Indecomposable representations of the conformal group: A nonsingular photon-Weyl graviton system

P. Furlan
International School for Advanced Studies (ISAS), Dipartimento di Fisica Teorica dell'Università, Trieste, Italy and Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Trieste, Trieste, Italy

V. B. Petkova[a)]
Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Trieste, Italy

G. M. Sotkov[a)]
International School for Advanced Studies (ISAS), Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Trieste, Trieste, Italy

(Received 12 March 1986; accepted for publication 27 August 1986)

A large class of nonelementary indecomposable representations of $G = SU(2,2)$ is constructed and the invariant integral operators are found. An example describing a photon and a linear Weyl graviton field along with some auxiliary fields is studied. A nonsingular Lagrangian for the system is given. The pure, linear Weyl gravity with a conformal invariant gauge fixing condition arises as a particular case.

I. INTRODUCTION

The use of indecomposable representations[1,2] of the conformal group has proved to be indispensable in the efforts to build conformally invariant gauge models such as the massless spinor electrodynamics.[3–11] The main idea was to extend the conformal invariance of the physical sector to the full indefinite metric space needed for a local and Poincaré-covariant formulation of a quantum gauge theory (see, e.g., Ref. 12). Such a formulation allows one to implement the usual consequences of the conformal symmetry like completely known, two- and three-point functions, etc. (see, e.g., Refs. 5 and 13 for an introduction to conformal invariance in QFT).

In the present paper we build a large class of *nonelementary* indecomposable representations of $G = SU(2,2)$. Nonelementary refers to the fact that they are induced by finite-dimensional reducible (indecomposable)—and hence nontrivial—representations of the nilpotent special conformal transformations subgroup N of G. The indecomposable *elementary* representations, induced by finite-dimensional irreducible representations of the maximal parabolic subgroup P of G, have been thorougly studied in the Euclidean[14] and Minkowski[15] cases and their structure was the essential ingredient of the physical applications.[5,7,16] Nevertheless, for the purposes of the covariant gauge models building, especially in the quantum case, they seem to be not enough suitable (see a discussion in Ref. 11).

The representations of $G = SU(2,2)$ built here (Sec. II), are the group analogs of some algeba representations (type Ib) of Ref. 2. The "fields" here are actually (finite) *multiplets* of the ordinary Poincaré-covariant fields grouped together by the conformal symmetry. After giving their general construction we show how the procedure of building the (integral) Knapp–Stein[17] intertwining operators can be generalized to this case. Their knowledge is important since they correspond to invariant two-point functions in the physical applications and provide as well invariant local action terms. Some of the multiplets in Sec. II are realized equivalently as the odd and even parts of "supermultiplets" related to the extended superconformal $SU(2,2/N)$-multiplets after reduction to $SU(2,2)$.

In Sec. III we apply our general results to a particular example which describes a multiplet of spin-1 (photon) and spin-2 (Weyl graviton) fields together with some auxiliary pure gauge partners as well as the corresponding currents multiplet. A nonsingular Lagrangian giving a G-invariant action for the system is written down. In a limiting case it provides a nonsingular Lagrangian (equivalently G-invariant gauge fixing conditions) for the linearized Weyl gravity. Unlike a previous realization[16] of this model based on elementary indecomposable representations, the action is invariant under local transformations of the fields. Other possible interpretations of the mixed model are given.

Finally in Sec. IV we relate our construction on one example of Sec. III to the manifestly covariant six-dimensional formalism,[2,18] where many of the formulas look simpler. The Minkowski space picture is reproduced by a standard reduction procedure.

We point out that the various realizations of the nonelementary "multiplet" representations provide us a natural method to introduce in the conformal invariant gauge theories the minimal set of auxiliary fields which are absolutely needed for the "off shell" description of both local and conformal (and Poincaré) invariant gauge conditions. The same problem arises in the string field theories[19]: to construct a local and reparametrization invariant string field action the infinite set of the so-called Stueckelberg auxiliary fields should be introduced. We believe that the construction of nonelementary representations given in the present paper can be extended to the case of reparametrization group and might be useful in the building and understanding of the string field theories.

[a)] On leave of absence from the Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia 1184, Bulgaria.

 0022-2488/87/020251-07$02.50 © 1987 American Institute of Physics

II. NONELEMENTARY REPRESENTATIONS OF SU(2,2): X-SPACE REALIZATION. KNAPP–STEIN INTERTWINING OPERATORS

(1) Let $\varphi(q;z,z^+)$ be a polynomial of maximal degree k in $q\in\mathbb{R}^4$ and a homogeneous polynomial of degrees $2j_1$, $2j_2$ in z, $z^+\in\mathbb{C}^2$, respectively. The space V_χ of polynomials $\varphi(q;z,z^+)$ carries a representation D_χ of the subgroup $P=\gamma AMN\simeq N\gamma AM\subset G(\gamma\simeq Z_4/Z_2,\ M\simeq \mathrm{SL}(2,\mathbb{C}),\ A,N$-subgroups of dilations and special conformal transformations) labeled by

$$\chi=(d;j_1,j_2;\lambda;k),\quad d \text{ real},\quad \lambda\in[0,2)\ (\mathrm{mod}\ 2),$$
$$\lambda=j_2-j_1\ (\mathrm{mod}\ 1), \tag{2.1}$$

and defined according to

$$(D_\chi(\gamma man)\varphi)(q;z,z^+)$$
$$=\rho^{-d}\exp(-i\pi\lambda\mathcal{N}_0)$$
$$\times\varphi(\rho l^{-1}\underline{q}(l^+)^{-1}-\underline{n};zl,l^+z^+). \tag{2.2}$$

Here[15] $\underline{q}=q^\mu\sigma_\mu$,

$$\gamma man=i^{\mathcal{N}_0}\begin{pmatrix} l & 0\\ 0 & (l^+)^{-1}\end{pmatrix}\begin{pmatrix}\sqrt{\rho} & 0\\ 0 & 1/\sqrt{\rho}\end{pmatrix}\begin{pmatrix}\mathbb{1}_2 & 0\\ i\tilde{n} & \mathbb{1}\end{pmatrix}$$
$$\in\gamma MAN; \tag{2.3}$$

$\mathcal{N}_0=0,1\ (\mathrm{mod}\ 2)$; $\tilde{n}=n_\mu\tilde{\sigma}^\mu$; $\tilde{\sigma}^\mu=-\sigma_\mu$ $\sigma_0=\mathbb{1}_2$, σ_i are the Pauli matrices; the metric $(-1,1,1,1)$ in M_4 is used; $l\in\mathrm{SL}(2,\mathbb{C})$, $\rho>0$. We use the standard realization (cf., e.g., Ref. 20)

$$\mathrm{SU}(2,2)=\left\{g\in\mathrm{SL}(4,\mathbb{C}),\ g^+wg=w,\quad w=\begin{pmatrix}0 & \mathbb{1}_2\\ \mathbb{1}_2 & 0\end{pmatrix}\right\}.$$

Note that the coefficients of the polynomial in q can be considered to represent elements of the irreducible finite-dimensional representation spaces of $\mathrm{SL}(2,\mathbb{C})$. Some of these coefficients can be chosen to be zero consistently with (2.2), thus specifying additionally the general polynomial introduced above.

The representation (2.1)–(2.3) of P induces a representation T_χ of the whole group G in the space C_χ of functions $f\in C^\infty(\mathbb{R}^4,V_\chi)$

$$[T_\chi(g)f](x;q,z,z^+)=[D_\chi^{-1}(nma\gamma)f](x';q;z,z^+), \tag{2.4}$$

where

$$x=\begin{pmatrix}\mathbb{1}_2 & i\underline{x}\\ 0 & \mathbb{1}_2\end{pmatrix},\quad x\in\mathbb{R}^4,$$

and

$x'=x'(x,g)$, $\rho(x,g)$, $l(x,g)$, $N_0(x,g)$, and $n(x,g)$ are determined from the decomposition

$$g^{-1}x=x'nma\gamma \tag{2.5}$$

which holds for all $(g^{-1}=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right),x)$ such that $\det(icx+d)\neq0$. As for the elementary representations, certain asymptotic behavior of the function f at infinity has to be required to extend (2.4) for $\det(icx+d)=0$, e.g.,

$$(T(w)f)(x_c;q,z,z^+)$$
$$\equiv\lim_{x\to x_c}\{[D_\chi^{-1}(p(x;w))f](x_w;q;z,z^+)\}<\infty,$$

$x_c\in X_c\equiv\{x\in X,\ x^2=0\}$, $i\underline{x}_w=(i\underline{x})^{-1}$. [Use (2.4), (2.5), and (2.2) to extract the needed asymptotic behavior of f.]

Equivalently, the representations above can be realized by the left regular action of G in the space of C^∞ functions on G satisfying the covariant condition

$$\mathcal{F}(gp;q;z,z^+)=(D_\chi^{-1}(p)\mathcal{F})(g;q;z,z^+),$$
$$\mathcal{F}(xp)=[T_\chi(p)f](x),\quad \mathcal{F}(wx_c)=[T_\chi(w)f](x_c),$$
$$p\in P,\quad g\in G. \tag{2.6}$$

We can rewrite (2.4) in another equivalent form

$$(T(g)f)(x;z;z^+;q)$$
$$=\rho^d\exp(i\pi\lambda\mathcal{N}_0)f(x_g;zl^{-1},(l^+)^{-1}z^+;(1/\rho)\underline{lql}^+ +\underline{n})$$
$$\equiv(T_E(g)\varphi)(x;z,z^+;\underline{q}+\rho l^{-1}\underline{n}(l^+)^{-1}). \tag{2.4'}$$

For $g=w$ (2.5) gives $gl^{-1}\underline{n}(l^+)^{-1}=\underline{x}/|x^2|$. The representation $T_E(g)$ defined in (2.4') is such that the coefficients of the polynomial in $\underline{q}+\rho l^{-1}\underline{n}(l^+)^{-1}$ in the rhs of (2.4') are given exactly by transformed initial coefficients that are elementary representations.

(2) From now on we shall always assume that there is one and only one nonvanishing term of highest degree k in the polynomial $\varphi(q)$. Furthermore any V_χ with $j_1-j_2=0$ (mod 1) can be imbedded in some space $V_{\chi'}$ of polynomials that start and end with scalar coefficients. Similarly for $|j_1-j_2|=\frac{1}{2}$ (mod 1), V_χ can be imbedded in a spinor $(j_1-j_2=\pm\frac{1}{2})$ field. We shall consider here in detail the scalar case $\chi=(d;0,0;\lambda,k)$.

We define a Knapp–Stein-type intertwining map[17,20]

$$W_{\tilde{\chi}}\colon C_\chi\to C_{\tilde{\chi}},\quad \tilde{\chi}=(4-d+k;0,0;\lambda,k),$$
$$C_\chi\ni f\mapsto(W_{\tilde{\chi}}\mathcal{F})(x_1,q_1)$$
$$=\int dx\,P_k\left(\underline{q}_1-\tilde{x},\frac{\partial}{\partial\underline{\tilde{q}}_2}\right)\mathcal{F}(x_1wx;q_2)\Big|_{q_2=0}, \tag{2.7}$$

where w is the Weyl inversion and $P_k(a,b)$ is a homogeneous polynomial of degree $k/2$ of the arguments $a^2,b^2,a\cdot b$. (Note that k is always even for the scalar multiplets considered here.)

The intertwining property of (2.7),

$$W_{\tilde{\chi}}\circ T_\chi=T_{\tilde{\chi}}\circ W_\chi, \tag{2.8}$$

is easily checked using that

$$x_1n\gamma amwx=x_1wx'((\gamma am)^+)^{-1},$$
$$\underline{x}'=\tilde{n}+(1/\rho)(l^+)^{-1}\underline{x}l^{-1},$$

and exploiting (2.2) and (2.6). The operator (2.7) establishes in general a partial equivalence of χ and $\tilde{\chi}$. If $P(a,b)=\mathrm{const}\cdot(b^2)^{k/2}$, the image of $W_{\tilde{\chi}}$ coincides with the elementary representation subspace of $C_{\tilde{\chi}}$ labeled by $\tilde{\chi}^{\mathrm{elem}}=(4-d+k,0,0,\lambda)$.

The operator (2.7) can be easily generalized to the case of polynomials starting and ending with the same tensor type of fields or, if these fields differ, by $(j_1,j_2)\to(j_2,j_1)$. In the more general cases, other invariants, i.e., like say

$$\frac{\partial}{\partial z}\tilde{a}\frac{\partial}{\partial\tilde{z}},\frac{\partial}{\partial z}\underline{b}\frac{\partial}{\partial z},$$

etc. should be exploited along with a^2, b^2, $a\cdot b$.

(3) For $k=2$ some of the representations above can be

realized equivalently in a way reminiscent of the $N=1$ supersymmetry. The space C consists of functions $\Phi(x;V;z,z^+)$ with the same properties with respect to z, z^+ which are polynomials of degree 2 with respect to

$$v_\mu = (1/\sqrt{2})\theta\sigma_\mu\theta^+, \quad \{\theta,\theta\} = 0 = \{\theta^+,\theta\} = \{\theta^+,\theta^+\}.$$

Instead of (2.2) we have here

$$[D_\chi(\gamma amn)\Phi](v,z,z^+) = \rho^{-d}\exp(-i\pi\lambda N)\exp((1/\rho)\mathrm{tr}\,\tilde{n}l^{-1}\underline{v}(l^+)^{-1})\cdot \Phi((1/\rho)l^{-1}\underline{v}(l^+)^{-1};zl,l^+z^+). \tag{2.9a}$$

More generally one can use polynomials of θ,θ^+. Then (2.9a) is replaced by

$$[D_\chi(\gamma amn)\Phi](\theta,\theta^+;z,z^+) = \rho^{-d}\exp(-i\pi\lambda N)\exp((1/\rho)\mathrm{tr}\,\tilde{n}l^{-1}\underline{v}l^+)\cdot \Phi((1/\sqrt{\rho})\theta l,l^+\theta^+(1/\sqrt{\rho});zl,l^+z^+). \tag{2.9b}$$

[In both (2.9a) and (2.9b) the parametrization (2.3) is used.] Obviously the odd and even parts of the multiplets split in an invariant way as implied by (2.9b). Note that here the multiplets start and end automatically with the same tensor type of fields. Unlike in (2.2) the dimension of the fields in the multiplets $\Phi(v)$ increases with v; let us give a sample example exploited in Refs. 8–11:

$$\chi = (2;0,0;0;2), \quad \varphi(q) = -\tfrac{1}{2}A_+ + A^\mu q_\mu + A_- q^2,$$
$$\chi_S = (0;0,0;0;2), \quad \Phi(v) = A_- + A^\mu V_\mu - A_+ V^2, \tag{2.10}$$

both give the same transformation laws for the multiplet of fields (A_-, A_μ, A_+).

The representation T_{χ_S} induced by (2.9) is (partially) equivalent to the representation $T_{\tilde{\chi}_S}$, $\tilde{\chi}_S = (4\text{-}d\text{-}2N;j_2,j_1;\lambda;2N)$ ($N=1$ here) via the intertwining map:

$$W_{\tilde{\chi}_S}: C_{\chi_S} \ni \Phi \to (W_{\tilde{\chi}_S}\Phi)(x;v_1;z,z^+) = \int d\mu(x,\theta,\theta^+)\exp(\mathrm{tr}\,\underline{x}\underline{v}_1) \times P(-\mathrm{tr}\,\underline{v}_1\underline{v}_2)\Phi(x_1 wx;v_2;z^+\epsilon,{}^t\epsilon z), \quad \epsilon = i\sigma_2, \tag{2.11}$$

where $P(a)$ is a polynomial of a and the integral in (2.11) is Berezin's integral. If we drop the $\exp(\mathrm{tr}\,\underline{x}\underline{v})$ term in (2.11), we get a direct sum of elementary representations spaces for the image of $W_{\tilde{\chi}_S}$. Both (2.9) and (2.11) can be generalized to $N>1$ using $v_i^\mu = \theta_i\sigma^\mu\theta_i$, $i=1,2,\ldots,N$ (no summation in i). The two-point function emerging in (2.11) has been found in Ref. 21 with the infinitesimal version of (2.9a) being used.

Having (2.11) one can define an invariant Hermitian sesquilinear form on $C_\chi \times C_\chi$:

$$(\varphi,\psi) = \frac{\text{const}}{(2j_1)!(2j_2)!}\int d\mu(x,\theta,\theta^+) \times \overline{\varphi\left(x,-v;\frac{\partial}{\partial z},\frac{\partial}{\partial z^+}\right)}(W\psi)x;v;z\epsilon,{}^t\epsilon z^+). \tag{2.12}$$

III. THE SPIN-2 EXAMPLE

Let $\chi = (2;0,0;0;4)$. Consider the multiplet

$$\varphi(x,q) = -\tfrac{1}{2}A_+(x) + q^\mu A_\mu(x) + q^2 A_-(x) + q_\mu q_\nu g^{\mu\nu}(x) - 2q^2 q^\mu C_\mu(x) - (q^2)^2 D(x). \tag{3.1}$$

The dimension of the component fields starts from $d_{A_+} = 2$ and goes down to $d_D = -2$. The field $g_{\mu\nu}(x)$, $d_{g_{\mu\nu}} = 0$ is supposed to be symmetric and traceless. Using definition (2.2) and (2.4'), one gets a set of inhomogeneous special conformal transformation laws for the components. Note that the tracelessness of the transformed $g_{\mu\nu}$ is not achieved automatically—a certain rearrangement of the shifted polynomial modifying also the transformation law for A_- can be done consistently. The resulting transformations include for a given component field all the lower (in scale dimensions) components. Only the field D transforms according to an elementary representation.

The dual multiplet characterized by $\tilde{\chi} = (6;0,0;0,4)$ is of the same type

$$\Phi(x,q) = -\tfrac{1}{2}\tilde{D}(x) + q^\mu\tilde{C}_\mu(x) + 2T^{\mu\nu}(x)q_\mu q_\nu + \tfrac{1}{2}q^2 H(x) - 2q^2 q^\mu J_\mu(x) - q^4 R(x), \tag{3.2}$$

and is related to (3.1) by (2.7). Here $T_{\mu\nu} = T_{\nu\mu}$, $T^\mu_\mu = 0$. An invariant form on $\chi\times\tilde{\chi}$ is provided by

$$\langle\varphi,\Phi\rangle = -\int d^4x[A_+R + A^\mu J_\mu + A_- H + \tfrac{1}{2}g^{\mu\nu}T_{\mu\nu} + C^\mu\tilde{C}_\mu + D\tilde{D}]. \tag{3.3}$$

The Euclidean version of (2.7) more explicitly reads

$$\Phi(x_1,q_1) = \lim_{\epsilon\to 0}\frac{1}{\Gamma(\epsilon)\pi^2}\int d^4x_2\left(\frac{1}{x_{12}^2}\right)^{2-\epsilon}P_2\left(q_1^\nu + \frac{x_{12}^\nu}{x_{12}^2}, \frac{\partial}{\partial q_2^\rho}\right)\Big\{-\frac{1}{2}A_+(x_2) + (A_\mu(x_2)q_{2\nu}r^{\mu\nu} + A^\mu(x_2)x_{12\mu})\frac{1}{x_{12}^2} + (q_2^2 + x_{12}^2 - 2x_{12}^\nu q_{2\nu})\frac{A_-(x_2)}{(x_{12}^2)^2} + (q_{2\mu}r^{\mu\nu}g_{\nu\rho}(x_2)r^{\rho\sigma}q_{2\sigma} + 2x_{12\mu}g^{\mu\nu}(x_2)r_{\rho\sigma}q_2^\sigma + x_{12\mu}g^{\mu\nu}(x_2)x_{12\nu})\frac{1}{(x_2^2)^2} - \frac{2}{(x_{12}^2)^3}(q_2^2 + x_{12}^2 - 2x_{12}^\rho q_{2\rho})(C_\mu(x_2)r^{\mu\nu}q_{2\nu} + C_\mu(x_2)x_{12}^\mu) - (q_2^2 + x_{12}^2 - 2x_{12}^\rho q_{2\rho})\frac{D(x_2)}{(x_{12}^2)^4}\Big\}\Big|_{q_2=0} \tag{3.4}$$

Here,

$$r_{\mu\nu} = \eta_{\mu\nu} - 2x_{12\mu}x_{12\nu}/x_{12}^2,$$

$$P_2(a,b) = C_1(a^2)^2 + C_2a^2a\cdot b + C_3a^2b^2 + C_4(a\cdot b)^2 + C_5(a\cdot b)b^2 + C_6(b^2)^2.$$

For each of the components of Φ Eq. (3.4) reduces to a differential operator acting on the fields in φ. Inserting the result in (3.3) we get an invariant form on $C_\chi \otimes C_\chi$:

$$(\varphi,\varphi) = \int d^4x\, \mathscr{L}^1(x), \tag{3.5}$$

$$\begin{aligned}\mathscr{L}^1 = C_1[&-\tfrac{1}{2}A_+^2 - \tfrac{1}{4}A_+\partial\cdot A + \tfrac{1}{4}A_+\Box A_- + \tfrac{1}{12}A_+\partial_\mu\partial_\nu g^{\mu\nu} + \tfrac{1}{12}A_+\Box\partial\cdot C \\ &+ \tfrac{1}{8}A^\mu\partial_\mu\partial_\nu A^\nu - \tfrac{1}{24}A^\mu\partial_\mu\partial_\nu\partial_\rho g^{\nu\rho} - \tfrac{1}{24}A^\mu\partial_\nu\Box\partial\cdot C - (1/4!12)g^{\mu\nu}\partial_\mu\partial_\nu\partial_\rho\partial_\sigma g^{\rho\sigma} \\ &- (1/4!6)g^{\mu\nu}\partial_\mu\partial_\nu\Box\partial\cdot C - \tfrac{1}{8}A^\mu\partial_\mu\Box A_- - (1/4!2)g^{\mu\nu}\partial_\mu\partial_\nu\Box A_- \\ &- \tfrac{1}{32}A_-\Box^2A_- - (1/4!2)A_-\Box^2\partial\cdot C + (1/4!12)C^\mu\partial_\mu\Box^2\partial\cdot C\,] + \text{(terms containing } D) + \mathscr{L}_0,\end{aligned} \tag{3.6}$$

where

$$\begin{aligned}\mathscr{L}_0 = &-\frac{\beta}{2}\left[A^\mu(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box)A^\nu - A^\mu\partial_\mu\Box A_- + \frac{1}{2}A^\mu(\Box\eta_{\mu\nu} - \partial_\mu\partial_\nu)\partial_\rho g^{\nu\rho}\right. \\ &\left.+\frac{1}{8}A^\mu(\Box\eta_{\mu\nu} - \partial_\mu\partial_\nu)\Box C^\nu - \frac{3}{8}A^\mu\partial_\mu\Box\partial\cdot C\right] - \frac{\alpha}{2}\left[A_-\Box^2A_- + \frac{3}{4}A_-\Box^2\partial\cdot C\right] \\ &+\frac{\beta}{8}\left[A_-\Box\partial_\mu\partial_\nu g^{\mu\nu} + \frac{3}{8}g^{\mu\nu}\partial_\mu\partial_\nu\Box\partial\cdot C - g^{\mu\nu}\partial_\mu(\Box\eta_{\nu\sigma} - \partial_\nu\partial_\sigma)\partial_\rho g^{\rho\sigma}\right. \\ &\left.+\frac{1}{8^2}C^\mu\Box^2(\Box\eta_{\mu\nu} - \partial_\mu\partial_\nu)C^\nu\right] + \frac{9\alpha}{2\cdot 8^2}C^\mu\partial_\mu\Box^2\partial_\nu C^\nu - \frac{\gamma}{2}\left[g^{\mu\nu}\Box^2g_{\mu\nu} - 2g^{\mu\nu}\partial_\mu\partial_\rho\Box g_\nu^\rho + \frac{2}{3}g^{\mu\nu}\partial_\mu\partial_\nu\partial_\rho\partial_\sigma g^{\rho\sigma}\right. \\ &\left.-g^{\mu\nu}\partial_\mu(\Box\eta_{\nu\rho} - \partial_\nu\partial_\rho)\Box C^\rho + \frac{1}{3}g^{\mu\nu}\partial_\mu\partial_\nu\Box\partial_\rho C^\rho\right] - \frac{\delta}{2}C^\mu\Box^2(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box)C^\nu + \left(\frac{5\gamma}{96} - \frac{\delta}{4}\right)C^\mu\partial_\mu\Box^2\partial^\nu C_\nu.\end{aligned} \tag{3.7}$$

(The constants $\alpha, \beta, \gamma, \delta$ are expressed by the initial constants C_i.)

To simplify the model we shall choose a subspace of C_χ such that A_+ is expressed through the rest of the fields in a covariant way

$$\begin{aligned}A_+ = &-\tfrac{1}{96}\Box^2D + \tfrac{1}{12}\Box\partial^\nu C_\nu + \tfrac{1}{12}\partial_\mu\partial_\nu g^{\mu\nu} \\ &+\tfrac{1}{4}\Box A_- - \tfrac{1}{2}\partial^\mu A_\mu,\end{aligned} \tag{3.8}$$

and further we shall choose $D = 0$ (which is possible since D is the lowest elementary field). Then we are left exactly with the form provided by (3.7). To the multiplet $(C_\mu, g_{\mu\nu}, A_-, A_\mu)$ thus obtained there corresponds a shortened dual multiplet $(J_\mu, H, T_{\mu\nu}, \tilde{C}_\mu)$. Adding to (3.7) what is left from the "interaction" terms (3.3) we finally get

$$\mathscr{L} = \mathscr{L}_0 + A^\mu J_\mu + \tfrac{1}{2}g^{\mu\nu}T_{\mu\nu} + A_-H + C^\mu\tilde{C}_\mu, \tag{3.9a}$$

which provides a conformal invariant action for the fields involved. We shall rewrite it in a compact form using an obvious notation:

$$\mathscr{L} = -\tfrac{1}{2}\mathscr{A}_a\mathscr{M}^{ab}(\partial)\mathscr{A}_b + \mathscr{A}_a\mathscr{J}^a. \tag{3.9b}$$

Then (3.9b) implies the following set of (classical) equations of motion in the presence of the external source:

$$\mathscr{M}^{ab}(\partial)\mathscr{A}_b = \mathscr{J}^a. \tag{3.10}$$

The free propagators (Schwinger functions) of the fields in $\mathscr{A}_a$ can be combined in a matrix $G_{ab}(x,y)$ satisfying

$$\mathscr{M}(\partial)\cdot G = \text{identity}. \tag{3.11}$$

The existence of G as an inverse of $\mathscr{M}$ is related to the fact that the Minkowski space Lagrangian corresponding to (3.9b) [as well as the more general (3.6)] is nonsingular for all generic values of the constants involved. There is still some arbitrariness in the choice of certain constants in G which may be fixed by requiring G to coincide with the invariant (matrix) two-point function obtained from the Euclidean version of the intertwining operator (2.7). Note that (2.7) itself reproduces the corresponding free Wightman two-point functions.

The model built above has various interpretations.

(1) It obviously comprises as a particular case the conformal electrodynamics model of Refs. 6 and 8–11. Indeed a choice $g_{\mu\nu} = 0 = C_\mu$ reduces the field multiplet to (A_μ, A_-), or to (A_+, A_μ, A_-).

(2) Integrating over the field A_- and then A_μ (or equivalently choosing $\beta = 0 = \alpha$) and postulating $J_\mu = 0 = H$, one gets a model described by ($\gamma = 1$)

$$\begin{aligned}\mathscr{L}^{\rm gr} = &\tfrac{1}{2}C^{\mu\nu}{}_{\rho\sigma}\left\{\tfrac{1}{2}C_{\mu\nu}{}^{\rho\sigma} - [\partial_{[\mu}\partial^{[\rho}g_{\nu]}^{\sigma]} - \tfrac{1}{3}(\eta_\nu^\sigma\eta_\mu^\rho - \eta_\mu{}^\sigma\eta_\nu{}^\rho)\right. \\ &\left.\times\partial^\lambda\partial^\kappa g_{\lambda\kappa} - \tfrac{1}{2}\eta_{[\mu}{}^{[\rho}(\Box g_{\nu]}{}^{\sigma]} - 2\partial_{\nu]}\partial^\lambda g_\lambda{}^{\sigma]})]\right\} \\ &+\tfrac{1}{2}g^{\mu\nu}\partial_\mu(\Box\eta_{\nu\rho} - \partial_\nu\partial_\rho)C^\rho - \tfrac{1}{2}g^{\mu\nu}\partial_\mu\partial_\nu\Box\partial^\rho C_\rho \\ &+(\delta/2)C^\mu\Box^2(\Box\eta_{\mu\nu} - \partial_\mu\partial_\nu)C^\nu \\ &+(\tfrac{5}{96} - \delta/4)C^\mu\partial_\mu\Box^2\partial^\nu C_\nu.\end{aligned} \tag{3.12}$$

This can be interpreted as a nonsingular Lagrangian for the

linearized Weyl gravity that provides a conformally invariant action ($C_{\mu\nu}{}^{\rho\sigma}$ is the Weyl tensor). We would arrive at (3.12) starting from a polynomial of the type $\varphi_{\mu\nu}(q,x) = g_{\mu\nu}(x) - 2q_{(\mu}C_{\nu)}(x) + \eta_{\mu\nu}q^\rho C_\rho(x)$ with dual multiplet $\Phi_\mu(x,q) = \tilde{C}_\mu(x) + 2T_{\mu\nu}(x)q^\nu$.

A nonsingular formulation of the linearized Weyl gravity has been constructed recently[16] using the indecomposable elementary representations of Ref. 11. Starting from $\mathscr{L}_0^{\rm gr}$ in (3.12) and integrating over the field $C_\mu(x)$ we will reproduce exactly the result of Ref. 16. Since effectively this expresses $C_\mu(x)$ in a nonlocal way by $g_{\mu\nu}$, one gets instead of the simple linear transformations for $(g_{\mu\nu}, C_\nu)$, the nonlocal law of Ref. 16 for $g_{\mu\nu}(x)$ which can be interpreted[14] as a sum of subrepresentations and of factor representations of elementary indecomposable representations. The experience with the analysis of the renormalized equations of motion in the massless quantum electrodynamics[11] suggests however that one can hardly go very far in the quantum case with the approach of Refs. 7 and 16.

The Weyl tensor $C_{\mu\nu\gamma\rho}$ is invariant under the gauge transformations (local deformations) of the metric

$$g_{\mu\nu} \to g_{\mu\nu} - \partial_{(\mu}\xi_{\nu)} + \tfrac{1}{2}\eta_{\mu\nu}\partial^\rho\xi_\rho. \tag{3.13}$$

(The solutions of the equation

$$\partial_{(\mu}\xi_{\nu)} - \tfrac{1}{2}\eta_{\mu\nu}\partial^\rho\xi_\rho = 0$$

parametrize the infinitesimal conformal transformations.) The equations of motion implied by (3.12) read

$$\begin{aligned}&\tfrac{1}{3}\partial^\kappa\partial^\rho C_{\mu\kappa\nu\rho} + \tfrac{2}{3}\partial_\mu\partial_\nu\Box\partial^\rho C_\rho - \tfrac{1}{4}\Box^2\partial_{(\mu}C_{\nu)} - \tfrac{1}{24}\eta_{\mu\nu}\Box^2\partial^\rho C_\rho\\ &\quad = \tfrac{1}{2}T_{\mu\nu} - \delta(\Box^3 C_\mu - \tfrac{3}{2}\Box^2\partial_\mu\partial^\rho C_\rho) - \tfrac{5}{48}\Box^2\partial_\mu\partial^\rho C_\rho\\ &\qquad + \tfrac{1}{2}\Box^2\partial^\kappa g_{\kappa\mu} - \tfrac{2}{3}\partial_\mu\Box\partial^\rho\partial^\kappa g_{\rho\kappa} = \tilde{C}_\mu.\end{aligned} \tag{3.14}$$

In the presence of a conserved energy-momentum tensor we have

$$\tfrac{2}{3}\partial_\nu\Box^2\partial^\rho C_\rho - \Box^3 C_\nu = 0, \tag{3.15a}$$

which leads to the conformal covariant gauge condition

$$-\tfrac{5}{48}\Box^2\partial_\mu\partial^\rho C_\rho + \tfrac{1}{2}\Box^2\partial^\nu g_{\nu\mu} - \tfrac{2}{3}\partial_\mu\Box\partial^\rho\partial^\nu g_{\rho\nu} = \tilde{C}_\mu. \tag{3.15b}$$

Note finally that as in Ref. 6 and 11 there exists covariant expression for the current $\tilde{C}_\mu$

$$\tilde{C}_\mu = (\partial^\nu S)T_{\nu\mu}, \tag{3.16}$$

in terms of the energy-momentum tensor and the field $\partial_\mu S$, which together with the constant field q transforms as the pair (A_μ, A_-). A kinetic term for S as in Refs. 6 and 11 should be added if (3.16) is assumed.

The free propagator matrix (3.11) reduces in this simple case to the expressions

$$\begin{aligned}&\langle g_{\mu\nu}(x_1)g_{\gamma\rho}(x_2)\rangle\\ &\quad = \int (dp)l^{ip\cdot x_{12}}\Big\{\Big[\frac{48\delta-6}{p^4}p_\mu p_\nu p_\gamma p_\rho + \Big(2\delta - \frac{3}{4}\Big)\eta_{\mu\nu}\eta_{\gamma\rho}\\ &\qquad - \Big(\frac{1}{2}+4\delta\Big)\frac{1}{p^2}(p_\mu p_\gamma\eta_{\nu\rho} + p_\nu p_\gamma\eta_{\mu\rho} + p_\mu p_\rho\eta_{\nu\gamma} + p_\nu p_\rho\eta_{\mu\gamma}) + \frac{2-8\delta}{p^2}(p_\mu p_\nu\eta_{\gamma\delta} + p_\gamma p_\rho\eta_{\mu\nu})\\ &\qquad + \frac{1}{2}(\eta_{\mu\gamma}\eta_{\nu\rho} + \eta_{\mu\rho}\eta_{\nu\gamma})\Big]\Big(\frac{1}{p^2}\Big)^2 + d\,[\eta_{\mu\nu}\eta_{\gamma\rho} - 2(\eta_{\gamma\mu}\eta_{\rho\nu} + \eta_{\gamma\nu}\eta_{\rho\mu})]\delta(p),\end{aligned} \tag{3.17a}$$

$$\begin{aligned}\langle C_\mu(x_1)g_{\gamma\rho}(x_2)\rangle &= i\int (dp)l^{ip\cdot x_{12}}\Big\{\Big(\frac{1}{p^2}\Big)^2\Big[-12\frac{p_\mu p_\gamma p_\rho}{p^4} + 2(p_\mu\eta_{\gamma\rho} + p_\gamma\eta_{\mu\rho} + p_\rho\eta_{\gamma\mu})\Big]\\ &\quad + ib\Big(\eta_{\gamma\mu}\partial_\rho^p + \eta_{\rho\mu}\partial_\gamma^p - \frac{1}{2}\eta_{\gamma\rho}\partial_\mu^p\Big)\delta(p)\Big\},\end{aligned} \tag{3.17b}$$

$$\langle C_\mu(x_1)C_\nu(x_2)\rangle = a\int (dp)l^{ip\cdot x_{12}}(\Box\eta_{\mu\nu} - 2\partial_\mu\partial_\nu)\delta(p), \quad (dp) = d^4p/(2\pi)^4. \tag{3.17c}$$

The two-point Schwinger function

$$\langle C_{\mu\eta\nu\kappa}(x_1)C_{\gamma\alpha\rho\beta}(x_2)\rangle$$

that results from (3.17a) coincides with the elementary representation invariant Knapp–Stein kernel. [We assume here that the (Euclidean) conformal group is enlarged to include space reflections thus ensuring the irreducibility of the unitary representation described by $C_{\mu\rho\nu\kappa}$.] Note that the corresponding Wightman two-point function of the Minkowski field $C_{\mu\rho\nu\kappa}(x)$ is well known to define an indefinite form, the relevant subrepresentation of the Minkowski space conformal group being nonunitary.[20]

(3) Finally the Lagrangian (3.9) describes a mixed system including the linear conformal gravity along with the electromagnetic field. Note that $\mathscr{L}_0$ in (3.9) can be looked upon as a Hertz-type formulation of the free electromagnetic field (see, e.g., Ref. 22). The fields C_μ and $g_{\mu\nu}$ give in general a contribution to the transversal as well as to the longitudinal parts of the A_μ-propagator. There exists however a choice of a constant $\delta = \delta(\gamma,\beta)$ such that only longitudinal terms are produced—thus the whole system in (3.9) can be looked as another version of the nonsingular massless electrodynamics. Note that analogously to (3.16) all higher "currents" in (3.9) can be realized explicitly in terms of the (elementary representation) electromagnetic current j_μ and the field $\partial_\mu S$. An alternative Hertz-type model can be built using instead of (3.1) the multiplet

$$A_\mu + q^\nu H_{\mu\nu} + 2q_\mu A_- - 2q^2 C_\mu + 4q_\mu q^\rho C_\rho,$$

with $H_{\mu\nu} = -H_{\nu\mu}$, $d_{H_{\mu\nu}} = 0$.

IV. MANIFESTLY O(4,2)-COVARIANT FORMALISM FOR THE PHOTON-WEYL GRAVITON SYSTEM

As is well known (see Ref. 10 and references therein) the manifestly O(4,2)-covariant [or O(5,1)-covariant in the Euclidean case] formalism[18] generates naturally just nonelementary representations, the elementary ones, being singled out by imposing subsidiary conditions.

In order to "translate" the results of Sec. III into the manifestly covariant language, let us introduce the (isotropic) six-vector $Q^a(\xi;q)$ with homogeneity degree (-1), whose components Q_M $(M=+,\mu,-)$, in the Aut $\mathscr{P}$ basis defined in Ref. 10, are given by

$$Q_+ = q^2, \quad Q_\mu = q_\mu, \quad Q_- = -\tfrac{1}{2}. \tag{4.1}$$

Let us consider the theory of a symmetric, traceless (and then irreducible) O(4,2) [or O(5,1)] tensor field $\mathscr{G}^{ab}(\xi)$ (six-dimensional Weyl graviton) with homogeneity degree zero and corresponding six-dimensional "energy-momentum tensor" $\mathscr{T}^{ab}(\xi)$ with homogeneity degree (-4). Then it is easy to get Eqs. (3,1) and (3,2) again, once we define the quadratic forms

$$\varphi(x;q) = \kappa^2 \mathscr{G}^{ab}(\xi)\mathscr{Q}_a(\xi;q)\mathscr{Q}_b(\xi;q), \tag{4.2}$$

$$\Phi(x,q) = \kappa^6 \mathscr{T}^{ab}(\xi)\mathscr{Q}_a(\xi;q)\mathscr{Q}_b(\xi;q), \tag{4.3}$$

with the following natural identification of the components of G_{MN} and t_{MN} $(M;N=+,\mu,-)$ of $\mathscr{G}$ and $\mathscr{T}$ in the Aut $\mathscr{P}$ basis[10]

$$\begin{aligned} &A_+ = -\tfrac{1}{2}G_{++}, \quad A_- = \tfrac{3}{4}G_\rho{}^\rho = -\tfrac{3}{2}G_{+-}, \\ &A_\mu = -G_{\mu+}, \quad C_\mu = -G_{\mu-}, \\ &g_{\mu\nu} = G_{\mu\nu} - \tfrac{1}{4}\eta_{\mu\nu}G_\rho{}^\rho, \quad D = -G_{--}, \end{aligned} \tag{4.4a}$$

and

$$\begin{aligned} &\tilde{D} = -\tfrac{1}{2}t_{++}, \quad H = \tfrac{1}{2}t_\rho{}^\rho = -t_{+-}, \\ &\tilde{C}_\mu = -t_{\mu+}, \quad J_\mu = -t_{\mu-}, \\ &T_{\mu\nu} = t_{\mu\nu} - \tfrac{1}{4}\eta_{\mu\nu}t_\rho{}^\rho, \quad R = -t_{--}. \end{aligned} \tag{4.4b}$$

The invariant form (3.3) is simply given by

$$\langle\varphi,\Phi\rangle = -\frac{1}{2}\int d\mu_n(\xi)\mathscr{G}_{ab}(\xi)\mathscr{T}^{ab}(\xi), \tag{4.5}$$

where

$$d\mu_n(\xi) = 2d^6\xi\,\delta(\xi^2)\delta(\xi\cdot n - 1).$$

In order to construct the invariant action for $\mathscr{G}^{ab}$ we should find the (Euclidean) two-point functions of the energy-momentum $\mathscr{T}^{ab}$. Since the procedure is just the same as in Ref. 10 (see Sec. III) we give here only the final expression for the general manifestly covariant Euclidean two-point function of $\mathscr{T}^{ab}$ (without tracelessness condition yet imposed)

$$\begin{aligned} &\langle\mathscr{T}^{ab}(\xi_1)\mathscr{T}^{cd}(\xi_2)\rangle_E \\ &= [A_1\xi_1^a\xi_1^b\xi_2^c\xi_2^d + A_2\xi_2^a\xi_2^b\xi_1^c\xi_1^d + A_3(\xi_1^a\xi_2^b + \xi_1^b\xi_2^a)(\xi_1^c\xi_2^d + \xi_1^d\xi_2^c)][1/(-2\xi_1\cdot\xi_2)^{d+4}] \\ &\quad + [B_1\eta^{ab}(\xi_1^c\xi_2^b + \xi_1^d\xi_2^c) + B_2\eta^{cd}(\xi_1^a\xi_2^b + \xi_1^b\xi_2^a) + B_3(\eta^{ac}\xi_1^b\xi_2^d + \eta^{ad}\xi_1^b\xi_2^c + \eta^{bc}\xi_1^a\xi_2^d + \eta^{bd}\xi_1^a\xi_2^c) \\ &\quad + B_4(\eta^{ac}\xi_2^b\xi_1^d + \eta^{ad}\xi_2^b\xi_1^c + \eta^{bc}\xi_2^a\xi_1^d + \eta^{bd}\xi_2^a\xi_1^c)][1/(-2\xi_1\cdot\xi_2)^{d+3}] \\ &\quad + [C_1\eta^{ab}\eta^{cd} + C_2(\eta^{ac}\eta^{bd} + \eta^{ad}\eta^{bc})][1/(-2\xi_1\cdot\xi_2)^{d+2}], \end{aligned} \tag{4.6}$$

where the limit $d\to 2$ has to be performed. We note that the limit $d\to -2$ in Eq. (4.6) gives us the "6-Weyl graviton" two-point function $\langle\mathscr{G}^{ab}(\xi_1)\mathscr{G}^{cd}(\xi_2)\rangle_E$. The well-known reduction procedure (see, e.g., Ref. 10) allows us to derive directly from (4.6) the x-space "current–current" (matrix) two-point function (as well as the corresponding four-space propagator matrix). They coincide with those ones obtained in Sec. III from the Lagrangian density (3.6), (3.7), once we relate the constants $\alpha,\beta,\gamma,\delta,C_1$ with those ones appearing in Eq. (4.6).

We have to remark that the three constants B_1, B_2, and C_1 do not appear in the expressions for the x-space current–current two-point functions. Furthermore, if we impose the tracelessness condition

$$\eta^{ab}\mathscr{T}_{ab} = 0 \tag{4.7}$$

on the two-point function (4.6), we see that it simply implies two relations for B_1 and B_2. Therefore the condition (4.7) has no influence on the x-space current–current functions expressions.

Since the x-space currents appearing in the multiplet (4.5) form an indecomposable representation of the conformal group the conformally invariant current conservation laws mix them in general. Nevertheless we are able to write such manifestly O(4,2)-covariant subsidiary conditions on $\mathscr{T}^{ab}(\xi)$ that they provide us with the usual (electromagnetic) current conservation

$$\partial^\mu J_\mu = 0. \tag{4.8}$$

In fact, from the condition

$$\delta_a\xi_b\mathscr{T}^{ab}(\xi) = 0 \tag{4.9}$$

(where δ_a is the interior derivative defined in Ref. 10), we get

$$\partial^\mu J_\mu(x) + \tfrac{1}{2}\Box R(x) = 0, \tag{4.10}$$

while the subsidiary condition

$$\xi_a\xi_b\mathscr{T}^{ab}(\xi) = 0 \tag{4.11}$$

gives simply $R(x)=0$ and therefore altogether conditions (4.9) and (4.11) imply Eq. (4.8).

To simplify the model we can impose here also condition (3.8). Its manifestly covariant counterpart is given by

$$\delta_a\delta_b\mathscr{G}^{ab}(\xi) = 0. \tag{4.12}$$

The conformal invariant constraint $D=0$ imposed in

Sec. III is easily translated into the condition

$$\xi_a \xi_b \mathcal{G}^{ab}(\xi) = 0. \tag{4.13}$$

Finally, we conclude that the manifestly covariant six-dimensional formalism recovers in compact form all results given in Sec. III and provides us with a simpler and easier method to construct nonsingular conformally invariant local Lagrangians (with invariant gauge-fixing terms). The six-dimensional formalism [as well as nonelementary multiplet representations $\varphi(x,q)$] seems to be the natural way to introduce in the conformal invariant gauge theories a minimal set of auxiliary fields needed to describe "off-shell" both local and conformal invariant gauge conditions.

ACKNOWLEDGMENTS

Two of us (V.B.P. and G.M.S.) would like to thank the local section of I.N.F.N. for the kind hospitality in Trieste where this work was completed. G. M. S. is grateful to Professor P. Budinich for his hospitality at the International School for Advanced Studies in Trieste.

The essential results of the paper were obtained two years ago and this work would have never appeared without the continuing insistence of Professor I. T. Todorov whom we thank also for useful discussions.

[1]P. Budinich, Nuovo Cimento **21**, 473 (1978).
[2]G. Mack and A. Salam, Ann. Phys. (NY) **53**, 174 (1969).
[3]D. H. Mayer, J. Math. Phys. **16**, 884 (1975).
[4]F. Bayen and M. Flato, J. Math. Phys. **17**, 1112 (1976).
[5]I. T. Todorov, M. C. Mintchev, and V. B. Petkova, *Conformal Invariance in Quantum Field Theory* (Scuola Normale Superiore, Pisa, 1978).
[6]G. Sotkov and D. Stoyanov, J. Phys. A **13**, 2807 (1980); JINR Report P2-81-665, Dubna, 1981 (in Russian); J. Phys. A **16**, 2817 (1983).
[7]E. S. Fradkin, A. A. Kozhevnikov, M. Ya. Palchik, and A. A. Ponceranski, Commun. Math. Phys. **91**, 529 (1983); A. A. Kozhevnikov, M. Ya Palchik, and A. A. Ponceranski, Sov. J. Nucl. Phys. **37**, 287 (1983).
[8]B. Binegar, C. Fronsdal, and W. Heidenreich, J. Math. Phys. **24**, 2828 (1983).
[9]R. P. Zaikov, JINR preprint E2-83-28, Dubna, 1983.
[10]P. Furlan, V. B. Petkova, G. M. Sotkov, and I. T. Todorov, Riv. Nuovo Cimento **8**, 3 (1985).
[11]V. B. Petkova, G. M. Sotkov, and I. T. Todorov, Commun. Math. Phys. **97**, 227 (1985).
[12]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974); **17**, 1930 (E) (1976).
[13]S. Ferrara, R. Gatto, and A. Grillo, *Conformal Algebra in Space-Time and OPE Springer Tracts in Modern Physics*, Vol. 67 (Springer, Berlin, 1973).
[14]V. K. Dobrev, G. Mack, V. B. Petkova, S. G. Petrova, and I. T. Todorov, *Lecture Notes in Physics*, Vol. **63** (Springer, Berlin, 1977); V. K. Dobrev and V. B. Petkova, Rep. Math. Phys. **13**, 233 (1978).
[15]V. B. Petkova and G. M. Sotkov, Lett. Math. Phys. **8**, 217 (1984).
[16]E. S. Fradkin and M. Ya Palchik, Class. Quantum Grav. **1**, 131 (1984).
[17]A. W. Knapp and E. M. Stein, Ann. Math. **93**, 489 (1971).
[18]P. A. M. Dirac, Ann. Math. **37**, 429 (1936).
[19]T. Banks and M. Peskin, SLAC-PUB-3740 preprint, 1985; W. Siegel and B. Zwiebach, Nucl. Phys. B **263**, 105 (1986); A. Neveu, J. Schwarz, and P. C. West, Phys. Lett. B **164**, 51 (1985).
[20]G. Mack, Commun. Math. Phys. **55**, 1 (1977).
[21]Ch. Makrelov, G. Sotkov, and R. P. Zaikov (unpublished).
[22]D. Zwanziger, Phys. Rev. D **19**, 3614 (1979); E. d'Emilio and M. C. Mintchev, Nuovo Cimento A **69**, 43 (1982).

Clebsch–Gordan coefficients for SU(5) unification models

M. A. del Olmo,[a] J. Patera, M. A. Rodriquez,[b] and C. Rousseau
Centre de recherches mathématiques, et Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec, Canada H3C 3J7

(Received 30 June 1986; accepted for publication 17 September 1986)

The Clebsch–Gordan coefficients for the product (1001) ⊗ (1001), where (1001) is the adjoint representation of SU(5), with respect to the group basis and the subgroup basis in the reduction $SU(5) \supset SU(3) \times SU(2) \times U(1)$ are computed. One of the basic tools in this computation is the exhaustive use of the Verma algorithm to find bases for the weight subspaces of dimension higher than 1. It allows for the construction of bases in a systematic way by using the so-called Verma inequalities. Only the coefficients for the dominant weights are calculated. The other ones can be obtained by using the elements of finite order (charge conjugation operators) of SU(5).

I. INTRODUCTION

The need to compute Clebsch–Gordan and related coefficients was the main avenue used by group representation theory to find its place in physics. As long as SU(2) was the relevant group, it was possible to seek the properties of the Clebsch–Gordan coefficients (CGC) in general, and in every detail.[1,2] The situation is already quite different for rank 2 Lie groups in spite of the fact that some CGC's are part of the everyday life of particle physicists and that general expressions for CGC are known.[3] For simple Lie algebras/groups of rank $n \geqslant 3$, many particular CGC's were calculated and a limited number of (infinite) series of special cases are published.[4]

Obstacles in deriving CGC in other cases invariably stem from the difficulty of building appropriate bases in representation spaces. More precisely, a basis of an irreducible space V_Λ of dim $V_\Lambda < \infty$, which decomposes into the direct sum

$$V_\Lambda = \bigoplus_{\lambda \in \Omega(\Lambda)} V_\Lambda(\lambda) \tag{1.1}$$

of subspaces $V_\Lambda(\lambda)$ labeled by weights λ (= sets of suitable chosen additive quantum numbers), is given by the weight system $\Omega(\Lambda)$ of the representation (Λ) as long as the subspaces $V_\Lambda(\lambda)$ are all one dimensional. Once one has dim $V_\Lambda(\lambda) > 1$ for some λ, the construction of a basis becomes considerably more involved.

The purpose of this article is to present the Clebsch–Gordan coefficients for an important particular case [24 ⊗ 24 of SU(5)] where the dimensions dim $V_\Lambda(\lambda)$ range up to 10. In that respect, our case is the most complicated ever worked out. Furthermore, our computation has two other objectives: (i) to provide the CGC in a basis which reduces naturally to the subgroup $SU(3) \times SU(2) \times U(1)$ of SU(5), and (ii) to proceed in a way not particular to SU(5) or SU(n)-type Lie groups (more precisely Lie algebras).

The first objective is clearly motivated by particle physics. Virtually all theories of unification of electromagnetic, weak, and strong interactions lead to groups such as SU(5) (or larger) containing $SU(3)^C \times SU(2)^W \times U(1)^W$ as a subgroup (for a detailed description of these models, see Ref. 4). The second objective is of general interest. Thus for instance, the well-known basis of Gel'fand–Zeitlin,[5] which could have been used to solve our problem here, would offer no advantages imposing a restriction to groups of the type SU(n) only.

The present article complements Ref. 6, where SU(5) CGC were found for lower representations. The method used here is that of Refs. 6 and 7. Its essential feature is a tabulation of only a small number of representative CGC. Any other CGC is obtained from our tables by application of the charge conjugation operators R of Refs. 6–8. The new feature here is a systematic exploitation of Verma bases[8] in $V_\Lambda(\lambda)$ of dimension > 1. These bases were not known at the publication of Refs. 6 and 7.

In Sec. II, we present a short summary of some mathematical tools used in this paper. A detailed account of them can be found in Refs. 6 and 7. We describe Verma bases in Sec. III and compute the bases corresponding to the cases studied here. Section IV contains some examples of the explicit computation of CGC in the group basis and the corresponding tables. The next section is devoted to CGC in the subgroups' basis [$SU(3) \times SU(2) \times U(1)$ of SU(5)]. Finally, in Sec. VI we comment on some properties of the method.

II. MATHEMATICAL PRELIMINARIES

In the following we will work with the Lie algebras su(n), with $n = 2,3,5$. We choose the $n^2 - 1$ generators of su(n) in the following way:

$$e_{\alpha_i},\ f_{\alpha_i},\ h_{\alpha_i},\quad i = 1,\ldots,n-1, \tag{2.1}$$

$$e_{(\alpha_i+\alpha_{i+1})},\quad f_{(\alpha_i+\alpha_{i+1})},\quad i = 1,\ldots,n-2, \tag{2.2}$$

$$\vdots$$

$$e_{(\alpha_1+\cdots+\alpha_{n-1})},\quad f_{(\alpha_1+\cdots+\alpha_{n-1})}, \tag{2.3}$$

where α_i, $i = 1, 2,\ldots,n-1$, are the simple roots of the algebra, considered as vectors of a real Euclidean space. The

[a] On leave of absence from Departamento de Física Teórica, Facultad de Ciencias, Universidad de Vallodolid, 47005 Valladolid, Spain.
[b] Present address: Departamento de Métodos Matemáticos de la Física, Facultad de Físicas, Universidad Complutense, 28040 Madrid, Spain.

 0022-2488/87/020258-14$02.50 © 1987 American Institute of Physics

TABLE I. Branching rules for $SU(5)\supset SU(3)\times SU(2)\times U(1)$.

$(2002)\supset(22)(0)(0)\oplus(11)(2)(0)\oplus(11)(0)(0)\oplus(21)(1)(5)\oplus(12)(1)(5)\oplus(10)(3)(5)$
$\oplus(01)(3)(5)\oplus(10)(1)(5)\oplus(01)(1)(5)\oplus(20)(2)(10)\oplus(02)(2)(10)\oplus(00)(4)(0)$
$\oplus(00)(2)(0)\oplus(00)(0)(0)$

$(0102)\supset(12)(1)(5)\oplus(03)(0)(0)\oplus(11)(2)(0)\oplus(11)(0)(0)\oplus(10)(3)(5)\oplus(02)(0)(10)$
$\oplus(02)(1)(5)\oplus(00)(2)(0)\oplus(01)(2)(10)\oplus(01)(1)(5)\oplus(10)(1)(5)$

$(2010)\supset(21)(1)(5)\oplus(30)(0)(0)\oplus(11)(2)(0)\oplus(11)(0)(0)\oplus(01)(3)(5)\oplus(20)(0)(10)$
$\oplus(20)(1)(5)\oplus(00)(2)(0)\oplus(10)(2)(10)\oplus(10)(1)(5)\oplus(01)(1)(5)$

$(0110)\supset(11)(2)(0)\oplus(11)(0)(0)\oplus(20)(1)(5)\oplus(02)(1)(5)\oplus(10)(1)(5)\oplus(01)(1)(5)$
$\oplus(10)(0)(10)\oplus(01)(0)(10)\oplus(00)(0)(0)$

$(1001)\supset(11)(0)(0)\oplus(10)(1)(5)\oplus(01)(1)(5)\oplus(00)(2)(0)\oplus(00)(0)(0)$

scalar product is

$$(\alpha_i,\alpha_j)=2\delta_{ij}-\delta_{1,|j-i|}\,,\quad i,j=1,2,\ldots,n-1\,. \tag{2.4}$$

The nonzero commutation relations of su(n) operators are

$$[e_\alpha,e_\beta]=\lambda_{\alpha\beta}e_{\alpha+\beta}\,,\quad [f_\alpha,f_\beta]=\lambda_{\alpha\beta}f_{\alpha+\beta}\,, \tag{2.5}$$

$$[e_\alpha,f_\alpha]=h_\alpha\,,\quad [h_\alpha,e_\alpha]=2e_\alpha,\quad [h_\alpha,f_\alpha]=-2f_\alpha\,, \tag{2.6}$$

where α,β, and $\alpha+\beta$ are roots of the algebra, and $\lambda_{\alpha\beta}$ are constants that can always be chosen as integers. For more details see Ref. 8. In the following we will write e_i instead of $e_{\alpha i}$, etc.

Each irreducible representation of a Lie algebra is characterized by the highest weight. The irreducible representation space is spanned by the weight vector which we denote by its weight. For instance, in the case of SU(5), $m=(m_1,m_2,m_3,m_4)$. The coordinates of the highest weight are non-negative integers. The weights of a representation with non-negative coordinates (in the basis of fundamental weights) are called dominant weights. There exists a standard algorithm to compute all the weights of a linear representation starting from the highest one.

