corresponding expressions in Eqs. (44). Six of the differential equations which are thereby obtained are integrated to yield Eqs. (34a), (34c), and (34d), and the remainder are Eqs. (37)–(42).

A detailed comparison[10] of our tetrad in Eqs. (34) with the corresponding orthonormal tetrad of Carter's[1] Hamilton–Jacobi separable space–times demonstrates the following statements:

(1) The set of all of Carter's Hamilton–Jacobi separable space–times for which his U_1 and U_{-1} are not constants is identical with the set of all of our space–times for which

(a) $E^{(1)}$ and $E^{(2)}$ are each independent of ϕ^3 and ϕ^4, and

(b) ϕ^3 and ϕ^4 can be chosen so that $F^{(i)}$ and $G_r^{(i)}$ are each independent of ϕ^3 and ϕ^4 ($i=1, 2$ and $r=3, 4$).

(2) For the aforementioned choice, our ϕ^3 and ϕ^4 are Carter's ϕ^2 and ϕ^{-2}.

Our λ^1 and λ^2 are Carter's U_1 and U_{-1}. Carter leaves the choice of his coordinates λ^1 and λ^{-1} open, and one possible choice (when his U_1 and U_{-1} are not constants) is $U_1=\lambda^1$ and $U_{-1}=\lambda^{-1}$.

(3) Our general ϕ^3, ϕ^4-dependent results, as given by Eqs. (34), has the same form as that of Carter if we let his U_1 and U_{-1} be our λ^1 and λ^2 and if we replace his arbitrary functions of λ^i by functions of λ^i, ϕ^3, ϕ^4 subject only to Eqs. (37) and (38).

The above statements constitute an explication of Theorem III.

We next prove Theorem IV. The premises of the theorem assert the existence of linearly independent commuting Killing vectors ξ^α and ζ^α such that

$$L_\xi K^{\alpha\beta}=L_\zeta K^{\alpha\beta}=0.$$

Since λ^1 and λ^2 are eigenvalues of $K^{\alpha\beta}$, their Lie derivatives with respect to the two Killing vectors are also zero. Therefore, the 2-surfaces of transitivity are the surfaces of constant λ^1 and λ^2. Since we are assuming that $d\lambda^1\wedge d\lambda^2$ is not null, these surfaces of transitivity are not null.

Moreover, the orthogonality of ξ and ζ to $d\lambda^1$ and $d\lambda^2$ implies that $\xi^1=\xi^2=\zeta^1=\zeta^2=0$ relative to the coordinate system $x^1=\lambda^1$, $x^2=\lambda^2$, $x^3=\phi^3$, $x^4=\phi^4$. Therefore, the (i, r) components of the equations

$$L_\xi g^{\alpha\beta}=L_\zeta g^{\alpha\beta}=0$$

yield the statement that ξ^3, ξ^4, ζ^3, ζ^4 are independent of the coordinates $x^1=\lambda^1$ and $x^2=\lambda^2$. (Recall that $g^{ir}=0$.) Therefore, we can now take advantage of the arbitrariness in ϕ^3 and ϕ^4 to select these coordinates so that $\xi^3=\zeta^4=0$ and $\xi^4=\zeta^3=1$. The rest of the proof of Theorem IV presents no difficulty.

4. WHEN HIDDEN SYMMETRIES IMPLY ISOMETRIES

One premise of Theorem V is the statement that the space–time admits a $[(11)(11)]$ Killing tensor with eigenvalues λ^1, λ^2 such that $d\lambda^1\wedge d\lambda^2$ is not null. The only additional premise is the vanishing of the Ricci tensor component $R^{\alpha\beta}(\nabla_\alpha\lambda^1)(\nabla_\beta\lambda^2)$, whereupon Eqs. (20), (30), and (27) imply

$$\delta_1\delta_2\neq 0,\quad \alpha_1=\alpha_2=0.$$

The above results and Eqs. (34d), (37) and (40) imply that $E^{(1)}$ and $E^{(2)}$ are each independent of ϕ^3 and ϕ^4. This completes the first round of our proof.

We next examine the implications of Eq. (42). For convenience, we select the coordinates ϕ^3 and ϕ^4 so that $G_3^{(2)}G_4^{(1)}>0$. Then, the forms given by Eqs. (35) and (36) are applicable, and the fact that $\delta_1\delta_2\neq 0$ and Eq. (42) imply

$$\frac{\partial J^{(j)}}{\partial\lambda^j}\neq 0. \tag{45}$$

Next, we look at Eq. (43) after refreshing our memory with a glance at Eq. (30). Equation (43) and the vanishing of R^{12} imply

$$\frac{\partial^2}{\partial\lambda^1\partial\lambda^2}\left[\ln\left(\frac{1-J^{(1)}J^{(2)}}{\lambda^2-\lambda^1}\right)\right]=0, \tag{46}$$

whose integral has the form

$$1-J^{(1)}J^{(2)}=(\lambda^2-\lambda^1)X^{(1)}X^{(2)} \tag{47}$$

where

$$\frac{\partial X^{(i)}}{\partial\lambda^j}=0\quad\text{if } i\neq j.$$

The general solution of the functional equation (47) subject to the condition (45) is

$$J^{(1)}=\frac{f+g\lambda^1}{h+k\lambda^1},\quad J^{(2)}=\frac{h+k\lambda^2}{f+g\lambda^2},\qquad fk-gh=\pm 1, \tag{48}$$

where f, g, h, k may depend on ϕ^3 and ϕ^4 but are independent of λ^1 and λ^2 and are independent of i. To prove Eqs. (48), take the partial derivative of Eq. (47) with respect to λ^2, and use the result to eliminate $X^{(1)}$ from Eq. (47) and thereby obtain an equation of the form

$$J^{(1)}=(f_1+g_1\lambda^1)(h_1+k_1\lambda^1)^{-1},$$

where f_1, g_1, h_1, k_1 are independent of λ^1. We may replace λ^2 by one of its values in the above expression for $J^{(1)}$, whereupon f_1, g_1, h_1, k_1 depend at most on ϕ^3 and ϕ^4. Since $\partial J^{(1)}/\partial\lambda^1$ cannot vanish, $f_1k_1-g_1h_1\neq 0$, and we can adjust the arbitrary common multiple of these functions to make $|f_1k_1-g_1h_1|=1$.

In continuation of the proof of Eqs. (48), a similar argument gives us $J^{(2)}=(h_2+k_2\lambda^2)(f_2+g_2\lambda^2)^{-1}$, where f_2, g_2, h_2, k_2 depend at most on ϕ^3 and ϕ^4, and $|f_2k_2-g_2h_2|=1$. The substitution of these expressions for $J^{(1)}$ and

$J^{(2)}$ back into the differential equation (46) and a straightforward analysis nets $f_1=f_2$, $g_1=g_2$, $h_1=h_2$, $k_1=k_2$.

That completes the proof of Eqs. (48). It is now useful to take inventory before the next step in the proof of Theorem V. Substitution of Eqs. (48) into Eqs. (35) yields the forms

$$\omega_y=\rho^{-1}K^{(1)}[(f+g\lambda^2)\,d\phi^3+(h+k\lambda^2)\,d\phi^4],$$
$$\omega_-=\rho^{-1}K^{(2)}[(f+g\lambda^1)\,d\phi^3+(h+k\lambda^1)\,d\phi^4], \tag{49}$$

where $K^{(1)}=\mp(h+k\lambda^1)H^{(1)}$ and $K^{(2)}=\mp(f+g\lambda^2)H^{(2)}$. From the above Eqs. (49), we can read off the expressions for the fields A_r and B_r, which are defined by Eq. (34b).

There is only one equation left to satisfy, and that is Eq. (38). Upon substituting from Eqs. (49) into (38) and using the linear independence of the terms of the zeroth and first degrees in λ^i, we obtain for $i=1,2$

$$f\frac{\partial K^{(i)}}{\partial\phi^4}+\frac{\partial f}{\partial\phi^4}K^{(i)}=h\frac{\partial K^{(i)}}{\partial\phi^3}+\frac{\partial h}{\partial\phi^3}K^{(i)},$$
$$g\frac{\partial K^{(i)}}{\partial\phi^4}+\frac{\partial g}{\partial\phi^4}K^{(i)}=k\frac{\partial K^{(i)}}{\partial\phi^3}+\frac{\partial k}{\partial\phi^3}K^{(i)}. \tag{50}$$

From the above equations and the relation $fk-gh=\pm1$, it is a simple matter to show that there exists a scalar field Φ which depends at most on ϕ^3 and ϕ^4, and there exist scalar fields $L^{(1)}$ and $L^{(2)}$ which depend only on λ^1 and λ^2 respectively, such that

$$K^{(i)}=\Phi L^{(i)}\quad(i=1,2). \tag{51}$$

Substitution of this result back into Eqs. (50) yield

$$\frac{\partial(\Phi f)}{\partial\phi^4}-\frac{\partial(\Phi h)}{\partial\phi^3}=\frac{\partial(\Phi g)}{\partial\phi^4}-\frac{\partial(\Phi k)}{\partial\phi^3}=0.$$

So, there exist scalar fields Φ^3 and Φ^4 such that

$$\Phi f=\frac{\partial\Phi^3}{\partial\phi^3},\quad \Phi h=\frac{\partial\Phi^3}{\partial\phi^4},\quad \Phi g=\frac{\partial\Phi^4}{\partial\phi^3},\quad \Phi k=\frac{\partial\Phi^4}{\partial\phi^4}. \tag{52}$$

Substitution of Eqs. (51) and (52) back into Eqs. (49) give us our final results

$$\omega_y=\rho^{-1}L^{(1)}(d\Phi^3+\lambda^2\,d\Phi^4),$$
$$\omega_-=\rho^{-1}L^{(2)}(d\Phi^3+\lambda^1\,d\Phi^4). \tag{53}$$

That completes the proof of Theorem V, because Φ^3 and Φ^4 are ignorable coordinates, i.e., the four remaining unspecified fields $E^{(i)}$ and $L^{(i)}$ $(i=1,2)$ are independent of Φ^3 and Φ^4. Moreover, Theorem VI now follows immediately from the fact that our Eq. (47) is equivalent to the following statement when there is independence of Φ^3 and Φ^4:

$$\rho^{-2}\Delta=(\text{function of }\lambda^1)\times(\text{function of }\lambda^2).$$

This statement is precisely Carter's necessary and sufficient condition for his Hamilton–Jacobi separable metric to be Schrödinger separable.[1]

5. PERSPECTIVES

Our key result, Theorem V, was originally suggested to us by a Newtonian couterpart. Though the Newtonian analogue is an imperfect reflection of our result and though it has probably been noted before in the long history of classical mechanics, it is sufficiently instructive to make its presentation here worthwhile. We first review the role of angular momentum in a force field derivable from a scalar potential V and a vector potential $\mathbf{A}$.

Consider any nonrelativistic test particle whose Hamiltonian has the form

$$H=\tfrac{1}{2}(\mathbf{p}-\mathbf{A})^2+V,$$

where V and $\mathbf{A}$ are time-independent scalar and vector functions of the radius vector $\mathbf{x}$. It is common knowledge that the statement that $\mathbf{l}=\mathbf{x}\times\mathbf{p}$ is a constant of the motion along all possible orbits consistent with the above Hamiltonian (analogous to specification of geodesic orbits in general relativity) is equivalent to the statement that V depends only on $r=|\mathbf{x}|$ and that $\mathbf{A}$ have the form $\mathbf{A}=\mathbf{x}f$ where f is also a function of r. Since $\mathbf{x}f$ is expressible as the gradient of a scalar field, a gauge transformation can remove this term from the scene, and we are left with $\mathbf{A}=0$ and $V=$ a central potential.

Now, suppose that the sources of the force field are such that $\mathbf{l}$ is no longer a constant of the motion, but $\mathbf{l}^2$ is still a constant of the motion. Then, the theory of Poisson brackets implies that

$$V=U(r)-\tfrac{1}{2}\mathbf{A}^2,\quad \mathbf{A}=\mathbf{x}f(r)+x\times\omega(r), \tag{54}$$

where $U(r)$, $f(r)$, and $\omega(r)$ depend only on r. Again, the term $\mathbf{x}f$ is removed by a gauge transformation. The assumption that $\mathbf{l}^2$ is a constant of the motion is a viable Newtonian counterpart of the statement that a $[(11)(11)]$ Killing tensor exists, though it is admittedly not the most general assumption of this kind. The dynamical conclusion expressed by Eqs. (54) is the corresponding analog of the metrical structure specified by Theorems II and III.

Next, suppose we assume that the entire source of the field $\mathbf{A}$ is in a bounded region of space and that we are external to this region. We interpret this assumption as meaning

$$\nabla^2\mathbf{A}=0,\quad \nabla\times\mathbf{A}\to0\ \text{ as } r\to\infty, \tag{55}$$

in that gauge where the central field $\mathbf{x}f$ is absent from $\mathbf{A}$. Then, $\omega(r)=\boldsymbol{\alpha}r^{-3}$, where $\boldsymbol{\alpha}$ is a constant axial vector parameter. The resulting Hamiltonian is

$$H=\tfrac{1}{2}\mathbf{p}^2+U(r)+\mathbf{l}\cdot\boldsymbol{\alpha}r^{-3}, \tag{56}$$

which is axially symmetric about $\boldsymbol{\alpha}$.

That is our rough Newtonian analog of Theorem V.[11] Of course, if V and A are electromagnetic potentials, $\mathbf{x}\times\boldsymbol{\alpha}r^{-3}$ is the vector potential due to a magnetic dipole source. It may also represent the analogous term due to a spinning mass in a post-Newtonian theory.

Theorem V has slightly embarrassed the authors, because we had contemplated the possibility of searching for those axially symmetric non-stationary vacuum solutions which admit a $[(11)(11)]$ Killing tensor. If we are to find a result which reduces to the Kerr metric as some physical parameter is turned off, then we must also assume that $d\lambda^1\wedge d\lambda^2$ is not null. Theorem V has nullified this class of solutions.

One possible alternative seems to be to look at space—times which admit Killing tensors with more than two distinct eigenvalues. This remains a relatively unexplored territory.

That brings us to the question of what remains to be done on $[(11)(11)]$ Killing tensors. Analysis of Eq. (38) is in progress, and we hope to arrive at a full solution of the problem corresponding to nonzero $(d\lambda^1 \wedge d\lambda^2) \cdot (d\lambda^1 \wedge d\lambda^2)$. It should be mentioned that we have constructed a solution which admits only one Killing vector which commutes with $K_{\alpha\beta}$, and we have constructed another which admits no Killing vector which commutes with $K_{\alpha\beta}$.

There are also the problems involving the cases for which $(d\lambda^1 \wedge d\lambda^2) \cdot (d\lambda^1 \wedge d\lambda^2) = 0$. Dietz[4] has included these cases in his recent work on axially symmetric stationary space—times which admit a $[(11)(11)]$ Killing tensor. However, the question of what happens when no isometries are assumed is still open, as far as we know.

Finally, and of greater physical interest than the preceding problems, is the study of the *interior solutions* for space—times admitting a $[(11)(11)]$ Killing tensor. This is still an open field.

ACKNOWLEDGMENTS

We thank Werner Dietz for preprints of his work on $[(11)(11)]$ Killing tensors. Also, we thank Fred Ernst for his encouragement and many helpful discussions.

[1]B. Carter, Commun. Math. Phys. 10, 280 (1968).

[2]Our notation for the null tetrad is used, e.g., by R.K. Sachs, in *Relativity, Groups and Topology*, edited by C. deWitt and B. deWitt (Gordon and Breach, New York, 1964), p. 530. Our signature is +2.

[3]For vacuums, therefore, the Goldberg—Sachs theorem implies that the conform tensor is type D, for which the conclusion of Theorem V is well known.

[4]W. Dietz (to be published).

[5]L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commn. Math. Phys. 27, 303 (1972); L. P. Hughston and P. Sommers, Commn. Math. Phys. 32, 147 (1973); 33, 129 (1973).

[6]See, e.g., L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N.J., 1926, 1960 Printing), pp. 96—142.

[7]For a description of our notations and the null tetrad formalism which we use, see Frederick J. Ernst, J. Math. Phys. **15**, 1409 (1974). Our P and Q are his $(u+u^*)/2$ and $(u-u^*)/2i$ respectively.

[8]Since $K_{\alpha\beta}$ is arbitrary up to multiplication by any nonzero real number and up to addition to any constant multiple of $g_{\alpha\beta}$, the eigenvalues have the corresponding arbitrariness $\lambda^i \rightarrow \alpha\lambda^i + \beta$, where $\alpha \neq 0$ and β are real parameters. For the Kerr metric and an appropriate choice of $K_{\alpha\beta}$, the square of ρ becomes $\rho^2 = r^2 + (a\cos\theta)^2$, where $\lambda^1 = -(a\cos\theta)^2$ and $\lambda^2 = (r)^2$ in terms of conventional notations. For fixed $K_{\alpha\beta}$, our results as given by Eqs. (33)—(43) and Eqs. (53) are applicable, as they stand, to regions in which $\rho^2 > 0$. If ρ is replaced by $|\rho|$, these equations become applicable to regions in which $\rho^2 < 0$.

[9]We have also found that $R_{kmkt} + R_{tt*kt} - \epsilon\ (R_{kmmt} - R_{tt*mt})$ is real.

[10]We did not normalize these 1-forms to unity, because we wanted to avoid a plethora of $\sqrt{2}$ factors in our calculations.

[11]It is not a satisfactory analog, since the condition $\nabla^2 A = 0$ is too stringent [compared with Eq. (3)].

Number of independent missing label operators*

A. Peccia and R. T. Sharp

Physics Department, McGill University, Montreal, Canada
(Received 8 March 1976)

When the bases for irreducible representations of a semisimple group are reduced according to a semisimple subgroup the number of functionally independent missing label operators available is just twice the number of missing labels. The argument presented suggests that the result holds for any Lie group.

1. INTRODUCTION

In the application of a Lie group to a physical problem it is usually desirable to reduce the irreducible representations (IR's) of the group into IR's of a subgroup. Often the subgroup does not provide enough labels to specify the basis states uniquely.

One resolution of this difficulty is to use as bases the common eigenstates of a complete set of commuting operators. Besides the generalized Casimir invariants of group and subgroup an appropriate number of "missing label" operators must be found. They should be subgroup scalars which are functions of the group generators. We shall show, for semisimple groups, that the number of functionally independent missing label operators available is just twice the number of labels actually missing. The argument used strongly suggests that the same result holds for all Lie groups.

Numerous examples of the result are known. In the case of $SU(3)\supset O(3)$ there is one missing label. Long ago Racah[1] proposed two independent missing label operators; he surmised and Judd, Miller, Patera, and Winternitz[2] later proved that they constitute the most general solution of the missing label problem. Recently, a similar result was shown to hold[3,4] for all cases of compact group—subgroups with one missing label. Two cases with two missing labels have been investigated, $SU(4)\supset SU(2)\times SU(2)$ (the Wigner supermultiplet model[5]) and $O(5)\supset O(3)$; in each case four missing label operators are available.

2. COUNTING SUBGROUP SCALARS

The invariants of semisimple groups were determined long ago.[6] Their number is l, the rank of the group, and they may be chosen to be homogeneous polynomials in the generators (i.e., Casimir operators). Their eigenvalues are necessary and sufficient to label unambiguously the IR's of the group (we ignore the possibility of using irrational combinations of Casimir operators; in any case this artificial way of reducing the number of labeling operators works only when, because of compactness, the eigenvalues are restricted to a discrete set). In order that our considerations apply to nonsemisimple groups, we make the plausible assumption[7,8] that the eigenvalues of the independent invariant functions of the generators unambiguously label the IR's of any Lie group (we ignore the possible existence of labels which assume only a finite number of values, such as the sign of the energy in the case of the Poincaré group). For nonsemisimple groups the invariants are not necessarily polynomials, but may be rational or even transcendental functions of the generators (generalized Casimir operators). They are determined[7,9,10] by solving a set of partial differential equations. Each equation corresponds to a row of the generator commutator table. The method of derivation shows that the number of invariants is[10]

$$l=r-R, \tag{1}$$

where r is the order of the group and R is the rank of the generator commutator table, considered as a matrix; for the purpose of computing R the generators are to be regarded as independent c-number variables.

A theorem due to Racah[1] shows that the number of internal labels needed to label unambiguously the basis states of the general IR of a Lie group is

$$i=\tfrac{1}{2}(r-l). \tag{2}$$

When a subgroup H is used to label basis states of IR's of a Lie group G, it provides $\tfrac{1}{2}(r_H+l_H)-l'$ labels. The appearance of l' allows for the possibility that the generalized Casimir operators of H and G may not be mutually independent; l' is the number of invariants of G which depend only on generators of H. The number of missing labels is thus

$$n=\tfrac{1}{2}(r_G-l_G-r_H-l_H)+l'. \tag{3}$$

Subgroup scalars which are functions of group generators may be determined, just as group invariants, by solving a set of partial differential equations. The equations are those corresponding to the first r_H rows (i.e., to the subgroup generators) of the group commutator table. The method of derivation shows that the number of subgroup scalar is r_G-R'. The arguments parallel those leading to (1). Here R' is the rank of the first r_H rows of the group commutator table. Subtracting the number l_G+l_H-l' of group and/or subgroup generalized Casimir invariants we find the number of available missing label operators to be

$$m=r_G-R'-l_G-l_H+l'. \tag{4}$$

We want to show that $m=2n$, i.e., that

$$l'=r_H-R'. \tag{5}$$

But Eq. (5) follows immediately from the definition of l'. Group scalars which depend only on subgroup generators are found by solving the r_G partial differential equations corresponding to the first r_H columns of the group commutator table. The number of solutions is just Eq. (5). Because of antisymmetry the first r_H columns have the same rank R' as the first r_H rows.

We complete this section by showing that for semisimple groups $R' = r_H$ (and hence $l' = 0$). We ignore the trivial possibility that $G = G' \otimes G''$ and $H = H' \otimes G''$; then G'' plays no role in the labeling problem and $l' = l_G''$. We may choose the generators of H in canonical fashion so that the first $r_H - l_H$ are Hermitian conjugate pairs and the last l_H are the weights (the Cartan subalgebra). The other generators of G may be taken to be irreducible tensor components with respect to H; our ignoring of the trivial case mentioned above implies that these tensors contain components whose weights span all directions in H-weight space.

To show that $R' = r_H$, we exhibit an $r_H \times r_H$ submatrix of the first r_H rows whose determinant does not vanish. The only elements of the first r_H rows which depend on diagonal (weight) generators of H are those at the intersection of a row and column corresponding to conjugate root generators of H. Such an element is[1] $\sum_i \alpha_j^i H_i$, where H_i are the diagonal generators and α_j^i are the weight components of the root in question (the jth). Ignoring the other elements of the first $r_H - l_H$ rows and columns (they cannot cancel the ones retained) we find the value of this subdeterminant to be $\prod_j^+ (\sum_i \alpha_j^i H_i)^2 \neq 0$. To complete the argument we need to choose l_H more columns whose intersection with rows $r_H - l_H + 1$ to r_H have nonzero determinant; this is easily done by choosing l_H columns corresponding to tensor components with l_H independent H weights.

3. EXAMPLES

As examples involving nonsemisimple Lie groups we consider two cases, one a group of order three, the other of order four. They are taken from Mubarakzyanov's[11] complete list of real algebras of dimension up to five. In Ref. 7 they are designated $A_{3,1}$ and $A_{4,7}$.

(a) $A_{3,1}$. This is the algebra of the quantum mechanical (Weyl) group in one dimension. The only nonzero commutator is $[e_2, e_3] = e_1$. There is one invariant operator e_1. Hence $r_G = 3$, $l_G = 1$. We consider the one-dimensional subgroup e_1. Then $r_H = l_H = l' = 1$. According to (3) there is one missing label. Hence there are two available missing label operators, which may be taken as e_2 and e_3.

(b) $A_{4,7}$. The nonzero commutators are $[e_2, e_3] = e_1$, $[e_1, e_4] = 2e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = e_2 + e_3$. There are no invariant operators, so we have $r_G = 4$, $l_G = 0$.

For each of the one-dimensional subalgebras e_1, e_2, e_3, e_4 we have $r_H = l_H = 1$, $l' = 0$, and hence one missing label. In each case there are two available missing label operators, as follows:

e_1: e_2 and e_3; e_2: e_1 and $e_1 e_4 - e_2 e_3$;

e_3: e_1 and $e_2^2 + 2e_1 e_4 + 2e_2 e_3$;

e_4: e_1, and $e_2 \exp[-e_3/e_2]$.

For each of the two-dimensional subalgebras $e_1 e_4$ and $e_2 e_4$ we have $r_H = 2$, $l_H = 0$, $l' = 0$, and hence one missing label. For $e_1 e_4$ the two missing label operators are e_2^2/e_1 and $e_2 \exp[-e_3/e_1]$; for $e_2 e_4$ they are e_2^2/e_1 and $e_2 \exp[(e_1 e_4 - e_2 e_3)/e_2^2]$.

4. DISCUSSION

The examples of Sec. 3 show that the method of partial differential equations is a practical way of determining missing label operators for Lie groups. It could be used to determine systematically such operators for group—subgroup combinations of interest in physics. The theorem of Sec. 2 shows that if there are n missing labels, there are $2n$ functionally independent missing label operators; in each case, when $n > 1$, there remains the problem of choosing a set of n mutually commuting functions of the $2n$ operators.

In many practical problems the number of missing labels, and of missing label ooperators, is reduced by restrictions on the group IR's being considered. Thus $O(5) \supset O(3)$ has two missing labels and four missing label operators for the general IR of O(5). But if attention is restricted to IR's of the form $(0, \lambda)$ (such states are needed in connection with nuclear quadrupole vibrations[12] and the Jahn—Teller effect[13]) there is only one missing label. We are trying to determine the missing label operators in this and other similar cases; the method described in this paper for counting and determining missing label operators is not directly applicable.

ACKNOWLEDGMENTS

We are grateful to J. Patera and P. Winternitz for helpful discussions.

*Supported in part by the Quebec Education Department and the National Research Council of Canada.

[1] G. Racah, lecture notes, "Group Theory and Spectroscopy," Institute for Advanced Study, Princeton, N.J. (1951); reprinted in Ergeb. Exakten Naturwiss. 37, 28 (1965).

[2] B.R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. 15, 1787 (1974).

[3] R.T. Sharp, J. Math. Phys. 16, 2050 (1975).

[4] J.D. Louck, Proceedings of the Colloquium on Mathematical Physics, Mexico City (1976). Louck points out a case overlooked in Ref. 3, SU(2) ×SU(2) ×SU(2) ⊃ SU(2). The two functionally independent missing label operators, in an obvious notation, are J_{12}^2 and J_{13}^2.

[5] W. Miller, Jr., private communication.

[6] G. Racah, Rend. Lincei 8, 108 (1950).

[7] J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 986 (1976).

[8] M. Perroud, to be published.

[9] E.G. Beltrametti and A. Blasi, Phys. Lett. 20, 62 (1966); M. Pauri and G.M. Prosperi, Nuovo Cimento 43, 533 (1966).

[10] L. Abellanos and L.M. Alonso, J. Math. Phys. 16, 1580 (1975).

[11] G.M. Mubarakzyanov, Izv. Vysshikh Uchebn. Zavedeniĭ Mat. No. 1 (32), 114; No. 3 (34), 99; No. 4 (35), 104 (1963).

[12] K.T. Hecht, Nucl. Phys. A 102, 11 (1967).

[13] B.R. Judd, Can. J. Phys. 52, 999 (1974).

Hertzian gravitational potentials for type D space–times

R. Güven*

Department of Physics, Middle East Technical University, Ankara, Turkey
(Received 29 December 1975; resubmitted 20 February 1976)

Hertzian gravitational potentials for type D vacuum space–times are defined. The Geroch–Held–Penrose formalism is expressed in the language of exterior forms, and by using its formal simplifications, the existence of Hertzian potentials for this class of nonlinear gravitational fields is demonstrated. The explicit form of these potentials possessing the spin and boost weights and the self-dual property of the field 2-form is constructed. The gauge freedom associated with these potentials is discussed, and the procedure is found to resemble the Debye reduction of the electromagnetic Hertzian potentials in that it enables us to construct a complex scalar superpotential from the Weyl spinor.

I. INTRODUCTION

In Maxwell's theory of electrodynamics the utility of defining bivector potentials has been recognized since the time of Hertz. It was observed that the integration of Maxwell's equations in the absence of sources could be reduced to the determination of a Hertzian bivector potential, the divergence of which gives the usual vector potential. When source distributions are present, one may also introduce stream potentials which are again bivectors, not necessarily satisfying the homogeneous field equations, but related to the 4-current in a similar fashion. The introduction of these potentials leads to a large freedom of gauge which can be exploited to exhibit the intrinsic degrees of freedom of the electromagnetic field. A well-known example is the characterization of the source-free Maxwell field in terms of two scalars satisfying the homogeneous wave equation. A detailed account of these potentials and the associated gauge freedom can be found in Nisbet.[1] Recently, it was shown that Hertzian potential formalism is convenient for integrating the Maxwell's equations in curved space as well.[2–4] In particular, Cohen and Kegeles[4] have given a covariant generalization of the procedure and constructed source-free electromagnetic fields in the test field approximation for a wide class of algebraically special background space—times.

The introduction of Hertzian potentials for the gravitational field itself has previously been dealt with only in the framework of the linearized theory of gravitation where a situation entirely analogous to the electromagnetic field exists. This was first observed by Sachs and Bergmann,[5] who constructed a supermetric with the symmetries of the conformal Weyl tensor as the Hertzian potential for the linearized gravitational field. Campbell and Morgan[6] have obtained Debye potentials which are particular Hertzian potentials for the linearized gravitational field. Penrose[7] has shown that for any spin-s, zero rest-mass field in flat space a Hertzian potential can be introduced, the tensor translation of which has the same symmetries and the number of indices as the field tensor. Quite recently Cohen and Kegeles[8] have announced an extension of the potentials of Penrose and reported Debye potentials for linearized gravitational perturbations in a curved space—time.

The problem which we shall consider in this paper is the introduction of Hertzian potentials for nonlinear gravitational fields themselves. We shall carry out such a program only for a class of nonlinear gravitational fields. By employing the Geroch—Held—Penrose[9] (GHP) formalism we shall establish that, for type D vacuum space—times, bivector potentials can fruitfully be introduced and provide all the necessary information about the gravitational field.

In Sec. II we give a review of the GHP formalism where the reduction of the connection and the resulting GHP covariant derivative are considered. The GHP set of equations are written in terms of exterior differential forms and the gravitational field is represented by curvature 2-forms. In Sec. III we consider the specialization of the formalism to type D vacuum space—times. We regard the connection 1-form (and its primed companion) as the gravitational analog of the electromagnetic potential 1-form. The Hertzian potential is then defined as the 2-form whose GHP covariant divergence gives the connection 1-form. We construct the explicit form of the Hertzian potentials that have the same symmetries of the curvature 2-forms. We discuss the gauge freedom associated with these potentials and identify a complex scalar superpotential as the gravitational Debye potential for type D space—times. The definition of Hertzian potentials for the Coulomb field which is the electromagnetic analog of our problem is discussed in the Appendix.

II. GHP FORMALISM

Consider a space—time M possessing a spinor structure.[10] The collection of all spin frames (o^A, ι^A) at all points of M is a principal fibre bundle Sp with structure group $SL(2, C)$. At each point of M the pair (o^A, ι^A) defines a null tetrad $(\mathbf{l}, \mathbf{n}, \mathbf{m}, \bar{\mathbf{m}})$, where $\mathbf{l}$ and $\mathbf{n}$ are real, $\mathbf{m}$ is complex and bar denotes complex conjugation. The basis vectors are normalized according to the condition $o_A \iota^A = 1$ and satisfy $\mathbf{l}\cdot\mathbf{n} = -\mathbf{m}\cdot\bar{\mathbf{m}} = 1$ with all other scalar products vanishing.

The differential geometry of the Newman—Penrose[11] (NP) formalism has been studied by Nutku,[12] who has observed that the connection on Sp can be described by the matrix Γ of complex-valued 1-forms

$$\Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_2 & -\gamma_0 \end{pmatrix}, \tag{1}$$

whose elements are given by

$$\gamma_0 = \gamma l + \epsilon n - \alpha m - \beta \bar{m},$$
$$\gamma_1 = -\tau l - \kappa n + \rho m + \sigma \bar{m}, \tag{2}$$
$$\gamma_2 = \nu l + \pi n - \lambda m - \mu \bar{m},$$

where l, n, m, $\bar{m}$ are the basis 1-forms dual to the vectors l, n, m, $\bar{\text{m}}$, respectively. The coefficients of the basis 1-forms are the NP spin coefficients. The Ricci identities are the defining relations of the curvature matrix of 2-forms[12]

$$R = d\Gamma - \Gamma^2, \tag{3}$$

and the Bianchi identities are expressed by

$$dR = \Gamma R - R\Gamma. \tag{4}$$

In these relations d denotes the exterior derivative and in the products matrix-exterior multiplications are to be understood.

Under a change of spin frame $(o^A, \iota^A) \to (o^A, \iota^A) S^T$ with $S \in SL(2, C)$ the tetrad transforms as

$$\sigma \to S\sigma S^\dagger, \tag{5}$$

where σ denotes the Hermitian matrix of basis 1-forms

$$\sigma = \begin{pmatrix} l & m \\ \bar{m} & n \end{pmatrix}. \tag{6}$$

The operations $\dagger$ and T are, respectively, the Hermitian conjugation and the transposition. The transformation law for the connection Γ and the curvature R for this change of frame is given by

$$\Gamma \to S\Gamma S^{-1} + dSS^{-1}, \tag{7}$$
$$R \to SRS^{-1}. \tag{8}$$

GHP formalism is a particular version of the NP formalism adapted to space—times in which two future pointing null directions are geometrically singled out. It rests on the observation that for these space—times the tetrad formalism can be reduced to a gauge invariant calculus with suitable operations defined in the graded algebra generated by a class of tetrad scalars.[13] Let (L, N) denote the pair of null direction fields. By choosing $\text{l} \in L$ and $\text{n} \in N$ at each point of M the allowable changes of spin frame will be restricted to the one parameter subgroup of $\text{SL}(2, C)$ which preserves these null directions. The typical element of this subgroup can be written as

$$S = \begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}, \tag{9}$$

and with the above choice of the tetrad the structure group will be reduced to the multiplicative group C^* of complex numbers. Hence it will be sufficient to consider only $B \subset \text{Sp}$ the principal fibre bundle with Abelian structure group C^*.

After this reduction all the scalars associated with the tetrad will fall into two categories. The interesting tetrad scalars will be the ones that undergo a transformation of the form

$$\eta \to z^p \bar{z}^q \eta, \tag{10}$$

where p and q are integers. We then call η to be a spin and boost weighted scalar of type (p, q). The scalars of a given type (p, q) form a vector space over the field of complex numbers and the string of all such vector spaces forms a graded algebra U. Thus, if η_1 is of the type (p_1, q_1) and η_2 is of the type (p_2, q_2), then $\eta_1\eta_2$ will be of the type $(p_1 + p_2,\ q_1 + q_2)$, the expression $\eta_1 + \eta_2$ will have a meaning only when $p_1 = p_2$, $q_1 = q_2$, and the sum will again be of the same type. Clearly the same distinction applies to exterior forms which are constructed from $l, n, m, \bar{m}$ together with the tetrad scalars. The exterior forms that have good spin and boost weights will obey the same product rule under exterior multiplication.

Returning back to the connection 1-forms, when S is given as in Eq. (9), the explicit form of the transformations (7) will be

$$\gamma_1 \to z^2 \gamma_1, \tag{11}$$
$$\gamma_2 \to z^{-2} \gamma_2, \tag{12}$$

and

$$\gamma_0 \to \gamma_0 + \frac{dz}{z}. \tag{13}$$

Thus, the types of γ_1 and γ_2 are respectively $(2, 0)$ and $(-2, 0)$. On the other hand, γ_0 does not have a definite type but defines a connection on B. The role played by γ_0 can be seen most easily by writing Eq. (3) in component form

$$\Theta_0 = d\gamma_0 - \gamma_1 \wedge \gamma_2, \tag{14}$$
$$\Theta_1 = d\gamma_1 - 2\gamma_0 \wedge \gamma_1, \tag{15}$$
$$\Theta_2 = d\gamma_2 + 2\gamma_0 \wedge \gamma_2, \tag{16}$$

where the 2-forms Θ_0, Θ_1, and Θ_2 denote the corresponding entries of the curvature matrix R and they are of the types $(0, 0)$, $(2, 0)$, $(-2, 0)$ respectively. The operation $\wedge$ denotes exterior multiplication. By inspecting Eqs. (15) and (16) one is led to define a covariant differential operator $\not{d}$ on the quantities that have definite types by

$$\not{d}\omega = d\omega - p\gamma_0 \wedge \omega - q\bar{\gamma}_0 \wedge \omega, \tag{17}$$

where ω is an arbitrary r-form of the type (p, q). Clearly $\not{d}$ is a derivation which maps ω to an $(r+1)$-form of the same type (p, q). The information contained in Eq. (14) will be incorporated as the integrability condition for the operator $\not{d}$.

At this stage it will be convenient to introduce the prime symmetry defined by

$$l' = n, \quad m' = \bar{m}, \quad n' = l, \quad \bar{m}' = m, \tag{18}$$

under which the set of NP spin coefficients is mapped onto itself. One easily verifies that $\gamma_2 = \gamma_1'$ and $\gamma_0 = \gamma_0'$. Furthermore, the GHP operator $\not{d}$ is invariant under the prime operation, and we shall have $\Theta_2 = \Theta_1'$, $\Theta_0 = \Theta_0'$. Therefore, from now on we shall omit the subscripts from γ_1 (Θ_1) and write it as γ (Θ). This will lead to no confusion with the spin coefficient γ. In general any quantity ω of the type (p, q) will be transformed by the prime operation to a quantity ω' of the type $(-p, -q)$ and will satisfy the properties

$$(\bar{\omega})' = \overline{(\omega')}, \quad (\omega')' = (-1)^{p+q}\omega. \tag{19}$$

Utilizing the prime symmetry, we now express the integrability condition of the operator $\not{d}$ by the relation

$$\not{d}\not{d}\omega = -p(\gamma \wedge \gamma' + \Theta_0) \wedge \omega - q(\bar{\gamma} \wedge \bar{\gamma}' + \bar{\Theta}_0) \wedge \omega. \qquad (20)$$

The curvature 2-form Θ_0 will explicitly be given by

$$\Theta_0 = (\Lambda - \Phi_{11} - \Psi_2) l \wedge l' + \Psi_3 l \wedge m + \Phi_{12} l \wedge m' - \Phi_{10} l' \wedge m - \Psi_1 l' \wedge m' + (\Psi_2 - \Phi_{11} - \Lambda) m \wedge m'.$$

where the coefficients are composed of curvature scalar and the components of trace-free Ricci and Weyl spinors.[9]

With the help of covariant operator $\not{d}$ the Ricci identities become expressible in the compact form

$$\Theta = \not{d}\gamma, \qquad (22)$$

where the curvature 2-form Θ is given by

$$\Theta = (\Psi_1 + \Phi_{01}) l \wedge l' - (\Psi_2 + 2\Lambda) l \wedge m - \Phi_{02} l \wedge m' + \Phi_{00} l' \wedge m + \Psi_0 l' \wedge m' + (\phi_{01} - \Psi_1) m \wedge m', \qquad (23)$$

and Eq. (16) will be seen to be the primed version of Eq. (22). Bianchi identities will be just the consequence of the general relation Eq. (20) when it is applied to γ. To express the contracted Bianchi identities, we define the currents

$$*j_1 = \Theta_0 \wedge l + \Theta \wedge m', \qquad *j_2 = \Theta_0 \wedge m + \Theta \wedge l', \qquad (24)$$

which will vanish for vacuum space—times. The operation $*$ denotes the Hodge star dual. Contracted Bianchi identities then take the form

$$\not{d}{*j_1} = \gamma \wedge *j_2' + \bar{\gamma} \wedge *j_2, \qquad \not{d}{*j_2} = \gamma \wedge *j_1' + \bar{\gamma}' \wedge *j_1, \qquad (25)$$

and if the primed sets are also taken into account we exhaust all the GHP set of equations. Finally we note that all the operations can be effected in the graded algebra U of spin and boost weighted scalars by defining the directional derivatives þ, þ′, ð, ð′ through the relation

$$\not{d} = l\text{þ}' + l'\text{þ} - m\eth' - m'\eth, \qquad (26)$$

and by considering the action of the operator $\not{d}$ on the basis 1-forms

$$\not{d}l = \gamma \wedge m' + \bar{\gamma} \wedge m, \qquad \not{d}m = \gamma \wedge l' + \bar{\gamma}' \wedge l, \qquad (27)$$

which will, of course, give us the set of scalar GHP equations.