Consider the subgroup $SU(3)\times SU(2)\times U(1)$ of SU(5) (see Table I). Its generators are linear combinations of SU(5) generators. In standard conventions, the SU(5) weights are related to those of the subgroup as follows:

$$m'=mP\,, \tag{2.7}$$

where P is the projection matrix of Ref. 6:

$$P=\begin{bmatrix}1&0&0&2\\1&0&1&\bar{1}\\0&1&1&1\\0&1&0&\bar{2}\end{bmatrix}. \tag{2.8}$$

In particular, the roots of SU(5) are projected into roots of the subgroup, and the generators of the three groups SU(3), SU(2), and U(1) are

$$\begin{aligned}&\text{SU(3):}\ \tilde{e}_1=e_{1+2}=[e_1,e_2]\,,\quad \tilde{e}_2=e_{3+4}=[e_3,e_4]\,,\\&\qquad\tilde{h}_1=h_1+h_2\,,\quad \tilde{f}_1=f_{1+2}=[f_1,f_2]\,,\\&\qquad\tilde{f}_2=f_{3+4}=[f_3,f_4]\,,\quad \tilde{h}_2=h_3+h_4\,;\end{aligned} \tag{2.9}$$

$$\begin{aligned}&\text{SU(2):}\ \hat{e}=e_{2+3}=[e_2,e_3],\quad f=f_{2+3}=[f_2,f_3],\\&\qquad\hat{h}=h_2+h_3\,;\end{aligned} \tag{2.10}$$

$$\text{U(1):}\ \bar{u}=2h_1-h_2+h_3-2h_4\,. \tag{2.11}$$

In the following the U(1) component of an $SU(3)\times SU(2)\times U(1)$ weight will not appear because it does not vary within an irreducible representation and for this reason, it is not necessary to compute CGC. It can always be found, for instance, using the generating function of Ref. 9.

In our tables, there are only the CGC of the dominant weight vectors, but one can easily compute the CGC of the nondominant ones, making use of the charge conjugation operators,[7] which are given by (α_i is a simple root)

$$\begin{aligned}R_i&=\exp(f_i)\exp(-e_i)\exp(f_i)\\&=(1+f_i+\cdots)(1-e_i+\cdots)(1+f_i+\cdots)\,.\end{aligned} \tag{2.12}$$

These operators generate the finite Demazure–Tits subgroup[10] N of SU(5). They act in the representation space V_Λ in the following way:

$$\begin{aligned}&R_1V_\Lambda(\lambda)=V_\Lambda(r_i\lambda)\,,\\&R_i|\lambda\rangle=\pm|r_i\lambda\rangle\quad\text{if } r_i\lambda\neq\lambda \text{ and } |\lambda\rangle\in V_\Lambda(\lambda)\,,\end{aligned} \tag{2.13}$$

where $|\lambda\rangle$ denotes a weight vector of weight λ and r_i is a reflection in a plane orthogonal to the simple root α_i (r_i is an element of the Weyl group). Determination of the sign is done by direct computation or from the prescription given in Ref. 6.

Finally, note that the product of two adjoint representations of SU(5) decomposes as

$$\begin{aligned}(1001)\otimes(1001)&=(2002)\oplus(0102)\oplus(2010)\\&\quad\oplus(0110)\oplus2(1001)\oplus(0000)\,.\end{aligned} \tag{2.14}$$

We choose the two representations (1001) as symmetric and skew-symmetric ones with respect to the permutation of factors on the left side of (2.14).

III. VERMA BASES FOR SU(5)

One of the most difficult tasks in representation theory of semisimple Lie algebras is to construct a basis in the subspaces $V_\Lambda(\Lambda)$ of a representation space V_Λ. In physics this has often been called the "internal state labeling problem." Until very recently there has been no effective method appli-

cable to all representations of semisimple Lie algebras of all types although a large number of particular cases has been solved.

Since Ref. 8, it is known how to construct a basis for any $V_\Lambda(\lambda)$ of a finite-dimensional representation of the classical Lie algebra G (and even for any rank 2 Lie or Kac–Moody algebra) and it appears quite likely that the same will soon be true for any exceptional simple Lie algebra. A complete basis of a subspace $V_\Lambda(\lambda)$ of V_Λ consists of the vectors

$$f_{i_N}^{a_N}\cdots f_{i_2}^{a_2} f_{i_1}^{a_1}|\Lambda\rangle\,, \quad 1\leqslant i_k\leqslant \text{rank}\,, \tag{3.1}$$

where the sequence of the subscripts $i_1,\ldots,i_N$ of the lowering generators is given by a chosen form of the opposite involution of G

$$\text{inv} = r_{i_N} r_{i_{N-1}}\cdots r_{i_2} r_{i_1}\,. \tag{3.2}$$

Here N is the number of positive roots of the algebra, which is also Racah's number of required labels.[11] In particular, a basis vector of an $SU(n)$-irreducible representation space with the highest weight Λ, can be written as an expression of the following type:

$$(f_1^{a_N} f_2^{a_{N-1}}\cdots f_n^{a_{N-n+1}})(f_1^{a_{N-n}}\cdots f_{n-1}^{a_{N-2n+2}})$$
$$\cdots(f_1^{a_3} f_2^{a_2}) f_1^{a_1}|\Lambda\rangle\,, \quad N = n(n+1)/2\,, \tag{3.3}$$

where the f_i are the lowering generators of the $SU(n)$ Lie algebra, and the a_i are integers limited by the following inequalities:

$$\begin{aligned}
&0\leqslant a_1\leqslant m_1\,, \quad 0\leqslant a_2\leqslant m_2+a_1\,, \quad 0\leqslant a_3\leqslant \min(m_2,a_2)\,,\\
&0\leqslant a_4\leqslant m_3+a_2\,, \quad 0\leqslant a_5\leqslant \min(m_3+a_3,a_4)\,,\\
&0\leqslant a_6\leqslant \min(m_3,a_5)\,, \quad 0\leqslant a_7\leqslant m_7+a_4\,,\\
&0\leqslant a_8\leqslant \min(m_4+a_5,a_7)\,, \quad 0\leqslant a_9\leqslant \min(m_4+a_6,a_8)\,,\\
&0\leqslant a_{10}\leqslant \min(m_4,a_9)\,.
\end{aligned} \tag{3.4}$$

These inequalities are a special case, $n=5$, of the general inequalities of Table I in Ref. 8. For $SU(3)$ [or $SU(2)$] one should take, respectively, only the first three (or one) inequalities of (3.4).

Examples of Verma bases for $SU(2)$ and $SU(3)$ representations were shown in Sec. II of Ref. 8. Since our task is to decompose the tensor product (1001) ⊗ (1001) of $SU(5)$, let us construct a basis in the representation space $V_{(1001)}$. It decomposes into a direct sum (2.1) of 21 different subspaces, the first one being the highest weight subspace $V_{(1001)}(1001)$. All but one of the subspaces are one dimensional and are related to $V_{(1001)}(1001)$ by the charge conjugation operators (2.13). The corresponding basis vectors are thus characterized by their $SU(5)$ weights so that one does not even need to use the fact that the inequalities (3.4) allow precisely one choice of exponents $a_1,a_2,\ldots$ for each basis vector. The subspace $V_{(1001)}(0000)$ is quite different. Its dimension is 4 because the multiplicity of the weight (0000) in the weight system $\Omega(\Lambda)$ of the representation $\Lambda=(1001)$ is equal to 4 (see Table II).[12]

TABLE II. Multiplicities of the dominant weights (1001) and (0000) in the representations (2002), (0102), (2010), (0110), and (1001) of $SU(5)$.

	(2002)	(0102)	(2010)	(0110)	(1001)
(1001)	4	3	3	2	1
(0000)	10	6	6	5	4

Let us construct a basis for $V_{(1001)}(0000)$ in detail (see Table III). The weight (0000) is obtained from the highest weight (1001) by subtracting four simple roots of A_4:

$$(1001) - \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4 = (0000)\,. \tag{3.5}$$

Recall that $\alpha_{1+}(2,\bar1,0,0)$, $\alpha_2=(\bar1,2,\bar1,0)$. $\alpha_3=(0,\bar1,2,\bar1)$, $\alpha_4=(0,0,\bar1,2)$. We denote by $|10001\rangle$ the highest weight vector of V_Λ, and by $|0000\rangle_1,|0000\rangle_2,|0000\rangle_3,|0000\rangle_4$ the basis vectors of $V_{(1001)}(0000)$ which we now want to write explicitly using (3.3) and (3.4). These are precisely the vectors (3.3) satisfying

$$\sum_{i=1}^{10} a_i = 4\,. \tag{3.6}$$

Furthermore, for the representation (1001) one has from

TABLE III. Verma bases for (1001) and (0000) weight vectors in the representations (2002), (0102), (2010), (0110), and (1001) of $SU(5)$. The notation represents the weight vector as a constant times a sequence of lowering operators applied to the highest weight.

	(2002)	(0102)	(2010)	(0110)	(1001)
$(1001)_i$	$(1/\sqrt{6})$ [4,3,2,1] $(1/2\sqrt{2})$ [3,4,2,1] $(1/2\sqrt{2})$ [2,3,4,1] $(1/\sqrt{6})$ [1,2,3,4]	$(1/\sqrt{3})$ [4,3,2] $(\frac{1}{2})$ [3,4,2] $(\frac{1}{2})$ [2,3,4]	$(\frac{1}{2})$ [3,2,1] $(\frac{1}{2})$ [2,3,1] $(1/\sqrt{3})$ [1,2,3]	$(1/\sqrt{2})$ [2,3] $(1/\sqrt{2})$ [3,2]	
$(0000)_i$	$(1/16\sqrt{6})$ [1,1,2,2,3,3,4,4] $(\frac{1}{32})$ [1,2,2,3,3,4,4,1] $(\frac{1}{16})$ [1,2,3,3,4,4,2,1] $(\frac{1}{8})$ [1,2,3,4,4,3,2,1] $(1/16\sqrt{6})$ [2,2,3,3,4,4,1,1] $(\frac{1}{32})$ [2,3,3,4,4,2,1,1] $(\frac{1}{16})$ [2,3,4,4,3,2,1,1] $(1/16\sqrt{6})$ [3,3,4,4,2,2,1,1] $(\frac{1}{32})$ [3,4,4,3,2,2,1,1] $(1/16\sqrt{6})$ [4,4,3,3,2,2,1,1]	$(1/8\sqrt{6})$ [1,2,2,3,3,4,4] $(1/8\sqrt{2})$ [1,2,3,3,4,4,2] $(1/4\sqrt{2})$ [1,2,3,4,4,3,2] $(1/4\sqrt{6})$ [2,3,3,4,4,1,2] $(1/4\sqrt{2})$ [2,3,4,4,3,1,2] $(1/2\sqrt{6})$ [3,4,4,2,3,1,2]	$(\frac{1}{4})$ [2,3,4,1,2,3,1] $(1/2\sqrt{2})$ [3,4,1,2,3,2,1] $(1/4\sqrt{2})$ [4,1,2,3,3,2,1] $(\frac{1}{8})$ [3,4,2,3,2,1,1] $(1/8\sqrt{2})$ [4,2,3,3,2,1,1] $(1/8\sqrt{6})$ [4,3,3,2,2,1,1]	$(\frac{1}{2})$ [2,3,4,1,2,3] $(\frac{1}{2})$ [3,4,1,2,3,2] $(\frac{1}{4})$ [4,1,2,3,3,2] $(\frac{1}{2})$ [3,4,2,3,1,2] $(\frac{1}{4})$ [4,2,3,3,1,2]	$(1/\sqrt{2})$ [1,2,3,4] $(1/\sqrt{2})$ [2,3,4,1] $(1/\sqrt{2})$ [3,4,2,1] $(1/\sqrt{2})$ [4,3,2,1]

TABLE IV. Verma bases for the (11), (10), (01), (00) weight vectors in the representations (22), (21), (12), and (11) of SU(3). Note: The lowering operators to be used in this table are f_i $i = 1,2$ of SU(3) (2.3).

	(22)	(21)	(12)	(11)
$(11)_i$	$(1/\sqrt{6})$ [1,2] $(1/\sqrt{6})$ [2,1]			
$(10)_i$		$(1/\sqrt{3})$ [1,2] $(\frac{1}{2})$ [2,1]		
$(01)_i$			$(\frac{1}{2})$ [1,2] $(1/\sqrt{3})$ [2,1]	
$(00)_i$	$(1/4\sqrt{6})$ [2,2,1,1] $(\frac{1}{4})$ [1,2,2,1] $(1/4\sqrt{6})$ [1,1,2,2]			$(1/\sqrt{2})$ [1,2] $(1/\sqrt{2})$ [2,1]

(3.4) that

$$a_3 = a_5 = a_6 = 0\,, \quad a_1 + a_3 + a_6 + a_{10} = 1\,,$$
$$a_2 + a_5 + a_9 = 1\,, \quad a_4 + a_8 = 1\,, \quad a_7 = 1\,. \tag{3.7}$$

This yields the following four linearly independent vectors spanning the $V_{(1001)}(0000)$ subspace

$$|0000\rangle_1 = 2^{-1/2} f_1 f_2 f_3 f_4 |1001\rangle\,,$$
$$|0000\rangle_2 = 2^{-1/2} f_2 f_3 f_4 f_1 |1001\rangle\,,$$
$$|0000\rangle_3 = 2^{-1/2} f_3 f_4 f_2 f_1 |1001\rangle\,, \tag{3.8}$$
$$|0000\rangle_4 = 2^{-1/2} f_4 f_3 f_2 f_1 |1001\rangle\,,$$

Verma bases are not orthogonal in general. The matrix

TABLE V. Inner products of the bases of weight spaces (as given in Table III with same order), with multiplicity greater than 1.

$$\begin{bmatrix}2002\\1001\end{bmatrix} \begin{bmatrix} 1 & 1/\sqrt{3} & 0 & 0 \\ 1/\sqrt{3} & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & 1/\sqrt{3} \\ 0 & 0 & 1/\sqrt{3} & 1 \end{bmatrix} \quad \begin{bmatrix}0102\\0000\end{bmatrix} \begin{bmatrix} 1 & 1/\sqrt{3} & 0 & \frac{1}{3} & 0 & 0 \\ 1/\sqrt{3} & 1 & \frac{1}{2} & 1/\sqrt{3} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 1 & 1/2\sqrt{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 1/\sqrt{3} & 1/2\sqrt{3} & 1 & 1/\sqrt{3} & \frac{1}{3} \\ 0 & \frac{1}{4} & \frac{1}{2} & 1/\sqrt{3} & 1 & 1/\sqrt{3} \\ 0 & 0 & 0 & \frac{1}{3} & 1/\sqrt{3} & 1 \end{bmatrix}$$

$$\begin{bmatrix}2002\\0000\end{bmatrix} \begin{bmatrix} 1 & 3/2\sqrt{6} & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ 3/2\sqrt{6} & 1 & \frac{1}{2} & 0 & 3/2\sqrt{6} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} & 1/\sqrt{6} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 1 & 0 & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & 3/2\sqrt{6} & 1/\sqrt{6} & 0 & 1 & 3/2\sqrt{6} & 0 & \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 3/2\sqrt{6} & 1 & \frac{1}{2} & 3/2\sqrt{6} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{2} & 1 & 1/\sqrt{6} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{6} & 3/2\sqrt{6} & 1/\sqrt{6} & 1 & 3/2\sqrt{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & 3/2\sqrt{6} & 1 & 3/2\sqrt{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 3/2\sqrt{6} & 1 \end{bmatrix}$$

$$\begin{bmatrix}2010\\0000\end{bmatrix} \begin{bmatrix} 1 & 1/\sqrt{2} & 1/4\sqrt{2} & \frac{3}{8} & 1/2\sqrt{2} & 1/2\sqrt{6} \\ 1/\sqrt{2} & 1 & \frac{1}{2} & 1/2\sqrt{2} & \frac{1}{4} & 0 \\ 1/4\sqrt{2} & \frac{1}{2} & 1 & 1/4\sqrt{2} & \frac{1}{2} & 0 \\ \frac{3}{8} & 1/2\sqrt{2} & 1/4\sqrt{2} & 1 & 1/\sqrt{2} & 2/\sqrt{6} \\ 1/2\sqrt{2} & \frac{1}{4} & \frac{1}{2} & 1/\sqrt{2} & 1 & 1/\sqrt{3} \\ 1/2\sqrt{6} & 0 & 0 & 2/\sqrt{6} & 1/\sqrt{3} & 1 \end{bmatrix}$$

$$\begin{bmatrix}0102\\1001\end{bmatrix} \begin{bmatrix} 1 & 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix} \quad \begin{bmatrix}2010\\1001\end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} & 1 \end{bmatrix}$$

$$\begin{bmatrix}0110\\1001\end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \quad \begin{bmatrix}0110\\0000\end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & 1 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}$$

of inner products of the basis vectors (3.8) is

$$[{}_i\langle 0000|0000\rangle_k] = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}. \tag{3.9}$$

The same procedure can be applied to any of the weight subspaces of dimension > 1. The bases in dominant weight subspaces relevant to our problem are given in Table III. The basis vectors in the table are written in a shortened way. Thus a vector

$$|1001\rangle_1 \in V_{(2002)}(1001) \subset V_{(2002)},$$

is found as

$$|1000\rangle_1 = 6^{-1/2}[4,3,2,1] \equiv 6^{-1/2} f_4 f_3 f_2 f_1 |2002\rangle. \tag{3.10}$$

Note that, e.g., for the same representation (2002) and dominant weight (1001), there are four vectors $|1001\rangle_i$, $i = 1,2,3,4$. Thus the first row corresponds to $|1001\rangle_1$, the second one to $|1001\rangle_2$, etc.

We will also encounter some SU(3)-weights with multiplicities higher than 1. In Table IV we give the bases for these cases, namely for the weights (11), (10), (01), and (00) of the representations with highest weights (22), (21), (12), and (11). In order to facilitate the use of Verma bases we also give the inner products among the vectors of those bases in Table V.

IV. CLEBSCH–GORDAN COEFFICIENTS IN THE GROUP BASIS

In this section we describe briefly how to compute CGC related to the tensor product (1001) ⊗ (1001). To do that, we will give some examples which illustrate how the method works.

First, we take an example from Table VI,

$$\begin{bmatrix} 2 & 0 & 0 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix} = (1001)(1001), \tag{4.1}$$

$$\begin{bmatrix} 2 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \end{bmatrix} = 2^{-1/2}\{(1001)(\bar{1}101) + (\bar{1}101)(1001)\}, \tag{4.2}$$

$$\begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 2 \end{bmatrix} = 2^{-1/2}\{(1001)(\bar{1}101) - (\bar{1}101)(1001)\}. \tag{4.3}$$

[Remark that ( ) ( ) is a short expression for ( ) ⊗ ( ).]

In this case the multiplicity for these three weights is 1. The vectors on the left side are written with the highest weight above the weight. That is sufficient to underline the irreducible subspace to which the vector belongs. On the

TABLE VI. CGC of the weight vectors with dominant weight different from (0000) in the group basis. The last row, when it exists under the list of weight vectors gives normalizing factors. In that case the corresponding whole column must be multiplied by that factor.

	(2002)											
	(2002)	(2002)	(0102)									
(1001)(1001)	1	(0102)	(0102)									
(1001)($\bar{1}$101)		$1/\sqrt{2}$	$1/\sqrt{2}$	(2002)	(2010)							
($\bar{1}$101)(1001)		$1/\sqrt{2}$	$-1/\sqrt{2}$	(2010)	(2010)							
		(1001)(10$\bar{1}$)		$1/\sqrt{2}$	$1/\sqrt{2}$	(2002)	(0102)	(2010)	(0110)			
		(10$\bar{1}$)(1001)		$1/\sqrt{2}$	$-1/\sqrt{2}$	(0110)	(0110)	(0110)	(0110)			
				(1001)($\bar{1}$1$\bar{1}$)		$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$			
				($\bar{1}$1$\bar{1}$)(1001)		$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$			
				($\bar{1}$101)(10$\bar{1}$)		$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$			
				(10$\bar{1}$)($\bar{1}$101)		$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$			

	(2002)				(0102)			(2010)			(0110)		$(1001)_s$	$(1001)_a$
	$(1001)_1$	$(1001)_2$	$(1001)_3$	$(1001)_4$	$(1001)_1$	$(1001)_2$	$(1001)_3$	$(1001)_1$	$(1001)_2$	$(1001)_3$	$(1001)_1$	$(1001)_2$	$(1001)_s$	$(1001)_a$
($\bar{1}$101)(2$\bar{1}$00)			$\frac{1}{2}$	$1/\sqrt{2}$			$-\frac{1}{2}$		$\frac{1}{2}$	$1/\sqrt{2}$	$-\frac{1}{2}$		$5/\sqrt{2}$	$1/\sqrt{2}$
(2$\bar{1}$00)($\bar{1}$101)			$\frac{1}{2}$	$1/\sqrt{2}$			$\frac{1}{2}$		$-\frac{1}{2}$	$-1/\sqrt{2}$	$-\frac{1}{2}$		$5/\sqrt{2}$	$-1/\sqrt{2}$
(101$\bar{1}$)(00$\bar{1}$2)	$1/\sqrt{2}$	$\frac{1}{2}$			$1/\sqrt{2}$	$\frac{1}{2}$		$-\frac{1}{2}$				$-\frac{1}{2}$	$5/\sqrt{2}$	$-1/\sqrt{2}$
(00$\bar{1}$2)(101$\bar{1}$)	$1/\sqrt{2}$	$\frac{1}{2}$			$-1/\sqrt{2}$	$-\frac{1}{2}$		$\frac{1}{2}$				$-\frac{1}{2}$	$5/\sqrt{2}$	$1/\sqrt{2}$
(0$\bar{1}$11)(11$\bar{1}$0)		$\frac{1}{2}$	$\frac{1}{2}$			$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
(11$\bar{1}$0)(0$\bar{1}$11)		$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(1001)(0000)_1$				1						1			-3	-1
$(0000)_1(1001)$				1						-1			-3	1
$(1001)(0000)_2$			$1/\sqrt{2}$				$1/\sqrt{2}$		$1/\sqrt{2}$		$1/\sqrt{2}$		1	1
$(0000)_2(1001)$			$1/\sqrt{2}$				$-1/\sqrt{2}$		$-1/\sqrt{2}$		$1/\sqrt{2}$		1	-1
$(1001)(0000)_3$		$1/\sqrt{2}$				$1/\sqrt{2}$		$1/\sqrt{2}$				$1/\sqrt{2}$	1	-1
$(0000)_3(1001)$		$1/\sqrt{2}$				$-1/\sqrt{2}$		$-1/\sqrt{2}$				$1/\sqrt{2}$	1	1
$(1001)(0000)_4$	1				1								-3	1
$(0000)_4(1001)$	1				-1								-3	-1
	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{105}$	$1/\sqrt{5}$

TABLE VII. CGC of the vectors with weight (0000) in the group basis. Normalizing factors for the columns are given in the last row.

	(2002)									
	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_6$	$(0000)_7$	$(0000)_8$	$(0000)_9$	$(0000)_{10}$
$(0000)_1(0000)_1$	2									
$(0000)_1(0000)_2$		$\frac{1}{2}$								
$(0000)_2(0000)_1$		$\frac{1}{2}$								
$(0000)_1(0000)_3$			$\frac{1}{2}$							
$(0000)_3(0000)_1$			$\frac{1}{2}$							
$(0000)_1(0000)_4$				$\frac{1}{2}$						
$(0000)_4(0000)_1$				$\frac{1}{2}$						
$(0000)_2(0000)_2$					2					
$(0000)_2(0000)_3$						$\frac{1}{2}$				
$(0000)_3(0000)_2$						$\frac{1}{2}$				
$(0000)_2(0000)_4$							$\frac{1}{2}$			
$(0000)_4(0000)_2$							$\frac{1}{2}$			
$(0000)_3(0000)_3$								2		
$(0000)_3(0000)_4$									$\frac{1}{2}$	
$(0000)_4(0000)_3$									$\frac{1}{2}$	
$(0000)_4(0000)_4$										2
$(1001)(\bar{1}00\bar{1})$				$\frac{1}{4}$						
$(\bar{1}00\bar{1})(1001)$				$\frac{1}{4}$						
$(\bar{1}101)(1\bar{1}0\bar{1})$				$\frac{1}{4}$			$\frac{1}{4}$			
$(1\bar{1}0\bar{1})(\bar{1}101)$				$\frac{1}{4}$			$\frac{1}{4}$			
$(101\bar{1})(\bar{1}0\bar{1}1)$			$\frac{1}{4}$	$\frac{1}{4}$						
$(\bar{1}0\bar{1}1)(101\bar{1})$			$\frac{1}{4}$	$\frac{1}{4}$						
$(0\bar{1}11)(01\bar{1}\bar{1})$							$\frac{1}{4}$		$\frac{1}{4}$	
$(01\bar{1}\bar{1})(0\bar{1}11)$							$\frac{1}{4}$		$\frac{1}{4}$	
$(\bar{1}11\bar{1})(1\bar{1}\bar{1}1)$			$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$			
$(1\bar{1}\bar{1}1)(\bar{1}11\bar{1})$			$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$			
$(11\bar{1}0)(\bar{1}\bar{1}10)$		$\frac{1}{4}$	$\frac{1}{4}$							
$(\bar{1}\bar{1}10)(11\bar{1}0)$		$\frac{1}{4}$	$\frac{1}{4}$							
$(2\bar{1}00)(\bar{2}100)$	1	$\frac{1}{4}$								
$(\bar{2}100)(2\bar{1}00)$	1	$\frac{1}{4}$								
$(\bar{1}2\bar{1}0)(1\bar{2}10)$		$\frac{1}{4}$	$\frac{1}{4}$		1	$\frac{1}{4}$				
$(1\bar{2}10)(\bar{1}2\bar{1}0)$		$\frac{1}{4}$	$\frac{1}{4}$		1	$\frac{1}{4}$				
$(0\bar{1}2\bar{1})(01\bar{2}1)$						$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{1}{4}$	
$(01\bar{2}1)(0\bar{1}2\bar{1})$						$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{1}{4}$	
$(00\bar{1}2)(001\bar{2})$									$\frac{1}{4}$	1
$(001\bar{2})(00\bar{1}2)$									$\frac{1}{4}$	1
	$1/\sqrt{6}$				$1/\sqrt{6}$			$1/\sqrt{6}$		$1/\sqrt{6}$

	(0102)						(2010)					
	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_6$	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_6$
$(0000)_1(0000)_1$												
$(0000)_1(0000)_2$	1						$-\frac{1}{2}$					
$(0000)_2(0000)_1$	-1						$\frac{1}{2}$					
$(0000)_1(0000)_3$		$\frac{1}{2}$						$-\frac{1}{2}$				
$(0000)_3(0000)_1$		$-\frac{1}{2}$						$\frac{1}{2}$				
$(0000)_1(0000)_4$			$\frac{1}{2}$						$-\frac{1}{2}$			
$(0000)_4(0000)_1$			$-\frac{1}{2}$						$\frac{1}{2}$			
$(0000)_2(0000)_2$												
$(0000)_2(0000)_3$				1			$-\frac{1}{2}$			$-\frac{1}{2}$		
$(0000)_3(0000)_2$				-1			$\frac{1}{2}$			$\frac{1}{2}$		
$(0000)_2(0000)_4$					$\frac{1}{2}$						$-\frac{1}{2}$	
$(0000)_4(0000)_2$					$-\frac{1}{2}$						$\frac{1}{2}$	
$(0000)_3(0000)_3$												
$(0000)_3(0000)_4$						1				$-\frac{1}{2}$		-1
$(0000)_4(0000)_3$						-1				$\frac{1}{2}$		1
$(0000)_4(0000)_4$												
$(1001)(\bar{1}00\bar{1})$			$\frac{1}{4}$						$\frac{1}{4}$			
$(\bar{1}00\bar{1})(1001)$			$-\frac{1}{4}$						$-\frac{1}{4}$			
$(\bar{1}101)(1\bar{1}0\bar{1})$			$\frac{1}{4}$		$\frac{1}{4}$				$\frac{1}{4}$		$\frac{1}{4}$	
$(1\bar{1}0\bar{1})(\bar{1}101)$			$-\frac{1}{4}$		$-\frac{1}{4}$				$-\frac{1}{4}$		$-\frac{1}{4}$	
$(101\bar{1})(\bar{1}0\bar{1}1)$		$\frac{1}{4}$	$\frac{1}{4}$					$\frac{1}{4}$	$\frac{1}{4}$			
$(\bar{1}0\bar{1}1)(101\bar{1})$		$-\frac{1}{4}$	$-\frac{1}{4}$					$-\frac{1}{4}$	$-\frac{1}{4}$			
$(0\bar{1}11)(01\bar{1}\bar{1})$					$\frac{1}{4}$	$\frac{1}{2}$				$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

TABLE VII. (*Continued.*)

	(0102)						(2010)					
	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_6$	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_6$
$(01\bar{1}\bar{1})(0\bar{1}11)$					$-\frac{1}{4}$	$-\frac{1}{2}$				$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$
$(\bar{1}11\bar{1})(1\bar{1}\bar{1}1)$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	
$(1\bar{1}\bar{1}1)(\bar{1}11\bar{1})$		$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$		$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	
$(11\bar{1}0)(\bar{1}\bar{1}10)$	$\frac{1}{2}$	$\frac{1}{4}$					$\frac{1}{4}$	$\frac{1}{4}$				
$(\bar{1}\bar{1}10)(11\bar{1}0)$	$-\frac{1}{2}$	$-\frac{1}{4}$					$-\frac{1}{4}$	$-\frac{1}{4}$				
$(2\bar{1}00)(\bar{2}100)$	$\frac{1}{2}$						$\frac{1}{4}$					
$(\bar{2}100)(2\bar{1}00)$	$-\frac{1}{2}$						$-\frac{1}{4}$					
$(\bar{1}2\bar{1}0)(1\bar{2}10)$	$\frac{1}{2}$	$\frac{1}{4}$		$\frac{1}{2}$			$\frac{1}{2}$	$\frac{1}{4}$		$\frac{1}{4}$		
$(1\bar{2}10)(\bar{1}2\bar{1}0)$	$-\frac{1}{2}$	$-\frac{1}{4}$		$-\frac{1}{2}$			$-\frac{1}{2}$	$-\frac{1}{4}$		$-\frac{1}{4}$		
$(0\bar{1}2\bar{1})(01\bar{2}1)$				$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$			$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$(01\bar{2}1)(0\bar{1}2\bar{1})$				$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$			$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{2}$
$(00\bar{1}2)(001\bar{2})$						$\frac{1}{2}$				$\frac{1}{4}$		$\frac{1}{2}$
$(001\bar{2})(00\bar{1}2)$						$-\frac{1}{2}$				$-\frac{1}{4}$		$-\frac{1}{2}$
	$1/\sqrt{3}$			$1/\sqrt{3}$		$1/\sqrt{3}$	$1/\sqrt{2}$			$1/\sqrt{2}$		$1/\sqrt{3}$

	(0110)					$(1001)_s$				$(1001)_a$				(0000)
	$(0000)_1$	$(0000)_2$	$(0000)_3$	$(0000)_4$	$(0000)_5$	$(0000)_{1s}$	$(0000)_{2s}$	$(0000)_{3s}$	$(0000)_{4s}$	$(0000)_{1a}$	$(0000)_{2a}$	$(0000)_{3a}$	$(0000)_{4a}$	(0000)
$(0000)_1(0000)_1$						-6								$\frac{4}{5}$
$(0000)_1(0000)_2$	$-\frac{1}{2}$					6	2							$-\frac{3}{5}$
$(0000)_2(0000)_1$	$-\frac{1}{2}$					6	2							$-\frac{3}{5}$
$(0000)_1(0000)_3$		$-\frac{1}{2}$				-4		2						$\frac{2}{5}$
$(0000)_3(0000)_1$		$-\frac{1}{2}$				-4		2						$\frac{2}{5}$
$(0000)_1(0000)_4$			$-\frac{1}{2}$			2			2					$-\frac{1}{5}$
$(0000)_4(0000)_1$			$-\frac{1}{2}$			2			2					$-\frac{1}{5}$
$(0000)_2(0000)_2$	1						2							$\frac{6}{5}$
$(0000)_2(0000)_3$	$-\frac{1}{2}$			$-\frac{1}{2}$			-4	-4						$-\frac{4}{5}$
$(0000)_3(0000)_2$	$-\frac{1}{2}$			$-\frac{1}{2}$			-4	-4						$-\frac{4}{5}$
$(0000)_2(0000)_4$					$-\frac{1}{2}$		2		-4					$\frac{2}{5}$
$(0000)_4(0000)_2$					$-\frac{1}{2}$		2		-4					$\frac{2}{5}$
$(0000)_3(0000)_3$				1				2						$\frac{6}{5}$
$(0000)_3(0000)_4$				$-\frac{1}{2}$				2	6					$-\frac{3}{5}$
$(0000)_4(0000)_3$				$-\frac{1}{2}$				2	6					$-\frac{3}{5}$
$(0000)_4(0000)_4$									-6					$\frac{4}{5}$
$(1001)(\bar{1}00\bar{1})$			$\frac{1}{4}$			$-\frac{5}{2}$			$-\frac{5}{2}$	$\frac{1}{2}$			$-\frac{1}{2}$	$\frac{1}{2}$
$(\bar{1}00\bar{1})(1001)$			$\frac{1}{4}$			$-\frac{5}{2}$			$-\frac{5}{2}$	$-\frac{1}{2}$			$\frac{1}{2}$	$\frac{1}{2}$
$(\bar{1}101)(1\bar{1}0\bar{1})$			$\frac{1}{4}$		$\frac{1}{4}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$\frac{5}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$
$(1\bar{1}0\bar{1})(\bar{1}101)$			$\frac{1}{4}$		$\frac{1}{4}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$\frac{5}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$
$(101\bar{1})(\bar{1}0\bar{1}1)$		$\frac{1}{4}$	$\frac{1}{4}$			$\frac{5}{2}$		$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{1}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(\bar{1}0\bar{1}1)(101\bar{1})$		$\frac{1}{4}$	$\frac{1}{4}$			$\frac{5}{2}$		$-\frac{5}{2}$	$-\frac{5}{2}$	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(0\bar{1}11)(01\bar{1}\bar{1})$				$\frac{1}{4}$	$\frac{1}{4}$		$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
$(01\bar{1}\bar{1})(0\bar{1}11)$				$\frac{1}{4}$	$\frac{1}{4}$		$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$(\bar{1}11\bar{1})(1\bar{1}\bar{1}1)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{5}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$(1\bar{1}\bar{1}1)(\bar{1}11\bar{1})$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
$(11\bar{1}0)(\bar{1}\bar{1}10)$	$\frac{1}{4}$	$\frac{1}{4}$				$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$		$\frac{1}{2}$
$(\bar{1}\bar{1}10)(11\bar{1}0)$	$\frac{1}{4}$	$\frac{1}{4}$				$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{2}$
$(2\bar{1}00)(\bar{2}100)$	$\frac{1}{4}$						$-\frac{5}{2}$			-1	$-\frac{1}{2}$			$-\frac{1}{2}$
$(\bar{2}100)(2\bar{1}00)$	$\frac{1}{4}$						$-\frac{5}{2}$			1	$\frac{1}{2}$			$-^2$
$(\bar{1}2\bar{1}0)(1\bar{2}10)$		$\frac{1}{4}$		$\frac{1}{4}$		$-\frac{5}{2}$		$\frac{5}{2}$		$\frac{1}{2}$	1	$\frac{1}{2}$		$-\frac{1}{2}$
$(1\bar{2}10)(\bar{1}2\bar{1}0)$		$\frac{1}{4}$		$\frac{1}{4}$		$-\frac{5}{2}$		$\frac{5}{2}$		$-\frac{1}{2}$	-1	$-\frac{1}{2}$		$-\frac{1}{2}$
$(0\bar{1}2\bar{1})(01\bar{2}1)$	$\frac{1}{4}$				$\frac{1}{4}$		$\frac{5}{2}$		$-\frac{5}{2}$		$1\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{2}$
$(01\bar{2}1)(0\bar{1}2\bar{1})$	$\frac{1}{4}$				$\frac{1}{4}$		$\frac{5}{2}$		$-\frac{5}{2}$		$\frac{1}{2}$	1	$\frac{1}{2}$	$-\frac{1}{2}$
$(00\bar{1}2)(001\bar{2})$				$\frac{1}{4}$				$-\frac{5}{2}$				$\frac{1}{2}$	1	$-\frac{1}{2}$
$(001\bar{2})(00\bar{1}2)$				$\frac{1}{4}$				$-\frac{5}{2}$				$-\frac{1}{2}$	-1	$-\frac{1}{2}$
						$1/\sqrt{105}$	$1/\sqrt{105}$	$1/\sqrt{105}$	$1/\sqrt{105}$	$1/\sqrt{5}$	$1/\sqrt{5}$	$1/\sqrt{5}$	$1/\sqrt{5}$	$1/\sqrt{6}$

TABLE VIII. CGC of the weight vectors with nonzero dominant weight in the subgroup basis (first part). Normalizing factors for the columns are given in the last row.

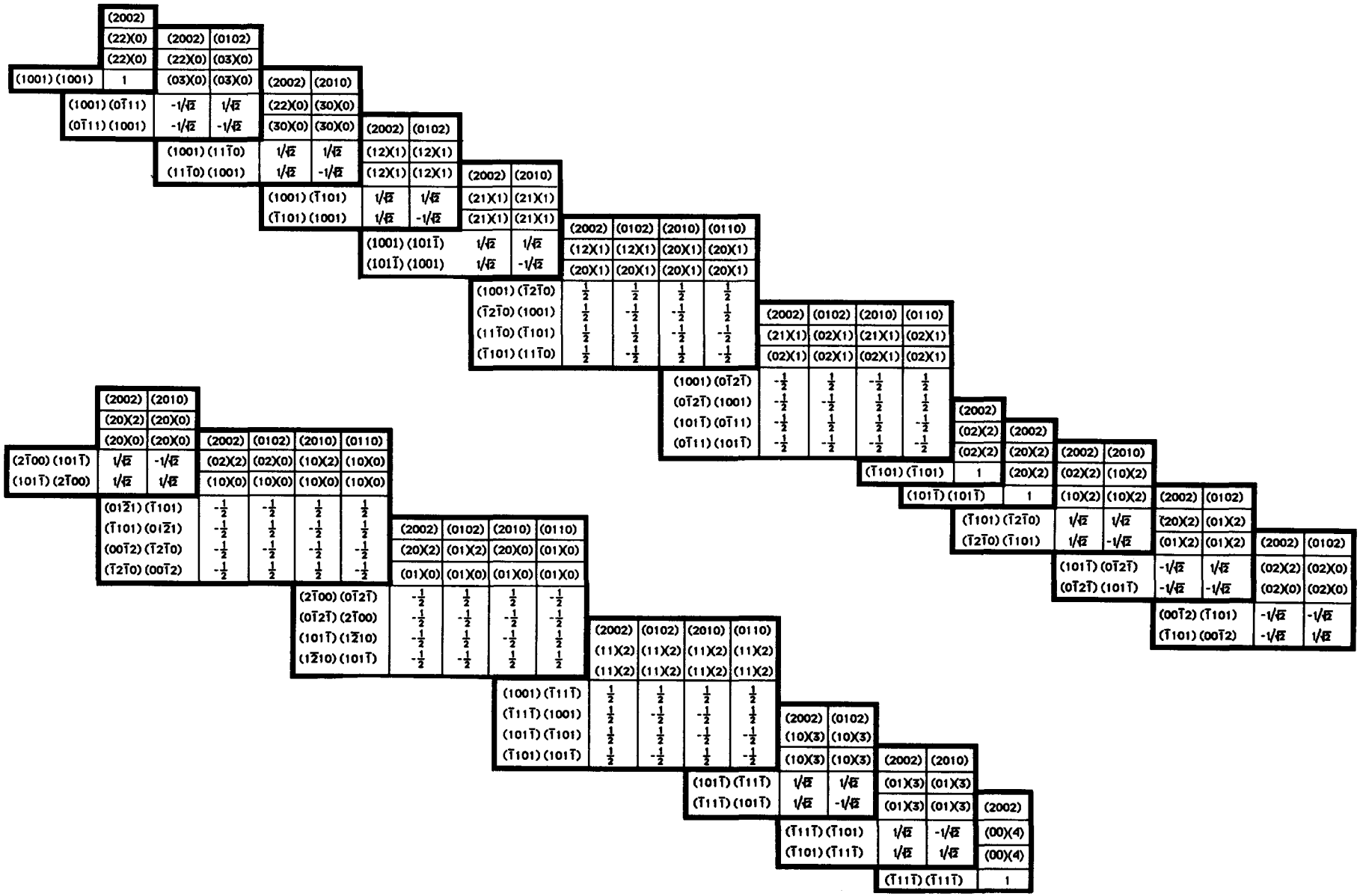


TABLE IX. CGC of the weight vectors with nonzero dominant weight in the subgroup basis (second part). Normalizing factors for the columns are given in the last row.

	(2002)				(0102)			(2010)			(0110)		$(1001)_s$	$(1001)_a$
	(12)(1)		(01)(3)	(01)(1)	(12)(1)		(01)(1)	(20)(1)	(01)(3)	(01)(1)	(20)(1)	(01)(1)	(01)(1)	(01)(1)
	$(01)_1(1)$	$(01)_2(1)$	(01)(1)	(01)(1)	$(01)_1(1)$	$(01)_2(1)$	(01)(1)	(01)(1)	(01)(1)	(01)(1)	(01)(1)	(01)(1)	(01)(1)	(01)(1)
$(1001)(\bar{2}100)$	$\frac{1}{2}$			$\frac{3}{4}$	$\frac{1}{2}$		$-\frac{1}{4}$	$\frac{1}{2}$		$3/2\sqrt{2}$	$\frac{1}{2}$	$-1/2\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(\bar{2}100)(1001)$	$\frac{1}{2}$			$\frac{3}{4}$	$-\frac{1}{2}$		$\frac{1}{4}$	$-\frac{1}{2}$		$-3/2\sqrt{2}$	$\frac{1}{2}$	$-1/2\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(0\bar{1}11)(\bar{1}2\bar{1}0)$	$-\frac{1}{2}$	$-1/\sqrt{2}$		$\frac{3}{4}$	$-\frac{1}{2}$	$-1/\sqrt{2}$	$-\frac{1}{4}$	$-\frac{1}{2}$		$3/2\sqrt{2}$	$-\frac{1}{2}$	$-1/2\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(\bar{1}2\bar{1}0)(0\bar{1}11)$	$-\frac{1}{2}$	$-1/\sqrt{2}$		$\frac{3}{4}$	$\frac{1}{2}$	$1/\sqrt{2}$	$\frac{1}{4}$	$\frac{1}{2}$		$-3/2\sqrt{2}$	$-\frac{1}{2}$	$-1/2\sqrt{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(\bar{1}11\bar{1})(00\bar{1}2)$			$-1/\sqrt{2}$	1			1		$1/\sqrt{2}$	$-1/\sqrt{2}$		$-1/\sqrt{2}$	$5/\sqrt{2}$	$-1/\sqrt{2}$
$(00\bar{1}2)(\bar{1}11\bar{1})$			$-1/\sqrt{2}$	1			-1		$-1/\sqrt{2}$	$1/\sqrt{2}$		$-1/\sqrt{2}$	$5/\sqrt{2}$	$1/\sqrt{2}$
$(\bar{1}101)(0000)_1$	$1/\sqrt{2}$			$3/2\sqrt{2}$	$-1/\sqrt{2}$		$1/2\sqrt{2}$	$1/\sqrt{2}$		$\frac{3}{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	2	
$(0000)_1(\bar{1}101)$	$1/\sqrt{2}$			$3/2\sqrt{2}$	$1/\sqrt{2}$		$-1/2\sqrt{2}$	$-1/\sqrt{2}$		$-\frac{3}{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	2	
$(\bar{1}101)(0000)_2$	$-1/\sqrt{2}$		1	$5/2\sqrt{2}$	$1/\sqrt{2}$		$-1/2\sqrt{2}$	$-1/\sqrt{2}$	1	$\frac{1}{2}$	$1/\sqrt{2}$	$-\frac{1}{2}$	1	1
$(0000)_2(\bar{1}101)$	$-1/\sqrt{2}$		1	$5/2\sqrt{2}$	$-1/\sqrt{2}$		$1/2\sqrt{2}$	$1/\sqrt{2}$	-1	$-\frac{1}{2}$	$1/\sqrt{2}$	$-1/\sqrt{2}$	1	-1
$(\bar{1}101)(0000)_3$		-1	-1	$-1/\sqrt{2}$		1	$1/\sqrt{2}$		-1	1		1	1	-1
$(0000)_3(\bar{1}101)$		-1	-1	$-1/\sqrt{2}$		-1	$-1/\sqrt{2}$		1	-1		1	1	1
$(\bar{1}101)(0000)_4$		1		$3/\sqrt{2}$		-1	$1/\sqrt{2}$						-3	1
$(0000)_4(\bar{1}101)$		1		$3/\sqrt{2}$		1	$-1/\sqrt{2}$						-3	-1
	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{21}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{5}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{15}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{105}$	$1/\sqrt{5}$

	(2002)				(0102)			(2010)			(0110)		$(1001)_s$	$(1001)_a$
	(21)(1)		(10)(3)	(10)(1)	(02)(1)	(10)(3)	(10)(1)	(21)(1)		(10)(1)	(02)(1)	(10)(1)	(10)(1)	(10)(1)
	$(10)_1(1)$	$(10)_2(1)$	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)	$(10)_1(1)$	$(10)_2(1)$	(10)(1)	(10)(1)	(10)(1)	(10)(1)	(10)(1)
$(1001)(00\bar{1}2)$		$\frac{1}{2}$		$\frac{3}{4}$	$-\frac{1}{2}$		$3/2\sqrt{2}$		$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{2}$	$1/2\sqrt{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(00\bar{1}2)(1001)$		$\frac{1}{2}$		$\frac{3}{4}$	$\frac{1}{2}$		$-3/2\sqrt{2}$		$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$1/2\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(11\bar{1}0)(0\bar{1}2\bar{1})$	$-1/\sqrt{2}$	$-\frac{1}{2}$		$\frac{3}{4}$	$\frac{1}{2}$		$3/2\sqrt{2}$	$-1/\sqrt{2}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$1/2\sqrt{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(0\bar{1}2\bar{1})(11\bar{1}0)$	$-1/\sqrt{2}$	$-\frac{1}{2}$		$\frac{3}{4}$	$-\frac{1}{2}$		$-3/2\sqrt{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{2}$	$1/2\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(\bar{1}11\bar{1})(2\bar{1}00)$			$1/\sqrt{2}$	1		$-1/\sqrt{2}$	$-1/\sqrt{2}$			-1		$1/\sqrt{2}$	$5/\sqrt{2}$	$1/\sqrt{2}$
$(2\bar{1}00)(\bar{1}11\bar{1})$			$1/\sqrt{2}$	1		$1/\sqrt{2}$	$1/\sqrt{2}$			1		$1/\sqrt{2}$	$5/\sqrt{2}$	$-1/\sqrt{2}$
$(101\bar{1})(0000)_1$	1			$3/\sqrt{2}$				-1		$-1/\sqrt{2}$			-3	-1
$(0000)_1(101\bar{1})$	1			$3/\sqrt{2}$				1		$1/\sqrt{2}$			-3	1
$(101\bar{1})(0000)_2$	-1		1	$-1/\sqrt{2}$		1	1	1		$-1/\sqrt{2}$		-1	1	1
$(0000)_2(101\bar{1})$	-1		1	$-1/\sqrt{2}$		-1	-1	-1		$1/\sqrt{2}$		-1	1	-1
$(101\bar{1})(0000)_3$		$-1/\sqrt{2}$	-1	$5/2\sqrt{2}$	$1/\sqrt{2}$	-1	$\frac{1}{2}$		$1/\sqrt{2}$	$1/2\sqrt{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	1	-1
$(0000)_3(101\bar{1})$		$-1/\sqrt{2}$	-1	$5/2\sqrt{2}$	$-1/\sqrt{2}$	1	$-\frac{1}{2}$		$-1/\sqrt{2}$	$-1/2\sqrt{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	1	1
$(101\bar{1})(0000)_4$		$1/\sqrt{2}$		$3/2\sqrt{2}$	$-1/\sqrt{2}$		$\frac{3}{2}$		$-1/\sqrt{2}$	$-1/2\sqrt{2}$	$1/\sqrt{2}$	$-\frac{1}{2}$	2	
$(0000)_4(101\bar{1})$		$1/\sqrt{2}$		$3/2\sqrt{2}$	$1/\sqrt{2}$		$-\frac{3}{2}$		$1/\sqrt{2}$	$1/2\sqrt{2}$	$1/\sqrt{2}$	$-\frac{1}{2}$	2	
	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{21}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{15}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{5}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{105}$	$1/\sqrt{5}$

	(2002)				(0102)			(2010)			(0110)		$(1001)_s$	$(1001)_a$
	(11)(2)		(00)(4)	(00)(2)	(11)(2)		(00)(2)	(11)(2)		(00)(2)	(11)(2)		(00)(2)	(00)(2)
	$(00)_1(2)$	$(00)_2(2)$	(00)(2)	(00)(2)	$(00)_1(2)$	$(00)_2(2)$	(00)(2)	$(00)_1(2)$	$(00)_2(2)$	(00)(2)	$(00)_1(2)$	$(00)_2(2)$	(00)(2)	(00)(2)
$(\bar{1}101)(00\bar{1}2)$		$\frac{1}{2}$		1		$-\frac{1}{2}$	$-1/\sqrt{2}$		$\frac{1}{2}$	$-1/\sqrt{2}$		$-\frac{1}{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(00\bar{1}2)(\bar{1}101)$		$\frac{1}{2}$		1		$\frac{1}{2}$	$1/\sqrt{2}$		$-\frac{1}{2}$	$1/\sqrt{2}$		$-\frac{1}{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(\bar{1}2\bar{1}0)(0\bar{1}2\bar{1})$	$-\frac{1}{2}$	$-\frac{1}{2}$		1	$\frac{1}{2}$	$\frac{1}{2}$	$-1/\sqrt{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(0\bar{1}2\bar{1})(\bar{1}2\bar{1}0)$	$-\frac{1}{2}$	$-\frac{1}{2}$		1	$-\frac{1}{2}$	$-\frac{1}{2}$	$1/\sqrt{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1/\sqrt{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(101\bar{1})(\bar{2}100)$	$\frac{1}{2}$			1	$\frac{1}{2}$		$1/\sqrt{2}$	$-\frac{1}{2}$		$1/\sqrt{2}$	$-\frac{1}{2}$		$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(\bar{2}100)(101\bar{1})$	$\frac{1}{2}$			1	$-\frac{1}{2}$		$-1/\sqrt{2}$	$\frac{1}{2}$		$-1/\sqrt{2}$	$-\frac{1}{2}$		$-5/\sqrt{2}$	$1/\sqrt{2}$
$(\bar{1}11\bar{1})(0000)_1$	$1/\sqrt{2}$			$2/\sqrt{2}$	$-1/\sqrt{2}$		-1	$-1/\sqrt{2}$		1	$1/\sqrt{2}$		2	
$(0000)_1(\bar{1}11\bar{1})$	$1/\sqrt{2}$			$2/\sqrt{2}$	$1/\sqrt{2}$		1	$1/\sqrt{2}$		-1	$1/\sqrt{2}$		2	
$(\bar{1}11\bar{1})(0000)_2$	$-1/\sqrt{2}$		1	$1/\sqrt{2}$	$1/\sqrt{2}$		1	$1/\sqrt{2}$		2	$-1/\sqrt{2}$		1	1
$(0000)_2(\bar{1}11\bar{1})$	$-1/\sqrt{2}$		1	$1/\sqrt{2}$	$-1/\sqrt{2}$		-1	$-1/\sqrt{2}$		-2	$-1/\sqrt{2}$		1	-1
$(\bar{1}11\bar{1})(0000)_3$		$-1/\sqrt{2}$	-1	$1/\sqrt{2}$		$1/\sqrt{2}$	-2		$1/\sqrt{2}$	-1		$-1/\sqrt{2}$	1	-1
$(0000)_3(\bar{1}11\bar{1})$		$-1/\sqrt{2}$	-1	$1/\sqrt{2}$		$-1/\sqrt{2}$	2		$-1/\sqrt{2}$	1		$-1/\sqrt{2}$	1	1
$(\bar{1}11\bar{1})(0000)_4$		$1/\sqrt{2}$		$2/\sqrt{2}$		$-1/\sqrt{2}$	-1		$-1/\sqrt{2}$	1		$1/\sqrt{2}$	2	
$(0000)_4(\bar{1}11\bar{1})$		$1/\sqrt{2}$		$2/\sqrt{2}$		$1/\sqrt{2}$	1		$1/\sqrt{2}$	-1		$1/\sqrt{2}$	2	
	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{21}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{15}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{15}$	$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{105}$	$1/\sqrt{5}$

TABLE IX. (*Continued.*)

	(2002)				(0102)			(2010)			(0110)		$(1001)_s$	$(1001)_a$
	(22)(0)		(11)(2)	(11)(0)	(03)(0)	(11)(2)	(11)(0)	(30)(0)	(11)(2)	(11)(0)	(11)(2)	(11)(0)	(11)(0)	(11)(0)
	$(11)_1(0)$	$(11)_2(0)$	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)	(11)(0)
$(\bar{1}101)(2\bar{1}00)$			$\frac{1}{2}$	$5/2\sqrt{2}$		$-\frac{1}{2}$	$-3/2\sqrt{2}$		$\frac{1}{2}$	$3/2\sqrt{2}$	$-\frac{1}{2}$	$-1/2\sqrt{2}$	$5/\sqrt{2}$	$1/\sqrt{2}$
$(2\bar{1}00)(\bar{1}101)$			$\frac{1}{2}$	$5/2\sqrt{2}$		$\frac{1}{2}$	$3/2\sqrt{2}$		$-\frac{1}{2}$	$-3/2\sqrt{2}$	$-\frac{1}{2}$	$-1/2\sqrt{2}$	$5/\sqrt{2}$	$-1/\sqrt{2}$
$(101\bar{1})(00\bar{1}2)$			$-\frac{1}{2}$	$5/2\sqrt{2}$		$-\frac{1}{2}$	$3/2\sqrt{2}$		$\frac{1}{2}$	$-3/2\sqrt{2}$	$\frac{1}{2}$	$-1/2\sqrt{2}$	$5/\sqrt{2}$	$-1/\sqrt{2}$
$(00\bar{1}2)(101\bar{1})$			$-\frac{1}{2}$	$5/2\sqrt{2}$		$\frac{1}{2}$	$-3/2\sqrt{2}$		$-\frac{1}{2}$	$3/2\sqrt{2}$	$\frac{1}{2}$	$-1/2\sqrt{2}$	$5/\sqrt{2}$	$1/\sqrt{2}$
$(0\bar{1}11)(11\bar{1}0)$	$-1/\sqrt{2}$	$-1/\sqrt{2}$		$1/\sqrt{2}$	$-1/\sqrt{2}$		$-1/\sqrt{2}$	$-1/\sqrt{2}$		$1/\sqrt{2}$		$-1/\sqrt{2}$	$-5/\sqrt{2}$	$-1/\sqrt{2}$
$(11\bar{1}0)(0\bar{1}11)$	$-1/\sqrt{2}$	$-1/\sqrt{2}$		$1/\sqrt{2}$	$1/\sqrt{2}$		$1/\sqrt{2}$	$1/\sqrt{2}$		$-1/\sqrt{2}$		$-1/\sqrt{2}$	$-5/\sqrt{2}$	$1/\sqrt{2}$
$(1001)(0000)_1$	1			2				1		2			-3	-1
$(0000)_1(1001)$	1			2				-1		-2			-3	1
$(1001)(0000)_2$	-1		$1/\sqrt{2}$	$\frac{1}{2}$		$1/\sqrt{2}$	$\frac{3}{2}$	-1	$1/\sqrt{2}$	$-\frac{1}{2}$	$1/\sqrt{2}$	$\frac{1}{2}$	1	1
$(0000)_2(1001)$	-1		$1/\sqrt{2}$	$\frac{1}{2}$		$-1/\sqrt{2}$	$-\frac{3}{2}$	1	$-1/\sqrt{2}$	$\frac{1}{2}$	$1/\sqrt{2}$	$\frac{1}{2}$	1	-1
$(1001)(0000)_3$		-1	$-1/\sqrt{2}$	$\frac{1}{2}$	1	$-1/\sqrt{2}$	$-\frac{1}{2}$		$-1/\sqrt{2}$	$\frac{3}{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	1	-1
$(0000)_3(1001)$		-1	$-1/\sqrt{2}$	$\frac{1}{2}$	-1	$1/\sqrt{2}$	$\frac{1}{2}$		$1/\sqrt{2}$	$-\frac{3}{2}$	$-1/\sqrt{2}$	$\frac{1}{2}$	1	1
$(1001)(0000)_4$		1		2	-1		2						-3	1
$(0000)_4(1001)$		1		2	1		-2						-3	-1
	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{35}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{15}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{15}$	$1/\sqrt{2}$	$1/\sqrt{3}$	$1/\sqrt{105}$	$1/\sqrt{5}$

right side, the weights of the representation (1001) denote the basis vector of the product space. The expression (4.2) is computed in the following way:

$$f_1\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}=2^{1/2}\begin{bmatrix}2&0&0&2\\0&1&0&2\end{bmatrix} \tag{4.4}$$

because $\left[\begin{smallmatrix}2002\\2002\end{smallmatrix}\right]$ is an element of a triplet of $SU(2)\alpha_1$ [subgroup of SU(5) associated to α_1] and, in the same way, the right side of (4.1) is

$$f_1\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}=\{f_1\}(1001)(1001)$$
$$=(\bar{1}101)(1001)+(1001)(\bar{1}101)\,, \tag{4.5}$$

where $\{f_i\}=f_i\otimes I+I\otimes f_i$. [Note that (1001) is an element of a doublet of $SU(2)_{\alpha_1}$]. Then we get (4.2). The highest weight vector $\left[\begin{smallmatrix}0102\\0102\end{smallmatrix}\right]$ does not belong to the subspace generated from $\left[\begin{smallmatrix}2002\\0102\end{smallmatrix}\right]$. It is then orthogonal to it. Choosing conveniently its phase, one gets (4.3).

One can compute all the CGC for weights with multiplicity 1 in the same way.

A more complicated case appears when the weights are degenerated. Let us take Table VI corresponding to the weight (1001), with multiplicity 14. The first vector gives

$$\begin{bmatrix}2&0&0&2\\1&0&0&1\end{bmatrix}=6^{-1/2}f_4f_3f_2f_1\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}. \tag{4.6}$$

Using our previous results (4.1) we can write

$$6^{-1/2}f_4f_3f_2f_1\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}$$
$$=6^{-1/2}\{f_4\}\{f_3\}\{f_2\}\{f_1\}(1001)(1001)$$
$$=3^{-1/2}[2^{-1/2}(101\bar{1})(00\bar{1}2)+2^{-1/2}(00\bar{1}2)(101\bar{1})$$
$$+(1001)(0000)_4+(0000)_4(1001)]\,. \tag{4.7}$$

The situation is similar for the other vectors with weight (1001) in (2002), (2010), (0102), and (0110). Finally, vectors $\left[\begin{smallmatrix}1001\\1001\end{smallmatrix}\right]_s$ and $\left[\begin{smallmatrix}1001\\1001\end{smallmatrix}\right]_a$ are chosen (symmetric and skew symmetric) in the subspace orthogonal to the subspace generated by the other vectors. With a convenient phase we obtain the values of Table VI. The symmetry and skew symmetry mean that the vectors of the tensorial product basis must appear like $[(a)(b)+(b)(a)]$ and $[(a)(b)-(b)(a)]$, respectively.

In Tables VI and VII only the dominant weight vectors appear. The CGC for a vector not in the tables can be obtained in two ways, as shown in the following examples: (i) via charge conjugation operators,

$$\begin{bmatrix}2&0&0&2\\\bar{2}&2&0&2\end{bmatrix}=R_{\alpha_1}\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}$$
$$=R_{\alpha_1}(1001)R_{\alpha_1}(1001)=(\bar{1}101)(\bar{1}101)\,, \tag{4.8}$$

$$\begin{bmatrix}2&0&0&2\\0&\bar{2}&2&2\end{bmatrix}=R_{\alpha_2}R_{\alpha_1}\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}=(0\bar{1}11)(0\bar{1}11)\,, \tag{4.9}$$

$$\begin{bmatrix}2&0&0&2\\0&0&\bar{2}&4\end{bmatrix}=R_{\alpha_3}R_{\alpha_2}R_{\alpha_1}\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}=(00\bar{1}2)(00\bar{1}2)\,, \tag{4.10}$$

(ii) via the expression of the vector in the Verma basis,

$$\begin{bmatrix}2&0&0&2\\\bar{2}&2&0&2\end{bmatrix}=2^{-1}f_1^2\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}$$
$$=2^{-1}\{f_1\}^2(1001)(1001)=(\bar{1}101)(\bar{1}101)\,. \tag{4.11}$$

Similarly,

$$\begin{bmatrix}2&0&0&2\\0&\bar{2}&2&2\end{bmatrix}=4^{-1}f_2^2f_1^2\begin{bmatrix}2&0&0&2\\2&0&0&2\end{bmatrix}=(0\bar{1}11)(0\bar{1}11)\,. \tag{4.12}$$

V. CLEBSCH–GORDAN COEFFICIENTS IN THE SUBGROUP BASIS

In this section we want to express the CGC for a basis of $SU(3)\times SU(2)\times U(1)$. Each basis vector of SU(5) has a definite $SU(3)\times SU(2)\times U(1)$ weight and, as in the pre-

TABLE X. CGC of the weight vectors with weight (00)(0) in the subgroup basis. Normalizing factors for the columns are given in the last row.

	(2002)									
	(22)(0) (00)$_1$(0)	(00)$_2$(0)	(00)$_3$(0)	(11)(2) (00)$_1$(0)	(00)$_2$(0)	(11)(0) (00)$_1$(0)	(00)$_2$(0)	(00)(4) (00)(0)	(00)(2) (00)(0)	(00)(0) (00)(0)
(0000)$_1$(0000)$_1$			2			4				2
(0000)$_1$(0000)$_2$			-2	$\frac{1}{2}$		$-\frac{3}{2}$			2	
(0000)$_2$(0000)$_1$			-2	$\frac{1}{2}$		$-\frac{3}{2}$			2	
(0000)$_1$(0000)$_3$		$-\frac{1}{2}$		$-\frac{1}{2}$		$\frac{3}{2}$	-1		-2	1
(0000)$_3$(0000)$_1$		$-\frac{1}{2}$		$-\frac{1}{2}$		$\frac{3}{2}$	-1		-2	1
(0000)$_1$(0000)$_4$		$\frac{1}{2}$				1	1			1
(0000)$_4$(0000)$_1$		$\frac{1}{2}$				1	1			1
(0000)$_2$(0000)$_2$			2	-1		-1		2	2	2
(0000)$_2$(0000)$_3$		$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$	$-\frac{3}{2}$	-2		-1
(0000)$_3$(0000)$_2$		$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$	$-\frac{3}{2}$	-2		-1
(0000)$_2$(0000)$_4$		$-\frac{1}{2}$			$\frac{1}{2}$	-1	$\frac{3}{2}$		2	1
(0000)$_4$(0000)$_2$		$-\frac{1}{2}$			$\frac{1}{2}$	-1	$\frac{3}{2}$		2	1
(0000)$_3$(0000)$_3$	2				1		-1	2	-2	2
(0000)$_3$(0000)$_4$	-2				$-\frac{1}{2}$		$-\frac{3}{2}$		-2	
(0000)$_4$(0000)$_3$	-2				$-\frac{1}{2}$		$-\frac{3}{2}$		-2	
(0000)$_4$(0000)$_4$	2						4			2
(1001)($\bar{1}$00$\bar{1}$)		$\frac{1}{4}$				$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}$
($\bar{1}$00$\bar{1}$)(1001)		$\frac{1}{4}$				$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{2}$
($\bar{1}$101)(1$\bar{1}$0$\bar{1}$)					$\frac{1}{4}$		$\frac{5}{4}$		1	1
(1$\bar{1}$0$\bar{1}$)($\bar{1}$101)					$\frac{1}{4}$		$\frac{5}{4}$		1	1
(101$\bar{1}$)($\bar{1}$0$\bar{1}$1)				$-\frac{1}{4}$		$\frac{5}{4}$			-1	1
($\bar{1}$0$\bar{1}$1)(101$\bar{1}$)				$-\frac{1}{4}$		$\frac{5}{4}$			-1	1
(0$\bar{1}$11)(01$\bar{1}\bar{1}$)	-1	$-\frac{1}{4}$				$-\frac{1}{2}$				$\frac{1}{2}$
(01$\bar{1}\bar{1}$)(0$\bar{1}$11)	-1	$-\frac{1}{4}$				$-\frac{1}{2}$				$\frac{1}{2}$
($\bar{1}$11$\bar{1}$)(1$\bar{1}\bar{1}$1)								-1		1
(1$\bar{1}\bar{1}$1)($\bar{1}$11$\bar{1}$)								-1		1
(11$\bar{1}$0)($\bar{1}\bar{1}$10)		$-\frac{1}{4}$	-1				$-\frac{1}{2}$			$\frac{1}{2}$
($\bar{1}\bar{1}$10)(11$\bar{1}$0)		$-\frac{1}{4}$	-1				$-\frac{1}{2}$			$\frac{1}{2}$
(2$\bar{1}$00)($\bar{2}$100)				$\frac{1}{4}$		$\frac{5}{4}$			1	1
($\bar{2}$100)(2$\bar{1}$00)				$\frac{1}{4}$		$\frac{5}{4}$			1	1
($\bar{1}$2$\bar{1}$0)(1$\bar{2}$10)				$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{5}{4}$	$-\frac{5}{4}$		1	1
(1$\bar{2}$10)($\bar{1}$2$\bar{1}$0)				$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{5}{4}$	$-\frac{5}{4}$		1	1
(0$\bar{1}$2$\bar{1}$)(01$\bar{2}$1)				$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{5}{4}$	$-\frac{5}{4}$		-1	1
(01$\bar{2}$1)(0$\bar{1}$2$\bar{1}$)				$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{5}{4}$	$-\frac{5}{4}$		-1	1
(00$\bar{1}$2)(001$\bar{2}$)					$-\frac{1}{4}$		$\frac{5}{4}$		-1	1
(001$\bar{2}$)(00$\bar{1}$2)					$-\frac{1}{4}$		$\frac{5}{4}$		-1	1
	$1/\sqrt{6}$		$1/\sqrt{6}$			$1/\sqrt{35}$	$1/\sqrt{35}$	$1/\sqrt{6}$	$1/\sqrt{42}$	$1/\sqrt{42}$

	(0102)						(2010)					
	(11)(2) (00)$_1$(0)	(00)$_2$(0)	(11)(0) (00)$_1$(0)	(00)$_2$(0)	(03)(0) (00)(0)	(00)(2) (00)(0)	(11)(2) (00)$_1$(0)	(00)$_2$(0)	(11)(0) (00)$_1$(0)	(00)$_2$(0)	(30)(0) (00)(0)	(00)(2) (00)(0)
(0000)$_1$(0000)$_1$												
(0000)$_1$(0000)$_2$	$\frac{1}{2}$		$\frac{3}{2}$			1	$\frac{1}{2}$		$\frac{3}{2}$			-1
(0000)$_2$(0000)$_1$	$-\frac{1}{2}$		$-\frac{3}{2}$			-1	$-\frac{1}{2}$		$-\frac{3}{2}$			1
(0000)$_1$(0000)$_3$	$-\frac{1}{2}$		$\frac{1}{2}$	1	1	-1	$-\frac{1}{2}$		$\frac{1}{2}$	1	1	1
(0000)$_3$(0000)$_1$	$\frac{1}{2}$		$-\frac{1}{2}$	-1	-1	1	$\frac{1}{2}$		$-\frac{1}{2}$	-1	-1	-1
(0000)$_1$(0000)$_4$			1	-1	-1				1	-1	-1	
(0000)$_4$(0000)$_1$			-1	1	1				-1	1	1	
(0000)$_2$(0000)$_2$												
(0000)$_2$(0000)$_3$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	-1	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1
(0000)$_3$(0000)$_2$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	1	1	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	1	-1
(0000)$_2$(0000)$_4$		$-\frac{1}{2}$	-1	$-\frac{1}{2}$	1	-1		$-\frac{1}{2}$	-1	$-\frac{1}{2}$	1	1
(0000)$_4$(0000)$_2$		$\frac{1}{2}$	1	$\frac{1}{2}$	-1	1		$\frac{1}{2}$	1	$\frac{1}{2}$	-1	-1
(0000)$_3$(0000)$_3$												
(0000)$_3$(0000)$_4$		$\frac{1}{2}$		$-\frac{3}{2}$		1		$\frac{1}{2}$		$-\frac{3}{2}$		-1
(0000)$_4$(0000)$_3$		$-\frac{1}{2}$		$\frac{3}{2}$		-1		$-\frac{1}{2}$		$\frac{3}{2}$		1
(0000)$_4$(0000)$_4$												

TABLE X. (*Continued.*)

	(0102)						(2010)					
	(11)(2)		(11)(0)		(03)(0)	(00)(2)	(11)(2)		(11)(0)		(30)(0)	(00)(2)
	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)
$(1001)(\bar{1}00\bar{1})$			$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$				$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
$(\bar{1}00\bar{1})(1001)$			$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$				$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	
$(\bar{1}101)(1\bar{1}0\bar{1})$		$-\frac{1}{4}$		$-\frac{3}{4}$		$-\frac{1}{2}$		$\frac{1}{4}$		$\frac{3}{4}$		$-\frac{1}{2}$
$(1\bar{1}0\bar{1})(\bar{1}101)$		$\frac{1}{4}$		$\frac{3}{4}$		$\frac{1}{2}$		$-\frac{1}{4}$		$-\frac{3}{4}$		$\frac{1}{2}$
$(101\bar{1})(\bar{1}0\bar{1}1)$	$-\frac{1}{4}$		$\frac{3}{4}$			$-\frac{1}{2}$	$\frac{1}{4}$		$-\frac{3}{4}$			$-\frac{1}{2}$
$(\bar{1}0\bar{1}1)(101\bar{1})$	$\frac{1}{4}$		$-\frac{3}{4}$			$\frac{1}{2}$	$-\frac{1}{4}$		$\frac{3}{4}$			$\frac{1}{2}$
$(0\bar{1}11)(01\bar{1}\bar{1})$			$-\frac{1}{2}$	-1	$\frac{1}{2}$				$-\frac{1}{2}$	1	$-\frac{1}{2}$	
$(01\bar{1}\bar{1})(0\bar{1}11)$			$\frac{1}{2}$	1	$-\frac{1}{2}$				$-\frac{1}{2}$	-1	$\frac{1}{2}$	
$(\bar{1}11\bar{1})(1\bar{1}\bar{1}1)$						$-\frac{3}{2}$						$-\frac{3}{2}$
$(1\bar{1}\bar{1}1)(\bar{1}11\bar{1})$						$\frac{3}{2}$						$\frac{3}{2}$
$(11\bar{1}0)(\bar{1}\bar{1}10)$			1	$\frac{1}{2}$	$\frac{1}{2}$				-1	$-\frac{1}{2}$	$-\frac{1}{2}$	
$(\bar{1}\bar{1}10)(11\bar{1}0)$			-1	$-\frac{1}{2}$	$-\frac{1}{2}$				1	$\frac{1}{2}$	$\frac{1}{2}$	
$(2\bar{1}00)(\bar{2}100)$	$\frac{1}{4}$		$\frac{3}{4}$			$\frac{1}{2}$	$-\frac{1}{4}$		$-\frac{3}{4}$			$\frac{1}{2}$
$(\bar{2}100)(2\bar{1}00)$	$-\frac{1}{4}$		$-\frac{3}{4}$			$-\frac{1}{2}$	$\frac{1}{4}$		$\frac{3}{4}$			$-\frac{1}{2}$
$(\bar{1}2\bar{1}0)(1\bar{2}10)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$		$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{3}{4}$		$-\frac{1}{2}$
$(1\bar{2}10)(\bar{1}2\bar{1}0)$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{3}{4}$		$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$		$\frac{1}{2}$
$(0\bar{1}2\bar{1})(01\bar{2}1)$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{3}{4}$		$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$		$-\frac{1}{2}$
$(01\bar{2}1)(0\bar{1}2\bar{1})$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$		$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{3}{4}$		$\frac{1}{2}$
$(00\bar{1}2)(001\bar{2})$		$\frac{1}{4}$		$-\frac{3}{4}$		$\frac{1}{2}$		$-\frac{1}{4}$		$\frac{3}{4}$		$\frac{1}{2}$
$(001\bar{2})(00\bar{1}2)$		$-\frac{1}{4}$		$\frac{3}{4}$		$-\frac{1}{2}$		$\frac{1}{4}$		$-\frac{3}{4}$		$-\frac{1}{2}$
			$1/\sqrt{15}$	$1/\sqrt{15}$	$1/\sqrt{3}$	$1/\sqrt{15}$			$1/\sqrt{15}$	$1/\sqrt{15}$	$1/\sqrt{3}$	$1/\sqrt{15}$

	(0110)					$(1001)_s$				$(1001)_a$				(0000)
	(11)(2)		(11)(0)		(00)(0)	(11)(0)		(00)(2)	(00)(0)	(11)(0)		(00)(2)	(00)(0)	(00)(0)
	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	(00)(0)
$(0000)_1(0000)_1$						-6			$-\frac{4}{5}$					$\frac{4}{5}$
$(0000)_1(0000)_2$	$\frac{1}{2}$		$\frac{1}{2}$		$-\frac{1}{3}$	4		2	$\frac{14}{15}$					$-\frac{3}{5}$
$(0000)_2(0000)_1$	$\frac{1}{2}$		$\frac{1}{2}$		$-\frac{1}{3}$	4		2	$\frac{14}{15}$					$-\frac{3}{5}$
$(0000)_1(0000)_3$	$-\frac{1}{2}$		$-\frac{1}{2}$	-1		-4	-2	-2	$-\frac{2}{5}$					$\frac{2}{5}$
$(0000)_3(0000)_1$	$-\frac{1}{2}$		$-\frac{1}{2}$	-1		-4	-2	-2	$-\frac{2}{5}$					$\frac{2}{5}$
$(0000)_1(0000)_4$			1	1	$-\frac{1}{3}$	2	2		$\frac{8}{15}$					$-\frac{1}{5}$
$(0000)_4(0000)_1$			1	1	$-\frac{1}{3}$	2	2		$\frac{8}{15}$					$-\frac{1}{5}$
$(0000)_2(0000)_2$	-1		-1		$\frac{2}{3}$	-2		2	$\frac{2}{15}$					$\frac{6}{5}$
$(0000)_2(0000)_3$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{2}{3}$	4	4		$-\frac{8}{15}$					$-\frac{4}{5}$
$(0000)_3(0000)_2$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{2}{3}$	4	4		$-\frac{8}{15}$					$-\frac{4}{5}$
$(0000)_2(0000)_4$		$\frac{1}{2}$	-1	$-\frac{1}{2}$		-2	-4	2	$-\frac{2}{5}$					$\frac{2}{5}$
$(0000)_4(0000)_2$		$\frac{1}{2}$	-1	$-\frac{1}{2}$		-2	-4	2	$-\frac{2}{5}$					$\frac{2}{5}$
$(0000)_3(0000)_3$		1		-1	$\frac{2}{3}$		-2	-2	$\frac{2}{15}$					$\frac{6}{5}$
$(0000)_3(0000)_4$		$-\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{3}$		4	-2	$\frac{14}{15}$					$-\frac{3}{5}$
$(0000)_4(0000)_3$		$-\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{3}$		4	-2	$\frac{14}{15}$					$-\frac{3}{5}$
$(0000)_4(0000)_4$							-6		$-\frac{4}{5}$					$\frac{4}{5}$
$(1001)(\bar{1}00\bar{1})$			$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$-\frac{2}{5}$	$\frac{1}{2}$	$-\frac{1}{2}$			$\frac{1}{2}$
$(\bar{1}00\bar{1})(1001)$			$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$		$-\frac{2}{5}$	$\frac{1}{2}$	$-\frac{1}{2}$			$\frac{1}{2}$
$(\bar{1}101)(1\bar{1}0\bar{1})$		$-\frac{1}{4}$		$-\frac{1}{4}$	$\frac{1}{6}$		$\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{1}{6}$		$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(1\bar{1}0\bar{1})(\bar{1}101)$		$-\frac{1}{4}$		$-\frac{1}{4}$	$\frac{1}{6}$		$\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{1}{6}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(101\bar{1})(\bar{1}0\bar{1}1)$	$\frac{1}{4}$		$-\frac{1}{4}$		$\frac{1}{6}$	$\frac{5}{2}$		$\frac{5}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(\bar{1}0\bar{1}1)(101\bar{1})$	$\frac{1}{4}$		$-\frac{1}{4}$		$\frac{1}{6}$	$\frac{5}{2}$		$\frac{5}{2}$	$-\frac{1}{6}$	$\frac{1}{2}$		$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(0\bar{1}11)(01\bar{1}\bar{1})$			$\frac{1}{2}$		$\frac{1}{6}$	$\frac{5}{2}$			$-\frac{2}{5}$	$-\frac{1}{2}$	-1			$\frac{1}{2}$
$(01\bar{1}\bar{1})(0\bar{1}11)$			$\frac{1}{2}$		$\frac{1}{6}$	$\frac{5}{2}$			$-\frac{2}{5}$	$\frac{1}{2}$	1			$\frac{1}{2}$
$(\bar{1}11\bar{1})(1\bar{1}\bar{1}1)$					$\frac{1}{2}$				1			-1		$\frac{1}{2}$
$(1\bar{1}\bar{1}1)(\bar{1}11\bar{1})$					$\frac{1}{2}$				1			1		$\frac{1}{2}$
$(11\bar{1}0)(\bar{1}\bar{1}10)$				$\frac{1}{2}$	$\frac{1}{6}$		$\frac{5}{2}$		$-\frac{2}{5}$	1	$\frac{1}{2}$			$\frac{1}{2}$
$(\bar{1}\bar{1}10)(11\bar{1}0)$				$\frac{1}{2}$	$\frac{1}{6}$		$\frac{5}{2}$		$-\frac{2}{5}$	-1	$-\frac{1}{2}$			$\frac{1}{2}$
$(2\bar{1}00)(\bar{2}100)$	$-\frac{1}{4}$		$-\frac{1}{4}$		$\frac{1}{6}$	$\frac{5}{2}$		$-\frac{5}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$		$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(\bar{2}100)(2\bar{1}00)$	$-\frac{1}{4}$		$-\frac{1}{4}$		$\frac{1}{6}$	$\frac{5}{2}$		$-\frac{5}{2}$	$-\frac{1}{6}$	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(\bar{1}2\bar{1}0)(1\bar{2}10)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$

TABLE X. (*Continued.*)

	(0110)					$(1001)_s$				$(1001)_a$				(0000)
	(11)(2)		(11)(0)		(00)(0)	(11)(0)		(00)(2)	(00)(0)	(11)(0)		(00)(2)	(00)(0)	(00)(0)
	$(00)_1(0)$	$(00)_2(0)$	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	$(00)_1(0)$	$(00)_2(0)$	(00)(0)	(00)(0)	(00)(0)
$(1\bar{2}10)(\bar{1}2\bar{1}0)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$-\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(0\bar{1}2\bar{1})(01\bar{2}1)$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$\frac{5}{2}$	$-\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$(01\bar{2}1)(0\bar{1}2\bar{1})$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$-\frac{5}{2}$	$-\frac{5}{2}$	$\frac{5}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(00\bar{1}2)(001\bar{2})$		$\frac{1}{4}$		$-\frac{1}{4}$	$\frac{1}{6}$		$\frac{5}{2}$	$\frac{5}{2}$	$-\frac{1}{6}$		$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$(001\bar{2})(00\bar{1}2)$		$\frac{1}{4}$		$-\frac{1}{4}$	$\frac{1}{6}$		$\frac{5}{2}$	$\frac{5}{2}$	$-\frac{1}{6}$		$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
			$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{2}$	$1/\sqrt{105}$	$1/\sqrt{105}$	$1/\sqrt{105}$	$1/\sqrt{7}$	$1/\sqrt{5}$	$1/\sqrt{5}$	$1/\sqrt{5}$	$1/\sqrt{3}$	$1/\sqrt{6}$

vious case belongs to a subspace, irreducible with respect to the subgroup.

Every SU(5) weight, $m=(m_1,m_2,m_3,m_4)$, corresponds to a definite subgroup weight $v=(v_1v_2)(v_3)(v_4)$ given by

$$v=mP=(m_1+m_2,m_3+m_4)(m_2+m_3)\times(2m_1-m_2+m_3-2m_4)\,, \tag{5.1}$$

where the parentheses indicate the SU(3), SU(2), and U(1) weights, respectively, and P is given in (2.8).

Each subgroup weight vector belongs to one irreducible subspace. This subspace is generated from the highest weight by the operator $\tilde{f}_1=[f_1,f_2],\tilde{f}_2=[f_3,f_4]$ for SU(3); $\hat{f}=[f_2,f_3]$ for SU(2); and $\bar{h}=2h_1-h_2+h_3-2h_4$ for U(1).

We also have the charge conjugation operators and reflections r_i and R_i, where i labels a root of SU(3) or SU(2). In this case there are more dominant weights and Tables VIII–X are bigger. We have omitted the U(1)-part of each weight to simplify the notation.

We present some examples. In the tables, the first line of each column indicates the SU(5) irreducible subspace. The second line gives the highest weight of each irreducible subspace with respect to the subgroup. The third line corresponds to a dominant weight of that subspace. Finally, the following lines give the linear combinations of the corresponding vector as a function of the tensorial product basis vectors, as in Sec. IV.

For the subspace of weight (2002), the highest weight is now (22)(0). The weight $(1\bar{1}12)$, which is not a dominant weight in the group basis, corresponds in the subgroup basis to the dominant weight (03)(0). In Table VIII we find

$$\begin{bmatrix}2&2\\2&2\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix}=(1001)(1001)\,, \tag{5.2}$$

$$\begin{aligned}\begin{bmatrix}2&2\\0&3\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix}&=2^{-1/2}\hat{f}_1\begin{bmatrix}2&2\\2&2\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix}\\&=2^{-1/2}\{[f_1,f_2]\}(1001)(1001)\\&=-2^{-1/2}[(1001)(0\bar{1}11)+(0\bar{1}11)(1001)]\,.\end{aligned} \tag{5.3}$$

Similarly,

$$\begin{aligned}\begin{bmatrix}0&2\\0&2\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix}&=2^{-1/2}\hat{f}\begin{bmatrix}0&2\\0&2\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix}\\&=2^{-1/2}\{[f_2,f_3]\}(\bar{1}101)(\bar{1}101)\,,\end{aligned} \tag{5.4}$$

where the expression of the vector $\begin{bmatrix}02\\02\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix}$ is given in Table VIII.

There are also degenerate weights. We consider the following example in the weight subspace (0102) in Table IX. In the subalgebra with highest weight (12)(1), there are two vectors belonging to the subspace weight (01)(1). Their expressions are

$$\begin{bmatrix}1&2\\0&1\end{bmatrix}_1\begin{bmatrix}1\\1\end{bmatrix}=2^{-1}\hat{f}_1\hat{f}_2\begin{bmatrix}1&2\\1&2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}, \tag{5.5}$$

$$\begin{bmatrix}1&2\\0&1\end{bmatrix}_2\begin{bmatrix}1\\1\end{bmatrix}=3^{-1/2}\hat{f}_2\hat{f}_1\begin{bmatrix}1&2\\1&2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}. \tag{5.6}$$

For the weight vectors which are not dominant, we can apply the charge conjugation operators, R_{β_1}, R_{β_2}, and R_γ to the vectors of the subspaces of SU(3) and SU(2), respectively.

Thus, in the weight subspace of (2002), we can take the vector $\begin{bmatrix}12\\32\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$ corresponding to $(03\bar{2}0)$. We have

$$\begin{aligned}\begin{bmatrix}1&2\\3&\bar{2}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}&=R_{\beta_2}\begin{bmatrix}1&2\\1&2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\&=R_{\alpha_3+\alpha_4}(2^{-1/2}(1001)(\bar{1}101)\\&\quad+2^{-1/2}(\bar{1}101)(1001))\\&=2^{-1/2}[(11\bar{1}0)(\bar{1}2\bar{1}0)+(\bar{1}2\bar{1}0)(11\bar{1}0)]\,.\end{aligned} \tag{5.7}$$

Another method is

$$\begin{aligned}\begin{bmatrix}1&2\\3&\bar{2}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}&=2^{-1}\hat{f}_2^2\begin{bmatrix}1&2\\1&2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\&=2^{-1}\{\hat{f}_2\}^2(2^{-1/2}(1001)(\bar{1}101)\\&\quad+2^{-1/2}(\bar{1}101)(1001))\\&=2^{-1/2}\hat{f}_2(2^{-1}[(1001)(\bar{1}2\bar{1}0)+(\bar{1}2\bar{1}0)(1001)\\&\quad+(11\bar{1}0)(\bar{1}101)+(\bar{1}101)(11\bar{1}0)])\,,\end{aligned} \tag{5.8}$$

where we have used the expression of the vector $\begin{bmatrix}12\\12\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$ or $\begin{bmatrix}12\\20\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$ accordingly to Table IX; and finally acting with f_2, we get (5.3).

VI. CONCLUSIONS

An extensive use of the Verma algorithm has been made through all the computations. Because of the high degeneracy of some weight subspaces, the difficulties of finding a

basis are overcome by using this algorithm. This method can be used with other Lie groups, different from SU(n). Though only CGC corresponding to dominant weights have been computed, the other ones can be found easily through charge conjugation operators. The results presented in this article complements those of Refs. 4 and 5, about CGC in unification theories. Some other calculations are found in recent papers[9] about CGC in supersymmetric theories. Our method can be easily applied to these cases.

ACKNOWLEDGMENTS

The research reported in this article was supported in part by the Natural Sciences and Engineering Research Council of Canada, the "Fonds FCAR du Gouvernement du Québec," and the "Ministerio de Educación y Ciencia" of Spain (grants to M. A. del Olmo and M. A. Rodríguez from the "Plan de Formación de Personal Investigador").

[1]L. C. Biedenharn and J. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, 1981); D. A. Varshalovich, A. N. Moskalenko, and V. K. Chersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975).
[2]R. T. Sharp and H. von Beyer, J. Math. Phys. **7**, 1105 (1966).
[3]For example, M. Hongoh, R. T. Sharp, and D. E. Tilley, J. Math. Phys. **15**, 782 (1974); J.-G. Chen, M.-J. Gao, and X.-G. Chen, J. Phys. A: Math. Nucl. Gen. **17**, 481 (1984).
[4]M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978); R. Slansky, Phys. Rep. **79**, 1 (1981).
[5]I. M. Gel'fand and M. L. Zeitlin, Dok. Akad. Nauk SSSR **71**, 825 (1950); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[6]I. -G. Koh, J. Patera, and C. Rousseau, J. Math. Phys. **24**, 1955 (1983); **25**, 2863 (1984).
[7]R. V. Moody and J. Patera, J. Math. Phys. **25**, 2838 (1984).
[8]S. -P. Li, R. V. Moody, M. Nicolescu, and J. Patera, J. Math. Phys. **27**, 668 (1986).
[9]J. Patera and R. T. Sharp, Phys. Rev. D **25**, 1141 (1982).
[10]L. Michel, J. Patera, and R. T. Sharp, "Demazure–Tits subgroup of a simple Lie group," submitted to J. Math. Phys.
[11]G. Racah, *Group Theory and Spectroscopy* (Institute of Adv. Study, Princeton, NJ, 1951), reprinted in Ergeb. Exakten Naturwiss. **37**, 28 (1965).
[12]M. R. Bremner, R. V. Moody, and J. Patera, *Tables of Dominant Weight Multiplicities for Representations of Simple Lie Algebras* (Dekker, New York, 1984.)

Finite-dimensional representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis. II. Nontypical representations

Tchavdar D. Palev
Institute of Nuclear Research and Nuclear Energy, 1184 Sofia, Bulgaria

(Received 7 October 1985; accepted for publication 27 August 1986)

All nontypical irreducible representations of the special linear Lie superalgebra sl(1,3) are considered. Explicit expressions for the transformation of the basis under the action of the generators are given. The results of this paper together with those obtained in Paper I [T. D. Palev, J. Math. Phys. **26**, 1640 (1985)] solve the problem of the finite-dimensional irreducible representations of sl(1,3). The results are compared with those obtained by the Young supertableau technique. A mapping of the supertableau basis on the Gel'fand–Zetlin basis is given.

I. INTRODUCTION

In Ref. 1 (hereafter referred to as I) we gave explicit expressions for all typical representations of the basic Lie superalgebra (LS) sl(1,3) in a Gel'fand–Zetlin basis. In the present paper we solve this problem for the nontypical representations.

We recall that the finite-dimensional irreducible representation (IR's) of any basic LS resolve into two classes: typical and nontypical.[2] The module V over the LS A (and, hence, the corresponding representation of A in V) is said to be typical, if, whenever V is a submodule of a larger A module W, there always exists a complement to V subspace V', which is also an A module, i.e., $W = V \oplus V', AV' \subset V'$. If this is not the case, i.e., there exists an A module W, containing V as a submodule and in the same time, it is impossible to determine a complement to a V submodule, then V (and also the representation of A in V) is called nontypical.

Let A be a basic LS, U its universal enveloping algebra, A_0 the even subalgebra, P_+ the linear span of all odd positive root vectors, and P the subspace sum of A_0 and P_+, $P = A_0 \oplus P_+$. Consider an arbitrary A_0 module V_0 and extend it to a P module, assuming $P_+ V_0 = 0$. Let $\overline{V}$ be the factor space of the tensor product $U \otimes V_0$ with respect to the linear envelope I of all elements $up \otimes v - u \otimes pv$, $u \in U$, $p \in P$, $v \in V_0$, i.e.,

$$\overline{V} = U \otimes V_0 / I. \tag{1.1}$$

The space $\overline{V}$ is turned into an A module in a natural way:

$$g(u \otimes v) = gu \otimes v, \quad g \in A, \quad u \otimes v \in \overline{V}. \tag{1.2}$$

The A module $\overline{V}$ is said to be induced from the A_0 module V_0.

Let F be the family of A modules $\overline{V}$, induced from all irreducible finite-dimensional A_0 modules V_0. The family F carries information about all finite-dimensional irreducible representations of the LS A (strictly speaking the statement below is true for the type I LS's, but with minor modification it holds for all other basis LS's) in the following sense. Denote by $F_t \subset F$ (resp. $F_{\text{red}} \subset F$) those induced A modules $\overline{V} \in F$ that contain no (resp. that do contain) nontrivial invariant subspaces: $F = F_t \cup F_{\text{red}}$. If $\overline{V} \in F_{\text{red}}$ and $\overline{I}$ is the maximal invariant subspace in $\overline{V}$, then the factor module $V = \overline{V}/\overline{I}$ is irreducible. Let F_{nt} be the set of all such modules, $F_{nt} = \{V \mid V = \overline{V}/\overline{I}, \ \overline{V} \in F_{\text{red}}\}$. Then it turns out that all A modules from F_t are typical, whereas those from F_{nt} are nontypical. The relevance of this construction stems from the observation that every typical (resp. nontypical) representation of A can be realized in a certain A module from F_t (resp. from F_{nt}).[2]

In I we have worked out the induced representations of the LS sl(1,3), whose even subalgebra is gl(3). We wrote the results in two different bases: the induced basis and the Gel'fand–Zetlin basis (GZ basis). Both of them may be assumed to be orthonormed. The transformation properties of the vectors from the induced basis are relatively simple [see I, (4.14)] and this is its advantage. This basis is inconvenient, however, for the construction of the nontypical representations, because most of the basis vectors have nonzero projections both on the maximal invariant gl(3) module $\overline{I}$ and on its orthogonal complement. This was the reason to introduce a new basis, which is reduced with respect to the even subalgebra. To this end we considered $\overline{V}$ as a representation space of $\mathrm{gl}(3) \subset \mathrm{sl}(1,3)$ and represented it as a direct sum of its irreducible gl(3) submodules V_n,

$$\overline{V} = \sum_n \oplus V_n. \tag{1.3}$$

As a basis Γ_n within every V_n we chose the Gel'fand–Zetlin basis[3] and defined an orthonormed Gel'fand–Zetlin basis in $\overline{V}$ to be $\Gamma = \cup_n \Gamma_n$. It turned out that $n \leqslant 8$.

In Sec. II we collect some of the results from I and give the action of the superalgebra generators on the GZ basis vectors. These relations, which in fact determine all typical representations, will serve as a starting point for the construction of the nontypical representations (Sec. III). In Sec. IV we show in a matrix form the lowest-dimensional nondecomposible representations. In Sec. V we relate our results with those obtained in the frame of the Young supertableau approach.[4–9] We establish a mapping of the supertableau basis (Weyl basis) on the Gel'fand–Zetlin basis in a similar way as this was done for the Lie algebra sl(n) (Ref. 10). This mapping is one to one for all nontypical representations.

II. INDUCED REPRESENTATIONS OF sl(1,3)

Let e_{AB}, $A,B = 0,1,2,3$, be a 4×4 matrix with 1 on the A th row and B th column and zero elsewhere. The LS sl(1,3)

 0022-2488/87/020272-20$02.50 © 1987 American Institute of Physics

can be defined as the linear span of the even generators

$$E_{ij} = e_{ij} + \delta_{ij}\, e_{00}, \quad i,j = 1,2,3, \tag{2.1}$$

which are the generators of the Lie algebra gl(3), and the odd generators

$$e_{0i}, e_{i0}, \quad i = 1,2,3. \tag{2.2}$$

The induced sl(1,3) modules are labeled by three complex numbers $[m_{13},m_{23},m_{33}] \equiv [m]_3$ with the only restriction that $m_{13} - m_{23}, m_{23} - m_{33}$ are non-negative integers, i.e., $m_{13} - m_{23} \in \mathbb{Z}_+$, $m_{23} - m_{33} \in \mathbb{Z}_+$. By $\overline{V}([m]_3) \in F$ we denote the corresponding to $[m]_3$ sl(1,3) module. In I we have shown that $\overline{V}([m]_3)$ is typical (and, hence, irreducible) iff

$$m_{i3} \neq i - 1, \quad i = 1,2,3. \tag{2.3}$$

There are three classes of reducible (but not completely reducible) modules

$$\overline{V}([0,m_{23},m_{33}]), \overline{V}([m_{13},1,m_{33}]), \overline{V}([m_{13},m_{23},2]) \in F_{\text{red}}. \tag{2.4}$$

In order to determine the corresponding nontypical representations, we have to factorize each space (2.4) with respect to its maximal nontrivial invariant subspace

$$\overline{I}([0,m_{23},m_{33}]), \quad \overline{I}([m_{13},1,m_{33}]), \quad \overline{I}([m_{13},m_{23},2]), \tag{2.5}$$

respectively.

Every induced sl(1,3) module from the family F is a direct space sum of (no more than) eight finite-dimensional irreducible gl(3) modules $V([m]_3)$. As a basis in each $V([m]_3)$ we choose a GZ basis[3]

$$\left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} \\ & & m_{11} \end{matrix} \right\rangle \equiv \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle. \tag{2.6}$$

The numbers $[m_{12},m_{22}] \equiv [m]_2$ and m_{11} label the basis vectors in $V([m]_3)$. They run over all possible values, consistent with the "betweenness condition"

$$m_{13} - m_{12}, m_{12} - m_{23}, m_{23} - m_{22}, m_{22} - m_{33},$$
$$m_{12} - m_{11}, m_{11} - m_{22} \in \mathbb{Z}_+. \tag{2.7}$$

Introduce the abbreviations

$$[m]_n^{\pm i} = [m_{1n} \pm \delta_{1i}, m_{2n} \pm \delta_{2i}, \ldots, m_{nn} \pm \delta_{ni}], \tag{2.8}$$

$$[m + c]_n = [m_{1n} + c, m_{2n} + c, \ldots, m_{nn} + c]. \tag{2.9}$$

Then the sl(1,3) module $\overline{V}([m]_3)$, induced from the gl(3) module $V([m]_3)$, reads

$$\overline{V}([m]_3) = V([m]_3) \oplus \sum_{i=1}^{3} \oplus V([m-1]_3^i) \oplus \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}) \oplus V([m-2]_3), \tag{2.10}$$

with

$$\overline{V}_0([m]_3) = V([m]_3) \oplus \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}), \tag{2.11}$$

and

$$\overline{V}_1([m]_3) = V([m-2]_3) \oplus \sum_{i=1}^{3} \oplus V([m-1]_3^{i}) \tag{2.12}$$

being the even and the odd subspaces of $\overline{V}([m]_3)$, respectively.

In order to write the transformation properties of the GZ basis under the action of the generators, it is convenient to introduce the notation

$$l_{ij} = m_{ij} - i, \tag{2.13}$$

$$S(i,j) = \begin{cases} 1, & \text{for } i \leqslant j, \\ -1, & \text{for } i > j, \end{cases} \tag{2.14}$$

$$\epsilon_{ijk} \text{ an antisymmetric tensor with } \epsilon_{123} = 1. \tag{2.15}$$

Then one has[1]

$$e_{0k} \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = 0, \quad k = 1,2,3, \tag{2.16}$$

$$e_{10} \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{2} S(i,j) S(j,1) \times \left| \frac{\Pi_{k \neq j=1}^{2} (l_{k2} - l_{11})(l_{k2} - l_{i3} - 1) \Pi_{k \neq i=1}^{3} (l_{k3} - l_{j2})}{(l_{12} - l_{22} - j + 1)(l_{12} - l_{22} - j + 2) \Pi_{k \neq i=1}^{3} (l_{k3} - l_{i3})} \right|^{1/2} \left| \begin{matrix} [m-1]_3^i \\ [m-1]_2^j \\ m_{11} \end{matrix} \right\rangle, \tag{2.17}$$

$$e_{20} \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{2} S(i,j) \times \left| \frac{(l_{j2} - l_{11} + 1) \Pi_{k \neq j=1}^{2} (l_{k2} - l_{i3} - 1) \Pi_{k \neq i=1}^{3} (l_{k3} - l_{j2})}{(l_{12} - l_{22} - j + 1)(l_{12} - l_{22} - j + 2) \Pi_{k \neq i=1}^{3} (l_{k3} - l_{i3})} \right|^{1/2} \left| \begin{matrix} [m-1]_3^i \\ [m-1]_3^j \\ m_{11} - 1 \end{matrix} \right\rangle, \tag{2.18}$$

$$e_{30}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{i=1}^{3}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{i}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle, \tag{2.19}$$

$$e_{01}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=(l_{s3}+1)\sum_{j=1}^{2}S(s,j)S(j,1)$$

$$\times\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{s3})\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m+1]_2^{-j}\\ m_{11}\end{matrix}\right\rangle, \tag{2.20}$$

$$e_{02}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=(l_{s3}+1)\sum_{j=1}^{2}S(s,j)$$

$$\times\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{s3})\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m+1]_2^{-j}\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.21}$$

$$e_{03}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=(l_{s3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3})}{\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]\\ [m+1]_2\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.22}$$

$$e_{10}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}S(i,j)S(j,1)$$

$$\times\left|\frac{(l_{s3}-l_{j2})(l_{l3}-l_{j2}-1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-\Lambda}\\ [m-1]_3^{j}\\ m_{11}\end{matrix}\right\rangle, \tag{2.23}$$

$$e_{20}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}S(i,j)$$

$$\times\left|\frac{(l_{11}-l_{j2}-1)(l_{s3}-l_{j2})(l_{l3}-l_{j2}-1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-\Lambda}\\ [m-1]_2^{j}\\ m_{11}-1\end{matrix}\right\rangle, \tag{2.24}$$

$$e_{30}\left|\begin{matrix}[m-1]_3^{s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{i=1}^{3}\epsilon_{sli}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-\Lambda}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle, \tag{2.25}$$

$$e_{0:}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}(l_{l3}+1)S(l,j)S(j,1)$$

$$\times\left|\frac{(l_{i3}-l_{j2})(l_{s3}-l_{j2}-1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{i}\\ [m+1]_2^{-j}\\ m_{11}\end{matrix}\right\rangle, \tag{2.26}$$

$$e_{02}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}(l_{l3}+1)S(l,j)$$

$$\times\left|\frac{(l_{i3}-l_{j2})(l_{s3}-l_{j2}-1)(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{i}\\ [m+1]_3^{-j}\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.27}$$

$$e_{03}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{j=1}^{3}(l_{j3}+1)\epsilon_{sji}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{j3}+1)}{\Pi_{k\neq j=1}^{3}(l_{k3}-l_{j3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{i}\\ [m+1]_2\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.28}$$

$$e_{10}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{j=1}^{2}S(s,j)S(j,1)$$

$$\times\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{s3}+1)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m-2]_3\\ [m-1]_2^{j}\\ m_{11}\end{matrix}\right\rangle, \tag{2.29}$$

$$e_{20}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{j=1}^{2}S(s,j)$$

$$\times\left|\frac{(l_{j2}-l_{11}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{s3}+1)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j2}-1)}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m-2]_3\\ [m-1]_2^{j}\\ m_{11}-1\end{matrix}\right\rangle, \tag{2.30}$$

$$e_{30}\left|\begin{matrix}[m-1]_3^{-s}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3}+1)}{\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m-2]_3\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle, \tag{2.31}$$

$$e_{01}\left|\begin{matrix}[m-2]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{j=1}^{2}(l_{l3}+1)S(l,j)S(j,1)$$

$$\times\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3}+2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{j2}-1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-}\\ [m+1]_2^{-j}\\ m_{11}\end{matrix}\right\rangle, \tag{2.32}$$

$$e_{02}\left|\begin{matrix}[m-2]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{j=1}^{2}(l_{l3}+1)S(l,j)$$

$$\times\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3}+2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{j2}-1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-}\\ [m+1]_2^{-j}\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.33}$$

$$e_{03}\left|\begin{matrix}[m-2]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=\sum_{l=1}^{3}(l_{l3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{l3}+2)}{\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-}\\ [m+1]_2\\ m_{11}+1\end{matrix}\right\rangle, \tag{2.34}$$

$$e_{k0}\left|\begin{matrix}[m-2]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=0,\quad k=1,2,3. \tag{2.35}$$

We do not write here the action of the even generators. These relations are known, they have been given in I, Eq. (3.22) and can be easily derived from the anticommutators

$$E_{ij} = \{e_{i0}, e_{0j}\}. \tag{2.36}$$

The dimensions of $\overline{V}([m]_3)$, its even subspace $\overline{V}_0([m]_3)$, and its odd subspace $\overline{V}_1([m]_3)$ are

$$\begin{aligned}\dim \overline{V}([m]_3) &= 2 \dim \overline{V}_0([m]_3) \\ &= 2 \dim \overline{V}_1([m]_3) \\ &= 4(m_{13} - m_{33} + 2)(m_{23} - m_{33} + 1) \\ &\quad \times (m_{13} - m_{23} + 1).\end{aligned} \tag{2.37}$$

For any admissible triple $[m_{13}, m_{23}, m_{33}]$ formulas (2.16)–(2.35) define an induced representation of sl(1,3) in an orthonormal GZ basis. If in addition the condition (2.3) holds, the representation is irreducible. Hence, the family F_t of all typical sl(1,3) modules is determined with Eqs. (2.16)–(2.35) and (2.3). This solves the problem about the typical representations of sl(1,3).

III. NONTYPICAL (IRREDUCIBLE) REPRESENTATIONS

If one of the conditions (2.3) is not fulfilled, the induced representation is indecomposible. The corresponding sl(1,3) module $\overline{V}([m]_3)$ contains a maximal invariant subspace $\overline{I}([m]_3)$, such that its orthogonal complement is not an invariant subspace. The factor module $\overline{V}([m]_3)/\overline{I}([m]_3)$ carries an irreducible nontypical representation of sl(1,3). In order to write the formulas (2.16)–(2.35) in the corresponding factor spaces, we now proceed to determine the maximal invariant subspaces (2.5). To this end we first prove some preliminary assertions.

Proposition 1: Let V be a finite-dimensional irreducible gl(3) module and U be the universal enveloping algebra of gl(3). Then for any nonzero vector $x \in V$,

$$Ux = V. \tag{3.1}$$

The proof is evident, since all finite-dimensional representations of gl(3) are completely reducible. Indeed, if $Ux \subset V$ and in the same time if (3.1) is not true, then Ux will be a proper gl(3) submodule in V, which is impossible.

Proposition 2: Let $\overline{I}([m]_3)$ be the maximal (nontrivial) invariant subspace in $\overline{V}([m]_3)$. Then [see (2.10)]

$$V([m-2]_3) \subset \overline{I}([m]_3). \tag{3.2}$$

Proof: It is easier to carry out the proof using the induced basis [I, (4.12)]. Suppose $0 \neq x \in \overline{I}([m]_3)$. Then

$$\begin{aligned}x = &\sum_{\theta_1,\theta_2,\theta_3 = 0,1} \sum_{(m)_3} \alpha(\theta_1,\theta_2,\theta_3;(m)_3) \\ &\times (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \otimes |(m)_3\rangle_0,\end{aligned} \tag{3.3}$$

where the second sum is over the basis in $V_0([m]_3)$, i.e., over all admissible GZ patterns $(m)_3$ in $V_0([m]_3)$ [remember that $V_0([m]_3)$ is the gl(3) module, inducing $\overline{V}([m]_3)$]. Suppose that all coefficients $\alpha(\theta_1,\theta_2,\theta_3;(m)_3)$ are equal to zero if $\theta_1 + \theta_2 + \theta_3 < k$, and that for certain $\theta_1^0, \theta_2^0, \theta_3^0$, $\theta_1^0 + \theta_2^0 + \theta_3^0 = k$ and for a certain GZ pattern $(m^0)_3$, $\alpha(\theta_1^0,\theta_2^0,\theta_3^0;(m^0)_3) \neq 0$. Then the first sum in (3.3) is over all $\theta_1,\theta_2,\theta_3$ such that $\theta_1 + \theta_2 + \theta_3 \geqslant k$. One can easily derive from [I, (4.12)] and [I, (5.12)] that

$$\begin{aligned}0 \neq &(e_{10})^{1-\theta_1^0}(e_{20})^{1-\theta_2^0}(e_{30})^{1-\theta_3^0} x \\ &\in e_{10}e_{20}e_{30} \otimes V_0([m]_3) \equiv V([m-2]_3).\end{aligned} \tag{3.4}$$

Thus,

$$\begin{aligned}0 \neq &(e_{10})^{1-\theta_1^0}(e_{20})^{1-\theta_2^0}(e_{30})^{1-\theta_3^0} x \\ &\in \overline{I}([m]_3) \cap V([m-2]_3)\end{aligned} \tag{3.5}$$

and, therefore, according to Proposition 1, (3.2) holds.

Proposition 3: The maximal invariant subspace $\overline{I}([m]_3)$ has zero intersection with the gl(3) module $V([m]_3)$:

$$\overline{I}([m]_3) \cap V([m]_3) = 0. \tag{3.6}$$

Proof: Suppose

$$0 \neq x \in \overline{I}([m]_3) \cap V([m]_3), \tag{3.7}$$

then, according to Proposition 1,

$$V([m]_3) \equiv \mathbf{1} \otimes V_0([m]_3) \subset \overline{I}([m]_3). \tag{3.8}$$

Hence, also

$$\begin{aligned}&\left[\sum_{\theta_1,\theta_2,\theta_3} (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3}\right](\mathbf{1} \otimes V_0([m]_3)) \\ &\quad = \sum_{\theta_1,\theta_2,\theta_3} (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \otimes V_0([m]_3) \\ &\quad = \overline{V}([m]_3) \subset \overline{I}([m]_3),\end{aligned}$$

i.e., $\overline{V}([m]_3) = \overline{I}([m]_3)$, which is impossible, since $\overline{I}([m]_3)$ is a proper subspace of $\overline{V}([m]_3)$. ■

Let g be a linear operator in a (finite-dimensional) Hilbert space $\overline{V}$, $\overline{I}$ be a proper g-invariant subspace of $\overline{V}$ with a basis $f_1,\ldots,f_m$, and $\overline{W} = \overline{V} \ominus \overline{I}$ be the orthogonal complement to $\overline{I}$ in $\overline{V}$ with a basis $e_1,\ldots,e_n$. Then

$$\begin{aligned}ge_i &= \sum_{j=1}^{n} A_{ji} e_j + \sum_{p=1}^{m} B_{pi} f_p, \\ gf_p &= \sum_{q=1}^{m} C_{qp} f_q.\end{aligned} \tag{3.9}$$

Consider the factor space $\overline{V}/\overline{I}$ and denote by $x' \in \overline{V}/\overline{I}$ the equivalence class containing $x \in \overline{V}$. The mapping $f(x) = x'$ of $\overline{V}$ on $\overline{V}/\overline{I}$ is linear. Moreover, its restriction on $\overline{W}$ is one-to-one. Indeed, let $x, y \in \overline{W}$ and suppose that $f(x) = f(y)$. Then $f(x-y) = 0$ and, therefore, $x - y \in \overline{I}$. Since, on the other hand, $x - y \perp \overline{I}$, one concludes that $x = y$. Thus, the linear spaces $\overline{W}$ and $\overline{V}/\overline{I}$ are isomorphic. The classes $e_1',\ldots,e_n'$ constitute a basis in $\overline{V}/\overline{I}$, whereas all $f_p' = 0, p = 1,\ldots,m$. In the factor space the operator g acts by definition as $gx' = (gx)'$. Therefore,

$$\begin{aligned}ge_i' = (ge_i)' &= \left(\sum_{j=1}^{n} A_{ji} e_j + \sum_{p=1}^{m} B_{pi} f_p\right)' \\ &= \sum_{j=1}^{n} A_{ji} e_j'.\end{aligned} \tag{3.10}$$

As usually we shall identify the spaces $\overline{W}$ and $\overline{V}/\overline{I}$, denoting with the same symbol, for instance x, both a vector from $\overline{W}$ and the corresponding to it equivalence class from $\overline{V}/\overline{I}$. Then g, considered as operator in the factor space (or, which is the

same, in $\overline{W}$), transforms the basis as

$$ge_i = \sum_{j=1}^{n} A_{ji} e_j. \tag{3.11}$$

Corollary: To obtain the transformation of $\overline{V}/\overline{I} = \overline{W}$ under the action of the operator g one has simply to replace in (3.9) all $f_1,\ldots,f_m$ by zero.

A. The class $m_{13}=0$ nontypical representations

As we have already proved in I, Proposition 3, the induced sl(1,3) module $\overline{V}([0,m_{23},m_{33}])$ is reducible.

Proposition 4: The subspace $I([0,m_{23},m_{33}])$ of $\overline{V}([0,m_{23},m_{33}])$ defined as

$$\begin{aligned} I([0,m_{23},m_{33}]) &= V([0,m_{23}-1,m_{33}-1]) \\ &\oplus V([-2,m_{23}-2,m_{33}-2]) \\ &\oplus V([-1,m_{23}-1,m_{33}-2]) \\ &\oplus V([-1,m_{23}-2,m_{33}-1]) \end{aligned} \tag{3.12}$$

is an sl(1,3)-invariant subspace of the maximal invariant subspace $\overline{I}([0,m_{23},m_{33}])$.

Proof: Using the transformation relations (2.16)–(2.35) one shows in a straightforward way that $I([0,m_{23},m_{33}])$ is an invariant subspace of $\overline{V}([0,m_{23},m_{33}])$. It remains to show that

$$I([0,m_{23},m_{33}]) \subset \overline{I}([0,m_{23},m_{33}]). \tag{3.13}$$

Since [see (3.2)]

$$V([-2,m_{23}-2,m_{33}-2]) \subset \overline{I}([0,m_{23},m_{33}]),$$

then

$$e_{03} V([-2,m_{23}-2,m_{33}-2]) \subset \overline{I}([0,m_{23},m_{33}]).$$

In particular,

$$e_{03} \left| \begin{matrix} -2, & m_{23}-2, & m_{33}-2 \\ & -2, & m_{23}-2 \\ & & -2 \end{matrix} \right\rangle$$

$$= c_1 \left| \begin{matrix} -1, & m_{23}-1, & m_{33}-2 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \equiv y,$$

where

$$c_1 = (l_{33}+1) \left| \frac{(l_{12}-l_{33}+2)(l_{22}-l_{33}+2)}{(1+l_{33})(l_{23}-l_{33})} \right|^{1/2} \neq 0.$$

Proposition 1 now yields that together with the nonzero vector $y \in V([-1,m_{23}-1,m_{33}-2]) \cap \overline{I}([0,m_{23},m_{33}])$,

$$V([-1,m_{23}-1,m_{33}-2]) \subset \overline{I}([0,m_{23},m_{33}]). \tag{3.14}$$

In the equality

$$e_{03} \left| \begin{matrix} -1, & m_{23}-1, & m_{33}-2 \\ & -1, & m_{23}-2 \\ & & -1 \end{matrix} \right\rangle$$

$$= c_2 \left| \begin{matrix} 0, & m_{23}-1, & m_{33}-1 \\ & 0, & m_{23}-1 \\ & & 0 \end{matrix} \right\rangle,$$

the constant c_2 is different from zero (if, certainly, $V([0,m_{23}-1,m_{33}-1]) \neq 0$) and, therefore,

$$V([0,m_{23}-1,m_{33}-1]) \subset \overline{I}([0,m_{23},m_{33}]). \tag{3.15}$$

The last inclusion we need is

$$V([-1,m_{23}-2,m_{33}-1]) \subset \overline{I}([0,m_{23},m_{33}]), \tag{3.16}$$

which follows, for instance, from

$$e_{03} \left| \begin{matrix} 0, & m_{23}-1, & m_{33}-1 \\ & m_{23}-1, & m_{23}-1 \\ & & m_{11} \end{matrix} \right\rangle$$

$$= c_3 \left| \begin{matrix} -1, & m_{23}-2, & m_{33}-1 \\ & m_{23}-2, & m_{23}-2 \\ & & m_{11}-1 \end{matrix} \right\rangle,$$

where $c_3 \neq 0$ if $V([-1,m_{23}-2,m_{33}-1]) \neq 0$.