III. HERTZIAN POTENTIALS FOR TYPE D SPACE-TIMES

In this section we shall consider the type D vacuum space—times. The GHP formalism is a very convenient tool for the study of these space—times because the Weyl tensor provides us with two degenerate principal null directions. We shall choose the basis vectors **l** and **n** to lie along these null directions so that

$$\sigma = \sigma' = \kappa = \kappa' = 0, \qquad \Psi_0 = \Psi_0' = \Psi_1 = \Psi_1' = 0, \qquad \Psi \equiv \Psi_2 \neq 0, \qquad (28)$$

and considerable simplifications in the expressions of Sec. II will result. The connection 1-form γ will be given by

$$\gamma = -\tau l + \rho m, \qquad (29)$$

and the type D gravitational fields will be represented by the curvature 2-form

$$\Theta = -\Psi l \wedge m \qquad (30)$$

This field 2-form Θ possesses the self-dual property

$$*\Theta = -i\Theta, \qquad (31)$$

which is invariant under the prime operation and, in fact, valid for all space—times with vanishing Ricci spinor Φ_{ij}.

We shall regard the connection γ (with its primed version) as the potential 1-form of the theory. To construct the Hertzian potentials for type D space—times, we shall look for a 2-form P whose covariant divergence will give us the gravitational potential 1-form

$$\gamma = \not{\delta} P, \qquad (32)$$

where we have introduced the co-derivative defined by $\not{\delta} \equiv *\not{d}*$. The co-derivative is, of course, invariant under the prime operation which we shall freely use. We note that the connection γ as given by Eq. (29) satisfies the "Lorentz condition"

$$\not{\delta}\gamma = 0. \qquad (33)$$

Consequently, the Hertzian potential P will be a 2-form of the type (2, 0) and will satisfy Eq. (33). We also require P to be self-dual as this is a symmetry of the field 2-form itself. These considerations lead us to the choice

$$P = \Sigma l \wedge m, \qquad (34)$$

for the Hertzian potential P where Σ is a complex scalar of type (0, 0). Equation (32) then gives us two first order differential equations for Σ:

$$\text{þ}\Sigma - \rho(\Sigma - 1) = 0, \qquad (35)$$

$$\eth\Sigma - \tau(\Sigma - 1) = 0, \qquad (36)$$

where þ and ð are the directional derivatives defined by Eq. (26). Taking also the primed versions of the above equations into account, we find that

$$P = (1 + c\Omega) l \wedge m, \qquad (37)$$

$$\Omega \equiv \Psi^{1/3}, \qquad (38)$$

where c is an arbitrary complex constant. A useful integral representation of Ω is given by

$$\Omega = A \exp\left(\int \omega\right), \qquad (39)$$

where ω denotes the 1-form

$$\omega = \rho' l + \rho l' - \tau' m - \tau m', \qquad (40)$$

which is of type (0, 0) and A is a certain complex constant.

Let us now consider the gauge freedom associated with these potentials. The GHP formalism already furnishes us with the way P transforms under the tetrad transformations. The remaining allowable transformations of P will be of the form

$$P \to P_1 = P + P_0, \tag{41}$$

where P satisfies the condition

$$\eth P_0 = 0, \tag{42}$$

so that the potential γ remains invariant. Returning back to Eq. (37) one may easily verify that

$$\eth(\Omega l \wedge m) = 0, \tag{43}$$

and one may identify this part of the Hertzian potential as the gauge term. The scalar Ω which appears as the coefficient of the gauge term satisfies the nonlinear equation

$$\Delta\Omega - 2\Omega^4 = 0, \tag{44}$$

where Δ denotes the harmonic operator

$$\Delta \equiv \delta d + d\delta. \tag{45}$$

We shall interpret the scalar Ω as the gravitational Debye superpotential for type D space—times. Note that there is no explicit reference to the spin coefficients in Eq. (37) and similar to the situation in electromagnetic theory the field 2-forms can be constructed directly from the Hertzian potential by differentiation

$$2\Theta = d\eth P + \eth d P, \tag{46}$$

together with the primed version of this equation. Thus for type D vacuum space—times the tetrad calculus can be formulated solely in terms of the Hertzian potential P. The introduction of the GHP directional derivatives then resolves the procedure into a set of scalar equations for the Debye superpotential Ω.

The coordinate representations of the gravitational Debye potentials according to the Kinnersley's[14] classification of type D vacuum metrics are given below in Table I.

TABLE I. Gravitational Debye potentials for all type D space—times. The notation and conventions we use are those of Kinnersley's catalog. The constant c_0 is defined by $c_0^3 = m + il$.

Case		Debye potential Ω
I		$-c_0(r+il)^{-1}$
II	A	$-c_0(r+il-ia\cos x)^{-1}$
	B	$-c_0(r-il+ia\cosh x)^{-1}$
	C	$-c_0(r-il+ia\sinh x)^{-1}$
	D	$-c_0(r-il+iae^x)^{-1}$
	E	$-c_0(r+ib+\frac{1}{2}ix^2)^{-1}$, $l=1$
	F	$-c_0(r+ix)^{-1}$, $l=0$
III	A	$-c_0 r^{-1}$, $l=0$
	B	$-c_0(\mathrm{dn}x - 2^{-1/2}i\,\mathrm{sn}x)(r+ia\,\mathrm{cn}x)^{-1}$
IV	A	$c_0(x+ia)^{-1}$
	B	$c_0 x^{-1}$, $l=0$

IV. CONCLUSIONS

In this paper we have presented a scheme for introducing Hertzian potentials for type D vacuum space—times. We have started by expressing the general GHP equations in terms of exterior differential forms. This has led to the recognition of the connection 1-forms which have well-defined spin and boost weights as the potential 1-forms of the theory. The corresponding curvature 2-forms were taken as the field variables. The Hertzian potentials for type D space—times were then defined through the introduction of the GHP coderivative. The explicit form of these potentials possessing the symmetries of the curvature 2-forms were constructed. The gauge freedom associated with the Hertzian potential has led us to identify a complex scalar superpotential which we have interpreted as the gravitational Debye potential. These gravitational Debye potentials are found to satisfy a second order nonlinear differential equation and the coordinate representations are seen to have $O(r^{-1})$ asymptotic behavior.[7] In fact, for the particular case of the Kerr solution (Case IIA with $l=0$) the known multipole moments[15,16] can be obtained by expanding the Debye potential in powers of angular momentum parameter a. In the Appendix we have shown that the present procedure is analogous to the construction of an electromagnetic Hertzian potential for the Coulomb field. We note that in both cases the field tensors provide us with two distinct null directions and gauge groups of the theories are Abelian. It should also be noted that although the Hertzian potentials are related to field 2-forms by the differential relations, the scalar potentials that are introduced through the gauge conditions are connected to the essential field variables by algebraic relations. In this the electromagnetic analog of our problem furnishes us with another interesting complex scalar (cf. Appendix)

$$\pi = \Phi_1^{1/2}, \tag{47}$$

which may be interpreted as an alternative Debye potential for the Coulomb field. This quantity has the appealing property that in all algebraically special space—times it satisfies the equation

$$\Delta\pi - 2\Psi\pi = 0, \tag{48}$$

which naturally reduces to $\Delta\pi = 0$ in flat space—time.

Some of the results obtained in Sec. III can directly be carried over to all algebraically special vacuum space—times where only $\sigma = \kappa = 0$. For example, the Hertzian potential P may be used to obtain the curvature 2-form Θ; however, the prime operation may not always be meaningful for this more general class of space—times. One can easily show that Eq. (44) satisfied by Ω also remains valid although the GHP equations cannot be exhausted solely from the definition of Ω. This can be done by expressing the harmonic operator in terms of GHP directional derivatives so that Eq. (44) becomes

$$þ'þ\Omega - \eth'\eth\,\Omega + \tau\eth'\Omega + \bar{\tau}\eth\,\Omega - \rho þ'\Omega - \bar{\rho}'þ\Omega + \Omega^4 = 0. \tag{49}$$

The Bianchi identities

$$þ\Omega = \rho\Omega, \tag{50}$$

$$\eth\,\Omega = \tau\Omega \tag{51}$$

reduce the above equation to the Ricci identity

$$þ'\rho - \eth'\tau = \rho\bar{\rho}' - \tau\bar{\tau} - \Omega^3, \tag{52}$$

which is true for all algebraically special vacuum space—times.

ACKNOWLEDGMENTS

The author wishes to express his deep gratitude to Professor Y. Nutku for his continued guidance and encouragement. He is indebted to Professor C. Arf, Professor F. Gürsey, Professor E. Inönü, and to Dr. H.K. Macomber for their kind interest. He thanks Professor F. Öktem for a critical reading of the manuscript. He also thanks A. Eris, M. Gürses, and M. Halil for stimulating discussions.

APPENDIX

It is well known that one cannot characterize the Coulomb field by a scalar superpotential through the usual Debye reduction of the electromagnetic Hertzian potential.[4] In this appendix we shall show that the procedure of Sec. III can be used to introduce an electromagnetic Hertzian potential which will accomplish this goal in flat space—time. This will also enable us to exhibit some similarities of the Coulomb field with the type D gravitational fields.

In the GHP formalism the electromagnetic field is most conveniently represented by the (0, 0) 2-form

$$\mathcal{J} = \tfrac{1}{2}(F + i{*F}), \tag{A1}$$

$$F = \tfrac{1}{2}F_{ab}\omega^a \wedge \omega^b, \tag{A2}$$

where F_{ab} is the Maxwell tensor and ω^a denote the basis 1-forms. Note that the field 2-form $\mathcal{J}$ is also self-dual

$$*\mathcal{J} = -i\mathcal{J}, \tag{A3}$$

and the source-free Maxwell's equations are given by

$$d\mathcal{J} = 0. \tag{A4}$$

We may write $\mathcal{J}$ explicitly as

$$\mathcal{J} = -\Phi_1(l \wedge n - m \wedge \bar{m}) - \Phi_0 n \wedge \bar{m} + \Phi_2 l \wedge m, \tag{A5}$$

where Φ_0, Φ_1, and Φ_2 are the components of the Maxwell spinor.[11]

Let us now specialize to the case when the basis vectors l and n of the null tetrad are chosen to lie along the distinct principal null directions of the Maxwell tensor. We shall then have $\Phi_0 = \Phi_2 = 0$, $\Phi_1 \neq 0$ and the relations

$$\text{þ}\Phi_1 = 2\rho\Phi_1, \tag{A6}$$

$$\eth\Phi_1 = 2\tau\Phi_1, \tag{A7}$$

together with their primed versions will constitute the field equations. Let us also introduce the spherical polar coordinates for flat space—time

$$ds^2 = dt^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{A8}$$

so that

$$\tau = \tau' = 0, \tag{A9}$$

$$\rho' = -\rho = (2^{1/2}r)^{-1}. \tag{A10}$$

In this coordinate system one can readily integrate the Maxwell's equations and obtain

$$\Phi_1 = -2a_0\rho\rho', \tag{A11}$$

where a_0 is an arbitrary complex constant.

We wish to determine an electromagnetic Hertzian potential P_E for the above case such that the field 2-form (A5) is to be constructed according to the relation

$$2\mathcal{J} = \Delta P_E. \tag{A12}$$

It can be shown very easily that the Hertzian potential

$$P_E = (a_0 + \Omega_E)(l \wedge n - m \wedge m), \tag{A13}$$

with the "gauge condition"

$$\eth[\Omega_E(l \wedge n - m \wedge \bar{m})] = 0 \tag{A14}$$

may be used to construct the required field 2-form. The coefficient Ω_E of the gauge term is a complex scalar of type (0, 0) and is, of course, proportional to the Maxwell spinor

$$\Omega_E = c\Phi_1, \tag{A15}$$

where c is an arbitrary complex constant. It is interesting to note that with the choice $c = a_0^{-1}$ the scalar Ω_E satisfies the equation

$$\Delta\Omega_E - 2\Omega_E^{\,2} = 0, \tag{A16}$$

and may be interpreted as a Debye superpotential.

*This paper will be presented as a thesis to the Department of Physics, Middle East Technical University, in partial fulfillment of the requirements for the Ph.D. degree.

[1]A. Nisbet, Proc. Roy. Soc. (London) A 231, 250 (1955).
[2]T.C. Mo and C.H. Papas, Phys. Rev. D 6, 2071 (1972).
[3]H. Stephani, J. Math. Phys. 15, 14 (1974).
[4]J.M. Cohen and L.S. Kegeles, Phys. Rev. D 10, 1070 (1974).
[5]R. Sachs and P.G. Bergmann, Phys. Rev. 112, 674 (1958).
[6]W.B. Campbell and T. Morgan, Physica (Netherlands) 53, 264 (1971).
[7]R. Penrose, Proc. Roy. Soc. (London) A 284, 159 (1965).
[8]J.M. Cohen and L.S. Kegeles, Phys. Lett. A 54, 5 (1975).
[9]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. 14, 874 (1973).
[10]R. Geroch, J. Math. Phys. 9, 1739 (1968).
[11]E. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[12]Y. Nutku, to be published.
[13]J. Ehlers, Commun. Math. Phys. 37, 327 (1974).
[14]W. Kinnersley, J. Math. Phys. 10, 1195 (1969).
[15]W.C. Hernandez, Phys. Rev. 159, 1070 (1967).
[16]R.O. Hansen, J. Math. Phys. 15, 46 (1974).

Quantum theory of anharmonic oscillators. II. Energy levels of oscillators with $x^{2\alpha}$ anharmonicity

F. T. Hioe, Don MacMillen, and E. W. Montroll

Institute for Fundamental Studies, Departments of Physics and Chemistry, University of Rochester, Rochester, New York 14627
(Received 26 January 1976)

This is an investigation of the energy levels of anharmonic oscillators characterized by the potentials $(1/2)x^2+\lambda x^{2\alpha}$ with $\alpha=2,3,\cdots$ and $\lambda>0$. Two regimes of (λ, n) space are distinguishable: In one, the energy levels differ only slightly from the harmonic ones $E_n=n+1/2$ and in the other they differ only slightly from the purely anharmonic oscillators with $E_n \simeq C_\alpha \lambda^{1/(1+\alpha)}(n+1/2)^{2\alpha/(1+\alpha)}$, C being a constant which depends on α. The magnitude of the combination $\eta \equiv \lambda(n+1/2)^{\alpha-1}$ determines the regime. If this combination is $\ll 1$, one is in the harmonic regime, and if it is $\gg 1$, one is in the anharmonic regime. As $n\to\infty$ the "boundary layer" between the two regimes narrows. The small parameter of a perturbation theory should thus be η rather than λ. Several rapidly convergent algorithms have been developed for the calculation of the energy levels of our anharmonic oscillators, and energy level tables and graphs are presented.

I. INTRODUCTION

There has been considerable recent interest in the quantum theory of anharmonic oscillators. This has been motivated by problems in quantum field theory and in molecular physics. Since traditionally these problems have been investigated by perturbation theory, there has been some concern with new results which show the limitations of that technique. The Hamiltonian

$$H_\alpha(\omega,\lambda) \equiv \tfrac{1}{2}(p^2+x^2\omega^2)+\lambda x^{2\alpha} \qquad (\mathrm{I.1})$$

is frequently used in model studies. In the Schrödinger representation it has the form

$$H_\alpha(\omega,\lambda) = -\tfrac{1}{2}d^2/dx^2+\tfrac{1}{2}x^2\omega^2+\lambda x^{2\alpha} \qquad (\mathrm{I.2})$$

An unexpected result in the application of perturbation theory to this class of Hamiltonian was discovered by Bender and Wu,[1] who showed that when $\alpha=2$ the perturbation series for energy levels in powers of λ diverges for all positive values of λ, no matter how small.

The B—W result is no surprise if one considers the Schrödinger equation for H_α in the momentum representation with

$p\sim p$ and $x\sim id/dp$,

$$\{\tfrac{1}{2}p^2+\tfrac{1}{2}\omega^2 d^2/dp^2+\lambda d^{2\alpha}/dp^{2\alpha}\}\psi=E\psi. \qquad (\mathrm{I.3})$$

The small parameter λ appears as the coefficient of the highest derivative of (I.3). It is known in the mathematical literature that perturbation expansions in integral powers of the small parameter are not valid in such cases. The general solution of the unperturbed ($\lambda=0$) equation has insufficient constants to be specialized to fit all the boundary conditions which might be given for the higher order perturbed equation. The construction of perturbation solutions of the Navier—Stokes equation for viscous fluids involves the same kind of small parameter problems. Hydrodynamic boundary layer theory has been developed to cope with such situations.

The Hamiltonian (I.1) can, through the transformation

$$x=y\lambda^{-1/2(1+\alpha)} \qquad (\mathrm{I.4})$$

(which was suggested first by Symanzik for the case $\alpha=2$), be transformed to

$$H_\alpha(\omega,\lambda)=\lambda^{1/(1+\alpha)}H_\alpha(\omega\lambda^{-1/(1+\alpha)},1). \qquad (\mathrm{I.5})$$

Since, as $\lambda\to\infty$,

$$H_\alpha(\omega,\lambda)\to\lambda^{1/(1+\alpha)}H_\alpha(0,1), \qquad (\mathrm{I.6})$$

the energy levels of (I.1) should have the form

$$E_n(\omega,\lambda)\sim C_n(\alpha)\,\lambda^{1/(1+\alpha)}[1+O(\lambda^{-2/(1+\alpha)})], \qquad (\mathrm{I.7})$$

$C_n(\alpha)$ being a set of numbers which do not depend on λ. On the other hand as $\lambda\to 0$ it might be expected that

$$E_n(\omega,\lambda)\sim(n+\tfrac{1}{2})\,\omega+O(\lambda). \qquad (\mathrm{I.8})$$

In the large λ regime (I.7) implies that our anharmonic oscillator has energy levels which deviate but slightly from a purely $2\alpha-ic$ oscillator, while (I.8) implies that in the small λ regime the energy levels deviate but slightly from a purely harmonic oscillator. The range of λ between those for which $2\alpha-ic$ and harmonic approximations are valid might be considered as the boundary "layer" or region between that of the two purer type oscillators.

TABLE I. Boundary regions for first 11 energy levels for quartic anharmonicity. The parameter λ_1 is the value of λ for which deviation from purely harmonic energy levels has reached 10%; λ_2 is the value of λ for which deviation from purely quartic oscillator energy levels has reached 10% (as λ is reduced from ∞).

Boundary of harmonic regime				Boundary of quartic regime		
n	λ_1	$E_n(\lambda_1)$	$\lambda_1 E_n(\lambda_1)$	λ_2	$E_n(\lambda_2)$	$\lambda_2 E_n(\lambda_2)$
0	0.0827	0.55	0.0455	2.946	1.069	3.150
1	0.0488	1.65	0.0806	1.754	3.175	5.570
2	0.0325	2.75	0.0893	1.039	5.233	5.437
3	0.0242	3.85	0.0930	0.7461	7.319	5.460
4	0.0190	4.95	0.0943	0.5821	9.409	5.477
5	0.0155	6.05	0.0938	0.4766	11.50	5.479
6	0.0128	7.15	0.0918	0.4034	13.58	5.479
7	0.0105	8.25	0.0869	0.3497	15.67	5.480
8	0.00933	9.35	0.0873	0.3087	17.76	5.482
9	0.00835	10.45	0.0873	0.2762	19.85	5.482
10	0.00756	11.55	0.0873	0.2499	21.94	5.482
∞			0.0888			5.482

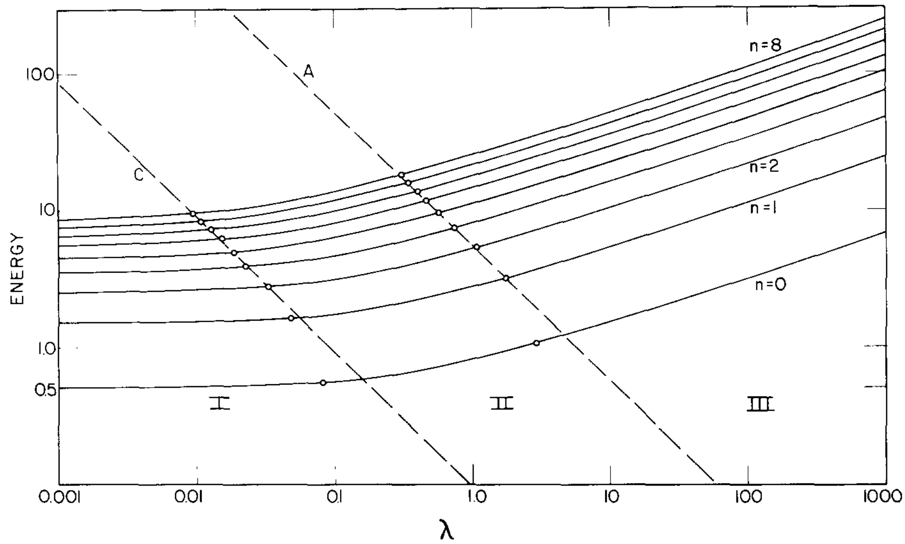


FIG. 1. Variation of energy levels with anharmonicity parameter λ for anharmonic oscillator with quartic anharmonicity. The quantum number of the level is specified by n. Regime I (to left of leftmost sets of dots) represents the (E,λ) range for which energy levels deviate by less than 10% from those of a purely harmonic oscillator. Regime III (to right of rightmost set of points) represent the (E,λ) range for which energy levels deviate by at most 10% from those of a purely quartic oscillator. The dotted lines separate the regimes as $n\to\infty$. Regime II is the "boundary layer" between almost quadratic and almost quartic regimes.

We now define the harmonic regime as that range of λ's for which energy levels differ from harmonic ones by less than 10% and the $2\alpha - ic$ regime as that range of λ's for which energy levels differ from the $2\alpha - ic$ ones by less than 10%. In Ref. 2 rapidly convergent algorithms were derived for the calculation of energy levels of H_2 (quartic anharomonicity) over the full range $0<\lambda<\infty$ and $n=0,1,2,\cdots$. Analytical formulas were derived for expansions such as (I.7) and (I.8). Tables of $E_n(\lambda)$ were also constructed for $n=0,1,2,\ldots,8$.

It is easy to determine the boundary of the harmonic regime (including the 10% deviation allowed in the definition) for the $\alpha=2$ case from this material.[2] This boundary is tabulated in Table I for the 11 energy levels. Notice that for the ground state the harmonic regime is restricted to $0\le\lambda\le 0.0827$. For the first excited state that regime ends at $\lambda\cong 0.0488$, and with increasing quantum number the regime continues to shrink, ending, for example, at $\lambda=0.00756$ for $n=10$. The regime boundaries, and the boundary region thickness as a function of E are clear on a $\log E-\log\lambda$ plot as given in Fig. 1. The plot indicates that the harmonic regime is limited to that part of the (E,λ) plane which is identified by I in the figure and lies between the lower curve C and the coordinate axes. The boundary regime lies between curves C and A, II, while the quartic oscillator regime (allowing for 10% deviation in the definition of the regime) lies above A in III. The dashed curve A is a hyperbola (on the usual E vs λ plot but is a straight line on the $\log E-\log\lambda$ plot given in Fig. 1) which defines the extent of the slightly perturbed quartic regime when the quantum number n is large.

As n increases, the harmonic regime becomes narrower and narrower as does the boundary region. The width of the boundary region for the ground state is $2.95-0.83=2.12$ while it reduces to $0.25-0.01=0.24$ for the excited state $n=10$. The effect of a small perturbation parameter λ gets strongly amplified as n increases.

The dashed boundary curve A in Fig. 1 was obtained from the large n, large λ asymptotic formula for the nth energy level derived in Ref. 2:

$$E_n(\lambda)\sim\lambda^{1/3}[C(n+\tfrac{1}{2})^{4/3}+a(n+\tfrac{1}{2})^{2/3}\lambda^{-2/3}+b\lambda^{-4/3}+\cdots],\tag{I.9}$$

where

$$C=1.3765,\quad a=0.26806,\quad\text{and}\quad b=-0.01167.$$

When n is large, the energy levels of a purely quartic oscillator are

$$\epsilon_n(\lambda)\sim C\lambda^{1/3}(n+\tfrac{1}{2})^{4/3}.\tag{I.10}$$

Then

$$[E_n(\lambda)-\epsilon_n(\lambda)]/\epsilon_n(\lambda)\sim(a/C)[\lambda(n+\tfrac{1}{2})]^{-2/3}+(b/C)[\lambda(n+\tfrac{1}{2})]^{-4/3}.\tag{I.11}$$

The relationship between the energy E and λ along the curve for which the deviation of $E_n(\lambda)$ from a pure quartic is 0.1 is given by

$$[\lambda(n+\tfrac{1}{2})]^{2/3}=\frac{1}{2}\left\{\frac{a}{0.1C}+\left[\left(\frac{a}{0.1C}\right)^2+\frac{4b}{0.1C}\right]^{1/2}\right\}=1.903\equiv\gamma,\tag{I.12a}$$

Hence from (I.9)

$$E\lambda=C\gamma^2+a\gamma+b=5.482,\tag{I.12b}$$

the dashed straight line A in Fig. 1.

When λ is small, the energy levels[2] have the form

$$E_n(\lambda)=(n+\tfrac{1}{2})+\tfrac{3}{4}\lambda\{1+2n(n+1)\}-\lambda^2\left(\frac{(n+1)(n+3/2)^2(n+2)}{2+3\lambda(2n+3)}-\frac{n(n-\frac{1}{2})^2(n-1)}{2+3\lambda(2n-1)}+\frac{(n+1)(n+2)(n+3)(n+4)}{16[4+6\lambda(2n+5)]}-\frac{n(n-1)(n-2)(n-3)}{16[4+6\lambda(2n-3)]}\right)+\cdots.\tag{I.13}$$

If we let $\epsilon_n(\lambda)\equiv n+\frac{1}{2}$ and take the limit $\lambda\to 0$, $n\to\infty$ with $\lambda n\equiv\beta$,

$$[E_n(\lambda)-\epsilon_n(\lambda)]/\epsilon_n(\lambda)\sim\tfrac{3}{2}\beta-\frac{17}{16}\,\frac{\beta^2(4+9\beta)}{(1+3\beta)^2}+\cdots. \qquad \text{(I. 14)}$$

The relationship between the energy E and λ along the curve for which the deviation of $E_n(\lambda)$ from a pure harmonic oscillator is 10% $(\Delta E_n/\epsilon_n(\lambda)=0.1)$, is obtained when $\beta=0.0808$, so that

$$E\lambda=\beta+\tfrac{3}{2}\beta^2+\cdots=0.0888. \qquad \text{(I. 15)}$$

This hyperbola is represented in Fig. 1 by the dashed straight line C on the log–log scale. The boundary width between the almost harmonic and almost quartic regions is then (for fixed E) comparing (I. 12b) and (1.15)

$$\Delta\lambda=5.39/E\to 0 \quad\text{as } E\to\infty. \qquad \text{(I. 16)}$$

Hence, in highly excited states, a system behaves either as a harmonic or quartic oscillator with but a very small range of λ being available for the transition from one regime to the other. The analytical approximation formulas for the boundary regimes corresponding to Eqs. (I. 11) and (I. 14) for a general value of α are given in Appendix C.

The quickly convergent algorithms used for the calculation of energy levels of $H_2(\omega,\lambda)$ and $H_2(\omega\lambda^{-1/3},1)$ in Ref. 2 were derived from the Bargmann representation of the Hamiltonian H_2. In the more general case (I. 1) this representation is found by first writing $H_\alpha(\omega,\lambda)$ in the second quantization form:

$$H_\alpha(\omega,\lambda)=\omega(a^\dagger a+\tfrac{1}{2})+2^{-\alpha}(\lambda\omega^{-\alpha})(a^\dagger+a)^{2\alpha} \qquad \text{(I. 17)}$$

In the Bargmann[3] representation one sets

$$a^\dagger=z \quad\text{and}\quad a=d/dz \qquad \text{(I. 18)}$$

so that the energy levels are characteristic values of the following differential equations:

$$[\omega\,(\tfrac{1}{2}+zd/dz)+2^{-\alpha}(\lambda\omega^{-\alpha})(z+d/dz)^{2\alpha}]\psi=E^{(\alpha)}\psi. \qquad \text{(I. 19)}$$

In the case of a purely harmonic system the characteristic values of $H_\alpha(\omega,0)$ are $\omega(n+\frac{1}{2})$ while the characteristic functions are z^n, since

$$\omega(\tfrac{1}{2}+zd/dz)\,z^n=\omega(\tfrac{1}{2}+n)\,z^n. \qquad \text{(I. 20)}$$

The wavefunctions of our anharmonic system characterized by (I. 1) are series expansions in z. Certain recursion formulas are found for various coefficients of powers of z. It is from the analysis of the determinants associated with these recurrence formula that our quickly convergent algorithms are generated for the calculation of energy levels.

In this report we calculate and tabulate the energy levels of a sextic oscillator over the complete positive λ range for the first six energy levels. We do the same for the first four energy levels of an octic oscillator ($\alpha=4$). Analytic expressions will be obtained for $E_n(\lambda)$ for general α when λ is small, when λ is large, and n is large.

The literature for the quartic ($\alpha=2$) anharmonicity is quite extensive and has been referenced in Ref. 2. However, for higher α values it is rather sparse.[4] The most extensive tables for $\alpha=3$ and $\alpha=4$ have been prepared by Biswas *et al.*,[5] who have tabulated the ground state and second excited state energy levels for the range $0<\lambda\le 50$. Lakshmanan and Prabhakaran[6] have coupled the classical dynamics of the $\alpha=3$ case with the Bohr–Sommerfeld quantum rule to find some asymptotic formulas for energy levels of high quantum number. Truong[7] has used the Weyl quantization prescription to study the sextic oscillator, but essentially no new numerical results were derived in the paper. A discussion of the mixed sextic and quartic anharmonicies has been made by Lakshmanan[8] in the same style as that used in Ref. 6.

In the following, formulas can be converted into those appropriate for

$$\left(\frac{\hbar^2}{2m}\,\frac{d^2}{dx^2}+(E'-\tfrac{1}{2}m\omega^2x^2-\lambda' x^{2\alpha})\right)\psi=0 \qquad \text{(I. 21)}$$

by noting that if we set

$$E=E'/\hbar\omega,\quad y=x(m\omega/\hbar)^{1/2},\quad \lambda=\lambda'\hbar^{\alpha-1}/m^\alpha\omega^{\alpha+1}, \qquad \text{(I. 22)}$$

then

$$\left(\frac{1}{2}\,\frac{d^2}{dy^2}+E-\tfrac{1}{2}y^2-\lambda y^{2\alpha}\right)\psi=0. \qquad \text{(I. 23)}$$

In traditional perturbation theory the small parameter of Eq. (I. 23) would be considered to be the coupling constant λ independently of the energy level n with which one is concerned. An important result of Sec. V is that the perturbation parameter for the nth level should really be $\Lambda_n=\lambda(2n+1)^{\alpha-1}$. For example, if values of $\lambda<1/10$ are to be considered as "small" for the ground state when $\alpha=4$ those for the quantum number $n=10$ should be $\lambda<\frac{1}{8}\times 10^{-4}$.

II. ON THE SMALL λ REGIME

The basic operator which appears in the Bargmann representation (I. 19) of our Hamiltonian H_α is

$$\mathcal{L}=2^{-1/2}(z+d/dz). \qquad \text{(II. 1)}$$

Notice, for example, that direct differentiation yields

$$\mathcal{L}^2z^m=\tfrac{1}{2}[z^{m+2}+(2m+1)z^m+m(m-1)z^{m-2}]. \qquad \text{(II. 2)}$$

Generally, for positive integral α and m,

$$\mathcal{L}^{2\alpha}z^m=\sum_{j=-\alpha}^{\alpha}c_m^{(2j)}(2\alpha)z^{m-2j}, \qquad \text{(II. 3)}$$

where, for example,

$$c_m^{(-4)}(4)=\tfrac{1}{4},\quad c_m^{(-2)}(4)=\tfrac{1}{2}(2m+3),$$

$$c_m^{(2)}(4)=\tfrac{1}{2}m(m-1)(2m-1),$$

$$c_m^{(0)}(4)=3[m^2+(m+1)^2]/4,$$

$$c_m^{(4)}(4)=[m!/(m-4)!]/4$$

and

$$c_m^{(-6)}(6)=1/8,\quad c_m^{(-4)}(6)=3(2m+5)/8,$$

$$c_m^{(-2)}(6)=15(m^2+3m+3)/8,$$

$$c_m^{(0)}(6)=5(4m^3+6m^2+8m+3)/8,$$

$$c_m^{(2)}(6) = 15m(m-1)(m^2-m+1)/8,$$

$$c_m^{(4)}(6) = 3(2m-3)[m!/(m-4)!]/8,$$

$$c_m^{(6)}(6) = [m!/(m-6)!]/8. \tag{II.4}$$

It is shown in Appendix A that generally

$$c_m^{(2h-2\alpha)}(2\alpha) = 2^{-\alpha} \sum_{j=0}^{h} \frac{(2\alpha)!}{(2\alpha-2j)!\,j!}\, 2^{-j} \binom{2\alpha-2j}{h-j} \times \prod_{l=0}^{h-j-1} (m-l), \quad h = 0, 1, \ldots, 2\alpha. \tag{II.5}$$

Now let us consider the differential equation (I.19). As in Ref. 2, we write the energy levels and wavefunctions of (I.19) in the form

$$E_n(\lambda) = (\tfrac{1}{2} + n) + A_n(\lambda), \tag{II.6a}$$

$$\psi_n(z) = f_n(\lambda, z) = z^n B_n(\lambda, z) \tag{II.6b}$$

with

$$B_n(\lambda, z) \equiv \sum_{k=-n}^{\infty} U_k^{(n)}(\lambda) z^k. \tag{II.6c}$$

After substituting (II.6) into (I.19) and employing (II.3), we find by equating coefficients of like powers, say z^{n+k}, on both sides of the resulting equations, that

$$\lambda \sum_{j=-\alpha}^{\alpha} U_{k+2j}\, c_{n+k+2j}^{(2j)}(2\alpha) + U_k[k - A_n(\lambda)] = 0, \tag{II.7}$$

$$k = -n, -(n-1), \ldots, 0, 1, \cdots.$$

For fixed n and α we suppress the explicit dependence on these quantities in the abbreviation

$$a_k^{(2j)} \equiv \lambda c_{n+k}^{(2j)}(2\alpha). \tag{II.8}$$

With the introduction of one more definition,

$$\gamma_k = \lambda c_{n+k}^{(0)}(2\alpha) + k - A_n(\lambda), \tag{II.9}$$

(II.7) has the matrix form,

$$\begin{pmatrix} \gamma_{-n} & & & & & & & & \\ & \gamma_{-n+2} & & & & & & & \\ & & \ddots & & & & & & \\ & & & \gamma_{-4} & & & & & \\ & & & a_{-4}^{(-2)} & \gamma_{-2} & a_0^{(2)} & \cdots & a_{2\alpha-2}^{(2\alpha)} & 0 & \cdots \\ a_{-2\alpha}^{(-2\alpha)} & \cdots & & a_{-2}^{(-2)} & \gamma_0 & a_2^{(2)} & \cdots & a_{2\alpha}^{(2\alpha)} & 0 \\ 0 & a_{-2\alpha+2}^{(-2\alpha)} & \cdots & a_0^{(-2)} & \gamma_2 & a_4^{(2)} & \cdots & a_{2\alpha+2}^{(2\alpha)} & \\ 0 & 0 & \ddots & & & \ddots & & & \ddots \end{pmatrix} \begin{pmatrix} U_{-n} \\ \vdots \\ U_{-4} \\ U_{-2} \\ U_0 \\ U_2 \\ \vdots \\ \vdots \end{pmatrix} = 0 \tag{II.10}$$

It should be noted that only U_k with even subscripts appear in the equations. For example, if $n=4$, the only U_k which appear are U_{-4}, U_{-2}, U_0, $U_2, \cdots$. One can arbitrarily set $U_{-3} = U_{-1} = U_1 = \cdots = 0$ since they do not appear in any formulas. If $n=3$, the only U_k which appear are U_{-2}, U_0, $U_2, \cdots$. Generally, for even n the lower index starts with $-n$ and for odd n the lower index starts with $-(n-1)$.

There are α nonvanishing diagonals above the main diagonal and the same number below. If we abbreviate (II.10) as

$$G(n)\, U(n) = 0, \tag{II.11}$$

then to get a nontrivial solution, we must set

$$\det G = 0 \tag{II.12}$$

or

$$A_n(\lambda) = a_0^{(0)} + \det G' / \det X, \tag{II.13}$$

TABLE IIa. Convergence of ground state energy shifts, $A_0(\lambda)$, with sextic anharmonicity, λx^6 (i.e., $\alpha = 3$).

Size of Determinants	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$
(4,3)	0.054 8565	0.088 6557	0.226 818	0.317 924
(8,7)	0.054 5480	0.087 0037	0.218 245	0.307 588
(12,11)	0.054 5437	0.086 9470	0.217 947	0.305 387
(16,15)	0.054 5435	0.086 9447	0.217 820	0.305 017
(20,19)	0.054 5435	0.086 9446	0.217 814	0.304 977
(24,23)	0.054 5435	0.086 9446	0.217 812	0.304 969
(28,27)	0.054 5435	0.086 9446	0.217 812	0.304 966
(32,31)	0.054 5435	0.086 9446	0.217 812	0.304 966

TABLE IIb. Convergence of ground state energy shifts, $A_0(\lambda)$, with octic anharmonicity, λx^8 (i.e., $\alpha = 4$).

Size of Determinants	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$
(7,6)	0.085351	0.12184	0.26203	0.34925
(15,14)	0.084507	0.12060	0.24634	0.32300
(23,22)	0.084486	0.12052	0.24560	0.32090
(31,30)	0.084486	0.12051	0.24552	0.32070
(39,38)	0.084486	0.12051	0.24551	0.32069
(47,46)	0.084486	0.12051	0.24551	0.32069

where G' is the same as G except that the element γ_0 is replaced by 0, and X_0 is obtained from G by striking out the row and column of the element γ_0. The explicit expressions of these matrix elements for $\alpha = 3$ and $\alpha = 4$ are given in Appendix B. Each of the determinants is a function of $A_n(\lambda)$ through the terms γ_l which appear in them.

An algorithm was presented in Ref. 2 for the calculation of $A_n(\lambda)$ from Eq. (II.13) in the case $\alpha = 2$. We use the same algorithm here for the higher α cases. In the algorithm the infinite determinants, G' and X_0, are truncated with G' being one order higher than X_0, say to, respectively, 4×4 and 3×3 determinants. Then an estimated $A_n(\lambda)$ [which might be taken as $a_0^{(0)}(\alpha)$] is substituted into the right-hand side of (II.13), yielding a new estimate of $A_n(\lambda)$ on the left. The new $A_n(\lambda)$ is then substituted into the right-hand side of (II.13) yielding another estimate of $A_n(\lambda)$. As this process is repeated, it is found from experience that a steady value of $A_n(\lambda)$ is quickly obtained. The truncation of the determinants is then made at a higher level, say to yield 8×8 and 7×7 determinants. The (4, 3) truncation estimation of $A_n(\lambda)$ is used as a first approximation in the (8, 7) truncation equation. After several iterations a new steady value of $A_n(\lambda)$ is obtained. This can be used as a first estimate in a (12, 11) truncation calculation, etc., until successive higher order truncations yield steady results. The rate of convergence of this process is exhibited in Table II for the sextic anharmonicity ($\alpha = 3$) ground state for several values of λ. Note that three significant figure accuracy is obtained with (8, 7) truncation when $\lambda \leq 0.5$ while a (24, 23) truncation is necessary for six significant figure accuracy.

With the aid of the Schweinsian expansion for the ratio of determinants which differ only by a single row or column, the above ideas can be used to find a formula for $A_0(\lambda)$ when λ is small. Since the method is developed in detail in Ref. 2, we merely state the results of such a calculation here for $A_0(\lambda)$ for the sextic anharmonicity ($\alpha = 3$):

$$A_0(\lambda) = \frac{15\lambda}{8} - \frac{5\lambda^2}{4}\left(\frac{405}{17(1+10\lambda)} + \frac{135}{8(1+60\lambda)} + \frac{144}{48(2+235\lambda)}\right) + O(\lambda^3). \tag{II.14}$$

Notice that one must refrain from expanding the terms such as $(1+10\lambda)^{-1}$ in powers of λ since the radius of con-

TABLE III. Values of the energies for $n = 0, 1, \ldots, 5$ of the oscillator with sextic anharmonicity.