From (3.2), (3.14), (3.15), and (3.16) one concludes that also the sum (3.12) is a subspace of $\overline{I}([0,m_{23},m_{33}])$, i.e.,

$$I([0,m_{23},m_{33}]) \subset \overline{I}([0,m_{23},m_{33}]). \tag{3.17}$$

which completes the proof.

Proposition 5: The linear subspace $I([0,m_{23},m_{33}])$ is the maximal sl(1,3)-invariant subspace in $\overline{V}([0,m_{23},m_{33}])$.

Proof: For simplicity we introduce the abbreviations

$$\begin{aligned} &\overline{V}([0,m_{23},m_{33}]) = \overline{V}, \quad \overline{I}([0,m_{23},m_{33}]) = \overline{I}, \\ &I([0,m_{23},m_{33}]) = I. \end{aligned} \tag{3.18}$$

Let x be an arbitrary vector from $\overline{I}$,

$$x \in \overline{I}([0,m_{23},m_{33}]) \equiv \overline{I}. \tag{3.19}$$

From (2.10) it follows that every vector from $\overline{V}$ and, in particular, the vector x can be represented uniquely as a sum

$$x = x_0 + \sum_{i=1}^{3} x_i + \sum_{i=1}^{3} x_{-i} + x_4 \tag{3.20}$$

of its projections

$$\begin{aligned} &x_0 \in V([0,m_{23},m_{33}]), \quad x_4 \in V([-2,m_{23}-2,m_{33}-2]), \\ &x_{\pm i} \in V([-1 \pm \delta_{1i}, m_{23}-1 \pm \delta_{2i}, m_{33}-1 \pm \delta_{3i}). \end{aligned} \tag{3.21}$$

Proposition 4 asserts that

$$x_1 + x_{-2} + x_{-3} + x_4 \in \overline{I}, \tag{3.22}$$

and, therefore, also

$$x_0 + x_2 + x_3 + x_{-1} \in \overline{I}. \tag{3.23}$$

We now proceed to show that each term in the sum (3.23) is equal to zero. Suppose first that $x_{-1} \neq 0$ and choose

$g \in U(\mathrm{gl}(3))$ (Proposition 1) to be such that

$$gx_{-1} = \left| \begin{matrix} -2, & m_{23}-1, & m_{33}-1 \\ & -2, & m_{23}-1 \\ & & -2 \end{matrix} \right\rangle. \tag{3.24}$$

Acting on both sides of (3.23) with g one has

$$g(x_0 + x_2 + x_3 + x_{-1}) \in \bar{I}. \tag{3.25}$$

Therefore, also

$$e_{03}E_{32}e_{03}\, g(x_0 + x_2 + x_3 + x_{-1}) \in \bar{I}. \tag{3.26}$$

An explicit computation gives ($c_3 \neq 0$)

$$\begin{aligned} & e_{03}E_{32}e_{03}\, g(x_0 + x_2 + x_3 + x_{-1}) \\ & = e_{03}E_{32}e_{03} \left| \begin{matrix} -2, & m_{23}-1, & m_{33}-1 \\ & -2, & m_{23}-1 \\ & & -2 \end{matrix} \right\rangle \\ & = c_1 e_{03}E_{32} \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23} \\ & & -1 \end{matrix} \right\rangle \\ & = c_2 e_{03} \left| \begin{matrix} -1, & m_{23}, m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \\ & = c_3 \left| \begin{matrix} 0, & m_{23}, & m_{33} \\ & 0, & m_{23} \\ & & 0 \end{matrix} \right\rangle \in \bar{I} \cap V([0, m_{23}, m_{33}]), \end{aligned} \tag{3.27}$$

i.e., $V([0,m_{23},m_{33}]) \cap \bar{I}([0,m_{23},m_{33}]) \neq 0$, which, according to Propositions 3 is impossible. Thus, the assumption that $x_{-1} \neq 0$ is wrong, i.e.,

$$x_{-1} = 0. \tag{3.28}$$

To go further we shall use the relations

$$e_{03} \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle = c_1 \left| \begin{matrix} 0, & m_{23}, & m_{33} \\ & 0, & m_{23} \\ & & 0 \end{matrix} \right\rangle,$$

$$c_1 \neq 0, \tag{3.29}$$

$$e_{03} \left| \begin{matrix} -1, & m_{23}-1, & m_{33} \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle = c_2 \left| \begin{matrix} 0, & m_{23}, & m_{33} \\ & 0, & m_{23} \\ & & 0 \end{matrix} \right\rangle,$$

$$c_2 \neq 0. \tag{3.30}$$

Assume that in (3.23) $x_2 \neq 0$. Choose $g \in U(\mathrm{gl}(3))$ such that

$$gx_2 = \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \tag{3.31}$$

and represent gx_3 as a sum of the highest weight vector (if gx_3 has a nonzero projection on it) of $V([-1,m_{23}-1,m_{33}])$ and an orthogonal-to-it vector y_3,

$$gx_3 = \alpha \left| \begin{matrix} -1, & m_{23}-1, & m_{33} \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle + y_3. \tag{3.32}$$

Then

$$\begin{aligned} g(x_0 + x_2 + x_3) = {} & gx_0 + y_3 \\ & + \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \\ & + \alpha \left| \begin{matrix} -1, & m_{23}-1, & m_{33} \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \in \bar{I}. \end{aligned} \tag{3.33}$$

In order to eliminate the last term in (3.33), we act on both sides with the operator $e_{03}E_{32}E_{23}$. Since

$$E_{23} \left| \begin{matrix} -1, & m_{23}-1, & m_{33} \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle = 0 \tag{3.34}$$

and

$$\begin{aligned} & E_{32}E_{23} \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle \\ & = \beta \left| \begin{matrix} -1, & m_{23}, & m_{33}-1 \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle, \quad \beta \neq 0, \end{aligned} \tag{3.35}$$

we obtain from (3.33)

$$\begin{aligned} z &= e_{03}E_{32}E_{23}(x_0 + x_2 + x_3) \\ &= e_{03}E_{32}E_{23}\, y_3 + \gamma \left| \begin{matrix} 0, & m_{23}, & m_{33} \\ & 0, & m_{23} \\ & & 0 \end{matrix} \right\rangle \\ &\in \bar{I} \cap V([0,m_{23},m_{33}]), \quad \gamma \neq 0. \end{aligned} \tag{3.36}$$

Because of the special choice of y_3 [see (3.32)], the term $e_{03}E_{32}E_{23}\, y_3$ in (3.36) is a linear combination of weight vectors from $V([0,m_{23},m_{33}])$, which does not contain the highest weight vector (= has zero projection on the highest weight vector). Therefore, $z \neq 0$, which is impossible. Hence, the assumption that $x_2 \neq 0$ cannot be true, i.e.,

$$x_2 = 0. \tag{3.37}$$

Suppose that $x_3 \neq 0$ [see (3.23)] and choose $g \in U(\mathrm{gl}(3))$ so that

$$gx_3 = \left| \begin{matrix} -1, & m_{23}-1, & m_{33} \\ & -1, & m_{23}-1 \\ & & -1 \end{matrix} \right\rangle. \tag{3.38}$$

Then ($c\neq 0$)

$$e_{03}\, g(x_0+x_3)=c\left|\begin{matrix}0, & m_{23}, & m_{33}\\ & 0, & m_{23}\\ & & 0\end{matrix}\right\rangle$$

$$\in\bar{I}\cap V([0,m_{23},m_{33}]), \tag{3.39}$$

which is impossible. Hence,

$$x_3=0 \tag{3.40}$$

and, therefore,

$$x_0\in\bar{I}\cap V([0,m_{23},m_{33}]) \tag{3.41}$$

has to be also zero,

$$x_0=0. \tag{3.42}$$

Inserting (3.28), (3.37), (3.40), and (3.42) into (3.20), we finally conclude that every vector $x\in\bar{I}$ can be represented as

$$x=x_1+x_{-2}+x_{-3}+x_4. \tag{3.43}$$

Therefore, $x\in I$ (Proposition 4), i.e., $\bar{I}([0,m_{23},m_{33}])\subset I([0,m_{23},m_{33}])$. This inclusion together with the inverse (3.17) implies that

$$I([0,m_{23},m_{33}])=\bar{I}([0,m_{23},m_{33}]), \tag{3.44}$$

which completes the proof.

Let

$$\begin{aligned}\bar{W}([0,m_{23},m_{33}])&=V([0,m_{23},m_{33}])\oplus V([-1,m_{23},m_{33}-1])\\&\quad\oplus V([-1,m_{23}-1,m_{33}])\\&\quad\oplus V([-2,m_{23}-1,m_{33}-1])\end{aligned} \tag{3.45}$$

be the orthogonal complement to $\bar{I}$ in $\bar{V}([0,m_{23},m_{33}])$. Then, according to the corollary, in order to obtain the transformation of the irreducible nontypical sl(1,3)-module $\bar{V}([0,m_{23},m_{33}])/\bar{I}([0,m_{23},m_{33}])=\bar{W}([0,m_{23},m_{33}])$, one has to insert $m_{13}=0$ ($l_{13}=-1$) in (2.16)–(2.35) and to replace everywhere in these relations the vectors

$$\left|\begin{matrix}[m-1]_3^1\\ [m]_2\\ m_{11}\end{matrix}\right\rangle,\quad\left|\begin{matrix}[m-1]_3^{-2}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle,$$

$$\left|\begin{matrix}[m-1]_3^{-3}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle,\quad\left|\begin{matrix}[m-2]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \tag{3.46}$$

by zero. The final result is the following.

Nontypical representations with $m_{13}=0$

$$e_{0k}\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}, & m_{22}\\ & m_{11}\end{matrix}\right\rangle=0,\quad k=1,2,3, \tag{3.47}$$

$$e_{10}\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}, & m_{22}\\ & m_{11}\end{matrix}\right\rangle=\sum_{i=2}^{3}\sum_{j=1}^{2}S(i,j)S(j,1)\left|\frac{(l_{j2}+1)\Pi_{k\neq i=2,3}(l_{k3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1\\ & m_{11}\end{matrix}\right\rangle, \tag{3.48}$$

$$e_{20}\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}, & m_{22}\\ & m_{11}\end{matrix}\right\rangle=\sum_{i=2}^{3}\sum_{j=1}^{2}S(i,j)\left|\frac{(l_{j2}+1)(l_{j2}-l_{11}+1)\Pi_{k\neq i=2,3}(l_{k3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1\\ & m_{11}-1\end{matrix}\right\rangle, \tag{3.49}$$

$$e_{30}\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}, & m_{22}\\ & m_{11}\end{matrix}\right\rangle=\sum_{i=2}^{3}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{(l_{i3}+1)(l_{23}-l_{33})}\right|^{1/2}\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ & m_{12}-1, & m_{22}-1\\ & & m_{11}-1\end{matrix}\right\rangle, \tag{3.50}$$

$$e_{01}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)S(j,1)$$

$$\times\left|\frac{(l_{j2}+1)\Pi_{k\neq s=2,3}(l_{k3}-l_{j2})\Pi^2_{k\neq j=1}(l_{k2}-l_{11}+1)(l_{k2}-l_{s3})}{(l_{s3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}0, & m_{23}, & m_{23}\\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ & m_{11} & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.51}$$

$$e_{02}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)$$

$$\times\left|\frac{(l_{j2}+1)(l_{j2}-l_{11})\Pi_{k\neq s=2,3}(l_{k3}-l_{j2})\Pi^2_{k\neq j=1}(l_{k2}-l_{s3})}{(l_{s3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.52}$$

$$e_{03}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\left|\frac{\Pi^2_{k=1}(l_{k2}-l_{s3})}{(l_{s3}+1)(l_{23}-l_{33})}\right|^{1/2}\left|\begin{matrix}0, & m_{23}, & m_{33}\\ m_{12}+1, & m_{22}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.53}$$

$$e_{10}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{1is}S(i,j)S(j,1)$$

$$\times\left|\frac{(l_{s3}-l_{j2})(l_{j2}+2)\Pi^2_{k\neq j=1}(l_{k2}-l_{11})(l_{k2}-l_{i3})}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11} & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.54}$$

$$e_{20}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{1is}S(i,j)$$

$$\times\left|\frac{(l_{j2}+2)(l_{11}-l_{j2}-1)(l_{s3}-l_{j2})\Pi^2_{k\neq j=1}(l_{k2}-l_{i3})}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.55}$$

$$e_{30}\left|\begin{matrix}-1, & m_{23}+\delta_{2s}-1, & m_{33}+\delta_{3s}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\epsilon_{1is}\left|\frac{\Pi^2_{k=1}(l_{k2}-l_{i3})}{(l_{i3}+1)(l_{23}-l_{33})}\right|^{1/2}\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ m_{12}-1, & m_{22}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \quad s=2,3, \tag{3.56}$$

$$e_{01}\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{1li}(l_{l3}+1)S(l,j)S(j,1)$$

$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3}+1)}{(l_{l3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ & m_{12}-\delta_{1j}+1, \quad m_{22}-\delta_{2j}+1 & \\ & m_{11} & \end{matrix}\right\rangle, \tag{3.57}$$

$$e_{02}\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{1li}(l_{l3}+1)S(l,j)$$

$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3}+1)}{(l_{l3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{23}-l_{33})}\right|$$

$$\times\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ & m_{12}-\delta_{1j}+1, \quad m_{22}-\delta_{2j}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \tag{3.58}$$

$$e_{03}\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=\sum_{i=1}^{3}\sum_{j=1}^{3}\epsilon_{1ji}(l_{j3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{j3}+1)}{(l_{j3}+1)(l_{23}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix}-1, & m_{23}+\delta_{2i}-1, & m_{33}+\delta_{3i}-1\\ & m_{12}+1, \quad m_{22}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \tag{3.59}$$

$$e_{k0}\left|\begin{matrix}-2, & m_{23}-1, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0, \quad k=1,2,3. \tag{3.60}$$

B. The class $m_{23}=1$ nontypical representations

In a manner similar to the previous section one proves the following.

Proposition 6: If $m_{23}=1$, the maximal invariant subspace $\bar{I}([m_{13},1,m_{33}])$ of $\bar{V}([m_{13},1,m_{33}])$ is

$$\bar{I}([m_{13},1,m_{33}])=V([m_{13}-1,1,m_{33}-1])\oplus V([m_{13}-2,0,m_{33}-1])$$

$$\oplus V([m_{13}-1,0,m_{33}-2])\oplus V([m_{13}-2,-1,m_{33}-2]). \tag{3.61}$$

In order to obtain the irreducible nontypical representaions of sl(1,3) from this class one has, according to the corollary, to insert $m_{23}=1$ ($l_{23}=-1$) everywhere in (2.16)–(2.35), assuming in addition

$$\left|\begin{matrix}m_{13}-1, & 1, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0, \quad \left|\begin{matrix}m_{13}-2, & -1, & m_{33}-2\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0,$$

$$\left|\begin{matrix}m_{13}-2, & 0, & m_{33}-1\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0, \quad \left|\begin{matrix}m_{13}-1, & 0, & m_{33}-2\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0. \tag{3.62}$$

The result gives all the nontypical representations with $m_{23}=1$:

$$e_{0k}\left|\begin{matrix}m_{13}, & 1, & m_{33}\\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle=0, \quad k=1,2,3, \tag{3.63}$$

$$e_{10}\left|\begin{matrix} m_{13}, & 1, & m_{33} \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1,3}\sum_{j=1}^{2} S(i,j)S(j,1)$$

$$\times\left|\frac{(l_{j2}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1,3}(l_{k3}-l_{j2})}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11} & \end{matrix}\right\rangle, \tag{3.64}$$

$$e_{20}\left|\begin{matrix} m_{13}, & 1, & m_{33} \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1,3}\sum_{j=1}^{2} S(i,j)$$

$$\times\left|\frac{(l_{j2}+1)(l_{j2}-l_{11}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1,3}(l_{k3}-l_{j3})}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \tag{3.65}$$

$$e_{30}\left|\begin{matrix} m_{13}, & 1, & m_{33} \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1,3}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{(l_{i3}+1)(l_{13}-l_{33})}\right|^{1/2}\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ & m_{12}-1, \quad m_{22}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \tag{3.66}$$

$$e_{01}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)S(j,1)$$

$$\times\left|\frac{(l_{j2}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{s3})\Pi_{k\neq s=1,3}(l_{k3}-l_{j2})}{(l_{s3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix} & m_{13}, & 1, & m_{33} & \\ m_{12}-\delta_{1j}+1, & & m_{12}-\delta_{2j}+1 & & \\ & & m_{11} & & \end{matrix}\right\rangle, \quad s=1,3, \tag{3.67}$$

$$e_{02}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)$$

$$\times\left|\frac{(l_{j2}+1)(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{s3})\Pi_{k\neq s=1,3}(l_{k3}-l_{j2})}{(l_{s3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{33})}\right|^{1/2}$$

$$\times\left|\begin{matrix} & m_{13}, & 1, & m_{33} & \\ m_{12}-\delta_{1j}+1, & & m_{22}-\delta_{2j}+1 & & \\ & & m_{11}+1 & & \end{matrix}\right\rangle, \quad s=1,3, \tag{3.68}$$

$$e_{03}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3})}{(l_{s3}+1)(l_{13}-l_{33})}\right|^{1/2}\left|\begin{matrix} m_{13}, & 1, & m_{33} \\ m_{12}+1, & m_{22}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \quad s=1,3, \tag{3.69}$$

$$e_{10}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{2is}S(i,j)S(j,1)$$
$$\times\left|\frac{(l_{s3}-l_{j2})(l_{j2}+2)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{33})(l_{i3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ m_{11} & & \end{matrix}\right\rangle, \quad s=1,3, \tag{3.70}$$

$$e_{20}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{2is}S(i,j)$$
$$\times\left|\frac{(l_{j2}+2)(l_{11}-l_{j2}-1)(l_{s3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})}{(l_{i3}+1)(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{33})}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ m_{11}-1 & & \end{matrix}\right\rangle, \quad s=1,3, \tag{3.71}$$

$$e_{30}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & 0, & m_{33}+\delta_{3s}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\epsilon_{2is}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{(l_{13}-l_{33})(l_{i3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}-1, & m_{22}-1 & \\ m_{11}-1 & & \end{matrix}\right\rangle, \quad s=1,3 \tag{3.72}$$

$$e_{01}\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{2li}(l_{l3}+1)S(l,j)S(j,1)$$
$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3}+1)}{(l_{l3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{33})}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ m_{11} & & \end{matrix}\right\rangle, \tag{3.73}$$

$$e_{02}\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{2li}(l_{l3}+1)S(l,j)$$
$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3}+1)}{(l_{l3}+1)(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{33})}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ m_{11}+1 & & \end{matrix}\right\rangle, \tag{3.74}$$

$$e_{03}\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ m_{12}, & m_{22} & \\ m_{11} & & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{3}(l_{j3}+1)\epsilon_{2ji}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{j3}+1)}{(l_{13}-l_{33})(l_{j3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & 0, & m_{33}+\delta_{3i}-1 \\ m_{12}+1, & m_{22}+1 & \\ m_{11}+1 & & \end{matrix}\right\rangle, \tag{3.75}$$

$$e_{k0}\left|\begin{matrix} m_{13}-1, & -1, & m_{33}-1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0, \quad k = 1,2,3.$$

C. The class $m_{33}=2$ nontypical representations

Proposition 7: If $m_{33} = 2$ the maximal sl(1,3)-invariant subspace $\bar{I}([m_{13},m_{23},2])$ of $\bar{V}([m_{13},m_{23},2])$ is

$$\bar{I}([m_{13},m_{23},2]) = V([m_{13}-1,m_{23}-1,2]) \oplus V([m_{13}-1,m_{23}-2,1]) \oplus V([m_{13}-2,m_{23}-1,1]) \oplus V([m_{13}-2,m_{23}-2,0]). \tag{3.76}$$

According to the corollary, one obtains this class of nontypical irreducible representations of sl(1,3) by inserting $m_{33} = 2$ in (2.16)–(2.35) and assuming that

$$\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 2 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0, \quad \left|\begin{matrix} m_{13}-2, & m_{23}-2, & 0 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0,$$

$$\left|\begin{matrix} m_{13}-1, & m_{23}-2, & 1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0, \quad \left|\begin{matrix} m_{13}-2, & m_{23}-1, & 1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0.$$

As a result one obtains all the nontypical representations with $m_{33} = 2$:

$$e_{0k}\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0, \quad k = 1,2,3, \tag{3.77}$$

$$e_{10}\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{2}\sum_{j=1}^{2} S(i,j)S(j,1)\left|\frac{(l_{j2}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{2}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ & m_{12}+\delta_{1j}-1, \quad m_{22}+\delta_{2j}-1 & \\ & m_{11} & \end{matrix}\right\rangle, \tag{3.78}$$

$$e_{20}\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{2}\sum_{j=1}^{2} S(i,j)\left|\frac{(l_{j2}+1)(l_{j2}-l_{11}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{2}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ & m_{12}+\delta_{1j}-1, \quad m_{22}+\delta_{2j}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \tag{3.79}$$

$$e_{30}\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{2}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ & m_{12}-1, \quad m_{22}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \tag{3.80}$$

$$e_{01}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)S(j,1)$$
$$\times\left|\frac{(l_{j2}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{s3})\Pi_{k\neq s=1}^{2}(l_{k3}-l_{j2})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{33})(l_{s3}+1)}\right|^{1/2}$$
$$\times\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ & m_{12}-\delta_{1j}+1, \quad m_{22}-\delta_{2j}+1 & \\ & m_{11} & \end{matrix}\right\rangle, \quad s = 1,2, \tag{3.81}$$

$$e_{02}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\sum_{j=1}^{2} S(s,j)$$

$$\times\left|\frac{(l_{j2}+1)(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{s3})\Pi_{k\neq s=1}^{2}(l_{k3}-l_{j2})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{23})(l_{s3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1) & \\ & m_{11}+1 & \end{matrix}\right\rangle, \quad s=1,2, \tag{3.82}$$

$$e_{03}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = (l_{s3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3})}{(l_{13}-l_{23})(l_{s3}+1)}\right|^{1/2}\left|\begin{matrix} m_{13}, & m_{23}, & 2 \\ m_{12}+1, & m_{22}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \quad s=1,2, \tag{3.83}$$

$$e_{10}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{3is}S(i,j)S(j,1)$$

$$\times\left|\frac{(l_{s3}-l_{j2})(l_{j2}+2)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11} & \end{matrix}\right\rangle, \quad s=1,2, \tag{3.84}$$

$$e_{20}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{3is}S(i,j)$$

$$\times\left|\frac{(l_{j2}+2)(l_{11}-l_{j2}-1)(l_{s3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ m_{12}+\delta_{1j}-1, & m_{22}+\delta_{2j}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \quad s=1,2, \tag{3.85}$$

$$e_{30}\left|\begin{matrix} m_{13}+\delta_{1s}-1, & m_{23}+\delta_{2s}-1, & 1 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\epsilon_{3is}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{(l_{13}-l_{23})(l_{i3}+1)}\right|^{1/2}\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ m_{12}-1, & m_{22}-1 & \\ & m_{11}-1 & \end{matrix}\right\rangle, \quad s=1,2, \tag{3.86}$$

$$e_{01}\left|\begin{matrix} m_{3}-1, & m_{23}-1, & 0 \\ m_{12}, & m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{3li}(l_{l3}+1)S(l,j)S(j,1)$$

$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{23})(l_{l3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ & m_{11} & \end{matrix}\right\rangle, \tag{3.87}$$

$$e_{02}\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{3li}(l_{l3}+1)S(l,j)$$

$$\times\left|\frac{(l_{j2}+2)(l_{i3}-l_{j2})(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)(l_{13}-l_{23})(l_{l3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ m_{12}-\delta_{1j}+1, & m_{22}-\delta_{2j}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \tag{3.88}$$

$$e_{03}\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{3}(l_{j3}+1)\epsilon_{3ji}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{j3}+1)}{(l_{13}-l_{23})(l_{j3}+1)}\right|^{1/2}$$

$$\times\left|\begin{matrix} m_{13}+\delta_{1i}-1, & m_{23}+\delta_{2i}-1, & 1 \\ & m_{12}+1, \quad m_{22}+1 & \\ & m_{11}+1 & \end{matrix}\right\rangle, \tag{3.89}$$

$$e_{k0}\left|\begin{matrix} m_{13}-1, & m_{23}-1, & 0 \\ & m_{12}, \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle = 0, \quad k=1,2,3. \tag{3.90}$$

IV. NONDECOMPOSIBLE REPRESENTATIONS OF LOWEST DIMENSION

It was already shown that the induced sl(1,3) modules $\overline{V}([0,m_{23},m_{33}])$, $\overline{V}([m_{13},1,m_{33}])$, and $\overline{V}([m_{13},m_{23},2])$ are reducible, but not completely reducible. Here we consider the lowest dimensional cases. There are three such modules, each one of dimension 8.

(1) The space

$$\overline{V}([0,0,0]) = \overline{W}([0,0,0]) \oplus \overline{I}([0,0,0]), \tag{4.1}$$

with a maximal invariant subspace

$$\overline{I}([0,0,0]) = V([-1,-1,-2]) \oplus V([0,-1,-1]) \oplus V([-2,-2,-2]) \tag{4.2}$$

and an orthogonal complement to $\overline{I}([0,0,0])$

$$\overline{W}([0,0,0]) = V([0,0,0]). \tag{4.3}$$

In this case

$$\dim \overline{I}([0,0,0]) = 7, \quad \dim \overline{W}([0,0,0]) = 1. \tag{4.4}$$

Therefore, in a proper homogeneous basis the elements of sl(1,3), written as matrices, will have the form

$$\mathrm{sl}(1,3) = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{28} \\ \vdots & \vdots & & \vdots \\ a_{81} & a_{82} & \cdots & a_{88} \end{pmatrix}. \tag{4.5}$$

(2) The space

$$\overline{V}([1,1,1]) = \overline{W}([1,1,1]) \oplus \overline{I}([1,1,1]), \tag{4.6}$$

with a maximal invariant subspace

$$\overline{I}([1,1,1]) = V([0,0,-1]) \oplus V([-1,-1,-1]) \tag{4.7}$$

and an orthogonal complement

$$\overline{W}([1,1,1]) = V([1,1,1]) \oplus V([1,0,0]); \tag{4.8}$$

$$\dim \overline{I}([1,1,1]) = \dim \overline{W}([1,1,1]) = 4. \tag{4.9}$$

In this representation the elements of sl(1,3), written as matrices, will have the form

$$\mathrm{sl}(1,3) = \begin{pmatrix} a_{11} & \cdots & a_{14} & 0 & \cdots & 0 \\ \vdots & & & \vdots & & \vdots \\ a_{41} & \cdots & a_{44} & 0 & \cdots & 0 \\ a_{51} & \cdots & a_{54} & a_{55} & \cdots & a_{58} \\ \vdots & & & & & \\ a_{81} & \cdots & a_{84} & a_{85} & \cdots & a_{88} \end{pmatrix}. \tag{4.10}$$

(3) The space

$$\overline{V}([2,2,2]) = \overline{W}([2,2,2]) \oplus \overline{I}([2,2,2]), \tag{4.11}$$

with a maximal invariant subspace

$$\overline{I}([2,2,2]) = V([0,0,0]) \tag{4.12}$$

and an orthogonal complement

$$\overline{W}([2,2,2]) = V([2,1,1]) \oplus V([1,1,0]) \oplus V([2,2,2]); \tag{4.13}$$

$$\dim \overline{I}([2,2,2]) = 1, \quad \dim \overline{W}([2,2,2]) = 7. \tag{4.14}$$

In a matrix form this representation reads

$$\mathrm{sl}(1,3) = \begin{pmatrix} a_{11} & \cdots & a_{17} & 0 \\ \vdots & & \vdots & \vdots \\ a_{71} & \cdots & a_{77} & 0 \\ a_{81} & \cdots & a_{87} & a_{88} \end{pmatrix}. \tag{4.15}$$

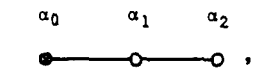
V. A MAPPING OF THE YOUNG SUPERTABLEAU BASIS (WEYL BASIS) ON THE GEL'FAND–ZETLIN BASIS

In Refs. 4 and 5 it has been shown that the Young tableau technique can be generalized also in the case of LS's. This approach has been developed in more detail for the LS sl(m,n) (Ref. 6) and along this line several important properties of the finite-dimensional representations of sl(m,n) have been derived.[7–9] In this section we translate our results in the Young supertableau language.[5] In a manner similar to that of Ref. 10 we construct a mapping of the Young supertableau basis on the GZ basis.

To begin with we recall that a given irreducible sl(1,3) module $W([m]_3)$ is characterized in our notation by the coordinate m_{13},m_{23},m_{33} of its highest weight Λ: $m_{i3}=\Lambda(E_{ii})$, which are the eigenvalues of the Cartan generators (2.1) on the highest weight vector x_Λ, i.e.,

$$E_{ii}x_\Lambda = m_{i3}x_\Lambda, \quad i=1,2,3. \tag{5.1}$$

In the notation of Kac[2] the same module is represented by a Kac–Dynkin diagram

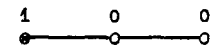


(5.2)

where $(\alpha_0,\alpha_1,\alpha_2)$ are the coordinates of Λ in the dual to $h_0=E_{11}$, $h_1=E_{11}-E_{22}$, $h_2=E_{22}-E_{33}$ basis, i.e., $\alpha_A=\Lambda(h_A)$, $A=0,1,2$, and, therefore, (see I.4.23)

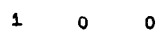
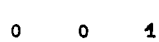
$$\alpha_0=m_{13}, \quad \alpha_1=m_{13}-m_{23}, \quad \alpha_2=m_{23}-m_{33}. \tag{5.3}$$

The defining (= the fundamental) representation of sl(1,3), given with the matrices (2.1) and (2.2) is realized in the space

$W([1,1,1]) =$ 

(5.4)

In the Bars–Balantekin notation $W([1,1,1])$ is denoted ☐ and its elements are called covariant tensors of first rank. The conjugate to $W([1,1,1])$ module $W([0,0,-1])$ is denoted ☐ and its elements are the contravariant tensors of first rank. Thus, we have

	Present paper	Kac	Balantekin Bars
Fundamental sl(1,3) module:	$W(1,1,1]) =$		= ☐ (covariant tensors),
Conjugate to the fundamental:	$W([0,0,-1]) =$		= ☐ (contravariant tensors).

(5.5)

The above relations indicate that the covariant and the contravariant tensors are transformed according to nontypical representations of sl(1,3). These representations [contrary to the case of the LA sl(n)] are inequivalent. Both ☐ and ☐ spaces are four dimensional. One can introduce a grading in them in two ways. Choose a basis e_A in ☐ (resp. g^A in ☐), $A=0,1,2,3$ in such a way that the B th coordinate of e_A is δ_{AB}, i.e.,

$$(e_A)_B=\delta_{AB} \quad [\text{resp. } (g^A)^B=\delta^{AB}]. \tag{5.6}$$

The representation of sl(1,3) in ☐ (resp. in ☐) is said to be class I, if the degree (A) of e_A (resp. of g^A) is

$$(A)=0 \text{ for } A=0 \text{ and } (A)=1 \text{ for } A=1,2,3. \tag{5.7}$$

If

$$(A)=1 \text{ for } A=0 \text{ and } (A)=0 \text{ for } A=1,2,3, \tag{5.8}$$

the representation is said to be of class II. The class I and the class II representations in ☐ (resp. in ☐) are equivalent. Therefore, without loss of generality we consider only class I representations.

The sl(1,3) generators transform the basis in ☐ as follows (we write e_{ij} instead of E_{ij}, $i,j=1,2,3$):

$$e_{AB}e_C=\delta_{BC}e_A, \quad A,B,C=0,1,2,3. \tag{5.9}$$

From (5.9) and (3.63)–(3.75) one obtains the relation between the basis e_A and the GZ basis in ☐:

$$e_0=\left|\begin{matrix}1,1,1\\ 1,1\\ 1\end{matrix}\right\rangle, \quad e_1=\left|\begin{matrix}1,0,0\\ 1,0\\ 0\end{matrix}\right\rangle,$$
$$e_2=\left|\begin{matrix}1,0,0\\ 1,0\\ 0\end{matrix}\right\rangle, \quad e_3=\left|\begin{matrix}1,0,0\\ 0,0\\ 0\end{matrix}\right\rangle. \tag{5.10}$$

Similarly,

$$e_{AB}\,g^C=-(-1)^{[(A)+(B)](C)}\delta_{AC}\,g^B, \tag{5.11}$$

which together with (3.47)–(3.60) yields

$$g^0=\left|\begin{matrix}-1,&-1,&-1\\ &-1,&-1\\ &&-1\end{matrix}\right\rangle,$$
$$g^1=\sqrt{3}\left|\begin{matrix}0,&0,&-1\\ &0,&-1\\ &&-1\end{matrix}\right\rangle,$$
$$g^2=-\sqrt{3}\left|\begin{matrix}0,&0,&-1\\ &0,&-1\\ &&0\end{matrix}\right\rangle, \tag{5.12}$$
$$g^3=\sqrt{3}\left|\begin{matrix}0,&0,&-1\\ &0,&0\\ &&0\end{matrix}\right\rangle.$$

A. Covariant representations

The sl(1,3) modules corresponding to these representations are tensor products of covariant rank-1 tensors. The pth tensor power of $\boxtimes$, namely

$$W(0;p) = \boxtimes \otimes \boxtimes \otimes \cdots \otimes \boxtimes \quad (p \text{ times}) , \tag{5.13}$$

is a carrier space of the covariant tensors of rank p with a homogeneous basis

$$e_{A_1} \otimes e_{A_2} \otimes \cdots \otimes e_{A_p}, \quad A_1,A_2,\ldots,A_p = 0,1,2,3. \tag{5.14}$$

In general this space is reducible. The key point for decomposing it into a direct sum of irreducible modules comes from the observation[4,5] that the LS commutes with the group $\tilde{S}_p$ of graded permutations, where $\tilde{S}_p$ gives a faithful representation of the permutation group S_p in $W(0,p)$. The operator $\pi(i,i+1) \in \tilde{S}_p$, corresponding to the neighbor transposition $(i,i+1) \in S_p$, acts as

$$\pi(i,i+1)\left[\cdots \otimes e_{A_i} \otimes e_{A_{i+1}} \otimes \cdots\right] = -(-1)^{(A_i)(A_{i+1})}\left[\cdots \otimes e_{A_{i+1}} \otimes e_{A_i} \otimes \cdots\right]. \tag{5.15}$$

To determine the action of any other permutation it suffices to represent it as a product of such neighbor transpositions, which is always possible. Thus, $W(0,p)$ is turned into an S_p module. Decompose it into a direct sum of irreducible S_p modules $W(0,[\lambda]_p)$,

$$W(0,p) = \sum_{[\lambda]} \oplus W(p;[\lambda]_p). \tag{5.16}$$

Then each term $W(0;[\lambda]_p)$ turns to be an irreducible module over the LS (which is sl(1,3) in our case) and it can be represented as

$$W(0;[\lambda]_p) = Y(0;[\lambda]_p)W(0,p), \tag{5.17}$$

where the Young symmetrizer $Y(0;[\lambda]_p)$ is a projection operator, corresponding to the Young tableau $[\lambda]_p$, which, following Ref. 5, will be called a supertableau. The decomposition (5.16) is rather standard. The sum is over all legal Young supertableaux. A given supertableau $[\lambda]_p$,

$$[\lambda]_p = \begin{array}{l} {\scriptstyle c_1 \ \cdots \ c_m} \\ \boxtimes\boxtimes\boxtimes\boxtimes\boxtimes\boxtimes\boxtimes \\ \boxtimes\boxtimes\boxtimes\boxtimes \\ \boxtimes\boxtimes \\ \boxtimes\boxtimes \end{array} \begin{array}{l} b_1 \\ \cdot \\ \cdot \\ b_n \end{array} , \tag{5.18}$$

contains p covariant boxes, where

b_i counts the boxes of the row i,

c_j counts the boxes of the column j,

and

$$b_1 \geqslant b_2 \geqslant \cdots \geqslant b_m > 0, \qquad c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n > 0. \tag{5.19}$$

For sl(1,3) the supertableau $[\lambda]_p$ is legal if[7]

$$b_2 \leqslant 3 \Leftrightarrow c_4 \leqslant 1. \tag{5.20}$$

The boxes are enumerated with the numbers from 1 to p $(1,2,\ldots,p)$ lexically: the numbers in each row increase as read from left to right and in column as read from top to bottom.

Turn now to the basis labeling problem in $W(0;[\lambda]_p)$. The action of the symmetrizer $Y(0;[\lambda]_p)$ is defined in a similar way as for the LA's (see, for instance, Ref. 10), however, because of the sign factor in (5.15) the odd indices appearing in one and the same row (resp. column) are antisymmetrized (resp. symmetrized). Therefore, the basis in $W(0;[\lambda]_p)$ is in one-to-one correspondence with the standard Weyl patterns $(\lambda)_p$ containing p boxes. For sl(1,3) a Weyl pattern $(\lambda)_p$ is a Young supertableau $[\lambda]_p$ in which the boxes have been "filled in" with integers selected from 1,2,3, and 0. It is convenient to assume that

$$1 < 2 < 3 < 0. \tag{5.21}$$

In this ordering the Weyl pattern $(\lambda)_p$ is standard if the sequence of integers 1,2,3 appearing in each row of $[\lambda]_p$ is strictly increasing as read from left to right followed by any admissible (from the shape of the diagram) number of zeros and the sequence of integers 1,2,3 appearing in each column is nondecreasing as read from top to bottom, followed by no more than one 0. For instance, if $p = 15$, then the table

$$(\lambda)_{15} = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 3 & 0 & 0 \\ \hline 1 & 2 & 3 \\ \cline{1-3} 1 & 2 & 3 \\ \cline{1-3} 1 & 2 & 0 \\ \cline{1-3} 0 \\ \cline{1-1} \end{array} \tag{5.22}$$

is an example of a Weyl pattern, i.e., a Weyl basis vector in $W(0;[\lambda]_{15})$.

We are now ready to define a mapping F of the Weyl basis of a given irreducible sl(1,3) module onto the GZ basis. The idea is the same as developed in Ref. 10 for the LA gl(n). The mapping F may be defined by three projection operators F_1, F_2, and F_3. To determine them take any weight vector $x \in W(0;[\lambda]_p)$.

(1) Consider all nonzero projections of x on the gl(3) irreducible submodules from $W(0;[\lambda]_p)$. Then F_3x is the projection of x on those gl(3) submodule $V([m]_3)$, which has the biggest gl(3)-highest weight.

(2) Similarly, F_2F_3x is the projection of F_3x on that gl(2) submodule $V([m]_2) \subset V([m]_3)$ which has the biggest highest weight among all those sl(2) submodules in $V([m]_3)$ on which F_3x has nonzero projections.

(3) $F_1F_2F_3x$ is the projection of F_2F_3x on that gl(1) submodule $V(m_{11}) \subset V([m]_2)$ which has the biggest highest weight among all those gl(1) submodules from $V([m]_2)$ on which F_2F_3x has nonzero projections.

Then Fx is the normed to unity vector $F_1F_2F_3x$, i.e., [see (2.6)]

$$Fx = \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle. \tag{5.23}$$

The operator F defines one to one mapping of the Weyl pattern basis onto the GZ basis. To write down the action of F explicitly one has to take into account the following.

(a) Each index $k = 1,\ldots,n$ appearing in the k th column of the Weyl pattern $(\lambda)_p$ can be transferred into k by a proper action of the gl(n) generators, $n = 1,2,3$.

(b) The basis vectors in $\boxtimes$ are weight vectors. The corre-

spondence with their weight is

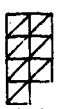
$$e_0 \leftrightarrow (1,1,1), \quad e_1 \leftrightarrow (1,0,0),$$
$$e_2 \leftrightarrow (0,1,0), \quad e_3 \leftrightarrow (0,0,1). \tag{5.24}$$

As a result one has

$$F(\lambda)_p = \left| \begin{matrix} m_{13}, & m_{23}, & m_{33} \\ & m_{12}, & m_{22} \\ & & m_{11} \end{matrix} \right\rangle, \tag{5.25}$$

where m_{ij} is equal to the number of the 1's, 2's,..., j's, which appear in the ith column of $(\lambda)_p$ plus the number of all zeros in $(\lambda)_p$.

From (5.25) one derives that the set of all Young supertableaux with two covariant columns corresponds to the class $m_{33} = 2$ nontypical modules. More precisely,



$$= W([m_{13},m_{23},2]). \tag{5.26}$$

$m_{23} - 1$ boxes in the second column

$m_{13} - 1$ boxes in the first column

B. Contravariant tensor representations

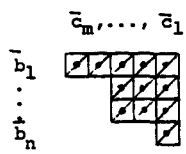
The representation space $W(q;0)$ of the contravariant tensors of rank q is

$$W(q;0) = \square \otimes \square \otimes \cdots \otimes \square \quad (q \text{ times}), \tag{5.27}$$

and it decomposes into a direct sum of irreducible sl(1,3) modules $W([\bar{\lambda}]_q;0)$:

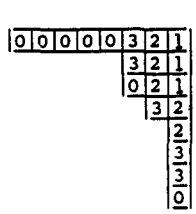
$$W(q;0) = \sum_{[\bar{\lambda}]_q} \oplus W([\bar{\lambda}]_q;0). \tag{5.28}$$

The sum is over all contravariant legal Young supertableaux $[\bar{\lambda}]_q$. The supertableau $[\bar{\lambda}]_q$ contains q boxes $\square$ ordered as



$$\equiv [\bar{\lambda}]_q, \tag{5.29}$$

and it is legal if the inequalities (5.19) and (5.20) hold for $\bar{b}_i$, $\bar{c}_j$, $i = 1,\ldots,n$ and $j = 1,\ldots,m$. All boxes of $[\bar{\lambda}]_q$ are enumerated lexically with 1, 2,..., q from right to left and from top to bottom. The definition of the Weyl pattern $(\bar{\lambda})_q$ corresponding to $[\bar{\lambda}]_q$ is as in the case of the covariant boxes. The only difference is that one "fills in" 1, 2, 3, 0 from right to left. All $(\bar{\lambda})_q$ corresponding to $[\bar{\lambda}]_q$ constitute a basis, the Weyl basis, in $W([\bar{\lambda}]_q;0)$. The diagram



(5.30)

is an example of a Weyl basis vector in $W([\bar{\lambda}]_{20};0)$.

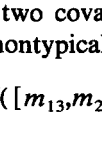
The mapping F of the Weyl patterns basis $(\bar{\lambda})_q$ onto the GZ basis is easily derived if one takes into account that the correspondence between the basis vectors g^A in $\square$ and their weight is

$$g^0 \leftrightarrow (-1,-1,-1), \quad g^1 \leftrightarrow (-1,0,0),$$
$$g^2 \leftrightarrow (0,-1,0), \quad g^3 \leftrightarrow (0,0,-1). \tag{5.31}$$

Then

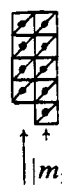
$$F(\bar{\lambda})_q = \left| \begin{matrix} m_{13}, & m_{23}, & m_{33} \\ & m_{12}, & m_{22} \\ & & m_{11} \end{matrix} \right\rangle, \tag{5.32}$$

where $(-m_{ij})$ is equal to the number of the 1's, 2's,..., j's, which appears in the $(j-i+1)+$th column of $(\bar{\lambda})_q$ plus all zeros in $(\bar{\lambda})_q$. For instance,



$$= \left| \begin{matrix} -1, & -4, & -7 \\ & -2, & -4 \\ & & -3 \end{matrix} \right\rangle \in W([0,-4,-6]). \tag{5.33}$$

The correspondence in this case is also one to one. The set of all Young supertableaux with two contravariant columns gives the class $m_{13} = 0$ nontypical modules:



$$= W([0,m_{23},m_{33}]). \tag{5.34}$$

$|m_{33}|$ boxes in the first column

$|m_{23}|$ boxes in the second column

C. Mixed covariant–contravariant tensor representations

The representation space of the tensors with q contravariant and p covariant indices is

$$W(q;p) = \underbrace{\square \otimes \cdots \otimes \square}_{q \text{ times}} \otimes \underbrace{\square \otimes \cdots \otimes \square}_{p \text{ times}}. \tag{5.35}$$

It resolves into a direct sum of sl(1,3) invariant subspaces $W([\bar{\lambda}]_q;0) \otimes W(0;[\lambda]_p)$:

$$W(q;p) = \sum_{[\bar{\lambda}]_q [\lambda]_p} \oplus W([\bar{\lambda}]_q;0) \otimes W(0;[\lambda]_p). \tag{5.36}$$

The sum is over all legal Young supertableaux $[\bar{\lambda}]_q$ and $[\lambda]_p$. Each mixed tensor

$$t \in W([\bar{\lambda}]_q;0) \otimes W(0;[\lambda]_p), \quad p,q \neq 0, \tag{5.37}$$

can be uniquely decomposed in terms of the tensor basis:

$$t = \sum_{\substack{A_1,\ldots,A_q \\ B_1,\ldots,B_p}} t^{A_1,\ldots,A_q}_{B_1,\ldots,B_p} g^{A_1} \otimes \cdots \otimes g^{A_q} \otimes e_{B_1} \otimes \cdots \otimes e_{B_p}. \tag{5.38}$$

Let

$$W([\bar{\lambda}]_q;[\lambda]_p) \subset W([\bar{\lambda}]_q;0) \otimes W(0;[\lambda]_p) \tag{5.39}$$

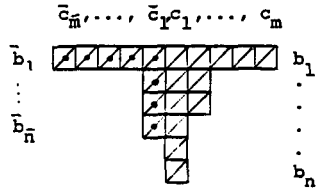
be the subspace of all tensors such that the supertrace with respect to any two pairs of covariant and contravariant in-

dices vanishes, i.e.,

$$\sum_{C=0}^{3}(-1)^{(C)}\, t^{\ldots,A_{r-1},C,A_{r+1},\ldots}_{\ldots,B_{s-1},C,B_{s+1},\ldots}=0,$$

$$\forall\, r=1,\ldots,q,\ s=1,\ldots,p. \quad (5.40)$$

Then $W([\bar{\lambda}\,]_q;[\lambda\,]_p)$ is either an irreducible or nondecomposible sl(1,3) module.[7-9] The Young supertableau corresponding to it is

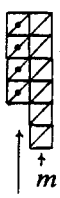


(5.41)

Such a supertableau is legal if[8]

$$b_1+b_2\leqslant 3, \quad \text{or} \quad b_2+b_1\leqslant 3$$

and its contravariant and covariant boxes are enumerated with the numbers 1, 2,..., q and 1,..., p as this was already described for the pure contravariant and pure covariant tensors. The mixed tensors with one covariant and one contravariant rows describe all nontypical representations from the class $m_{23}=1$. The correspondence is

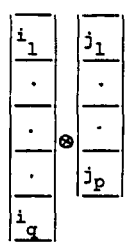


$$=W([m_{13},1,m_{33}]). \quad (5.42)$$

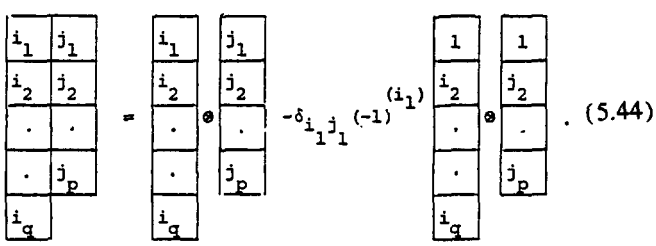
m_{13} covariant boxes

$1-m_{33}$ contravariant boxes

In this case one cannot define a Weyl pattern in the way this was done for the purely covariant or purely contravariant tensors. One possible generalization of this concept is the following. Consider for simplicity only Young supertableaux of the shape (5.42). The tensor product of the Weyl patterns,

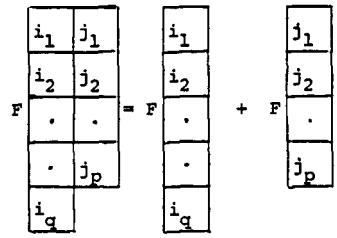
contravariant boxes →  ← covariant boxes, (5.43)

constitute a basis in $W([\bar{\lambda}\,]_q;0)\otimes W(0;[\lambda\,]_p)$. We set



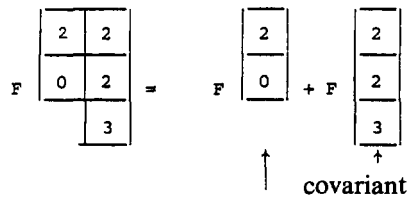
$$-\delta_{i_1 j_1}(-1)^{(i_1)} \quad (5.44)$$

The patterns (5.44) correspond to traceless tensors and constitute a basis in $W([\bar{\lambda}\,]_q;[\lambda\,]_p)$. We call them Weyl patterns. Then the one-to-one mapping of the Weyl pattern basis onto the GZ basis is defined as



(5.45)

For instance,



$$=\left|\begin{matrix}-1,&-1,&-1\\&-1,&-2\\&&-1\end{matrix}\right\rangle+\left|\begin{matrix}3,&0,&0\\&2,&0\\&&0\end{matrix}\right\rangle$$

$$=\left|\begin{matrix}2,&-1,&-2\\&-1,&-2\\&&-1\end{matrix}\right\rangle. \quad (5.46)$$

We conclude that the Young supertableaux describe all nontypical representations of the Lie superalgebra sl(1,3). With this technique one can construct several other representations, irreducible or nondecomposible, all of them corresponding to integer coordinates $[m_{13},m_{23},m_{33}]$ of the highest weight.

VI. CONCLUDING REMARKS

In Ref. 1 and in the present paper we have constructed all finite-dimensional irreducible representations of the Lie superalgebra sl(1,3). More precisely, we wrote down explicit expressions for the transformation of the basis within every finite-dimensional irreducible sl(1,3) module under the action of the generators. In solving the problem we have essentially used the results of Kac[2] on the induced representations of the basic Lie superalgebras. The main difficulty we had to overcome was to introduce a basis in such a way that every basis vector has a nonzero projection only on one irreducible gl(3) submodule. To this end, we have essentially used the tensor properties of the odd generators under the adjoint representation of the even subalgebra gl(3). This allowed us to establish that every induced sl(1,3) module $\overline{V}([m_{13},m_{23},m_{33}])$ can be considered as a direct sum of four gl(1,3)-invariant subspaces, each such subspace being a tensor product of two irreducible gl(3) modules. Therefore, the coefficients, connecting the tensor-product basis (the induced basis) with the basis we were looking for (the GZ basis), can be chosen to be Clebsch–Gordan coefficients of gl(3). Since we knew the action of the generators on the induced basis, the problem to express the transformation of the GZ basis was reduced to a rather standard transforma-

tion from one basis (the induced one) to another basis (the GZ one). The main difficulty along this way was of a technical nature—one had to sum several terms (the coefficients in front of the basis vectors) and it was not clear from some general considerations that the summation can be carried out. Luckily, because of several, presumably not accidental cancellations, we succeeded in writing the final result in a rather simple form. The similarity of the formulas (2.16)–(2.35) suggests that it should be possible to go further and to unify them (as in case of Lie algebras), expressing all of them in terms of only two relations—one for the positive root generators and one for the negative root generators. We leave this task for the future.

The method that has been used here can be immediately applied also to the Lie superalgebra sl(1,n), since in this case the necessary for the computation Clebsch–Gordan coefficients are known. From a technical point of view the problem will be more difficult. We believe, however, that also in this more general case it will be possible to carry out all summations and to write down the final results in a closed form.

[1]T. D. Palev, "Finite-dimensional representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis. I. Typical representations," J. Math. Phys. **26**, 1640 (1985).

[2]V. G. Kac, Lect. Notes Math. **676**, 597 (1978).

[3]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).

[4]P. H. Dondi and P. D. Jarvis, Z. Phys. C **4**, 201 (1980); J. Phys. A **14**, 547 (1981).

[5]A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149 (1981); **22**, 1810 (1981).

[6]A. B. Balantekin and I. Bars, J. Math. Phys. **23**, 1239 (1982); A. B. Balantekin, J. Math. Phys. **23**, 486 (1982).

[7]I. Bars, B. Morel, and H. Ruegg, J. Math. Phys. **24**, 2253 (1983); I. Bars and M. Günaydin, Commun. Math. Phys. **91**, 31 (1983); I. Bars, Physica D **15**, 42 (1985).

[8]F. Delduc and M. Gourdin, J. Math. Phys. **25**, 1651 (1984).

[9]F. Delduc and M. Gourdin, J. Math. Phys. **26**, 1865 (1985).

[10]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); see also *Angular Momentum in Quantum Physics. Theory and Applications*, edited by L. C. Biedenharn and J. D. Louck (Addison–Wesley, Reading, MA, 1981).

Regular subalgebras of Lie superalgebras and extended Dynkin diagrams

Joris Van der Jeugt[a)]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B9000 Gent, Belgium

(Received 19 March 1986; accepted for publication 8 October 1986)

Using the method of extended Cartan matrices and extended Dynkin diagrams, a classification of maximal regular semisimple subalgebras of the basic classical Lie superalgebras is obtained. Especially in the case of exceptional Lie superalgebras, some curious inclusion relations are discovered.

I. INTRODUCTION

Since the work of Corwin, Ne'eman, and Sternberg,[1] Lie superalgebras have become increasingly important in theoretical physics.[2] Simple Lie superalgebras were classified completely,[3,4] and it was shown that basic classical Lie superalgebras can be described by a Cartan matrix or, equivalently, by a Kac–Dynkin diagram.

In this paper we make a first step in the classification of subalgebras of basic classical Lie superalgebras. Besides having a mathematical interest, the investigation of subalgebras of Lie superalgebras is also important for physicists. Indeed, in physical models where Lie algebras or Lie superalgebras are used, one is very often concerned with a chain of subalgebras.

The first class of subalgebras to be considered are the so-called regular subalgebras. These are generated by some root vectors of the original Lie superalgebra L. The aim of this paper is to obtain a classification of all maximal regular semisimple subalgebras of basic classical Lie superalgebras. The method we use is similar to the one introduced by Dynkin[5] in his classification of regular subalgebras of simple Lie algebras. The Lie superalgebras we investigate here are $A(m,n)$ $(m\neq n)$, $\mathrm{spl}(m,m)$, $B(m,n)$, $C(n)$, $D(m,n)$, $D(2,1;\alpha)$, $G(3)$, and $F(4)$, where we have used the notation of Kac.[3] Note the difference in notation between a Lie algebra and a Lie superalgebra: $C(n)$ is the Lie superalgebra $\mathrm{osp}(2,2n-2)$, whereas C_n is the Lie algebra $\mathrm{sp}(2n)$. All Lie algebras are denoted by means of an index.

In his classification, Dynkin used the method of extended simple root systems and extended Dynkin diagrams. In Sec. II we show that a similar method can be used here. There is, however, one main difference: for simple Lie algebras, all simple root systems are W equivalent (W = Weyl group). For Lie superalgebras this is not the case: in fact there exists a so-called distinguished choice of the simple root system such that only one simple root is odd.[3,6] Besides the distinguished choice, there are other choices possible, which give rise to a different Cartan matrix and Dynkin diagram. In our analysis, all possible simple root systems have to be taken into account. But we shall give details only for the distinguished choice, and simply state the results for all other choices.

In Secs. III–VIII we analyze systematically the semisimple regular subalgebras of the basic classical Lie superalgebras. The main results are summarized in Table I, and are particularly remarkable in the case of inclusions among exceptional Lie superalgebras. The construction of the extended Cartan matrices and Dynkin diagrams for the basic Lie superalgebras may also be important from another point of view. Indeed, it is well known that the extended Dynkin diagrams and the extended Cartan matrices of the simple Lie algebras are precisely the Dynkin diagrams and Cartan matrices of the so-called nontwisted affine Kac–Moody algebras.[7] It seems natural to ask whether there correspond any infinite-dimensional "affine" Lie superalgebras with the Cartan matrices listed in this paper. This question, however, falls beyond the scope of the present work.

Let us finally mention that the Lie superalgebra bracket is always denoted by $[x,y]$ throughout the paper. Whether $[x,y]$ must be interpreted as a commutator or as an anticommutator depends on the degree of x and y, and is always clear from the context.

II. GENERAL METHOD TO CONSTRUCT REGULAR SUBALGEBRAS

Throughout this paper, a simple Lie superalgebra L is always a basic classical Lie superalgebra of type $A(m,n)$ $(m\neq n)$, $B(m,n)$, $C(n)$, $D(m,n)$, $D(2,1;\alpha)$, $G(3)$, or $F(4)$. These Lie superalgebras and their properties have been described by Kac.[3] We say that a Lie superalgebra is *semisimple* if it is the direct sum of components which are either of the previously mentioned simple type, or simple Lie algebras, or else of type $\mathrm{spl}(n+1,n+1)$. Then the semisimple Lie superalgebras coincide[3] with the finite-dimensional contragredient Lie superalgebras $G(A,\tau)$, described by a Cartan matrix A and a subset τ of the index set $I=\{1,\ldots,r\}$ (for the notation, see Ref. 3). The matrix A satisfies

$$a_{ij}=0\Leftrightarrow a_{ji}=0. \tag{2.1}$$

Moreover, if

$$\forall i,j\in I, \text{ there exists a sequence } i_1,\ldots,i_t\in I \text{ for which } a_{ii_1}a_{i_1i_2}\cdots a_{i_tj}\neq 0, \tag{2.2}$$

then the contragredient Lie superalgebra is either simple or else of type $\mathrm{spl}(n+1,n+1)$. If (2.2) is not satisfied, then A splits into the direct sum of matrices of type (2.2), and hence $G(A,\tau)$ is semisimple (according to our definition of semisimplicity).