λ	E_0	E_1	E_2	E_3	E_4	E_5
0.0001	0.500187	1.50131	2.50464	3.51163	4.52364	5.54194
0.0005	0.500924	1.50639	2.52224	3.55505	4.60929	5.68890
0.001	0.501825	1.51248	2.54307	3.60389	4.70204	5.84189
0.003	0.505225	1.53454	2.61429	3.76303	4.98831	6.29153
0.005	0.508371	1.55399	2.67371	3.88885	5.20417	6.61764
0.01	0.515443	1.59544	2.79382	4.13169	5.60608	7.20873
0.05	0.554544	1.79802	3.32220	5.11894	7.15352	9.40088
0.10	0.586945	1.95042	3.69082	5.77373	8.14755	10.7796
0.30	0.666448	2.30027	4.49838	7.16978	10.2336	13.6439
0.50	0.717812	2.51670	4.98331	7.99472	11.4551	15.3113
0.70	0.757449	2.68065	5.34619	8.60815	12.3602	16.5441
1.0	0.804966	2.87467	5.77197	9.32486	13.4152	17.9788
2.0	0.915219	3.31756	6.73353	10.9350	15.7785	21.1187
3.0	0.991891	3.62146	7.38764	12.0258	17.3758	23.3523
4.0	1.05231	3.85931	7.89731	12.8740	18.6165	25.0332
5.0	1.10286	4.05742	8.32061	13.5776	19.6447	26.4254
10.0	1.28232	4.75605	9.80691	16.0429	23.2442	31.2948
20	1.50016	5.59753	11.5883	18.9910	27.5433	37.1067
50	1.85849	6.97310	14.4887	23.7825	34.5234	46.5371
100	2.19334	8.25314	17.1803	28.2239	40.9893	55.2691
200	2.59418	9.78157	20.3891	33.5149	48.6890	65.6649
300	2.86398	10.8088	22.5437	37.0662	53.8560	72.6400
400	3.07309	11.6044	24.2117	39.8148	57.8545	78.0374
500	3.24618	12.2627	25.5912	42.0878	61.1611	82.5005
1000	3.85087	14.5606	30.4053	50.0182	72.6959	98.0691
3000	5.05525	19.1330	39.9780	65.7835	95.6232	129.011
5000	5.73940	21.7289	45.4106	74.7290	108.630	146.565
8000	6.45138	24.4297	51.0619	84.0339	122.160	164.824
20 000	8.10586	30.7039	64.1884	105.645	153.584	207.222

TABLE IV. Values of the energies for $n=0,1,2,3$ of the oscillator with octic anharmonicity.

λ	$E_0(\lambda)$	$E_1(\lambda)$	$E_2(\lambda)$	$E_3(\lambda)$
0.0001	0.500 64	1.5056	2.5242	3.5711
0.0005	0.502 93	1.5240	2.5955	3.7538
0.001	0.505 43	1.5424	2.6602	3.9044
0.003	0.513 43	1.5957	2.8295	4.2676
0.005	0.519 75	1.6343	2.9432	4.4979
0.01	0.532 10	1.7047	3.1398	4.8812
0.05	0.584 49	1.9699	3.8200	6.1406
0.1	0.620 51	2.1377	4.2265	6.8686
0.3	0.698 86	2.4847	5.0391	8.2987
0.5	0.745 51	2.6844	5.4969	9.0956
0.7	0.780 21	2.8308	5.8293	9.6718
1.0	0.820 69	2.9998	6.2106	10.330
2.0	0.911 09	3.3722	7.0435	11.764
3.0	0.971 79	3.6195	7.5925	12.705
4.0	1.0186	3.8091	8.0120	13.424
5.0	1.0573	3.9649	8.3556	14.011
10.0	1.1909	4.5003	9.5322	16.020
20.0	1.3473	5.1221	10.892	18.338
50.0	1.5943	6.0975	13.017	21.951
100.0	1.8163	6.9697	14.911	25.168
200.0	2.0731	7.9760	17.092	28.870
300.0	2.2415	8.6345	18.517	31.288
400.0	2.3698	9.1358	19.602	33.127
500.0	2.4748	9.5454	20.488	34.629
1000.0	2.8331	10.943	23.508	39.749
3000.0	3.5159	13.602	29.248	49.475
5000.0	3.8891	15.053	32.380	54.782
8000.0	4.2683	16.528	35.560	60.168
20 000.0	5.1194	19.835	42.693	72.247

vergence of the first term in the square bracket is 1/10, that of the second term is 1/60 while that of the third term is only 1/117.5. Certain terms in the λ^3 component have even smaller radii of convergence.

We have used the algorithm described above to make tables of the first six energy levels of a sextic oscillator in the small λ range (see Table III) and the first four levels for the octic oscillator in Table IV. As λ increases, the size of the truncated determinants increases for a preassigned accuracy of the results. For a given energy level there is value of λ, say λ_c, such that when $\lambda \geq \lambda_c$ our algorithm should be applied to $H_\alpha(\omega\lambda^{-1(1+\alpha)}, 1)$, [see Eq. (I.5)] rather than $H_\alpha(\omega, \lambda)$ for faster convergence. This will be discussed in detail in Sec. III.

If one wishes to compare our results of Table III with those of Biswas *et al.*[5] in their Table IV for E_0 and E_2, the only energy levels which they discuss), their λ is twice our λ as are their energy levels. Their Hamiltonian differs from (I.1) by the omission of the factor $\frac{1}{2}$ on $\frac{1}{2}(p^2+\omega^2x^2)$.

The small λ expansion for arbitrary quantum number n and general α is

$$A_n(\lambda,\alpha)=a_0^{(0)}-\sum_{j=1}^{\alpha}\left(\frac{a_0^{(2j)}a_{-2j}^{(-2j)}}{\zeta_{-2j}}+\frac{a_0^{(-2j)}a_{2j}^{(2j)}}{\zeta_{2j}}\right)+\cdots, \qquad \text{(II.15)}$$

where $\zeta_k=k+a_k^{(0)}-a_0^{(0)}$. The a's are defined by (II.8) and (II.5).

As n and α increase, the range of λ for which this formula is valid decreases rapidly.

III. ON THE LARGE λ REGIME

In the large λ regime, we start with (I.5) and use our determinant ratio algorithm to find the energy levels of $H_\alpha(\omega\lambda^{-1/(1+\alpha)}, 1)$. These multiplied by $\lambda^{1/(1+\alpha)}$ give the required levels for $H_\alpha(\omega,\lambda)$. The matrix equation (II.10) is basic for this purpose if the $a_k^{(2j)}$ are replaced by $a_k^{(2j)\prime}$, where the primed variables are defined by

$$a_k^{(0)\prime}=a_k^{(0)}-\tfrac{1}{2}\epsilon[1+2(n+k)],$$
$$a_k^{(-2)\prime}=a_k^{(-2)}-\tfrac{1}{2}\epsilon,$$
$$a_k^{(2)\prime}=a_k^{(2)}-\tfrac{1}{2}\epsilon(n+k)(n+k-1), \qquad \text{(III.1)}$$

and all λ in (II.10) are set equal to 1 while

$$\epsilon=\tfrac{1}{2}[1-\lambda^{-2/(\alpha+1)}]. \qquad \text{(III.2)}$$

All other $a_k^{(2\alpha)}$ remain unchanged except that $\lambda=1$. The analog of (II.13) for the ground state energy shift is

$$A_0(\epsilon)=a_0^{(0)\prime}+\det G''/\det X_0', \qquad \text{(III.3)}$$

where the primes indicate that the elements of the determinants are the $a_k^{(2j)\prime}$ coefficients.

The simplest case to consider first is the regime of very large λ (i.e., $\epsilon=\frac{1}{2}$) which corresponds to a purely $2\alpha-ic$ oscillator. Then the ground state energy is

$$E_0(\lambda)\sim\lambda^{1/(\alpha+1)}[\tfrac{1}{2}+A_0(\infty)]. \qquad \text{(III.4)}$$

TABLE V. The convergence of $A_0(\epsilon)$ for $\epsilon = \frac{1}{2}$, i.e., $\lambda = \infty$ for the sextic ($\alpha = 3$) and octic ($\alpha = 4$) anharmonicities as the sizes of determinants increase.

Size of determinant	$A_0(\infty)$ $\alpha = 3$	Size of determinant	$A_0(\infty)$ $\alpha = 4$
(4,3)	0.200 620	(11,10)	0.247 65
(12,11)	0.181 375	(19,18)	0.245 68
(20,19)	0.180 723	(27,26)	0.245 52
(28,27)	0.180 704	(35,34)	0.245 51
(35,34)	0.180 704	(43,42)	0.245 51

TABLE VI. Values of ϵ_n, α_n, and β_n of Eq. (III.7) for sextic anharmonicity.

n	ϵ_n	α_n	β_n
0	0.680 707	0.129 39	−0.0052
1	2.579 75	0.301 93	−0.0071
2	5.394 89	0.380 29	−0.0032
3	8.880 51	0.446 89	−0.0026
4	12.9113	0.505 93	−0.0022
5	17.4216	0.558 67	−0.0020

The values of $A_0(\lambda)$ which have been obtained for various truncations of the determinants are given in Table V for the sextic anharmonicity ($\alpha = 3$). For seven figure accuracy one must apply the determinant algorithm to 30th order determinants. However, since we use a sparce determinant simplification in the computation, the actual computer time for the calculation is only a few seconds.

We have calculated the $A_0(\lambda)$ for the combined harmonic and sextic oscillator in

$$E_0(\lambda) = \lambda^{1/(\alpha+1)}[\tfrac{1}{2} + A_0(\lambda)] \qquad \text{(III.5)}$$

by our algorithms for some sample values of λ in the range $1 \leq \lambda < \infty$. These are listed in Table III for $\alpha = 3$, along with the values for $0 \leq \lambda < 1$ calculated in the manner discussed in Sec. II. The results listed in the table as well as a number of others for a denser set of λ's than those of the table have been used to find the following asymptotic formula:

$$E_0(\lambda) = \lambda^{1/4}(0.680707 + 0.12939\lambda^{-1/2} - 0.0052\lambda^{-1} + \cdots). \qquad \text{(III.6)}$$

The precise scheme for the calculation of the coefficients has been outlined in Sec. III of Ref. 2.

The large λ determinant truncation algorithm has also been employed for the determination of the next five energy levels for our sextic anharmonicity. The results are tabulated in Table III along with those obtained from the small λ algorithm. Curves are plotted in Fig. 2. Asymptotic formula analogous to (III.6) have also been found empirically for $n = 1, 2, \ldots, 5$:

$$E_n(\lambda) = \lambda^{1/4}(\epsilon_n + \alpha_n\lambda^{-1/2} + \beta_n\lambda^{-1} + \cdots). \qquad \text{(III.7)}$$

The coefficients ϵ_n, α_n, and β_n are listed in Table VI.

We have listed a selection of energy levels for our oscillator with octic anharmonicity in Table IV for the first four energy levels. Curves are plotted in Fig. 3. For large λ the asymptotic expansion of $E_n(\lambda)$ is

$$E_n(\lambda) = \lambda^{1/5}(\epsilon_n + \alpha_n\lambda^{-2/5} + \beta_n\lambda^{-4/5} + \cdots). \qquad \text{(III.8)}$$

The various values of ϵ_n, α_n, and β_n are given in Table VII.

Since, as n increases, the size of the determinant required for the truncation algorithm increases rapidly, it is desirable to find an alternative procedure for the calculation of energy levels of highly excited states. A WKB type of analysis will be used for this purpose.

IV. LARGE n REGIME

It is known from experience with WKB calculations that the nth energy level associated with certain one-dimensional potentials, of which those considered here are examples, depend on n through the combination $(n + \frac{1}{2})$. In the style presented in Ref. 2 in the discussion of quartic anharmonicities, we scale the parameters ϵ_n, α_n, and β_n of (III.7) and and (III.8) with powers of $(n + \frac{1}{2})$. In Table VIII we list the following combination, as a function of n, for sextic anharmonicities; $\epsilon_n/(n+\frac{1}{2})^{3/2}$, $\alpha_n/(n+\frac{1}{2})^{1/2}$, and $\beta_n(n+\frac{1}{2})^{1/2}$. Notice that, with increasing n, limits seem to be approached. Indeed, even for n only 5 the ratios are very close to

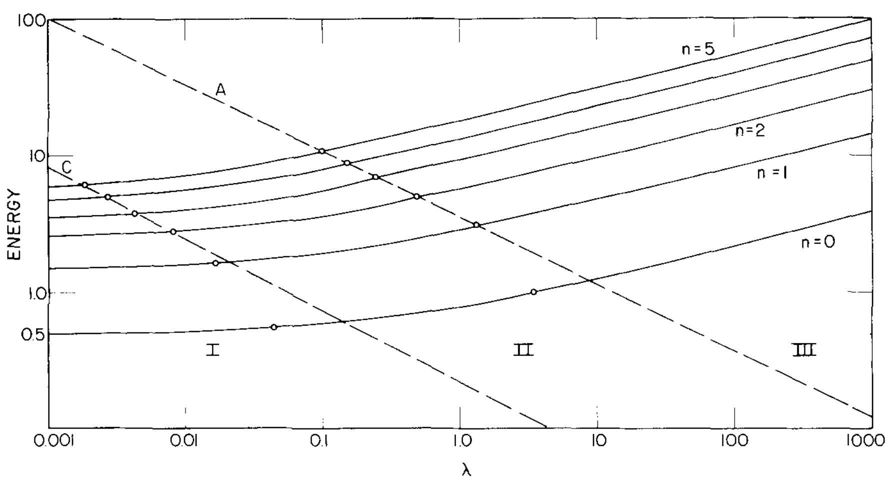


FIG. 2. Same material as Fig. 1 but for sextic ($\alpha = 3$) oscillator.

TABLE VII. Values of ϵ_n, α_n, and β_n of Eq. (III.8) for octic anharmonicity.

n	ϵ_n	α_n	β_n
0	0.70405	0.1205	−0.0039
1	2.7315	0.2730	−0.0047
2	5.8842	0.3276	−0.0012
3	9.9611	0.3699	−0.0006
4	14.824		
5	20.389		

TABLE VIII. The approach to WKB results for sextic anharmonicity. [see Eq. (IV.23) with $\alpha = 3$].

n	$C_n = \epsilon_n/(n+\frac{1}{2})^{3/2}$	$a_n = \alpha_n/(n+\frac{1}{2})^{1/2}$	$b_n = \beta_n(n+\frac{1}{2})^{1/2}$
0	1.92533	0.18299	−0.0037
1	1.40424	0.24653	−0.0086
2	1.36481	0.24052	−0.0054
3	1.35624	0.23887	−0.0049
4	1.35254	0.23850	−0.0047
5	1.35066	0.23822	−0.0047
WKB	1.346760	0.238075	−0.0045919

those which have been obtained from the WKB type calculation described below.

Since we are especially concerned with the large λ regime, we apply the WKB method to the Hamiltonian $H_\alpha(\omega\lambda^{-1/(\alpha+1)}, 1)$ rather than to $H_\alpha(\omega, \lambda)$. We first examine the case in which λ is so large that we need only consider the purely $2\alpha - ic$ oscillator with Hamiltonian $H_\alpha(0,1)$. A formula due to Titchmarsh (Ref. 9, p. 151, Eq. 7.7.4) is immediately applicable. Let $\mu_0, \mu_1, \cdots$ be eigenvalues of

$$\psi'' + [\mu - q(v)]\psi = 0 \quad \text{with } q(v) \to \infty \quad \text{as } v = \pm\infty, \qquad \text{(IV.1)}$$

$$\psi(\pm\infty) = \psi'(\pm\infty) = 0. \qquad \text{(IV.2)}$$

Then, if v_n and v_n' are roots of $q(v) = \mu_n$,

$$\frac{1}{\pi}\int_{v_n'}^{v_n} [\mu_n - q(v)]^{1/2}\, dv = n + \tfrac{1}{2} + O(1/n). \qquad \text{(IV.3)}$$

We proceed to follow the scheme outlined above by starting with the equation

$$(d^2/dy^2 + 2E - y^2 - 2\lambda y^{2\alpha})\psi = 0 \qquad \text{(IV.4)}$$

and letting

$$y = (2\lambda)^{-1/(1+\alpha)}\, v \quad \text{and} \quad \mu = 2E(2\lambda)^{-1/(1+\alpha)}. \qquad \text{(IV.5)}$$

Then

$$[d^2/dv^2 + (\mu - v^{2\alpha}) - v^2(2\lambda)^{-2/(1+\alpha)}]\psi = 0. \qquad \text{(IV.6)}$$

In the very large λ regime in which we can neglect the term proportional to v^2, (IV.6) has the form (IV.1) so that we can apply the Titchmarsh formula to the resulting equation with

$$q(v) \equiv v^{2\alpha}. \qquad \text{(IV.7)}$$

In this case $v_n = -v_n' = \mu_n^{1/2\alpha}$.

Titchmarsh[9] has evaluated the resulting integral (left-hand side) (IV.3), which now has the form

$$\mu_n^{(1+\alpha)/2\alpha}\, \pi^{-1} \int_0^1 (1 - x^{2\alpha})^{1/2}\, dx$$

$$= \mu_n^{(1+\alpha)/2\alpha}\left(\frac{\Gamma(\frac{3}{2})\Gamma(1/2\alpha)}{2\alpha\pi\Gamma(\frac{3}{2} + 1/2\alpha)}\right). \qquad \text{(IV.8)}$$

When this expression is combined with (IV.5) and (IV.3), we obtain the following formula for the energy levels:

$$E_n(\lambda) \sim \tfrac{1}{2}(2\lambda)^{1/(1+\alpha)}\left(\frac{(n+\frac{1}{2})\, 2\pi\alpha\Gamma(\frac{3}{2} + 1/2\alpha)}{\Gamma(\frac{3}{2})\Gamma(1/2\alpha)}\right)^{2\alpha/(1+\alpha)}. \qquad \text{(IV.9)}$$

From this expression we can obtain the WKB value of ϵ_n for special values of α:

α	$\epsilon_n/(n+\frac{1}{2})^{2\alpha/(1+\alpha)} \equiv C$
2	1.376507
3	1.346760
4	1.326607
5	1.312307

These values are consistent with the numbers in Tables VIII and IX and in Table V in Ref. 2.

Equation (IV.9) is an asymptotic result for very large n. It can be improved by referring back to the Titchmarsh formula (IV.3), in which the right-hand

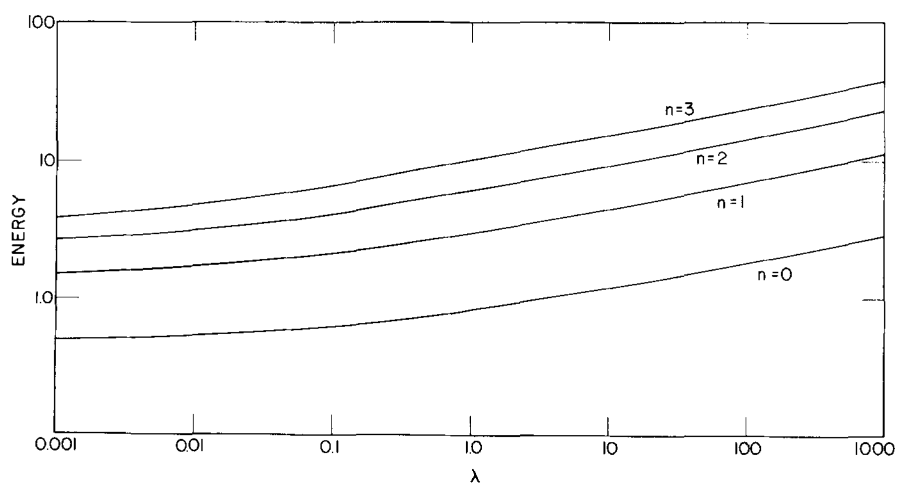


FIG. 3. Variation of first four energy levels as a function of anharmonicity constant λ for octic oscillator.

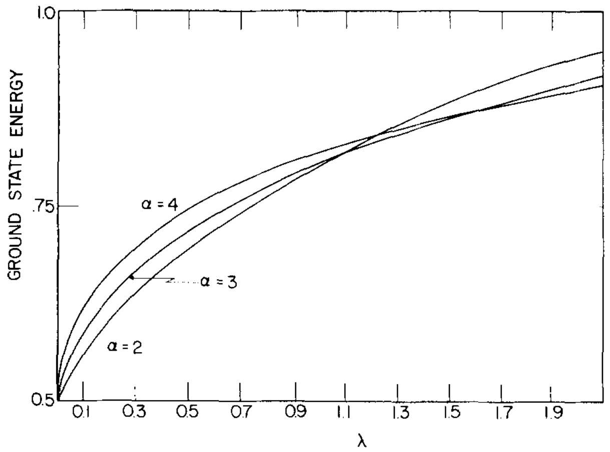


FIG. 4. Ground state energy levels for anharmonic oscillators with quartic ($\alpha = 2$), sextic ($\alpha = 3$), and octic ($\alpha = 4$) oscillators.

side of the equation has a term of order $1/n$ which has been neglected. Let us suppose that this term is written as $\delta_\alpha/[n+\frac{1}{2}]$ where the parameters δ_α depend on the power 2α which appears in the anharmonic force law. Let C_α be the constant multiplying $(n+\frac{1}{2})^{2\alpha/(1+\alpha)}$ in Eq. (IV.9). Then, if we refer back to equations such as (III.7) and (III.8) and if we assume that in (IV.9) we replace $(n+\frac{1}{2})$ by

$$(n+\tfrac{1}{2}) + \delta_\alpha/(n+\tfrac{1}{2}),$$

we could, from our numerical data, determine the δ_α or

$$\delta_\alpha = \lim_{n\to\infty}(n+\tfrac{1}{2})[(\epsilon_n/C_\alpha)^{(1+\alpha)/2\alpha} - (n+\tfrac{1}{2})]. \qquad \text{(IV.9')}$$

In Ref. 2 this was found for $\alpha = 2$ to be

$$\delta_2 = 0.02651$$

Estimates of δ_α for $\alpha = 3$ and $\alpha = 4$ are given in Table X. The WKB values of α_n and β_n in Tables VIII and IX are obtained by substituting

$$q(v) \equiv v^{2\alpha} + v^2(2\lambda)^{-2/(1+\alpha)} \qquad \text{(IV.10)}$$

into (IV.3), which then becomes

$$n+\tfrac{1}{2}+O(1/n) = \pi^{-1}\mu_n^{(1+\alpha)/2\alpha}F(\beta,\alpha), \qquad \text{(IV.11)}$$

where

$$F(\beta,\alpha) \equiv \int_0^{x_n}[1-\beta x^2 - x^{2\alpha}]^{1/2}\,dx, \qquad \text{(IV.12a)}$$

$$\beta \equiv \mu_n^{1/\alpha}/2\lambda^{2/(1+\alpha)}, \qquad \text{(IV.12b)}$$

with β being a small number and x_n the root near 1 of

$$1-\beta x_n^2 - x_n^{2\alpha} = 0. \qquad \text{(IV.13)}$$

It is easy to show that

$$x_n = 1-(\beta/2\alpha) + (5-2\alpha)(\beta^2/8\alpha^2) + \cdots. \qquad \text{(IV.14)}$$

If we let $z = x/x_n$, then (1) becomes

$$F(\beta,\alpha) = x_n^{1+\alpha}\int_0^1[(1-z^{2\alpha}) + \gamma(1-z^2)]^{1/2}\,dz, \qquad \text{(IV.15)}$$

where

$$\gamma = \beta x_n^{2(1-\alpha)}.$$

Since, for $0 \leq z \leq 1$ and $\gamma \leq 1$,

$$\gamma(1-z^2) \leq (1-z^{2\alpha}) \quad \text{for } \alpha \geq 1, \qquad \text{(IV.16)}$$

TABLE IX. The approach to the WKB results for octic anharmonicity.

n	$C_n = \epsilon_n/(n+\frac{1}{2})^{8/5}$	$a_n = \alpha_n/(n+\frac{1}{2})^{2/5}$	$b_n = \beta_n(n+\frac{1}{2})^{4/5}$
0	2.1343	0.1590	−0.0022
1	1.4278	0.2321	−0.0065
2	1.3583	0.2271	−0.0025
3	1.3421	0.2241	−0.0016
4	1.3360		
5	1.3330		
WKB	1.326 607	0.222 270	−0.001 607

$$F(\beta,\alpha) = x_n^{1+\alpha}\int_0^1 (1-z^{2\alpha})^{1/2} \times\left(1+\tfrac{1}{2}\gamma\,\frac{(1-z^2)}{(1-z^{2\alpha})} - \tfrac{1}{8}\gamma^2\,\frac{(1-z^2)^2}{(1-z^{2\alpha})^2}+\cdots\right)dz$$

$$= x_n^{1+\alpha}(I_0 + \tfrac{1}{2}\gamma I_1 - \tfrac{1}{8}\gamma^2 I_2 + \cdots), \qquad \text{(IV.17)}$$

where I_0 and I_1 are immediately expressible in terms of beta functions since

$$\int_0^1 y^{t-1}(1-y)^{s-1}\,dy = \mathrm{B}(s,t) = \Gamma(s)\Gamma(t)/\Gamma(s+t), \quad \text{when } t>0,\ s>0. \qquad \text{(IV.18)}$$

Then

$$I_0 = (1/2\alpha)\,\mathrm{B}(1/2\alpha, 3/2), \qquad \text{(IV.19a)}$$

$$I_1 = (1/2\alpha)\,\{\mathrm{B}(1/2\alpha,\tfrac{1}{2}) - \mathrm{B}(3/2\alpha,\tfrac{1}{2})\}. \qquad \text{(IV.19b)}$$

While I_2 is also expressible in terms of beta functions, a little preparation is first necessary. We write

$$I_2 = \int_0^1 \frac{(1-z^2)^2[(1-z^{2\alpha}) + z^{2\alpha}]\,dz}{(1-z^{2\alpha})^{3/2}} = \int_0^1\frac{(1-z^2)^2}{(1-z^{2\alpha})^{1/2}}\,dz + \frac{1}{\alpha}\int_0^1 z(1-z^2)^2\,d(1-z^{2\alpha})^{-1/2}.$$

Integration by parts and rearrangement leads to

$$I_2 = \left(1-\frac{1}{\alpha}\right)\int_0^1\frac{(1-z^2)^2}{(1-z^{2\alpha})^{1/2}}\,dz + \frac{4}{\alpha}\int_0^1 z^2\,\frac{(1-z^2)}{(1-z^{2\alpha})^{1/2}}\,dz$$

$$= \left(1-\frac{1}{\alpha}\right)\int_0^1\frac{dz}{(1-z^{2\alpha})^{1/2}} - 2\left(1-\frac{3}{\alpha}\right)\int_0^1\frac{z^2\,dz}{(1-z^{2\alpha})^{1/2}} + \left(1-\frac{5}{\alpha}\right)\int_0^1\frac{z^4\,dz}{(1-z^{2\alpha})^{1/2}}$$

$$= \frac{1}{2\alpha}\left[\left(1-\frac{1}{\alpha}\right)\mathrm{B}\left(\frac{1}{2\alpha},\frac{1}{2}\right) - 2\left(1-\frac{3}{\alpha}\right)\mathrm{B}\left(\frac{3}{2\alpha},\frac{1}{2}\right) + \left(1-\frac{5}{\alpha}\right)\mathrm{B}\left(\frac{5}{2\alpha},\frac{1}{2}\right)\right]. \qquad \text{(IV.19c)}$$

TABLE X. Estimates of δ_α for $\alpha = 3$ and 4 from Eq. (IV.9').

n	δ_3	δ_4
0	0.067 216	0.086 518
1	0.063 570	0.105 765
2	0.055 722	0.092 811
3	0.057 406	0.089 426
4	0.057 917	0.089 816
5	0.058 330	0.090 660

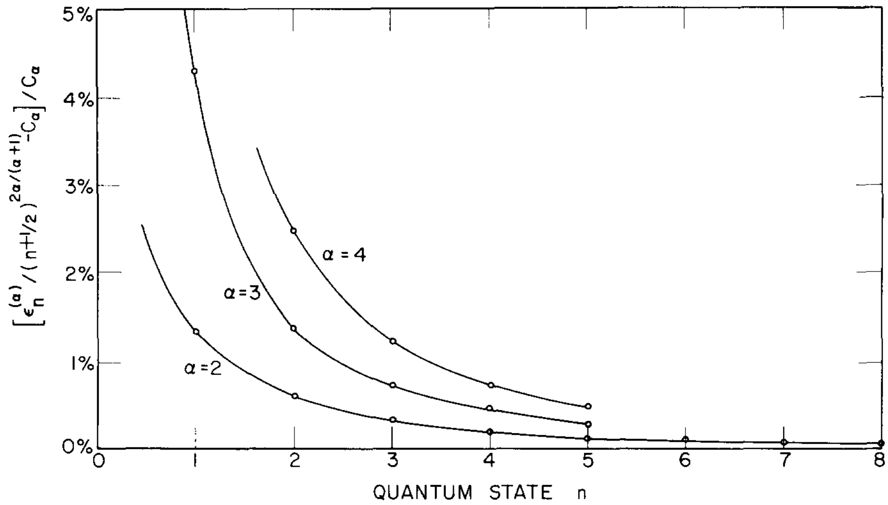


FIG. 5. Percentage difference of the coefficient ϵ_n for the $2\alpha - ic$ oscillator from the WKB coefficient as the quantum number n increases.

Hence

$$F(\beta,\alpha)=\frac{x_n^{1+\alpha}}{2\alpha}\left\{B\left(\frac{1}{2\alpha},\frac{3}{2}\right)+\frac{\gamma}{2}\left[B\left(\frac{1}{2\alpha},\frac{1}{2}\right)-B\left(\frac{3}{2\alpha},\frac{1}{2}\right)\right]-\frac{\gamma^2}{8}\left[\left(1-\frac{1}{\alpha}\right)B\left(\frac{1}{2\alpha},\frac{1}{2}\right)-2\left(1-\frac{3}{\alpha}\right)B\left(\frac{3}{2\alpha},\frac{1}{2}\right)+\left(1-\frac{5}{\alpha}\right)B\left(\frac{5}{2\alpha},\frac{1}{2}\right)\right]+\cdots\right\}. \tag{IV.20}$$

It is to be noted from (IV.14) that

$$\gamma=\beta x_n^{2(1-\alpha)}=\beta[1-2\beta^{1-\alpha}/2\alpha+\cdots]. \tag{IV.21}$$

By introducing (IV.14) and (IV.21) into (IV.20) we finally obtain the following expression for $F(\beta,\alpha)$:

$$F(\beta,\alpha)=\left(\frac{\alpha}{1+\alpha}\right)B\left(\frac{1}{2\alpha},\frac{1}{2}\right)-\frac{\beta}{8}B\left(\frac{1}{2\alpha},\frac{1}{2}\right)+\frac{(5-\alpha)}{8\alpha}\beta^2 B\left(\frac{5}{2\alpha},\frac{1}{2}\right)+\cdots. \tag{IV.22}$$

When this form is inserted into (IV.11) [remembering that through (IV.12b) β depends on μ_n and $1/\lambda$] the resulting equation can be solved for μ_n in the form of a series expansion in inverse powers of λ. That value of μ_n can, from (IV.5), be converted into the following formula for the nth energy level in the large n, large λ regime:

$$E_n=\lambda^{1/(\alpha+1)}[C(n+\tfrac{1}{2})^{2\alpha/(1+\alpha)}+a\lambda^{-2/(1+\alpha)}(n+\tfrac{1}{2})^{2/(\alpha+1)}+b\lambda^{-4/(\alpha+1)}(n+\tfrac{1}{2})^{-2(\alpha-2)/(\alpha+1)}+\cdots], \tag{IV.23}$$

where now

$$C=2^{(\alpha-2)/(\alpha+1)}[\pi(\alpha+1)\gamma_1]^{2\alpha/(\alpha+1)}, \tag{IV.24a}$$

$$a=2^{-(\alpha-2)/(\alpha+1)}[\pi(\alpha+1)\gamma_1]^{2/(\alpha+1)}\gamma_1\gamma_3^{-1}, \tag{IV.24b}$$

$$b=2^{-(5\alpha-4)/(\alpha+1)}(5\alpha^{-1}-1)[\pi(\alpha+1)\gamma_1]^{-2(\alpha-2)/(\alpha+1)}\times(\gamma_1^2\gamma_3^{-2}-\gamma_1\gamma_5^{-1}), \tag{IV.24c}$$

and

$$\gamma_j=\Gamma(j/\alpha)/[\Gamma(j/2\alpha)]^2. \tag{IV.24d}$$

The numerical values for the various constants in some special cases are

i. $\alpha=2$	ii. $\alpha=3$	
$C=1.376\,507$	$C=1.346\,760$	
$a=0.268\,055$	$a=0.238\,075$	
$b=-0.011\,675$	$b=-0.004\,591$	
iii. $\alpha=4$	**iv. $\alpha=5$**	**v. $\alpha=\infty$**
$C=1.326\,607$	$C=1.312\,307$	$C=1.233\,700$
$a=0.222\,270$	$a=0.212\,348$	$a=0.166\,666$
$b=-0.001\,607$	$b=0$	$b=0.004\,503$

V. SMALL λ, LARGE n REGIME

Since we have completed our investigation of the energy levels associated with large n and large λ, it would be natural to examine the WKB approach to the small λ regime. In our method of doing that, we will incidentally introduce a broader strategy which is motivated by results presented in Ref. 10 in discussion of classical anharmonic oscillators; with this strategy we will obtain formulas for energy levels (for large n) for the full range $0\le\lambda\le\infty$. Our starting point is the unscaled Schrödinger equation (I.23):

$$\psi''+(\mu-y^2-2\lambda y^{2\alpha})\psi=0 \quad \text{with } \mu=2E.$$

This can be rewritten as

$$\psi''+(\mu-\Omega^2y^2+\epsilon[(\Omega^2-1)y^2-2\lambda y^{2\alpha}])\psi=0, \tag{V.1}$$

where Ω can be identified as a renormalized frequency which will be specified later and ϵ is an expansion parameter which will finally be set equal to 1. We will choose Ω in such a way that when $\epsilon=0$ the resulting harmonic oscillator system better mimics the anharmonic oscillator. If $\mu_n=2E_n$ and $\psi=\psi_n$, n being the quantum number of the energy level, then Ω also depends on n as well as λ.

Titchmarsh's formula (IV.3) now has the form

$$n+\tfrac{1}{2}+O(1/n)=(2/\pi)\int_0^{v_n}\{\mu_n-\Omega_n^2y^2-\epsilon[(1-\Omega_n^2)y^2+2\lambda y^{2\alpha}]\}^{1/2}dy, \tag{V.2}$$

where v_n is the value of y which makes the integrand vanish. Now, if we consider the large n regime so that for practical purposes we drop terms of $O(1/n)$ and if we set

$$\Omega_n^2 y^2 = \mu_n x^2 \quad \text{and} \quad \Omega_n^2 v_n^2 = \mu_n x_n^2,$$

then

$$(n+\tfrac{1}{2}) = (2\mu_n/\pi\Omega_n)\int_0^{x_n}[1 - x^2 - \epsilon(Ax^2 + \lambda B x^{2\alpha})]^{1/2}\,dx \tag{V.3}$$

and

$$(1 - x_n^2) - \epsilon(Ax_n^2 + \lambda B x_n^{2\alpha}) = 0, \tag{V.4}$$

where

$$A \equiv (1-\Omega_n^2)/\Omega_n^2 \quad \text{and} \quad B = (2/\mu_n)(\mu_n/\Omega_n^2)^\alpha. \tag{V.5}$$

When the 1 in the integrand of (V.3) is replaced by the combination of parameters which equal the leftmost 1 in (V.4), (V.3) becomes

$$(n+\tfrac{1}{2}) = (2\mu_n/\pi\Omega_n)\int_0^{x_n}\{(x_n^2 - x^2) + \epsilon[A(x_n^2 - x^2) + B\lambda(x_n^{2\alpha} - x^{2\alpha})]\}^{1/2}\,dx. \tag{V.6}$$

If we set $t = (x/x_n)^2$, we have

$$(n+\tfrac{1}{2}) = (\mu_n x_n^2/\pi\Omega_n)\int_0^1 t^{-1/2}(1-t)^{1/2} \times\{1 + \epsilon(A + \lambda B x_n^{2(\alpha-1)}[(1-t^\alpha)/(1-t)])\}^{1/2}\,dt. \tag{V.7}$$

Now let us expand

$$\mu_n = \mu_n^{(0)} + \epsilon\mu_n^{(1)} + \cdots. \tag{V.8a}$$

Then, since B depends on λ according to (V.5),

$$B = B_0 + \epsilon B_1 + \cdots, \tag{V.8b}$$

where

$$B_0 = 2[\mu_n^{(0)}]^{\alpha-1}/\Omega_n^{2\alpha}, \tag{V.9a}$$

$$B_1 = 2(\alpha-1)[\mu_n^{(0)}]^{\alpha-2}\mu_n^{(1)}/\Omega_n^{2\alpha}, \cdots. \tag{V.9b}$$

Also, from (V.4)

$$x_n^2 = 1 - \epsilon(A + \lambda B_0) + \epsilon^2[-\lambda B_1 + (A+\lambda B_0)(A + \lambda\alpha B_0)] + \cdots. \tag{V.10}$$

When (V.10) is substituted into (V.7) and the square root is expanded, it is found to first order in ϵ [noting that $1 - t^\alpha = (1-t) + t(1-t^{\alpha-1})$] that

$$[(n+\tfrac{1}{2})/\mu_n^{(0)}][1 - \epsilon(\mu_n^{(1)}/\mu_n^{(0)}) + O(\epsilon^2)] = (\pi\Omega_n)^{-1}\int_0^1 t^{-1/2}(1-t)^{1/2}\,dt\,\{1 - \tfrac{1}{2}\epsilon(A + \lambda B_0) + \tfrac{1}{2}\epsilon[(t - t^\alpha)/(1-t)]\lambda B_0 + O(\epsilon^2)\}; \tag{V.11}$$

the basic integral

$$\int_0^1 (1-t)^{z-1} t^{w-1}\,dt = \Gamma(z)\Gamma(w)/\Gamma(z+w) \equiv \mathrm{B}(z,w) \tag{V.12}$$

implies that

$$[(n+\tfrac{1}{2})/\mu_n^{(0)}][1 - \epsilon(\mu_n^{(1)}/\mu_n^{(0)}) + \cdots] = (2\Omega_n)^{-1}\{1 - \tfrac{1}{2}\epsilon[A + \lambda B_0(2\alpha)!/2^{2\alpha-1}\alpha!\,\alpha!] + O(\epsilon^2)\}. \tag{V.13}$$

At this point we choose the as yet unspecified value of Ω_n in such a manner that the coefficient of the term proportional to ϵ in (V.6) vanishes; i.e., recalling the definition of A and B_0, we have

$$\Omega_n^{2(\alpha-1)}(\Omega_n^2 - 1) = \lambda(\mu_n^{(0)})^{\alpha-1}C_\alpha \equiv \lambda(2E_n^{(0)})^{\alpha-1}C_\alpha, \tag{V.14}$$

where

$$C_\alpha = (2\alpha)!/[2^{2(\alpha-1)}\alpha!\,\alpha!] \quad \text{and} \quad A = -\tfrac{1}{2}\lambda C_\alpha B_0. \tag{V.15}$$

The vanishing of the term proportional to ϵ on the right-hand side of (V.13) implies the vanishing of the corresponding term on the left; i.e.,

$$\mu_n^{(1)} = 0. \tag{V.16}$$

Some special values of C_α are

$$C_2 = 3/2, \quad C_4 = 35/32, \quad C_3 = 5/4, \quad C_5 = 63/64, \quad \text{etc.} \tag{V.17a}$$

Stirling's approximation of $\alpha!$ yields the large α form

$$C_\alpha \sim 4(\alpha\pi)^{-1/2}. \tag{V.17b}$$

We obtain the zeroth approximation to the nth energy level by equating the constant terms on each side of (V.13):

$$E_n^{(0)} = \Omega_n(n+\tfrac{1}{2}) \quad (\text{since } \mu_n^{(0)} = 2E_n^{(0)}), \tag{V.18}$$

which is the standard form for a harmonic oscillator energy level; however, the Ω_n is a renormalized frequency which is to be determined from (V.14) and which itself depends on the quantum number n. When (V.18) is substituted into (V.14), it is found that

$$\Omega_n^{\alpha-1}(\Omega_n^2 - 1) = \lambda C_\alpha(2n+1)^{\alpha-1}. \tag{V.19}$$

When $\alpha = 2$ the resulting cubic form of this equation is of the "reduced" type for which simple solutions exist. With $C_2 = 3/2$, the form of the solution depends on whether

$$\tfrac{9}{4}\lambda(2n+1)\sqrt{3} > 1 \tag{V.20a}$$

or

$$\tfrac{9}{4}\lambda(2n+1)\sqrt{3} < 1. \tag{V.20b}$$

In case (V.20a) one has the Cardan solution

$$\Omega_n = \{\tfrac{3}{4}\lambda(2n+1) + [\tfrac{9}{16}\lambda^2(2n+1)^2 - \tfrac{1}{27}]^{1/2}\}^{1/3} + \{\tfrac{3}{4}\lambda(2n+1) - [\tfrac{9}{16}\lambda^2(2n+1)^2 - \tfrac{1}{27}]^{1/2}\}^{1/3}, \tag{V.21a}$$

which in case (V.20b) the trigonometric solution is more convenient:

$$\Omega_n = (2/\sqrt{3})\cos\{\tfrac{1}{3}\cos^{-1}[\tfrac{9}{4}\sqrt{3}\,\lambda(2n+1)]\}. \tag{V.21b}$$

When $\alpha = 3$, Eq. (V.19) is quadratic in Ω_n^2 so that

$$\Omega_n^2 = \tfrac{1}{2}\{1 + [1 + 5\lambda(2n+1)^2]^{1/2}\}. \tag{V.22}$$

When $\alpha = 5$, Eq. (V.19) becomes a Cardan "reduced" cubic in the variable Ω_n^{-2}. The Cardan solution yields the result

$$\Omega_n = \left[\left\{\frac{1}{2f} + \left[\left(\frac{1}{3f}\right)^3 + \left(\frac{1}{2f}\right)^2\right]^{1/2}\right\}^{1/3} + \left\{\frac{1}{2f} - \left[\left(\frac{1}{3f}\right)^3 + \left(\frac{1}{2f}\right)^2\right]^{1/2}\right\}^{1/3}\right]^{-2}, \tag{V.23a}$$

where

$$f=\tfrac{63}{64}\lambda(2n+1)^4. \quad (V.23b)$$

Generally, the solution of Eq. (V.19) for any positive integral α depends on the combination

$$\eta_\alpha=\lambda(2n+1)^{\alpha-1}\sim(2n)^{\alpha-1}\lambda \quad \text{for large } n. \quad (V.24)$$

When η_α is either large or small a systematic expansion for Ω_n is easily obtained.

When $\eta_\alpha \ll 1$, $\Omega_n\approx 1$ and iteration of (V.19) yields

$$\Omega_n^2=1+\frac{C_\alpha\eta_\alpha}{[1+C_\alpha\eta_\alpha(1+\eta_\alpha C_\alpha\Omega_n^{1-\alpha})]^{(\alpha-1)/2}}, \quad (V.25)$$

which is, to second order,

$$\Omega_n^2\sim 1+C_\alpha\eta_\alpha[1+C_\alpha\eta_\alpha(1+\eta_\alpha C_\alpha)]^{(1-\alpha)/2}. \quad (V.26)$$

When $\eta_\alpha\gg 1$, $\Omega_n\sim[C_\alpha\eta_\alpha]^{1/(\alpha+1)}$. Iteration yields

$$\Omega_n=\frac{(C_\alpha\eta_\alpha)^{1/(1+\alpha)}}{[1-(C_\alpha\eta_\alpha)^{-2/(1+\alpha)}(1-\Omega_n^{-2})^{2/(1+\alpha)}]^{1/(1+\alpha)}}, \quad (V.27)$$

which to second order is

$$\Omega_n\sim\frac{(C_\alpha\eta_\alpha)^{1/(1+\alpha)}}{[1-(C_\alpha\eta_\alpha)^{-2/(1+\alpha)}]^{1/(1+\alpha)}}. \quad (V.28)$$

Equation (V.18) and the various forms of Ω_n given above give zeroth order estimates of the nth energy level, $E_n^{(0)}$. The frequency Ω_n was chosen so that terms first order in ϵ vanish. We now proceed to the calculation of the second order correction.

Let us substitute (V.10) into (V.7), put μ_n on the left-hand side of the resulting equation and retain terms up to order ϵ^2 on the right. Then we find [again noting that $1-t^\alpha\equiv(1-t)+(t-t^\alpha)$]

$$(n+\tfrac{1}{2})\Omega_n/\mu_n=\{1+\tfrac{1}{2}\epsilon^2(A+\lambda B_0)[\tfrac{3}{4}A+(\alpha-\tfrac{1}{4})\lambda B_0]\}\mathrm{B}(\tfrac{1}{2},\tfrac{3}{2})$$
$$-\tfrac{1}{2}(\alpha+\tfrac{1}{2})\epsilon^2\lambda B_0(A+\lambda B_0)J_2-\tfrac{1}{8}\epsilon^2\lambda^2B_0^2J_3, \quad (V.29)$$

where J_2 and J_3 are the integrals

$$J_2=\int_0^1 t^{-1/2}(t-t^\alpha)(1-t)^{-1/2}\,dt=\mathrm{B}(\tfrac{1}{2},\tfrac{3}{2})-\mathrm{B}(\tfrac{1}{2},\alpha+\tfrac{1}{2}), \quad (V.30a)$$

$$J_3=\int_0^1 t^{-1/2}(t-t^\alpha)^2(1-t)^{-3/2}\,dt$$
$$=-2\int_0^1 t^{-1/2}(t-t^\alpha)^2\,d(1-t)^{-1/2}$$
$$=-2[\tfrac{3}{2}\mathrm{B}(\tfrac{1}{2},\tfrac{3}{2})-2(\alpha+\tfrac{1}{2})\mathrm{B}(\tfrac{1}{2},\alpha+\tfrac{1}{2})$$
$$+(2\alpha-\tfrac{1}{2})\mathrm{B}(\tfrac{1}{2},2\alpha-\tfrac{1}{2})]. \quad (V.30b)$$

After these integral expressions are substituted into (V.29) and B_0 is related to A through (V.15), it is found that

$$2(n+\tfrac{1}{2})\Omega_n/\mu_n=2(n+\tfrac{1}{2})(\Omega_n/\mu_n^0)[1-\epsilon^2(\mu_n^{(2)}/\mu_n^{(0)})+\cdots]$$
$$=1+\tfrac{1}{2}\epsilon^2(\alpha-\tfrac{1}{4})[-A^2+2B_0^2\lambda^2\pi^{-1}\mathrm{B}(\tfrac{1}{2},2\alpha-\tfrac{1}{2})]+O(\epsilon^3). \quad (V.31)$$

When the beta function $\mathrm{B}(s,t)$ is expressed in terms of gamma function [see Eq. (V.18)] and A and B_0 are related by

$$A=-\lambda B_0\mathrm{B}(\tfrac{1}{2},\alpha+\tfrac{1}{2})/\mathrm{B}(\tfrac{1}{2},\tfrac{3}{2}), \quad (V.32)$$

Eq. (V.31) reduces to

$$[(2n+1)\Omega_n/\mu_n^{(0)}][1-\epsilon^2(\mu_n^{(2)}/\mu_n^{(0)})-\cdots]$$
$$=1-\tfrac{1}{8}\epsilon^2A^2(4\alpha-1)\left(1-\frac{\pi^{1/2}}{2}\,\frac{\Gamma(2\alpha-\tfrac{1}{2})\Gamma^2(\alpha+1)}{\Gamma(2\alpha)\Gamma^2(\alpha+\tfrac{1}{2})}\right)+\cdots$$
$$=1-\tfrac{1}{8}\epsilon A^2(4\alpha-1)\left[1-\frac{2\alpha}{(4\alpha-1)}(4\alpha)!\left(\frac{\alpha!}{(2\alpha)!}\right)^4\right]+\cdots. \quad (V.33)$$

If we write the nth energy level as

$$E_n=E_n^{(0)}+\epsilon^2E_n^{(2)}+\epsilon^3E_n^{(3)}+\cdots, \quad (V.34a)$$

and remember that $2E_n\equiv\mu_n$, by equating the coefficients of ϵ^2 of both sides of (V.33) we find [also using the definition of A, (V.32)]

$$E_n^{(2)}/E_n^{(0)}=[(4\alpha-1)/8][(1-\Omega_n^2)\Omega_n^{-2}]^2\{1-[2\alpha/(4\alpha-1)]$$
$$\times(4\alpha)!\,[\alpha!/(2\alpha)!]^4\}. \quad (V.34b)$$

This yields the following results for $\alpha=2,3,$ and 4:

for $\alpha=2$ $\quad E_n^{(2)}/E_n^{(0)}=-(7/72)(1-\Omega_n^{-2})^2,$ (V.35a)

for $\alpha=3$ $\quad E_n^{(2)}/E_n^{(0)}=-(143/400)(1-\Omega_n^{-2})^2,$ (V.35b)

for $\alpha=4$ $\quad E_n^{(2)}/E_n^{(0)}=-(1473/1960)(1-\Omega_n^{-2})^2.$ (V.35c)

So that to second order, for example,

for $\alpha=2$ $\quad E_n=(n+\tfrac{1}{2})\Omega_n[1-\tfrac{7}{72}\epsilon^2(1-\Omega_n^{-2})^2+\cdots]$ (V.36)

Of course, for each α, Ω_n has a different form, as discussed in Eqs. (V.19)–(V.27).

After a straightforward but lengthy calculation, the following expression can be found for $E_n^{(3)}$:

$$E_n^{(3)}/E_n^{(0)}=[(2/C_\alpha)(1-\Omega_n^{-2})]^3(\tfrac{1}{128})$$
$$\times[8(6\alpha-1)(2\alpha-1)R_3+20\alpha(1-4\alpha)R_2C_\alpha$$
$$+(8\alpha^2+3\alpha-1)C_\alpha^3], \quad (V.37)$$

where

$$R_2=\frac{2[2(6\alpha-1)]!}{2^{2(2\alpha-1)}(2\alpha-1)!\,(2\alpha-1)!}, \quad (V.38a)$$

$$R_3=\frac{2(6\alpha-4)!}{2^{2(3\alpha-2)}(3\alpha-2)!\,(3\alpha-2)!}. \quad (V.38b)$$

This yields the following results for $\alpha=2,3,4$:

for $\alpha=2$ $\quad E_n^{(3)}/E_n^{(0)}=\tfrac{1}{8}(1-\Omega_n^{-2})^3,$ (V.39a)

for $\alpha=3$ $\quad E_n^{(3)}/E_n^{(0)}=(449/400)(1-\Omega_n^{-2})^3,$ (V.39b)

for $\alpha=4$ $\quad E_n^{(3)}/E_n^{(0)}=(402\,069/98\,000)(1-\Omega_n^{-2})^3,$ (V.39c)

so that to third order, for example, for $\alpha=2$

$$E_n=(n+\tfrac{1}{2})\Omega_n[1-\tfrac{7}{72}\epsilon^2(1-\Omega_n^{-2})^2+\tfrac{1}{8}\epsilon^3(1-\Omega_n^{-2})^3+\cdots]. \quad (V.40)$$

Let us now consider the numerical aspects of Eqs. (V.18) and (V.35)–(V.40). We first examine the case $\alpha=2$, of course setting $\epsilon=1$ in various expansions. When λ is small, (V.26), the zeroth approximation to E_n is

$$E_n = (n+\tfrac{1}{2})\sqrt{1+3\lambda(n+\tfrac{1}{2})}\,[1-O(\lambda^2)]$$
$$= (n+\tfrac{1}{2}) + \tfrac{3}{2}\lambda(n^2+n+\tfrac{1}{4}) + O(\lambda^2). \quad \text{(V.41)}$$

The term proportional to λ is to be compared with the exact result

$$\tfrac{3}{2}\lambda(n^2+n+\tfrac{1}{2}) \quad \text{(V.42)}$$

obtained in Eq. (I.13). Although (V.41) was derived for the large n regime as small as 4, (V.41) yields the combination

$$n^2+n+\tfrac{1}{4} = 20.25.$$

This deviates only by about a percent from the exact combination

$$n^2+n+\tfrac{1}{2} = 20.50.$$

While the expansions in this section were intended to be useful in the small λ regime, it is amusing to test them in the large λ range. As $\lambda \to \infty$, we use Eq. (V.27) to note that to "zeroth order" (V.40) yields (since $\alpha = 2$ in this discussion)

$$\Omega_n \sim [3\lambda(n+\tfrac{1}{2})]^{1/3}. \quad \text{(V.43)}$$

Hence, to zeroth order

$$E_n^{(0)} \sim 3^{1/3}\lambda^{1/3}(n+\tfrac{1}{2})^{4/3} = (1.4422)\,\lambda^{1/3}(n+\tfrac{1}{2})^{4/3}, \quad \text{(V.44)}$$

where the coefficient $3^{1/3}$ is to be compared to the exact coefficient

$$C = 1.3765\cdots \quad \text{(V.45)}$$

listed below Eq. (IV.24d). It is remarkable that the exponents on λ and $(n+\frac{1}{2})$ are correct in this zeroth order approximation, and that the estimation of C of 1.4422 is only in error by about 5%.

Some improvement in the estimation of C might be expected through the use of the expansion (V.40). In the large λ limit Ω_n becomes large so that Ω_n^{-2} can be neglected compared with 1. Hence, with $\epsilon = 1$ we must consider the correction

$$1-(7/72)+(1/8)-(673/3456)+\cdots \quad \text{(V.46)}$$

[we have found the term of $O(\epsilon^4)$ to be included in the curly brackets to be

$$-(673/3456)\epsilon^4(1-\Omega_n^{-2})^4]. \quad \text{(V.47)}$$

The disappointing feature of (V.46) is that the correction terms of alternating sign are all of about the same order so that there is no indication of a convergent result. However, since the series seems to be somewhat like

$$1/(1+x) = 1-x+x^2\cdots$$

as $x \to 1$, it is suggested that we note the formal equivalence of (V.40) with

$$E_n = (n+\tfrac{1}{2})\,\Omega_n\{1-\tfrac{7}{72}\epsilon^2(1-\Omega_n^{-2})^2[1+\tfrac{9}{7}\epsilon(1-\Omega_n^{-2})$$
$$-\tfrac{823}{2358}\epsilon^2(1-\Omega_n^{-2})^2+\cdots]^{-1}\}. \quad \text{(V.48)}$$

At least in this form we have reached a stage in which terms seem to be starting to decrease in the ϵ expansion as $\epsilon \to 1$ so that there is some hope for the convergence of the series.

In the large λ limit (with $\epsilon = 1$) the terms in the curly bracket become

$$1-(\tfrac{7}{72})(1+\tfrac{9}{7}-\tfrac{823}{2352}+\cdots)^{-1} = 0.94978. \quad \text{(V.49)}$$

When multiplied by the 1.4422 which appears in Eq. (V.44) this gives the estimate 1.3688 to C, a value differs from the exact C of Eq. (V.45) by only 68 parts in 13 000. While this estimate of c is a slight overshoot of the exact value, we expect that further terms in (V.49) would correct for it. Actually the error is slightly less if the term (823/2352) is omitted in (V.49), the estimate of C in that case being 1.3808 which is 43 parts per 13 000 too large. Were the exact C not known, one could not have too much confidence in this result since without the neglected term in (V.49) no indication of a convergent series would have appeared.

When $\alpha = 3$, the equation equivalent to (V.48) is

$$E_n = (n+\tfrac{1}{2})\,\Omega_n\{1-\tfrac{143}{400}\epsilon^2(1-\Omega_n^{-2})^2$$
$$\times[1+\tfrac{449}{143}\epsilon(1-\Omega_n^{-2})-\cdots]^{-1}\}. \quad \text{(V.50)}$$

In this case we have not yet calculated the fourth order term, but in the limit $\lambda \to \infty$ when $\Omega_n^{-1} \to 0$, we have (with $\epsilon = 1$) using (V.24), (V.28), and (V.17a)

$$E_n = (n+\tfrac{1}{2})^{3/2}\lambda^{1/4}5^{1/4}[1-\tfrac{143}{400}(1+\tfrac{449}{143}-\cdots)^{-1}]$$
$$\approx 1.366\ (n+\tfrac{1}{2})^{3/2}\lambda^{1/4}.$$

The coefficient 1.366 compares favorably with the exact result 1.347 listed as C in the $\alpha = 3$ table below Eq. (IV.24d).

When $\alpha = 4$,

$$E_n = (n+\tfrac{1}{2})\,\Omega_n\{1-(1473/1960)\epsilon^2(1-\Omega_n^{-2})^2$$
$$\times[1+(143\,023/24\,550)\epsilon(1-\Omega_n^{-2})-\cdots]^{-1}\}. \quad \text{(V.51)}$$

As $\lambda \to \infty$, with $\epsilon = 1$ we obtain

$$E_n = (n+\tfrac{1}{2})^{3/5}\lambda^{1/5}(35/4)^{1/5}\{1-(1473/1960)$$
$$\times[1+(143\,023/24\,550)-\cdots]^{-1}\}$$
$$\approx 1.373\lambda^{1/5}\ (n+\tfrac{1}{2})^{3/5}.$$

When the coefficient 1.373 is compared to the exact result 1.327, it is evident that a calculation terminated at third order when $\alpha = 4$ is not as accurate as similar calculations for $\alpha = 2$ and 3. This is not surprising because the larger α, the more difficult is should be to find an equivalent harmonic oscillator.

Equations (V.48), (V.50), and (V.51) can be rewritten as

$$\left(\frac{(n+\frac{1}{2})\Omega_n}{E_n-(n+\frac{1}{2})\Omega_n}\right) = \epsilon^2u^2[1+\zeta_1(\alpha)\,\epsilon u+\zeta_2(\alpha)\,\epsilon^2u^2+\cdots],$$

where $\zeta_j(\alpha)$ is a set of coefficients in a series in ϵu with

$$u = (1-\Omega_n^{-2}).$$

Our numerical analysis indicates that while the series expansion of the relative deviation of E_n from $(n+\frac{1}{2})\Omega_n$ converges slowly if at all, that for the inverse seems to converge more rapidly.

An important feature of the general formulas of this section is that the quantity $E_n/(n+\frac{1}{2})$ depends only on the combination of n and λ in the form of η_α [see Eq. (V.24)]

$$2^{1-\alpha}\eta_\alpha = \lambda(n+\tfrac{1}{2})^{\alpha-1}.$$

Hence one should not consider λ to be the perturbation parameter in our anharmonic oscillator system, but rather the combination $\lambda(n+\frac{1}{2})^{\alpha-1}$. The "weak" coupling range for λ thus depends on the energy level that one is considering. The larger n, the smaller λ must be to be in the weak coupling range. This is consistent with classical theory of anharmonic oscillator dynamics in that the frequency of oscillation of an anharmonic oscillator depends on the initial conditions on the oscillator displacement from equilibrium. Big initial displacement and velocities are the analog of the large quantum number regime. The importance of the combination λn in the case $\alpha=2$ was noticed independently by Halpern.[11]

APPENDIX A

In this appendix, we first show how $(a^\dagger + a)^p$, where $a^\dagger$ and a are the boson creation and annihilation operators, can be expressed in terms of the normally ordered products, and from this we show how the matrix elements given in Eq. (II.5) are readily obtained.

The normally ordered product of $(a^\dagger+a)^p$ is defined as

$$:(a^\dagger+a)^p: = \sum_{j=0}^{p}\binom{p}{j}(a^\dagger)^{p-j}a^j. \tag{A1}$$

By repeated use of the commutation relation

$$[a, a^\dagger] = 1 \tag{A2}$$

we readily find, for the first few values of $p=2,3,4$ (and writing ϕ for $a^\dagger + a$),

$$\phi^2 = :\phi^2: + 1, \tag{A3}$$

$$\phi^3 = :\phi^3: + 3:\phi: \tag{A4}$$

$$\phi^4 = :\phi^4: + 6:\phi^2: + 3. \tag{A5}$$

We now postulate that, for any $p \geq 1$, ϕ^p can be written as

$$\phi^p = \sum_{j=0}^{[p/2]} c_j^p :\phi^{p-2j}: \tag{A6}$$

and we attempt to determine the possibility of a solution for c_j^p.

Multiplying (A6) from the left by ϕ and using the relation

$$a(a^\dagger)^m = m(a^\dagger)^{m-1} + (a^\dagger)^m a, \tag{A7}$$

we find

$$\begin{aligned}(a^\dagger+a)(a^\dagger+a)^p &= \sum_{j=0}^{[p/2]} c_j^p \sum_{i=0}^{p-2j}\binom{p-2j}{i}\,[(a^\dagger)^{p-2j-i+1}a^i \\ &\quad + (p-2j-i)(a^\dagger)^{p-2j-i-1}a^i + (a^\dagger)^{p-2j-i}a^{i+1}] \\ &= \sum_{j=0}^{[(p+1)/2]} c_j^p \sum_{i=0}^{p+1-2j}\binom{p+1-2j}{i}(a^\dagger)^{p+1-2j-i}a^i \\ &\quad + (p-2j)\sum_{i=0}^{p-1-2j}\binom{p-1-2j}{i}(a^\dagger)^{p-1-2j-i}a^i\Big)\end{aligned} \tag{A8}$$

or

$$\begin{aligned}\phi^{p+1} &= \sum_{j=0}^{[(p+1)/2]}[c_j^p:\phi^{p+1-2j}: + (p-2j)\,c_j^p:\phi^{p-1-2j}:] \\ &= \sum_{j=0}^{[(p+1)/2]}[c_j^p + (p-2j+2)\,c_{j-1}^p]:\phi^{p+1-2j}: \\ &= \sum_{j=0}^{[(p+1)/2]} c_j^{p+1}:\phi^{p+1-2j}:,\end{aligned} \tag{A9}$$

from which we find that c_j^p must satisfy the recurrence relation

$$c_j^p + \{p-2(j-1)\}c_{j-1}^p = c_j^{p+1}. \tag{A10}$$

This recurrence relation is somewhat similar to the recurrence relation for the binomial coefficients

$$C_j^p + C_{j-1}^p = C_j^{p+1}. \tag{A11}$$

After a little thought, we find

$$c_j^p = [p!/(p-2j)!\,j!\,]\,2^{-j}. \tag{A12}$$

Thus we obtain

$$(a^\dagger+a)^p = \sum_{j=0}^{[p/2]}\frac{p!}{(p-2j)!\,j!}\,2^{-j}\sum_{i=0}^{p-2j}\binom{p-2j}{i}(a^\dagger)^{p-2j-i}a^i. \tag{A13}$$

The expression (A12) was stated by Baker[12] and was known to others.

Now, to derive our basic matrix equation using the Bargmann representation, we note that the operator $z^{g_1}(d/dz)^{g_2}$ with $g_1 - g_2 = 2\alpha - 2h$ in the Hamiltonian would contribute a term

$$\prod_{l=0}^{g_2-1}(n+k-l)$$

to the matrix element

$$c_{n+k}^{(2h-2\alpha)}(2\alpha)\ \left(\prod_{l=0}^{-1}(\cdots)\text{ being defined as }1\right)$$

and the collection of operators with $g_1 - g_2 = 2\alpha - 2h$ is, from (A13), given by (letting $p = 2\alpha$)

$$\frac{1}{2^\alpha}\sum_{j=0}^{h}\frac{(2\alpha)!}{(2\alpha-2j)!\,j!}\,2^{-j}\binom{2\alpha-2j}{h-j}\,z^{2\alpha-h-j}\left(\frac{d}{dz}\right)^{h-j} \tag{A14}$$

and hence

$$c_{n+k}^{(2h-2\alpha)}(2\alpha) = \frac{1}{2^\alpha}\sum_{j=0}^{h}\frac{(2\alpha)!}{(2\alpha-2j)!\,j!}\,2^{-j}\binom{2\alpha-2j}{h-j} \times \prod_{l=0}^{h-j-1}(n+k-l),\quad h=0,1,\ldots,2\alpha. \tag{A15}$$

APPENDIX B

In this appendix, we give the explicit expressions for the matrix elements of G in Eq. (II.11) for $\alpha = 3$ and 4.

$\alpha = 3$

$$a_k^{(-6)} = \tfrac{1}{8}\lambda,$$
$$a_k^{(-4)} = \tfrac{3}{4}\lambda(n+k+\tfrac{5}{2}),$$
$$a_k^{(-2)} = \tfrac{15}{8}\lambda[(n+k)(n+k+3)+3],$$
$$a_k^{(0)} = \tfrac{5}{4}\lambda\{(n+k)[(n+k)(2n+2k+3)+4]+\tfrac{3}{2}\},$$
$$a_k^{(2)} = \tfrac{15}{8}\lambda\{(n+k)(n+k-1)[(n+k+1)(n+k-2)+3]\},$$
$$a_k^{(4)} = \tfrac{3}{4}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3)(n+k-\tfrac{3}{2}),$$
$$a_k^{(6)} = \tfrac{1}{8}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3)(n+k-4)\times(n+k-5),$$
$$\gamma_k = k + a_k^{(0)} - A(\lambda).$$

$\alpha = 4$

$$a_k^{(-8)} = \tfrac{1}{16}\lambda,$$
$$a_k^{(-6)} = \tfrac{1}{2}\lambda(n+k+\tfrac{7}{2}),$$
$$a_k^{(-4)} = \tfrac{7}{8}\lambda[2(n+k)(n+k+5)+15],$$
$$a_k^{(-2)} = \tfrac{7}{4}\lambda\{(n+k)[(n+k-1)(2n+2k+11)+30]+15\},$$
$$a_k^{(0)} = \tfrac{35}{16}\lambda\{2(n+k)[(n+k-1)((n+k-2)(n+k+5)+18)+12]+3\},$$
$$a_k^{(2)} = \tfrac{7}{4}\lambda(n+k)(n+k-1)\{(n+k-2)[(n+k-3)\times[2(n+k-4)+15]+30]+15\},$$
$$a_k^{(4)} = \tfrac{7}{8}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3)\times[2(n+k-4)(n+k+1)+15],$$
$$a_k^{(6)} = \tfrac{1}{4}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3)(n+k-4)\times(n+k-5)(2n+2k-5),$$
$$a_k^{(8)} = \tfrac{1}{16}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3)\times(n+k-4)(n+k-5)(n+k-6)(n+k-7),$$
$$\gamma_k = k + a_k^{(0)} - A(\lambda).$$

APPENDIX C: ANALYTICAL APPROXIMATIONS FOR THE BOUNDARY LAYERS OF THE OSCILLATOR WITH 2α-ic ANHARMONICITY

1. Boundary layer for the deviation from the purely harmonic oscillator in the limit $n\to\infty$, $\lambda\to 0$, $n^{\alpha-1}\lambda = \text{const}$. We shall derive that [see Eqs. (C1–9) below]

$$\frac{E_n - n}{n} = \frac{A(\lambda)}{n} \approx \frac{1}{2^\alpha}\frac{(2\alpha)!}{(\alpha!)^2}(n^{\alpha-1}\lambda) - \frac{1}{2^{2\alpha+1}}\left\{\frac{(4\alpha)!}{[(2\alpha)!]^2} - \frac{[(2\alpha)!]^2}{(\alpha!)^4}\right\} \times\left(\frac{2\alpha + (1/2^\alpha)[(2\alpha)!/(\alpha!)^2]\alpha(\alpha+1)(n^{\alpha-1}\lambda)}{\{1+(1/2^\alpha)[(2\alpha)!/(\alpha!)^2]\alpha(n^{\alpha-1}\lambda)\}^2}\right)(n^{\alpha-1}\lambda)^2.$$

If this is set to be equal to 10%, say, then solving (a cubic equation) for $n^{\alpha-1}\lambda$, we find the solution $n^{\alpha-1}\lambda = \beta$, from which we determine

$$\lambda_1 E_n^{\alpha-1}(\lambda_1) \approx (n^{\alpha-1}\lambda)\left(1 + (\alpha-1)\frac{A(\lambda)}{n} + \frac{(\alpha-1)(\alpha-2)}{2!}\left(\frac{A(\lambda)}{n}\right)^2 + \cdots\right)$$

by substituting $n^{\alpha-1}\lambda = \beta$ in the expression.

2. Boundary layer for the deviation from the purely $2\alpha - ic$ oscillator. From Eq. (IV.23) in the text, we find

$$\frac{E_n - Cn^{2\alpha/(\alpha+1)}\lambda^{1/(\alpha+1)}}{Cn^{2\alpha/(\alpha+1)}\lambda^{1/(\alpha+1)}} \approx \frac{an^{2/(\alpha+1)}\lambda^{-1/(\alpha+1)} + bn^{-2(\alpha-2)/(\alpha+1)}\lambda^{-3/(\alpha+1)}}{Cn^{2\alpha/(\alpha+1)}\lambda^{1/(\alpha+1)}} = \frac{a}{C}\frac{1}{(n^{\alpha-1}\lambda)^{2/(\alpha+1)}} + \frac{b}{C}\frac{1}{(n^{\alpha-1}\lambda)^{4/(\alpha+1)}} = \frac{a}{C}\frac{1}{x} + \frac{b}{C}\frac{1}{x^2},$$

where $x = (n^{\alpha-1}\lambda)^{2/(\alpha+1)}$. Thus if this is set to be equal to 10%, we get a quadratic equation in x:

$$x^2 - \left(\frac{a}{0.1C}\right)x - \frac{b}{0.1C} = 0,$$

from which the desired solution is

$$x = \frac{1}{2}\left\{\frac{a}{0.1C} + \left[\left(\frac{a}{0.1C}\right)^2 + \frac{4b}{0.1C}\right]^{1/2}\right\}.$$

Then we determine

$$\lambda_2^{1/(\alpha-1)}E_n(\lambda_2) \approx Cx^{\alpha/(\alpha-1)} + ax^{1/(\alpha-1)} + bx^{-(\alpha-2)/(\alpha-1)}.$$

Now we outline the steps leading to our expression for $A(\lambda)/n$ to the second order in $n^{\alpha-1}\lambda$ in the limit $n\to\infty$, $\lambda\to 0$, $n^{\alpha-1}\lambda = $ small constant.

The Schweinsian expansion for $A(\lambda)$ to the order λ^2 is given by

$$A(\lambda) = a_0^{(0)} - \left(\frac{a_0^{(2\alpha)}a_{-2\alpha}^{(-2\alpha)}}{\zeta_{-2\alpha}} + \frac{a_0^{(2\alpha-2)}a_{-2\alpha+2}^{(-2\alpha+2)}}{\zeta_{-2\alpha+2}} + \cdots + \frac{a_0^{(2)}a_{-2}^{(-2)}}{\zeta_{-2}} + \frac{a_0^{(-2)}a_2^{(2)}}{\zeta_2} + \cdots + \frac{a_0^{(-2\alpha+2)}a_{2\alpha-2}^{(2\alpha-2)}}{\zeta_{2\alpha-2}} + \frac{a_0^{(-2\alpha)}a_{2\alpha}^{(2\alpha)}}{\zeta_{2\alpha}}\right), \tag{C1}$$

where

$$\zeta_k = k + a_k^{(0)} - a_0^{(0)} \tag{C2}$$

and

$$a_k^{(2h-2\alpha)} = \frac{\lambda}{2^\alpha}\sum_{j=0}^{h}\left[\frac{(2\alpha)!}{(2\alpha-2j)!\,j!}2^{-j}\binom{2\alpha-2j}{h-j}\times\prod_{l=0}^{h-j-1}(n+k-l)\right],\quad h=0,1,\ldots,2\alpha. \tag{C3}$$

From Eq. (C3), we find

$$a_k^{(2h-2\alpha)} = \frac{\lambda}{2^\alpha}\left[\binom{2\alpha}{h}\left(n^h + n^{h-1}\sum_{l=0}^{h-1}(k-l)+\cdots\right) + \frac{(2\alpha)!}{(2\alpha-2)!\,1!}2^{-1}\binom{2\alpha-2}{h-1}(n^{h-1}+\cdots)+\cdots\right] = \frac{\lambda}{2^\alpha}\left[\binom{2\alpha}{h}n^h + \frac{1}{2}\binom{2\alpha}{h}\times h(2\alpha+2k-2h+1)n^{h-1}+\cdots\right], \tag{C4}$$

$$a_0^{(2\alpha-2h)}a_{2h-2\alpha}^{(2h-2\alpha)} \equiv a_0^{2(-h+2\alpha)-2\alpha}a_{2h-2\alpha}^{(2h-2\alpha)}$$

$$= \frac{\lambda^2}{2^{2\alpha}} \left[\binom{2\alpha}{-h+2\alpha} n^{-h+2\alpha} + \frac{1}{2} \binom{2\alpha}{-h+2\alpha} \times (-h+2\alpha)(2h-2\alpha+1) n^{-h+2\alpha-1} + \cdots \right] \times \left[\binom{2\alpha}{h} n^h + \frac{1}{2} \binom{2\alpha}{h} h(2h-2\alpha+1) n^{h-1} + \cdots \right]$$

$$= \frac{\lambda^2}{2^{2\alpha}} \binom{2\alpha}{h} \binom{2\alpha}{-h+2\alpha} \times [n^{2\alpha} + \alpha(2h-2\alpha+1) n^{2\alpha-1} + \cdots],$$

$$h = 0, 1, \ldots, 2\alpha. \quad \text{(C5)}$$

To get ζ_k in Eq. (C2), we find

$$a_k^{(0)} = \frac{\lambda}{2^\alpha} \left[\binom{2\alpha}{\alpha} \left(n^\alpha + n^{\alpha-1} \sum_{l=0}^{\alpha-1} (k-l) + n^{\alpha-2} \sum_{l_1=0}^{\alpha-1} \sum_{l_2>l_1}^{\alpha-1} (k-l_1)(k-l_2) + \cdots \right) + \frac{(2\alpha)!}{(2\alpha-2)!\,1!} 2^{-1} \binom{2\alpha-2}{\alpha-1} \left(n^{\alpha-1} + n^{\alpha-2} \sum_{l=0}^{\alpha-2} (k-l) + \cdots \right) + \frac{(2\alpha)!}{(2\alpha-4)!\,2!} 2^{-2} \binom{2\alpha-4}{\alpha-2} (n^{\alpha-2} + \cdots) + \cdots \right],$$

$$a_k^{(0)} - a_0^{(0)} = \frac{\lambda}{2^\alpha} \left[0 + \binom{2\alpha}{\alpha} \left(n^{\alpha-1} \sum_{l=0}^{\alpha-1} k + n^{\alpha-2} \times \sum_{l_1=0}^{\alpha-1} \sum_{l_2>l_1}^{\alpha-1} [k^2 - (l_1+l_2)k] + \cdots \right) + \frac{1}{2} \binom{2\alpha}{\alpha} \alpha^2 \left(0 + n^{\alpha-2} \sum_{l=0}^{\alpha-2} k + \cdots \right) + \cdots (\cdots + \cdots) \right]. \quad \text{(C7)}$$

Now

$$\sum_{l=0}^{\alpha-1} k = \alpha k, \quad \sum_{l=0}^{\alpha-2} k = (\alpha-1)k, \quad \sum_{l_1=0}^{\alpha-1} \sum_{l_2>l_1}^{\alpha-1} k^2 = \frac{\alpha(\alpha-1)}{2} k^2,$$

and

$$\sum_{l_1=0}^{\alpha-1} \sum_{l_2>l_1}^{\alpha-1} (l_1+l_2) k = \frac{\alpha(\alpha-1)^2}{2} k.$$

Thus we find, after substituting these into Eq. (C7),

$$a_k^{(0)} - a_0^{(0)} = \frac{\lambda}{2^\alpha} \binom{2\alpha}{\alpha} [\alpha k n^{\alpha-1} + \tfrac{1}{2}\alpha(\alpha-1)k(k+1) n^{\alpha-2} + \cdots]. \quad \text{(C8)}$$

A straightforward algebra now gives

$$\frac{a_0^{(-2h+2\alpha)} a_{2h-2\alpha}^{(2h-2\alpha)}}{\zeta_{2h-2\alpha}} + \frac{a_0^{(2h-2\alpha)} a_{2\alpha-2h}^{(2\alpha-2h)}}{\zeta_{2\alpha-2h}} = \frac{\lambda^2}{2^{2\alpha}} \binom{2\alpha}{h} \binom{2\alpha}{2\alpha-h} \times \left(\frac{2\alpha n^{2\alpha-1} + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha(\alpha+1) n^{3\alpha-2}}{[1 + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha n^{\alpha-1}]^2} \right),$$

$$A(\lambda) = \frac{\lambda}{2^\alpha} \binom{2\alpha}{\alpha} n^\alpha - \sum_{h=0}^{\alpha-1} \left(\frac{a_0^{(-2h+2\alpha)} a_{2h-2\alpha}^{(2h-2\alpha)}}{\zeta_{2h-2\alpha}} + \frac{a_0^{(2h-2\alpha)} a_{2\alpha-2h}^{(2\alpha-2h)}}{\zeta_{2\alpha-2h}} \right)$$

$$= \frac{\lambda}{2^\alpha} \binom{2\alpha}{\alpha} n^\alpha - \frac{\lambda^2}{2^{2\alpha}} \times \left(\frac{2\alpha n^{2\alpha-1} + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha(\alpha+1) n^{3\alpha-2}}{[1 + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha n^{\alpha-1}]^2} \right) \times \sum_{h=0}^{\alpha-1} \binom{2\alpha}{h} \binom{2\alpha}{2\alpha-h} + \cdots$$

$$= \frac{\lambda}{2^\alpha} \binom{2\alpha}{\alpha} n^\alpha - \frac{\lambda^2}{2^{2\alpha+1}} \left(\frac{(4\alpha)!}{[(2\alpha)!]^2} - \frac{[(2\alpha)!]^2}{[\alpha!]^4} \right) \times \left(\frac{2\alpha n^{2\alpha-1} + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha(\alpha+1) n^{3\alpha-2}}{[1 + (\lambda/2^\alpha)\binom{2\alpha}{\alpha} \alpha n^{\alpha-1}]^2} \right) + \cdots \quad \text{(C9)}$$

APPENDIX D

In this appendix, we present the elliptic integral solution for the energy levels of the oscillator with mixed quartic and sextic anharmonicities. Consider

$$H = \tfrac{1}{2}(p^2 + x^2) + \lambda' x^4 + \lambda x^6, \quad \lambda', \lambda > 0, \quad \text{(D1)}$$

or by a scale change

$$H = \lambda^{1/4} [\tfrac{1}{2}(p^2 + \lambda^{-1/2} x^2) + \lambda' \lambda^{-3/4} x^4 + x^6]. \quad \text{(D2)}$$

Consider Titchmarsh's formula

$$n + \tfrac{1}{2} + O(1/n) = (2/\pi) \int_0^{x_0} (\mu_n - \xi x^2 - \rho x^4 - \eta x^6)^{1/2} \, dx, \quad \text{(D3)}$$

x_0 being the positive root of $\mu_n - \xi x^2 - \rho x^4 - \eta x^6 = 0$. We can discuss the small λ case with the choice

$$\xi = 1, \quad \rho = 2\lambda', \quad \text{and} \quad \eta = 2\lambda, \quad \text{(D4)}$$

in which case

$$E_n = \tfrac{1}{2}\mu_n; \quad \text{(D5)}$$

or we can discuss the large λ case with the choice

$$\xi = \lambda^{-1/2}, \quad \rho = 2\lambda' \lambda^{-3/4}, \quad \text{and} \quad \eta = 2, \quad \text{(D6)}$$

in which case

$$E_n = \tfrac{1}{2}\lambda^{1/4} \mu_n. \quad \text{(D7)}$$

By a change of variable $t = x^2$, Eq. (D3) may be written as

$$n + \tfrac{1}{2} + O(1/n) = \frac{1}{\pi \eta^{1/2}} \int_0^{t_0} (\mu_n - \xi t - \rho t^2 - \eta t^2) \times \left[t \left(\frac{\mu_n}{\eta} - \frac{\xi}{\eta} t - \frac{\rho}{\eta} t^2 - t^3 \right) \right]^{-1/2} dt. \quad \text{(D8)}$$

The discriminant of the cubic equation

$$t^3 + \frac{\rho}{\eta} t^2 + \frac{\xi}{\eta} t - \frac{\mu_n}{\eta} = 0 \quad \text{(D9)}$$

is

$$\Delta = \frac{1}{4} \left(\frac{2\rho^3}{27\eta^3} - \frac{\rho\xi}{3\eta^2} - \frac{\mu_n}{\eta} \right)^2 + \frac{1}{27} \left(\frac{\xi}{\eta} - \frac{\rho^2}{3\eta^2} \right)^3. \quad \text{(D10)}$$

It follows that if

$$3\eta\xi \geq \rho^2, \quad \text{(D11)}$$

then Δ is always positive, and hence the cubic equation (D9) has one real root and two complex roots and Eq. (D8) may be written as

$$n + \tfrac{1}{2} + O\left(\frac{1}{n}\right) = \frac{1}{\pi \eta^{1/2}} \int_0^a (\mu_n - \xi t - \rho t^2 - \eta t^2) \times [t(a-t)(t-c)(t-\bar{c})]^{-1/2} dt, \quad \text{(D12)}$$

where by defining

$$X=\left\{\frac{\mu_n}{2\eta}+\frac{\rho\xi}{6\eta^2}-\frac{\rho^3}{27\eta^3}+\left[\frac{1}{4}\left(\frac{\mu_n}{\eta}+\frac{\rho\xi}{3\eta^2}-\frac{2\rho^3}{27\eta^3}\right)^2+\frac{1}{27}\left(\frac{3\eta\xi-\rho^2}{3\eta^2}\right)^3\right]^{1/2}\right\}^{1/3}, \quad \text{(D13)}$$

$$Y=\left\{\frac{\mu_n}{2\eta}+\frac{\rho\xi}{6\eta^2}-\frac{\rho^3}{27\eta^3}-\left[\frac{1}{4}\left(\frac{\mu_n}{\eta}+\frac{\rho\xi}{3\eta^2}-\frac{2\rho^3}{27\eta^3}\right)^2+\frac{1}{27}\left(\frac{3\eta\xi-\rho^2}{3\eta^2}\right)^3\right]^{1/2}\right\}^{1/3}, \quad \text{(D14)}$$

the roots a, c, and $\bar{c}$ are given by

$$a=-\frac{\rho}{3\eta}+X+Y, \quad \text{(D15)}$$

$$c=-\frac{\rho}{3\eta}-\frac{X+Y}{2}+\frac{X-Y}{2}\sqrt{3}i, \quad \text{(D16)}$$

$$\bar{c}=-\frac{\rho}{3\eta}-\frac{X+Y}{2}-\frac{X-Y}{2}\sqrt{3}i. \quad \text{(D17)}$$

Following Byrd and Friedman's *Table of Elliptic Integrals* (B—F), p. 133, we define

$$A^2=\tfrac{3}{4}[3(X+Y)^2+(X-Y)^2], \quad \text{(D18a)}$$

$$B^2=[\rho/3\eta+\tfrac{1}{2}(X+Y)]^2+\tfrac{3}{4}(X-Y)^2, \quad \text{(D18b)}$$

$$k^2=(1/4AB)[(-\rho/3\eta+X+Y)^2-(A-B)^2], \quad \text{(D19)}$$

$$\alpha=(A-B)/(A+B), \quad \text{(D20)}$$

$$g=1/\sqrt{AB}. \quad \text{(D21)}$$

By using formula 259.03 on p. 133 of B—F, Eq. (D12) becomes

$$n+\tfrac{1}{2}+O(1/n)=(1/\pi\eta^{1/2})(\mu_n S_0-\xi S_1-\rho S_2-\eta S_3), \quad \text{(D22)}$$

where

$$S_m=\frac{g(\alpha B)^m}{(A-B)^m}\sum_{j=0}^{m}\frac{(-1)^{m-j}(\alpha+1)^j\, m!}{(m-j)!\,j!}R_j \quad \text{(D23)}$$

and

$$R_0=2K(k) \quad \text{(D24)}$$

$$R_1=\frac{2}{1-\alpha^2}\,\Pi\left(\frac{\alpha^2}{\alpha^2-1},k\right) \quad \text{(D25)}$$

$$R_2=\frac{1}{(\alpha^2-1)(k^2+\alpha^2k'^2)}\{[\alpha^2(2k^2-1)-2k^2]R_1+2(k^2+\alpha^2k'^2)K(k)-2\alpha^2E(k)\} \quad \text{(D26)}$$

$$R_3=\frac{1}{2(\alpha^2-1)(k^2+\alpha^2k'^2)}\{-3[\alpha^2(1-2k^2)+2k^2]R_2+(6k^2+\alpha^2-2k^2\alpha^2)R_1-2k^2R_0\}, \quad \text{(D27)}$$

and where $K(k)$, $E(k)$ and $\Pi(\alpha^2/(\alpha^2-1),k)$ are the complete elliptic integrals of the first, second, and third kinds, respectively, and $k'^2=1-k^2$. On the other hand, if

$$3\eta\xi<\rho^2, \quad \text{(D28)}$$

then two different cases may arise:

Case (i):

$$\mu_n>(1/27\eta^2)\{\rho(2\rho^2-9\eta\xi)+2[(\rho^2-3\eta\xi)^3]^{1/2}\}. \quad \text{(D29)}$$

Then the discriminant Δ remains positive, and thus the previous analysis is valid and the final formula is Eq. (D22).