Definition 1: A subalgebra L' of a simple Lie superalgebra L is regular if, for a proper choice of the Cartan subalgebra H of L, the relation

[a)] Senior Research Assistant N.F.W.O. (Belgium).

 0022-2488/87/020292-10$02.50 © 1987 American Institute of Physics

$$\left(h+\sum_{\alpha\in\Delta} e_\alpha\right)\in L',\quad h\in H,\quad e_\alpha\in L^\alpha,$$

where Δ is the set of roots of L relative to H and L^α is the root space, implies that

$$h\in L' \quad \text{and} \quad e_\alpha\in L'.$$

An equivalent defintion is given by the following.

Definition 1': A subalgebra L' of L is regular if there exists a basis of L' consisting of elements of some Cartan subalgebra H of L, and of root vectors of L relative to H.

Let L be a simple Lie superalgebra and L' a regular subalgebra. Then L' can be written in the following form:

$$L'=H'\oplus\sum_{\alpha\in\Delta'} L^\alpha, \tag{2.3}$$

with $H'\subset H$, $\Delta'\subset\Delta$. If $\alpha,\beta\in\Delta'$, and e_α, e_β are corresponding root vectors, then $[e_\alpha,e_\beta]\in L'$. Hence if $\alpha+\beta\in\Delta$, it also belongs to Δ'. If α and $-\alpha$ belong to Δ', then $e_{\pm\alpha}\in L'$, and

$$[e_\alpha,e_{-\alpha}]=(e_\alpha,e_{-\alpha})h_\alpha. \tag{2.4}$$

Herein, $(\ ,\ )$ is the unique (up to a constant factor) nondegenerate invariant supersymmetric bilinear form (see Proposition 2.5.5 of Kac[3]), and h_α is defined by

$$(h_\alpha,h)=\alpha(h),\quad \forall h\in H. \tag{2.5}$$

In (2.4), $h_\alpha\in L'\cap H$, hence $h_\alpha\in H'$. Therefore, if L' is a regular subalgebra written in the form (2.3), then the following conditions are valid:

$$\begin{aligned}&\text{(a) } \alpha,\beta\in\Delta' \quad\text{and}\quad \alpha+\beta\in\Delta\Rightarrow\alpha+\beta\in\Delta';\\ &\text{(b) } \alpha,-\alpha\in\Delta'\Rightarrow h_\alpha\in H'.\end{aligned} \tag{2.6}$$

Conversely, suppose that L' is a subspace of L of the form (2.3) such that (2.6) are satisfied. Then $[h',h'']=0\in L'$, for all $h',h''\in H'$. Further, $[h,e_\alpha]=\alpha(h)e_\alpha\in L'$ for $h\in H'$, $\alpha\in\Delta'$. Finally, for $\alpha,\beta\in\Delta'$, $[e_\alpha,e_\beta]=0$ if $\alpha+\beta\notin\Delta$, $[e_\alpha,e_\beta]=e_{\alpha+\beta}$ if $\alpha+\beta\in\Delta$, but then also $\alpha+\beta\in\Delta'$, and $[e_\alpha,e_\beta]\sim h_\alpha$ if $\beta=-\alpha$. Hence L' is a subalgebra, and obviously it is regular.

A Cartan subalgebra of L is a Cartan subalgebra of its even part, the Lie algebra $L_{\bar 0}$. All Cartan subalgebras of $L_{\bar 0}$ are conjugate under inner automorphisms of $L_{\bar 0}$. Hence, in order to find all regular subalgebras (up to conjugacy) of L, we can use the following procedure: (1) consider a particular Cartan subalgebra H of L and the corresponding root system Δ; (2) consider all possible subspaces of the form (2.3); and (3) determine which subspaces also satisfy (2.6). From now on, we may suppose that H (and consequently also Δ) is fixed.

Definition 2: Let Δ be the root system of L. A subset Γ of Δ is called a regular subsystem if

$$\begin{aligned}&\text{(a) } \alpha,\beta\in\Gamma\Rightarrow\alpha-\beta\notin\Delta;\\ &\text{(b) } \Gamma \text{ is linearly independent.}\end{aligned} \tag{2.7}$$

Proposition: Let H be the Cartan subalgebra and Δ the root system of L. Let $\Gamma=\{\alpha_1,\ldots,\alpha_m\}$ be a regular subsystem of Δ, and let L' be the subalgebra of L generated by the root vectors $e_{\alpha_1},\ldots,e_{\alpha_m},e_{-\alpha_1},\ldots,e_{-\alpha_m}$. Then L' is a regular semisimple subalgebra of L and Γ is a system of simple roots for L'. Conversely, every regular semisimple subalgebra L' of L (up to conjugacy) can be obtained by such a construction.

Proof: The subalgebra L' is regular since (2.6) is fulfilled. Put

$$\begin{aligned}&e_i=e_{\alpha_i},\quad f_i=e_{-\alpha_i},\\ &h_i=(e_{\alpha_i},e_{-\alpha_i})h_{\alpha_i}.\end{aligned} \tag{2.8}$$

The elements h_i are linearly independent, since Γ is a regular subsystem. Then L' is the contragredient Lie superalgebra $G(A,\tau)$ with A determined by

$$[h_i,e_j]=a_{ij}e_j, \tag{2.9}$$

and τ consisting of those i for which α_i is an odd root. From (2.8) it follows that A satisfies (2.1). Hence, L' is semisimple, and obviously Γ is a set of simple roots of L'. Conversely, let L' be a semisimple regular subalgebra of L. We may suppose that the Cartan subalgebra of L' is contained in H:

$$L'=H'+\sum_{\alpha\in\Delta'} L^\alpha, \tag{2.10}$$

and (2.6) is satisfied. Since L' is semisimple, we have

$$\alpha\in\Delta'\Leftrightarrow-\alpha\in\Delta'. \tag{2.11}$$

Let Γ be a set of simple roots for L'. Then Γ is linearly independent and $\alpha,\beta\in\Gamma$ implies $\alpha-\beta\notin\Delta'$. But if $\alpha,\beta\in\Gamma$, then (2.11) gives $\alpha,-\beta\in\Delta'$. Now (2.6a) implies

$$\alpha,-\beta\in\Delta',\quad \alpha+(-\beta)\notin\Delta'\Rightarrow\alpha+(-\beta)\in\Delta. \tag{2.12}$$

Hence (2.7a) is valid and Γ is a regular subsystem. Obviously, L' is generated by the root vectors of the simple roots and their negatives.

The problem of finding semisimple regular subalgebras of L is now reduced to the following: determine all the regular subsystems Γ of the root system Δ of L. When L is a Lie superalgebra of rank n, every regular subsystem will always be contained in a regular subsystem of order n. Hence, we only have to classify the "maximal" regular subsystems.

Definition 3: A regular semisimple subalgebra L' ($\neq L$) of L is called maximal if there does not exist any other regular semisimple algebra L'' such that $L'\subset L''\subset L$ (all inclusions are strict).

Let Π be a simple root system of L: obviously, Π is a regular subsystem. Extend Π by a root γ of Δ to $\Pi^e=\Pi\cup\{\gamma\}$ such that (2.7a) is still satisfied. We call Π^e the extended simple root system. It follows from the following sections that there is in general only one way to extend Π (only in some situations Π can be extended by either γ or else by 2γ, if both γ and 2γ belong to Δ). Then, one obtains a maximal regular subsystem $\Gamma^{(1)}$ by deleting one root from Π^e. Now, $\Gamma^{(1)}$ can be extended by a root γ', and again deleting a root gives another regular maximal subsystem $\Gamma^{(2)}$, and so on. All possible maximal regular subsystems are obtained in this way. However, the regular subalgebra determined by $\Gamma^{(2)}$ is contained in the one determined by $\Gamma^{(1)}$. Hence, in order to find the maximal semisimple regular subalgebras, only the first extension has to be considered, unless $\Gamma^{(1)}$ gives rise to a subalgebra isomorphic to L.

The method to find all maximal semisimple regular subalgebras of L is now clear. The only difference with Lie algebras is the following: for Lie algebras all simple root systems are W equivalent (W = Weyl group), whereas for Lie superalgebras there are in general several nonequivalent

simple root systems. In Secs. III–VIII we shall construct the extended simple root systems, the corresponding extended Cartan matrices, and the extended Dynkin diagrams for the distinguished choice[3,6] of Π for all basic classical Lie superalgebras L. Moreover, we shall state the results for all other simple root systems, and hence obtain all maximal semisimple regular subalgebras of L.

III. THE LIE SUPERALGEBRAS $A(m,n)$ $(m\neq n)$ OR $\mathrm{spl}(n+1,n+1)$

The special linear Lie superalgebras are defined by

$$\mathrm{spl}(m+1,n+1) = \left\{ x = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| \mathrm{str}(x) = \mathrm{tr}(a) - \mathrm{tr}(d) = 0 \right\}, \quad (3.1)$$

where a, b, c, and d are $(m+1)\times(m+1)$, $(m+1)\times(n+1)$, $(n+1)\times(m+1)$, and $(n+1)\times(n+1)$ matrices, respectively. The even elements are of the form $\left[\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right]$, and the odd elements of the form $\left[\begin{smallmatrix} 0 & b \\ c & 0 \end{smallmatrix}\right]$. The Cartan subalgebra H is a subspace of diagonal matrices of (3.1). For a diagonal matrix $D = \mathrm{diag}(d_{11},d_{22},\ldots,d_{m+n+2,m+n+2})$, we define

$$\begin{aligned} &\epsilon_i(D) = d_{ii} \quad (i = 1,\ldots,m+1), \\ &\delta_i(D) = d_{m+i+1,m+i+1} \quad (i = 1,\ldots,n+1). \end{aligned} \quad (3.2)$$

Then the even roots of $\mathrm{spl}(m+1,n+1)$ are

$$\begin{aligned} \Delta_0 = \{ &\epsilon_i - \epsilon_j \ (i,j = 1,\ldots,m+1);\ \delta_i - \delta_j \\ &(i,j = 1,\ldots,n+1)\}, \end{aligned} \quad (3.3)$$

and the odd roots are given by

$$\Delta_1 = \{ \pm(\epsilon_i - \delta_j); i = 1,\ldots,m+1,\ j = 1,\ldots,n+1\}. \quad (3.4)$$

The distinguished positive simple root system Π is determined by[3]

$$\begin{aligned} \Pi = \{ &\epsilon_1 - \epsilon_2, \epsilon_2 - \epsilon_3, \ldots, \epsilon_m - \epsilon_{m+1}, \epsilon_{m+1} \\ &- \delta_1, \delta_1 - \delta_2, \ldots, \delta_n - \delta_{n+1}\}. \end{aligned} \quad (3.5)$$

The corresponding root vectors are

$$e_i = E_{i,i+1}, \quad f_i = E_{i+1,i} \quad (i = 1,\ldots,m+n+1). \quad (3.6)$$

Then, one can check that $[e_i,f_j] = \delta_{ij}h_i$, with

$$\begin{aligned} &h_i = E_{ii} - E_{i+1,i+1} \quad (i = 1,\ldots,m,m+2,\ldots,m+n+1), \\ &h_{m+1} = E_{m+1,m+1} + E_{m+2,m+2}. \end{aligned} \quad (3.7)$$

In (3.6) and (3.7), E_{ij} stands for the matrix of $\mathrm{spl}(m+1,n+1)$ with zeros everywhere, except a 1 in the entry (i,j). From (3.6) and (3.7) the Cartan matrix and Dynkin diagram of $\mathrm{spl}(m+1,n+1)$ can be determined.

Now, we shall extend Π. It is easy to see from (3.3)–(3.5) that the only root γ by which Π is extended such that (2.7a) is still satisfied, is

$$\gamma = \delta_{n+1} - \epsilon_1. \quad (3.8)$$

We call γ the zero root of Π^e. In order to find the extended Cartan matrix, we have to construct the root vectors corresponding to γ and $-\gamma$, and their (anti-) commutation relations with the $\mathrm{spl}(m+1,n+1)$ generators. One finds

$$\begin{aligned} &e_0 = E_{m+n+2,1}, \quad f_0 = E_{1,m+n+2}, \\ &[e_0,f_0] = h_0 = h_1 + h_2 + \cdots + h_{m+1} \\ &\qquad - h_{m+2} - \cdots - h_{m+n+1}, \end{aligned} \quad (3.9)$$

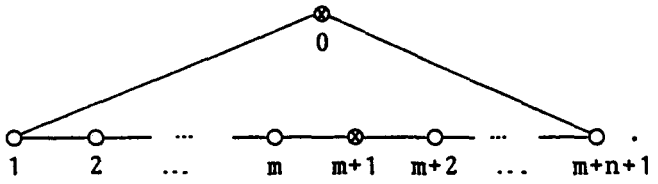
and

$$\begin{aligned} &[h_0,e_0] = 0, \\ &[h_i,e_0] = -\delta_{i,1}e_0 - \delta_{i,m+n+1}e_0, \\ &[h_0,e_i] = \delta_{i,1}e_1 - \delta_{i,m+n+1}e_{m+n+1} \\ &\qquad (i = 1,\ldots,m+n+1). \end{aligned} \quad (3.10)$$

In fact, (3.10) determines the "zero row and column" of the extended Cartan matrix A. We obtain

$$\begin{array}{c|ccccccccccc} & 0 & 1 & 2 & \cdots & m+1 & & \cdots & & m+n+1 \\ \hline 0 & 0 & 1 & 0 & \cdots & & & & 0 & -1 \\ -1 & 2 & -1 & \cdots & & & & & & \\ 0 & -1 & 2 & -1 & & & & & & \\ & & & \ddots & & & & & & \\ & & & & -1 & & & & & \\ & & & & -1 & 2 & -1 & & & \\ & & & & & -1 & 0 & 1 & & \\ & & & & & & -1 & 2 & -1 & \\ & & & & & & & -1 & 2 & \\ & & & & & & & & \ddots & \\ 0 & & & & & & & & 2 & -1 \\ -1 & & & & & & & & -1 & 2 \end{array}, \quad \tau = \{0,m+1\}. \quad (3.11)$$

Hence, the extended Dynkin diagram is



(3.12)

When a single node is deleted from (3.12) (which corresponds to deleting a root from Π^e), the corresponding diagram is again a Dynkin diagram for $\mathrm{spl}(m+1,n+1)$. When two adjacent nodes are deleted, the remaining Dynkin diagram is that for $\mathrm{spl}(m,n+1)$ or $\mathrm{spl}(m+1,n)$. When two nonadjacent nodes are deleted, there remains the Dynkin diagram of $\mathrm{spl}(k,l)\oplus\mathrm{spl}(i,j)$ with $k+i=m+1$ and $l+j=n+1$. Note that i or j can be zero: $\mathrm{spl}(r,0)=\mathrm{spl}(0,r)=\mathrm{sl}(r)=A_{r-1}$.

The most general system of simple roots of $\mathrm{spl}(m+1,n+1)$ is determined[3] by two increasing sequences $S=\{1\leqslant s_1<s_2<\cdots\}$ and $T=\{1\leqslant t_1<t_2<\cdots\}$ and a sign,

$$\Pi_{S,T}=\pm\{\epsilon_1-\epsilon_2,\epsilon_2-\epsilon_3,\ldots,\epsilon_{s_1}-\delta_1,\delta_1-\delta_2,\ldots,\delta_{t_1}-\epsilon_{s_1+1},\ldots\}. \tag{3.13}$$

Let $\pm(\eta_i-\eta'_j)$ be the last element of $\Pi_{S,T}$ (η and η' can be ϵ or δ). Then $\gamma=\pm(\eta'_j-\epsilon_1)$, and it is again uniquely determined. Deleting one node from $\Pi^e_{S,T}$ gives rise to a Dynkin diagram for $\mathrm{spl}(m+1,n+1)$, and we find similar results as for the distinguished choice of Π.

The conclusion is the following: for $L=\mathrm{spl}(m+1,n+1)$, the only semisimple regular subalgebra of rank $m+n+1$ is the algebra L itself. Hence, the maximal regular semisimple subalgebras are of rank $m+n$ and are of the form $\mathrm{spl}(k,l)\oplus\mathrm{spl}(i,j)$ with $k+i=m+1$ and $l+j=n+1$.

IV. THE ORTHOSYMPLECTIC SUPERALGEBRAS $B(m,n)$

$B(m,n)=\mathrm{osp}(2m+1,2n)$ is the subalgebra of $\mathrm{spl}(2m+1,2n)$ consisting of those $x=\left[\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right]$ for which the even and odd components x_ξ ($\xi=\bar{0},\bar{1}$) satisfy

$$x_\xi^T B+(-1)^\xi B x_\xi=0, \tag{4.1}$$

where $x^T=\left[\begin{smallmatrix}a^t&-c^t\\b^t&d^t\end{smallmatrix}\right]$ is the supertranspose of x, and

$$B=\left[\begin{array}{ccc:ccc}0&1_m&0&&&\\1_m&0&0&&&\\0&0&1&&&\\\hdashline&&&0&1_n&\\&&&-1_n&0&\end{array}\right]. \tag{4.2}$$

Consider first the situation with $m>0$. The Cartan subalgebra is spanned by diagonal matrices D, and we put

$$\epsilon_i(D)=d_{ii}\quad(i=1,\ldots,m),$$
$$\delta_i(D)=d_{2m+i+1,2m+i+1}\quad(i=1,\ldots,n). \tag{4.3}$$

One can check that the even roots are given by

$$\Delta_0=\{\pm(\epsilon_j-\epsilon_k),\pm(\epsilon_j+\epsilon_k),\pm\epsilon_j\ (1\leqslant j<k\leqslant m);\pm(\delta_j-\delta_k),\pm(\delta_j+\delta_k),\pm2\delta_j\ (1\leqslant j<k\leqslant n)\}, \tag{4.4}$$

and the odd roots by

$$\Delta_1=\{\pm(\epsilon_j-\delta_k),\pm(\epsilon_j+\delta_k),\pm\delta_k,\ (1\leqslant j\leqslant m,\ 1\leqslant k\leqslant n)\}. \tag{4.5}$$

The distinguished set of simple roots is equal to[3]

$$\Pi=\{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_n-\epsilon_1,\epsilon_1-\epsilon_2,\ldots,\epsilon_{m-1}-\epsilon_m,\epsilon_m\}. \tag{4.6}$$

The corresponding root vectors, in the same notation as in Sec. III, are

$$\begin{aligned}
e_i&=E_{2m+i+1,2m+i+2}-E_{2m+n+i+2,2m+n+i+1},\\
f_i&=E_{2m+i+2,2m+i+1}-E_{2m+n+i+1,2m+n+i+2}\\
&\quad(i=1,\ldots,n-1);\\
e_n&=E_{m+1,2m+2n+1}+E_{2m+n+1,1},\\
f_n&=E_{1,2m+n+1}-E_{2m+2n+1,m+1};\\
e_{n+i}&=E_{i,i+1}-E_{m+i+1,m+i},\\
f_{n+i}&=E_{i+1,i}-E_{m+i,m+i+1}\quad(i=1,\ldots,m-1);\\
e_{n+m}&=\sqrt{2}(E_{m,2m+1}-E_{2m+1,2m}),\\
f_{n+m}&=\sqrt{2}(E_{2m+1,m}-E_{2m,2m+1}).
\end{aligned} \tag{4.7}$$

Then, with $[e_i f_j]=\delta_{ij}h_j$, one obtains

$$\begin{aligned}
h_i&=E_{2m+i+1,2m+i+1}-E_{2m+i+2,2m+i+2}\\
&\quad-E_{2m+n+i+1,2m+n+i+1}\\
&\quad+E_{2m+n+i+2,2m+n+i+2}\quad(i=1,\ldots,n-1),\\
h_n&=E_{11}-E_{m+1,m+1}+E_{2m+n+1,2m+n+1}\\
&\quad-E_{2m+2n+1,2m+2n+1},\\
h_{n+i}&=E_{ii}-E_{i+1,i+1}-E_{m+i,m+i}\\
&\quad+E_{m+i+1,m+i+1}\quad(i=1,\ldots,m-1),\\
h_{n+m}&=2(E_{m,m}-E_{2m,2m}).
\end{aligned} \tag{4.8}$$

These elements determine the Cartan matrix and hence also the Dynkin diagram of $B(m,n)$ completely.

Let us now try to extend Π by a root γ such that (2.7a) is still valid. It turns out that there are two solutions: either $\gamma=-2\delta_1$, or else $\gamma=-\delta_1$. Consider first the situation with $\gamma=-2\delta_1$. Then

$$e_0=E_{2m+n+2,2m+2},\quad f_0=E_{2m+2,2m+n+2}, \tag{4.9}$$

and

$$\begin{aligned}
h_0&=E_{2m+n+2,2m+n+2}-E_{2m+2,2m+2}\\
&=-h_1-h_2-\cdots-h_n+h_{n+1}+h_{n+2}\\
&\quad+\cdots+h_{n+m-1}+\tfrac{1}{2}h_{n+m}.
\end{aligned} \tag{4.10}$$

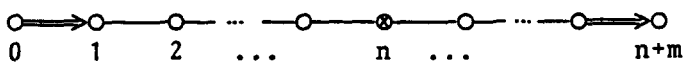
Here, one finds

$$\begin{aligned}
&[h_0,e_0]=2e_0,\\
&[h_0,e_i]=-\delta_{i,1}e_i,\\
&[h_i,e_0]=-2\delta_{i,1}e_0\quad(i=1,..,n+m).
\end{aligned} \tag{4.11}$$

Therefore, the extended Cartan matrix is

$$
\begin{array}{c|cccccccccc}
 & 0 & 1 & & & & n & & & & n+m \\
\hline
 & 2 & -1 & 0 & \cdots & & & & & & 0 \\
 & -2 & 2 & -1 & & & & & & & \\
 & 0 & -1 & 2 & & & & & & & \\
 & & & & \ddots & & & & & & \\
 & \vdots & & & & 2 & -1 & & & & \\
 & & & & & -1 & 2 & -1 & & & \\
\hline
 & & & & & & -1 & 0 & 1 & & \\
\hline
 & & & & & & & -1 & 2 & -1 & \\
 & & & & & & & & -1 & 2 & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & 2 & -1 \\
 & 0 & & & & & & & & -2 & 2
\end{array}
\quad , \quad \tau = \{n\}. \tag{4.12}
$$

This corresponds to the following Dynkin diagram:



(4.13)

The subalgebras obtained by deleting the 0th, 1st, 2nd,..., $(n+m)$th node are $B(m,n)$, $A_1 \oplus B(m,n-1)$, $C_2 \oplus B(m,n-2)$,..., $C_{n-1} \oplus B(m,1)$, $C_n \oplus B_m$, $C(n+1) \oplus B_{m-1}$, $D(2,n) \oplus B_{m-2}$,..., $D(m-1,n) \oplus A_1$, $D(m,n)$. Note that these subalgebras are all of the form $\mathrm{osp}(0,2k) \oplus \mathrm{osp}(2m+1,2n-2k)$ or $\mathrm{osp}(2j,2n) \oplus \mathrm{osp}(2m+1-2j,0)$, where $\mathrm{osp}(2r+1,0) = \mathrm{so}(2r+1) = B_r$ and $\mathrm{osp}(0,2r) = \mathrm{sp}(2r) = C_r$.

Next, consider the situation where $\gamma = -\delta_1$. Then

$$
\begin{aligned}
e_0 &= \sqrt{2}(E_{2m+1,2m+2} - E_{2m+n+2,2m+1}), \\
f_0 &= -\sqrt{2}(E_{2m+2,2m+1} + E_{2m+1,2m+n+2}), \\
h_0 &= -2(h_1 + h_2 + \cdots + h_n - h_{n+1} - \cdots \\
&\quad - h_{n+m-1} - \tfrac{1}{2}h_{n+m}),
\end{aligned} \tag{4.14}
$$

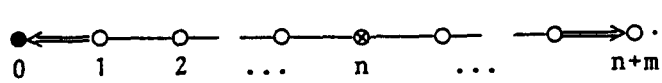
and the relations (4.11) become

$$
\begin{aligned}
&[h_0,e_0] = 2e_0, \\
&[h_0,e_i] = -2\delta_{i,1}e_i, \\
&[h_i,e_0] = -\delta_{i,1}e_i \quad (i = 1,\ldots,m+n).
\end{aligned} \tag{4.15}
$$

The new extended Cartan matrix has obviously the same $(n+m)\times(n+m)$ part as in (4.12), and hence it is determined by

$$
\begin{array}{c|cccc}
 & 0 & 1 & & n+m \\
\hline
 & 2 & -2 & 0 \cdots & 0 \\
\hline
 & -1 & & & \\
 & 0 & & \text{same as} & \\
 & \vdots & & (4.12) & \\
 & 0 & & &
\end{array}
\quad , \quad \tau = \{0,n\}. \tag{4.16}
$$

The corresponding Dynkin diagram is



(4.17)

Now, the subalgebras obtained by deleting the 0th, 1st, 2nd,..., $(n+m)$th node are $B(m,n)$, $B(0,1) \oplus B(m,n-1)$, $B(0,2) \oplus B(m,n-2)$,..., $B(0,n) \oplus B_m$, $B(1,n) \oplus B_{m-1}$,..., $B(m-1,n) \oplus A_1$, $B(m,n)$. Note that every subalgebra from the previous series with $\gamma = -2\delta_1$ is contained in one of the algebras of the series with $\gamma = -\delta_1$. This is not remarkable, since $[e_{-\delta_1}, e_{-\delta_1}] = -4e_{-2\delta_1}$.

Another system of simple roots is of the form $\Pi_{S,T}$ (see Kac[3]), and detailed investigations show that the most general maximal semisimple regular subalgebras of $B(m,n)$ are $B(k,l) \oplus B(i,j)$, with $k+i=m$, $l+j=n$, and $D(m,n)$.

The Lie superalgebras $B(0,n) = \mathrm{osp}(1,2n)$ form a special case of the algebras $B(m,n)$. We treat them separately because the distinguished choice of the simple roots for $B(0,n)$ is different than the analog of (4.6). The roots of $B(0,n)$ are, in the same notation as in (4.4),

$$
\Delta_0 = \{\pm(\delta_j - \delta_k), \pm(\delta_j + \delta_k), \pm 2\delta_j \quad (1 \leqslant j < k \leqslant n)\}, \tag{4.18}
$$

$$
\Delta_1 = \{\pm\delta_j \quad (1 \leqslant j \leqslant n)\}. \tag{4.19}
$$

The distinguished set of simple roots is given by

$$
\Pi = \{\delta_1 - \delta_2, \delta_2 - \delta_3, \ldots, \delta_{n-1} - \delta_n, \delta_n\}. \tag{4.20}
$$

The $B(0,n)$ generators are

$$
\begin{aligned}
e_i &= E_{i+1,i+2} - E_{n+i+2,n+i+1}, \\
f_i &= E_{i+2,i+1} - E_{n+i+1,n+i+2} \quad (i = 1,\ldots,n-1), \\
e_n &= \sqrt{2}(E_{1,2n+1} + E_{n+1,1}), \\
f_n &= \sqrt{2}(E_{1,n+1} - E_{2n+1,1}),
\end{aligned} \tag{4.21}
$$

and

$$
\begin{aligned}
h_i &= E_{i+1,i+1} - E_{i+2,i+2} - E_{n+i+1,n+i+1} \\
&\quad + E_{n+i+2,n+i+2} \quad (i = 1,\ldots,n-1), \\
h_n &= 2(E_{n+1,n+1} - E_{2n+1,2n+1}).
\end{aligned} \tag{4.22}
$$

From (4.21) and (4.22), the usual Cartan matrix and Dynkin diagram of $B(0,n)$ are obtained.

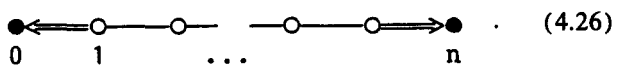
Just as in (4.9) and (4.14), Π can be extended in exactly

two different ways, either by $\gamma = -2\delta_1$ or else by $\gamma = -\delta_1$. Because the extension with $-2\delta_1$ gives rise to a series of subalgebras which are all contained in subalgebras of the series with $-\delta_1$, we consider only the second extension. Hence, $\gamma = -\delta_1$, and

$$e_0 = \sqrt{2}(E_{1,2} - E_{n+2,1}),$$
$$f_0 = -\sqrt{2}(E_{1,n+2} + E_{2,1}), \tag{4.23}$$

$$\begin{array}{cccccccc}
0 & 1 & & \cdots & & & & n \\
2 & -2 & 0 & & & & & 0 \\
-1 & 2 & -1 & & & & & \\
 & -1 & 2 & -1 & & & & \\
 & & & \ddots & & & & \\
 & & & & 2 & -1 & & \\
 & & & & -1 & 2 & -1 & \\
 & & & & & -2 & 2 &
\end{array}$$

and the corresponding Dynkin diagram is given by



(4.26)

Hence, the regular subalgebras obtained from (4.26) by deleting one node are of the form $B(0,k) \oplus B(0,n-k)$. The regular subalgebras derived from $\Pi \cup \{-2\delta_1\}$ are of the form $C_k \oplus B(0,n-k)$, and are all contained in the former series. However, for $k = n$ we find the inclusion $C_n \subset B(0,n)$, and this is also maximal since the series with $\gamma = -\delta_1$ only gives the algebra $B(0,n)$ itself as "greater than" C_n. Here, any other choice of simple roots for $B(0,n)$ is W equivalent[3] to Π, and consequently the list of maximal regular subalgebras is exhausted.

V. THE ORTHOSYMPLECTIC SUPERALGEBRAS $D(m,n)$ $(m \neq 1)$ AND $C(n)$

The Lie superalgebra $D(m,n) = \mathrm{osp}(2m,2n)$ is the subalgebra of $\mathrm{spl}(2m,2n)$ consisting of those x for which

$$x_\xi^T B + (-1)^\xi B x_\xi = 0, \tag{5.1}$$

with

$$B = \begin{bmatrix} 0 & 1_m & & \\ 1_m & 0 & & \\ & & 0 & 1_m \\ & & -1_m & 0 \end{bmatrix}. \tag{5.2}$$

As usual, we define the forms ϵ_i and δ_i by

$$\epsilon_i(D) = d_{ii} \quad (i = 1,...,m),$$
$$\delta_i(D) = d_{2m+i,2m+i} \quad (i = 1,...,n). \tag{5.3}$$

The roots of $D(m,n)$ are given by

$$\Delta_0 = \{\pm(\epsilon_j - \epsilon_k), \pm(\epsilon_j + \epsilon_k) \ (1 \leqslant j < k \leqslant m);$$
$$\pm(\delta_j - \delta_k), \pm(\delta_j + \delta_k), \pm 2\delta_j \ (1 \leqslant j < k \leqslant n)\}, \tag{5.4}$$
$$\Delta_1 = \{\pm(\delta_j - \epsilon_k), \pm(\delta_j + \epsilon_k), \ (1 \leqslant j \leqslant n, \ 1 \leqslant k \leqslant m)\}.$$

$$h_0 = -2(h_1 + h_2 + \cdots + h_{n-1} + \tfrac{1}{2}h_n).$$

The relevant (anti-) commutation relations are

$$[h_0,e_0] = 2e_0,$$
$$[h_0,e_i] = -2\delta_{i,1}e_i, \tag{4.24}$$
$$[h_i,e_0] = -\delta_{i,1}e_0 \quad (i = 1,...,n).$$

Hence, the extended Cartan matrix is

$$\tau = \{0,n\}, \tag{4.25}$$

The distinguished choice for the simple root system Π is given by[3,6]

$$\Pi = \{\delta_1 - \delta_2, \delta_2 - \delta_3, ..., \delta_n - \epsilon_1, \epsilon_1 - \epsilon_2, ..., \epsilon_{m-2} - \epsilon_{m-1}, \epsilon_{m-1} - \epsilon_m, \epsilon_{m-1} + \epsilon_m\}. \tag{5.5}$$

The corresponding root vectors are determined by

$$e_i = E_{2m+i,2m+i+1} - E_{2m+n+i+1,2m+n+i},$$
$$f_i = E_{2m+i+1,2m+i} - E_{2m+n+i,2m+n+i+1}$$
$$(i = 1,...,n-1);$$
$$e_n = E_{m+1,2m+2n} + E_{2m+n,1},$$
$$f_n = E_{1,2m+n} - E_{2m+2n,m+1}; \tag{5.6}$$
$$e_{n+i} = E_{i,i+1} - E_{m+i+1,m+i},$$
$$f_{n+i} = E_{i+1,i} - E_{m+i,m+i+1} \quad (i = 1,..,m-1);$$
$$e_{n+m} = E_{m-1,2m} - E_{m,2m-1},$$
$$f_{n+m} = E_{2m,m-1} - E_{2m-1,m}.$$

Then $[e_i f_j] = \delta_{ij}h_j$, and the h_j span the Cartan subalgebra H,

$$h_i = E_{2m+i,2m+i} - E_{2m+i+1,2m+i+1} - E_{2m+n+i,2m+n+i} + E_{2m+n+i+1,2m+n+i+1} \quad (i = 1,...,n-1),$$
$$h_n = E_{11} - E_{m+1,m+1} + E_{2m+n,2m+n} - E_{2m+2n,2m+2n},$$
$$h_{n+i} = E_{ii} - E_{i+1,i+1} - E_{m+i,m+i} + E_{m+i+1,m+i+1} \quad (i = 1,...,m-1), \tag{5.7}$$
$$h_{n+m} = E_{m-1,m-1} + E_{m,m} - E_{2m-1,2m-1} - E_{2m,2m}.$$

The Cartan matrix for $D(m,n)$ is obtained from (5.6) and (5.7) and the relations $[h_i,e_j] = a_{ij}e_j$.

Let us now investigate the extension of Π. We find that there is only one solution for $\Pi^e = \Pi \cup \{\gamma\}$, namely $\gamma = -2\delta_1$. Hence,

$$e_0 = E_{2m+n+1,2m+1}, \quad f_0 = E_{2m+1,2m+n+1}, \tag{5.8}$$

and

$$h_0 = [e_0,f_0]$$

$$= -h_1 - h_2 - \cdots - h_n + h_{n+1} + \cdots + h_{n+m-2} + \tfrac{1}{2}(h_{n+m-1} + h_{n+m}). \tag{5.9}$$

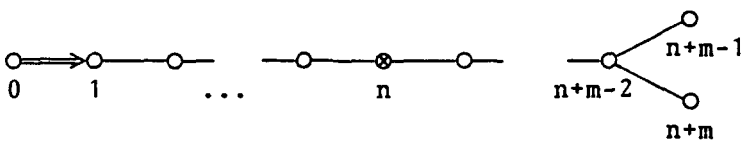
Then, the relevant relations to determine the extended Cartan matrix are

$$\begin{aligned} &[h_0,e_0] = 2e_0, \\ &[h_i,e_0] = -2\delta_{i,1}e_0, \\ &[h_0,e_i] = -\delta_{i,1}e_i \quad (i = 1,\ldots,m+n). \end{aligned} \tag{5.10}$$

This gives rise to

$$\begin{array}{c|ccccccccccc} & 0 & 1 & \cdots & & n & & \cdots & & & & n+m \\ \hline 0 & 2 & -1 & 0 & & & & & & & & 0 \\ 1 & -2 & 2 & -1 & & & & & & & & \\ & 0 & -1 & 2 & & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & 2 & -1 & & & & & & \\ & & & & -1 & 2 & -1 & & & & & \\ n & & & & & -1 & 0 & 1 & & & & \\ & & & & & & -1 & 2 & -1 & & & \\ & & & & & & & -1 & 2 & & & \\ & & & & & & & & & \ddots & & \\ & & & & & & & & & 2 & -1 & -1 \\ & & & & & & & & & -1 & 2 & 0 \\ n+m & 0 & & & & & & & & -1 & 0 & 2 \end{array} \quad , \quad \tau = \{n\}. \tag{5.11}$$

Hence, the extended Dynkin diagram is given by



(5.12)

The regular subalgebras of $D(m,n)$ obtained by deleting the 0th, 1st,..., $(n+m)$th node from (5.12) are $D(m,n)$, $A_1 \oplus D(m,n-1)$, $C_2 \oplus D(m,n-2)$,..., $C_{n-1} \oplus D(m,1)$, $C_n \oplus D_m$, $C(n+1) \oplus D_{m-1}$, $D(2,n) \oplus D_{m-2}$,..., $D(m-2,n) \oplus A_1 \oplus A_1$, $D(m,n)$, $D(m,n)$. All these subalgebras are of the form $\mathrm{osp}(0,2k) \oplus \mathrm{osp}(2m,2n-2k)$ or $\mathrm{osp}(2k,2n) \oplus \mathrm{osp}(2m-2k,0)$. There is one regular subalgebra of rank $m+n-1$ which is not contained in a proper regular subalgebra of rank $m+n$, namely $\mathrm{spl}(m,n)$. This subalgebra is found by deleting node 0 and node $n+m$.

For $D(m,n)$, however, several choices are possible for the simple root system.[3] They are determined by two increasing sequences S and T and a number: $\Pi^{(i)}_{S,T}$ $(i = 1 \text{ or } 2)$. Detailed investigations of the extensions of those nondistinguished simple root systems showed that the most general maximal semisimple regular subalgebras are of the form $D(k,l) \oplus D(i,j)$ with $k+i=m$, $l+j=n$, or are equal to $\mathrm{spl}(m,n)$. Of course, the latter notation includes all "degenerate" cases such as $D(1,r) = C(r-1)$, $D(0,r) = C_r$, $D(r,0) = D_r$, $D(2,0) = A_1 \oplus A_1$, etc.

The Lie superalgebras $C(n) = \mathrm{osp}(2,2n-2)$ form a special case of the $D(m,n)$ series. The roots of $C(n)$ are given by (there is only one ϵ form, hence we omit the index)

$$\begin{aligned} \Delta_0 &= \{ \pm(\delta_j - \delta_k), \pm(\delta_j + \delta_k), \pm 2\delta_j \\ &\qquad (1 \leqslant j < k \leqslant n-1)\}, \\ \Delta_1 &= \{ \pm(\epsilon + \delta_j), \pm(\epsilon - \delta_j) \ \ (1 \leqslant j \leqslant n-1)\}. \end{aligned} \tag{5.13}$$

The distinguished set of simple roots is

$$\Pi = \{\epsilon - \delta_1, \delta_1 - \delta_2, \delta_2 - \delta_3, \ldots, \delta_{n-2} - \delta_{n-1}, 2\delta_{n-1}\}, \tag{5.14}$$

and the corresponding root vectors are determined by

$$\begin{aligned} &e_1 = E_{13} - E_{n+2,2}, \quad f_1 = E_{2,n+2} + E_{31}; \\ &e_i = E_{i+1,i+2} - E_{n+i+1,n+i}, \\ &f_i = E_{i+2,i+1} - E_{n+i,n+i+1} \quad (i = 2,\ldots,n-1), \\ &e_n = E_{n+1,2n}, \quad f_n = E_{2n,n+1}. \end{aligned} \tag{5.15}$$

Then $[e_i,f_j] = \delta_{ij}h_j$, and the h_j span the Cartan subalgebra H,

$$\begin{aligned} &h_1 = E_{11} - E_{22} + E_{33} - E_{n+2,n+2}, \\ &h_i = E_{i+1,i+1} - E_{i+2,i+2} - E_{n+i,n+i} + E_{n+i+1,n+i+1} \\ &\qquad (i = 2,\ldots n-1), \\ &h_n = E_{n+1,n+1} - E_{2n,2n}. \end{aligned} \tag{5.16}$$

From (5.15) and (5.16), the Cartan matrix of $C(n)$ can be constructed.

For the extension of Π, it turns out that there is again only one possibility, namely, $\gamma = -\epsilon - \delta_1$. This implies

$$\begin{aligned} &e_0 = E_{23} - E_{n+2,1}, \\ &f_0 = E_{1,n+2} + E_{32}, \\ &h_0 = [e_0,f_0] = -h_1 + 2(h_2 + \cdots + h_n). \end{aligned} \tag{5.17}$$

The zero row and column of the extended Cartan matrix are determined by

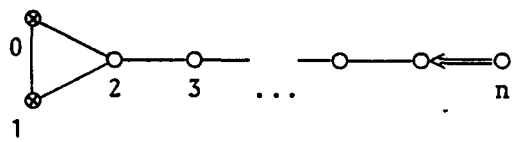
$$[h_0,e_0]=0,$$

$$[h_i,e_0]=-2\delta_{i,1}e_0-\delta_{i,2}e_0, \tag{5.18}$$
$$[h_0,e_i]=-2\delta_{i,1}e_i+\delta_{i,2}e_i \quad (i=1,\ldots,n).$$

Hence, we find

$$\begin{bmatrix} 0 & 1 & \cdots & & & & n \\ 0 & -2 & 1 & 0 & & & 0 \\ -2 & 0 & 1 & 0 & & & \\ -1 & -1 & 2 & -1 & & & \\ & & -1 & 2 & & & \\ & & & & \ddots & & \\ & & & & 2 & -1 & \\ & & & & -1 & 2 & -2 \\ 0 & & & & & -1 & 2 \end{bmatrix}, \quad \tau=\{0,1\}. \tag{5.19}$$

The Dynkin diagram corresponding to a Cartan matrix of this form is given by



(5.20)

It is easy to check that the regular subalgebras of $C(n)$ obtained from (5.20) by deleting the 0th, 1st,...,nth node are $C(n)$, $C(n)$, $A(1,0)\oplus C_{n-2}$, $C(3)\oplus C_{n-3}$,..., $C(n-1)\oplus A_1$, $C(n)$. These are all of the form $\mathrm{osp}(0,2k)\oplus\mathrm{osp}(2,2n-2k-2)$, since $\mathrm{osp}(0,2k)=C_k$ and $\mathrm{osp}(2,2)=\mathrm{spl}(2,1)=A(1,0)$. There are two other nonequivalent choices for the simple root system,[3] but the regular subalgebras obtained by means of these alternative choices coincide with the previous series.

VI. THE EXCEPTIONAL SERIES $D(2,1;\alpha)$

For the exceptional Lie superalgebra $D(2,1;\alpha)$ ($\alpha\in\mathbb{C}\setminus\{0,-1\}$), we use the realization given in Ref. 8. Here $D(2,1;\alpha)$ is determined by its Cartan matrix

$$A=\begin{bmatrix} 0 & 1 & \alpha \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix}, \quad \tau=\{1\}, \tag{6.1}$$

and the corresponding generators e_i,f_i,h_i $(i=1,2,3)$. From now on, we use the following shorthand notation:

$$e_{i_1i_2\cdots i_s}=[\cdots[[e_{i_1},e_{i_2}],e_{i_3}]\ldots,e_{i_s}] \tag{6.2}$$

and similarly for $f_{i_1\cdots i_s}$. The roots of $D(2,1;\alpha)$ are expressed in terms of linear functions[3] $\epsilon_1,\epsilon_2,\epsilon_3$ (of $A_1\oplus A_1\oplus A_1=L_{\bar0}$),

$$\Delta_0=\{\pm2\epsilon_i\} \quad (i=1,2,3),$$
$$\Delta_1=\{\pm\epsilon_1\pm\epsilon_2\pm\epsilon_3\} \quad \text{(independent } \pm \text{ signs)}. \tag{6.3}$$

The distinguished system of simple roots Π, corresponding to the Cartan matrix (6.1), is given by[3]

$$\Pi=\{\epsilon_1-\epsilon_2-\epsilon_3,2\epsilon_2,2\epsilon_3\}. \tag{6.4}$$

The only root by which Π can be extended turns out to be $\gamma=-2\epsilon_1$. The corresponding root vectors can be chosen as follows[8]:

$$e_0=[1/(1+\alpha)]f_{1231}, \quad f_0=-[1/(1+\alpha)]e_{1231}. \tag{6.5}$$

Then

$$h_0=[e_0,f_0]=-[1/(1+\alpha)](2h_1-h_2-\alpha h_3). \tag{6.6}$$

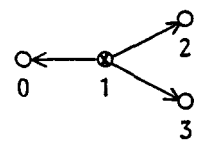
The relevant commutation relations are given by

$$[h_0,e_0]=2e_0,$$
$$[h_i,e_0]=-(1+\alpha)\delta_{i,1}e_0, \tag{6.7}$$
$$[h_0,e_i]=-\delta_{i,1}e_i \quad (i=1,2,3).$$

Hence, the extended Cartan matrix is

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1-\alpha & 0 & 1 & \alpha \\ 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{bmatrix}, \quad \tau=\{1\}. \tag{6.8}$$

To this Cartan matrix, there corresponds the following Dynkin diagram:



(6.9)

Deleting node 1 from (6.9) gives rise to the subalgebra $A_1\oplus A_1\oplus A_1$. Deleting node 0, 2, or 3 gives rise to the Dynkin diagram for $D(2,1;\alpha)$, $D(2,1;-1-\alpha)$, and $D(2,1;-\alpha/(1+\alpha))$, respectively, but these are all three isomorphic. Hence, the only nontrivial regular subalgebra of rank 3 is $A_1\oplus A_1\oplus A_1$, the even subalgebra of L. The regular subalgebras of rank 2 are $A_1\oplus A_1$ and $A(1,0)$. Note that $A(1,0)$ is not contained in any semisimple regular subalgebra of rank 3, and so it is also maximal. It is easy to check that the other choice for the simple root system does not give rise to any new regular subalgebras.

VII. THE EXCEPTIONAL LIE SUPERALGEBRA $G(3)$

The Lie superalgebra $G(3)$ is a contragredient Lie superalgebra, and it is determined by the following Cartan matrix:

$$A=\begin{bmatrix} 0 & 1 & 0 \\ -1 & 2 & -3 \\ 0 & -1 & 2 \end{bmatrix}, \quad \tau=\{1\}. \tag{7.1}$$

The roots of $G(3)$ are given by[3]

$$\begin{aligned} &\Delta_0=\{\epsilon_j-\epsilon_k, \pm\epsilon_j; \pm 2\delta\} \\ &\Delta_1=\{\pm(\epsilon_j+\delta), \pm(\epsilon_j-\delta), \pm\delta\} \end{aligned} \quad (j,k=1,2,3), \tag{7.2}$$

where $\epsilon_j-\epsilon_k, \pm\epsilon_j$ are the roots of G_2 (satisfying $\epsilon_1+\epsilon_2+\epsilon_3=0$) and $\pm 2\delta$ are the roots of A_1 in $G(3)_{\bar{0}} = G_2 \oplus A_1$. The simple roots corresponding to (7.1) are

$$\Pi=\{\delta+\epsilon_1, \epsilon_2, \epsilon_3-\epsilon_2\}. \tag{7.3}$$

Then, it is easy to verify that the only root by which Π can be extended is $\gamma=-2\delta$. Using the same convention as in (6.2), one can verify that f_{1232} is a root vector with root $-\delta$. Hence, we put

$$\begin{aligned} e_0&=\tfrac{1}{8}[f_{1232}, f_{1232}], \\ f_0&=-\tfrac{1}{8}[e_{1232}, e_{1232}]. \end{aligned} \tag{7.4}$$

One can compute that

$$h_0=[e_0, f_0]=-\tfrac{1}{2}(h_1-2h_2-3h_3). \tag{7.5}$$

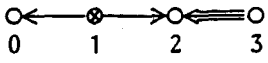
Now, the relevant commutations relations are

$$\begin{aligned} &[h_0, e_0]=2e_0, \\ &[h_i, e_0]=-4\delta_{i,1}e_0, \\ &[h_0, e_i]=-\delta_{i,1}e_i \quad (i=1,2,3). \end{aligned} \tag{7.6}$$

This gives rise to the following extended Cartan matrix:

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ 0 & -1 & 2 & -3 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \quad \tau=\{1\}, \tag{7.7}$$

with corresponding Dynkin diagram



(7.8)

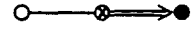
Deleting node 0 gives the Dynkin diagram for $G(3)$. When node 1 is deleted, the corresponding regular subalgebra is $A_1 \oplus G_2$. When node 2 is deleted, we find the Dynkin diagram of $A(1,0) \oplus A_1$. Finally, deleting the last node in (7.8) corresponds to omitting the last row and column in (7.7). The remaining matrix is one of type D_{-4} (see proposition 2.5.6 of Ref. 3). Hence, the corresponding regular subalgebra is $D(2,1;-4)$, which is isomorphic[3,8] to $D(2,1;3)$. In a previous paper,[8] we have shown that $D(2,1;3)$ contains a 14-dimensional representation [in general, $D(2,1;\alpha)$ has a $(4\alpha+2)$-dimensional irreducible representation for $\alpha \in \mathbb{N}$]. So, we can check on the dimensions that $G(3) \supset D(2,1;3)$: the dimension of $D(2,1;3)$ is 17, and $17+14=31$, which is the dimension of $G(3)$.

For $G(3)$, there exists another choice for the simple roots which is not equivalent to (7.3), namely

$$\Pi=\{\epsilon_3-\epsilon_1, \epsilon_1+\delta, -\delta\}. \tag{7.9}$$

Then, the corresponding Cartan matrix and Dynkin diagram are given by

$$A=\begin{bmatrix} 2 & 3 & 0 \\ -1 & 0 & -2 \\ 0 & -2 & 2 \end{bmatrix}, \quad \tau=\{2,3\}, \tag{7.10}$$

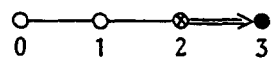


(7.11)

Actually, this situation is not being described in Ref. 3. In this case, $\gamma=\epsilon_2-\epsilon_3$, and after extending (7.9) by γ, one arrives at the following extended Cartan matrix:

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 3 & 0 \\ 0 & -1 & 0 & -2 \\ 0 & 0 & -2 & 2 \end{bmatrix}, \quad \tau=\{2,3\}, \tag{7.12}$$

with Dynkin diagram



(7.13)

When node 1 is deleted, the corresponding regular subalgebra is $A_1 \oplus B(1,1)$, which actually contains the previously found $A_1 \oplus A(1,0)$. When node 2 is deleted, we find the regular subalgebra $A_2 \oplus B(0,1)$. Finally, deleting node 3 gives rise to the regular subalgebra $A(2,0)$. Up to equivalence, (7.3) and (7.9) are the only simple root systems for $G(3)$, hence we find as maximal regular subalgebras: $A_1 \oplus G_2$, $D(2,1;3)$, $A_1 \oplus B(1,1)$, $A_2 \oplus B(0,1)$, and $A(2,0)$.

VIII. THE EXCEPTIONAL LIE SUPERALGEBRA $F(4)$

The Lie superalgebra $F(4)$ is determined by its Cartan matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 2 & -2 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \quad \tau=\{1\}. \tag{8.1}$$

The roots are expressed in terms of linear functions $\epsilon_1, \epsilon_2, \epsilon_3$, corresponding to B_3, and δ, corresponding to A_1 in $F(4)_{\bar{0}} = B_3 \oplus A_1$,

$$\begin{aligned} \Delta_0=\{&\pm(\epsilon_j+\epsilon_k), \pm(\epsilon_j-\epsilon_k), \pm\epsilon_j \\ &(1\leqslant j<k\leqslant 3); \pm 2\delta\}, \\ \Delta_1=\{&\pm\delta+\tfrac{1}{2}(\pm\epsilon_1\pm\epsilon_2\pm\epsilon_3)\}(\text{independent} \pm \text{signs}). \end{aligned} \tag{8.2}$$

The system of simple roots corresponding to (8.1) is[3]

$$\Pi=\{\delta+\tfrac{1}{2}(-\epsilon_1-\epsilon_2-\epsilon_3), \epsilon_3, \epsilon_2-\epsilon_3, \epsilon_1-\epsilon_2\}. \tag{8.3}$$

There is a unique way by which Π can be extended, namely $\gamma=-2\delta$. Using the same convention as in (6.2), we can choose the following form for the root vectors corresponding to -2δ and $+2\delta$, respectively:

$$\begin{aligned} e_0&=\tfrac{1}{3}[f_{1232}, f_{1234}], \\ f_0&=\tfrac{1}{3}[e_{1232}, e_{1234}]. \end{aligned} \tag{8.4}$$

After some calculations, one finds

$$h_0=[e_0, f_0]=-\tfrac{1}{3}(2h_1-3h_2-4h_3-2h_4). \tag{8.5}$$

Then, the relevant commutation relations are

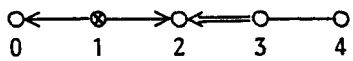
$$[h_0, e_0]=2e_0,$$

$$[h_i,e_0] = -3\delta_{i,1}e_0, \qquad (8.6)$$
$$[h_0,e_i] = -\delta_{i,1}e_i \quad (i=1,2,3,4).$$

Hence, the extended Cartan matrix becomes

$$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -3 & 0 & 1 & 0 & 0 \\ 0 & -1 & 2 & -2 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}, \quad \tau=\{1\}, \qquad (8.7)$$

with corresponding Dynkin diagram



(8.8)

Deleting node 0 gives the original algebra $F(4)$. When node 1 is deleted, we find $A_1 \oplus B_3$, which is the even subalgebra of $F(4)$. When node 2 is deleted, the corresponding regular subalgebra is $A(1,0) \oplus A_2$. Deleting node 3 corresponds to omitting row 3 and column 3 in (8.7): the remaining Cartan matrix is one of type $D_{-3} \oplus (2)$. Hence, the corresponding regular subalgebra is $D(2,1;-3) \oplus A_1$, or $D(2,1;2) \oplus A_1$. We have already noted that $D(2,1;2)$ contains a ten-dimensional irreducible representation [10], and obviously A_1 has a two-dimensional representation [2]. So, we can check the dimensions again: $\dim(D(2,1;2) \oplus A_1) = 17+3 = 20$, $\dim([10] \otimes [2]) = 20$, and $20+20 = 40 = \dim F(4)$. Finally, one can verify that deleting the last node in (8.8) again gives rise to a Dynkin diagram of $F(4)$. However, $F(4)$ has a regular subalgebra of rank 3 which is not contained in any proper regular subalgebra of rank 4, namely, $C(3)$. This algebra is obtained by deleting node 0 and 4 in (8.8). Also, note that any other choice for the simple root system of $F(4)$ does not yield any new regular subalgebras.

TABLE I. Maximal regular semisimple subalgebras of the basic classical Lie superalgebras.

L	L'
$\mathrm{spl}(m,n)$	$\mathrm{spl}(k,l) \oplus \mathrm{spl}(i,j)$, $k+i=m$, $l+j=n$;
$B(m,n)$	$B(k,l) \oplus B(i,j)$, $k+i=m$, $l+j=n$;
	$D(m,n)$;
$C(n)$	$C(k) \oplus C_j$, $k+j=n$;
$D(m,n)$	$D(k,l) \oplus D(i,j)$, $k+i=m$, $l+j=n$;
	$\mathrm{spl}(m,n)$;
$D(2,1;\alpha)$	$A_1 \oplus A_1 \oplus A_1$,
	$A(1,0)$;
$G(3)$	$A_1 \oplus G_2$,
	$D(2,1;3)$,
	$A_1 \oplus B(1,1)$,
	$A_2 \oplus B(0,1)$,
	$A(2,0)$;
$F(4)$	$A_1 \oplus B_3$,
	$A(1,0) \oplus A_2$,
	$D(2,1;2) \oplus A_1$,
	$C(3)$.

IX. CONCLUSION

In Table I we list all the maximal regular semisimple subalgebras of the basic classical Lie superalgebras. Note that isomorphisms such as $C_1 = A_1$ and $D(2,0) = D_2 = A_1 \oplus A_1$ have to be taken into account.

As far as the Lie superalgebras $\mathrm{spl}(m,n)$, $B(m,n)$, $C(n)$, and $D(m,n)$ are concerned, Table I does not yield any surprises. The results are more exciting when the exceptional Lie superalgebras are considered, with inclusions such as $G(3) \supset D(2,1;3)$ and $F(4) \supset D(2,1;2) \oplus A_1$.

[1] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[2] I. Bars, Physica (Utrecht) D **15**, 42 (1985), and references therein.
[3] V. G. Kac, Adv. Math. **26**, 8 (1977).
[4] M. Scheunert, *Springer Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).
[5] E. B. Dynkin, Am. Math. Soc. Transl. **6**, 111 (1957).
[6] V. G. Kac, *Springer Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 579.
[7] V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U.P., Cambridge, 1985).
[8] J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).

Differential-difference AKNS equations and homogeneous Heisenberg algebras

M. J. Bergvelt
Institute for Theoretical Physics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

A. P. E. ten Kroode
Mathematical Institute, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

(Received 5 March 1986; accepted for publication 17 September 1986)

The well-known evolution equations associated to the homogeneous Heisenberg algebras of Kac–Moody algebras $g^{(1)}$ (AKNS systems) are extended by differential-difference equations that can be written in zero curvature form.

I. INTRODUCTION

It is well known that the Toda lattice and the nonlinear Schrödinger equation are intimately related, see, e.g., Jimbo and Miwa[1] and Flaschka.[2] In Ref. 3 we showed how in the representation theoretic approach to soliton equations one can associate in a natural way a hierarchy of differential-difference equations to the homogeneous Heisenberg algebra of the Kac–Moody algebra $A_1^{(1)}$. The first and second nontrivial members of the hierarchy are the equation for the Toda lattice and the nonlinear Schrödinger equation, respectively.