Case (ii):

$$(1/27\eta^2)\{\rho(2\rho^2-9\eta\xi)+2[(\rho^2-3\eta\xi)^3]^{1/2}>\mu_n>0. \quad \text{(D30)}$$

In this case, Δ becomes negative, and the cubic Eq. (D9) has three real roots given by

$$a=-\frac{\rho}{3\eta}+2\left[\frac{1}{27}\left(\frac{\rho^2-3\eta\xi}{3\eta^2}\right)^3\right]^{1/6}\cos\left(\frac{\theta}{3}\right), \quad \text{(D31)}$$

$$c=-\frac{\rho}{3\eta}+2\left[\frac{1}{27}\left(\frac{\rho^2-3\eta\xi}{3\eta^2}\right)^3\right]^{1/6}\cos\left(\frac{\theta+4\pi}{3}\right), \quad \text{(D32)}$$

$$d=-\frac{\rho}{3\eta}+2\left[\frac{1}{27}\left(\frac{\rho^2-3\eta\xi}{3\eta^2}\right)^3\right]^{1/6}\cos\left(\frac{\theta+2\pi}{3}\right), \quad \text{(D33)}$$

$\cos\theta$ being given by

$$\cos\theta=\left(\frac{\mu_n}{2\eta}+\frac{\rho\xi}{6\eta^2}-\frac{\rho^3}{27\eta^3}\right)\Bigg/\left[\frac{1}{27}\left(\frac{\rho^2-3\xi\eta}{3\eta^2}\right)^3\right]^{1/2}. \quad \text{(D34)}$$

Note that a is positive while c and d are negative and $a>c>d$. Thus, in this case, we write Eq. (D8) as

$$n+\tfrac{1}{2}+O\left(\frac{1}{n}\right)=\frac{1}{\pi\eta^{1/2}}\int_0^a(\mu_n-\xi t-\rho t^2-\eta t^3)\times[(a-t)\,t(t-c)(t-d)]^{-1/2}\,dt. \quad \text{(D35)}$$

Following Byrd and Friedman's *Table of Elliptic Integrals*, p. 124, let us define

$$k^2=\frac{a(c-d)}{(a-c)(-d)},\quad \alpha^2=\frac{a}{d}<0,\quad\text{and}\quad g=\frac{2}{[(a-c)(-d)]^{1/2}} \quad \text{(D36)}$$

By using formulas 257.11 on p. 125 of B—F, Eq. (D35) becomes

$$n+\tfrac{1}{2}+O(1/n)=(g/\pi\eta^{1/2})(\mu_nZ_0-\xi aZ_1-\rho a^2Z_2-\eta a^3Z_3), \quad \text{(D37)}$$

where

$$Z_m=\frac{1}{\alpha^{2m}}\sum_{j=0}^{m}\frac{(\alpha^2-1)^j m!}{j!\,(m-j)!}V_j \quad \text{(D38)}$$

and

$$V_0=K(k),$$
$$V_1=\Pi(\alpha^2,k),$$
$$V_2=[2(\alpha^2-1)(k^2-\alpha^2)]^{-1}[\alpha^2E(k)+(k^2-\alpha^2)K(k)+(2\alpha^2k^2+2\alpha^2-\alpha^4-3k^2)\,\Pi(\alpha^2,k)]$$
$$V_3=[4(1-\alpha^2)(k^2-\alpha^2)]^{-1}[k^2V_0+2(\alpha^2k^2+\alpha^2-3k^2)V_1+3(\alpha^4-2\alpha^2k^2-2\alpha^2+3k^2)V_2]. \quad \text{(D39)}$$

TABLE XI. Boundary regions for the first six energy levels for sextic anharmonicity. The parameter λ_1 is the value of λ for which deviation from purely harmonic energy levels has reached 10%; λ_2 is the value of λ for which deviation from purely sextic oscillator energy levels has reached 10% (as λ is reduced from ∞).

	Boundary of harmonic regime			Boundary of sextic regime		
n	λ_1	$E_n(\lambda_1)$	$\lambda_1^{1/2}E_n(\lambda_1)$	λ_2	$E_n(\lambda_2)$	$\lambda_2^{1/2}E_n(\lambda_2)$
0	0.04535	0.55	0.1170	3.459	1.021	1.899
1	0.01658	1.65	0.2124	1.314	3.038	3.483
2	0.008176	2.75	0.2486	0.4850	4.952	3.449
3	0.004382	3.85	0.2550	0.2474	6.889	3.426
4	0.002732	4.95	0.2587	0.1501	8.840	3.425
5	0.001926	6.05	0.2655	0.1005	10.79	3.424
∞			0.2814			3.424

A discussion of the mixed quartic and sextic case was also given by Lakshmanan[8] from the point of view of quantizing the classical result.

For oscillator with only the sextic anharmonicity, the formula is thus (D22) with $\rho = 0$.

[1] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Rev. D **7**, 1620 (1973).

[2] F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975). Misprints in this paper: Eq. (I.19): Last inequality should read if $\lambda \geq 1$. Table II B.: Second column heading should be $E_3(\lambda)$. Table IV.: The value of ϵ_0 should be 0.667 986 259.

[3] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); Rev. Mod. Phys. **34**, 829 (1962).

[4] C. M. Bender, J. Math. Phys. **11**, 796 (1970).

[5] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).

[6] M. Lakshmanan and J. Prabhakaran, Lett. Nuovo Cimento **7**, 689 (1973).

[7] T. T. Truong, J. Chem. Phys. **16**, 1034 (1975).

[8] M. Lakshmanan, Lett. Nuovo Cimento **8**, 743 (1973).

[9] E. C. Titchmarsh, *Eigenfunction Expansions, Part I* (Oxford U. P., Oxford, 1961), 2nd ed.

[10] C. Eminhizer, R. Helleman, and E. Montroll, J. Math. Phys. **17**, 121 (1976).

[11] F. R. Halpern, private communication.

[12] G. A. Baker, J. Math. Phys. **16**, 1324 (1975).

Cross between Born and WKB approximations: Variational solutions of nonlinear forms of the Schrödinger equation

A. R. P. Rau

Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803
(Received 14 October 1975)

Dashen, taking as his starting point the fact that the Born approximation for the phase shift has a wider range of applicability than would be expected from the usual criterion that the phase shift should be small, investigated a nonlinear form of the "one-dimensional" Schrödinger equation and arrived at an expression which can best be described as "a cross between the Born and WKB approximations." This expression for the phase shift reduces to the Born result even when the phase shift is not small. We investigate this result from a different point of view and establish that the Dashen expression is a variational principle for the nonlinear equation considered by him. This serves to explain the accuracy of the approximation. We also construct similar variational expressions for the more usual nonlinear Riccati equation and contrast it with the WKB series solution and recently proposed alternatives to this series. Contact is made between the techniques of functional analysis that Dashen used in arriving at his result and the unified formulation of variational principles that we adopt. An appendix deals with the principles of "invariant imbedding" in the particular context of such nonlinear versions of the one-dimensional Schrödinger equation.

I. INTRODUCTION

Some time ago, Dashen[1] took as his starting point the feature that the Born approximation for the phase shift,

$$\delta \simeq -k\int_0^\infty U(r)\,[rj_l(kr)]^2\,dr, \tag{1.1}$$

where U is the potential [more precisely, it is the potential multiplied by $(2m/\hbar^2)$] and $\frac{1}{2}k^2$ the energy (in atomic units), often seems to do very well even when there is no reason to expect this, that is, even when the phase shift is not very small. He investigated a nonlinear form of the radial Schrödinger equation (not quite the familiar Riccati equation) and through an integral equation he derived, he arrived at an expression for the phase shift which combined a judicious mixture of Born and WKB arguments applied over different ranges of r values. This "cross between the Born and WKB approximations," in the words of Ref. 1, reduces to the Born expression in (1.1) even without assuming that $\delta \ll 1$, so long as $|U/k^2| \ll 1$. The arrangement of this paper is as follows. Through a simple and straightforward general procedure that has been recently proposed,[2] we construct in Sec. II variational principles and identities for nonlinear equations. On applying these general results to the particular equation considered by Dashen, the variational identity coincides with his integral equation and the expression he derived for the phase shift on the basis of physical arguments is shown to be a variational principle. This provides some further insight into his approach and its success in numerical tests conducted by him. In Sec. III we turn to a similar application to the usual Riccati equation which is the starting point of the WKB series solution. Here again, alternatives to the WKB series have been proposed in the recent literature[3] and we contrast these results with the variational principle that follows from our procedure which again has features intermediate between Born and WKB solutions. A specific and simple numerical test is also presented. All combined, it is suggested that our simple variational procedure for these various forms of the one-dimensional Schrödinger equation (and for other more general nonlinear equations) is a powerful analytical and numerical tool, particularly when the usual Born and WKB approximations fail. Section IV comments on the relation between the procedure we follow and the methods of functional analysis (due to Bellman and Kalaba[4]) that Dashen used to arrive at his integral equation. An Appendix deals with the powerful principle of "invariant imbedding" which is known[5] to lead naturally to Riccati type equations. Specifically, it is pointed out that the alternative equation of Ref. 3 results naturally from the invariant imbedding approach and that the quantities involved have direct physical significance. The principle is also applied to show how the so-called phase-amplitude method for potential scattering[6] follows from the Born result in Eq. (1.1). Finally, interconnections between the phase and amplitude functions as adjoint functions in the variational formalism are established.

II. THE DASHEN NONLINEAR EQUATION

We review first the procedure and results of Dashen[1] and then show how variational principles and identities lead to the same conclusions. This is only a brief review of Ref. 1, just sufficient for our purposes and mainly to establish the notation; we strongly recommend the original reference to the reader. The radial Schrödinger equation,

$$u''(r) + q^2(r)\,u(r) = 0, \tag{2.1a}$$

$$q^2(r) \equiv k^2 - U(r) - l(l+1)/r^2, \tag{2.1b}$$

has admissible solutions obeying the conditions

$$u(r) \propto r^{l+1} \quad \text{as } r \to 0, \tag{2.2a}$$

$$u(r) \sim \sin(kr - \tfrac{1}{2}l\pi + \delta) \quad \text{as } r \to \infty. \tag{2.2b}$$

(Primes will represent throughout differentiation with respect to the argument.) Dashen defines the function

$$\alpha(r) \equiv \tan^{-1}[q(r)u(r)/u'(r)], \tag{2.3}$$

which satisfies the nonlinear equation

$$\alpha'(r) = q + (q'/2q)\sin 2\alpha, \quad \alpha \sim kr - \tfrac{1}{2}l\pi + \delta. \tag{2.4}$$

Similarly, the function

$$\alpha_0(r) \equiv \tan^{-1}\{q(r) r j_l(kr)/[r j_l(kr)]'\}, \tag{2.5a}$$

satisfies the equation

$$\alpha_0'(r) = q + (q'/2q)\sin 2\alpha_0 + (U/q)\sin^2\alpha_0,$$
$$\alpha_0 \sim kr - \tfrac{1}{2}l\pi. \tag{2.5b}$$

The physical significance of $\alpha(r_1)$ and $\alpha_0(r_1)$ at any arbitrary r_1 is that they are related to the phase shifts from potentials [in the former case, the angular momentum potential plus $U(r)$, whereas angular momentum alone in the latter] which are truncated at r_1. Defining

$$\beta \equiv k\alpha/q, \quad \beta_0 \equiv k\alpha_0/q, \tag{2.6}$$

one finds

$$\beta' = k - (q'/q)[\beta - (k/2q)\sin(2q\beta/k)], \tag{2.7}$$

$$\beta_0' = k - (q'/q)[\beta_0 - (k/2q)\sin(2q\beta_0/k)] + (kU/q^2)\sin^2(q\beta_0/k), \tag{2.8}$$

$$\beta(0) = \beta_0(0) = 0. \tag{2.9}$$

Dashen combines Eqs. (2.7) and (2.8) and uses mathematical techniques developed by Kalaba,[4] and earlier by Bellman,[7] to derive the integral equation

$$\beta(r) - \beta_0(r) = -\int_0^r dx\{[kU(x)/q^2(x)]\sin^2\alpha_0(x) + A(x)\} \times \exp\left(2\int_r^x dy\{[q'(y)/q(y)]\sin^2\alpha_0(y) + B(y)\}\right), \tag{2.10}$$

where

$$A \equiv (q'/q)[(\beta - \beta_0)\cos 2\alpha - (k/q)\sin(\alpha - \alpha_0)\cos(\alpha + \alpha_0)], \tag{2.11a}$$

$$B \equiv (q'/q)\sin(\alpha - \alpha_0)\sin(\alpha + \alpha_0). \tag{2.11b}$$

He then neglects A and B in Eq. (2.10), a step that is justified not only when $(\beta - \beta_0)$ is small but also when (q'/q) is small. This is the crucial step because these are, respectively, the Born and WKB restrictions. Near the origin, $|\beta - \beta_0|$ will be small, whereas further out q will often be slowly varying, so that (q'/q) is small even though $\beta - \beta_0$ may not be small in that region. Thus, neglect of A and B expresses a judicious application of Born and WKB conditions over different ranges of r in the radial evolution of the wavefunction, a point of view that these first-order, nonlinear formalisms such as the phase amplitude method particularly lend themselves to.

The expression obtained from (2.10) by dropping A and B and letting $r \to \infty$, namely,

$$\delta = \lim_{r\to\infty}[\beta(r) - \beta_0(r)] \simeq -\int_0^\infty dx (kU/q^2)\sin^2\alpha_0 \exp\left[2\int_\infty^x dy (q'/q)\sin^2\alpha_0\right], \tag{2.12}$$

or, alternatively, with the aid of (2.5a),

$$\delta \simeq -k\int_0^\infty dx \frac{U(x)}{q^2(x) + [(xj_l(kx))'/xj(kx)]^2} \times \exp\int_\infty^x \left(dy \frac{q^{2\prime}(y)}{q^2(y) + [(yj_l(ky))'/yj_l(ky)]^2}\right), \tag{2.13}$$

is a cross between the Born and WKB methods. Further, without assuming $\delta \ll 1$, Eq. (2.13) reduces to the Born result in Eq. (1.1), provided only

$$|U/k^2| \ll 1. \tag{2.14}$$

Thus, for long range potentials, when δ may be large but inequality (2.14) is valid, the Born approximation is still applicable as Dashen explicitly demonstrates through a numerical example. What is happening, of course, is that there is a slow accumulation of phase shift from large r values resulting in a final δ that is large but, throughout this range, the wavelength is varying slowly so that the step of considering (q'/q) as small is completely justified. Even when the Born approximation is invalid, Eq. (2.13) has a wider range of validity and provides a convenient method for computing δ. We refer the reader to the original paper of Dashen for demonstrations of this.

Our aim now is to establish the basic equations, (2.10) and (2.12), and view the passage from the former to the latter somewhat differently. We will see that Eq. (2.10) is a "variational identity" and Eq. (2.12) the associated variational principle so that the terms that are dropped in the passage, A and B, must be second-order terms. This provides an alternative justification for the Dashen procedure. Our method, though equivalent in the context of such nonlinear equations to the methods of functional analysis used by Dashen, has the advantage of simplicity and of fitting into a common framework[2] that applies to a very wide range of variational principles and identities for properties of bound states, scattering theory, homogeneous and inhomogeneous equations, linear and nonlinear equations, integral, differential and integro-differential equations, etc. The derivation proceeds as follows. Consider the equation

$$Z'(r) = f(Z; r), \quad Z(r_0) \text{ known}, \tag{2.15}$$

where f is a nonlinear function of Z, and r_0 is some specific value at which the value of Z is specified. Following the procedure of Ref. 2, a variational estimate of Z at some arbitrary value of r is written as

$$Z_v(r) = Z_t(r) - \int_{r_0}^r dx\, L_t(r,x)\{Z_t'(x) - f(Z_t; x)\}, \tag{2.16}$$

where Z_t is a trial estimate of Z and L_t is a trial approximation to an exact "Lagrange multiplier" which will be so defined that all first-order terms in

$$\delta Z \equiv Z_t - Z \quad \text{and} \quad \delta L \equiv L_t - L \tag{2.17}$$

on the right-hand side of Eq. (2.16) explicitly vanish. Substitution of Eq. (2.17) in (2.16) shows that we must have

$$\delta Z(r) - \int_{r_0}^r dx\, L(r,x)\left(\delta Z'(x) - \frac{\delta f}{\partial Z}\delta Z\right) = 0, \tag{2.18}$$

where $\delta f/\delta Z$ is the formal derivative of $f(Z;x)$ with respect to Z. From Eq. (2.18), the defining equations for L are extracted by simple manipulations. Transferring the derivative with respect to x from δZ on to L generates surface terms which can be made to vanish first of all by choosing $Z_t(r)$ such that $Z_t(r_0)$ is equal to the known value of $Z(r_0)$ and by setting a boundary condition on $L(r,x)$ at the upper limit $x = r$. We finally have

$$\frac{dL(r,x)}{dx} = -L(r,x)\frac{\delta f}{\delta Z}, \quad L(r,r)=1. \tag{2.19}$$

As in all applications of this method, the equation for L is necessarily linear and, as illustrated here, in many examples involving nonlinear equations for Z, the equation for the "adjoint" Lagrange function admits a closed-form solution. We have

$$L(r,x) = \exp\left(-\int_r^x dy \frac{\delta f}{\delta Z}(Z;x)\right). \tag{2.20}$$

A trail solution L_t of this (an obvious choice is to replace Z by Z_t in the argument of the exponential) when inserted into Eq. (2.16) provides the desired variational principle. Again, as in general,[2] rearrangement of Eq. (2.18) provides the corresponding variational identity which takes the form

$$Z(r) = Z_t(r) - \int_{r_0}^r dx\, L(r,x)\{Z_t'(x) - (Z_t - Z)\frac{\delta f}{\delta Z}(Z;x) - f(Z;x)\}. \tag{2.21}$$

The variational principle Eq. (2.16) is connected to its associated identity Eq. (2.21) in the obvious way—through neglect of terms of second order on the right-hand side of Eq. (2.21). We note for future use the feature that the identity is linear in Z_t which is a feature characteristic of the general method.

This completes the general derivation according to our method; its contact with the methods of functional analysis will be taken up in Sec. IV. For now, we apply the above considerations to the Dashen equation, Eq. (2.7). From Eqs. (2.20) and (2.7), the adjoint function is

$$L(r,x) = \exp[2\int_r^x dy (q'/q)\sin^2\alpha]. \tag{2.22}$$

As trial functions in the variational principle if we choose $\beta_t = \beta_0$ and L_t as given by Eq. (2.22) with the corresponding $\alpha_0 = q\beta_0/k$ in place of α, we have from Eq. (2.16), ($r_0 = 0$ now)

$$\beta_v(r) = \beta_0(r) - \int_0^r dx (kU/q^2)\sin^2\alpha_0(x) \times \exp[2\int_r^x dy (q'/q)\sin^2\alpha_0(y)], \tag{2.23}$$

and from Eq. (2.21) the identity

$$\beta(r) = \beta_0(r) - \int_0^r dx[(kU/q^2)\sin^2\alpha_0 + A] \times \exp\{2\int_r^x dy(q'/q)\sin^2\alpha\}, \tag{2.24}$$

with A as defined before in Eq. (2.11a). Together with the easily established relation,

$$(q'/q)\sin^2\alpha = (q'/q)\sin^2\alpha_0 + B,$$

Eq. (2.24) coincides with Dashen's integral equation in Eq. (2.10). The expression he uses for calculating phase shifts by neglecting A and B is precisely the variational principle in Eq. (2.23). A and B are, therefore, second-order terms in the sense of the variational principle which justifies their neglect. The rapid convergence in the numerical examples considered by Dashen can be understood in these same terms as the expression of a variational principle in action. We remark on the characteristic structure of the variational expression in Eq. (2.23) which involves the potential both directly in the integrand and again under an integral in exponential. Since the former feature is characteristic of Born-type expressions for the phase shift [see, for instance, Eq. (1.1)], whereas a well-known feature of WKB expressions for the phase shift is an exponential involving an integral over the potential, we see again that Eq. (2.23) has the appearance of a cross between Born and WKB approximations; the structure is also reminiscent of the Glauber approximation.

III. THE USUAL RICCATI EQUATION

In this section, we consider the familiar Riccati form of the one-dimensional Schrödinger equation and examine a variational solution for it that follows from Eq. (2.16), contrasting this with other methods of solution. It is well-known that the nonlinear transformation, $u = \exp(\pm iS_\pm)$ transforms the Schrödinger equation,

$$\ddot{u}(r) + q^2(r)u(r) = 0, \tag{3.1}$$

to the first-order, nonlinear equation for $Z_\pm \equiv S_\pm'$,

$$\pm iZ_\pm' + q^2 - Z_\pm^2 = 0. \tag{3.2}$$

We will concentrate on Z_+ which we will call Z. This familiar Riccati equation is, of course, the starting point for generating the WKB series:

$$\begin{aligned} Z^{(0)}(r) &= q(r), \\ Z^{(1)}(r) &= Z^{(0)}(r) + iq'/2q, \\ Z^{(2)}(r) &= Z^{(1)}(r) + (3q'^2/8q^3) - (q''/4q^2), \\ &\cdots . \end{aligned} \tag{3.3}$$

An alternative to Eq. (3.2) has been suggested recently[3] which uses a new variable ϕ, related to Z according to

$$Z \equiv q(1-\phi)(1+\phi)^{-1}, \tag{3.4}$$

and obeying an alternative nonlinear equation,

$$2q\phi' + q'(\phi^2 - 1) + 4iq^2\phi = 0, \tag{3.5}$$

as a better starting point. Equation (3.5) admits a similar series solution

$$\begin{aligned} \phi^{(0)} &= 0, \\ \phi^{(1)} &= -iq'/4q^2, \\ \phi^{(2)} &= \phi^{(1)} - (q'^2/4q^4) - (iq'^3/64q^6) + (q''/8q^3), \\ &\cdots . \end{aligned} \tag{3.6}$$

It is easily verified that substitution of Eq. (3.6) in (3.4) yields, to the same level of approximation, the WKB series in (3.3). In certain numerical applications, Eq. (3.6) has advantages over the series in (3.3).[3]

If we were to seek a variational principle and identity for Z, we need only apply the results in Eqs. (2.16)–(2.21) with the specific form of f in Eq. (3.2), namely, $f = i(q^2 - Z^2)$. We have from Eq. (2.20)

$$L(r,x) = \exp[2i\int_r^x dy\, Z(y)]. \tag{3.7}$$

From Eq. (2.16), with L_t chosen again to be given as in Eq. (3.7) with Z_t in the integrand, the variational principle is

$$Z_v(r)=Z_t(r)-\int_{r_0}^{r}dx[Z_t'(x)-iq^2(x)+iZ_t^2(x)]\times\exp[2i\int_r^x dy\,Z_t(y)], \quad (3.8)$$

and from Eq. (2.21), the identity takes the form

$$Z(r)=Z_t(r)-\int_{r_0}^{r}dx[Z_t'(x)-iq^2(x)-iZ^2(x)+2iZ(x)Z_t(x)]\times\exp[2i\int_r^x dy\,Z(y)]. \quad (3.9)$$

Once again, it is explicitly clear that passage from Eq. (3.9) to Eq. (3.8) involves the neglect of second-order terms. We note the similarity in the structure of the variational expression in Eq. (3.8) to the previous one in (2.23).

The choice of Z_t is unrestricted, subject only to the boundary condition that $Z_t(r_0)$ be equal to the known value, $Z(r_0)$. A natural choice and one that affords ready comparison with the results of Eqs. (3.3) and (3.6) is $Z_t(r)=q(r)$. We have, as a result, from Eq. (3.8),

$$Z_v(r)=q(r)-\int_{r_0}^{r}dx\,q'(x)\exp[2i\int_r^x q(y)\,dy]. \quad (3.10)$$

It is of interest to view this expression in somewhat different ways. Carrying out an integration by parts according to a particular scheme, we find

$$Z_v(r)=q(r)-\int_{r_0}^{r}dx[q'(x)/2iq(x)]\left\{2iq(x)\exp\left[2i\int_r^x q(y)\,dy\right]\right\}$$
$$=q(r)+\frac{iq'(r)}{2q(r)}-\frac{iq'(r_0)}{2q(r_0)}\exp\left[\int_r^{r_0}2iq(y)\,dy\right]$$
$$-\int_{r_0}^{r}dx\left(\frac{iq'}{2q}\right)'\exp\left(2i\int_r^x q(y)\,dy\right). \quad (3.11)$$

The first two terms agree with the first two of the WKB result in (3.3) and if, according to the usual WKB criterion, successive derivatives are small, we see that Eq. (3.10) contains in it the WKB series as one would expect. On the other hand, when the WKB criterion does not hold so that $q(r)$ is a poor starting choice we see that the second two terms in Eq. (3.11) are important or, alternatively, that the variational expression can correct for the poor choice as follows. Integrating Eq. (3.10) by parts as it stands we find

$$Z_v(r)=q(r)-q(r)+q(r_0)\exp[2i\int_r^{r_0}q(y)\,dy]+2i\int_{r_0}^{r}dx\,q^2(x)\exp[2i\int_r^x q(y)\,dy]. \quad (3.12)$$

Equations (3.11) and (3.12) are, of course, entirely equivalent and are only alternative forms of Eq. (3.10) arranged to bring out explicitly the structure of the variational expression for cases where WKB may be a good approximation and for those where it is not. In the latter case, as we see from Eq. (3.12) the variational result can correct for the poor starting choice and as such it has advantages over Eqs. (3.5) or (3.6) which are, after all, just an alternative (even though more convenient) way of getting the WKB series.

It is also of interest to examine the identity in Eq. (3.9) with the specific choice $Z_t=q$. We have

$$Z(r)=q(r)-\int_{r_0}^{r}dx[q'(x)-i(q-Z)^2]\exp[2i\int_r^x q(y)\,dy]. \quad (3.13)$$

Though the exact Z also appears on the right-hand side of the identity, the identity is useful because it helps to display the second-order errors left out by the variational principle. In particular, the appearance of a quadratic term in $(q-Z)^2$, which will usually be of a well-defined sign, suggests that in many problems the second-order errors may be of a definite sign [we note that it is $iZ(r)$ that will be of interest] so that the variational expression may in fact be an extremal or a variational bound, which is a more powerful result. We illustrate this with a simple numerical example. The choice $q^2(r)=-2/r^2$ permits an exact solution of Eq. (3.2) so that $Z(r)=i/r$. On the other hand, the starting point of the series solutions in Eqs. (3.3) and (3.6), is $q(r)=\sqrt{2}\,i/r$. Taking the boundary condition to be $Z(\infty)=0$, we apply the variational principle in Eq. (3.8) with $Z_t(z)=q(r)$ as the initial trial choice and iterate using successive results of Eq. (3.8) as the trial solutions for the next step of interaction. The results are contrasted in Table I with the solutions given by Eqs. (3.3) and (3.6). We note the faster, and quadratic convergence of the variational results. The oscillations in the WKB series are larger than the results for the series in Eq. (3.6). The variational results on the other hand show monotonic convergence indicating that we have a variational upper bound. This is as expected from inspection of second-order terms in Eq. (3.13) for this particular example.

IV. RELATION TO THE METHOD OF KALABA AND BELLMAN

Dashen arrived at his integral equation, quoted here in Eq. (2.10), by using a mathematical procedure developed by Bellman[7] and Kalaba.[4] We have shown in Sec. II that, using the very general procedure of Ref. 2, the same integral equation can be established as being the variational identity associated with a variational principle. This connection between the Kalaba–Bellman method and that of our Ref. 2 seems to hold true quite generally as we now discuss, even though their starting point is very different from ours. In fact, they start by observing[4] that the connection between the theory of linear differential equations and variational problems is well known and is often exploited to establish the existence of solutions of such equations by the methods of the calculus of variations making use of principles such as the Rayleigh–Ritz principle. On the other hand, their approach to nonlinear equations is, they say, entirely different and proceeds by establishing representation theorems for solutions of various nonlinear equations in terms of "the maximum operation on solutions of a set of associated linear equations" that they define. For an equation such as Eq. (2.15) they begin by observing that for $f(Z;r)$, a strictly convex function in Z, that is, $\delta^2 f/\delta Z^2>0$, we have

$$f(Z)=\max_v\left[f(v)+(Z-v)\frac{\delta f}{\delta Z}(v)\right]. \quad (4.1)$$

By dropping the maximum operation and writing

$$\omega'=f(v)+(\omega-v)\frac{\delta f}{\delta Z}(v), \quad (4.2)$$

where ω and v are any two functions in the same space

TABLE I. Comparison of various approximation methods applied to the solution of $iZ'(r) - (2/r^2) - Z^2(r) = 0$. The exact solution is $rZ/i = 1$. We contrast the convergence towards this result of the variational results of this paper with other approximation methods, taking the same initial starting point of $rZ_t/i = \sqrt{2}$ for each method.

WKB	Results from Eq. (3.6)	Variational principle
1.4142136	1.4142136	1.4142136
0.9142136	0.9893242	1.0448156
1.0026019	1.0006605	1.0006500
	0.9999600	1.0000001

of functions as Z, and $\omega(r_0) = Z(r_0)$, they have in Eq. (4.2) an associated linear equation with the original nonlinear one in Eq. (2.15). A solution of the original equation is expressed in terms of ω by

$$Z(r) = \max_v \omega(v; r). \tag{4.3}$$

The maximization is over all v and is, in fact, attained when $v = Z$ when one also has $\omega = Z$. Kalaba and Bellman also show how Eq. (4.2) can be used to set up monotone sequences that converge quadratically to the correct solution which makes the method analogous to Newton's method for finding roots of algebraic equations.

We can now readily see the contact between this method and ours in Eqs. (2.15)–(2.21). Doing an integration by parts in Eq. (2.16) and using Eq. (2.19), we have

$$Z_v(r) = Z(r_0) L_t(r, r_0) + \int_{r_0}^{r} dx\, L_t(r, x) \left[f(Z_t; x) - Z_t \frac{\delta f}{\delta Z}(Z_t; x) \right], \tag{4.4}$$

where L_t is defined as in Eq. (2.20) with Z_t replacing Z in the argument of the exponential. We can now observe that Eq. (4.4) is the solution of Eq. (4.1) with ω identified with Z_v and v with Z_t. Hence the Kalaba–Bellman method as expressed by Eqs. (4.1) and (4.2) is equivalent to the variational principle in Eq. (2.16) and, of course, as discussed there Z_v becomes Z when $Z_t = Z$—this was in fact the starting point in writing down Eq. (2.16). The next question is what the associated linear equation in Eq. (4.2) corresponds to in the variational formalism. We observe that the variational identity in Eq. (2.21) has inside the curly brackets an expression linear in Z_t. Setting this expression equal to zero means that the solution thus obtained for Z_t coincides with Z as is immediate from Eq. (2.21). We now note that this linear equation for Z_t is precisely the Kalaba–Bellman one in Eq. (4.2) with ω identified with Z_t and v with Z. Seen from this angle, it is again an immediate conclusion that since $v = Z$, ω will also coincide with the exact Z. Hence the associated linear equation of Kalaba–Bellman is a statement of the variational identity in Eq. (2.21) and is the expression that multiplies L_t in such an identity. That this expression is always linear in Z_t regardless of whether the original equation for Z is linear or nonlinear is an aspect of the general variational procedure of Ref. 2 that we have emphasized before.[2]

Having seen the complete equivalence of the two methods for handling nonlinear equations, a few comments contrasting the two are of interest. The general variational formulation applies equally to many problems in mathematical physics—to equations that may be linear or nonlinear, differential or integral, etc. The variational principle is constructed (and the construction is very straightforward and simple for any problem of interest) using the original equations as well as a set of adjoint equations for Lagrange multipliers which are *always linear* equations. Further, with each variational principle one has a corresponding identity which contains products of trial Lagrange multipliers with expressions that are *linear* in trial estimates of the original unknown functions. It is these latter linear expressions that arise as the associated linear equations of the Kalaba–Bellman method for solving nonlinear problems. The Kalaba–Bellman method is restricted to functions $f(Z; r)$ which are either strictly convex or concave in which case it gives upper and lower bounds to the true solutions. The variational procedure applies to arbitrary functions f and gives a stationary principle even when f is neither convex nor concave. The variational procedure also has the merit that it views nonlinear problems on the same footing as linear problems and thus on par with familiar principles such as the Rayleigh–Ritz principle.[8] Kalaba's remarks about their method being analogous to Newton's method also fits readily into the framework of our general variational formulation because we regard Newton's method as one of the simplest illustrations of our formulation.[2]

APPENDIX: INVARIANT IMBEDDING AND EQUATIONS FOR THE PHASE SHIFT

The powerful mathematical principle of invariant imbedding, which views a problem as one of a family of similar problems, is known[5] to lead quite naturally to nonlinear initial-value type problems instead of the equivalent linear eigenvalue problems that result from conventional analysis. In fact, Dashen was motivated by the principle of invariant imbedding in writing down Eq. (2.4). In this Appendix we emphasize the particularly close relationship of invariant imbedding to a host of different nonlinear formulations for the phase shift. The authors of Ref. 3, in writing down ϕ in Eq. (3.5) and the series in Eq. (3.6) as an alternative to the usual WKB series, leave open the question of why ϕ fares better than the original Z. This can be understood by looking at the physical significance of ϕ and the origins of the nonlinear equation satisfied by it. It is shown in Ref. 5 that if one views the problem of a wave of unit

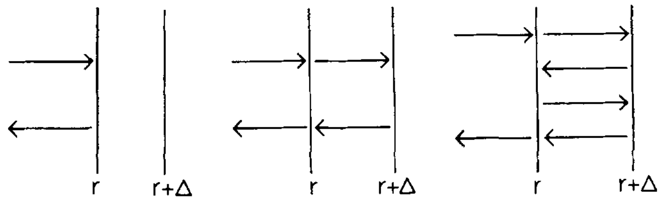


FIG. 1. The three contributions to the reflected wave at r up to first order in Δ.

amplitude incident from the left on a slab (r, b) (the left boundary of the slab is at r and the right boundary at b) with varying wave vector in the slab, and asks for the coefficient of reflection, this problem can be imbedded in a family of such problems by imagining a plane at $r+\Delta$. The reflected wave at r can now be built up by viewing Δ as a small quantity so that the plane at r separates regions of wave vector $q(r)$ and $q(r+\Delta)$. Reflection and transmission of this elementary problem are trivially written down and to first order in Δ there are three contributions of interest as shown in Fig. 1: the immediately reflected wave at r, the wave resulting from transmission through r followed by reflection at $r+\Delta$ and subsequent transmission through r and, finally, the case when there are four "traversals" of the region between r and $r+\Delta$. In this manner, on taking the limit $\Delta \to 0$, a nonlinear equation results (see Sec. 17 of Ref. 5) for the reflection coefficient which is precisely Eq. (3.5) for ϕ given in Ref. 3. Thus ϕ has the physical significance of being this reflection coefficient and the special status of the particular nonlinear equation in Eq. (3.5) can be appreciated since it is the form one is naturally led to through invariant imbedding.

As yet another illustration of the power of this principle as applied to phase shifts for potential scattering, we will now demonstrate that the phase equation of the so-called phase-amplitude method[6] can be derived through this method, starting simply from the Born expression in Eq. (1.1). The phase amplitude method starts by writing $u(r)=A(r)\exp[i\delta(r)]$ in terms of an amplitude and a phase function. The phase function or more accurately, $t(r)=\tan\delta(r)$ satisfies a well-known first-order nonlinear differential equation. $\delta(\infty)$ is what is usually called the phase shift but $\delta(r)$ has the nice feature of being physically significant even at any arbitrary r, being the phase shift that would result from a potential truncated at that point and set equal to zero beyond it [see similar remarks with regard to $\alpha(r)$ in the discussion following Eq. (2.5)]. We will now derive the equation for $t(r)$ by considering the problem of scattering from a given potential $U(r)$ as imbedded in the problem of scattering by a "slightly extended" potential $U(r+\Delta)$. As regards the infinitesimally small potential between r and $r+\Delta$, the Born approximation is certainly applicable and we know that in this approximation, it is given by

$$-k\int_r^{r+\Delta} U(r)\,u^2(r)\,dr, \qquad (A1)$$

where $u(r)$ is the "unperturbed wave" for this part of the problem, namely, the wave resulting from the potential $U(r)$. The usual expression, as in Eq. (1.1), for the Born phase shift (or tangent of the phase shift) regards the unperturbed wave to be the free wave in the absence of the potential and thus uses $rj_l(kr)$ in place of $u(r)$. By the principle of invariant imbedding we can now write $t(r+\Delta)$ as the part due to the potential up to r alone which is $t(r)$ plus the piece in Eq. (A1). We have

$$t(r+\Delta)=t(r)-k\int_r^{r+\Delta} U(r)\,u^2(r)\,dr. \qquad (A2)$$

With $u(r)$ given by the usual definition as $rj_l(kr)+t(r)\,rn_l(kr)$, passage to the limit $\Delta\to 0$ in Eq. (A2) gives

$$t'(r)=-kU(r)\,[rj_l(kr)+t(r)\,rn_l(kr)]^2, \qquad (A3)$$

which is the equation in the phase-amplitude method as derived by other procedures.[6,9] We note the rather remarkable feature that through the principle of invariant imbedding we have succeeded in deriving an exact equation for the phase shift [no approximation is involved in Eq. (A3)] using only an approximate relationship, namely, the Born expression for the phase shift due to an infinitesimally weak potential. This illustrates the power of this technique.

This Appendix seems an appropriate place to record also variational principles for the phase-amplitude functions which follow from the general formulation in Sec. 2. We can start with the basic equations for the phase function $\delta(r)$ and the amplitude function $A(r)$ in a more general form than, for instance, given in Eq. (A3) where specific "reference functions," $rj_l(kr)$ and $rn_l(kr)$, were used. With more general reference functions, a regular solution $f(r)$ and an irregular solution $g(r)$, the equations take the form[10]

$$\delta(r)=W^{-1}\int_0^r ds\; U(s)\,[f(s)\cos\delta(s)-g(s)\sin\delta(s)]^2, \qquad (A4)$$

$$\ln[A(r)/A(\infty)]=W^{-1}\int_\infty^r ds\; U(s)\,[f(s)\cos\delta(s)-g(s)\sin\delta(s)] \times[f(s)\sin\delta(s)+g(s)\cos\delta(s)], \qquad (A5)$$

where W is the Wronskian of f and g and $\delta(0)\equiv 0$. Writing Eq. (A4) is differential form,

$$\frac{d\delta(r)}{dr}=W^{-1}\,U(r)\,[f(r)\cos\delta(r)-g(r)\sin\delta(r)]^2, \qquad (A6)$$

so that it has the general structure given in Eq. (2.15), a variational principle of the form given in Eq. (2.16) immediately follows if we seek the value of, say, $\delta(\infty)$. We have

$$\delta_v(\infty)=\delta_t(\infty)-\int_0^\infty dr\,L_t(r)\Big(\frac{d}{dr}\,\delta_t(r) - W^{-1}U(r)\,[f(r)\cos\delta_t(r)-g(r)\sin\delta_t(r)]\Big)^2,$$

and $L(r)$ is seen, from Eq. (2.19), to obey

$$\frac{dL(r)}{dr}=2L(r)\,W^{-1}U(r)\,[f(r)\cos\delta(r)-g(r)\sin\delta(r)] \times[f(r)\sin\delta(r)+g(r)\cos\delta(r)],$$
$$L(\infty)=1. \qquad (A7)$$

The solution of $L(r)$ can be written down immediately and, on comparison with Eq. (A5), we can make the identification

$$L(r)=A^2(r)/A^2(\infty). \qquad (A8)$$

The square of the amplitude function and the phase function are, therefore, adjoint functions in the variational formalism. As a corollary to this, we note from the very structure of Eqs. (A4) and (A5) that, whereas $\delta(r)$ evolves from its zero value at $r=0$ to the physically significant value $\delta(\infty)$ at $r=\infty$, the amplitude function or, alternatively, $A^2(r)$ evolves from its reference value at $r=\infty$ [usually taken to be 1 to represent a wave of unit amplitude incident on the potential] to, finally, the physically significant "density of the wavefunction at the origin." It is also of interest to note that were we to write a variational principle for $A(0)$ given the defining equation (A5) or its equivalent first-order linear differential equation, we would find that the adjoint function in this case is $A(0)/A(r)$. For completeness, we note that other derivations of variational principles for the phase amplitude method are given by Calogero in Ref. 6. The merit of our procedure of Ref. 2 is that it is very simple. We also note that phase-amplitude equations for multichannel scattering are available[6] and similar variational principles for their solution can be readily constructed.

ACKNOWLEDGMENT

I thank the Raman Research Institute and the Centre for Theoretical Studies, Bangalore, India, for their hospitality while this work is being completed.

[1]R. Dashen, Nuovo Cimento 28, 229 (1963).

[2]E. Gerjuoy, A.R.P. Rau, and L. Spruch, Phys. Rev. A 8, 662 (1973); J. Math. Phys. 13, 1797 (1972); 16, 1104 (1975), and a review article to be published.

[3]W.I. Newman and W.R. Thorson, Phys. Rev. Lett. 29, 1350 (1972); Can. J. Phys. 50, 2997 (1972); W.R. Thorson, Can. J. Phys. 52, 2509 (1974).

[4]R. Kalaba, J. Math. Mech. 8, 519 (1959) and references therein.

[5]R. Bellman and R. Kalaba, J. Math. Mech. 8, 683 (1959).

[6]See, for instance, F. Calogero, *The Variable Phase Approach to Potential Scattering* (Academic, New York, 1967); V.V. Babikov, Usp. Fiz. Nauk. 92, 3 (1967) [Sov. Phys. Usp. 10, 271 (1967)]; J.L. Dehmer and U. Fano, Phys. Rev. A 2, 304 (1970).

[7]R. Bellman, Proc. Natl. Acad. Sci. 41, 743 (1955).

[8]For other applications of variational principles to nonlinear equations see, for instance, L. Rosenberg and S. Tolchin, Phys. Rev. A 8, 1702 (1973); E. Gerjuoy, A.R.P. Rau, and L. Spruch, in Proceedings of the Third International Conference on Atomic Physics, Boulder, 1972, and A.R.P. Rau and A.K. Rajagopal, Phys. Rev. B 4, 3604 (1975), and references therein.

[9]L. Spruch, in *Lectures in Theoretical Physics, Boulder*, 1961, edited by W.E. Brittin, W.B. Downs, and J. Downs (Wiley, New York, 1962), Vol. 4, p. 205.

[10]See, for instance, Sec. 4 of U. Fano, C.E. Theodosiou, and J.L. Dehmer, Rev. Mod. Phys. 48, 49 (1976).

A gauge invariant formulation of time-dependent dynamical symmetry mappings and associated constants of motion for Lagrangian particle mechanics. I

Gerald H. Katzin* and Jack Levine

Department of Physics and Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607*

(Received 11 February 1976)

In this paper (part I of two parts), which is restricted to classical particle systems, a study is made of time-dependent symmetry mappings of Lagrange's equations (a) $\Lambda_i(L)=0$, and the constants of motion associated with these mappings. All dynamical symmetry mappings we consider are based upon infinitesimal point transformations of the form (b) $\bar{x}^i = x^i+\delta x^i$ $[\delta x^i \equiv \xi^i(x,t)\delta a]$ with associated changes in trajectory parameter t defined by (c) $\bar{t}=t+\delta t$ $[\delta t \equiv \xi^0(x,t)\delta a]$. The condition (d) $\delta\Lambda_i(L)=0$ for a symmetry mapping may be represented in the equivalent form (e) $\Lambda_i(N)=0$, where (f) $N\delta a \equiv \delta L + L d(\delta t)/dt$. We consider two subcases of these symmetry mappings which are referred to as R_1, R_2 respectively. Associated with R_1 mappings [which are satisfied by a large class of Lagrangians including all $L=L(\dot{x},x)$] is a time-dependent constant of motion (g) $C_1 \equiv (\partial N/\partial \dot{x}^i)\dot{x}^i - N + (\partial/\partial t)[(\partial L/\partial \dot{x}^i)\xi^i - E\xi^0] + \gamma_1(x,t)$, where γ_1 is determined by R_1. The R_2 subcase is the familiar Noether symmetry condition and hence has associated with it the well-known Noether constant of motion which we refer to as C_2. For symmetry mappings which satisfy both R_1 and R_2 it is shown that (h) $C_1=\partial C_2/\partial t+\gamma_1$. The various forms of symmetry equations and constants of motion considered are shown to be invariant under the Lagrangian gauge transformation (i) $L \rightarrow L' = L + d\psi(x,t)/dt$.

1. INTRODUCTION

In a previous paper[1] we considered time-independent dynamical symmetries and associated time-independent constants of motion for classical particle systems. These dynamical symmetries were based upon infinitesimal point mappings $\bar{x}^i = x^i + \delta x^i$ $[\delta x^i \equiv \xi^i(x)\delta a]$, with associated change in the independent variable defined by $\bar{t} = t + \delta t$ $(\delta t = \{\int 2\phi[x(t)]dt + c\}\delta a)$. These symmetries were formulated directly at the level of the dynamical equations which were taken in the form of Newton's equations, Lagranges's equations, Hamilton's equations, and the Hamilton—Jacobi equation. In a subsequent paper[2] we treated the time-dependent theory of symmetries and constants of motion for Hamilton's equations [based upon the phase-space version of mappings (1.1) and (1.2)].

The purpose of this paper is to determine the time-dependent symmetry mappings of Lagrange's equations and determine associated time-dependent constants of motion. Included in this investigation is a comparison with the well-known Noether approach to symmetries and constants of motion. This work which is in two parts is an extension and generalization of the time-independent theory of Ref. 1. Illustrations and applications will be included in Paper II.

We define two types of time-dependent dynamical symmetries based upon infinitesimal point mappings which map the set of trajectories of a dynamical system into itself (with associated change in trajectory parameter[3] t). These two types of mappings are[4]

Type I:

$$\bar{x}^i = x^i + \delta x^i, \quad \delta x^i \equiv \xi^i(x,t)\delta a, \quad (\delta a \equiv \text{infinitesimal}), \qquad (1.1)$$

$$\bar{t} = t + \delta t, \quad \delta t \equiv \xi^0(x,t)\delta a; \qquad (1.2)$$

Type II:

$$\bar{x}^i = x^i + \delta x^i, \quad \delta x^i \equiv \xi^i(x,t)\delta a, \qquad (1.1)$$

$$\bar{t} = t + \delta t, \quad \delta t \equiv \{\int 2\phi[x(t),t]dt + c\}\delta a. \qquad (1.3)$$

The notation $\phi[x(t),t]$ indicates the function $\phi(x,t)$ is to be evaluated along a trajectory.

It should be noted that the infinitesimal point mappings considered in Ref. 1 are a special case of the Type II mappings. In this paper we consider mappings of Type I only and defer a discussion of Type II mappings to a later paper.

In Sec. 3 based upon infinitesimal Type I mappings (1.1) and (1.2) we formulate directly at the level of Lagrange's equations two equivalent equations [(3.3), (3.10)] for the existence of Type I dynamical symmetry mappings. It is shown that every Type I Noether symmetry mapping is a Type I symmetry mapping (but not conversely).

In Sec. 4 we consider three restricted Type I symmetries, these restrictions being defined respectively by (4.5) [Type R_1], (4.7) [Type R_2, Noether symmetry condition], and (4.5) and (4.7) [both R_1 and R_2]. Besides the familiar Noether constant of motion C_2 [given by (4.8)] which is associated with the Type R_2 restriction, it is shown there exists a constant of motion C_1, in general distinct from C_2, [given by (4.6)] concomitant with the Type R_1 restriction. Corresponding to the combined R_1, R_2 restriction there exists a concomitant constant of motion C_{12} [given by (4.10)] which is a reduced form of C_1 which is essentially the partial time derivative of the Noether constant of motion C_2. The restriction R_1 is always satisfied for dynamical systems characterized by Lagrangians with no explicit time dependence. For such systems the constant of motion C_1 is expressible in the form $C_1 = \delta E$, where

δE is a Type I symmetry deformation of the well-known constant of motion $E \equiv (\partial L/\partial \dot{x}^i)\dot{x}^i - L$.

A time-dependent related integral theorem is stated for dynamical systems which admit Type I symmetry mappings. The above-mentioned relation $C_1 = \delta E$ is an illustration of this theorem.

In Sec. 5 it is shown that the equivalent equations (3.3) and (3.10) for the existence of dynamical symmetry mappings, the restrictions R_1, R_2, and the three constants of motion C_1, C_2, C_{12} are all gauge invariant under the Lagrangian gauge transformation (5.1).

2. BASIC FORMULAS

In this section we list several basic formulas which will be used in the analysis of Type I mappings.

Based upon the infinitesimal point mapping (1.1) and (1.2) we first define the δ-variations[5]

$$\delta \dot{x}^i \equiv \frac{d\bar{x}^i}{d\bar{t}} - \frac{dx^i}{dt} = (\dot{\xi}^i - \dot{x}^i\dot{\xi}^0)\delta a, \tag{2.1}$$

$$\delta \ddot{x}^i \equiv \frac{d^2\bar{x}^i}{d\bar{t}^2} - \frac{d^2x^i}{dt^2} = (\ddot{\xi}^i - \dot{x}^i\ddot{\xi}^0 - 2\ddot{x}^i\dot{\xi}^0)\delta a, \tag{2.2}$$

$$\delta G(\ddot{x}, \dot{x}, x, t) \equiv \frac{\partial G}{\partial \ddot{x}^i}\delta \ddot{x}^i + \frac{\partial G}{\partial \dot{x}^i}\delta \dot{x}^i + \frac{\partial G}{\partial x^i}\delta x^i + \frac{\partial G}{\partial t}\delta t, \tag{2.3}$$

where

$$G(\ddot{x}, \dot{x}, x, t) \equiv G(\ddot{x}^1, \ldots, \ddot{x}^n; \dot{x}^1, \ldots, \dot{x}^n; x^2, \ldots, x^n; t).$$

For any function $F \equiv F(\dot{x}, x, t)$ the total time derivative is defined by

$$\frac{dF}{dt} \equiv \frac{\partial F}{\partial \dot{x}^i}\ddot{x}^i + \frac{\partial F}{\partial x^i}\dot{x}^i + \frac{\partial F}{\partial t}. \tag{2.4}$$

It can be shown by use of the above definitions that

$$\delta\left(\frac{dF}{dt}\right) \equiv \frac{d(\delta F)}{dt} - \frac{dF}{dt}\frac{d(\delta t)}{dt}, \tag{2.5}$$

$$\delta\left(\frac{\partial F}{\partial x^i}\right) - \frac{\partial(\delta F)}{\partial x^i} \equiv -\frac{\partial F}{\partial x^j}\frac{\partial(\delta x^j)}{\partial x^i} - \frac{\partial F}{\partial \dot{x}^j}\frac{\partial(\delta \dot{x}^j)}{\partial x^i} - \frac{\partial F}{\partial t}\frac{\partial(\delta t)}{\partial x^i}, \tag{2.6}$$

$$\delta\left(\frac{\partial F}{\partial \dot{x}^i}\right) - \frac{\partial(\delta F)}{\partial \dot{x}^i} \equiv -\frac{\partial F}{\partial \dot{x}^j}\frac{\partial(\delta \dot{x}^j)}{\partial \dot{x}^i}, \tag{2.7}$$

$$\frac{\partial(\delta \dot{x}^j)}{\partial \dot{x}^i} \equiv -\delta^j_i\frac{d(\delta t)}{dt} - \dot{x}^j\frac{\partial}{\partial \dot{x}^i}\left[\frac{d(\delta t)}{dt}\right] + \frac{\partial(\delta x^j)}{\partial x^i}, \tag{2.8}$$

$$\frac{\partial}{\partial x^j}\left(\frac{dF}{dt}\right) \equiv \frac{d}{dt}\left(\frac{\partial F}{\partial x^j}\right), \tag{2.9}$$

$$\frac{d}{dt}\frac{\partial F}{\partial t} \equiv \frac{\partial}{\partial t}\frac{dF}{dt}, \tag{2.10}$$

where in operations involving partial differentiation the arguments $(\ddot{x}, \dot{x}, x, t)$ are considered as independent variables.

In addition for a function $F(\dot{x}, x, t)$ we write the Euler–Lagrange operator in the form

$$\Lambda_i(F) \equiv \left(\frac{d}{dt}\frac{\partial}{\partial \dot{x}^i} - \frac{\partial}{\partial x^i}\right)F = \ddot{x}^j\frac{\partial^2 F}{\partial \dot{x}^j\partial \dot{x}^i} + \dot{x}^j\frac{\partial^2 F}{\partial x^j\partial \dot{x}^i} + \frac{\partial^2 F}{\partial t\partial \dot{x}^i} - \frac{\partial F}{\partial x^i}. \tag{2.11}$$

It is found that

$$\Lambda_i(F)\dot{x}^i \equiv \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}^i}\dot{x}^i - F\right) + \frac{\partial F}{\partial t}, \tag{2.12}$$

$$\Lambda_i\left[\frac{d\psi(x,t)}{dt}\right] \equiv 0, \tag{2.13}$$

$$\Lambda_i(L+M) \equiv \Lambda_i(L) + \Lambda_i(M). \tag{2.14}$$

3. TYPE I SYMMETRY EQUATIONS BASED ON LAGRANGE'S EQUATIONS

In this section we obtain two (equivalent) forms for the time-dependent Type I symmetry equations of a dynamical system described by Lagrange's equations.

From (2.11) we may express Lagrange's equations in the form

$$\Lambda_i(L) = 0, \tag{3.1}$$

where the Lagrangian $L \equiv L(\dot{x}^1, \ldots, \dot{x}^n, x^1, \ldots, x^n, t)$. By means of (2.11) with $F = L$ and (2.3) we find that

$$\begin{aligned}\delta\Lambda_i(L) = {} & \left[\frac{\partial^2 L}{\partial \dot{x}^k\partial \dot{x}^i}\right]\delta \ddot{x}^k \\ & + \left[\frac{\partial^3 L}{\partial x^j\partial \dot{x}^k\partial \dot{x}^i}\dot{x}^j + \frac{\partial^3 L}{\partial \dot{x}^k\partial \dot{x}^j\partial \dot{x}^i}\ddot{x}^j + \frac{\partial^3 L}{\partial \dot{x}^k\partial \dot{x}^i\partial t} + \frac{\partial^2 L}{\partial x^k\partial \dot{x}^i} - \frac{\partial^2 L}{\partial x^i\partial \dot{x}^k}\right]\delta \dot{x}^k \\ & + \left[\frac{\partial^3 L}{\partial x^k\partial x^j\partial \dot{x}^i}\dot{x}^j + \frac{\partial^3 L}{\partial x^k\partial \dot{x}^j\partial \dot{x}^i}\ddot{x}^j + \frac{\partial^3 L}{\partial x^k\partial \dot{x}^i\partial t} - \frac{\partial^2 L}{\partial x^k\partial x^i}\right]\delta x^k \\ & + \left[\frac{\partial^3 L}{\partial x^j\partial \dot{x}^i\partial t}\dot{x}^j + \frac{\partial^3 L}{\partial \dot{x}^j\partial \dot{x}^i\partial t}\ddot{x}^j + \frac{\partial^3 L}{\partial \dot{x}^i\partial t\partial t} - \frac{\partial^2 L}{\partial x^i\partial t}\right]\delta t.\end{aligned} \tag{3.2}$$

To obtain the explicit form of the symmetry equations, we start with the basic requirement

$$\delta\Lambda_i(L) = 0. \tag{3.3}$$

Then in (3.3) the quantities $\delta\ddot{x}^k$, $\delta\dot{x}^k$, δx^k, δt are replaced by their definitions (2.2), (2.1), (1.1), (1.2), respectively. The terms $\ddot{x}^k$ in the resulting equations are eliminated by the use of Lagrange's equations (3.1) [which are assumed to be solvable for the $\ddot{x}^k$]. The equations thus obtained when considered as identically zero in the $\dot{x}$ variables will give the symmetry equations in the mapping functions $\xi^i(x,t)$, $\xi^0(x,t)$.[6]

We now express (3.2) in an alternative form which has the advantage of leading to the immediate construction of a constant of motion associated with a Type I symmetry.[7] From (2.11) with $F = L$ we have

$$\delta\Lambda_i(L) = \delta\left(\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}\right) - \delta\left(\frac{\partial L}{\partial x^i}\right). \tag{3.4}$$

By use of (2.5) with $F=\partial L/\partial\dot{x}^i$ and (2.7) with $F=L$ we expand the first term on the right-hand side of (3.4) and make use of (2.8) in the resulting expansion. We rewrite the second term of the right-hand side of (3.4) by use of (2.6) with $F=L$. Equation (3.4) then takes the form

$$\delta\Lambda_i(L)\equiv\Lambda_i(\delta L) + \left[\dot{x}^j\left(\frac{d}{dt}\frac{\partial L}{\partial\dot{x}^j}\right)+\ddot{x}^j\frac{\partial L}{\partial\dot{x}^j}\right]\frac{\partial}{\partial\dot{x}^i}\left(\frac{d(\delta t)}{dt}\right) + \frac{\partial L}{\partial\dot{x}^j}\left[\frac{\partial(\delta\dot{x}^j)}{\partial x^i}-\frac{d}{dt}\left(\frac{\partial(\delta x^j)}{\partial x^i}\right)\right] - \Lambda_j(L)\frac{\partial(\delta x^j)}{\partial x^i} + \frac{\partial L}{\partial\dot{x}^i}\frac{d^2(\delta t)}{dt^2}+\dot{x}^j\frac{\partial L}{\partial\dot{x}^j}\frac{d}{dt}\left[\frac{\partial}{\partial\dot{x}^i}\left(\frac{d(\delta t)}{dt}\right)\right] + \frac{\partial L}{\partial t}\frac{\partial(\delta t)}{\partial x^i}. \quad (3.5)$$

If we make use of (2.4) and (2.11) (with $F=L$) in the second term of the right-hand side of (3.5), and of (2.9) (with $F=L$) and (2.1) in the third term of the right-hand side of (3.5), and then we make use of the identity

$$\frac{\partial L}{\partial\dot{x}^i}\frac{d^2(\delta t)}{dt^2}+\frac{dL}{dt}\frac{\partial}{\partial\dot{x}^i}\left(\frac{d(\delta t)}{dt}\right) \equiv\Lambda_i\left(L\frac{d(\delta t)}{dt}\right)-\Lambda_i(L)\frac{d(\delta t)}{dt}-L\frac{d}{dt}\left[\frac{\partial}{\partial\dot{x}^i}\left(\frac{d(\delta t)}{dt}\right)\right] + L\frac{\partial}{\partial x^i}\left(\frac{d(\delta t)}{dt}\right), \quad (3.6)$$

it will follow that we may express (3.5) in the form

$$\delta\Lambda_i(L)\equiv\Lambda_i\left(\delta L+L\frac{d(\delta t)}{dt}\right)+E\Lambda_i\left(\frac{d(\delta t)}{dt}\right)+\Lambda_j(L)B^j_i, \quad (3.7)$$

where

$$B^j_i\equiv\dot{x}^j\frac{\partial}{\partial\dot{x}^i}\left(\frac{d(\delta t)}{dt}\right)-\frac{\partial(\delta x^j)}{\partial x^i}-\delta^j_i\frac{d(\delta t)}{dt}, \quad (3.8)$$

$$E\equiv\dot{x}^i\frac{\partial L}{\partial\dot{x}^i}-L. \quad (3.9)$$

By (2.13) it follows that the coefficient of E in (3.7) is zero for the Type I mappings.

The above-mentioned alternative form of the symmetry condition can now be obtained from (3.7) by requiring that $\delta\Lambda_i(L)=0$ for all solutions of (3.1). This gives as the alternative form

$$\Lambda_i\left(\delta L+L\frac{d(\delta t)}{dt}\right)=0. \quad (3.10)$$

We summarize the above results in the theorem to follow.

Theorem:3.1: A necessary and sufficient condition that the Type I mapping (1.1) and (1.2) define a (Type I) symmetry of a dynamical system described by Lagrange's equations (3.1) with $L=L(\dot{x},x,t)$ is that (3.3) or its equivalent (3.10) be satisfied for all solutions of Lagrange's equations (3.1) (where the δ operator is defined in Sec. 2).

Remarks: (a) See comments following (3.3); (b) the sufficient condition of Theorem 3.1 follows immediately from (3.7).

A Type I mapping (1.1) and (1.2) which satisfies the condition[8]

$$N\delta a\equiv\delta L+L\frac{d(\delta t)}{dt}=-\frac{d(\delta\Omega)}{dt} \quad (3.11)$$

for some function $\Omega(x,t)$ is called a Type I Noether mapping. It follows from (3.11), (2.13), and (3.10) that such mappings are Type I symmetries, which we refer to as Type I Noether symmetries.

However every Type I symmetry is not necessarily a Type I Noether symmetry, that is, every solution of (3.10) will not satisfy (3.11). To show this it is sufficient to consider the example for which $L=\frac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j$.

Hence we may state the following theorem.

Theorem 3.2: Every type I Noether symmetry [based upon mapping (1.1) and (1.2)] as defined by (3.11) will be a symmetry mapping of Lagrange's equations as defined in Theorem 3.1. However, every Type I symmetry of Lagrange's equations will in general not be a Type I Noether symmetry.

As an application of Theorem 3.1 we choose the Lagrangian L to be[9]

$$L\equiv\tfrac{1}{2}\eta_{AB}x^A\dot{x}^B-H(x^A,t),\quad A,B=1,\ldots,2n. \quad (3.12)$$

(For this choice of L it is known that Lagrange's equations $\Lambda_B(L)=0$ reduce Hamilton's equations $\dot{x}^A=\eta^{AB}H_{,B}$.) If (3.10) is expanded for L defined by (3.12) we obtain the phase-space symmetry equation[10]

$$\xi^A_{,B}H^B-\xi^BH^A_{,B}+\xi^A_{,t}-\xi^0_{,B}H^BH^A-\xi^0H^A_{,t}-\xi^0_{,t}H^A=0, \quad (3.13)$$

where $H^A\equiv\eta^{AB}H_{,B}$.

4. CONSTANTS OF MOTION ASSOCIATED WITH TYPE I SYMMETRIES

We now derive a formula for constants of motion associated with Type I symmetries (as defined by Theorem 3.1) for all Lagrangians of the form $L(\dot{x},x)$ and for a large class of Lagrangians of the form $L(\dot{x},x,t)$.

Contract (3.7) with $\dot{x}^i$ and use (2.12) with $F=N$ [where N is defined in (3.11)] to obtain[7]

$$\delta\Lambda_i(L)\dot{x}^i=\left[\frac{d}{dt}\left(\frac{\partial N}{\partial\dot{x}^i}\dot{x}^i-N\right)+\frac{\partial N}{\partial t}\right]\delta a+E\Lambda_i\left(\frac{d(\delta t)}{dt}\right) + \Lambda_j(L)B^j_i\dot{x}^i. \quad (4.1)$$

For a Type I mapping the well-known Noether identity[8] may be expressed in the form

$$N=\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{x}^k}\xi^k-E\xi^0\right)-\Lambda_k(L)(\xi^k-\dot{x}^k\xi^0). \quad (4.2)$$

If in (4.1) we eliminate the term $\partial N/\partial t$ by means of (4.2), make use of (2.10) with $F=(\partial L/\partial\dot{x}^k)\delta x^k-E\delta t$, and use (3.8), we obtain

$$\delta\Lambda_i(L)\dot{x}^i = \left[\frac{dM_1}{dt} - \frac{\partial\Lambda_k(L)}{\partial t}(\xi^k - \dot{x}^k\xi^0) + E\Lambda_k(\xi^0)\dot{x}^k - \Lambda_k(L)\dot{\xi}^k\right]\delta a \qquad (4.3)$$

where

$$M_1 \equiv \left(\frac{\partial N}{\partial \dot{x}^i}\dot{x}^i - N\right) + \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{x}^k}\xi^k - E\xi^0\right). \qquad (4.4)$$

For Type I mappings which define symmetries the left-hand side and the fourth term of the right-hand side of (4.3) will vanish by Theorem (3.1). The third term on the right-hand side of (4.3) vanishes by means of (2.13). It follows that if there exists a function $\gamma_1(x,t)$ such that the second term of the right-hand side satisfies the condition (along a trajectory)

$$R_1 \equiv \frac{\partial\Lambda_k(L)}{\partial t}(\xi^k - \dot{x}^k\xi^0) + \frac{d\gamma_1}{dt} = 0, \qquad (4.5)$$

then along a trajectory $d(M_1+\gamma_1)/dt = 0$, and hence

$$C_1 \equiv M_1 + \gamma_1 = \left(\frac{\partial N}{\partial \dot{x}^i}\dot{x}^i - N\right) + \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{x}^k}\xi^k - E\xi^0\right) + \gamma_1 \qquad (4.6)$$

will be a constant of motion.[11]

Express the Type I Noether symmetry condition (3.11) in the equivalent form

$$R_2 \equiv \delta L + L\frac{d}{dt}(\delta t) + \frac{d\gamma_2}{dt}\delta a = 0, \quad (\gamma_2(x,t)), \qquad (4.7)$$

where $\delta\Omega = \gamma_2\delta a$. Then [as is well known[12] by use of the Noether identity (4.2)] if the condition (4.7) is satisfied the function C_2 defined by

$$C_2 \equiv \frac{\partial L}{\partial \dot{x}^k}\xi^k - E\xi^0 + \gamma_2 \qquad (4.8)$$

will be a (Noether) constant of motion.

Suppose, finally, that a dynamical system admits a Type I symmetry which in addition satisfies both conditions (4.5) and (4.7). In this case it can be shown that the function M_1 of (4.4) is expressible in the form

$$M_1 \equiv M_{12} = \frac{\partial C_2}{\partial t} \quad (R_1 = 0, R_2 = 0). \qquad (4.9)$$

The constant of motion C_1 of (4.6) will then take the form[13]

$$C_1 \equiv C_{12} = M_{12} + \gamma_1 = \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{x}^k}\xi^k - E\xi^0 + \gamma_2\right) + \gamma_1, \quad (R_1 = R_2 = 0). \qquad (4.10)$$

A sufficient condition, independent of the symmetry mapping, that (4.5) holds is that $\gamma_1 = 0$ and

$$\frac{\partial\Lambda_k(L)}{\partial t} = 0 \quad \text{(along a trajectory)}. \qquad (4.11)$$

A very general class of Lagrangians for which (4.11) holds is given by those $L = L(\dot{x}, x, t)$ which satisfy equations of the form[14]

$$\frac{\partial\Lambda_i(L)}{\partial t} = A_i^j\Lambda_j(L), \quad A_i^j(\ddot{x}, \dot{x}, x, t). \qquad (4.12)$$

An important class of Lagrangians L for which (4.11) is satisfied is given by $L = L(\dot{x}, x)$. It may be verified for such Lagrangians that the constant of motion (4.6) (with $\gamma_1 = 0$) may be written in the form [see (3.9)]

$$C_1 = \delta E, \quad (\gamma_1 = 0). \qquad (4.13)$$

We summarize the above in the following theorem.

Theorem 4.1: If a dynamical system admits a Type I symmetry as described in Theorem 3.1, and

(a) if (4.5) holds (case $R_1 = 0$), then the dynamical system admits a constant of motion $C_1 = M_1 + \gamma_1$ [defined by (4.6)];

(b) if the symmetry is a Noether symmetry, that is (4.7) holds (case $R_2 = 0$), then the dynamical system admits the Noether constant of motion[12] C_2 defined by (4.8);

(c) if the symmetry satisfies both (4.5) and (4.7) [case $R_1 = 0$, $R_2 = 0$], then the dynamical system admits the constant of motion $C_{12} = M_{12} + \gamma_1$ defined by (4.10).

Corollary 4.1: If a dynamical system admits a Type I symmetry and in addition there exists functions $A_i^j(\ddot{x}, \dot{x}, x, t)$ such that the Lagrangian satisfies (4.12), then the system admits the constant of motion C_1 defined by (4.6) (with $\gamma_1 = 0$). In particular if $L = L(\dot{x}, x)$, (4.12) is satisfied with $A_i^j = 0$ and then C_1 may be expressed in the form $C_1 = \delta E$ where E is defined by (3.9).

In a previous paper (Ref. 2) a related integral theorem was obtained for time-dependent dynamical systems based upon Hamilton's equations which admitted (time-dependent) symmetry mappings. A similar theorem can be proved for the Lagrangian formulation of such dynamical systems which admit Type I symmetry mappings. Such a theorem may be stated in the following form.

Theorem 4.2: If a dynamical system (3.1) based upon Lagrangian $L(\dot{x}, x, t)$ admits a Type I symmetry mapping $\delta x^i, \delta t$ (as described in Theorem 3.1) and also admits a constant of motion $K(\dot{x}, x, t)$, then (in general) the system will admit an additional constant of motion

$$\delta K \equiv \frac{\partial K}{\partial \dot{x}^i}\delta\dot{x}^i + \frac{\partial K}{\partial x^i}\delta x^i + \frac{\partial K}{\partial t}\delta t. \qquad (4.14)$$

This theorem may be proved in a manner similar to that of a corresponding theorem of Ref. 1, Sec. 4.

We note that, in the case $L = L(\dot{x}, x)$, (4.13) illustrates Theorem 4.2 since E is then a constant of the motion of the dynamical system.

In order to gain further insight into the significance of the condition (4.5), $R_1 = 0$, we again make use of the Lagrangian (3.12). [Recall that for this case the symmetry equation (3.10) takes the form (3.13).] Based upon this Lagrangian we find from (4.4) (with appropriate changes in index range) that M_1 takes the form

$$M_1(x^A, t) = H_{,B}\xi^B, \quad A, B = 1, \ldots, 2n. \qquad (4.15)$$

In addition it is found that (4.5) becomes

$$R_1 = H_{,Bt}(\xi^B - \dot{x}^B\xi^0) + \frac{d}{dt}\gamma_1(x^A, t) = 0. \qquad (4.16)$$

It follows from (4.16) that if $H=H(x^A)$ and we choose $\gamma_1=0$, then $R_1=0$ and the constant of motion C_1 given by (4.6) takes the form

$$C_1=M_1=H_{,B}\xi^B=\delta H(x^A)/\delta a. \tag{4.17}$$

This result was previously obtained by another method in Ref. 2, Sec. 3.

We now again consider the case in which $H=H(x^A,t)$. It was shown in Ref. 2, Sec. 3 that in order for $\delta H(x^A,t)/\delta a$ to be a constant of motion the condition given by (3.8) of Ref. 2 must be satisfied. If now we choose

$$\gamma_1=H_{,t}\xi^0, \tag{4.18}$$

then from (4.6) and (4.15)

$$C_1=M_1+\gamma_1=H_{,A}\xi^A+H_{,t}\xi^0=\delta H(x,t)/\delta a, \tag{4.19}$$

and the condition (4.16), $R_1=0$, reduces precisely to (3.8) of Ref. 2.

5. GAUGE INVARIANCE

It is well known that Lagrange's equations corresponding to a given Lagrangian $L(\dot{x},x,t)$ remain unchanged under the gauge transformation[15]

$$L(\dot{x},x,t)\to L'(\dot{x},x,t)\equiv L+\frac{d}{dt}\psi(x,t), \tag{5.1}$$

in that

$$\Lambda_i(L')=\Lambda_i\left(L+\frac{d\psi}{dt}\right)=\Lambda_i(L), \tag{5.2}$$

as follows from (2.13) and (2.14).

We now show that the restricted Type I symmetry equations as defined by the pair of equations (3.3) and (4.5) are likewise gauge invariant. In the process it will be shown that (3.3) and (4.5) are individually gauge invariant.

In terms of the Lagrangian L' given by (5.1) the above-mentioned restricted Type I symmetry equations are defined by

$$\delta'\Lambda_k(L')=0, \tag{5.3}$$

$$R_1'\equiv\frac{\partial\Lambda_k(L')}{\partial t}(\xi'^k-\dot{x}^k\xi'^0)+\frac{d\gamma_1'}{dt}=0, \tag{5.4}$$

where the δ' variation is based upon mappings of the form (1.1) and (1.2) with

$$\delta' x^i\equiv\xi'^i(x,t)\delta a, \quad \delta' t\equiv\xi'^0(x,t)\delta a. \tag{5.5}$$

By use of (5.1), (2.13), and (2.14) it follows that the left-hand sides of (5.3) and (5.4) may be expressed respectively in the forms

$$\delta'\Lambda_k(L')=\delta'\Lambda_k(L), \tag{5.6}$$

$$\frac{\partial\Lambda_k(L')}{\partial t}(\xi'^k-\dot{x}^k\xi'^0)+\frac{d\gamma_1'}{dt}=\frac{\partial\Lambda_k(L)}{\partial t}(\xi'^k-\dot{x}^k\xi'^0)+\frac{d\gamma_1'}{dt}. \tag{5.7}$$

From (5.3) and (5.6) and from (5.4) and (5.7) we have respectively

$$\delta'\Lambda(L)=0, \tag{5.8}$$

$$\frac{\partial\Lambda_k(L)}{\partial t}(\xi'^k-\dot{x}^k\xi'^0)+\frac{d\gamma_1'}{dt}=0. \tag{5.9}$$

A comparison of (5.8) with (3.3) and (5.9) with (4.5) shows the pair (5.8) and (5.9) defines the same set of restricted Type I symmetry mappings as the pair (3.3) and (4.5), that is, $\delta'=\delta$, $\gamma_1'=\gamma_1$.

Hence from (5.6) and (5.7) we may write

$$\delta'\Lambda_k(L')=\delta\Lambda_k(L), \tag{5.10}$$

$$\frac{\partial\Lambda_k(L')}{\partial t}(\xi'^k-\dot{x}^k\xi'^0)+\frac{d\gamma_1'}{dt}=\frac{\partial\Lambda_k(L)}{\partial t}(\xi^k-\dot{x}^k\xi^0)+\frac{d\gamma_1}{dt}. \tag{5.11}$$

Thus (3.3) and (4.5) are gauge invariant.

Since the alternative form of the symmetry equation (3.10) is equivalent to (3.3), it is clear that (3.10) will also exhibit gauge invariance, that is under the transformation (5.1)

$$\Lambda_k\left(\delta'L'+L'\frac{d(\delta't)}{dt}\right)=\Lambda_k\left(\delta L+L\frac{d(\delta t)}{dt}\right). \tag{5.12}$$

This, in fact, can be shown directly by starting with the left-hand side of (5.12), making use of (5.1), (2.14), (2.5), (2.13), and employing the rationale which led to (5.10).

We collect the above results in the form of the following theorem.

Theorem 5.1: The Type I symmetry equations (3.3) or their equivalent (3.10) are gauge invariant under the transformation (5.1). The restricted Type I symmetry equations defined by (3.3) and (4.5) or their equivalent (3.10) and (4.5) are also gauge invariant. These statements imply (5.10), (5.11), and (5.12) hold under the transformation (5.1).

A theorem similar to Theorem 5.1 can be proved for Type I Noether symmetry mappings.[16] To see this we consider the Noether symmetry condition (3.11) expressed in terms of Lagrangian L',

$$\delta'L'+L'\frac{d(\delta't)}{dt}=-\frac{d(\delta'\Omega')}{dt}, \tag{5.13}$$

where the δ' variation is defined by (5.5). By use of (5.1) and (2.5) it follows from (5.13) that

$$\delta'L+L\frac{d(\delta't)}{dt}=-\frac{d[\delta'(\Omega'+\psi)]}{dt}. \tag{5.14}$$

A comparison of the Noether symmetry condition (3.11) and (5.14) shows that both define the same set of Type I Noether symmetry mappings, with $\Omega\equiv\Omega'+\psi$, and hence $\delta'=\delta$. Therefore the Type I Noether symmetry mappings are gauge invariant.

We may hence state the following.

Theorem 5.2: The set of Type I Noether symmetry mappings as defined by (3.11) is invariant under the gauge transformation (5.1).