In a different approach soliton equations are obtained as conditions of commutativity of a set of covariant derivatives D_{t_i} (see for example Newell[4] and Drinfeld and Sokolov[5]). In this paper we extend these equations by introducing, apart from the variables t_i and associated covariant derivatives D_{t_i}, a lattice and covariant derivatives D_δ on the lattice. The equations $[D_{t_i}, D_\delta] = 0$ are the differential-difference equations supplementing the evolution equations of the form $[D_{t_i}, D_{t_j}] = 0$.

In this setup the continuous covariant derivatives are constructed from the positive generators of the homogeneous Heisenberg algebra of an affine Kac–Moody algebra $\mathsf{g}^{(1)}$. The lattice covariant derivatives are defined using elements of the centralizer of this Heisenberg algebra in the loop group associated to $\mathsf{g}^{(1)}$. At the same time the structure of the centralizer forces the lattice introduced here to be the coroot lattice of the underlying finite-dimensional Lie algebra g.

For the case of $A_1^{(1)}$ we once more find the connection between the Toda lattice and the nonlinear Schrödinger equation, but from a point of view apparently completely different from that in Ref. 3.

II. CONTINUOUS ZERO CURVATURE EQUATIONS

In this section we recall the construction of evolution equations related to the homogeneous Heisenberg algebra of an affine Kac-Moody algebra $\mathsf{g}^{(1)}$. For background we refer to the review by Drinfeld and Sokolov.[5]

Let g be a simple finite-dimensional Lie algebra, h a Cartan subalgebra of g, $(|)$ the Killing–Cartan form on g, and

$$L(\mathsf{g}) = \underset{i\in\mathbb{Z}}{\oplus} \lambda^i \otimes \mathsf{g}, \tag{2.1}$$

the loop algebra associated to g. The algebra $L(\mathsf{g})$ contains a maximal Abelian subalgebra

$$\mathsf{s} = \underset{i\in\mathbb{Z}}{\oplus} \lambda^i \otimes \mathsf{h}. \tag{2.2}$$

In the sequel we will often suppress the tensor product symbol for elements of (2.1) or (2.2).

Let $\{h^a, a = 1,2,\ldots,\text{rank } \mathsf{g}\}$ be an orthonormal basis of h. Then s is generated by

$$\mathsf{h}, \quad p_i^a = \lambda^i h^a, \quad q_i^a = \lambda^{-i} h^a / i, \quad i > 0. \tag{2.3}$$

[The p_i^a and q_i^a together with a central element c generate the homogeneous Heisenberg algebra of $\mathsf{g}^{(1)} = L(\mathsf{g}) \oplus \mathbb{C}c \oplus \mathbb{C}d$, the affine Kac–Moody algebra of type $k = 1$ associated to g, see Ref. 6. In the rest of this paper we will only work with $L(\mathsf{g})$, not with the full affine algebra $\mathsf{g}^{(1)}$.]

On $L(\mathsf{g})$ we define a bilinear form

$$(\lambda^j x | \lambda^k y) = \delta_{j+k,0}(x|y). \tag{2.4}$$

With respect to (2.4), we have an orthogonal decomposition

$$L(\mathsf{g}) = \mathsf{s} \oplus \mathsf{s}^\perp. \tag{2.5}$$

An element h of h is called regular if for all roots α in h^* we have $\langle \alpha, h \rangle \neq 0$. Fix some regular vector $h = h^1$, extend this to an orthonormal basis $\{h^a, a = 1,2,\ldots,\text{rank } \mathsf{g}\}$ of h, and introduce a covariant derivative

$$D_x = \partial_x - \lambda h - v(x) = \partial_x - R_x. \tag{2.6}$$

The field $v(x)$ takes values in the orthocomplement of h in g.

We want to introduce other covariant derivatives $D_t = \partial_t - R_t$ such that the condition $[D_x, D_t] = 0$, which can be rewritten as

$$\partial_t v = [D_x, R_t], \tag{2.7}$$

gives an evolution equation for the field v.

The construction of R_t's leading to sensible evolution equations is well known (see, e.g., Refs. 5 and 7). One first constructs a resolvent[8] $R(x)$ of D_x. This is a function with values in $L(\mathsf{g})$ such that

$$[D_x, R] = 0. \tag{2.8}$$

 0022-2488/87/020302-05$02.50 © 1987 American Institute of Physics

[To be more precise the resolvent takes values in the Lie algebra

$$\mathrm{g}((\lambda^{-1})) = \mathbb{C}((\lambda^{-1})) \otimes \mathrm{g},$$

with $\mathbb{C}((\lambda^{-1}))$ the algebra of formal power series of the form

$$\sum_{i=0}^{\infty} c_i \lambda^{n-i}, \quad c_i \text{ in } \mathbb{C},$$

and n some positive number.[5] We will be somewhat careless about the distinction between $L(\mathrm{g})$ and $g((\lambda^{-1}))$.]

Then one makes a decomposition

$$R = R_- + R_+, \tag{2.9}$$

where R_- contains only negative powers in λ and R_+ only non-negative powers. Finally one proves that one can take for R_t in (2.7) the part R_+ or $-R_-$ of any resolvent R.

To construct resolvents we perform a gauge transformation on the covariant derivative D_x. As is well known there exists a unique function $k(x)$ taking values in $\mathrm{s}^{\perp}$ of the form $k = \lambda^{-1}k_1 + \lambda^{-2}k_2 + \cdots$, such that the gauge transformed covariant derivative

$$\overline{D}_x = e^{adk}(D_x) = \partial_x - \overline{R}_x, \tag{2.10}$$

has a gauge potential $\overline{R}_x$ that takes values in s (see, e.g., Ref. 5, Proposition 4.1). We refer to the situation after the gauge transformation by e^k as "the diagonal gauge." Objects in the diagonal gauge will be distinguished by an overbar, as in (2.10).

The resolvents of $\overline{D}_x$ can be shown to be the constant elements of s and hence the resolvents of D_x are of the form

$$R(s) = e^{-adk}(s), \quad s \in \mathrm{s}. \tag{2.11}$$

Only resolvents of D_x containing positive powers in λ will give nontrivial equations (2.7). Hence these equations are linear combinations of

$$\partial_{t_i^a}(v) = [D_x, R_+(p_i^a)]. \tag{2.12}$$

A short calculation yields

$$R_x = R_+(\lambda h), \tag{2.13}$$

and hence $\partial/\partial x = \partial/\partial t_1^1$. One can also show that

$$[D_{t_i^a}, D_{t_j^b}] = 0, \quad [D_{t_i^a}, R(p_j^b)] = 0, \tag{2.14}$$

and that in the diagonal gauge all $D_{t_i^a}$ are diagonal, with

$$D_{t_i^a} = \partial_{t_i^a} - R_+(p_i^a). \tag{2.15}$$

Note that we can also concentrate on covariant derivatives

$$\hat{D}_{t_i^a} = \partial_{t_i^a} + R_-(p_i^a), \tag{2.16}$$

without changing the above results.

III. COVARIANT DERIVATIVES ON A LATTICE

Consider an m-dimensional lattice. On the vertices we have fields ψ^l, with l in $\mathbb{Z}^m$, taking values in some representation space of a gauge group. Under (local) gauge transformation we have

$$\psi^l \to \tilde{\psi}^l = g^l \psi^l. \tag{3.1}$$

In the continuum covariant derivatives define infinitesimal parallel transport. On the lattice parallel transport is defined by specifying a collection of parallel transport operators U_δ^l taking values in the gauge group, with δ a step (i.e., a path) on the lattice. Parallel transporting the field ψ^l by U_δ^l yields a field $\hat{\psi}^{l+\delta}$ in the point $l+\delta$:

$$\hat{\psi}^{l+\delta} = U_\delta^l \psi^l. \tag{3.2}$$

The field ψ^l is called covariantly constant if we have

$$\psi^{l+\delta} = U_\delta^l \psi^l. \tag{3.3}$$

Then define the covariant derivative on the lattice by

$$D_\delta^l \psi^l = (U_\delta^l)^{-1} \psi^{l+\delta} - \psi^l. \tag{3.4}$$

Under gauge transformations we have

$$D_\delta^l \psi^l \to \tilde{D}_\delta^l \tilde{\psi}^l = g^l D_\delta^l \psi^l, \tag{3.5}$$

if and only if under gauge transformations

$$U_\delta^l \to \tilde{U}_\delta^l = g^{l+\delta} U_\delta^l (g^l)^{-1}. \tag{3.6}$$

Now suppose that the field ψ^l also depends on a continuous parameter t and that there is a covariant derivative $D_t^l = \partial_t - R_t^l$. If we demand that the field ψ^l is covariantly constant both in a lattice direction δ and in the continous t direction, we must have

$$D_\delta \psi^l = 0, \quad D_t \psi^l = 0. \tag{3.7}$$

The compatibility condition for (3.7) is

$$R_t^{l+\delta} = \partial_t (U_\delta^l)(U_\delta^l)^{-1} + U_\delta^l (R_t^l)(U_\delta^l)^{-1}. \tag{3.8}$$

IV. DIFFERENTIAL-DIFFERENCE ZERO CURVATURE EQUATIONS

In this section we will extend the theory of Sec. II by introducing, apart from the continuous variables t_i^a, discrete variables l belonging to some m-dimensional lattice $\mathbb{Z}^m$, with m as yet unspecified.

In every point of the lattice we have a field v^l and a covariant derivative

$$D_x^l := \partial_x - R_x^l := \partial_x - (\lambda h + v^l). \tag{4.1}$$

There exists a unique $k^l = \Sigma_{i>0} k_i^l \lambda^{-i}$ in $\mathrm{s}^{\perp}$ such that

$$e^{k^l} D_x^l e^{-k^l} = \partial_x - \overline{R}_x^l, \quad \overline{R}_x^l \text{ in } \mathrm{s}. \tag{4.2}$$

Resolvents $R^l(s)$ of D_x^l are given by

$$R^l(s) = e^{-k^l} s e^{k^l}, \quad s \text{ in } \mathrm{s}. \tag{4.3}$$

Using the positive part of $R^l(p_i^a)$ we define covariant derivatives

$$D_{t_i^a}^l = \partial_{t_i^a} - R_+^l(p_i^a). \tag{4.4}$$

The zero curvature conditions $[D_x^l, D_{t_i^a}^l] = 0$ give evolution equations for the fundamental fields v^l of the form

$$\partial_{t_i^a} v^l = F(v^l, \partial_x v^l, \ldots). \tag{4.5}$$

The theory of this section is up to now just a trivial extension of the theory of Sec. II; we have provided all fields with a multi-index l, but the fields v^l and $v^{l+\delta}$ are completely unrelated. We will change this situation by introducing as in Sec. III a covariant derivative D_δ^l and by imposing the compatibility condition

$$[D_\delta^l, D_x^l] = 0, \tag{4.6}$$

or, equivalently [see (3.8)]

$$R^{l+\delta}_{+}(\lambda h) = (\partial_x U^l_\delta)(U^l_\delta)^{-1} + U^l_\delta R^{l}_{+}(\lambda h)(U^l_\delta)^{-1}. \tag{4.7}$$

This equation tells us how the field $v^{l+\delta}$ is coupled to the field v^l.

In order for (4.7) to be a sensible equation, we will have to make a suitable choice for the parallel transport operators U^l_δ. We will be able to choose the U^l_δ's such that the compatibility conditions $[D_\delta, D_{t^a_i}] = 0$ are automatically satisfied. Then these conditions will not lead to new relations between the fundamental fields v^l and $v^{l+\delta}$, in the same way as in the continuum the conditions $[D_{t^a_i}, D_{t^b_j}] = 0$ do not lead to new equations.

To find U^l_δ's we proceed in a similar way as in Sec. II. There we introduced a resolvent R as an element of the loop algebra that gives a trivial evolution of the fundamental field v [see (2.7)]. Here we define what we will call—for lack of a better name—a lattice resolvent U^l as an element of the loop group $L(\mathsf{G})$, with G the simply connected Lie group whose Lie algebra is g. We demand that the lattice resolvent substituted in (4.7) gives a trivial evolution

$$R^{l}_{+}(\lambda h) = \partial_x U^l (U^l)^{-1} + U^l (R^{l}_{+}(\lambda h))(U^l)^{-1}. \tag{4.8a}$$

[To be more precise; the lattice resolvent will take its values not in the loopgroup $L(\mathsf{G})$ associated to $L(\mathsf{g})$ (see, e.g., Ref. 9), but in the group $\mathsf{G}(C((\lambda^{-1})))$ associated to the Lie algebra $\mathsf{g}((\lambda^{-1}))$. Again we will be careless about this distinction.]

In Sec. II we saw that $-R^{l}_{-}(\lambda h)$ plays essentially the same role as $R^{l}_{+}(\lambda h)$, and therefore it is natural to require, in addition to (4.8a),

$$-R^{l}_{-}(\lambda h) = \partial_x (U^l)(U^l)^{-1} - U^l (R^{l}_{-}(\lambda h))(U^l)^{-1}. \tag{4.8b}$$

Combining (4.8a) and (4.8b) we find

$$R^l(\lambda h) = U^l R^l(\lambda h)(U^l)^{-1}. \tag{4.9}$$

In the diagonal gauge (4.9) reads

$$\lambda h = \overline{U}^l \lambda h (\overline{U}^l)^{-1}, \tag{4.10}$$

where $\overline{U}^l = e^{k^l} U^l e^{-k^l}$.

So $\overline{U}^l$ must belong to the centralizer of λh. Because h is regular, we may conclude that $\overline{U}^l$ is an element of the (Abelian) centralizer S of the whole algebra s (see Ref. 9). Combining this with equations (4.8a) and (4.8b) in the diagonal gauge and recalling that $\overline{R}^{l}_{\pm}(\lambda h)$ take values in s we conclude that $\overline{U}^l$ must be a constant element of S. All lattice resolvents are therefore of the form

$$U^l = e^{-k^l} S e^{k^l}, \quad S \text{ constant in } \mathsf{S}. \tag{4.11}$$

Conversely it is not difficult to show that (4.11) implies both equations (4.8a) and (4.8b).

Now suppose that we have found some lattice resolvent U^l and that moreover U^l admits a factorization

$$U^l = U^l_- U^l_+ \tag{4.12}$$

with

$$\begin{aligned} U^l_- &= 1 + U^l_{-1}\lambda^{-1} + U^l_{-2}\lambda^{-2} + \cdots, \\ U^l_+ &= U^l_0 + U^l_1 \lambda + U^l_2 \lambda^2 + \cdots. \end{aligned} \tag{4.13}$$

The inverses of U^l_-, U^l_+ will also have the form (4.13). The factorization (4.12) which takes place in the loop group is the analog of the decomposition (2.9) on the level of the loop algebra. If it exists it is unique.

Substituting (4.12) in (4.8a) we find

$$\begin{aligned} &\partial_x U^l_+ (U^l_+)^{-1} + U^l_+ R^{l}_{+}(\lambda h)(U^l_+)^{-1} \\ &\quad = \partial_x (U^l_-)^{-1} U^l_- + (U^l_-)^{-1} R^{l}_{+}(\lambda h) U^l_-. \end{aligned} \tag{4.14}$$

The left-hand side of this equation contains only non-negative powers of λ, while the right-hand side has λh as the term containing the highest power of λ. Therefore (4.14) is an expression of the form $A_0 + \lambda h$. Using this we find, see expansion (4.13),

$$\begin{aligned} &(\partial_x U^l_+)(U^l_+)^{-1} + U^l_+ R^{l}_{+}(\lambda h)(U^l_+)^{-1} \\ &\quad = \lambda h + v^l + [h, U^l_{-1}]. \end{aligned} \tag{4.15}$$

Since $[h, U^l_{-1}]$ belongs to $\mathsf{h}^\perp$ we can choose $U^l_\delta = U^l_+$ in (4.7). We obtain

$$v^{l+\delta} = v^l + [h, U^l_{-1}], \tag{4.16}$$

and $v^{l+\delta}$ takes value in $\mathsf{h}^\perp$.

The rest of this section will be devoted to the calculation of the commutators $[D^l_\delta, D^l_{t^a_i}]$ and $[D_\delta, \hat{D}_{t^a_i}]$, where

$$\hat{D}^l_{t^a_i} = \partial_{t^a_i} + R^l_-(p^a_i). \tag{4.17}$$

First of all we remark that the condition $[D^l_\delta, D^l_x] = 0$ implies that the expression $U^l_+ R^l(\lambda h)(U^l_+)^{-1} = (U^l_-)^{-1} R^l(\lambda h) U^l_-$ is a resolvent for $D^{l+\delta}_x$. In particular we have

$$[D^{l+\delta}_x, (U^l_-)^{-1} R^l(\lambda h) U^l_-] = 0. \tag{4.18}$$

Therefore we must have in the diagonal gauge

$$(\overline{U}^l_-)^{-1} \lambda h (\overline{U}^l_-) \text{ in } \mathsf{s}, \tag{4.19}$$

where $\overline{U}^l_- = e^{k^l} U^l_- e^{-k^{l+\delta}}$. Here $\overline{U}^l_-$ and $(\overline{U}^l_-)^{-1}$ have expansions

$$\overline{U}^l_- = 1 + \sum_{i=1}^{\infty} V_i \lambda^{-i}, \quad (\overline{U}^l_-)^{-1} = 1 + \sum_{i=1}^{\infty} W_i \lambda^{-i}. \tag{4.20}$$

The V_i and W_i's are related by

$$\begin{aligned} &W_1 + V_1 = 0, \\ &V_i + W_i + \sum_{k=1}^{i-1} W_k V_{i-k} = 0, \quad i > 1. \end{aligned} \tag{4.21}$$

Using this in (4.19) we find

$$\begin{aligned} &[h, V_1] \text{ in } \mathsf{h}, \\ &[h, V_i] + \sum_{k=1}^{i-1} W_k [h, V_{i-k}] \text{ in } \mathsf{h} \quad i > 1. \end{aligned} \tag{4.22}$$

Using the regularity of h we prove inductively that V_i belongs to the universal enveloping algebra of h, therefore, $[h, V_i] = 0$ and (4.19) yields

$$(\overline{U}^l_-)^{-1} (\lambda h) \overline{U}^l_- = \lambda h. \tag{4.23}$$

Again using the regularity of h we find that $\overline{U}^l_-$ belongs to

the centralizer S of s. Since $\overline{U}^l$ also does, this is even true for $\overline{U}^l{}_+$:

$$\overline{U}^l{}_+(p_i^a)(\overline{U}^l{}_+)^{-1} = (\overline{U}^l{}_-)^{-1}(p_i^a)\overline{U}^l{}_- = p_i^a. \tag{4.24}$$

From this it follows that

$$\begin{aligned} R^{l+\delta}(p_i^a) &= U^l{}_+ R^l(p_i^a)(U^l{}_+)^{-1} \\ &= (U^l{}_-)^{-1}R^l(p_i^a)U^l{}_-. \end{aligned} \tag{4.25}$$

Using (4.25), the decomposition $R^l = R^l{}_- + R^l{}_+$ and the factorization $U^l = U^l{}_- U^l{}_+$, one easily derives

$$\begin{aligned} &R^{l+\delta}_+(p_i^a) - \partial_{t_i^a}U^l{}_+(U^l{}_+)^{-1} - U^l{}_+ R^l{}_+(p_i^a)(U^l{}_+)^{-1} \\ &\quad = -R^{l+\delta}_- - \partial_{t_i^a}(U^l{}_-)^{-1}U^l{}_- \\ &\qquad + (U^l{}_-)^{-1}R^l{}_-(p_i^a)U^l{}_-. \end{aligned} \tag{4.26}$$

Since the left-hand side of (4.26) contains only non-negative powers of λ and the right-hand side only negative powers, both sides must be equal to zero, which is equivalent to

$$[D^l_\delta, D^l_{t_i^a}] = [D^l_\delta, \hat{D}^l_{t_i^a}] = 0. \tag{4.27}$$

V. PRODUCTS OF FACTORIZING LATTICE RESOLVENTS

With (4.11) we have constructed all lattice resolvents. However, it is not at all clear that these admit a factorization of the form (4.12). In this section we show that if two lattice resolvents $U^l(S_1)$ and $U^l(S_2)$ factorize their product

$$U^l(S_1S_2) = U^l(S_1)U^l(S_2) = U^l(S_2)U^l(S_1), \tag{5.1}$$

also does and that there are simple relations between all factors.

Let $\delta_{1,2}$ be the steps on the lattice associated to parallel transport operators $U^l{}_+(S_{1,2})$. Since in the diagonal gauge $\overline{U}^l{}_+(S_{1,2})$ belongs to S, and S is Abelian we have

$$S_1 = \overline{U}^l{}_+(S_2)S_1\overline{U}^l{}_+(S_2)^{-1}, \tag{5.2}$$

and therefore

$$U^{l+\delta_2}(S_1) = U^l{}_+(S_2)U^l(S_1)U^l{}_+(S_2)^{-1}. \tag{5.3}$$

Performing the factorization of both lattice resolvents in (5.3) we obtain

$$\begin{aligned} &U^{l+\delta_2}_-(S_1)U^{l+\delta_2}_+(S_1)U^l{}_+(S_2) \\ &\quad = U^l{}_+(S_2)U^l{}_-(S_1)U^l{}_+(S_1). \end{aligned} \tag{5.4}$$

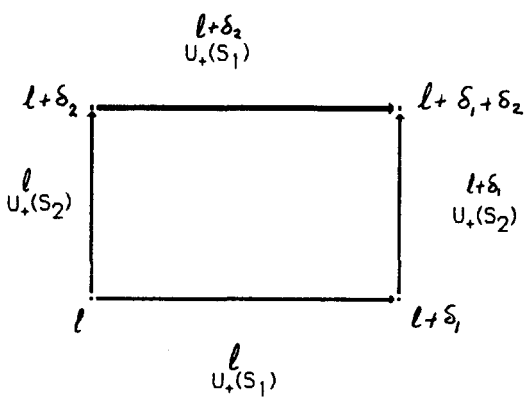


FIG. 1. The commutativity of lattice covariant derivatives.

Next consider the resolvent (5.1). Using (5.4) and the equation obtained by interchanging 1 and 2, we see that (5.1) admits a factorization with

$$\begin{aligned} U^l{}_-(S_1S_2) &= U^l{}_-(S_2)U^{l+\delta_2}_-(S_1) \\ &= U^l{}_-(S_1)U^{l+\delta_1}_-(S_2), \\ U^l{}_+(S_1S_2) &= U^{l+\delta_2}_+(S_1)U^l{}_+(S_2) \\ &= U^{l+\delta_1}_+(S_2)U^l{}_+(S_1). \end{aligned} \tag{5.5}$$

Note that with (5.5) we have proven the commutativity of covariant lattice derivatives (see Fig. 1). This means that the imposition of conditions $[D_{\delta_1}, D_{\delta_2}] = 0$ does not lead to new equations; they are automatically satisfied.

VI. STRUCTURE OF S AND OF THE LATTICE

In the previous section we found that the product of two factorizing lattice resolvents again factorizes. If we can write any element S of S as the product of some S_i's with factorizing lattice resolvent we have proven that any $U^l(S)$ factorizes.

The group S can—according to Kac and Peterson[9]—be described as follows. Let α be an element of the coroot lattice Q^v of g and β an arbitrary element of h. Then S is the collection of loops $h(\alpha,\beta)(\lambda)$ in the loopgroup $L(G)$ of the form

$$h(\alpha,\beta)(\lambda) = \exp(i\phi\alpha + 2\pi i\beta), \quad \lambda = \exp i\phi. \tag{6.1}$$

First consider the subgroup of loops $\hat{\beta} = h(0,\beta)$. The lattice resolvent associated to $\hat{\beta}$ certainly factorizes, since $\hat{\beta}$ is λ independent:

$$U^l(\hat{\beta}) = e^{-k^l}\hat{\beta}e^{k^l} = (U^l(\hat{\beta})\hat{\beta}^{-1})(\hat{\beta}). \tag{6.2}$$

Consider the transformation (4.7) induced by $U^l{}_+(\hat{\beta}) = \hat{\beta}$. Since $\hat{\beta}$ is constant we have

$$R^{l+\delta}_x = \hat{\beta}R^l_x\hat{\beta}^{-1}, \tag{6.3}$$

or, in terms of the field v^l,

$$v^{l+\delta} = \hat{\beta}v^l\hat{\beta}^{-1}. \tag{6.4}$$

This is a trivial linear transformation on the components of the field v^l. Therefore we will discard these transformations by quotienting the group S by $\{h(0,\beta)\}$. The resulting group

$$T = \{\exp i\phi\alpha | \alpha \text{ in } Q^v\}, \tag{6.5}$$

is a discrete group called the translation group.[6] It is isomorphic to the coroot lattice and generated by the elements corresponding to a set of simple coroots α_i in Q^v:

$$T_i = \exp i\phi\alpha_i. \tag{6.6}$$

The lattice introduced in Sec. IV will therefore be taken to be the coroot lattice Q^v of g.

Next we should show that the lattice resolvents $U^l(T_i)$ factorize. We have not yet proved this in full generality. In the next section we will discuss the case of $A_1^{(1)}$.

VII. LATTICE EQUATION ASSOCIATED TO $A_1^{(1)}$

We now concentrate on the case $g = A_1 = sl(2,C)$. The covariant derivative (2.5) is taken to be

$$D^l_x = \partial_x - \frac{\lambda}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} - \begin{pmatrix}0 & q^l\\ r^l & 0\end{pmatrix}. \tag{7.1}$$

The first terms of the diagonalizing element k^l of (2.10) are

$$k^l = \begin{pmatrix} 0 & q^l \\ -r^l & 0 \end{pmatrix}\lambda^{-1} + \begin{pmatrix} 0 & \partial_x q^l \\ \partial_x r^l & 0 \end{pmatrix}\lambda^{-2} + \cdots . \tag{7.2}$$

The element T_1 corresponding to the single simple coroot α_1 is

$$T_1 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}. \tag{7.3}$$

The lattice resolvent associated to T_1 is

$$U^l(T_1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\lambda + \begin{pmatrix} 0 & q^l \\ r^l & 0 \end{pmatrix} + \begin{pmatrix} -q^l r^l & \partial_x q^l \\ -\partial_x r^l & 1 + q^l r^l \end{pmatrix}\lambda^{-1} + \cdots . \tag{7.4}$$

This lattice resolvent admits a factorization with

$$U^l_+(T_1) = \begin{pmatrix} \lambda - \partial_x \ln q^l & q^l \\ -1/q^l & 0 \end{pmatrix}. \tag{7.5}$$

The differential-difference equation (4.16) obtained from (7.5) reads

$$\begin{aligned} q^{l+\delta} &= -(q^l)^2 r^l + q^l \partial_x^2 \ln(q^l), \\ r^{l+\delta} &= -1/q^l. \end{aligned} \tag{7.6}$$

Substituting $q^l = e^{u^l}$ we find

$$\partial_x^2 u^l = e^{u^{l+\delta} - u^l} - e^{u^l - u^{l-\delta}}, \tag{7.7}$$

the equation for the Toda lattice.[2]

The first nontrivial evolution equation (2.12) associated to the covariant derivative (7.1) is a system of coupled nonlinear Schrödinger equations, see Ref. 4.

Note that U^l_+ will become singular if we let q^l become zero in (7.5). One can take two points of view on this problem.

First, one can consider q^l, r^l to be elements of a differential algebra (generated by q^l, r^l, and a derivation ∂_x). Then U^l_+ will belong to the field of fractions of this algebra. In this purely algebraic interpretation q^l cannot become zero and there will be no problem.

On the other hand, if one is interested in solutions to soliton equations one is forced to interpret $q^l(x)$ as an honest function. The conclusion then is that as soon as q^l (or r^l if one considers the lattice resolvent associated to T_1^{-l}), becomes zero, the buildup of the lattice stops and the lattice will be of finite extent.

This can also be understood in terms of τ functions. A τ function for the AKNS hierarchy has components τ^l, l in $\mathbb{Z}$ (see, e.g., Ref. 3) and is related to the fields q^l, r^l by

$$q^l = \tau^{l+1}/\tau^l, \quad r^l = -\tau^{l-1}/\tau^l.$$

In general, most of the components will be zero and hence only a finite number of fields q^l, r^l will make sense. For instance, the vacuum solution has a τ function with $\tau^0 = 1$, $\tau^i = 0$, $i \neq 0$, corresponding to $q^0 = 0$, $r^0 = 0$ and the other q^i, r^i undefined.

VIII. REMARKS

(1) It would be interesting to investigate whether the Hamiltonian structures and conservation laws[5] of the equations (2.12) can be extended to cover the lattice equations introduced in this paper. Also other aspects of soliton equations, such as the dressing method, Miura transformations, two-dimensional Toda field equations, etc. deserve further study in the light of the results obtained above.

(2) The equations described in this paper are related to the homogeneous Heisenberg algebra of $\mathfrak{g}^{(1)}$. Recently Kac and Peterson[9] have classified all conjugacy classes of Heisenberg algebras of any affine Kac–Moody algebra $g(A)$. In a forthcoming paper[10] we will use this result to associate a hierarchy of zero curvature equations to any such conjugacy class, containing in general also lattice equations. In the case of the principal Heisenberg algebra one obtains the equations of Drinfeld and Sokolov,[11] without any lattice equations, because of the special structure of the centralizer of this Heisenberg algebra.

(3) Kac and Peterson[9] also show that one can associate to any conjugacy class of Heisenberg subalgebras a vertex realization of the basic representation $L(\Lambda_0)$ of $g(A)$. Each realization will, we expect, lead to a realization of the defining equations of the group orbit through the vacuumvector in $L(\Lambda_0)$, i.e., it will lead to a new hierarchy of soliton equations (Jimbo and Miwa[12] and Kac[6]). This is presently under investigation.

(4) It has become increasingly clear that there are many similarities between the zero curvature construction of soliton equations and the representation theoretic approach. The major difference seems to be the absence of the central extension of $L(\mathfrak{g})$ in the zero curvature construction. The precise connection remains to be clarified. For attempts in this direction see Refs. 4 and 13.

[1] M. Jimbo and T. Miwa, Physica D **2**, 306, 407 (1981); D **4**, 26 (1981).
[2] H. Flaschka, in *Non-Linear Integrable Systems—Classical Theory and Quantum theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).
[3] A. P. E. ten Kroode and M. J. Bergvelt, Lett. Math. Phys. **12**, 139 (1986).
[4] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[5] V. G. Drinfeld and V. V. Sokolov, J. Sov. Math. **30**, 1975 (1985).
[6] V. G. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Boston, 1983).
[7] G. Wilson, Ergod, Theory Dynamical Systems, **1**, 361 (1981).
[8] L. A. Dickey, Commun. Math. Phys. **82**, 345 (1981).
[9] V. G. Kac and D. H. Peterson, "112 constructions of the basic representations of the loopgroup of E_8," to appear in proceedings, *Anomalies, Geometry, Topology* (World Scientific, Singapore).
[10] M. J. Bergvelt and A. P. E. ten Kroode, in preparation.
[11] V. G. Drinfeld and V. V. Sokolov, Sov. Math. Dokl. **23**, 457 (1981).
[12] M. Jimbo and T. Miwa, Publications R.I.M.S. **18**, 1077 (1982), **19**, 943 (1983).
[13] G. Wilson, C. R. Acad. Sci. Paris **299**, 587 (1984).

On invariance properties of the wave equation

George Bluman
Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Y4

Sukeyuki Kumei
Faculty of Textile Science, Shinsu University, Tokida 3-15-1, Ueda, Japan

(Received 5 December 1985; accepted for publication 8 October 1986)

A complete group classification is given of both the wave equation $c^2(x)u_{xx} - u_{tt} = 0$ (I) and its equivalent system $v_t = u_x$, $c^2(x)v_x = u_t$ (II) when the wave speed $c(x) \neq$ const. Equations (I) and (II) admit either a two- or four-parameter group. For the exceptional case, $c(x) = (Ax + B)^2$, equation (I) admits an infinite group. Equations (I) and (II) do not always admit the same group for a given $c(x)$: The group for (I) can have more parameters or fewer parameters than that for (II); moreover, the groups can be different with the same number of parameters. Separately for (I) and (II), all possible $c(x)$ that admit a four-parameter group are found explicitly. The corresponding invariant (similarity) solutions are considered. Some of these wave speeds have realistic physical properties: $c(x)$ varies monotonically from one positive constant to another positive constant as x goes from $-\infty$ to $+\infty$.

I. INTRODUCTION

In this paper we consider invariance properties of second-order hyperbolic partial differential equations (PDE's) (wave equations)

$$c^2(x)u_{xx} - u_{tt} = 0 \tag{1.1}$$

and corresponding hyperbolic systems

$$v_t = u_x, \quad u_t = c^2(x)v_x. \tag{1.2}$$

Their invariance properties are used to construct solutions of these PDE's for various classes of wave speeds $c(x)$.

An important related equation is

$$(c^2(x)v_x)_x - v_{tt} = 0. \tag{1.3}$$

Many physical problems lead to (1.1)–(1.3). Equation (1.1) arises in the study of small transverse vibrations of a string with variable density, system (1.2) in the study of transmission lines with variable capacitance or variable resistance, and Eq. (1.3) in the study of small longitudinal vibrations of a rod with variable Young's modulus.

Equations (1.1)–(1.3) are equivalent in the following senses [(1.4)–(1.7)]:

if $\{u(x,t), v(x,t)\}$ satisfy (1.2),
then $u(x,t)$ solves (1.1)
and $v(x,t)$ solves (1.3); (1.4)

if $u = F(x,t)$ satisfies (1.1),
then $(u,v) = (F_t, F_x)$ solves (1.2)
and $v = F_x$ solves (1.3); (1.5)

if $v = G(x,t)$ satisfies (1.3),
then $(u,v) = (c^2(x)G_x, G_t)$ solves (1.2)
and $u = c^2(x)G_x$ solves (1.1). (1.6)

Under the transformation

$$y = \int c^2(x)dx,$$

Eq. (1.3) can be rewritten as an equation of the form (1.1), namely,

$$c^{-1}(x(y))v_{yy} - v_{tt} = 0. \tag{1.7}$$

In spite of the apparent equivalence of a single PDE and a corresponding system of PDE's it does not necessarily follow that their respective invariance groups of point transformation are the same. It could happen that the group of point transformations leaving invariant the system is larger than that leaving invariant the single equation; also the converse could be true. We will show that this is indeed the case for the single equation (1.1) and the corresponding system (1.2). For example we show that if $c(x) = (Ax + B)^2$, then (1.1) is invariant under an infinite Lie group of point transformations, whereas the Lie group of point transformations leaving invariant (1.2) has only four parameters; if $c(x) = \sqrt{A + Be^{kx}}$, then the Lie group of (1.1) has two parameters and that of (1.2) has four parameters.

Consequently it follows that invariant (similarity) solutions of a system of PDE's lead to noninvariant solutions of a corresponding equivalent single PDE and vice versa. In Sec. IV of this paper we construct such noninvariant solutions for (1.1).

It is important to note that under the hodograph transformation (the interchange of dependent and independent variables), system (1.2) is equivalent to the nonlinear system

$$v_t = u_x, \quad u_t = c^2(v)v_x. \tag{1.8}$$

Consequently if $\{u(x,t), v(x,t)\}$ solve (1.8) then $v(x,t)$ solves

$$(c^2(v)v_x)_x - v_{tt} = 0, \tag{1.9}$$

and introducting the potential $\phi(x,t)$, where $(u,v) = (\phi_t, \phi_x)$, the system (1.8) reduces to

$$c^2(\phi_x)\phi_{xx} - \phi_{tt} = 0. \tag{1.10}$$

The rest of this paper is organized as follows.

In Sec. II the Lie group of point transformations admitted by (1.1) is derived for all possible wave speeds $c(x)$. The corresponding invariant solutions are constructed.

In Sec. III the Lie group of point transformations ad-

 0022-2488/87/020307-12$02.50 © 1987 American Institute of Physics

mitted by (1.2) is derived for all possible $c(x)$. If $c(x)$ satisfies the ordinary differential equation

$$cc'(c/c')'' = \pm\lambda^2, \quad \lambda \neq 0, \tag{1.11}$$

then (1.2) admits a larger group than (1.1). (Throughout this paper a prime denotes differentiation of a function of a single variable.) Invariant solutions of (1.2) and hence solutions of (1.1) are constructed for $c(x)$ satisfying (1.11).

In Sec. IV we discuss the differences between the invariance properties of the single equation (1.1) and the system (1.2). We show that in general the Lie group of point transformation leaving invariant (1.2) [(1.1)] does not necessarily correspond to a Lie group of point transformations or Lie–Bäcklund transformations leaving invariant (1.1) [(1.2)].

In Sec. V we find the equivalence classes of wave speeds $c(x)$ for the wave equation (1.1).

II. THE INVARIANCE PROPERTIES OF THE WAVE EQUATION AS A SINGLE EQUATION

Lie[1,2] proved that a second-order linear hyperbolic PDE with two independent variables admits a group of point transformations containing at most four parameters if it does not admit an infinite group. Lie did not study specifically the wave equation (1.1).

A. Infinitesimal transformations

By using Lie's algorithm,[2,3] one can find the generators of the invariance group of point transformations of (1.1). If the point transformation

$$\begin{aligned} X &= x + \epsilon\xi(x,t) + O(\epsilon^2), \\ T &= t + \epsilon\tau(x,t) + O(\epsilon^2), \\ U &= u + \epsilon f(x,t)u + O(\epsilon^2), \end{aligned} \tag{2.1}$$

leaves (1.1) invariant, then its infinitesimals $\{\xi,\tau,f\}$ satisfy the *determining equations*

$$\xi_t - c^2(x)\tau_x = 0; \tag{2.2a}$$

$$c(x)[\tau_t - \xi_x] + c'(x)\xi = 0; \tag{2.2b}$$

$$\tau_{tt} - 2f_t - c^2(x)\tau_{xx} = 0; \tag{2.2c}$$

$$\xi_{tt} + c^2(x)[2f_x - \xi_{xx}] = 0; \tag{2.2d}$$

$$f_{tt} - c^2(x)f_{xx} = 0. \tag{2.2e}$$

Solving (2.2a) for τ_x and (2.2b) for τ_t and setting $\tau_{tx} = \tau_{xt}$, one finds that

$$\xi_{xx} - (1/c^2)\xi_{tt} - [H(x)\xi]_x = 0, \tag{2.3}$$

where $H(x) = c'/c$.

The solution of (2.3), (2.2c), and (2.2d) for f leads to

$$f = \tfrac{1}{2}H\xi + s, \quad s = \text{const}. \tag{2.4}$$

Substituting (2.4) into (2.2e), one obtains

$$[(2H' + H^2)\xi^2]_x = 0. \tag{2.5}$$

From Eq. (2.5) there follow three cases.

Case I: $2H' + H^2 = 0$

In this case

$$c(x) = (Ax + B)^2, \tag{2.6}$$

where A and B are arbitrary constants. It is easy to show that here an infinite group leaves invariant (1.1). In particular for *any* solution $\xi(x,t)$ of the corresponding equation (2.3), one can find $\{\tau(x,t), f(x,t)\}$ solving (2.2a)–(2.2e),

$$\tau = \int(\xi_x - H\xi)dt, \quad f = \frac{A\xi(x,t)}{Ax + B}. \tag{2.7}$$

Case II: $2H' + H^2 \neq 0$, $\xi \neq 0$

From (2.5) it follows that ξ can be expressed in the separable form

$$\xi(x,t) = \alpha(x)\beta(t), \tag{2.8}$$

where

$$\alpha^2(x) = [2H' + H^2]^{-1} \tag{2.9}$$

and $\beta(t)$ is to be determined.

Substituting (2.4) and (2.8) into (2.2d), one finds that

$$\frac{\beta''(t)}{\beta(t)} = \frac{c^2(\alpha' - H\alpha)'}{\alpha} = \text{const} = \sigma^2. \tag{2.10}$$

Note that α, σ could be real or imaginary.

Case II(a): The subcase $\sigma = 0$

Here $c(x)$ must satisfy the differential equation

$$(\alpha' - H\alpha)' = 0 \tag{2.11}$$

and correspondingly

$$\beta(t) = p + qt, \tag{2.12}$$

where p and q are arbitrary constants.

The substitution of (2.4) and (2.8) into (2.2e) leads to

$$(\alpha H)'' = 0. \tag{2.13}$$

Thus it is necessary and sufficient that the wave speed $c(x)$ satisfy Eqs. (2.11) and (2.13). The general solution of these equations is

$$\alpha = Bx^2 + Cx + D, \tag{2.14}$$

$$\alpha H = A + 2Bx, \tag{2.15}$$

where $\{A, B, C, D\}$ are arbitrary constants. Consequently

$$c(x) = (Bx^2 + Cx + D) \times \exp\left((A - C)\int(Bx^2 + Cx + D)^{-1}dx\right). \tag{2.16}$$

It is easy to show that

$$\tau = (C - A)\left(pt + \frac{1}{2}qt^2\right) + q\int\frac{\alpha}{c^2}dx + r, \tag{2.17}$$

where r is another arbitrary constant.

If $B = 0$ in Eq. (2.16), then this expression reduces to the general form

$$c(x) = (Ax + B)^C, \tag{2.18}$$

where $\{A, B, C\}$ are arbitrary constants, $C \neq 0, 2$.

If $B = C = 0$ in (2.16), then the corresponding wave speeds are of the general form

$$c(x) = Ae^{Bx}, \tag{2.19}$$

where A and B are arbitrary constants.

Case II(b): The subcase $\sigma \neq 0$

Here Eq. (2.10) leads to $c(x)$ solving

$$c^2(\alpha' - H\alpha)' = \sigma^2\alpha, \tag{2.20}$$

where $H = c'/c$ and α is given by (2.9). Equation (2.20) can be integrated to give

$$(\alpha' - H\alpha)^2 - (\sigma\alpha/c)^2 = \text{const} = K. \tag{2.21}$$

$\beta(t)$ solves $\beta'' = \sigma^2\beta$, i.e., $\beta = pe^{\sigma t} + qe^{-\sigma t}$.

Thus in this subcase the infinitesimals of (2.1) become

$$\begin{aligned}\xi &= \alpha(x)\,[pe^{\sigma t} + qe^{-\sigma t}],\\ \tau &= \sigma^{-1}[\alpha' - H\alpha]\,[pe^{\sigma t} - qe^{-\sigma t}] + r,\\ f &= \tfrac{1}{2}\alpha H\,[pe^{\sigma t} + qe^{-\sigma t}] + s,\end{aligned} \tag{2.22}$$

where the group parameters $\{p, q, r, s\}$ are arbitrary constants. The solution of Eq. (2.9), (2.21) for the wave speed $c(x)$ is given in Appendix A. In Case II, if $\xi \neq 0$, the wave equation (1.1) is invariant under a four-parameter Lie group of point transformations.

Case III: $\xi = 0$

From the determining equations (2.2a)–(2.2e) it follows immediately that

$$\tau = \text{const} = r, \quad f = \text{const} = s,$$

and hence (1.1) is invariant only under translations in t and scalings of u. In particular for any wave speed $c(x)$ that does not solve the system (2.9), (2.20) for any σ (zero or nonzero), the wave equation (1.1) is invariant only under this trivial two-parameter Lie group of point transformations.

Hence the following theorem has been proved.

Theorem: The wave equation (1.1), whose wave speed $c(x)$ is a solution of system (2.9), (2.21) for any σ (zero or nonzero), is invariant under a four-parameter Lie group of point transformations. The group becomes infinite if and only if $c(x) = (Ax + B)^2$. All other wave speeds $c(x)$ admit the two-parameter group of translations in t and scalings of u.

B. Group generators and their Lie algebras in the finite parameter cases

If (2.1) leaves invariant (1.1), the corresponding group generator is

$$L = \xi(x,t)\frac{\partial}{\partial x} + \tau(x,t)\frac{\partial}{\partial t} + f(x,t)u\frac{\partial}{\partial u}. \tag{2.23}$$

To the parameters $\{p, q, r, s\}$ of the group there correspond generators $\{L_p, L_q, L_r, L_s\}$. The generators form a Lie algebra. The generators for all possible wave speeds $c(x)$ follow. Cases (i)–(iv) relate to $\sigma = 0$.

Case (i): $c(x) = (Bx^2 + Cx + D)\exp[(A - C)\int (Bx^2 + Cx + D)^{-1}dx]$

Here

$$\begin{aligned}L_p &= [Bx^2 + Cx + D]\frac{\partial}{\partial x}\\ &\quad + [C - A]t\frac{\partial}{\partial t} + \frac{1}{2}[A + 2Bx]u\frac{\partial}{\partial u},\\ L_q &= t[Bx^2 + Cx + D]\frac{\partial}{\partial x}\\ &\quad + \left[\frac{1}{2}(C - A)t^2 + \int\frac{Bx^2 + Cx + D}{c^2(x)}dx\right]\frac{\partial}{\partial t}\\ &\quad + \frac{1}{2}t[A + 2Bx]u\frac{\partial}{\partial u},\\ L_r &= \frac{\partial}{\partial t}, \quad L_s = u\frac{\partial}{\partial u}.\end{aligned} \tag{2.24}$$

The commutator table for the Lie algebra is

$$\begin{aligned}&[L_p, L_q] = (C - A)L_q; \quad [L_p, L_r] = (A - C)L_r;\\ &[L_q, L_r] = -L_p; \quad \begin{bmatrix}L_p, L_s\\ q\\ r\end{bmatrix} = 0.\end{aligned} \tag{2.25}$$

It is easy to show that this group is isomorphic to SO(2,1) when $A - C \neq 0$. An interesting special case is $A = C$ where $c(x) = Bx^2 + Cx + D$.

Case (ii): $c(x) = (Ax + B)^C$, $C \neq 0,1,2$

Here

$$\begin{aligned}L_p &= (Ax + B)\frac{\partial}{\partial x} + A(1 - C)t\frac{\partial}{\partial t} + \frac{1}{2}ACu\frac{\partial}{\partial u},\\ L_q &= (Ax + B)t\frac{\partial}{\partial x}\\ &\quad + \frac{1}{2}\left[A(1 - C)t^2 + \frac{(Ax + B)^{2-2C}}{A(1 - C)}\right]\frac{\partial}{\partial t}\\ &\quad + \frac{1}{2}ACtu\frac{\partial}{\partial u},\\ L_r &= \frac{\partial}{\partial t}, \quad L_s = u\frac{\partial}{\partial u}.\end{aligned} \tag{2.26}$$

The commutator table for the Lie algebra is the same as (2.25) with $(C - A)$ replaced by $A(1 - C)$.

Case (iii): $c(x) = Ax + B$

Here

$$\begin{aligned}L_p &= (Ax + B)\frac{\partial}{\partial x} + \frac{1}{2}Au\frac{\partial}{\partial u},\\ L_q &= (Ax + B)t\frac{\partial}{\partial x} + \left[\frac{1}{A}\log(Ax + B)\right]\frac{\partial}{\partial t}\\ &\quad + \frac{1}{2}Atu\frac{\partial}{\partial u},\\ L_r &= \frac{\partial}{\partial t}, \quad L_s = u\frac{\partial}{\partial u}.\end{aligned} \tag{2.27}$$

The corresponding commutator table is

$$\begin{aligned}&[L_p, L_q] = L_r; \quad [L_p, L_r] = 0;\\ &[L_q, L_r] = -L_p; \quad \begin{bmatrix}L_p, L_s\\ q\\ r\end{bmatrix} = 0.\end{aligned}$$

Case (iv): $c(x) = Ae^{Bx}$

Here

$$\begin{aligned}L_p &= A\frac{\partial}{\partial x} - ABt\frac{\partial}{\partial t} + \frac{1}{2}ABu\frac{\partial}{\partial u},\\ L_q &= At\frac{\partial}{\partial x} - \frac{1}{2}\left[ABt^2 + \frac{1}{AB}e^{-2Bx}\right]\frac{\partial}{\partial t}\\ &\quad + \frac{1}{2}ABtu\frac{\partial}{\partial u},\\ L_r &= \frac{\partial}{\partial t}, \quad L_s = u\frac{\partial}{\partial u}.\end{aligned} \tag{2.28}$$

The commutator table is the same as (2.25) with $A - C$ replaced by AB.

Cases (ii)–(iv) can result as limiting cases for the constants $\{A, B, C, D\}$ of case (i).

Case (v): $c(x)$ for $\sigma \neq 0$

From (2.22)

$$L_p = e^{\sigma t}\left[\alpha \frac{\partial}{\partial x} + \sigma^{-1}(\alpha' - H\alpha)\frac{\partial}{\partial t} + \frac{1}{2}\alpha H u \frac{\partial}{\partial u}\right],$$
$$L_q = e^{-\sigma t}\left[\alpha \frac{\partial}{\partial x} - \sigma^{-1}(\alpha' - H\alpha)\frac{\partial}{\partial t} + \frac{1}{2}\alpha H u \frac{\partial}{\partial u}\right], \tag{2.29}$$
$$L_r = \frac{\partial}{\partial t}, \quad L_s = u\frac{\partial}{\partial u}.$$

The corresponding commutator table is

$$[L_p, L_q] = 2\sigma^{-1}KL_r; \quad [L_p, L_r] = -\sigma L_p;$$
$$[L_q, L_r] = \sigma L_q; \quad \begin{bmatrix} L_p, L_s \\ q \quad r \end{bmatrix} = 0. \tag{2.30}$$

Recall that K is given by (2.21).

Clearly this group is isomorphic to SO(2,1) when $K \neq 0$. When σ is imaginary, appropriate linear combinations of L_p and L_q will yield the corresponding real Lie algebra.

Case (iv): All other $c(x)$

Here the generators are only

$$L_r = \frac{\partial}{\partial t}; \quad L_s = u\frac{\partial}{\partial u}. \tag{2.31}$$

C. The infinite group case: $c(x)=(Ax+B)^2$

In this case the wave equation (1.1) becomes

$$(Ax + B)^2 u_{xx} - u_{tt} = 0. \tag{2.32}$$

Equation (2.32) can be mapped into the wave equation ($A \neq 0$)

$$U_{XT} = 0 \tag{2.33}$$

by the transformation[4]

$$X = [1/(Ax + B)] + At,$$
$$T = [1/(Ax + B)] - At, \tag{2.34}$$
$$U = (Ax + B)^{-1}u.$$

Hence the general solution of (2.32) is

$$u = (Ax + B)[F(X) + G(T)], \tag{2.35}$$

where F and G are arbitrary twice differentiable functions of their respective arguments.

D. Similarity solutions of the wave equation (1.1)

A *similarity solution* (invariant solution)[2,3] of (1.1) is a solution $u = \theta(x,t)$ of (1.1) satisfying the characteristic equations

$$\frac{dx}{\xi(x,t)} = \frac{dt}{\tau(x,t)} = \frac{du}{f(x,t)u}, \tag{2.36}$$

corresponding to an admitted group (2.1). The *similarity variable* $z(x,t)$ is the constant of integration of the first equality of (2.36).

For all of our cases, similarity solutions for $r \neq 0$ can always be obtained from similarity solutions for $r = 0$ by replacing t by $t + r$. For the cases where $\sigma = 0$, the class of similarity solutions for $\{q = 1,\ r$ arbitrary, s arbitrary, $p = 0\}$ is identical to the class of similarity solutions for $\{q = 1, p, r, s,$ arbitrary$\}$ since the commutator of L_q with L_r generates L_p. Next we discuss similarity solutions of (1.1) keeping in mind the above remarks.

Case (i): Similarity solutions of (1.1) for $p = q = 0$, $r = 1$, s arbitrary

Here (2.36) becomes

$$\frac{dx}{0} = \frac{dt}{1} = \frac{du}{su}. \tag{2.37}$$

The similarity variable $z = x$, and the similarity form for the similarity solutions is

$$u = e^{st}F(x;s), \tag{2.38}$$

where $F(x;s)$ is a function of x and the parameter s. Substituting (2.38) into (1.1), one find that $F(x;s)$ satisfies the ordinary differential equation (ODE)

$$c^2(x)F_{xx}(x;s) - s^2F(x;s) = 0, \tag{2.39}$$

If $\{F_1(x;s), F_2(x;s)\}$ are linearly independent solutions of (2.39) for any s, then any linear superposition

$$u = \sum_s e^{st}[A_1(s)F_1(x;s) + A_2(s)F_2(x;s)] \tag{2.40}$$

solves (1.1) for arbitrary $\{A_1(s), A_2(s)\}$. Note that the sum in (2.40) can be replaced by an integral with respect to s.

Now we consider all cases for invariance of (1.1) under a four-parameter group. The following cases (ii)–(v) correspond to $\sigma = 0$ in Eq. (2.10).

Case (ii): $c(x) = x^C$, $C \neq 0,1,2$

The substitutions $Ax + B \to x$, $t \to A^{-1}t$, make the PDE

$$(Ax + B)^{2C}u_{xx} - u_{tt} = 0 \tag{2.41}$$

equivalent to the PDE

$$x^{2C}u_{xx} - u_{tt} = 0. \tag{2.42}$$

1. Similarity solutions of (2.42) for $q=r=0$, $p=1$, s arbitrary

Here (2.36) becomes equivalently

$$\frac{dx}{x} = \frac{dt}{(1 - C)t} = \frac{du}{su}. \tag{2.43}$$

The similarity variable is

$$z = x^{C-1}t. \tag{2.44}$$

The similarity form for the solutions is

$$u = x^sF(z;s). \tag{2.45}$$

$F(z;s)$ satisfies the ODE

$$[1 - (C - 1)^2z^2]F_{zz}(z;s) + (1 - C)(s + C - 1)zF_z(z;s) + s(1 - s)F(z;s) = 0. \tag{2.46}$$

Linearly independent solutions of (2.46) are

$$F_1(z;s) = F(\alpha,\beta;\gamma;\zeta)$$

and (2.47)

$$F_2(z;s) = \zeta^{1-C}F(1 + \alpha - \gamma, 1 + \beta - \gamma; 2 - \gamma;\zeta),$$

where $F(\alpha,\beta;\gamma;\zeta)$ is the hypergeometric function,

$$\alpha=\frac{s}{1-C},\quad \beta=\frac{s-1}{1-C},\quad \gamma=\frac{1}{2}\frac{(2s+C-2)}{C-1},$$
$$\zeta=\frac{1}{2}+\frac{1}{2}(C-1)z. \tag{2.48}$$

2. Similarity solutions of (2.42) for p=r=0, q=1, s arbitrary

In this case (2.36) is equivalent to

$$\frac{dx}{2tx}=\frac{dt}{[x^{2-2C}/(1-C)]+(1-C)t^2}=\frac{du}{(Ct+[s/(C-1)])u}. \tag{2.49}$$

The similarity variable is

$$z=(C-1)^2t^2x^{C-1}-x^{1-C}. \tag{2.50}$$

The similarity solutions are of the form

$$u=x^{C/2}e^{sx^{C-1}t/z}F(z;s). \tag{2.51}$$

$F(z;s)$ satisfies the ODE

$$4(C-1)^2[z^2F_{zz}(z;s)+2zF_z(z;s)]+[C(C-2)-4sz^{-2}]F(z;s)=0. \tag{2.52}$$

If $[1/(C-1)]\neq$integer, then linearly independent solutions of (2.52) are

$$F(z;s)=z^{-1/2}I_{\pm\nu}(\zeta), \tag{2.53}$$

where $I_\nu(\zeta)$ is a modified Bessel function of order ν,

$$\nu=\frac{1}{2(C-1)},\quad \zeta=\frac{sz^{-1}}{C-1}. \tag{2.54}$$

Case (iii): $c(x)=x$

Here we consider the PDE

$$(Ax+B)^2u_{xx}-u_{tt}=0 \tag{2.55}$$

equivalent to the PDE

$$x^2u_{xx}-u_{tt}=0. \tag{2.56}$$

3. Similarity solutions of (2.56) for q=r=0, p=1, s arbitrary

The characteristic equations (2.36) are equivalently

$$\frac{dx}{x}=\frac{dt}{0}=\frac{du}{su}.$$

The similarity variable is

$$z=t \tag{2.57}$$

with corresponding similarity form

$$u=x^sF(t;s). \tag{2.58}$$

$F(t;s)$ satisfies the ODE

$$F_{tt}(t;s)+s(1-s)F(t;s)=0. \tag{2.59}$$

The resulting superposition of similarity solutions is

$$u(x,t)=\sum_s x^s[A_1(s)e^{\sqrt{s(s-1)}t}+A_2(s)e^{-\sqrt{s(s-1)}t}]. \tag{2.60}$$

These solutions are of the form (2.40).