We now examine the effect of the gauge transformation (5.1) on the constants of motion discussed in Theorem 4.1.

First we consider the constant of motion $C_1=M_1+\gamma_1$

defined in (4.6). In terms of the Lagrangian L' we have

$$C_1' \equiv M_1' + \gamma_1' \equiv \left(\frac{\partial N'}{\partial \dot{x}^i}\dot{x}^i - N'\right) + \frac{\partial}{\partial t}\left(\frac{\partial L'}{\partial \dot{x}^i}\xi'^i - E'\xi'^0\right) + \gamma_1', \qquad (5.15)$$

where ξ'^i, ξ'^0, δ', and γ_1' are those used in the proof of Theorem 5.1. From (5.1), (3.11), (3.9), and the fact that $\delta' = \delta$, $\gamma_1' = \gamma_1$ we have

$$N'\delta a = N\delta a + \frac{d}{dt}(\delta\psi), \qquad (5.16)$$

$$E' = E - \frac{\partial \psi}{\partial t}. \qquad (5.17)$$

By use of (5.16) and (5.17) in (5.15) we find $M_1' = M_1$, and hence we may write $M_1' + \gamma_1' = M_1 + \gamma_1$, which shows the constant of motion $C_1 = M_1 + \gamma_1$ is gauge invariant. Note also that M_1 [as defined in (4.4) is gauge invariant].

We consider next the effect of the gauge transformation on the Noether constant of motion C_2 defined in (4.8). In terms of the Lagrangian L' we have

$$C_2'\delta a = \frac{\partial L'}{\partial \dot{x}^k}\delta' x^k - E'\delta' t + \delta'\Omega', \quad (\delta'\Omega' \equiv \gamma_2'\delta a), \qquad (5.18)$$

where δ' and Ω' are those used in the proof of Theorem 5.2.

By use of (5.1) and (5.17), the relation $\Omega' = \Omega - \psi$, and the fact $\delta' = \delta$ (refer to the proof of Theorem 5.2), we may reduce (5.18) to the form $C_2' = C_2$, which shows the constant of motion C_2 is gauge invariant.

It is clear that the constant of motion C_{12} defined in (4.10) will also be gauge invariant.

The above results on constants of motion are stated in the form of the following Theorem.

Theorem 5.3: The three constants of motion C_1, C_2, C_{12} of Theorem 4.1 are gauge invariant with respect to the transformation (5.1).

[1]G.H. Katzin and J. Levine, J. Math. Phys. **15**, 1460 (1974).
[2]G.H. Katzin and J. Levine, J. Math. Phys. **16**, 548 (1975).
[3]See Ref. 1, Secs. 2 and 3 and Ref. 2, Sec. 1.
[4]Indices will have the range 1 through n (unless otherwise indicated), and the Einstein summation notation is used unless otherwise indicated.
[5]A dot (·) indicates total time derivative d/dt.
[6]For a somewhat more detailed discussion of this procedure see Ref. 1, Sec. 3 which treats the time-independent problem.
[7]In many of the derivations to follow the term $d(\delta t)/dt$ is left unexpanded in order to facilitate a latter modification of these derivations to Type II symmetries.
[8]The middle term of (3.11) is recognized as part of the expression of the Noether identity. For a simple derivation of this identity see Sec. 8 and Ref. 34 of Ref. 1.
[9]See Sec. 7 of Ref. 1 or Ref. 2 for notation.
[10]See Sec. 2 of Ref. 2.
[11]An analogous technique based upon time-independent Type II mappings {refer to (1.1), (1.3) with $\xi^i = \xi^i(x)$, $\phi = \phi[x(t)]$} was used in our earlier paper (Ref. 1) to obtain a similar (time-independent) constant of motion for a Lagrangian system with $L = L(\dot{x}, x)$. A similar technique to obtain time-independent constants of motion based upon a Newtonian formulation with time-independent forces $\mathbf{F}(x)$ was used by K.H. Mariwalla, Lett. Nuovo Cimento **12**, 253 (1975). (Apparently he was unaware of our work, Ref. 1.)
[12]See Sec. 8 of Ref. 1.
[13]It is also worth noting that C_{12} of (4.10) can be obtained directly by forming the partial time derivative of the Noether identity (4.2) and then making use of (3.11), (2.10), and (4.5).
[14]Examples of Lagrangians which satisfy equations of the form (4.12) will be included in Paper II.
[15]We adopt the terminology "gauge" transformation for such transformations of the Lagrangian. With respect to this nomenclature refer to H. Rund, *The Hamilton—Jacobi Theory in the Calculus of Variations* (Van Nostrand, London, 1966), p. 163, or E.J. Konopinski, *Classical Descriptions of Motion* (Freeman, San Francisco, 1969), p. 173.
[16]The gauge invariance of Type I Noether symmetry mappings is probably known. For completeness, however, we give a short derivation.

Inverse scattering problems in absorbing media*

M. Jaulent[†]

Département de Physique Mathématique, Université des Sciences et Techniques du Languedoc, 34060-Montpellier Cedex, France
(Received 3 October 1975)

We study inverse scattering problems which occur in various fields of physics (transmission lines theory, electromagnetism, elasticity theory), and in which the inhomogeneous media considered are absorbing. We suppose that waves propagate in a z direction from $z=0$ to $z=\infty$ and are totally reflected at $z=0$, the input data being the values of the reflection coefficient to the right $S^+(k)$ for all frequencies k (note that the case where waves propagate from $z=-\infty$ to $z=\infty$ can be studied in a similar way). These problems are reduced to an inverse scattering problem for the radial s-wave Schrödinger equation with an energy-dependent potential. The case where $S^+(k)$ is close to 1 and the case where the absorption is weak are specially investigated, and a class of exactly solvable examples is given.

I. INTRODUCTION

In many areas of physics the interest is determinating the characteristics of a medium from limited knowledge of fields propagating through this medium. Evidently, this inverse scattering problem is closely related to a synthesis problem wherein the task is to construct a special medium reproducing some prescribed properties of the fields. In certain cases, where waves propagate in a z direction, this inverse problem can be reduced by means of the Liouville transformation[1] to an inverse scattering problem for the radial s wave, or for the one-dimensional Schrödinger equation with an energy-independent potential. So the works of Gel'fand and Levitan[2] and Agranovich and Marchenko[3] in the radial case, and those of Kay,[4] Kay and Moses,[5] and Faddeev[6] in the one-dimensional case, are powerful theoretical tools of investigation. The input data necessary for this investigation are in most cases the values of a reflection coefficient for all frequencies of the waves. In the case of a nonuniform transmission line one can thus seek the inductance $L(z)$ and the capacitance $C(z)$ per unit length, z being the physical position on the line (for a survey of this inverse problem, see Kay[7]). In electromagnetic wave propagation through an inhomogeneous medium, plane stratified in the z direction, the desired quantity is the dielectric constant $\epsilon(z)$ (see Moses and De Ridder[8]). In the analogous problem in elastic wave propagation the desired quantities are the density $\rho(z)$ and the elastic parameter $E(z)$ (see Ware and Aki[9]).

A remarkable common feature of the above inverse problems is that all media considered are lossless. This is an important condition for the results in Refs. 2–6 to be applicable. In this paper we consider inverse problems in absorbing media (some results of this paper were announced in Ref. 10). So in a transmission line we consider, in addition to nonuniform $L(z)$ and $C(z)$, a nonuniform series resistance $R(z)$ per unit length or a nonuniform shunt conductance $G(z)$ per unit length. In a similar way, we introduce a nonuniform conductivity $\sigma(z)$ in electromagnetic scattering and a nonuniform absorption coefficient $\alpha(z)$ (according to the Maxwell solid theory) in elastic wave propagation. For the sake of convenience, we consider in detail only the case where waves propagate in the z direction from $z=0$ to $z=\infty$ and are totally reflected at $z=0$. The input data are the values of the "reflection coefficient to the right" for all frequencies. Such inverse problems will be called radial inverse problem in absorbing media (RIPAM). We will indicate briefly in Sec. IV how this study can be easily transposed to the case of one-dimensional inverse problems in absorbing media with waves propagating in the z direction from $z=-\infty$ to $z=\infty$. Note that these inverse problems can be formulated in an equivalent way in terms of the time-dependent wave equations. In this form, Weston[11] and Weston and Krueger[12] have studied the case corresponding to one-dimensional media in which characteristics do not vary outside of a compact support. Their resolving method is different from the one given here and is obtained by treating the problem as a Cauchy initial value problem and using the Riemann function to deduce a dual set of integral equations of the Gel'fand-Levitan type.

In Sec. II by means of the Liouville transformation, we show that the RIPAM can be reduced to an inverse scattering problem (called RIPEP) for the radial s-wave Schrödinger equation ($x \geq 0$)

$$y^{+\prime\prime} + [k^2 - V^+(k,x)]y^+ = 0, \tag{1.1}$$

with the energy-dependent potential

$$V^+(k,x) = U(x) + 2kQ(x) \tag{1.2}$$

(in quantum mechanics, $E=k^2$ is the energy). $U(x)$ must be real and $Q(x)$ purely imaginary. In the more general case where $U(x)$ and $Q(x)$ are complex, we gave in collaboration with Jean (see Ref. 13) a method of solving the RIPEP. It generalizes the Agranovich-Marchenko method[3] valid for $Q(x)=0$. The input data is the "scattering matrix" $S^+(k)$ ($k \in \mathbb{R}$). Note that only the part $S^+(k)$ ($k>0$) has a physical meaning. In general the part $S^+(k)$ ($k \leq 0$) is unknown and plays the role of a parameter in the solution of the "physical" RIPEP. We gave no general proof for the existence and uniqueness of a solution of the RIPEP, though arguments to expect them in many cases have been given. Indeed, for real $U(x)$ and $Q(x)$, these questions can be completely answered.[14,15] Section III is devoted to the investigation of the RIPEP for real $U(x)$ and purely imaginary $Q(x)$. A peculiar point of this case is that, thanks to the complex conjugate relations $\overline{S^+(k)} = S^+(-k)$, the scattering matrix $S^+(k)$ ($k \in \mathbb{R}$) which coincides with the reflection coefficient to the right of the RIPAM (k is then the frequency) is completely determined from its physical part $S^+(k)$ $(k>0)$. So the RIPEP and the "physical" RIPEP coincide in this case. In Sec. IIIA we improve some steps of the general study given in Ref. 13 but we do not obtain as refined results as for real $U(x)$ and $Q(x)$. $\mathcal{V}$ being a large class of pairs $(U(x), Q(x))$, we state that the scattering matrix associated with a pair $(U(x), Q(x))$ in $\mathcal{V}$ belongs to the class $\mathcal{S}$ of functions $S^+(k)$ $(k\in\mathbf{R})$ satisfying certain conditions I_1 and I_2. In particular, these imply the existence of two real integrable functions $s^+(t)$ and $s^-(t)$ $(t\in\mathbf{R})$ and two strictly positive numbers F_0^+ and $F_0^-=[F_0^+]^{-1}$ such that $S^+(k)$ and $S^-(k)=[S^+(-k)]^{-1}$ can be written in the form

$$S^{\pm}(k)=(F_0^{\pm})^2+\int_{-\infty}^{\infty}s^{\pm}(t)\exp(-ikt)\,dt,\quad k\in\mathbf{R}. \qquad (1.3)$$

We prove that the RIPEP has a unique solution in $\mathcal{V}$ if the input function $S^+(k)$ $(k\in\mathbf{R})$ belongs to a subclass of $\mathcal{S}$, $\mathcal{S}'$, consisting of functions satisfying certain additional complicated conditions I_3, I_4, I_5, and I_6. Our method is based on the solution of the following system S of equations in which $s^+(t)$, $s^-(t)$, F_0^+, and F_0^- are known:

$$A^+(x,t)=F^-(x)s^+(x+t)+\int_x^{\infty}A^-(x,u)s^+(u+t)\,du,\quad t\geq x\geq 0, \qquad (1.4)$$

$$A^-(x,t)=F^+(x)s^-(x+t)+\int_x^{\infty}A^+(x,u)s^-(u+t)\,du,\quad t\geq x\geq 0, \qquad (1.5)$$

$$F^+(x)F^-(x)=1,\quad x\geq 0,\quad F^{\pm}(x)>0,$$
$$F^{\pm}(0)=F_0^{\pm},\quad F^{\pm\prime}(\infty)=0, \qquad (1.6)$$

$$f^+(x)F^-(x)=f^-(x)F^+(x),\quad x\geq 0, \qquad (1.7)$$

where

$$f^{\pm}(x)=F^{\pm\prime\prime}(x)-2\frac{d}{dx}A^{\pm}(x,x)+2F^{\pm\prime}(x)[F^{\pm}(x)]^{-1}A^{\pm}(x,x),\quad x\geq 0. \qquad (1.8)$$

The pair $(U(x), Q(x))$, the solution of the RIPEP, is obtained from the solution $(F^+(x), F^-(x), A^+(x,t), A^-(x,t))$ of the system S by formulas

$$Q(x)=\mp iF^{\pm\prime}(x)[F^{\pm}(x)]^{-1},\quad U(x)=f^{\pm}(x)[F^{\pm}(x)]^{-1},\quad x\geq 0. \qquad (1.9)$$

Note that for the radial s-wave Schrödinger equation with the potential $V^{\pm}(k,x)=U(x)\pm 2kQ(x)$, the Jost solution $f^{\pm}(k,x)$, i.e., the solution whose asymptotic behavior as $x\to\infty$ is $\exp(-ikx)$, is given by

$$f^{\pm}(k,x)=F^{\pm}(x)\exp(-ikx)+\int_x^{\infty}A^{\pm}(x,t)\times\exp(-ikt)\,dt,\quad \mathrm{Im}k\leq 0,\quad x\geq 0. \qquad (1.10)$$

We solve S in two steps. We first prove that Eqs. (1.4) and (1.5) are equivalent to the following equation:

$$A^{\pm}(x,t)=F^{\mp}(x)\alpha^{\pm}(x,t)+F^{\pm}(x)\beta^{\mp}(x,t),\quad t\geq x\geq 0, \qquad (1.11)$$

where the functions $\alpha^{\pm}(x,t)$ and $\beta^{\mp}(x,t)$ are completely determined by the data. Then substituting (1.11) into Eqs. (1.6) and (1.7) and writing $F^{\pm}(x)$ in the form $F^{\pm}(x)=\exp[\mp y(x)]$, we are led to solve the differential equation

$$y'=-2\alpha^+(x,x)\exp y+2\alpha^-(x,x)\exp(-y)+2\beta^+(x,x)-2\beta^-(x,x),\quad x\geq 0, \qquad (1.12)$$

with the condition $y(0)=-2\ln F_0^+$.

In Secs. IIIB, IIIC, and IIID we are given a function $S^+(k)$ $(k\in\mathbf{R})$ in $\mathcal{S}$ and we prove that some of the additional conditions I_3, I_4, I_5, and I_6, which have to be imposed on the input function for our inversion procedure to work, are sometimes automatically satisfied. In particular, in Sec. IIIC we prove that this happens for I_3, I_4, and I_5 if there is no absorption. In Secs. IIIB and IIID, respectively, we consider the cases where $S^+(k)$ is sufficiently close to unity and where the absorption is sufficiently weak. These notions are defined by introducing two semidistances d and $\tilde{d}$ in $\mathcal{S}$. In the first case, I_3, I_4, and I_6 are automatically satisfied. In the second case this is still true for I_3 and I_4 but I_6 is only satisfied in certain situations. In Sec. IIIB we also give approximation formulas for the inversion procedure when $S^+(k)$ is very close to unity: we see that $U(x)$ depends only on $\arg S^+(k)$ and $Q(x)$ depends only on $|S^+(k)|$. In Sec. IIIE we give a class of functions $S^+(k)$ belonging to $\mathcal{S}'$ for which the RIPEP can be exactly solved.

II. REDUCTION OF THE RIPAM TO THE RIPEP

A. Definition of the RIPEP

Let us first recall some facts about the scattering problem associated with (1.1) and (1.2) in the radial case. As seen in Ref. 13 it is useful to consider simultaneously both radial Schrödinger equations $(x\geq 0)$

$$y^{\pm\prime\prime}+[k^2-V^{\pm}(k,x)]y^{\pm}=0, \qquad (2.1)$$

with the energy-dependent potentials

$$V^{\pm}(k,x)=U(x)\pm 2kQ(x). \qquad (2.2)$$

If $U(x)$ and $Q(x)$ are complex functions satisfying certain conditions and if $k\in\mathbf{R}^*$, we know from Ref. 13 that there exists a unique solution $\psi^{\pm}(k,x)$ of (2.1) which vanishes for $x=0$ and a unique complex number $S^{\pm}(k)$ such that $\psi^{\pm}(k,x)$ has the following asymptotic behavior as $x\to\infty$:

$$\psi^{\pm}(k,x)=\exp(-ikx)-S^{\pm}(k)\exp(ikx)+o(1). \qquad (2.3)$$

We set $S^{\pm}(0)=1$. The function $S^+(k)$ $(k\in\mathbf{R})$ is called the "scattering matrix" associated with the pair $(U(x), Q(x))$. $f^{\pm}(k,x)$ $(\mathrm{Im}k\leq 0)$ being the Jost solution of (2.1)—i.e., the solution whose asymptotic behavior as $x\to\infty$ is $\exp(-ikx)-f^{\pm}(k)$ $(\mathrm{Im}k\leq 0)$ being the Jost function $f^{\pm}(k,0)$ $(\mathrm{Im}k\leq 0)$ and $\phi^{\pm}(k,x)(k\in\mathbb{C})$ being the regular solution—i.e., the solution defined by the conditions $\phi^{\pm}(k,0)=0$, $\phi^{\pm\prime}(k,0)=1-$, we have the relations

$$S^{\pm}(k)=f^{\pm}(k)[f^{\mp}(-k)]^{-1},\quad k\in\mathbf{R}, \qquad (2.4)$$

$$S^-(k)=[S^+(-k)]^{-1},\quad k\in\mathbf{R}, \qquad (2.5)$$

$$|S^{\pm}(k)|^2=1+k^{-1}\int_0^{\infty}|\psi^{\pm}(k,x)|^2\,\mathrm{Im}V^{\pm}(k,x)\,dx,\quad k\in\mathbf{R}^*, \qquad (2.6)$$

$$\psi^{\pm}(k,x)=-2ik[f^{\mp}(-k)]^{-1}\phi^{\pm}(k,x),\quad x\geq 0,\quad k\in\mathbf{R}^*. \qquad (2.7)$$

In what follows we will be interested in the class $\mathcal{V}$ of pairs of potentials $(U(x), Q(x))$ which satisfy the following conditions:

(i) the function $U(x)$ is continuously differentiable for $x\geq 0$; $xU(x)$ and $xU'(x)$ are integrable in $\mathbf{R}^+$;

(ii) the function $Q(x)$ is twice continuously differentiable for $x \geq 0$; $Q(x)$, $xQ'(x)$, and $xQ''(x)$ are integrable in $\mathbb{R}^+$;

(iii) the functions $f^+(k)$ and $f^-(k)$ have no zero for Im$k \leq 0$ [so there is no bound state for Eqs. (2.1) (+) and (2.1) (−), i.e., there is no square integrable solution $\phi^+(k,x)$ and $\phi^-(k,x)$];

(iv) $U(x)$ is real and $Q(x)$ is purely imaginary;

(v) Im$Q(x) \leq 0$ for $x \geq 0$;

(vi) $f^+(0,x) > 0$ for $x \geq 0$ [note that $f^+(0,x) = f^-(0,x)$].

In that case we have the following complex conjugate relations:

$$\overline{f^\pm(k,x)} = f^\pm(-\bar{k},x), \quad \text{Im}k \leq 0, \tag{2.8}$$

$$\overline{S^\pm(k)} = S^\pm(-k), \quad k \in \mathbb{R}, \tag{2.9}$$

and the following inequality which follows from (2.6):

$$|S^+(k)| \leq 1, \quad k \in \mathbb{R}. \tag{2.10}$$

The RIPEP that we consider here is the construction of the pairs $(U(x), Q(x))$ belonging to $\mathcal{V}$ whose scattering matrix is a given function $S^+(k)$ $(k \in \mathbb{R})$.

It is interesting to remark that the following part of condition (iii) of the class $\mathcal{V}$: $f^-(k)$ has no zero for Im$k < 0$, is in fact a consequence of (i), (ii), (iv), (v), and (vi)—so in particular, if $Q(x) = 0$, (iii) follows from (i), (iv), and (vi); the proof is given in Appendix A.

B. List of scattering problems in absorbing media

The typical situation in which we are interested is that of an inhomogeneous absorbing medium which extends in a z direction from $z=0$ to $z=\infty$, which has characteristics $A(z) > 0$, $B(z) > 0$, and $\Gamma(z) \geq 0$, $\Gamma(z)$ being the characteristic responsible of absorption, and in which a field $u(z,t)$ propagates according to $(z \geq 0)$

$$\frac{\partial}{\partial z}\left(A(z)\frac{\partial u}{\partial z}\right) - B(z)\frac{\partial^2 u}{\partial t^2} - \Gamma(z)\frac{\partial u}{\partial t} = 0, \tag{2.11}$$

with the following condition which expresses the hypothesis that the medium is totally reflecting at $z=0$:

$$u(0,t) = 0. \tag{2.12}$$

For a wave of frequency k, i.e., for $u(z,t) = u(k,z)\exp(-ikt)$, (2.11) and (2.12) are written $(z \geq 0)$

$$\frac{d}{dz}\left(A(z)\frac{du}{dz}\right) + k^2 B(z)u + ik\Gamma(z)u = 0, \tag{2.13}$$

$$u(k,0) = 0. \tag{2.14}$$

In what follows we give examples where such situations occur. This is the case if we consider a nonuniform transmission line extending in a z direction from $z=0$ to $z=\infty$, open at the origin $z=0$, having an inductance $L(z)$, a capacitance $C(z)$ and a series resistance $R(z)$ per unit length. Then we have $u(k,z) \equiv I(k,z)$ (intensity of the current), $A(z) \equiv [C(z)]^{-1}$, $B(z) \equiv L(z)$, and $\Gamma(z) \equiv R(z)$. This is also the case when the line is short circuited at the origin and $R(z)$ is replaced by a shunt conductance $G(z)$ per unit length. Then we have $u(k,z) \equiv V(k,z)$ (voltage), $A(z) \equiv [L(z)]^{-1}$, $B(z) \equiv C(z)$, and $\Gamma(z) \equiv G(z)$.

As another example we can also consider an inhomogeneous medium plane stratified in the z direction for $z \geq 0$, with dielectric constant $\epsilon(z)$, permeability μ_0 (independent of z), and conductivity $\sigma(z)$, limited at $z=0$ by a medium of infinite conductivity, where an electric field $E(k,z)$ polarized in a direction perpendicular to the z direction propagates. We have $u(k,z) \equiv E(k,z)$, $A(z) \equiv 1$, $B(z) \equiv \epsilon(z)\mu_0$, and $\Gamma(z) \equiv \sigma(z)\mu_0$.

Finally, we consider longitudinal propagation in an elastic inhomogeneous medium plane stratified in the z direction for $z \geq 0$, with density $\rho(z)$, elastic parameter $E(z)$ $[=\lambda(z) + 2\mu(z)]$, absorption coefficient $\alpha(z)$ (according to the Maxwell solid theory), and limited at $z=0$ by a rigid surface. Then $u(k,z)$ is the displacement and we have $A(z) \equiv E(z)$, $B(z) \equiv \rho(z)$, and $\Gamma(z) \equiv \alpha(z)$.

C. Definition and reduction of the RIPAM

We set the following assumptions for the functions $A(z)$, $B(z)$, and $\Gamma(z)$ $(z \geq 0)$:

(a) $A(z)$ and $B(z)$ are strictly positive, twice continuously differentiable and have strictly positive finite limits $A(\infty)$ and $B(\infty)$ as $z \to \infty$;

(b) $\Gamma(z)$ is non-negative and continuous.

Let us set (Liouville transformation)

$$x(z) = \int_0^z [A(u)]^{-1/2}[B(u)]^{1/2}\,du, \tag{2.15}$$

$$y^+(k,x) = [A(x)B(x)]^{1/4}u(k,x), \tag{2.16}$$

where we use the convention $u(k,z(x)) = u(k,x)$, $A(z(x)) = A(x)$, etc., which is justified by the one-to-one correspondence between z and x, $x(z)$ varying from $x(0)=0$ to $x(\infty)=\infty$. It is easy to prove that $y^+(k,x)$ satisfies the radial Schrödinger equation (2.1) (+) with the energy-dependent potential (2.2) (+) where

$$U(x) = [A(x)B(x)]^{-1/4}\frac{d^2}{dx^2}[A(x)B(x)]^{1/4},$$

$$Q(x) = -i\Gamma(x)[2B(x)]^{-1}. \tag{2.17}$$

Furthermore, condition (2.14) for $u(k,z)$ implies that $y^+(k,x)$ must vanish at $x=0$. It follows from (2.17) that

$$[A(x)B(x)]^{1/4} = [A(\infty)B(\infty)]^{1/4}f^+(0,x), \quad x \geq 0. \tag{2.18}$$

Let us call $\mathcal{C}$ the class of triplets $(A(z), B(z), \Gamma(z))$ $(z \geq 0)$ which satisfy conditions a and b and for which the pair $(U(x), Q(x))$ defined in (2.17) satisfies the conditions (i), (ii), and (iii) of the class $\mathcal{V}$. It is clear from (2.17) and (2.18) that if $(A(z), B(z), \Gamma(z))$ belongs to $\mathcal{C}$, then $(U(x), Q(x))$ belongs to $\mathcal{V}$. We see from (2.3), (2.15), and (2.16) that $[A(z)]^{1/2}[B(z)]^{-1/2}$ is the local wave velocity, that $x(z)$ is the travel time of waves from the position z to the origin, and that the function $S^+(k)$ $(k \in \mathbb{R})$ can be called the "reflection coefficient to the right" [note that because of (2.9), the knowledge of the function $S^+(k)$ $(k \in \mathbb{R})$ is equivalent to that of its "physical part" $S^+(k)$ $(k > 0)$]. If $\Gamma(z) = 0$, then $Q(x) = 0$ and we have $|S^+(k)| = 1$. In the general case we only have inequality (2.10). Therefore, the quantity $1 - |S^+(k)|$ represents the absorption. The RIPAM that we consider is the construction of the triplets $(A(z), B(z), \Gamma(z))$ belonging to $\mathcal{C}$ whose reflection coefficient to the right is a given function

$S^+(k)$ ($k \in \mathbf{R}$). We shall see below that in many cases there is a unique solution $(U(x), Q(x))$ to the RIPEP. The set of all solutions of the RIPAM is the set of all triplets $(A(z), B(z), \Gamma(z))$ satisfying conditions a and b, which are connected to $(U(x), Q(x))$ by the relations (2.17) and (2.15). The functions $A(x)B(x)[A(\infty)B(\infty)]^{-1}$ ($x \geq 0$) and $\Gamma(x)[B(x)]^{-1}$ ($x \geq 0$) do not depend on the solution considered. These are quantities which are determined in a unique way in the RIPAM. To construct a solution to the RIPAM from $(U(x), Q(x))$ we can arbitrarily choose $B(\infty)$ and $A(x)$ for instance, $B(\infty)$ being any strictly positive real number and $A(x)$ any twice continuously differentiable strictly positive function having a strictly positive finite limit $A(\infty)$ as $x \to \infty$. $B(x)$ and $\Gamma(x)$ are then determined uniquely from (2.18) and (2.17). Equation (2.15) gives the relation $x = x(z)$. Hence the solution $(A(z), B(z), \Gamma(z))$. Varying $B(\infty)$ and $A(x)$ we obtain in this way all solutions of the RIPAM. A special mention should be made for the case where $A(z)$ does not vary and has a known value (this occurs in the example of electromagnetism given in Sec. II B). Then the RIPAM has a unique solution $(B(z), \Gamma(z))$ if we are also given $B(\infty)$.

III. SOLUTION OF THE RIPEP

A. General Solution

Our methods in this paragraph are quite similar to those followed in Refs. 13—15. So we only state the principal results.

First we assert that the scattering matrix associated with a pair $(U(x), Q(x))$ in $\mathcal{V}$ belongs to the class $\mathcal{S}$ of functions $S^+(k)(k \in \mathbf{R})$, which satisfy the following conditions I_1 and I_2, where $S^-(k)$ ($k \in \mathbf{R}$) is the function obtained from $S^+(k)$ ($k \in \mathbf{R}$) by formula (2.5):

(I_1) (a) $S^\pm(k) \neq 0$ ($k \in \mathbf{R}$), $S^\pm(0) = 1$, $|S^+(k)| \leq 1$ ($k \in \mathbf{R}$), $\overline{S^\pm(k)} = S^\pm(-k)$ ($k \in \mathbf{R}$) and $\arg S^\pm(k)|_{-\infty}^{\infty} = 0$;

(b) there exist two real integrable functions $s^+(t)$ and $s^-(t)$ ($t \in \mathbf{R}$) and two strictly positive numbers F_0^+ and $F_0^- = [F_0^+]^{-1}$ such that

$$S^\pm(k) = (F_0^\pm)^2 + \int_{-\infty}^{\infty} s^\pm(t) \exp(-ikt)\, dt, \quad k \in \mathbf{R} \tag{3.1}$$

[note that F_0^+ and F_0^- are defined in a unique way by (3.1), that $F_0^+ \leq 1$ and that condition I_1 for $S^-(k)$ follows from condition I_1 for $S^+(k)$];

(I_2) $s^\pm(t)$ ($t \geq 0$) is twice continuously differentiable and $ts^{\pm\prime}(t)$ ($t \geq 0$) and $ts^{\pm\prime\prime}(t)$ ($t \geq 0$) are integrable.

On the other hand, the Jost solution $f^\pm(k, x)$ of (2.1) is generated by two real functions $F^\pm(x)$ and $A^\pm(x, t)$:

$$f^\pm(k, x) = F^\pm(x) \exp(-ikx) + \int_x^\infty A^\pm(x, t) \exp(-ikt)\, dt, \quad \operatorname{Im} k \leq 0, \quad x \geq 0, \tag{3.2}$$

where

$$F^\pm(x) = \exp\left(\mp i \int_x^\infty Q(t)\, dt\right), \quad x \geq 0. \tag{3.3}$$

$(U(x), Q(x))$ is obtained from $(F^+(x), F^-(x), A^+(x, t), A^-(x, t))$ by the formulas

$$Q(x) = \mp i F^{\pm\prime}(x)[F^\pm(x)]^{-1}, \quad U(x) = f^\pm(x)[F^\pm(x)]^{-1}, \quad x \geq 0, \tag{3.4}$$

where

$$f^\pm(x) = F^{\pm\prime\prime}(x) - 2\frac{d}{dx}A^\pm(x, x) + 2F^{\pm\prime}(x)[F^\pm(x)]^{-1}A^\pm(x, x), \quad x \geq 0. \tag{3.5}$$

$(F^+(x), F^-(x), A^+(x, t), A^-(x, t))$ is solution of the following system S of equations:

$$A^+(x, t) = F^-(x)s^+(x + t) + \int_x^\infty A^-(x, u)s^+(u + t)\, du, \quad t \geq x \geq 0, \tag{3.6}$$

$$A^-(x, t) = F^+(x)s^-(x + t) + \int_x^\infty A^+(x, u)s^-(u + t)\, du, \quad t \geq x \geq 0, \tag{3.7}$$

$$F^+(x)F^-(x) = 1, \quad x \geq 0, \tag{3.8}$$

$$f^+(x)F^-(x) = f^-(x)F^+(x), \quad x \geq 0, \tag{3.9}$$

where $A^+(x, t)(t \geq x \geq 0)$ and $A^-(x, t)(t \geq x \geq 0)$ are twice continuously differentiable real functions bounded by a function $\tilde{\sigma}[(x + t)/2]$—$\tilde{\sigma}(x)$ ($x \geq 0$) being a nonincreasing and integrable positive function—and where $F^+(x)$ can be written in the form

$$F^+(x) = \exp[-y(x)/2], \quad x \geq 0, \tag{3.10}$$

where $y(x)$ is a three-times continuously differentiable real function such that $y'(\infty) = 0$ and $y(0) = -2\ln F_0^+$.

Now we consider the RIPEP. We are given a function $S^+(k)$ ($k \in \mathbf{R}$) in $\mathcal{S}$ and we propose to solve the system S associated with this input function. For $x \geq 0$, let $M_x^\pm$ be the linear operator defined as

$$(M_x^\pm y)(t) = \int_x^\infty s^\pm(t + u)y(u)\, du, \quad t \geq x, \tag{3.11}$$

in the Banach space $L^1(x, \infty)$ of classes of real functions $y(t)$ integrable in $[x, \infty[$, equipped with the norm $\|y\| = \int_x^\infty |y(t)|\, dt$. Let also $\mathcal{M}_x^+$ be the linear operator defined as

$$\mathcal{M}_x^+(y^+, y^-) = (M_x^+y^-, M_x^-y^+), \tag{3.12}$$

in the Banach space $L^1(x, \infty) \times L^1(x, \infty)$ equipped with the norm $\|(y^+, y^-)\| = \max(\|y^+\|, \|y^-\|)$. M_x^+ and M_x^- being compact by standard arguments, $\mathcal{M}_x^+$ is also compact. Now we set the following additional supposition on $S^+(k)$ ($k \in \mathbf{R}$):

(I_3) For any $x \geq 0$, the system of coupled integral equations

$$y^+(t) = \int_x^\infty s^+(t + u)y^-(u)\, du, \quad t \geq x, \tag{3.13}$$

$$y^-(t) = \int_x^\infty s^-(t + u)y^+(u)\, du, \quad t \geq x, \tag{3.14}$$

has the unique solution $(y^+, y^-) = (0, 0)$ in $L^1(x, \infty) \times L^1(x, \infty)$. From the Fredholm alternative theorem, I_3 is equivalent to assume that the operator $I - \mathcal{M}_x^+$ has an inverse in $L^1(x, \infty) \times L^1(x, \infty)$ for any $x \geq 0$ (I is the identity operator). As a consequence, for fixed $F^+(x)$ and $F^-(x)$, the system of coupled Fredholm integral equations (3.6) and (3.7) has a unique solution $(A^+(x, t), A^-(x, t))$ in the space of pairs of functions of (x, t) ($t \geq x \geq 0$) which are, for fixed $x \geq 0$, continuous and integrable in t for $t \geq x$. Let

$(a_1^+(x,t), a_1^-(x,t))$ be the solution of (3.6)–(3.7) corresponding to $F^\pm(x)=1$ and $(a_2^+(x,t), a_2^-(x,t))$ be the solution corresponding to $F^\pm(x)=\mp i$. Equations (3.6) and (3.7) are equivalent to the following relations:

$$A^\pm(x,t)=F^\mp(x)\alpha^\pm(x,t)+F^\pm(x)\beta^\mp(x,t), \quad t\geq x\geq 0, \qquad (3.15)$$

where

$$\alpha^\pm(x,t)=\tfrac{1}{2}[a_1^\pm(x,t)\mp ia_2^\pm(x,t)],$$
$$\beta^\mp(x,t)=\tfrac{1}{2}[a_1^\pm(x,t)\pm ia_2^\pm(x,t)]. \qquad (3.16)$$

The functions $a_1^\pm(x,t)$, $ia_2^\pm(x,t)$, $\alpha^\pm(x,t)$, and $\beta^\pm(x,t)$ are real. Inserting (3.15) in Eq. (3.9), and using (3.8), (3.10), and the condition $y'(\infty)=0$, we find the differential equation

$$y'=-2\alpha^+(x,x)\exp y+2\alpha^-(x,x)\exp(-y)$$
$$+2\beta^+(x,x)-2\beta^-(x,x), \quad x\geq 0, \qquad (3.17)$$

with the condition

$$y(0)=-2\ln F_0^+. \qquad (3.18)$$

We assume that $S^+(k)$ $(k\in\mathbb{R})$ satisfies the additional condition I_4:

(I_4) The differential equation (3.17) with the condition (3.18) admits a bounded solution in $[0,\infty]$.

It is easy to prove that such a solution $y(x)$ is unique. Then constructing $(F^+(x), F^-(x), A^+(x,t), A^-(x,t))$ from $y(x)$ by using formulas (3.10), (3.8), and (3.15) one can conclude that we have thus obtained the unique solution of system S. From (F^+, F^-, A^+, A^-) we define the pair $(U(x), Q(x))$ by the formulas (3.4). Note that

$$Q(x)=\tfrac{1}{2}[iy'(x)], \quad x\geq 0. \qquad (3.19)$$

One can prove that $(U(x), Q(x))$ satisfies the conditions (i), (ii), (iii), and (iv) of the class $\mathcal{V}$ and admits the input function $S^+(k)$ $(k\in\mathbb{R})$ as its scattering matrix. The proof is sketched in Appendix B. It follows from this proof that $y(x)$ can alternatively be defined as the unique solution of (3.17) which satisfies the condition $y(\infty)=0$.

Let us write the following equation which results from the equality of $f^+(0,x)$ and $f^-(0,x)$ and from (3.2):

$$F^+(x)+\int_x^\infty A^+(x,t)\,dt=F^-(x)+\int_x^\infty A^-(x,t)\,dt, \quad x\geq 0. \qquad (3.20)$$

Inserting (3.15) into (3.20) we find

$$F^+(x)\left(1-\int_x^\infty[\alpha^-(x,t)-\beta^-(x,t)]\,dt\right)$$
$$=F^-(x)\left(1-\int_x^\infty[\alpha^+(x,t)-\beta^+(x,t)]\,dt\right), \quad x\geq 0; \qquad (3.21)$$

(3.21) and (3.8) permit to determine $F^+(x)$ for the values of x which do not cancel the second factor of the left-hand side. This is clearly true at least for x sufficiently large. One can prove (see Appendix C) that if the following condition I_4' is satisfied, then I_4 is also satisfied and solution of the system S can be replaced by solution of the system S', obtained from S by replacing Eq. (3.9) by (3.20) and suppressing condition $y(0)=-2\ln F_0^+$:

$$(I_4')\ 1-\int_x^\infty[\alpha^\pm(x,t)-\beta^\pm(x,t)]\,dt>0, \quad x\geq 0. \qquad (3.22)$$

Let us call $\mathcal{S}'$ the class of functions $S^+(k)$ $(k\in\mathbb{R})$ which satisfy the above conditions I_1, I_2, I_3, I_4, and the following conditions I_5 and I_6:

(I_5) $\operatorname{Im}Q(x)\leq 0$ for $x\geq 0$, $Q(x)$ being the potential constructed from the solution of system S.

(I_6) $f^+(0,x)>0$ for $x\geq 0$, $f^+(0,x)$ being the Jost solution constructed from the solution of system S.

We note that as a consequence of I_1 and (2.6) the quantity $\int_0^\infty|\psi^\pm(k,x)|^2\operatorname{Im}Q(x)\,dx$ must be negative for every $k\in\mathbb{R}^*$. So we can expect that I_5 is not a very strong condition on $S^+(k)$. It is clear from what seen above that the RIPEP has a unique solution $(U(x), Q(x))$ in $\mathcal{V}$ when the input function $S^+(k)$ $(k\in\mathbb{R})$ belongs to $\mathcal{S}'$. Unfortunately, I_3, I_4, I_5, and I_6 are very complicated conditions on the data. Besides, I_3 has not been proved to be necessary for a function $S^+(k)$ $(k\in\mathbb{R})$ to be the scattering matrix associated with a pair $(U(x), Q(x))$ in $\mathcal{V}$. We shall see in next paragraphs that the situation can be improved in particular cases. Finally, we remark that the Jost solution $f^+(0,x)$ is more interesting than $U(x)$ in the associated RIPAM [see (2.18)] and that it can be directly obtained from solution of system S by formula (3.2).

B. Case where $S^+(k)$ $(k\in\mathbb{R})$ is close to unity

In Secs. IIIB and IIID we use the following semi-distances d and $\tilde{d}$ in $\mathcal{S}$:

$$d(S_a^+, S_b^+)=\int_0^\infty t\sup(|s_b^{+\prime}(t)-s_a^{+\prime}(t)|, |s_b^{-\prime}(t)-s_a^{-\prime}(t)|)\,dt, \qquad (3.23)$$

$$\tilde{d}(S_a^+, S_b^+)=|(F_{0a}^+)^2-(F_{0b}^+)^2|+d(S_a^+, S_b^+), \quad (S_a^+, S_b^+)\in\mathcal{S}\times\mathcal{S}. \qquad (3.24)$$

Note that $\tilde{d}$ is a distance in $\mathcal{S}'$ because $\tilde{d}(S_a^+, S_b^+)=0$ implies successively the equality of F_{0a}^+ and F_{0b}^+ and of $s_a^\pm(t)$ and $s_b^\pm(t)$ for $t\geq 0$, that of $(U_a(x), Q_a(x))$ and $(U_b(x), Q_b(x))$ (by the inversion procedure), and finally that of S_a^+ and S_b^+. Let us set

$$\sigma_{ab}(x)=\int_x^\infty\sup(|s_b^{+\prime}(t)-s_a^{+\prime}(t)|, |s_b^{-\prime}(t)-s_a^{-\prime}(t)|)\,dt, \quad x\geq 0. \qquad (3.25)$$

$\sigma_{ab}(x)$ $(x\geq 0)$ is continuous, positive, nonincreasing, and approaches zero as $x\to\infty$; $x\sigma_{ab}(x)$ $(x\geq 0)$ is bounded and $\sigma_{ab}(x)$ $(x\geq 0)$ is integrable. $|s_b^\pm(x)-s_a^\pm(x)|$ is bounded by $\sigma_{ab}(x)$ for $x\geq 0$ and we have the equality

$$d(S_a^+, S_b^+)=\int_0^\infty\sigma_{ab}(x)\,dx. \qquad (3.26)$$

Now we are given a function $S^+(k)$ $(k\in\mathbb{R})$ in $\mathcal{S}$. We will show that if $S^+(k)$ $(k\in\mathbb{R})$ is sufficiently close to 1, in the sense that $d(S^+,1)$ is sufficiently small, then conditions I_3, I_4, and I_6 are automatically satisfied. So in this case, we only have to make the additional supposition I_5 to be able to assert that $S^+(k)$ belongs to $\mathcal{S}'$ and therefore that the RIPEP has a unique solution in $\mathcal{V}$. Our proof is based on the reasoning made in Ref. 13, Sec. 6, paragraph 1. First it is clear from this that if $d(S^+,1)<1$, then condition I_3 is satisfied and, for fixed $F^+(x)$ and $F^-(x)$, the solution $(A^+(x,t), A^-(x,t))$ of the system of coupled Fredholm integral equations (3.6) and (3.7) can be obtained by the method of successive approximations.

Furthermore, if $d(S^+, 1) < \frac{1}{2}$, then condition I_4' (which implies condition I_4) is satisfied and $F^+(x)$ can be obtained from (3.21) and (3.8). Finally, it is easy to see that $f^+(0, x)$ can be written in the form

$$f^+(0, x) = F^+(x)[1 + g(x)], \quad x \geq 0, \tag{3.27}$$

where the function $g(x)$ $(x \geq 0)$ can be strictly bounded by 1 when $d(S^+, 1)$ is sufficiently small. Therefore condition I_6 is also satisfied in this case.

We will now give approximation formulas for the inversion procedure when $S^+(k)$ is very close to 1. Let us write $S^+(k)$ in the form

$$S^+(k) = 1 + ia(k) + b(k), \quad k \in \mathbb{R}, \tag{3.28}$$

where $a(k)$ and $b(k)$ are real functions. Using the first order approximation in $a(k)$ and $b(k)$ we find, for $k \in \mathbb{R}$,

$$b(k) = |S^+(k)| - 1, \quad 1 + ia(k) = S^+(k)|S^+(k)|^{-1}. \tag{3.29}$$

On the other hand using (3.1) and (2.5) we obtain the approximate relation

$$s^-(t) = -s^+(-t), \quad t \in \mathbb{R}. \tag{3.30}$$

Equations (3.1), (3.29), and (3.30) easily yield the formulas $(k \in \mathbb{R})$

$$|S^+(k)| = S^+(\infty) + \int_{-\infty}^{\infty} \tfrac{1}{2}(s^+(t) - s^-(t)) \exp(-ikt)\, dt, \tag{3.31}$$

$$S^+(k)|S^+(k)|^{-1} = 1 + \int_{-\infty}^{\infty} \tfrac{1}{2}(s^+(t) + s^-(t)) \exp(-ikt)\, dt. \tag{3.32}$$

Using the inversion procedure to the first-order approximation in s^+ and s^- we find the following formulas:

$$a_1^{\pm}(x, t) = \alpha^{\pm}(x, t) = A^{\pm}(x, t) = s^{\pm}(x, t), \quad t \geq x \geq 0, \tag{3.33}$$

$$a_2^{\pm}(x, t) = \pm i s^{\pm}(x + t), \quad \beta^{\pm}(x, t) = 0, \quad t \geq x \geq 0, \tag{3.34}$$

$$iQ(x) = s^+(2x) - s^-(2x), \quad U(x) = -(s^{+\prime}(2x) + s^{-\prime}(2x)), \quad x \geq 0. \tag{3.35}$$

Using (3.35) and supposing that Fourier inversion formulas can be used in (3.31) and (3.32), we obtain the approximate formulas, for $x \geq 0$,

$$Q(x) = -i\pi^{-1} \int_{-\infty}^{\infty} (|S^+(k)| - S^+(\infty)) \exp(2ikx)\, dx, \tag{3.36}$$

$$U(x) = -\pi^{-1} \frac{d}{dx} \int_{-\infty}^{\infty} (S^+(k)|S^+(k)|^{-1} - 1) \exp(2ikx)\, dk. \tag{3.37}$$

Formulas (3.36) and (3.37) show that when $S^+(k)$ is very close to 1, $Q(x)$ depends only on $|S^+(k)|$ and $U(x)$ depends only on $\arg S^+(k)$. Formulas (3.36) and (3.37) can be obtained in a shorter but less rigorous way by considering the inverse problem for the Born approximation [$U(x)$ and $Q(x)$ small] of the scattering problem. To show this, we start from the exact formulas

$$f^{\pm}(k) = 1 + \int_0^{\infty} k^{-1} \sin kx (U(x) \pm 2kQ(x)) f^{\pm}(k, x)\, dx. \tag{3.38}$$

In the Born approximation we can replace $f^{\pm}(k, x)$ by $\exp(-ikx)$ in (3.38). Using then (2.4) we obtain without difficulty, for $k \in \mathbb{R}$,

$$|S^+(k)| - S^+(\infty) = 2i \int_0^{\infty} Q(x) \cos 2kx\, dx, \tag{3.39}$$

$$S^+(k)|S^+(k)|^{-1} - 1 = -2i \int_0^{\infty} \left(\int_x^{\infty} U(y)\, dy \right) \sin 2kx\, dx. \tag{3.40}$$

Fourier inversion formulas for (3.39) and (3.40) give readily (3.36) and (3.37).

C. Case where there is no absorption

In this paragraph we recover the results of the well-known particular case corresponding to $Q(x) = 0$. Let us call $\mathcal{V}_0$ the set of pairs $(U(x), Q(x))$ such that $Q(x) = 0$ and the conditions (i), (iv), and (vi) of the class $\mathcal{V}$ are satisfied. We know from Sec. IIA that $\mathcal{V}_0$ is a subset of $\mathcal{V}$. The scattering matrix $S^+(k)$ $(k \in \mathbb{R})$ associated with an element of $\mathcal{V}_0$ belongs to the subclass $\mathcal{S}_0$ of $\mathcal{S}$ of functions satisfying the condition

$$|S^+(k)| = 1, \quad k \in \mathbb{R}. \tag{3.41}$$

(3.41) implies that $F_0^+ = 1$ and is equivalent to say that $S^+(k)$ and $S^-(k)$ coincide for $k \in \mathbb{R}$. So we use the same notation $S(k)$ for $S^+(k)$ and $S^-(k)$. Similarly, $s^+(t)$ and $s^-(t)$ in (3.1) will be noted $s(t)$. $\mathcal{S}_0$ will be called the set of elements in $\mathcal{S}$ with no absorption. Conversely, we are given a function $S(k)$ in $\mathcal{S}_0$ and we apply our inversion procedure to this function. Let us show that condition I_3 is automatically satsified. Adding (3.13) and (3.14) we find an homogeneous equation in $[y^+(t) + y^-(t)]$. By using (3.41) we can prove (see for instance Ref. 14, Sec. 3) that this equation has only the trivial solution $y^+(t) + y^-(t) = 0$. Then (3.13) can be transformed in an homogeneous equation in $y^+(t)$. By the same method we can show that $y^+(t) = 0$, which completes the proof. It is easy to see that the functions $a_1^+(x, t)$ and $a_1^-(x, t)$, $a_2^+(x, t)$ and $-a_2^-(x, t)$, $\alpha^+(x, t)$ and $\alpha^-(x, t)$, $\beta^+(x, t)$ and $\beta^-(x, t)$, are respectively equal, so that $a_1^+(x, t)$, $a_2^+(x, t)$, $\alpha^+(x, t)$, and $\beta^+(x, t)$ will be simply noted $a_1(x, t)$, $a_2(x, t)$, $\alpha(x, t)$, and $\beta(x, t)$, respectively. Equations (3.17) and (3.18) become

$$y' = -4\alpha(x, x) \sinh y \quad (x \geq 0), \quad y(0) = 0. \tag{3.42}$$

Equation (3.42) admitting $y(x) = 0$ for solution, I_4 is satisfied and $(U(x), Q(x))$ is given by

$$Q(x) = 0, \quad U(x) = -2 \frac{d}{dx} A(x, x), \quad x \geq 0, \tag{3.43}$$

where $A(x, t)$ $(= a_1(x, t))$ is the solution of the integral equation

$$A(x, t) = s(x + t) + \int_x^{\infty} A(x, u) s(u + t)\, du, \quad t \geq x \geq 0. \tag{3.44}$$

Finally, we can assert that a necessary and sufficient condition for a function $S(k)$ $(k \in \mathbb{R})$ to be the scattering matrix associated with a pair $(U(x), Q(x))$ in $\mathcal{V}_0$ is that $S(k)$ belongs to the subclass $\mathcal{S}_0'$ of $\mathcal{S}_0$ of functions satisfying condition I_6. This pair is the only one in $\mathcal{V}$—and therefore in $\mathcal{V}_0$—admitting this function as its scattering matrix.

D. Case where the absorption is weak

Let $S^+(k)$ $(k \in \mathbb{R})$ be a function in $\mathcal{S}$. We will show that if the absorption is sufficiently weak, in the sense that there exists a function $S(k)$ $(k \in \mathbb{R})$ in $\mathcal{S}_0$ with $\tilde{d}(S^+, S)$

sufficiently small, then conditions I_3 and I_4 are automatically satisfied [in practice one can try to show that $S^+(k)|S^+(k)|^{-1}$ belongs to $\mathcal{S}_0$ and is sufficiently close to $S^+(k)$]. Furthermore, if $\mathcal{S}_0$ is replaced by $\mathcal{S}_0'$ in this statement, we will see that condition I_6 is also satisfied. So in this last case we only have to make the additional supposition I_5 to be able to assert that the RIPEP has a unique solution in $\mathcal{V}$.

Before proving this statement we make the following remark which ensures its usefulness in non trivial cases, i.e., cases with non-null absorption: $S(k)$ being any element in $\mathcal{S}_0$, any neighborhood of $S(k)$ in $\mathcal{S}$ (equipped with the semidistance $\tilde{d}$) contains at least one element which is not in $\mathcal{S}_0$. To show this, it is sufficient to consider the function $S_\epsilon^+(k)$ defined as

$$S_\epsilon^+(k)=S(k)T_\epsilon^+(k), \quad k\in\mathbf{R}, \quad \epsilon>0, \tag{3.45}$$

$$T_\epsilon^+(k)=ba^{-1}(k-ia)(k-ib)^{-1}, \quad b=a(1-\epsilon), \tag{3.46}$$

where a is a fixed positive number. Using the fact that $T_\epsilon^+(k)$ belongs to $\mathcal{S}$ and the relation

$$T_\epsilon^\pm(k)=b^{\pm1}a^{\mp1}+\int_0^{\pm\infty}b^{\pm1}a^{\mp1}(a-b)\exp\binom{-b}{a}t \times\exp(-ikt)\,dt, \quad k\in\mathbf{R}, \tag{3.47}$$

it is a straightforward matter to verify that $S_\epsilon^+(k)$ belongs to $\mathcal{S}$ and does not satisfy (3.41), and that $\tilde{d}(S,S_\epsilon^+)\to0$ as $\epsilon\to0$.

Now we prove our statement. Let $S(k)$ in $\mathcal{S}_0$ and $S^+(k)$ in $\mathcal{S}$ be given. It is convenient to suppose that $S(k)$ is fixed. Since $\|(I-\mathcal{M}_x)^{-1}\|$ is a continuous function of x and $\|\mathcal{M}_x\|\to0$ as $x\to\infty$ the quantity $\sup\|(I-\mathcal{M}_x)^{-1}\|$ $(x\geq0)$ is finite. Let us call C the inverse quantity. Let us show that if $d(S^+,S)<C$, I_3 is satisfied. Indeed this inequality, the following ones [where $\sigma(x)$ is the function defined in (3.25) for $S_a^+=S^+$ and $S_b^+=S$]:

$$\|\mathcal{M}_x^+-\mathcal{M}_x\|\leq\sup(\|M_x^+-M_x\|, \ \|M_x^--M_x\|)\leq\int_{2x}^\infty\sigma(t)\,dt, \quad x\geq0, \tag{3.48}$$

and (3.26) imply that $\|\mathcal{M}_x^+-\mathcal{M}_x\|$ is bounded by $\|(I-\mathcal{M}_x)^{-1}\|^{-1}$ for $x\geq0$. Therefore, the inverse $(I-\mathcal{M}_x^+)^{-1}$ exists for $x\geq0$. Now we derive some bounds in order to investigate condition I_4. Starting from the inequality

$$\|(I-\mathcal{M}_x^+)^{-1}-(I-\mathcal{M}_x)^{-1}\|\leq\|(I-\mathcal{M}_x)^{-1}\|^2\|\mathcal{M}_x^+-\mathcal{M}_x\| \times(1-\|(I-\mathcal{M}_x)^{-1}\|\|\mathcal{M}_x^+-\mathcal{M}_x\|)^{-1}, \tag{3.49}$$

we obtain the bound

$$\|(I-\mathcal{M}_x^+)^{-1}-(I-\mathcal{M}_x)^{-1}\|\leq C^{-1}[C-d(S^+,S)]^{-1}\int_{2x}^\infty\sigma(t)\,dt. \tag{3.50}$$

Noting that the functions $\|a_1^\pm(x,t)-a_1(x,t)\|$ and $\|a_2\pm(x,t)\mp a_2(x,t)\|$ $(x\geq0)$ are bounded by

$$\|(I-\mathcal{M}_x^+)^{-1}-(I-\mathcal{M}_x)^{-1}\|\max\begin{pmatrix}\|s^+(x+t)\|\\ \|s^-(x+t)\|\end{pmatrix} +\|(1-\mathcal{M}_x)^{-1}\|\max\begin{pmatrix}\|s^+(x+t)-s(x+t)\|\\ \|s^-(x+t)-s(x+t)\|\end{pmatrix},$$

we deduce that they are bounded by

$$C^{-1}[C+d(S,1)][C-d(S^+,S)]^{-1}\int_{2x}^\infty\sigma(t)\,dt.$$

It is then easy to see that the function $|a_1^\pm(x,x)-a_1(x,x)|$, $|a_2^\pm(x,x)\mp a_2(x,x)|$, $|a_1^+(x,x)-a_1^-(x,x)|$, $|a_2^+(x,x)+a_2^-(x,x)|$, $|\alpha^+(x,x)-\alpha^-(x,x)|$, $|\beta^+(x,x)-\beta^-(x,x)|$, and $f(x,-2\ln F_0^+)$, where $f(x,y)$ is the right-hand side in Eq. (3.17), are bounded for $x\geq0$ by a function $\sigma_1(x)$ which is continuous, positive, nonincreasing, integrable and approaches zero as $x\to\infty$. $\sigma_1(x)$ is also such that $d_1=\int_0^\infty\sigma_1(t)\,dt$ is a rational function of $\tilde{d}(S^+,S)$ which approaches zero as $\tilde{d}(S^+,S)\to0$. Similarly, $4|\alpha^\pm(x,x)|$ is bounded by a function $\sigma_2(x)$ with the same properties as $\sigma_1(x)$ except that $d_2=\int_0^\infty\sigma_2(t)\,dt$ is a rational function of $\tilde{d}(S^+,S)$ which approaches a constant as $\tilde{d}(S^+,S)\to0$. We have the following bound for $(y_1,y_2)\in\mathbf{R}^2$:

$$|f(x,y_2)-f(x,y_1)|\leq\sigma_2(x)|y_2-y_1|\max(\exp|y_2|,\exp|y_1|). \tag{3.51}$$

Let us consider the sequence $y_n(x)$ defined as

$$y_0(x)=-2\ln F_0^+, \quad y_n(x)=y_0(x)+\int_0^x f(t,y_{n-1}(t))\,dt, \quad n\geq1. \tag{3.52}$$

For $n\geq1$, $x\geq0$, we can prove that $|y_n(x)-y_{n-1}(x)|$ is bounded by v_n, v_n being defined as

$$v_1=d_1, \quad v_n=d_1\exp[(n-1)v_1+(n-2)v_2+\cdots+v_{n-1}] \times\frac{[(F_0^-)^2d_2]^{n-1}}{(n-1)!}, \quad n\geq2. \tag{3.53}$$

To investigate the convergence of the series v_n we consider the quantity $A_n=v_{n+1}[v_n]^{-1}$ $(n\geq1)$. We have

$$A_n=A_{n-1}(n-1)n^{-1}\exp(d_1A_1A_2\cdots A_{n-1}), \quad n\geq2. \tag{3.54}$$

Noticing that there exists $\epsilon>0$ such that, for $\tilde{d}(S^+,S)<\epsilon$ and $n\geq2$, we have

$$\exp[d_1(2A_1)^{n-1}(n!)^{-1}]\leq n^2(n^2-1)^{-1}, \tag{3.55}$$

one can prove by induction that A_n obeys the following bound:

$$A_n\leq A_{n-1}n(n+1)^{-1}, \quad n\geq2. \tag{3.56}$$

Equation (3.56) shows that A_n is bounded by $2A_1(n+1)^{-1}$ and therefore approaches zero as $n\to\infty$. For $\tilde{d}(S^+,S)<\epsilon$ the series v_n is therefore convergent and $y_n(x)$ converges to a bounded function $y(x)$ which is the solution of Eq. (3.17) with condition (3.18). Therefore I_4 is satisfied. Since $\|\alpha^\pm(x,t)-\alpha(x,t)\|$ and $\|\beta^\pm(x,t)-\beta(x,t)\|$ approaches zero uniformly in $x\geq0$ as $\tilde{d}(S^+,S)\to0$, it is clear that $f^+(0,x)\to f(0,x)$ uniformly in $x\geq0$ as $\tilde{d}(S^+,S)\to0$. Then, if $S(k)$ belongs to $\mathcal{S}_0'$, it is easy to conclude that I_6 is satisfied when $\tilde{d}(S^+,S)$ is sufficiently small.

E. Class of exactly solvable examples

We are given the function

$$S^+(k)=\prod_{p=1}^n b_pa_p^{-1}(k-ia_p)(k-ib_p)^{-1}, \quad k\in\mathbf{R}, \tag{3.57}$$

where a_p and b_p are strictly positive numbers such that $a_p>b_p>a_{p+1}$. $S^+(k)$ belongs to the class $\mathcal{S}$ with $s^+(t)$ $(t<0)$ and $s^-(t)$ $(t>0)$ equal to zero and

$$s^+(t)=\sum_{r=1}^{n}\alpha_r\exp(-b_r t),\quad t\geq 0, \tag{3.58}$$

$$\alpha_r=-\left(\prod_{p=1}^{n}b_p a_p^{-1}\right)(b_r-a_r)\prod_{p\neq r}\frac{(b_r-a_p)}{(b_r-b_p)}. \tag{3.59}$$

(b_r-a_p) and (b_r-b_p) having the same sign, α_r is positive and therefore $s^+(t)$ is positive. Condition I_3 is trivially satisfied and we have

$$\alpha^+(x,t)=s^+(x+t),\quad \alpha^-(x,t)=\beta^{\mp}(x,t)=0,\quad t\geq x\geq 0. \tag{3.60}$$

On the other hand, making $k=0$ in Eq. (3.1), we see that the quantity $1-\int_0^\infty s^+(t)\,dt$ is positive. We conclude that I_4' (and therefore also I_4) is satisfied. The solution of system S and the pair $(U(x), Q(x))$ are given by the formulas

$$F^{\pm}(x)=\left(1-\sum_{r=1}^{n}\alpha_r b_r^{-1}\exp(-2b_r x)\right)^{\pm 1/2},$$

$$A^+(x,t)=F^-(x)s^+(x,t),\quad A^-(x,t)=0, \tag{3.61}$$

$$Q(x)=-is^+(2x)[F^-(x)]^2,$$

$$U(x)=-2s^{+\prime}(2x)[F^-(x)]^2+3[s^+(2x)]^2[F^-(x)]^4. \tag{3.62}$$

Let us also give the formulas

$$f^+(k)=F^-(0)S^+(k),\quad f^-(k)=F^-(0),\quad \mathrm{Im}k\leq 0, \tag{3.63}$$

$$f^{\pm}(0,x)=F^-(x),\quad x\geq 0. \tag{3.64}$$

Formulas (3.62) and (3.64) show that I_5 and I_6 are satisfied. Therefore, $S^+(k)$ belongs to $\mathcal{S}'$ and $(U(x), Q(x))$ given by (3.62) is the unique solution of the RIPEP in the class $\mathcal{V}$ (note that it is easy to verify directly that it is a solution.

IV. ONE-DIMENSIONAL INVERSE PROBLEMS IN ABSORBING MEDIA

In this section we indicate briefly how one-dimensional inverse problems in absorbing media (called OIPAM) can be solved exactly in the same way as the RIPAM, the analogous radial cases. First by means of the Liouville transformation we reduce the OIPAM to an inverse scattering problem (called OIPEP) for the one-dimensional Schrödinger Eq. (1.1) ($x\in\mathbb{R}$) with the energy-dependent potential (1.2). As in the analogous reduction of the RIPAM to the RIPEP we find that $U(x)$ must be real and $Q(x)$ purely imaginary. Making use of our work on the OIPEP for real $U(x)$ and $Q(x)$ in Ref. 16 and of our work on the RIPEP for real $U(x)$ and purely imaginary $Q(x)$ in this paper, it is straightforward to obtain the solution of the OIPEP for real $U(x)$ and purely imaginary $Q(x)$.

The common input data of the OIPEP and the OIPAM is a 2×2 matrix-valued function

$$S^+(k)=\begin{pmatrix} s^+_{11}(k) & s^+_{21}(k)\\ s^+_{12}(k) & s^+_{22}(k)\end{pmatrix}\quad (k\in\mathbb{R}),$$

which represents the "scattering matrix" in the OIPEP. Each function $s^+_{ij}(k)$ $(k\in\mathbb{R})$ $(i=1,2;\ j=1,2)$ is completely determined by its physical part $s^+_{ij}(k)$ $(k>0)$ because of the relation $\overline{s^+_{ij}(k)}=s^+_{ij}(-k)$, and has analogous physical meanings in the OIPEP and the OIPAM. $s^+_{21}(k)$ $(k\in\mathbb{R})$ is the reflection coefficient to the right. $s^+_{12}(k)$ $(k\in\mathbb{R})$ is the reflection coefficient to the left. $s^+_{11}(k)[=s^+_{22}(k)]$ $(k\in\mathbb{R})$ is the transmission coefficient. $\mathcal{V}_1$ being a large class of pairs $(U(x), Q(x))$ with real $U(x)$ and purely imaginary $Q(x)$ we can prove that the scattering matrix associated with a pair $(U(x), Q(x))$ in $\mathcal{V}_1$ belongs to a certain class $\mathcal{S}_1$ of 2×2 matrix-valued functions $S^+(k)$ $(k\in\mathbb{R})$. Note that, as in the radial case, it is useful in order to define this class to introduce the function

$$S^-(k)=\begin{pmatrix} s^-_{11}(k) & s^-_{21}(k)\\ s^-_{12}(k) & s^-_{22}(k)\end{pmatrix}=[{}^tS^+(-k)]^{-1}\quad (k\in\mathbb{R}),$$

where t means transposed.

We can prove that the OIPEP has a unique solution in $\mathcal{V}_1$ if the input function $S^+(k)$ $(k\in\mathbb{R})$ belongs to a subclass of $\mathcal{S}_1$, $\mathcal{S}_1'$, consisting of functions satisfying certain complicated conditions. On the one hand, these conditions ensure that a certain system S_1 of equations (analogous to the system S in the radial case), with data $s^+_{21}(k)$ $(k\in\mathbb{R})$ and $s^-_{21}(k)$ $(k\in\mathbb{R})$, admits a unique solution $(F_1^+(x), F_1^-(x), A_1^+(x,t), A_1^-(x,t))$. On the other hand, they ensure that another and similar system S_2 of equations, with data $s^+_{12}(k)$ $(k\in\mathbb{R})$ and $s^-_{12}(k)$ $(k\in\mathbb{R})$, admits a unique solution $(F_2^+(x), F_2^-(x), A_2^+(x,t), A_2^-(x,t))$. The pair $(U(x), Q(x))$, the solution of the OIPEP, can be obtained from the solution of S_1 or S_2 by formulas similar to Eq. (1.9). Note that for the one-dimensional Schrödinger equation with the potential $V^{\pm}(k,x)=U(x)\pm 2kQ(x)$, the Jost solution at ∞ $f_1^{\pm}(k,x)$—i.e., the solution whose asymptotic behavior as $x\to\infty$ is $\exp(-ikx)$—and the Jost solution at $-\infty$ $f_2^{\pm}(k,x)$—i.e., the solution whose asymptotic behavior as $x\to-\infty$ is $\exp(ikx)$—are given by

$$f_1^{\pm}(k,x)=F_1^{\pm}(x)\exp(-ikx)+\int_x^\infty A_1^{\pm}(x,t)\exp(-ikt)\,dt,$$
$$x\in\mathbb{R},\quad \mathrm{Im}k\leq 0, \tag{4.1}$$

$$f_2^{\pm}(k,x)=F_2^{\pm}(x)\exp(ikx)+\int_{-\infty}^x A_2^{\pm}(x,t)\exp(ikt)\,dt,$$
$$x\in\mathbb{R},\quad \mathrm{Im}k\leq 0. \tag{4.2}$$

Though we need to solve only one of the systems S_1 or S_2 to obtain $(U(x), Q(x))$, we stress that we also need to know that the other system has a unique solution in order to be able to assert that the pair thus obtained is the solution of the OIPEP.

We remark that in contrast to the case where $U(x)$ and $Q(x)$ are real we can avoid any ambiguity in the existence question of the OIPEP. This is due to the fact that now $F_1^+(x)$ and $F_2^+(x)$ are known to be positive functions. [Recall that for real $U(x)$ and $Q(x)$ we find, in general, that the pair $(U(x), Q(x))$ obtained from S_1 and S_2 admits for scattering matrix either the input function $S^+(k)$ $(k\in\mathbb{R})$ or the function obtained from it by changing the sign of the diagonal coefficients.]

As a striking difference with the case where $U(x)$ and $Q(x)$ are real, and therefore with the case where $U(x)$ is real and $Q(x)=0$, we find, in general, that the scattering matrix $S^+(k)$ $(k\in\mathbb{R})$ associated with a pair in $\mathcal{V}_1$ is not determined by the reflection coefficient to the right $s^+_{21}(k)$ $(k\in\mathbb{R})$. [In general, we do not have the relation $s^-_{21}(k)=\overline{s^+_{21}(-k)}$ as for real $U(x)$ and $Q(x)$ and therefore we can only assert that if the system S_1 admits a unique solution, then $S^+(k)$ ($k\in\mathbf{R}$) is determined from $s^+_{21}(k)$ ($k\in\mathbf{R}$) and $s^-_{21}(k)$ ($k\in\mathbf{R}$).] Nevertheless, note that for real $U(x)$ and $Q(x)$ we do not have the relation $\overline{s^+_{21}(k)} = s^+_{21}(-k)$ [except of course if $Q(x)=0$] and so the part $s^+_{21}(k)$ ($k<0$) has no physical meaning.

ACKNOWLEDGMENT

I would like to thank Professor P.C. Sabatier for fruitful discussions.

APPENDIX A

We prove our last assertion in Sec. IIA. We suppose that $f^-(k)$ has a zero $k_0(\mathrm{Im}k_0<0)$ and we start from the equality $(x\geq 0)$

$$\frac{d}{dx}[\phi^-(k_0,x)\,\overline{\phi^{-\prime}(k_0,x)}-\phi^{-\prime}(k_0,x)\overline{\phi^-(k_0,x)}\,]$$
$$=2i|\phi^-(k_0,x)|^2[\mathrm{Im}(k_0^2)+2\,\mathrm{Im}(k_0Q(x))]. \qquad \text{(A1)}$$

Integrating both sides of (A1) from 0 to ∞ we easily find

$$(\mathrm{Re}k_0)\int_0^\infty|\phi^-(k_0,x)|^2[\mathrm{Im}k_0+\mathrm{Im}Q(x)]\,dx=0. \qquad \text{(A2)}$$

Hence, $\mathrm{Re}k_0=0$. So k_0 can be written in the form $k_0=ib_0$ $(b_0<0)$. Note that up to now we have not used (vi). Now we start from the equality $(x\geq 0)$

$$\frac{d}{dx}[f^-(0,x)f^{-\prime}(ib_0,x)-f^-(ib_0,x)f^{-\prime}(0,x)]$$
$$=f^-(0,x)f^-(ib_0,x)[b_0^2-2ib_0Q(x)]. \qquad \text{(A3)}$$

Noting that the continuous function $f^-(ib_0,x)$ $(x\geq 0)$ is real, that it admits a zero for $x=0$, and that it is positive for large enough x since it behaves as $\exp(bx)$ as $x\to\infty$, we conclude that there exists a number $x_0\in\mathbf{R}^+$ such that $f^-(ib_0,x_0)=0$ and $f^-(ib_0,x)>0$ for $x>x_0$, and therefore such that $f^{-\prime}(ib_0,x_0)\geq 0$. Integrating both sides of (A3) from x_0 to ∞ we find

$$-f^-(0,x_0)f^{-\prime}(ib_0,x_0)=\int_{x_0}^\infty f^-(0,x)f^-(ib_0,x)$$
$$\times[b_0^2-2ib_0Q(x)]\,dx. \qquad \text{(A4)}$$

The left-hand side is negative and the right-hand side is strictly positive. Therefore, $f^-(k)$ cannot have zero for $\mathrm{Im}k<0$.

APPENDIX B

We sketch the proof that the pair $(U(x), Q(x))$ constructed in Sec. IIIA from solution of S solves the RIPEP except for conditions (v) and (vi) of the class $\mathcal{V}$. First we can prove that $(U(x), Q(x))$ satisfies the conditions (i), (ii), and (iv) of the class $\mathcal{V}$ and that the pair $(a^+(x,t), a^-(x,t))$ defined from (F^+, F^-, A^+, A^-) by

$$a^\pm(x,t)=\frac{\partial^2A^\pm}{\partial x^2}(x,t)-\frac{\partial^2A^\pm}{\partial t^2}(x,t)$$
$$\pm 2iQ(x)\frac{\partial A^\pm}{\partial t}(x,t), \quad t\geq x\geq 0, \qquad \text{(B1)}$$

is the solution of the equations obtained by replacing $F^\pm(x)$ by $f^\pm(x)$ in (3.6) and (3.7). Therefore, we have

$$a^\pm(x,t)=f^\mp(x)\alpha^\pm(x,t)+f^\pm(x)\beta^\mp(x,t), \quad t\geq x\geq 0. \qquad \text{(B2)}$$

Recalling (3.4) and (3.15) we find the partial differential equation

$$a^\pm(x,t)=U(x)A^\pm(x,t), \quad t\geq x\geq 0. \qquad \text{(B3)}$$

Equations (B3) and (3.4) show that the function $f^\pm(k,x)$ defined from (F^+, F^-, A^+, A^-) by formula (3.2) is solution of Eq. (2.1). The limits $y(\infty)$ and $F^\pm(\infty)$ exist. Therefore, $F^\mp(\infty)f^\pm(k,x)$ is the Jost solution of (2.1). Writing the equality of the Jost solutions of (2.1) (+) and (2.1) (−) for $k=0$, we obtain

$$F^-(\infty)f^+(0,x)=F^+(\infty)f^-(0,x). \qquad \text{(B4)}$$

On the other hand, from the fact that $S^+(k)$ ($k\in\mathbf{R}$) satisfies condition I_1, one can prove by arguments similar to those given in Ref. 14, Sec. 2, that there exist two functions $f^+(k)$ and $f^-(k)$ ($\mathrm{Im}k\leq 0$) having no zero for $\mathrm{Im}k\leq 0$, which satisfy the relation (2.4) and can be written in the form

$$f^\pm(k)=F_0^\pm+\int_0^\infty b^\pm(t)\exp(-ikt)\,dt, \quad \mathrm{Im}k\leq 0, \qquad \text{(B5)}$$

where $b^\pm(t)$ is a real integrable function. By using properties of Fourier transforms and relations (2.4), (3.1), and (B5) one can easily show that $(b^+(t), b^-(t))$ is solution of the equations deduced from (3.6) and (3.7) by putting $x=0$ and replacing $F^\pm(0)$ by $F_0^\pm$. The equality of $F^\pm(0)$ and $F_0^\pm$, which follows from (3.10) and (3.18), implies the equality of $A^\pm(0,t)$ and $b^\pm(t)$, and therefore

$$f^\pm(k,0)=f^\pm(k), \quad \mathrm{Im}k\leq 0. \qquad \text{(B6)}$$

The assumption $S^+(0)=1$ and the equality (2.4) show that

$$f^+(0)=f^-(0). \qquad \text{(B7)}$$

It follows from (B4), (B6), and (B7) that $F^\pm(\infty)=1$, so that $f^\pm(k,x)$ is the Jost solution of (2.1), $f^\pm(k)$ is the Jost function, and the input function $S^+(k)$ ($k\in\mathbf{R}$) is the scattering matrix of the constructed pair $(U(x), Q(x))$ which satisfies the conditions (i), (ii), (iii), and (iv) of the class $\mathcal{V}$.

APPENDIX C

We prove that condition I_4' implies condition I_4. Indeed, if we assume I_4', it is clear from (3.21) and (3.8) that the system S' has a unique solution (F^+, F^-, A^+, A^-). It is easy to prove that (f^+, f^-, a^+, a^-) defined from (F^+, F^-, A^+, A^-) by (3.5), (B1), and (3.4), is a solution of Eqs. (3.6), (3.7), and (3.20), so that (B2) is true, the same as the equation obtained by replacing $F^+(x)$ by $f^+(x)$ in (3.21). We conclude that Eq. (3.9) is satisfied. Equation (3.17) with condition $y(\infty)=0$ is also satisfied for $y(x)$. Starting from (B7) and (B5) it is possible to prove that (3.21) holds for $x=0$ with $F^+(0)$ replaced by F_0^+. Hence the equality of $F^+(0)$ and F_0^+ and the equality (3.18). So we have proved that I_4 is satisfied and that the solutions of S' and S coincide.

*Work done as part of the program "Recherche Coopérative sur Programme n°264: Etude interdisciplinaire des problèmes inverses."

†Physique Mathématique et Théorique, Equipe de recherche associée au CNRS n° 154.

[1]R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1 (Interscience, New York, 1953), p. 292.

[2]I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk S.S.S.R. 15, 309 (1951); [Am. Math. Soc. Transl. 1, 253 (1955)].
[3]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[4]I. Kay, Research Report No. EM-74-NY-University, Inst. Math. Sci., Electromagnetic Research, 1955 (unpublished).
[5]I. Kay and H. E. Moses, Nuovo Cimento 3, 276 (1956).
[6]L. D. Faddeev, Trudy Mat. Inst. Steklov 73, 314 (1964) [Am. Math. Soc. Transl. 2 65, 139 (1967)].
[7]I. Kay, in Proceeding of a Workshop held at Ames Research Center Moffett Field, Calif. July 12–16, 1971, NASA TM X-62, 150 (unpublished).
[8]H. E. Moses and C. M. DeRidder, MIT Lincoln Lab. Tech. Rept. No. 322, 1–47, 1963 (unpublished).
[9]J. A. Ware and K. Aki, J. Acoust. Soc. Am. 45, 911 (1969).
[10]M. Jaulent, Report of the meeting Etudes interdisciplinaires des problèmes inverses, Montpellier 7 et 8 décembre 1972, 97–106 (unpublished).
[11]V. H. Weston, J. Math. Phys. 13, 1952 (1972).
[12]V. H. Weston and R. J. Krueger, J. Math. Phys. 14, 406 (1973); 14, 1296 (1973).
[13]M. Jaulent and C. Jean, Commun. Math. Phys. 28, 177 (1972).
[14]M. Jaulent, Ann. Inst. Henri Poincaré 17, 363 (1972).
[15]M. Jaulent, C.R. Acad. Sci. Paris 280, 1467 (1975).
[16]M. Jaulent and C. Jean, Ann. Inst. Henri Poincaré (to be published).

Addendum: Analysis of the dispersion of low frequency uniaxial waves in heterogeneous periodic elastic media [J. Math. Phys. 16, 1383 (1975)]

George N. Balanis

Applied Theory, Inc., 1010 Westwood Blvd., Los Angeles, California 90024
(Received 20 April 1976)

Due to a regrettable oversight the author neglected to acknowledge his appreciation for the helpful comments and suggestions provided by the reviewer, Professor George Gavalas of the California Institute of Technology, and Dr. Michael El Raheb of Tetra Tech., Inc.

[2]I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk S.S.S.R. 15, 309 (1951); [Am. Math. Soc. Transl. 1, 253 (1955)].
[3]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[4]I. Kay, Research Report No. EM-74-NY-University, Inst. Math. Sci., Electromagnetic Research, 1955 (unpublished).
[5]I. Kay and H. E. Moses, Nuovo Cimento 3, 276 (1956).
[6]L. D. Faddeev, Trudy Mat. Inst. Steklov 73, 314 (1964) [Am. Math. Soc. Transl. 2 65, 139 (1967)].
[7]I. Kay, in Proceeding of a Workshop held at Ames Research Center Moffett Field, Calif. July 12–16, 1971, NASA TM X-62, 150 (unpublished).
[8]H. E. Moses and C. M. DeRidder, MIT Lincoln Lab. Tech. Rept. No. 322, 1–47, 1963 (unpublished).
[9]J. A. Ware and K. Aki, J. Acoust. Soc. Am. 45, 911 (1969).
[10]M. Jaulent, Report of the meeting Etudes interdisciplinaires des problèmes inverses, Montpellier 7 et 8 décembre 1972, 97–106 (unpublished).
[11]V. H. Weston, J. Math. Phys. 13, 1952 (1972).
[12]V. H. Weston and R. J. Krueger, J. Math. Phys. 14, 406 (1973); 14, 1296 (1973).
[13]M. Jaulent and C. Jean, Commun. Math. Phys. 28, 177 (1972).
[14]M. Jaulent, Ann. Inst. Henri Poincaré 17, 363 (1972).
[15]M. Jaulent, C.R. Acad. Sci. Paris 280, 1467 (1975).
[16]M. Jaulent and C. Jean, Ann. Inst. Henri Poincaré (to be published).

Addendum: Analysis of the dispersion of low frequency uniaxial waves in heterogeneous periodic elastic media [J. Math. Phys. 16, 1383 (1975)]

George N. Balanis

Applied Theory, Inc., 1010 Westwood Blvd., Los Angeles, California 90024
(Received 20 April 1976)

Due to a regrettable oversight the author neglected to acknowledge his appreciation for the helpful comments and suggestions provided by the reviewer, Professor George Gavalas of the California Institute of Technology, and Dr. Michael El Raheb of Tetra Tech., Inc.

ERRATA

Erratum: Scattering by singular potentials with a perturbation—theoretical introduction to Mathieu functions [J. Math. Phys. 16, 961 (1975)]

H. H. Aly

Department of Physics, Southern Illinois University at Edwardsville, Edwardsville, Illinois

H. J. W. Müller-Kirsten

Fachbereich Physik, Universität Trier-Kaiserslautern, Kaiserslautern, Germany

N. Vahedi-Faridi

Tehran University, Tehran, Iran
(Received 26 May 1975)

The byline address for the third author is incorrect in the article and should be as above.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) after the page number indicates an Erratum.