4. Similarity solutions of (2.56) for p=r=0, q=1, s arbitrary

Now (2.36) is equivalently

$$\frac{dx}{2tx}=\frac{dt}{2\log x}=\frac{du}{(t+2s)u}. \tag{2.61}$$

The similarity variable is

$$z=t^2-(\log x)^2. \tag{2.62}$$

The corresponding form of the similarity solutions is

$$u=x^{1/2}|\log x+t|^sF(z;s). \tag{2.63}$$

$F(z;s)$ satisfies the ODE

$$16z^2F_{zz}(z;s)+16(1+s)zF_z(z;s)-zF(z;s)=0. \tag{2.64}$$

If $2s\neq$integer, linearly indepedent solutions of (2.64) are

$$F(z;s)=z^{s/2}I_{\pm\nu}(\zeta), \tag{2.65}$$

where

$$\nu=s,\quad \zeta=\tfrac{1}{4}z^{1/2}. \tag{2.66}$$

Case (iv): $c(x)=e^{-x/2}$

The substitutions $x\to -x/2B$, $t\to t/2AB$, make the PDE

$$A^2e^{2Bx}u_{xx}-u_{tt}=0 \tag{2.67}$$

equivalent to the PDE

$$e^{-x}u_{xx}-u_{tt}=0. \tag{2.68}$$

5. Similarity solutions of (2.68) for q=r=0, p=1, s arbitrary

The characteristic equations (2.36) are equivalent to

$$\frac{dx}{2}=\frac{dt}{t}=\frac{du}{2su}. \tag{2.69}$$

The similarity variable is

$$z=te^{-x/2}. \tag{2.70}$$

The similarity solutions are of the form

$$u=e^{sx}F(z;s). \tag{2.71}$$

$F(z;s)$ satisfies the ODE

$$(4-z^2)F_{zz}(z;s)+(4s-1)zF_z(z;s)-4s^2F(z;s)=0. \tag{2.72}$$

Linearly independent solutions of (2.72) are of the hypergeometric form (2.47), where

$$\alpha=\beta=-2s,\quad \gamma=\tfrac{1}{2}(1-4s), \tag{2.73}$$

and

$$\zeta=\tfrac{1}{2}+\tfrac{1}{4}z. \tag{2.74}$$

6. Similarity solutions of (2.68) for p=r=0, q=1, s arbitrary

Now (2.36) is equivalent to

$$\frac{dx}{4t}=\frac{dt}{t^2+4e^x}=\frac{du}{(s-t)u}. \tag{2.75}$$

The similarity variable is

$$z=t^2e^{-x/2}-4e^{x/2}. \tag{2.76}$$

The corresponding similarity solutions are of the form

$$u = \exp(-[\tfrac{1}{4}x + stz^{-1}e^{-x/2}])F(z;s). \tag{2.77}$$

$F(z;s)$ satisfies the ODE

$$4z^2F_{zz}(z;s) + 8zF_z(z;s) + (1 - 16s^2z^{-2})F(z;s) = 0. \tag{2.78}$$

This equation has linearly independent solutions

$$F_1(z;s) = z^{-1/2}I_0(\zeta), \quad F_2(z;s) = z^{-1/2}K_0(\zeta), \tag{2.79}$$

where $\{I(\zeta), K(\zeta)\}$ are modified Bessel functions of order 0, and

$$\zeta = 2sz^{-1}. \tag{2.80}$$

Case (v): $c(x) = (Bx^2 + Cx + D)\exp[(A - C)\int(Bx^2 + Cx + D)^{-1}dx]$

By appropriate scalings and translations in x and scalings in t, the corresponding wave equation (1.1) is equivalent to one of the five canonical forms (2.42), (2.68), or

$$[(x^2 + 1)^2 e^{4A \arctan x}]u_{xx} - u_{tt} = 0, \tag{2.81}$$

$$[(1 - x)^{2+2A}(1 + x)^{2-2A}]u_{xx} - u_{tt} = 0, \tag{2.82}$$

$$[x^4 e^{2/x}]u_{xx} - u_{tt} = 0. \tag{2.83}$$

In Eqs. (2.81), (2.82), A is an arbitrary constant.

Case (va): $c(x) = (x^2 + 1)e^{2A \arctan x}$

7. Similarity solutions of (2.81) for q=r=0, p=1, s arbitrary

The characteristic equations (2.36) are

$$\frac{dx}{1 + x^2} = \frac{dt}{-2At} = \frac{du}{(x + s)u}. \tag{2.84}$$

The similarity variable is

$$z = te^{2Ay}, \tag{2.85}$$

where

$$y = \arctan x. \tag{2.86}$$

The corresponding similarity form is

$$u = \sqrt{1 + x^2}e^{sy}F(z;s). \tag{2.87}$$

$F(z;s)$ solves the ODE

$$(4A^2z^2 - 1)F_{zz}(z;s) + 4A(A + s)zF_z(z;s) + (1 + s^2)F(z;s) = 0 \tag{2.88}$$

whose general solution can be expressed in terms of hypergeometric functions.

In the special case $A = 0$, the resulting superposition of similarity solutions is

$$u(x,t) = \sqrt{x^2 + 1}\sum_s e^{sy}[A_1(s)e^{\sqrt{s^2+1}\,t} + A_2(s)e^{-\sqrt{s^2+1}\,t}]. \tag{2.89}$$

These solutions are of the form (2.40).

8. Similarity solutions of (2.81) for p=0, r=1/4A, q=1, s arbitrary

The characteristic equations are

$$\frac{dx}{t(1 + x^2)} = \frac{4A\,dt}{-4A^2t^2 - e^{-4Ay} + 1} = \frac{du}{[t(A + x) + s]u}. \tag{2.90}$$

The similarity variable is

$$z = 2A^2t^2e^{2Ay} - \cosh 2Ay, \tag{2.91}$$

where

$$y = \arctan x. \tag{2.92}$$

The resulting similarity form is

$$u = \sqrt{1 + x^2}e^{Ay}|z + e^{2Ay}(1 + 2At)|^s F(z;s). \tag{2.93}$$

$F(z;s)$ satisfies the ODE

$$4A^2(z^2 - 1)F_{zz}(z;s) + 8A^2(1 + s)zF_z(z;s) + \{1 + [A(1 + 2s)]^2\}F(z;s) = 0. \tag{2.94}$$

Linearly independent solutions of (2.94) are of the hypergeometric form (2.47), where

$$\alpha = \frac{1}{2} + s + \frac{i}{2A}, \quad \beta = \frac{1}{2} + s - \frac{i}{2A}, \tag{2.95}$$

$$\gamma = 1 + s, \quad \zeta = \tfrac{1}{2}(1 + z).$$

In the special case $A = 0$, the similarity variable becomes

$$z = -t^2 + y^2. \tag{2.96}$$

Here the similarity form reduces to

$$u = \sqrt{x^2 + 1}(t + \arctan x)^s F(z;s). \tag{2.97}$$

$F(z;s)$ satisfies the ODE

$$4zF_{zz}(z;s) + 4(s + 1)F_z(z;s) + F(z;s) = 0. \tag{2.98}$$

Solutions of (2.98) can be expressed in terms of Bessel functions:

$$F(z;s) = z^{-s/2}J_{\pm\nu}(\zeta), \tag{2.99}$$

where

$$\nu = s, \quad \zeta = z^{-1/2}. \tag{2.100}$$

Case (vb): $c(x) = (1 - x)^{1+A}(1 + x)^{1-A}$

9. Similarity solutions of (2.82) for q=r=0, p=1, s arbitrary

The characteristic equations are equivalent to

$$\frac{dx}{x^2 - 1} = \frac{dt}{-2At} = \frac{du}{(x + 2s)u}. \tag{2.101}$$

The similarity variable is

$$z = ty^A, \tag{2.102}$$

where

$$y = (1 - x)/(1 + x). \tag{2.103}$$

The similarity form is

$$u = \sqrt{1 - x^2}y^s F(z;s). \tag{2.104}$$

$F(z;s)$ satisfies the ODE

$$(4A^2z^2-1)F_{zz}(z;s)+4A(A+2s)zF_z(z;s)+(4s^2-1)F(z;s)=0. \tag{2.105}$$

Linearly independent solutions of (2.105) are of the hypergeometric form (2.47), where

$$\alpha=\frac{1}{A}\left[2s-\frac{1}{2}\right],\quad \beta=\frac{1}{2}A,\quad \gamma=\frac{1}{2}+\frac{s}{A},\quad \zeta=\frac{1}{2}+Az. \tag{2.106}$$

In the special case $A=0$, the resulting superposition of similarity solutions, which is of the form (2.40), is

$$u=\sqrt{1-x^2}\sum_s y^s\left[A_1(s)e^{\sqrt{4s^2-1}\,t}+A_2(s)e^{-\sqrt{4s^2-1}\,t}\right]. \tag{2.107}$$

10. Similarity solutions of (2.82) for p=0, r=1/A, q=1, s arbitrary

Here the characteristic equations are

$$\frac{dx}{(x^2-1)t}=\frac{4A\,dt}{1-4A^2t^2-y^{-2A}}=\frac{du}{[(A+x)t+s]u}. \tag{2.108}$$

The similarity variable is

$$z=2A^2t^2y^A-\tfrac{1}{2}(y^A+y^{-A}), \tag{2.109}$$

where

$$y=(1-x)/(1+x). \tag{2.110}$$

The resulting similarity solutions are of the form

$$u=\sqrt{1-x^2}y^{A/2}|(1+2At)y^A+z|^sF(z;s). \tag{2.111}$$

$F(z;s)$ satisfies the ODE

$$4A^2(z^2-1)F_{zz}(z;s)+8A^2(s+1)zF_z(z;s)+[A^2(2s+1)^2-1]F(z;s)=0. \tag{2.112}$$

Linearly independent solutions of (2.112) are of the hypergeometric form (2.47), where

$$\alpha=s+\frac{1}{2}+\frac{1}{2A},\quad \beta=s+\frac{1}{2}-\frac{1}{2A},\quad \gamma=s+1,\quad \zeta=\tfrac{1}{2}(z+1). \tag{2.113}$$

Case (vc): $c(x)=x^2e^{1/x}$

11. Similarity solutions of (2.83) for q=r=0, p=1, s arbitrary

The characteristic equations are equivalent to

$$\frac{dx}{x^2}=\frac{dt}{t}=\frac{du}{(x-s)u}. \tag{2.114}$$

The similarity variable is

$$z=te^{1/x} \tag{2.115}$$

and the corresponding similarity form is

$$u=xe^{s/x}F(z;s). \tag{2.116}$$

$F(z;s)$ solves the ODE

$$(z^2-1)F_{zz}(z;s)+(2s+1)zF_z(z;s)+s^2F(z;s)=0. \tag{2.117}$$

Linearly independent solutions of (2.117) are of the hypergeometric form (2.47) where

$$\alpha=\beta=s,\quad \gamma=\tfrac{1}{2}+s,\quad \zeta=\tfrac{1}{2}(1+z). \tag{2.118}$$

12. Similarity solutions of (2.83) for p=r=0, q=1, s arbitrary

Here the characteristic equations are

$$\frac{dx}{2tx^2}=\frac{dt}{t^2+e^{-2/x}}=\frac{du}{[(2x-1)t+2s]u}. \tag{2.119}$$

The similarity variable is

$$z=t^2e^{1/x}-e^{-1/x}, \tag{2.120}$$

and the resulting similarity form is

$$u=xe^{1/2x}e^{-2ste^{1/x}z^{-1}}F(z;s). \tag{2.121}$$

$F(z;s)$ satisfies the ODE

$$4z^2F_{zz}(z;s)+8zF_z(z;s)+(1-16s^2z^{-2})F(z;s)=0. \tag{2.122}$$

Linearly independent solutions of (2.122) can be expressed in terms of the modified Bessel functions:

$$F_1(z;s)=z^{-1/2}I_0(\zeta),\quad F_2(z;s)=z^{-1/2}K_0(\zeta), \tag{2.123}$$

where

$$\zeta=2sz^{-1}. \tag{2.124}$$

Case (iv): $c(x)$ for $\sigma\neq 0$

The corresponding characteristic equations are

$$\frac{dx}{2\alpha(x)\beta(t)}=\frac{\sigma^2dt}{2(\alpha'-H\alpha)\beta'(t)}=\frac{du}{[\alpha H\beta(t)+s]u}, \tag{2.125}$$

where

$$\beta(t)=pe^{\sigma t}+qe^{-\sigma t}, \tag{2.126}$$

and $\alpha(x)$, $H=c'/c$ satisfy Eqs. (2.9) and (2.21). The similarity variable is

$$z=(c/\alpha)(pe^{\sigma t}-qe^{-\sigma t}). \tag{2.127}$$

The corresponding form for the similarity solutions is

$$u=\sqrt{c}\left[\frac{w}{\beta\sqrt{K}+2\sqrt{pq(K+\sigma^2w^2)}}\right]^\rho F(z;s), \tag{2.128}$$

where

$$w=\alpha/c,\quad \rho=s/4\sqrt{pqK}. \tag{2.129}$$

$F(z;s)$ satisfies the ODE

$$(K^2z^2-4pq\sigma^2)F_{zz}(z;s)+2K(1-\rho)zF_z(z;s)+\{\tfrac{1}{4}+\rho(\rho-1)K\}F(z;s)=0. \tag{2.130}$$

III. THE INVARIANCE PROPERTIES OF THE SYSTEM

Clearly (1.2) is always invariant under translations in t and uniform scalings of u and v.

If the point transformation

$$\begin{aligned}X&=x+\epsilon\xi(x,t)+O(\epsilon^2),\\ T&=t+\epsilon\tau(x,t)+O(\epsilon^2),\\ U&=u+\epsilon[f(x,t)u+g(x,t)v]+O(\epsilon^2),\\ V&=v+\epsilon[k(x,t)v+l(x,t)u]+O(\epsilon^2),\end{aligned} \tag{3.1}$$

leaves invariant (1.2), then $\{\xi,\tau,f,g,k,l\}$ satisfy determining equations which reduce to

$$k_t - g_x = 0, \tag{3.2a}$$
$$l_t - f_x = 0, \tag{3.2b}$$
$$c^2(x)l - g = 0, \tag{3.2c}$$
$$c^2(x)\tau_x - \xi_t = 0, \tag{3.2d}$$
$$c^2(x)k_x - g_t = 0, \tag{3.2e}$$
$$c^2(x)l_x - f_t = 0, \tag{3.2f}$$
$$c(x)[\tau_t - \xi_x] + c'(x)\xi = 0, \tag{3.2g}$$
$$\xi_x - \tau_t + k - f = 0. \tag{3.2h}$$

The consistency of Eqs. (3.2b), (3.2c), and (3.2f) leads to $g(x,t)$ satisfying

$$g_x H + gH' = 0, \tag{3.3}$$

where $H = c'/c$. Then $g(x,t)$ satisfies

$$g(x,t) = -a(t)/2H. \tag{3.4}$$

Moreover if $a(t) \neq 0$, then it is necessary that $\{c(x), a(t)\}$ satisfy

$$cc'(c/c')'' = a''(t)/a(t) = \text{const} = \lambda^2. \tag{3.5}$$

If $a(t) = 0$, then either $c(x)$ solves (3.5) with $\lambda = 0$ or (1.2) is only invariant under above-mentioned scalings of u and v and translations in t.

In the following subsections we will show that system (1.2) is invariant under a four-parameter Lie group of point transformations of the form (3.1) if and only if $c(x)$ satisfies the ODE (3.5), namely,

$$cc'(c/c')'' = \lambda^2. \tag{3.6}$$

The general solution of (3.6) is derived in Appendix B. It turns out that if $\lambda \neq 0$, the general solution of (3.6) does not solve (2.9), (2.21). Note that λ can be real or imaginary. The case $\lambda = 0$ will be considered in the following subsection and the case $\lambda \neq 0$ in Sec. III B.

A. The case $\lambda = 0$

The general solution of

$$(c/c')'' = 0 \tag{3.7}$$

leads to the consideration of three separate subcases.

Case (i): $c(x) = (Ax + B)^C$, $C \neq 0,1$

The same subsitutions that reduced (2.41) to (2.42) lead here to the equivalent system

$$v_t = u_x, \quad u_t = x^{2C} v_x. \tag{3.8}$$

The solution of the determining equations (3.2a)–(3.2h) leads to

$$\begin{aligned}
&\xi = px + 2qxt,\\
&\tau = p(1-C)t + q[(1-C)t^2 + x^{2-2C}/(1-C)] + r,\\
&f = q(2C-1)t + s,\\
&g = -qx,\\
&k = -pC - qt + s,\\
&l = -qx^{1-2C},
\end{aligned} \tag{3.9}$$

where p, q, r, and s are arbitrary constants.

1. Similarity solutions of (3.8) for $q=r=0$, $p=1$, s arbitrary

The corresponding characteristic equations are

$$\frac{dx}{x} = \frac{dt}{(1-C)t} = \frac{du}{su} = \frac{dv}{(s-C)v}. \tag{3.10}$$

Comparing (2.43) and (3.10), one sees that the similarity solutions for u are of the form (2.44), (2.45). The corresponding solutions for v are of the form

$$v = x^{s-C} G(z;s). \tag{3.11}$$

Substituting (2.45) and (3.11) into the system (3.8), one finds that

$$\begin{aligned}
&G_z(z;s) = sF(z;s) + (C-1)zF_z(z;s),\\
&(s-C)G(z;s) + (C-1)zG_z(z;s) = F_z(z;s).
\end{aligned} \tag{3.12}$$

If one eliminates $G(z;s)$ from (3.12), then $F(z;s)$ solves (2.46). Correspondingly

$$G(z;s) = \frac{[1-(C-1)^2z^2]F_z(z;s) + (1-C)szF(z;s)}{s-C}. \tag{3.13}$$

2. Similarity solutions of (3.8) for $p=r=0$, $q=1$, s arbitrary

First we find the global transformation (3.1) corresponding to (3.9) for $p = r = s = 0$, $q = 1$. Then it is easy to obtain the global transformation for arbitrary s. This global transformation leads to the similarity form of the solutions.

The global transformation for $p = r = s = 0$, $q = 1$, is found by solving the characteristic differential equations

$$\begin{aligned}
\frac{dX}{2XT} &= \frac{dT}{(1-C)T^2 + X^{2-2C}/(1-C)}\\
&= \frac{dU}{(2C-1)TU - XV}\\
&= \frac{dV}{-[TV + X^{1-2C}U]} = d\epsilon,
\end{aligned} \tag{3.14}$$

where $X = x$, $T = t$, $U = u$, $V = v$, at $\epsilon = 0$.

The first equality in (3.14) leads to

$$\begin{aligned}
&(1-C)T^2X^{C-1} - \frac{1}{(1-C)X^{C-1}}\\
&\quad = \text{const} = (1-C)t^2x^{C-1} - \frac{1}{(1-C)x^{C-1}} = z.
\end{aligned} \tag{3.15}$$

Next we consider the differential equations

$$\frac{dU}{d\epsilon} = (2C-1)TU - XV; \tag{3.16}$$

$$\frac{dV}{d\epsilon} = -TV - X^{1-2C}U. \tag{3.17}$$

One can show that

$$\frac{d^2V}{d\epsilon^2} = -2CT\frac{dV}{d\epsilon} + \left[\frac{C}{C-1}X^{2-2C} - CT^2\right]V. \tag{3.18}$$

Let $V = X^{-C/2}W$. Then (3.18) reduces to

$$\frac{d^2W}{d\epsilon^2} = 0. \tag{3.19}$$

Hence

$$V = X^{-C/2}(F\epsilon + G), \tag{3.20}$$

where F and G are constants. Equation (3.17) leads to

$$U = X^{3C/2-1}[(C-1)T(F\epsilon + G) - F]. \tag{3.21}$$

The solution of

$$\frac{dX}{2XT} = d\epsilon \tag{3.22}$$

leads to

$$[1 + z(1-C)X^{C-1}]^{1/2} = z(C-1)(\epsilon + E), \tag{3.23}$$

where E is a constant.

The global transformation for arbitrary s, $p = r = 0$, $q = 1$, follows:

$$\begin{aligned}
&[1 + z(1-C)X^{C-1}]^{1/2} = z(C-1)(\epsilon + E),\\
&(1-C)T^2X^{C-1} - \frac{1}{(1-C)X^{C-1}} = z,\\
&U = e^{s\epsilon}X^{3C/2-1}[(C-1)T(F\epsilon + G) - F],\\
&V = e^{s\epsilon}X^{-C/2}(F\epsilon + G),
\end{aligned} \tag{3.24}$$

where the constants $\{z, E, F, G\}$ can be expressed in terms of $\{x, t, u, v\}$ by solving (3.24) at $\epsilon = 0$. The explicit form of the global transformation is easily found by solving (3.24) for $\{X, T, U, V\}$.

The corresponding similarity solutions are found by letting z play the role of the similarity variable, and letting $\{E, F, G\}$ be arbitrary functions of z and s. Without loss of generality one can set $E = 0$. Solving the first two equations of (3.24) for ϵ, one then finds that the resulting similarity form is

$$\begin{aligned}
u = {} & e^{stx^{C-1}z^{-1}}x^{3C/2-1}[(C-1)t\{x^{C-1}z^{-1}F(z;s)\\
& + G(z;s)\} - F(z;s)],\\
v = {} & e^{stx^{C-1}z^{-1}}x^{-C/2}[tx^{C-1}z^{-1}F(z;s) + G(z;s)].
\end{aligned} \tag{3.25}$$

If one substitutes (3.25) into the system (3.8) then $F(z;s)$ and $G(z;s)$ satisfy a corresponding system of coupled first-order linear ODE's.

Case (ii): $c(x) = x$

Here the system (1.2) becomes

$$v_t = u_x, \quad u_t = x^2 v_x. \tag{3.26}$$

The solution of the determining equations (3.2a)–(3.2h) leads to

$$\begin{aligned}
&\xi = px + 2qxt, \quad \tau = 2q\log x + r,\\
&f = qt + s, \quad g = -qx,\\
&k = -p - qt + s, \quad l = -qx^{-1}.
\end{aligned} \tag{3.27}$$

3. Similarity solutions of (3.26) for q=r=0, p=1, s arbitrary

The resulting similarity solutions are easily found to be of the form

$$u = x^s F(t;s), \quad v = x^{s-1}G(t;s). \tag{3.28}$$

$F(t;s)$ is any solution of (2.59) and

$$G(t;s) = (s-1)^{-1}F_t(t;s). \tag{3.29}$$

4. Similarity solutions of (3.26) for p=r=0, q=1, s arbitrary

Here the same procedure is followed as in Case (i). The resulting global transformation can be written as

$$\begin{aligned}
&T + \log X = Ee^{2\epsilon},\\
&T^2 - (\log X)^2 = z,\\
&U = e^{2\epsilon s}X^{1/2}(e^{2\epsilon}F + G),\\
&V = e^{2\epsilon s}X^{-1/2}(G - e^{2\epsilon}F).
\end{aligned} \tag{3.30}$$

The resulting similarity form is

$$\begin{aligned}
u &= x^{1/2}|t + \log x|^s[|t + \log x|F(z;s) + G(z;s)],\\
v &= x^{-1/2}|t + \log x|^s[G(z;s) - |t + \log x|F(z;s)],
\end{aligned} \tag{3.31}$$

where $\{F(z;s), G(z;s)\}$ are to be determined by substitution of (3.31) into (3.26).

Case (iii): $c(x) = e^{-x/2}$

Here the system (1.2) is

$$v_t = u_x, \quad u_t = e^{-x}v_x. \tag{3.32}$$

The solution of the determining equations (3.2a)–(3.2h) leads to

$$\begin{aligned}
&\xi = 2p + 4qt, \quad \tau = pt + q(t^2 + 4e^x) + r,\\
&f = -2qt + 2s, \quad g = -2q,\\
&k = p + 2s, \quad l = -2qe^x.
\end{aligned} \tag{3.33}$$

5. Similarity solutions of (3.32) for q=r=0, p=1, s arbitrary

The similarity variable is

$$z = te^{-x/2}. \tag{3.34}$$

The form of the solutions is

$$u = e^{sx}F(z;s), \quad v = e^{(s+1/2)x}G(z;s). \tag{3.35}$$

$F(z;s)$ is any solution of (2.72) and

$$G(z;s) = (2s+1)^{-1}[(2 - \tfrac{1}{2}z^2)F_z(z;s) + szF(z;s)]. \tag{3.36}$$

6. Similarity solutions of (3.32) for p=r=0, q=1, s arbitrary

The resulting global transformation (3.1) can be written as

$$\begin{aligned}
&T^2e^{-x/2} - 4e^{x/2} = z,\\
&2z^{-1}\sqrt{4 + ze^{-x/2}} = \epsilon + E,\\
&U = e^{\epsilon s}e^{-3x/4}[F - \tfrac{1}{2}T(F\epsilon + G)],\\
&V = e^{\epsilon s}e^{x/4}(F\epsilon + G).
\end{aligned} \tag{3.37}$$

The resulting similarity form is

$$\begin{aligned}
u &= -e^{2sz^{-1}te^{-x/2}}e^{-3x/4}[4z^{-1}e^{x/2}F(z;s) + \tfrac{1}{2}tG(z;s)],\\
v &= e^{2sz^{-1}te^{-x/2}}e^{x/4}[2z^{-1}te^{-x/2}F(z;s) + G(z;s)],
\end{aligned} \tag{3.38}$$

where $\{F(z;s), G(z;s)\}$ are determined by substitution of (3.38) into (3.32).

B. The case $\lambda \neq 0$

By appropriate scalings of c and x, Eq. (3.6) reduces to (see Appendix B)

$$c' = \nu^{-1} \sinh(\nu \log c) \tag{3.39}$$

or

$$c' = \nu^{-1} \sin(\nu \log c) \tag{3.40}$$

for $\lambda^2 > 0$. For $\lambda^2 < 0$, Eq. (3.6) reduces to

$$c' = \nu^{-1} \cosh(\nu \log c). \tag{3.41}$$

In Eqs. (3.39)–(3.41), ν is an arbitrary real constant. If $\nu = 1$, then $c(x) = \sqrt{1 + e^x}$ solves (3.39).

In the cases {(3.39), (3.40)}, the solution of the determining equations (3.2a)–(3.2h) leads to

$$\begin{aligned} \xi &= (2c/c')[pe^t + qe^{-t}], \\ \tau &= 2[(c/c')' - 1][pe^t - qe^{-t}] + r, \\ f &= [2 - (c/c')'][pe^t + qe^{-t}] + s, \\ g &= -(c/c')[pe^t - qe^{-t}], \\ k &= -(c/c')'[pe^t + qe^{-t}] + s, \\ l &= -(1/cc')[pe^t - qe^{-t}]. \end{aligned} \tag{3.42}$$

The similarity solutions for wave speeds $c(x)$ satisfying (3.39), (3.40), or (3.41) will be constructed in a future paper.

IV. INVARIANCE PROPERTIES OF THE SINGLE EQUATION VIS-A-VIS THE SYSTEM WHEN $c(x) \neq$ const

The single equation (1.1) is invariant under a four-parameter Lie group of point transformations, $\{p,q,r,s\}$, if and only if $c(x)$ solves Eqs. (2.21) and (2.9). This corresponds to a five-parameter family for $c(x)$.

If

$$c = \Psi(x,\sigma,K) \tag{4.1}$$

is a solution of {(2.21), (2.9)}, it follows from their invariance properties that

$$c = k_1 \Psi(k_3 x + k_2, \sigma, K) \tag{4.2}$$

is the general solution of {(2.21), (2.9)}, where $\{k_1, k_2, k_3\}$ are arbitrary constants.

The system (1.2) is invariant under a four-parameter Lie group of point transformations if and only if $c(x)$ solves Eq. (3.6). This corresponds to a four-parameter family for $c(x)$. If

$$c = \Phi(x,\nu) \tag{4.3}$$

solves (3.39), (3.40), or (3.41) then it follows that

$$c = k_1 \Phi((k_1/\lambda)x + k_2, \nu) \tag{4.4}$$

is the general solution of (3.6) where $\{k_1, k_2, \nu\}$ are arbitrary constants.

One can show that the single equation (1.1) and the system of equations (1.2), for the same $c(x)$, admit a four-parameter Lie group of point transformation if and only if

$$c(x) = (A + Bx)^C, \tag{4.5}$$

or the limiting case

$$c(x) = Ae^{Bx}, \tag{4.6}$$

where $\{A, B, C\}$ are arbitrary constants. However, it could still follow that an invariant solution of (1.2) maps into a noninvariant solution of the wave equation (1.1) under the mapping (1.4). In fact if $c(x)$ is of the form (4.5) or (4.6), an invariant solution of (1.2) maps into an invariant solution of (1.1), under the mapping (1.4), if and only if the invariant solution of (1.2) has $q = 0$.

The group leaving invariant the single equation (1.1) is infinite if and only if

$$c(x) = (A + Bx)^2. \tag{4.7}$$

The group leaving invariant the system (1.2) contains at most four parameters.

Any Lie group of point transformations (3.1), leaving invariant (1.2), can be expressed in the equivalent form

$$\begin{aligned} &X = x, \quad T = t, \\ &U = u + \epsilon\eta(x,t,u,v,u_x,u_t) + O(\epsilon^2), \\ &V = v + \epsilon\zeta(x,t,u,v,u_x,u_t) + O(\epsilon^2), \end{aligned} \tag{4.8}$$

where

$$\eta = f(x,t)u + g(x,t)v - \xi(x,t)u_x - \tau(x,t)u_t, \tag{4.9}$$

$$\zeta = k(x,t)v + l(x,t)u - \tau(x,t)u_x - \xi(x,t)c^{-2}(x)u_t. \tag{4.10}$$

The symmetry (4.8) of the system (1.2) is the symmetry

$$X = x, \quad T = t, \quad U = u + \epsilon\tilde{\eta} + O(\epsilon^2), \tag{4.11}$$

of (1.1), where

$$\tilde{\eta} = \eta(x,t,u,D_t^{-1}u_x,u_x,u_t), \tag{4.12}$$

and D_t^{-1} is the operator inverse to the total derivative operator D_t defined by

$$D_t = \frac{\partial}{\partial t} + u_t\frac{\partial}{\partial u} + u_{tt}\frac{\partial}{\partial u_t} + u_{tx}\frac{\partial}{\partial u_x} + \cdots. \tag{4.13}$$

If η depends explicitly on v, i.e., $g \neq 0$ in (4.9), then accordingly $\tilde{\eta}$ depends explicitly on $D_t^{-1}u_x$ and consequently the resulting transformation is neither a Lie group of point transformations nor more generally a Lie–Bäcklund transformation.[5,6] If the group parameter $q \neq 0$ in (3.9), (3.27), (3.33), and (3.42), then $g \neq 0$. If η is independent of v, i.e., $g = 0$ in (4.9), then the symmetry (4.11) corresponds to a Lie group of point transformations admitted by the wave equation (1.1), and in this case the invariant (similarity) solutions of (1.2) map into invariant solutions of (1.1) under the mapping (1.4).

Conversely, let

$$\begin{aligned} &X = x, \quad T = t, \\ &U = u + \epsilon\eta(x,t,u,u_x,u_t) + O(\epsilon^2), \end{aligned} \tag{4.14}$$

be a Lie group of point transformations, equivalent to (2.1), leaving invariant (1.1). Then

$$\eta = f(x,t)u - \xi(x,t)u_x - \tau(x,t)u_t. \tag{4.15}$$

The corresponding symmetry of (1.2) is

$$\begin{aligned} &X = x, \quad T = t, \quad U = u + \epsilon\eta + O(\epsilon^2), \\ &V = v + \epsilon\zeta + O(\epsilon^2), \end{aligned} \tag{4.16}$$

where ζ satisfies the compatible system of PDE's

$$D_t\zeta = D_x\eta, \quad D_x\zeta = c^{-2}(x)D_t\eta, \tag{4.17}$$

and D_x is the total derivative operator

$$D_x = \frac{\partial}{\partial x} + u_x\frac{\partial}{\partial u} + u_{xx}\frac{\partial}{\partial u_x} + u_{tx}\frac{\partial}{\partial u_t} + \cdots. \tag{4.18}$$

Although (4.17) always has a solution ζ for any η of the form (4.15), ζ cannot necessarily be expressed in terms of $\{x,t,u,v\}$ and the partial derivatives of u. If this is the case the symmetry (4.16) is not a Lie–Backlünd transformation.

V. EQUIVALENCE CLASSES OF THE SINGLE EQUATION

A natural question arises as to whether PDE's of the form (1.1) or (1.2), admitting a four-parameter Lie group of point transformations, are equivalent to each other in the sense that there exists a point transformation mapping one PDE into the other. Lie[1,2] gave a criterion applicable to the single PDE (1.1). When Eq. (1.1) is invariant under a four-parameter Lie group of point transformations, Lie's criterion reduces simply to the following statement.

Wave equations of the form (1.1) admitting a four-parameter group are equivalent if and only if the corresponding wave speeds c(x) have the same value for the integration constant K in Eq. (2.21).

For $\sigma = 0$ and any value of K, $-\infty < K < \infty$, there exists a solution $c(x)$ of system $\{(2.9), (2.21)\}$. As noted previously $\alpha(x)$ can be imaginary. Hence the wave speed $c(x)$ for any $\sigma \neq 0$ is equivalent to some wave speed $c(x)$ for $\sigma = 0$.

For $\sigma = 0$, the following wave speeds $c(x)$ are equivalent, modulo scalings in c and x and translations in x:

(a) $c(x) = x, x^2 + 1, x^2 - 1$;

(b) $c(x) = e^x, x^2e^{1/x}$;

(c) $c(x) = x^C, x^{2-C}, (1-x)^C(1+x)^{2-C}$, for any C.

VI. CONCLUSIONS

In this paper we have given the complete group classification of the wave equation (1.1) and the corresponding system (1.2). We have shown that for a wide class of wave speeds $c(x)$, (1.2) is invariant under a larger group than (1.1). Consequently for such wave speeds, whose canonical equations are (3.39), (3.40), and (3.41), there exist invariant (similarity) solutions of (1.2) which are noninvariant solutions of (1.1).

In a future paper we will discuss some interesting solutions of (1.1) for wave speeds $c(x)$ solving (3.39), (3.40), or (3.41). These include solutions for a class of wave speeds with the following physically significant properties:

(a) $c(x)$ is monotone on $(-\infty, \infty)$;

(b) $\lim_{x\to-\infty} c(x) = A$, $\lim_{x\to+\infty} = B$;

(c) $\max_{x\in\mathbf{R}}|c'(x)| = C$;

(d) $c(0) = D$;

where $\{A,B,C,D\}$ are arbitrary positive constants, provided D is between A and B.

In another future paper we will shown how to use the invariance properties of the system (1.2) to linearize some nonlinear systems of PDE's which cannot be linearized by hodograph transformations, applying procedures outlined in Ref. 7.

APPENDIX A: THE GENERAL SOLUTIONS OF EQS. (2.9) and (2.21)

Here we find the general solution for $c(x)$ of the system

$$(\alpha' - H\alpha)^2 - \sigma^2\alpha^2/c^2 = K, \tag{A1}$$

$$\alpha^2 = (2H' + H^2)^{-1}, \tag{A2}$$

where $H = c'/c$, $\sigma \neq 0$. An integration of Eq. (2.20) to Eq. (A1) resulted from taking the commutator of L_p with L_q in (2.30). Without loss of generality, $\sigma = 1$, by an obvious scaling of c.

First we factor (A1) as

$$(\alpha' - H\alpha + \alpha/c)(\alpha' - H\alpha - \alpha/c) = K. \tag{A3}$$

Now let

$$h(x) = \alpha' - \alpha c'/c + \alpha/c. \tag{A4}$$

Then

$$\alpha' - \alpha(c'/c) - \alpha/c = K/h(x). \tag{A5}$$

Equations (A4) and (A5) lead to

$$c = h/h', \quad \alpha = \tfrac{1}{2}[(h^2 - K)/h']. \tag{A6}$$

Thus the problem of finding $c(x)$ is equivalent to finding $h(x)$ satisfying (A2) which now becomes

$$(h^2 - K)[2h'''h'h^2 - 3(h'')^2h^2 + (h')^4] = (h')^4h^2. \tag{A7}$$

Equation (A7) is invariant under arbitrary scalings and translations in x. Hence[3] one can reduce (A7) to a first-order ODE by choosing corresponding differential invariants

$$u = h, \quad v = h''/(h')^2. \tag{A8}$$

Then (A7) becomes the Riccati equation

$$2\frac{dv}{du} + v^2 + \frac{1}{u^2 - K} + \frac{1}{u^2} = 0. \tag{A9}$$

After v is solved explicitly in terms of u, $v = v(u)$, (A8) becomes

$$h''/h' = v(h)h'. \tag{A10}$$

Thus

$$\log h' = \int v(h)dh + k_1 = -\log M(h), \tag{A11}$$

where k_1 is an arbitrary constant. Then

$$\int M(h)dh = x + k_2, \tag{A12}$$

where k_2 is an arbitrary constant. After solving (A12) for $h(x)$, (A6) leads to

$$c(x) = h(x)M(h(x)). \tag{A13}$$

It should be noted that the transformation

$$v = 2w^{-1}\frac{dw}{du} \tag{A14}$$

reduces (A9) to the second-order linear ODE

$$4\frac{d^2w}{du^2} + \left(\frac{1}{u^2-k} + \frac{1}{u^2}\right)w = 0. \tag{A15}$$

Equation (A15) can be solved in terms of hypergeometric functions.

APPENDIX B: THE GENERAL SOLUTION OF EQ. (3.6)

Here we find the general solution of Eq. (3.6) when $\lambda \neq 0$. Without loss of generality, $\lambda = 1$ or i, by an appropriate scaling of $c(x)$. Hence we consider

$$cc'(c/c')'' = \pm 1. \tag{B1}$$

This ODE can be fully integrated using group methods described in Ref. 3.

Since (B1) is invariant under scalings $x^* = \mu x$, $c^* = \mu c$, and translations in x, we choose new variables[3]

$$u = c', \quad v = cc'', \tag{B2}$$

which are differential invariants with respect to this two-parameter family of symmetries. Consequently (B1) becomes

$$\frac{dv}{du} = \frac{2v}{u} \mp \frac{u}{v}. \tag{B3}$$

Equation (B3) is homogeneous in u and v. Using this fact, one finds that the general solution of (B3) is

$$u^{-2}[(v/u)^2 \mp 1] = \text{const} = \nu^2. \tag{B4}$$

Now we choose new variables c and u, invariants under translations in x, so that (B4) becomes

$$\frac{du}{dc} = \frac{\sqrt{\nu^2u^2 \pm 1}}{c}. \tag{B5}$$

The general solution of (B5) is then

$$u = (1/2\nu)[(\rho c)^\nu \mp (\rho c)^{-\nu}], \tag{B6}$$

where ρ is an arbitrary constant. After scaling c and x so that ρ becomes 1, (B1) reduces to

$$c' = (1/2\nu)[c^\nu \mp c^{-\nu}], \tag{B7}$$

i.e.,

$$c' = \nu^{-1}\sinh(\nu \log c) \text{ or } \nu^{-1}\cosh(\nu \log c). \tag{B8}$$

If ν^2 is replaced by $-\nu^2$ in (B4) then (B7) reduces to

$$c' = \nu^{-1}\sin(\nu \log c). \tag{B9}$$

Equation (B8) can be integrated out if ν is any rational number.

[1] S. Lie, Arch. Math. **6**, 328 (1881).
[2] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[3] G. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, Berlin, 1974).
[4] G. W. Bluman, SIAM J. Appl. Math. **43**, 1259 (1983).
[5] R. L. Anderson, S. Kumei, and C. E. Wulfman, Rev. Mex. Fis. **21**, 1, 35 (1972); Phys. Rev. Lett. **28**, 988 (1972).
[6] R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[7] S. Kumei and G. W. Bluman, SIAM J. Appl. Math. **42**, 1157 (1982).

On the inverse problem and prolongation structure for the modified anisotropic Heisenberg spin chain

Archan K. De and A. Roy Chowdhury
High Energy Physics Division, Department of Physics, Jadavpur University, Calcutta-700 032, India

(Received 5 February 1986; accepted for publication 27 August 1986)

A modified form of the anisotropic Heisenberg spin chain has been considered. By use of the prolongation structure technique of Wahlquist and Estabrook, the elliptic Lax pair associated with this equation has been deduced. A variant of the Reimann–Hilbert problem on the torus is used to indicate the way to the solution of the inverse problem.

I. INTRODUCTION

In the last decade there has been a tremendous amount of progress in the theoretical study of nonlinear phenomena. Applicability of such methodologies to various classes of physical situations made the research worth pursuing. One of the most important equations occurring in the domain of solid state physics is that of the Heisenberg spin chain.[1] An interesting feature of many such integrable equations is that many of them do possess a similar but slightly different integrable form known as the modified form. The most familiar example is that of the KdV and MKdV cases.[2] Also, it has been found that it is possible to find a mapping between the usual and the modified form of the KdV equation—a phenomenon known as deformation.[3] In this paper we propose to study a modified anisotropic Heisenberg spin chain equation, which, as far as these authors' knowledge goes, was not previously known. We have made a prolongation analysis to deduce the Lax pair and then study the inverse problem with the technique of the Reimann–Hilbert problem on the torus[4] because the Lax pair obtained is a doubly periodic function of the spectral parameter. The problem that we do not touch upon is that of the mapping between the anisotropic Heisenberg spin chain and our equation.

II. FORMULATION

The equation we propose to study is written as

$$\frac{\partial \boldsymbol{\mu}}{\partial t} = -[\boldsymbol{\mu}\times\boldsymbol{\mu}_{xx}] - \tfrac{1}{2}\{\boldsymbol{\mu}(\boldsymbol{\mu},B\boldsymbol{\mu})\}_x + (\alpha + B)\boldsymbol{\mu}_x , \quad (1)$$

where $\boldsymbol{\mu}$ is a three-vector $(\mu_1(x),\mu_2(x),\mu_3(x))$ coupled by the constraint

$$\mu_1^2(x) + \mu_2^2(x) + \mu_3^2(x) = 1 , \quad (2)$$

where α is an arbitrary constant, $B = \text{diag}\,(b_1,b_2,b_3)$, (a,b) denotes the scalar product, and $(a\times b)$ denotes the cross product. To deduce the inverse scattering equation, we follow the methodology laid down by Wahlquist and Estabrook. In this approach we first convert the set of equations (1) in the language of the differential form.

If we now introduce the new set of independent variables, $\mu_{1x} = p$, $\mu_{2x} = q$, and $\mu_{3x} = r$, then we observe that our given equation set is equivalent to the following differential forms on proper sectioning.

We name these differential two-forms as α_i:

$$\begin{aligned}
\alpha_1 &= d\mu_1\wedge dt - p\,dx\wedge dt ,\\
\alpha_2 &= d\mu_2\wedge dt - q\,dx\wedge dt ,\\
\alpha_3 &= d\mu_3\wedge dt - r\,dx\wedge dt ,\\
\alpha_4 &= d\mu_1\wedge dx + \mu_2\,dr\wedge dt - \mu_3\,dq\wedge dt\\
&\quad - \{(\tfrac{3}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2) + \alpha + b_1\}d\mu_1\wedge dt\\
&\quad - \tfrac{1}{2}b_2\mu_1\mu_2\,d\mu_2\wedge dt - \tfrac{1}{2}b_3\mu_1\mu_3\,d\mu_3\wedge dt ,\\
\alpha_5 &= d\mu_2\wedge dx + \mu_3\,dp\wedge dt - \mu_1\,dr\wedge dt\\
&\quad - \{(\tfrac{1}{2}b_1\mu_1^2 + \tfrac{3}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2) + \alpha + b_2\}d\mu_2\wedge dt\\
&\quad - \tfrac{1}{2}b_1\mu_1\mu_2\,d\mu_1\wedge dt - \tfrac{1}{2}b_3\mu_2\mu_3\,d\mu_3\wedge dt ,\\
\alpha_6 &= d\mu_3\wedge dx + \mu_1\,dq\wedge dt - \mu_2\,dr\wedge dt\\
&\quad - \{(\tfrac{1}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{3}{2}b_3\mu_3^2) + \alpha + b_3\}d\mu_3\wedge dt\\
&\quad - \tfrac{1}{2}b_1\mu_3\mu_1\,d\mu_1\wedge dt - \tfrac{1}{2}b_2\mu_3\mu_2\,d\mu_2\wedge dt .
\end{aligned} \quad (3)$$

It is interesting to observe that this set of forms generates a closed ideal; that is,

$$d\alpha_i = \sum f_{ij}\wedge\alpha_j , \quad (4)$$

where f_{ij} are some functions. The basic assumption of the prolongation theory is that we may extend the set of variables from $(x,t,p,q,r,\mu_1,\mu_2,\mu_3)$ to $(x,t,p,q,r,\mu_1,\mu_2,\mu_3$ and $yi)$ (for the moment i is unspecified), where these extra independent variables (whose number is not fixed to start with) are the prolongation variables. With this setting in mind we proceed to search for a one-form

$$\Omega_k = dy_k + F_k\,dx + G_k\,dt , \quad (5)$$

with

$$\begin{aligned}
F_k &= F_k(x,t,p,q,r,\mu_1,\mu_2,\mu_3,y_i) ,\\
G_k &= G_k(x,t,p,q,r,\mu_1,\mu_2,\mu_3,y_i) ,
\end{aligned}$$

such that the exterior derivative of Ω_k, that is, $d\Omega_k$, also remains in the extended ideal comprising of (α_k,Ω_k). Written explicitly, this condition reads

$$d\Omega_k = \sum f_i^k\alpha_i + \sum (a_i^k\,dx + b_i^k\,dt)\wedge d\Omega_i . \quad (6)$$

On the other hand,

$$d\Omega = \sum \frac{\partial F}{\partial x_\mu}dx_\mu\wedge dx + \sum \frac{\partial G}{\partial x_\mu}dx_\mu\wedge dt , \quad (7)$$

where x_μ denotes the full set of independent variables

$$x_\mu \equiv \{x,t,p,q,r,\mu_1,\mu_2,\mu_3,y_i\} . \quad (8)$$

So comparing Eqs. (6) and (7), equating coefficients of different two-forms such as $dx\wedge dt$, $dx\wedge dp$, etc., we get the

 0022-2488/87/020319-04$02.50 © 1987 American Institute of Physics

following conditions restricting the structure of the functions F and G:

$$aG - bF - pf_1 - qf_2 - rf_3 = 0, \tag{9}$$

$$G_p = \mu_3 F_{\mu_2} - \mu_2 F_{\mu_3},$$
$$G_q = \mu_1 F_{\mu_3} - \mu_3 F_{\mu_1}, \tag{10}$$
$$G_r = \mu_2 F_{\mu_1} - \mu_1 F_{\mu_2},$$

$$G_{\mu_1} = f_1 - f_4\left[\tfrac{3}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2 + \alpha + b_1\right] - \tfrac{1}{2}f_5 b_1\mu_1\mu_2 - \tfrac{1}{2}f_6 b_1\mu_3\mu_1, \tag{11}$$

$$G_{\mu_2} = f_2 - f_5\left[\tfrac{1}{2}b_1\mu_1^2 + \tfrac{3}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2 + \alpha + b_2\right] - \tfrac{1}{2}f_4 b_2\mu_1\mu_2 - \tfrac{1}{2}f_6 b_2\mu_3\mu_2, \tag{12}$$

$$G_{\mu_3} = f_3 - f_6\left[\tfrac{1}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{3}{2}b_3\mu_3^2 + \alpha + b_3\right] - \tfrac{1}{2}f_5 b_3\mu_2\mu_3 - \tfrac{1}{2}f_4 b_3\mu_1\mu_3, \tag{13}$$

$$F_p = G_q = G_r = 0,$$
$$G_{pp} = G_{qq} = G_{rr} = 0, \tag{14}$$
$$G_{p\mu_2} = -F_{\mu_3}, \quad G_{p\mu_3} = -F_{\mu_2}, \text{ etc.},$$

along with the condition

$$pG_{\mu_1} + pF_{\mu_1}\left[\tfrac{3}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2 + \alpha + b_1\right] + (p/2)b_1\mu_1\mu_2 F_{\mu_2} + (p/2)b_1\mu_3\mu_1 F_{\mu_3} + qG_{\mu_2}$$
$$+ qF_{\mu_2}\left[\tfrac{1}{2}b_1\mu_1^2 + \tfrac{3}{2}b_2\mu_2^2 + \tfrac{1}{2}b_3\mu_3^2 + \alpha + b_2\right] + (q/2)b_2\mu_1\mu_2 F_{\mu_1} + (q/2)b_2\mu_3\mu_2 F_{\mu_3} + rG_{\mu_3} \tag{15}$$
$$+ rF_{\mu_3}\left[\tfrac{1}{2}b_1\mu_1^2 + \tfrac{1}{2}b_2\mu_2^2 + \tfrac{3}{2}b_3\mu_3^2 + \alpha + b_3\right] + (r/2)b_3\mu_2\mu_3 F_{\mu_2} + (r/2)b_3\mu_1\mu_3 F_{\mu_3} + [F,G] = 0,$$

where

$$[F,G] = F_k \frac{\partial G}{\partial y_i} - G_k \frac{\partial F}{\partial y_i}.$$

It is then interesting to observe that the above equations dictate the following forms of F and G, the factors represented by $X_i(y)$ denote the dependence on the prolongation variables y_i that is still unknown,

$$F = \sigma_1\mu_1 X_1(y_i) + \sigma_2\mu_2 X_2(y_i) + \sigma_3\mu_3 X_3(y_i),$$
$$G = \omega_1^{-1}\{s\mu_1 + \mu_1(\alpha - (\Lambda/2)) - (\mu_2\mu_{3x} - \mu_3\mu_{2x})\}X_1(y_i) + \omega_2^{-1}\{s\mu_2 + \mu_2(\alpha - (\Lambda/2)) \tag{16}$$
$$- (\mu_3\mu_{1x} - \mu_1\mu_{3x})\}X_2(y_i) + \omega_3^{-1}\{s\mu_3 + \mu_3(\alpha - (\Lambda/2)) - (\mu_2\mu_{1x} - \mu_1\mu_{2x})\}X_3(y_i).$$

In Eq. (16), s, σ_i, and ω_i^{-1} $(i = 1,2,3)$ are arbitrary constants to be determined. Substitution of these forms of F and G in (15) leads to the following: (i) if we consider the constants σ_i and ω_i such that

$$\sigma_1\omega_1 = \sigma_2\omega_2 = \sigma_3\omega_3 = 1, \tag{17}$$

and (ii) the $X_i(y)$ thought of as operators in y satisfy

$$[X_i, X_j] = \epsilon_{ijk}X_k, \tag{18}$$

then the equation satisfied by μ_1 (for example) is

$$\mu_{1t} + \left(\frac{\alpha}{2} - s - \frac{\sigma_2}{\omega_3\sigma_1}\right)\mu_{1x} + \frac{\mu_1}{2}(\boldsymbol{\mu}, B\boldsymbol{\mu})_x + \epsilon_{ijk}\mu_j\mu_{kxx} = 0, \tag{19}$$

after we take care of the condition $\mu_1 p + \mu_2 q + \mu_3 r = 0$, which is deduced from $\mu_1^2 + \mu_2^2 + \mu_3^2 = 1$. If we now compare Eq. (19) with the first component of the original nonlinear Equation (1) [the same procedure is to be followed for the 2nd and 3rd component of (1) also], then we arrive at

$$b_1 = -\tfrac{3}{2}\alpha + s + \omega_1/\omega_2\omega_3,$$
$$b_2 = -\tfrac{3}{2}\alpha + s + \omega_2/\omega_1\omega_3, \tag{20}$$
$$b_3 = -\tfrac{3}{2}\alpha + s + \omega_3/\omega_1\omega_2.$$

Hence the consistency of the prolongation equations leads to the nonlinear equations we started with.

III. INTRODUCTION OF A SPECTRAL PARAMETER

A derivation of the linear problem associated with a nonlinear equation can never be thought to be complete until a spectral parameter can be introduced. Here we show how our linear problem associated with (F,G) has elliptic dependence on the spectral parameter.

From Eq. (20) we obtain

$$b_1 - b_2 = \frac{1}{\omega_1\omega_2\omega_3}(\omega_1^2 - \omega_3^2) = \omega_1^{12} - \omega_3^{12}, \tag{21}$$

and

$$b_2 - b_3 = \omega_2^{12} - \omega_3^{12}, \quad b_3 - b_1 = \omega_3^{12} - \omega_1^{12},$$

with

$$\omega_i^1 = \omega_i/(\omega_1\omega_2\omega_3)^{1/2} \quad (i = 1,2,3).$$

Equation (21) immediately suggests that the ω_i's are soluble in terms of elliptic functions:

$$\omega_1^1 = \frac{\rho}{\mathrm{sn}(\lambda,k)}, \quad \omega_2^1 = \frac{\rho\,\mathrm{dn}(\lambda,k)}{\mathrm{sn}(\lambda,k)}, \quad \omega_3^1 = \frac{\rho\,\mathrm{cn}(\lambda,k)}{\mathrm{sn}(\lambda,k)}, \tag{22}$$

with

$$b_1 - b_2 = k^2\rho^2,$$
$$b_2 - b_3 = (1 - k^2)\rho^2, \tag{23}$$

so that the x part of our Lax pair is written as

$$L = [\rho^2/\mathrm{sn}^2(\lambda,k)][\mu_1 \ln(\lambda,k)\mathrm{dn}(\lambda,k)X_1 + \mu_2\,\mathrm{cn}(\lambda,k)X_2 + \mu_3\,\mathrm{dn}(\lambda,k)X_3]. \tag{24}$$

The time part is given by G. The common parameter λ occurring in the elliptic functions serves as the spectral parameter. It is important to note that L is a doubly periodic function of λ defined on a torus.

IV. SOLUTION OF THE INVERSE PROBLEM

The usefulness of the Lax pair deduced above can be ascertained if and only if we can effectively use them to determine the nonlinear fields $\mu_i(x)$ from a suitably given spectral data. The main steps for the inverse problem for a doubly periodic Lax operator has been given in detail by Rodin. Due to the occurrence of the elliptic functions in Eq. (24), we have

$$\begin{aligned} &L(\lambda+2k)=\sigma_3 L\sigma_3\,, \\ &L(\lambda+2ik')=\sigma_1 L\sigma_1\,, \end{aligned} \tag{25}$$

where k and ik' are the two periods of the elliptic functions and Eq. (25) defines an involution. Also if we assume the asymptotic condition

$$\mu_1(x),\mu_2(x)\to 0\,,\quad \mu_3(x)\to 1\quad \text{as } x\to\infty\,,$$

then Eq. (24) yields

$$\psi_x(\infty)\cong[i\rho^2/\mathrm{sn}^2(\lambda,k)]\,\mathrm{dn}(\lambda,k)X_3\psi(\infty)\,,$$

whose straightforward solution is seen to be

$$\psi(\infty)\sim\exp\left[\frac{xi\rho^2}{\mathrm{sn}^2(\lambda,k)}\,\mathrm{dn}(\lambda,k)\right]\begin{pmatrix}1\\0\end{pmatrix}. \tag{26}$$

To extract the asymptotic behavior from (24) we set

$$\phi(x)=\exp\left(-i\rho^2\frac{\mathrm{dn}(\lambda,k)\sigma_3 x}{\mathrm{sn}^2(\lambda,k)}\right)\psi(x)\,,$$

which shows that ϕ satisfies

$$\begin{aligned}\phi_x=\Big\{&i\rho^2\frac{\mathrm{dn}(\lambda,k)}{\mathrm{sn}^2(\lambda,k)}\sigma_3+\frac{\rho^2}{\mathrm{sn}^2(\lambda,k)}(\mu_1\sigma_1\,\mathrm{cn}(\lambda,k)\mathrm{dn}(\lambda,k)\\&+\mu_2\sigma_2\,\mathrm{cn}(\lambda,k)+\mu_3\sigma_3\,\mathrm{dn}(\lambda,k))\Big\}\phi\,.\end{aligned} \tag{27}$$

Now it is well known from the properties of the elliptic functions that

$$\mathrm{cn}(\lambda,k)\,,\quad \mathrm{dn}(\lambda,k)\to 1\,,\quad \mathrm{sn}(\lambda,k)\to\lambda\,,\quad \text{for small } \lambda\,.$$

So for Eq. (27) to be free from singularity at $\lambda=0$, we must have the residue at $\lambda=0$ to be equal to zero, so

$$\begin{aligned}&\mu_i\sigma_i\phi(\lambda=0)+i\phi(\lambda=0)\sigma_3=0\,,\\ &\text{or}\\ &\mu_i\sigma_i=-i\phi(\lambda=0)\sigma_3\phi^{-1}(\lambda=0)\,.\end{aligned} \tag{28}$$

So Eq. (28) is nothing but the key equation for the inverse problem. Because, if $\phi(x)$ is known from some spectral considerations, then $\mu_i(x)$ can be obtained via (28).

V. THE ANALYTIC STRUCTURE OF ϕ AND ITS USE

In the next part of our discussion the two main properties of ϕ that we will use are (i) the pole structure in the λ plane and (ii) the double periodicity. The explicit form of ϕ that we will use is constructed on the basis of these two properties and was first used in Ref. 5. It is actually an expansion in terms of the Riemann zeta function:

$$\phi(\lambda)=N_1\left[I+\sum_{\alpha=1}^{N}\sum_{i=1}^{4}B_i^{\alpha}f_\alpha^i(\lambda)\right]B\,, \tag{29}$$

where

$$f_\alpha^i(\lambda)=\zeta(\lambda-\mu_\alpha)-c^i\,,$$

ζ is the Riemann zeta function, the c^i are constants, and the B's are the matrices satisfying the conditions

$$\begin{aligned}&B_1^{\alpha}=\sigma_1 B_2^{\alpha *}\sigma_1=\sigma_3 B_3^{\alpha}\sigma_3=\sigma_3\sigma_1 B_4^{\alpha *}\sigma_1\sigma_3\,,\\ &B=\sigma_1 B^{*}\sigma_1=\sigma_3\sigma_1 B^{*}\sigma_1\sigma_3\,,\end{aligned} \tag{30}$$

where N_1 is a normalizing diagonal matrix.

It is also important to note

$$\begin{aligned}&f_\alpha^s(\lambda-2k)=f_\alpha^{s+2}(\lambda)-\eta^i\,,\\ &f_\alpha^{2s-1}(\lambda+2ik')=f_\alpha^{2s^*}(\lambda^*)\,,\quad s=1,2\,.\end{aligned} \tag{31}$$

After substitution of (29) in (24), if we demand that the resulting expression should not have pole at $\lambda=\mu_i$, we arrive at

$$m^{*\beta}\phi(\lambda_\beta)=0\,, \tag{32}$$

showing that matrix B is degenerate and can be represented in the form

$$(B_1^{\alpha})_{ab}=(m^{\alpha})_a(Y^{\alpha})_b\,. \tag{33}$$

This explicit form of (B_1) yields the following sets of equations:

$$\begin{aligned}m^{*\beta}\sigma_1+\sum_{\alpha=1}^{N}\Big[&(m^{*\beta}\sigma_1 m^{\alpha})f_\alpha^{(1)}(\lambda_\beta)Y^{\alpha}+(m^{*\beta}\sigma_1\sigma_3 m^{\alpha})\\&\times f_\alpha^{(3)}(\lambda_\beta)Y^{\alpha}\sigma_3+(m^{*\beta}m^{*\alpha})f_\alpha^{(2)}(\lambda_\beta)Y^{*\alpha}\sigma_1\\&+(m^{*\beta}\sigma_3 m^{*\alpha})f_\alpha^{(4)}(\lambda_\beta)Y^{*\alpha}\sigma_3\sigma_1\Big]=0\,.\end{aligned} \tag{34}$$

If we now set $Y'=(_b^a)$ and $m=(m_1,m_2)$, then (34) can be recast as

$$m^{*\beta}\sigma_1+U^{\alpha}Y^{\alpha}+V^{\alpha}Y^{\alpha}\sigma_3+W^{\alpha}Y^{\alpha}\sigma_3+Z^{\alpha}Y^{\alpha}\sigma_3\sigma_1=0, \tag{35}$$

with

$$\begin{aligned}&U^{\alpha}=(m^{*\beta}\sigma_1 m^{\alpha})f_\alpha^{(1)}(\lambda_\beta)\,,\\ &V^{\alpha}=(m^{*\beta}\sigma_1\sigma_3 m^{\alpha})f_\alpha^{(3)}(\lambda_\beta)\,,\\ &W^{\alpha}=(m^{*\beta}m^{*\alpha})f_\alpha^{(2)}(\lambda_\beta)\,,\\ &Z^{\alpha}=(m^{*\beta}\sigma_3 m^{*\alpha})f_\alpha^{(4)}(\lambda_\beta)\,.\end{aligned} \tag{36}$$

The set of equations in (35) can be solved for Y. So the vector X^i is explicitly known in terms of the values of ζ-functions, where the only arbitrariness is in the choice of m, which can be chosen properly.

And hence we can determine the eigenfunction ϕ, which in turn determines μ_i through Eq. (28).

VI. DISCUSSION

In our above analysis we have discussed the prolongation approach to a new nonlinear equation and have shown how an elliptic Lax pair can result from such an analysis. Lastly, a variant of the Riemann–Hilbert problem on the torus has been used to solve explicitly the full inverse problem.

[1]R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons And Nonlinear Wave Equations* (Academic, New York, 1982); H. C. Foged-

by, "Theoretical aspects of mainly low dimensional magnetic systems," in *Lecture Notes in Physics*, Vol. 131 (Springer, Berlin, 1982).
[2]See the lecture by R. K. Bullough in the volume *Soliton—Current Topics in Physics*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).
[3]B. Kuperschmidt, preprints Los Alamos National Laboratory, 1980, 1981.
[4]W. Koppelman, Commun. Pure Appl. Math. **12**, 13 (1959).
[5]Y. L. Rodin, Lett. Math. Phys. **110**, 90 (1984).

Singular points from Taylor series

Francisco M. Fernández, Gustavo A. Arteca, and Eduardo A. Castro[a)]
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), División Química Teórica, Sucursal 4, Casilla de Correo 16, (1900) La Plata, Argentina

(Received 13 March 1986; accepted for publication 30 April 1986)

A simple and accurate method is developed for calculating singular points from Taylor series. It consists of finding the least-squares deviation of the Taylor coefficients from a proposed asymptotic form. Sequences are obtained that converge quickly to the closest singularity to the origin. Some simple mathematical examples and physically interesting eigenvalue problems are discussed to illustrate the procedure. The branch points of the eigenvalues for the solutions of period 2π of the Mathieu equation and those of the Stark shifts for rigid symmetric-top molecules, which were not obtained before, are shown.

I. INTRODUCTION

It often happens that the only available approach to a physical problem is a Taylor series. In that case one has to obtain as much information as possible from it. For example, when the expansion (supposed, without loss of generality, to be about the origin) is known to have a finite radius of convergence, it is of great interest to determine the number, kind (i.e., pole, branch point, etc.), and position of the closest singularities to the origin. The accuracy of the singular points obtained from power series expansions depends on the number of available Taylor coefficients (and their rapidity in reaching the asymptotic behavior[1]).

Calculation of singularities from Taylor series is of utmost importance in the examples below.

(a) Some quantum-mechanical eigenvalue problems. In this case the convergence radius of the perturbation series is determined by branch points due to level crossings in the complex plane (see Refs. 2–6, and references therein).

(b) Critical phenomena and phase transitions. High- and low-temperature expansions and virial series prove to be suitable for calculation of phase transitions in spin-lattice models[7,8] and fluid–solid phase transitions in imperfect-gas continuum models,[9–12] respectively.

Most of the methods used in finding singular points by means of Taylor series apply when the singularity nearest to the origin is real. However, the convergence radius of the power series is frequently determined by, at least, a pair of complex conjugate branch points, as in the case of some critical phenomena[7,13] and quantum theory problems.[6]

The aim of the present paper is to develop a new way of obtaining, from the Taylor series, the parameters (critical parameters from now on) characterizing the closest singular point (real or complex) to the origin. The method is developed in Sec. II and its connection to other techniques is discussed in Sec. III. The procedure is checked in Sec. IV by means of some simple mathematical examples and applied to the bounded delta atom in Sec. V. The branch points of the characteristic values of the Mathieu equation of period 2π and of the Stark shifts for a rigid symmetric-top molecule are discussed in Sec. VI. Conclusions are found in Sec. VII.

II. THE METHOD

Let $f(z)$ be a function of the complex variable $z = x + iy$ that has singular points at $z_1, z_2, \ldots, z_j = x_j + iy_j$, numbered in such a way that $|z_1| \leqslant |z_2| \leqslant \cdots$. Owing to this, the Taylor series about the origin

$$f(z) = \sum_{n=0}^{\infty} f_n z^n \tag{1}$$

will converge in $|z| < |z_1|$. It is further supposed that $f(x)$ is real, which means that the coefficients f_n are real and each singular point z_j is either real or a complex conjugate of another one. It will be shown below how to obtain the closest singularity to the origin from the expansion (1).

For the sake of simplicity, we first discuss the algebraic singularities and then extend the method to other cases. If z_j is a branch point of order $k-1$ or a pole ($k=1$), $f(z)$ can be expanded, in a neighborhood of z_j, as[14]

$$f(z) = \sum_{n=m}^{\infty} F_n q_j^n, \tag{2}$$

where m is a positive or negative integer and $q_j = (z - z_j)^{1/k}$. To illustrate the procedure let us suppose that $z_2 = z_1^*$. Upon neglecting terms of order larger than m we have $f(z) \simeq F_0 + F_1 q_1^m$, which, on the same level of approximation, can be rewritten as

$$f(z) \simeq F_0 + A\, q_1^m q_2^m, \tag{3}$$

where $A = F_1/(z_1 - z_2)^a$, $a = m/k$, and $q_1^m q_2^m = (z^2 - 2x_1 z + x_1^2 + y_1^2)^a$. The singular points z_1 and z_2 are characterized by the critical parameters (CP) (x_1, y_1), a, and A, called the critical position, exponent, and amplitude, respectively.

On the other hand, the coefficients Y_n of the Taylor expansion about the origin for the generating function (GF)

$$Y(z) = B(z^2 - 2uz + r^2)^b, \tag{4}$$

obey the three-term recursion relationship

$$(n - 2b - 1)Y_{n-1} + 2u(b - n)Y_n + (n+1)r^2 Y_{n+1} = 0, \tag{5}$$

[a)] To whom all correspondence should be addressed.

 0022-2488/87/020323-07$02.50 © 1987 American Institute of Physics

where $Y_0 = Br^{2b}$ and $Y_n = 0$ if $n < 0$.

The present technique, called the generating function method (GFM) from now on, rests upon the supposition that the Taylor coefficients f_n tend to obey Eq. (5) as n tends to infinity provided $u = x_1$, $r^2 = x_1^2 + y_1^2$, and $b = a$. Therefore, if Y_n is replaced by f_n in Eq. (5) for $n = N, N+1$, and $N+2$ and the resulting set of equations is solved for u, r^2, and b, we obtain three sequences, namely u_N, r_N^2, and b_N, that converge towards x_1, $x_1^2 + y_1^2$, and a, respectively, as N tends to infinity. The recursion relationship (5) also applies when $f(z)$ has a pair of complex conjugate branch points with the same irrational exponent a. This is due to the fact that, though the argument leading to Eq. (3) does not hold in this case, $f(z)$ is expected to behave approximately as const $+ A\ (z - z_1)^a (z - z_1^*)^a$ in the neighborhood of either z_1 or $z_2 = z_1^*$ and the generating function (4) proves to be suitable.

Though we are at present unable to prove rigorously that our sequences converge towards the actual CP, extensive numerical investigation covering a very large number of mathematical examples and physical problems suggests that the GFM is always successful provided an appropriate GF is chosen. This is confirmed in later sections. For example, in the particular case just discussed (i.e., $z_2 = z_1^*$) the sequences converge quickly if z_3 is far enough from the convergence circle.

Since the exact a value is known beforehand for all the problems studied in this paper we restrict ourselves to this simpler case. A straightforward algebraic manipulation of Eq. (5) with $Y_n = f_n$ and $b = a$ leads to

$$u_N = \Delta(W_N/U_N)/\Delta(V_N/U_N), \tag{6a}$$

$$r_N^2 = \Delta(W_N/V_N)/\Delta(U_N/V_N), \tag{6b}$$

where $U_N = 2(a - N)$, $V_N = (N+1)Q_N$, $W_N = (2a + 1 - N)/Q_N$, $Q_N = f_N/f_{N-1}$, and $\Delta P_N = P_{N+1} - P_N$.

If A is real we can calculate it as follows: Let u_L, r_L^2, and b_L the approximate limits of the u, r^2, and b sequences, respectively, computed somehow. Then we can use the recursion relationship (5), to obtain $Y_N(b_L, u_L, r_L^2, B = 1)$. Therefore, since $Y_N(B) = BY_N(B = 1)$, it is assumed that the sequence

$$B_N = f_N/Y_N(b_L, u_L, r_L^2, B = 1) \tag{7}$$

must converge towards A as N increases. When A is not real, the B sequence is found to be strongly oscillatory, as shown later on in Sec. IV.

The CP sequences can be obtained another way. The recursion relationship (5) enables one to calculate $Y_n(u, r^2, b, B)$ quickly for large enough n values. Then the values of the adjustable CP u, r^2, b, and B can be selected so that the smallest square deviation

$$S_{M,N} = \sum_{n=M}^{N} (f_n - Y_n)^2, \quad N > M + 2, \tag{8}$$

is obtained. This leads to sequences that are believed to converge towards the actual CP as M and N tend to infinity (the B sequence converges provided A is real). This form of the GFM can be viewed as a generalization of the asymptotic least-squares method recently developed to obtain the real singular points of the compressibility factor of imperfect-gas continuum models from the virial series.[15] Though this calculation scheme leads to greatly revealing two-entry tables,[15] in this paper we mostly use the much simpler one discussed above that is accurate enough for our purposes. However, we want to emphasize the very striking fact that when A is not real the critical exponent and position sequences obtained from the least-squares version of the GFM are convergent in spite of the fact that the minimum equations depend on the divergent B sequence.

The GFM applies to cases other than those discussed above provided the GF is chosen accordingly. For example, if z_1 is real and $|z_2| > |z_1|$ the simplest GF is

$$Y(z) = B(1 - z/u)^b, \tag{9}$$

whose Taylor coefficients obey

$$(n+1)Y_{n+1} = (n - b)Y_n/u. \tag{10}$$

After replacing Y_n by f_n we obtain the equations of the well-known ratio method (Chap. 4 in Ref. 7):

$$u_N^{-1} = \Delta(NQ_N), \tag{11a}$$

$$b_N = -1 - N(N+1)u_N \Delta Q_N. \tag{11b}$$

These sequences prove to be convergent if the conditions above are fulfilled.

When z_2 lies close to the convergence circle the ratio method converges too slowly (interfering singularities). In such a case, assuming for the sake of simplicity that z_2 is real, we try a GF of the form

$$Y(z) = B(1 - z/u)^b (1 - z/v)^{b'}, \tag{12}$$

or

$$Y(z) = B(1 - z/u)^b + B'(1 - z/v)^{b'}. \tag{13}$$

They may be useful, for example, in studying phase transitions in antiferromagnetic spin-lattice models[7] or when there are confluent singularities.[4,8] The specific heat and magnetic susceptibility of two-dimensional loose-packed spin-lattice Ising models exhibit logarithmic singular points.[7] These problems can be treated by means of appropriate logarithmic GF 's.

If the form of the singularity closest to the origin is known beforehand, the appropriate GF is easily built. Otherwise, we try different ones till we obtain quickly convergent CP sequences. Since the examples discussed in this paper exhibit a pair of complex conjugate branch points of order 1 ($k = 2$) closest to the origin, then Eqs. (6) and (7) with $a = b_L = \frac{1}{2}$ can be used.

III. RELATION WITH BERNOULLI'S ITERATION ALGORITHM

Let $z_1, z_2, \ldots, z_n$ (numbered as before) be the roots of the algebraic equation $g(z) = 0$, where

$$g(z) = g_0 + g_1 z + \cdots + g_n z^n. \tag{14}$$

If $g_0 \neq 0$ the function $f(z) = 1/g(z)$, which has a pole at each z_j, can be expanded in Taylor series about the origin. The coefficients f_i can be obtained from

$$f_0 = 1/g_0, \tag{15a}$$

$$f_n = -g_n^{-1}\sum_{i=0}^{n} g_i f_{n-i}, \quad n>0. \tag{15b}$$

Therefore, the method developed in Sec. II can be used to calculate the zero of $g(z)$ nearest to the origin (and also the other ones after the appropriate polynomial factorization). Though there are better ways of doing this in the mathematical literature, we discuss the GFM here in order to show that it reduces to Bernoulli's algorithm.[16]

First of all it must be noticed that, due to the way Eq. (14) is written, present roots are exactly the inverse of those in Ref. 16. Let us consider three different cases (cf. Ref. 16).

(a) z_1 is real and $|z_1|<|z_j|$, $j>1$. The GF (10) with $b=-1$ leads to

$$z_1 = \lim_{i\to\infty} f_{i-1}/f_i. \tag{16}$$

(b) $z_1 = z_2^*$ and $|z_1|<|z_j|$, $j>2$. It follows from Eq. (6) with $b=-1$ that

$$x_1 = \lim_{i\to\infty} (f_{i-1}f_i - f_{i-2}f_{i+1})/\{2(f_i^2 - f_{i-1}f_{i+1})\}, \tag{17a}$$

$$x_1^2 + y_1^2 = \lim_{i\to\infty} (f_{i-1}^2 - f_i f_{i-2})/(f_i^2 - f_{i-1}f_{i+1}). \tag{17b}$$

(c) $z_1 = z_2$ are real and $|z_1|<|z_j|$, $j>2$. Since (4) is the appropriate GF with $r^2 = u$ and $b=-1$, we can use Eq. (17a) to obtain z_1.

Other exceptional cases, in which several roots have the same minimum absolute value, can be treated in a similar way.

This procedure is also suitable for obtaining the roots of infinite series. For example, Aguilera-Navarro and Aguilera-Navarro[17] derived Eqs. (15) and (16) from Padé approximants. However, these authors were not able to explain why the algorithm always yields the closest zero to the origin.[17] A very simple proof is given in Ref. 16.

We are at present unable to prove rigorously that there is an appropriate GF for each problem leading to convergent CP sequences. However, it is our aim to give satisfactory enough evidence that this is so. To begin with, notice that in certain cases the GFM reduces to two well-known convergent algorithms such as that of Bernoulli[16] and the ratio method (Chap. 4 in Ref. 6). Besides, it will be shown in the next sections that the GFM applies successfully to a number of examples.

IV. SIMPLE MATHEMATICAL EXAMPLES

In order to verify the statements in Sec. II we will discuss some simple mathematical examples, which are nontrivial in the sense that $Y(z)\neq f(z)$ even when the adjustable CP equal the actual ones.

Let us consider an implicit equation of the form $G(z,f(z))=0$ so that $f(z)$ and $z(f)$ are analytic at $z=0$ and $f=F_0=f(z_1)$, respectively. Therefore, if $(\partial^k z/\partial f^k)(f=F_0)=0$ and $(\partial^{k+1}z/\partial f^{k+1})(f=F_0)\neq 0$ then z_1 is a branch point of $f(z)$ of order $k-1$ (order k in the notation of Ref. 6 and references therein). Some particular cases are considered below.

If

$$G(z,f) = z + z^2 - fe^{f+1}, \tag{18}$$

then $k=2$, $z_1 = -(1\pm 3^{1/2}i)/2$ (for the sake of simplicity we use z_1 to represent both z_1 and $z_2=z_1^*$), $f(z_1)=-1$, and $(\partial^2 f/\partial g^2)(z=z_1)=1$, $g=z+z^2$. Therefore,

$$f(z) \simeq -1 + 2^{1/2}(z^2+z+1)^{1/2}, \tag{19}$$

for z close enough to z_1. This shows plainly that $a=\frac{1}{2}$, $x_1=-\frac{1}{2}$, $|y_1|=3^{1/2}/2$, and $A=2^{1/2}$.

To obtain the coefficients of the Taylor series for $f(z)$ about the origin we notice that $f(z)$ obeys

$$(z^2 - z + z^2 f + zf)f' - f(1+2z) = 0, \tag{20}$$

where $f' = \partial f/\partial z$. The expansion of Eq. (20) in powers of z leads to the recursion relationship

$$f_n = \frac{1}{1-n}\left\{(n-3)f_{n-1} + \sum_{j=0}^{n-1} j f_j (f_{n-j} + f_{n-j-1})\right\}, \tag{21}$$

where $f_0=0$, and $f_1 = 1/e$, which enables one to compute all the Taylor coefficients quickly and accurately.

On using Eqs. (6) and (7), with $b_L = a$, we obtain the CP sequences in Table I. Their limits can be estimated from $1/N$ extrapolations. Since the B sequence converges in a stepwise manner we must select the corresponding member of each step before extrapolating. Thus we obtain $x_1 = -0.500\,000 \pm 3\times 10^{-6}$, $|y_1| = 0.866\,025 \pm 2\times 10^{-6}$, and $A = 1.414\,21 \pm 4\times 10^{-5}$ that closely agree with the actual CP. As happens when using other techniques the critical-amplitude sequence is always the most slowly convergent one.

The GFM is very promising in calculating complex singularities from power series because the most widely used algorithms, namely, the Padé approximants and their variants, such as the N point fits, do not appear to be so easy to handle.[7,13] Besides, the GFM is expected to be more accurate since it takes into account explicity the form of the singularity.

The next example is

$$G(z,f) = z + 2 - (f^2+2)e^f, \tag{22}$$

for which $k=2$, $z_1 = 2^{3/2}e^{-1}(\{\sin(1+\pi/4)-2\} \pm i\sin(1-\pi/4))$, $f(z_1) = -1 \pm i$, and $A = i(e/2y_1)^{1/2} \times \exp(\mp i/2)$. The Taylor coefficients can be evaluated as in the previous case and the CP sequences are shown in Table II. As expected the B sequence is strongly oscillating and does not converge. The CP estimates, $x_1 = -0.983\,37 \pm 2\times 10^{-5}$ and $|y_1| = 0.221\,58 \pm 10^{-5}$, are in close agreement with the exact values.

TABLE I. CP sequences for the example in Eq. (18).

N	$-u_N$	r_N	$(r_N^2-u_N^2)^{1/2}$	B_N
495	0.499 997 15	0.999 997 16	0.866 023 767	1.410 877
496	0.499 997 18	0.999 997 18	0.866 023 773	1.418 781
497	0.499 997 18	0.999 997 20	0.866 023 791	1.408 023
498	0.499 997 19	0.999 997 19	0.866 023 787	1.410 897
499	0.499 997 21	0.999 997 21	0.866 023 793	1.418 754
500	0.499 997 22	0.999 997 23	0.866 023 811	1.408 060

TABLE II. CP sequences for the example in Eq. (22).

N	$-u_N$	r_N	$(r_N^2-u_N^2)^{1/2}$	B_N
495	0.983 346 754	1.008 005 854	0.221 596 39	−0.195 032
496	0.983 346 749	1.008 005 833	0.221 596 32	0.469 296
497	0.983 346 760	1.008 005 828	0.221 596 25	1.414 442
498	0.983 346 785	1.008 005 839	0.221 596 19	3.068 908
499	0.983 346 821	1.008 005 862	0.221 596 14	7.377 252
500	0.983 346 860	1.008 005 892	0.221 596 10	84.007 040

V. THE BOUNDED DELTA-POTENTIAL ATOM

Bounded quantum-mechanical models prove to be useful in simulating some physical phenomena[18] (and references therein). The Kato–Rellich theorem[2,3] assures us that in such cases the perturbation expansion[18] has a non-null radius of convergence. Although there is no rigorous proof about the singularities determining it, it appears reasonable to think that they are branch points of order 1. To verify this assumption in this section we apply the GFM to a very simple one-dimensional example. More complex problems will be treated elsewhere in a forthcoming paper.

Let us consider a one-dimensional delta-potential hydrogenlike atom (atomic units are used throughout)

$$-\tfrac{1}{2}\psi''(x)-Z\delta(x)\psi(x)=E\psi(x), \tag{23}$$

within a box with impenetrable walls at $x=\pm b$, i.e., $\psi(\pm b)=0$. The odd-parity solutions of Eq. (23) are those of the particle in a box (and we, therefore, completely neglect them) whereas the energy eigenvalues for the even-parity states are easily shown to be the roots of[19,20]

$$z=v\cot v, \tag{24}$$

where $z=Zb$ and $v=b(2E)^{1/2}$.

By reasoning as in the previous section we conclude that $v(z)$ has an infinite number of branch points of order 1 at $z=z_0, z_1,\ldots$ given by the roots of

$$v=\tfrac{1}{2}\sin 2v, \tag{25}$$

and that

$$\frac{\partial^2 z}{\partial v^2}(v=w_0)=-\tfrac{2}{3}, \tag{26a}$$

$$\frac{\partial^2 z}{\partial v^2}(v=w_j)=-2,\quad j>0, \tag{26b}$$

where $w_j=v(z_j)$. The branch point at $z_0=1$ ($w_0=0$) can be neglected because it is due merely to a change of sign in the energy. The remaining ones are singularities in $E(z)$ and can be obtained as accurately as required from Eq. (25).

To obtain the energy perturbation series we first notice that $f(z)=v^2(z)$ obeys the following differential equation:

$$(f+z^2-z)f'+2f=0. \tag{27}$$

Therefore, the coefficients of the Taylor series about the origin for $f(z)$ can be obtained from

$$f_{n+1}=(n+1)^{-1}f_0^{-1}\Big\{(n-2)f_n+(n-1)f_{n-1}-\sum_{j=0}^{n-1}(j+1)f_{j+1}f_{n-j}\Big\}, \tag{28}$$

TABLE III. Branch points in the energy eigenvalues of the bounded delta-potential atom. Exact values are given within the parentheses.

j	x_1	y_1
1	1.895 27 (1.895 282)	3.719 45 (3.719 436)
2	2.180 19 (2.180 2181)	6.932 99 (6.932 967)
3	2.360 51 (2.360 58)	10.107 34 (10.107 30)
4	2.492 84 (2.492 953)	13.268 11 (13.268 063)

where $f_0=(2j+1)^2\pi^2/4$ and $j=0,1,\ldots$ is the quantum number. The perturbation series for the lowest eigenvalue ($j=0$) converges for all z values because of what was said above, while the radius of convergence of the jth state ($j>1$) is given by $|z_j|$.

On using Eqs. (6) (without extrapolation) with $N=500$ we obtain the results in Table III, which closely agree with the roots of Eq. (25) (between parentheses). This suggests that the GFM may be useful in dealing with the many other bounded quantum-mechanical models for which a very large number of energy perturbation corrections can be calculated easily.[18]

VI. PERIODIC EIGENVALUE PROBLEMS

In this section we will study two simple but nontrivial (in the sense that they are not exactly solvable) periodic eigenvalue equations of great physical importance. They are very useful for our purposes because the Kato–Rellich theorem[2,3] assures us that in both cases the perturbation series have non-null convergence radii. The first one is the Mathieu equation[21]

$$\psi''(\theta)+(\alpha-2q\cos 2\theta)\psi(\theta)=0. \tag{29}$$

Following standard notation we will use α and β to label the characteristic values for even- and odd-parity solutions, respectively, and even and odd subscripts to designate solutions of period π and 2π, respectively.

The characteristic values for the solutions of period π are known to have an infinite number of conjugate branch points of order 1 on the imaginary axis in the complex q plane. It is found that α_{4j-4} (β_{4j-2}), and α_{4j-2} (β_{4j}), $j=1,2,\ldots$, have a common real value at q_j^α (q_j^β). These singularities can be calculated easily from continued fractions expansions[21] or through a determinantal recurrence relation.[22]

On the other hand, no systematic study of the singularities of the characteristic values for the solutions of period 2π has been carried out as far as we know (see note added in proof). However, there is no doubt that no real singular point exists.[21] Since $\alpha_{2j+1}(q)=\beta_{2j+1}(-q)$ (see Ref. 21) we only consider the even-parity solutions. Diagonalization of the matrix of the linear operator $H(q)=d^2/d\theta^2-2q\cos\theta$ in the basis set of eigenfunctions of $H(0)$ for complex q values shows that α_1 and α_3 have a common complex value at ($q=z$, $q_j=z_j$, etc.) $x_1=1.931\,3926$, $|y_1|=3.237\,6385$ (see Ref. 23). Unfortunately, this brute-

force procedure is extremely cumbersome, lengthy, and restricted to eigenvalue problems. A better approach is required.

The Taylor series about the origin for the characteristic values of Eq. (29) are easily calculated by standard perturbation theory[21] (see the Appendix). According to our computational facilities we have been able to calculate $\alpha_{2j+1}^{(s)}$ for s (perturbation order) <40. It is found that $\alpha_1^{(s)}/\alpha_3^{(s)} \to -1$ as $s \to \infty$, which is in agreement with the fact that α_1 and α_3 are the branches of the same branch point. Actually, Eqs. (6) show that the same critical position is obtained when using either $\alpha_1^{(N)}$ or $\alpha_3^{(N)}$ for large enough N values. On the other hand, it is expected that $A(\alpha_1) = -A(\alpha_3)$ because of the two signs of the square root (branch point of order 1).

The GFM sequences obtained from Eqs. (6) are shown in Table IV. Since they converge in a stepwise manner, appropriate subsequences must be carefully chosen before extrapolating. On doing this we obtain $x_1 = 1.931\,40 \pm 5\times10^{-5}$ and $|y_1| = 3.237\,65 \pm 5\times10^{-5}$, which closely agree with the exact result.

The A value coming from Eq. (7) is not accurate enough because the B_N sequence converges rather slowly. To improve it we calculate the u, r^2, b, and B values leading to the minimum of $S_{M,N}$ [Eq. (8)]. Only the $B_{M,N}$ sequences are considered because the other ones lead to nearly the same results shown above. From the two-entry (M,N) table in Table V we can estimate $A(\alpha_3) = -A(\alpha_1) = 1.100 \pm 0.003$.

When $j = 2,4,\ldots$, $\alpha_{2j+1}^{(s)}/\alpha_{2j+3}^{(s)}$ does not approach -1 as s increases, suggesting that the singular points for these states must be different from the previous ones. In fact, the CP sequences obtained from the GF (4) (with b variable) are not found to be convergent. This may be due to the occurrence of interferent singularities; i.e., other singular points close to the convergence circle. It is not unreasonable to assume that there could be two pairs of conjugate branch points of equal order and with nearly the same absolute value. If this were true an appropriate GF would be

$$Y(z) = B(C_0 + C_1 z + C_2 z^2 + C_3 z^3 + z^4)^b, \tag{30}$$

whose Taylor coefficients obey

$$(n+1)C_0 Y_{n+1} + (n-b)C_1 Y_n + (n-2b-1)C_2 Y_{n-1} + (n-3b-2)C_3 Y_{n-2} + (n-4b-3)Y_{n-3} = 0. \tag{31}$$

Upon replacing Y_n by $\alpha_5^{(n)}$ in Eq. (31) with $n = N, N+1, N+2, N+3$ we are led to a set of linear equations from which we can obtain the coefficients C_i as functions of b. Although the sequences $C_{iN}(b)$ are oscillating, they are found to be convergent. We have tried several b values but the smoothest sequences appear to be those for $b = \frac{1}{2}$. Their limits, estimated from first 35 perturbation corrections, are $C_0 = (1.80 \pm 0.02)\times10^4$, $C_1 = -(1.480 \pm 0.002)\times10^3$, $C_2 = (1.380 \pm 0.004)\times10^2$, and $C_3 = -12.55 \pm 0.03$. It is worth noticing that the coefficients of z^3 and z^4 in Eq. (30) are much smaller than the other ones, showing that the polynomial in parentheses does not differ too much from a quadratic one near the origin. The third- and fourth-degree terms take into account the effect of the interferent singularities.

TABLE IV. CP sequences for the solutions of period 2π of the Mathieu equation.

N	u_N	$(r_N^2 - u_N^2)^{1/2}$	$-B_N$
31	1.931 539 65	3.237 932 16	1.127 5851
32	1.931 511 26	3.237 899 23	1.029 6439
33	1.931 519 62	3.237 874 09	1.092 8564
34	1.931 509 40	3.237 876 34	1.124 5137
35	1.931 487 88	3.237 853 45	1.018 6699
36	1.931 493 60	3.237 833 36	1.091 3069
37	1.931 487 51	3.237 834 98	1.121 7909
38	1.931 470 81	3.237 818 71	1.004 3702
39	1.931 474 70	3.237 802 40	1.089 8195

The closest singular points to the origin of α_5, obtained as the roots of the polynomial in Eq. (30), are approximately given by $z_{1,2} = -3.385 \pm 10.725\ i$ and $z_{3,4} = 9.660 \pm 7.000\ i$, where $|z_{1,2}| \simeq 11.246$ and $|z_{3,4}| \simeq 11.930$. It is clear that the interferent singularities, namely z_3 and z_4, are quite close to the convergence region. Other b values lead to more strongly oscillating sequences but the singularity positions are not substantially altered. We cannot therefore be sure of the actual critical exponent value.

When replacing Y_n by $\alpha_7^{(n)}$ in Eq. (31), nonconvergent sequences C_{iN} are found, which suggests that the singularity pattern may be more complex. We will not go on discussing the characteristic values of the Mathieu function because it is not the aim of the present paper. However, since the subject is of great theoretical and practical interest, a more detailed description will be published elsewhere in a forthcoming paper, which will render the basis for a rigorous mathematical investigation.

Another physically interesting problem is the Stark effect in a polar rigid symmetric-top molecule. The stationary Schrödinger equation in appropriate units can be reduced to[24]

$$\left\{ -\sin^{-1}\theta \frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + (M^2 + K^2)\sin^{-2}\theta - 2KM\cos\theta\sin^{-2}\theta - q\cos\theta \right\}\psi_{K,M,J} = E_{K,M,J}\psi_{K,M,J}, \tag{32}$$

where θ is the angle between the dipole moment and the electric field, $E_{K,M,J}$ and q are proportional to the energy and

TABLE V. (M,N) table for the critical amplitude of the lowest eigenvalues of the Mathieu equation.

N \ $N-M$	3	4	5	6
24	1.099 214	1.089 958	1.103 390	1.101 085
25	1.129 822	1.107 666	1.099 050	1.113 317
26	1.063 091	1.106 629	1.099 113	1.088 068
27	1.097 271	1.088 452	1.102 008	1.099 147
28	1.126 868	1.106 112	1.098 659	1.112 094
29	1.060 146	1.107 426	1.097 904	1.087 485

field intensity, respectively, and $J = 0,1,2,\ldots$, $|K| = 0,1,\ldots,J$, and $|M| = 0,1,\ldots,J$ (see Ref. 24).

The singularities of $E_{0,M,J}(q)$ and those of the characteristic values of α_{2j} and β_{2j} of Eq. (29) are similar and some of them recently have been calculated accurately.[22,25] [Notice that when $K = 0$, Eq. (32) becomes that corresponding to a linear rotator.] Numerical evidence suggests that, for each M value, $E_{0,M,|M|+2j-2}$ and $E_{0,M,|M|+2j-1}$, $j = 1,2,\ldots$, have a common real value for a purely imaginary q value (say q_j).

On the other hand, no result has been reported regarding the singular points of $E_{K,M,J}(q)$ when $K \neq 0$ but it is not unreasonable to think that they can resemble those for the characteristic values α_{2j+1} and β_{2j+1} of Eq. (29). This proves to be the case as shown below.

A very large number of perturbation coefficients for this problem was reported by Røeggen some years ago.[26] Unfortunately, they are not accurate enough for our purposes and we have had to recalculate them by means of the equations in the Appendix. The required matrix elements can be found, for example, in Ref. 24. Due to our rather limited computational facilities we were not able to handle more than 25 perturbation coefficients.

Results for some states with positive K and M values are shown in Table VI. The singular points for other eigenvalues can also be obtained from Table VI just remembering that

$$E_{K,M,J}(q) = E_{M,K,J}(q) = E_{-M,-K,J}(q) = E_{-K,-M,J}(q)$$
$$= E_{-K,M,J}(-q) = E_{K,-M,J}(-q)$$

(see Ref. 24). Present numerical investigation is accurate enough to enable us to suggest the following properties for the eigenvalues of Eq. (32).

(a) For each pair of K and M values, the eigenvalues with $J = |M|$ and $J = |M| + 1$ have a common value at a branch point of order 1. For larger J values the singularity pattern is more complex as in the case of the characteristic values α_{2j+1} and β_{2j+1} of the Mathieu equation when $j > 1$.

(b) When $J = |M|$ and $J = |M| + 1$ the critical amplitude is real and appears to be K independent. Its absolute value decreases as M increases.

(c) The convergence radius of the perturbation series increases slowly with K and strongly with M.

A systematic study of the singular points of the eigenvalues of Eq. (32), using a larger number of Taylor coefficients and improved GF 's will be published elsewhere in a forthcoming paper.

TABLE VI. Singular points of the eigenvalues $E_{K,M,M}$ and $E_{K,M,M+1}$, where $K, M = 1,2,3$ and 4.

(K, M, J)	$-x_1$	z_1	A
(1,1,1)	2.784 ± 0.002	5.341 ± 0.001	−0.39 ± 0.01
(1,1,2)	2.783 ± 0.001	5.3407 ± 0.0003	0.390 ± 0.006
(1,2,2)	3.659 ± 0.005	10.347 ± 0.003	−0.34 ± 0.02
(1,2,3)	3.660 ± 0.003	10.347 ± 0.003	0.34 ± 0.02
(2,2,2)	7.019 ± 0.001	10.115 ± 0.005	−0.36 ± 0.03
(2,2,3)	7.022 ± 0.003	10.112 ± 0.002	0.36 ± 0.03
(1,3,3)	4.51 ± 0.01	16.80 ± 0.01	−0.28 ± 0.02
(1,3,4)	4.51 ± 0.01	16.80 ± 0.01	0.28 ± 0.02
(2,3,3)	8.80 ± 0.02	16.58 ± 0.01	−0.29 ± 0.02
(2,3,4)	8.81 ± 0.01	16.59 ± 0.01	0.29 ± 0.02
(3,3,3)	12.58 ± 0.01	16.14 ± 0.01	−0.31 ± 0.02
(3,3,4)	12.60 ± 0.05	16.15 ± 0.05	0.30 ± 0.02
(1,4,4)	5.32 ± 0.01	24.66 ± 0.02	−0.25 ± 0.01
(1,4,5)	5.29 ± 0.02	24.64 ± 0.02	0.25 ± 0.02
(2,4,4)	10.50 ± 0.02	24.43 ± 0.02	−0.25 ± 0.02
(2,4,5)	10.5 ± 0.1	24.46 ± 0.03	0.25 ± 0.02
(3,4,4)	15.19 ± 0.03	23.96 ± 0.05	−0.25 ± 0.02
(3,4,5)	15.0 ± 0.2	24.1 ± 0.2	0.25 ± 0.02
(4,4,4)	19.40 ± 0.01	23.40 ± 0.01	−0.28 ± 0.02
(4,4,5)	19.45 ± 0.05	23.40 ± 0.07	0.27 ± 0.03

VII. CONCLUSIONS

The GFM seems to be very promising in obtaining the closest singular points to the origin from Taylor series. Any sort of singularities can be dealt with provided the behavior of the function in its neighborhood is approximately known. Since the GFM takes into account the form and number of singular points nearest to the origin explicitly, it proves to be preferable to other techniques. The more we know about the singularities the more accurate the GF that can be used and the larger the convergence rate of the CP sequences.

The singularity pattern for the eigenvalue problems in Sec. VI is quite interesting since it is very different from those discussed previously (cf. Sec. V and Refs. 21, 22, and 25). A more detailed numerical investigation and a rigorous mathematical study would be of great value due to the physical importance of both models.

During the last years there has been a great deal of interest in perturbation expansions for some bounded systems[18] (and references therein). Upper bounds to their convergence radii have been estimated[27] that can, in principle, be checked very easily by means of the GFM.

The GFM also can be useful in studying critical phenomena in spin-lattice models.[7,8] In this case some real singular points of the thermodynamic functions are found to be related to phase transitions and sometimes complex singularities occur that interfere with their calculation.[7]

Note added in proof: After the present article was sent to press we came across the papers by Blanch and Clemm[28] and Hunter and Guerrieri,[29] where the branch points of the eigenvalues for all the solutions of the Mathieu equation are fully discussed.

APPENDIX

For the sake of completeness we will develop here a standard large-order nondegenerate perturbation theory that is necessary to deal with the problems in Sec. VI. To this end let us consider the eigenvalue equation

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \tag{A1}$$

where $H = H_0 + \lambda V$ is a Hermitian operator for all $0 \leqslant \lambda < \infty$ and the eigensolutions of H_0 are known; i.e.,

$$H_0|i\rangle = E_i^{(0)}|i\rangle. \tag{A2}$$

It is supposed that $E_i^{(0)} \neq E_j^{(0)}$ if $i \neq j$ and that $\{|i\rangle\}$ is a

complete set of orthonormal vectors.

On expanding $|\psi_n\rangle$ as

$$|\psi_n\rangle = C_{0n}|0\rangle + C_{1n}|1\rangle + \cdots, \tag{A3}$$

it is found that Eq. (A1) becomes

$$\lambda \sum_{i=0}^{\infty} V_{ji} C_{in} = (E_n - E_j^{(0)}) C_{jn}, \tag{A4}$$

where $V_{ji} = \langle j|V|i\rangle$. If C_{nn} is arbitrarily chosen equal to unity, reflecting the fact that $|\psi_n\rangle(\lambda = 0) = |n\rangle$, and E_n and C_{in} are expanded in powers of λ,

$$E_n = \sum_{s=0}^{\infty} E_n^{(s)} \lambda^s, \quad C_{in} = \sum_{s=0}^{\infty} C_{in}^{(s)} \lambda^s, \tag{A5}$$

we obtain

$$C_{jn}^{(p)} = (E_n^{(0)} - E_j^{(0)})^{-1} \left\{ \sum_{i=0}^{\infty} V_{ji} C_{in}^{(p-1)} - \sum_{s=0}^{p-1} E_n^{(p-s)} C_{jn}^{(s)} \right\}, \quad j \neq n, \tag{A6a}$$

$$E_n^{(p)} = \sum_{i=0}^{\infty} V_{ni} C_{in}^{(p-1)}. \tag{A6b}$$

When $V_{ij} = 0$ for $|i-j| > I$, Eqs. (A6) and the starting point $C_{jn}^{(0)} = \delta_{jn}$ enable one to obtain a very large number of perturbation corrections exactly. For the two examples in Sec. VI, $I = 2$.

[1] E. Borel, *Leçons sur les Series Divergentes* (Gauthier-Villars, Paris, 1923).
[2] F. Rellich, *Perturbation Theory of Eigenvalue Problems* (Gordon and Breach, New York, 1969).
[3] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976).
[4] A. Pellegati, J. Čížek, and J. Paldus, Int. J. Quantum Chem. **21**, 147 (1982).
[5] J. N. Silverman, Phys. Rev. A **23**, 441 (1981).
[6] P. R. Certain and W. Byers-Brown, Int. J. Quantum Chem. **6**, 131 (1972).
[7] *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1974), Vol. 3.
[8] *Proceedings of the 1980 Cargèse Summer Institute on Phase Transitions*, edited by M. Levy, J. C. Le Guillou, and J. Zinn-Justin (Plenum, New York, 1982).
[9] F. H. Ree and W. E. Hoover, J. Chem. Phys. **46**, 4181 (1967).
[10] A. Baram and M. Luban, J. Phys. C **12**, L 659 (1979).
[11] V. C. Aguilera-Navarro, M. Fortes, M. de Llano, and A. Plastino, J. Chem. Phys. **76**, 749 (1982).
[12] M. Alexanian, Phys. Rev. B **31**, 7160 (1985).
[13] A. J. Guttmann, J. Phys. C **2**, 1900 (1969).
[14] A. I. Markushevich, *Theory of Functions of a Complex Variable* (Prentice–Hall, New York, 1967), Vol. III, pp. 276–294.
[15] G. A. Arteca, F. M. Fernández, and E. A. Castro, J. Chem. Phys. (in press).
[16] F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill, New York, 1956), pp. 458–462.
[17] V. C. Aguilera-Navarro and M. C. K. Aguilera-Navarro, J. Comp. Phys. **53**, 193 (1984).
[18] F. M. Fernández and E. A. Castro, Int. J. Quantum Chem. **19**, 521 (1981).
[19] D. A. Atkinson and H. W. Crater, Am. J. Phys. **43**, 301 (1975).
[20] G. A. Arteca, F. M. Fernández, and E. A. Castro, Nuovo Cimento B **72**, 246 (1982).
[21] N. W. McLachlan, *Theory and Application of the Mathieu Functions* (Oxford U. P., Oxford, 1947).
[22] F. M. Fernández and E. A. Castro, Phys. Lett. A **107**, 215 (1985).
[23] S. A. Maluendes, A. M. Mesón, F. M. Fernández, and E. A. Castro, "Singular points for some periodic eigenvalue problems" (unpublished).
[24] J. H. Shirley, J. Chem. Phys. **38**, 2896 (1963).
[25] S. A. Maluendes, F. M. Fernández, and E. A. Castro, J. Chem. Phys. **83**, 4599 (1985).
[26] I. Røeggen, Atomic Data **4**, 289 (1972).
[27] V. C. Aguilera-Navarro, J. F. Gromes, A. H. Zimerman, and E. Ley-Koo, J. Phys. A **16**, 2943 (1983).
[28] G. Blanch and D. S. Clemm, Math. Comp. **23**, 97 (1969).
[29] C. Hunter and B. Guerrieri, Stud. Appl. Math. **64**, 113 (1981).

Fractional approximation to elliptic functions

Kley Visentin and Pablo Martin
Departamento de Física, Universidad Simón Bolívar Apartado 80659, Caracas 1080-A, Venezuela

(Received 9 February 1986; accepted for publication 1 October 1986)

The Jacobi functions $\mathrm{sn}(x/m)$ have been approximated by a quotient of polynomials of first to fourth degrees. The method used here is an extension to nonlinear differential equations of one previously published for first- and second-order linear differential equations. That method uses power series and asymptotic expansions simultaneously. The accuracy here obtained is very good (the absolute error is lower than 10^{-8} for $m<0.8$) except for values of the parameter m near 1. They are much better for several orders of magnitude than those based on the Padé method for the same number of parameters to be determined.

I. INTRODUCTION

A method to obtain fractional approximations that uses power series and asymptotic expansions simultaneously has been published recently.[1-3] One of the main differences with other methods is that this method requires the use of a suitable variable[3] instead of any independent variable as customarily used. The method has been applied to functions defined by linear differential equations.[1-3]

In this paper we consider the Jacobi functions of the first kind, $\mathrm{sn}(x/m)$, which are defined by a first-order differential equation of second degree. In the case of nonlinear differential equations the choice of a suitable independent variable becomes more crucial, and our analysis shows that the variable is unique and is obtained from a differential equation derived from the original one. Once the suitable variable is defined, the approximations attain very good accuracy. With only a fourth-degree polynomial we can obtain an approximation of greater accuracy than the usual table of Jacobi functions for any value of the parameter m, provided m is not very close to 1 ($m<0.92$).[4] We have compared our approximations with those obtained by the Padé method[5] and our accuracy is much better for the same order of approximation.

We have arranged the material of this paper in the following way. In Sec. II, we detail how the suitable variable is obtained. The procedure leads to an auxiliary differential equation of the Riccati type, whose solution gives the adequate independent variable. Fractional approximations in this variable are calculated in Sec. III. Each coefficient of the approximation is determined as a function of the module m and a quarter of period $K(m)$ of the Jacobi functions. The accuracy of the approximations from first to fourth degree is computed in Sec. IV. The degree of the polynomial denominator is always one degree higher than that of the numerator. At least six exact digits are obtained for the fourth-degree approximation for $m<0.92$. This shows that the approximations give the accuracy of the usual Jacobi tables[4] and the computations require only a desk calculator. The graphs in this section will help to choose the adequate approximation, for the accuracy needed.

The last section of this paper is devoted to the conclusions and the discussion of the main results.

II. SUITABLE VARIABLE FOR ELLIPTIC FUNCTIONS

The differential equation for elliptic functions of the first kind is

$$\left(\frac{dy}{dx}\right)^2 = 1-(1+m)y^2+my^4\,, \tag{1}$$

where $y=\mathrm{sn}(x/m)$ and $m=k^2$ is the characteristic parameter of the elliptic function. The function y has a period $4K$, where K is a function of m.

Suppose we approximate y by the quotient of polynomials of degrees n and l, respectively, then

$$y\approx P_n(x)/Q_l(x)\,, \tag{2}$$

$$\frac{dy}{dx}=\frac{P_n'Q_l-Q_l'P_n}{Q_l^2}\,. \tag{3}$$

By substituting into (1) and rationalizing, we obtain

$$(P_n'Q_l-Q_l'P_n)^2=Q_l^4-(1+m)P_n^2Q_l^2+mP_n^4\,. \tag{4}$$

Here the degree of each of the terms are, respectively, $2(n+l-1), 4l, 2(l+n), 4n$. Since in our method we want to compare the highest powers, we should have the higher degree in at least two terms. We should analyze the possibilities $l\lesseqgtr n$; however, since the results are the same in the three cases, we will analyze in detail only the case $l>n$. In this case the largest degree on the right-hand side is $4l$ and the difference of the degrees on both sides of the equation is $2(l-n)+2$. Since $l-n$ is larger than zero, the least difference will be 4.

In order to equalize the degrees of both sides we have to change the variables, thus arranging that the degree of the left-hand side of the Eq. (4) increases by four units when we arrive to the fractional approximation. By denoting the new variable by t, we should have

$$t=t(x)\,,\quad y(x)=u(t)\,. \tag{5}$$

From our previous considerations dt/dx should be a second-order polynomial in t

$$\frac{dt}{dx}=\mu^2(t+\alpha)^2+\beta^2\,, \tag{6}$$

where μ, α, and β are coefficients to be determined. In this way, t is obtained from a Riccati equation with constant coefficients. The solution will be trigonometric or a hyperbo-

 0022-2488/87/020330-04$02.50 © 1987 American Institute of Physics

lic function. Since we want to keep the periodicity of the elliptic function, we consider only the trigonometric solution, that is, we take μ real. One of the adequate solutions to our problem will be

$$t+\alpha=(\beta/\mu)\tan[\beta\mu(x+c)]\,. \tag{7}$$

In order to apply our method to fractional approximations we should use the interval $(0,\infty)$ for the new independent variable, this the interval $(0,2K)$ for x is mapped to the interval $(0,\infty)$ for t, and the same assumption for the interval $(2K,4K)$. Thus, we have that α and c must be zero, and

$$\beta\mu=\frac{\pi}{4K},\quad t=\frac{4K}{\pi}\beta\tan\left(\frac{\pi}{4K}x\right). \tag{8}$$

We shall see later [Eq. (25)] that if we choose $\beta=1$, the approximation exactly reproduces the Jacobi functions $\mathrm{sn}(x/m)$ for the parameter $m=0$, that is, $\sin(x)$. With this selection we have defined the suitable variable as

$$t=(1/\mu)\tan(\mu x),\quad \mu=\pi/4K\,. \tag{9}$$

Using this change of variable, Eq. (1) becomes

$$(1+\mu^2t^2)^2\left(\frac{du}{dt}\right)^2=1-(1+m)u^2+mu^4\,. \tag{10}$$

III. FRACTIONAL APPROXIMATIONS

Once Eq. (10) has been obtained we can proceed to obtain a direct approximation of u, writing it as a quotient of polynomials $\hat{P}_n(t)/\hat{Q}_l(t)$. The boundary conditions in our case are that $y(x)$ is zero for $x=0$ and $x=2K$, which means that $u(t)$ is zero for $t=0$ and $t=\infty$. From the first condition the independent term of $\hat{P}_n(t)$ is zero. From the second condition the degree of $\hat{Q}_l(t)$ must be higher than the degree of $\hat{P}_n(t)$.

The structure of the polynomial $\hat{P}_n(t)$ and the symmetry of $\mathrm{sn}(x/m)$ suggest an additional change of variable, which simplifies the calculation of the approximations

$$t=z^2\,, \tag{11}$$

$$v(z)=u(t)/t\,. \tag{12}$$

Equation (10) is now transformed to

$$(1+\mu^2z)^2\left(v+2z\frac{dv}{dz}\right)^2=1-(1+m)zv^2+mz^2v^4\,. \tag{13}$$

Now we replace $v(z)$ by a fractional approximation

$$v(z)=P_n(z)/Q_l(z)\,. \tag{14}$$

Equating now for the degree of the highest power in both sides of Eq. (13) we obtain $2+2(n+1)=4l$, that is, $l=n+1$.

The recursion relation for the coefficients of the power series for v, i.e.,

$$v(z)=\sum_{k=0}^{\infty}a_kz^k\,, \tag{15}$$

will be

$$\begin{aligned}2(1+2k)a_0a_k=&-\sum_{i=0}^{k-1}[1+m+2\mu^2(1+2i)(2k-1-2i)]a_ia_{k-1-i}-\sum_{i=1}^{k-1}(1+2i)(1+2k-2i)a_ia_{k-i}\\&+\sum_{i=0}^{k-2}\left\{m\left(\sum_{r=0}^{i}a_ra_{i-r}\right)\left(\sum_{s=0}^{k-2-i}a_sa_{k-2-i-s}\right)-\mu^4(1+2i)(2k-2i-3)a_ia_{k-2-i}\right\},\quad n\geqslant 2\,,\end{aligned} \tag{16}$$

and for $n=0,1$

$$a_0^2=1\,,\quad 6a_0a_1=-(1+m)a_0^2-2\mu^2a_0^4\,. \tag{17}$$

According to Jacobi's power series, a_0 must be 1.

In Table I the first four coefficients of the power series are given, these will be needed for the best approximation analyzed in this paper.

The asymptotic expansion is obtained by inserting

$$v(z)=\frac{1}{z}\sum_{k=0}^{\infty}b_kz^{-k} \tag{18}$$

in Eq. (13). However, this is more easily obtained from the preceding power series by noting that a transformation of the type

$$s=1/\mu^4z\,,\quad w(s)=v(z)/\mu^2s \tag{19}$$

leaves Eq. (13) invariant. Therefore

$$b_k=(1/\mu^2)(a_k/\mu^{4k})\,. \tag{20}$$

The approximation will be

$$\hat{v}(z)=\frac{\sum_{j=0}^{n-1}p_jz^j}{1+\sum_{j=1}^{n}q_jz^j}, \tag{21}$$

where $n=1,2,3,\ldots$.

Here the p's and q's will be obtained from the equality of the coefficients

$$\sum_{j=0}^{n-1}p_jz^j\cong\left(\sum_{k=0}^{\infty}a_kz^k\right)\left(1+\sum_{j=1}^{n}q_jz^j\right), \tag{22}$$

$$\sum_{j=0}^{n-1}p_{n-j-1}\left(\frac{1}{z}\right)^j\cong\left(\sum_{k=0}^{\infty}b_k\left(\frac{1}{z}\right)^k\right)\sum_{j=0}^{n-1}q_{n-j}\left(\frac{1}{z}\right)^j+\frac{1}{z^n}\Bigg). \tag{23}$$

We have to obtain $2n+2$ equations. Our analysis shows that the highest accuracy is obtained when we choose equal number of coefficients in Eq. (22) and Eq. (23) ($n+1$ coef-

TABLE I. First four coefficients of the power series as a function of m and μ.

$$a_0=1$$

$$a_1=-\tfrac{1}{6}(1+m+2\mu^2)$$

$$a_2=\frac{1}{120}(1+m)^2+\frac{\mu^2}{6}(1+m)+\frac{m}{10}+\frac{\mu^4}{5}$$

$$a_3=-\frac{1}{5020}(1+m)^3-\left[\frac{11}{30}+\frac{7\mu^2}{36}\right]\frac{(1+m)^2}{14}+\left[\frac{11}{30}-\frac{7}{3}\mu^2-\frac{98}{45}\mu^4\right]\frac{(1+m)}{14}+\frac{7}{3}\mu^2-2\mu^6$$

ficients from each equation). In this case the results are given in Table II for the first four values of n.

It is important to point out that for $m = 0$ the approximation reduces to

$$v(z) = 1/(1 + z/2) . \tag{24}$$

This result has been verified for all of the orders n here analyzed. When Eq. (24) is expressed in terms of the original variable x, it becomes

$$y(x) = \frac{2\tan(x/2)}{1 + \tan^2(x/2)} = \sin(x) . \tag{25}$$

This corresponds to the exact value of $\mathrm{sn}(x/m)$ for $m = 0$. Thus the approximation becomes the exact function. This is a consequence of our choice of $\beta = 1$ [see Eq. (8)].

IV. RESULTS

We have analyzed the approximations for values of n from 1–4 and for any value of the module m. For a given n the best results are obtained when we take the same numbers of terms from the asymptotic expansion as from the power series. The values of this p's and q's for three cases are given in Table II as a function of the coefficients a's given in Table I.

The largest error always happens for $x = K(m)$. An example of this is shown in Fig. 1 for $m = 0.75$ ($n = 2, 3$). We have shown only a half-period, because of the symmetry of the function. The errors near $x = 0$ and $x = 2K$ are very small. The maximum absolute errors are 0.72×10^{-2} and 0.17×10^{-4} for $n = 2, 3$, respectively, which correspond to relative errors of 0.33% and 0.0008%, respectively.

TABLE II. Fractional parameters for first- to fourth-degree approximations as a function of μ and the power series coefficients.

$n = 1$	
$p_0 = 1$	$q_1 = \mu^2$
$n = 2$	
$p_0 = 1$	$q_1 = \mu^2 - a_1$
$p_1 = \mu^2$	$q_2 = \mu^4$
$n = 3$	
$p_0 = 1$	$q_1 = \mu^4 - a_1$
$p_1 = \dfrac{\mu^4 + a_1\mu^2 + a_1^2 - a_2}{\mu^2 + a_1}$	$q_2 = \mu^2\left[\dfrac{\mu^4 - a_2}{\mu^2 + a_1}\right]$
$p_2 = \mu^4$	$q_3 = \mu^6$
$n = 4$	
$p_0 = 1$	
$p_1 = \dfrac{-\mu^6 + a_1\mu^4 + 2a_1^2\mu^2 - a_1^3 + a_3}{\mu^4 + a_1\mu^2 + a_2 - a_1^2}$	
$p_2 = \mu^2\left[\dfrac{-\mu^6 + a_1\mu^4 + 2a_1^2\mu^2 - a_1^3 + a_3}{\mu^4 + a_1\mu^2 + a_2 - a_1^2}\right]$	
$p_3 = \mu^6$	
$q_1 = \dfrac{-\mu^6 + a_1^2\mu^2 + a_3 - a_1a_2}{\mu^4 + a_1\mu^2 + a_2 - a_1^2}$	
$q_2 = \dfrac{-\mu^8 + a_2\mu^4 + (a_3 - a_1a_2)\mu^2 + a_2^2 - a_1a_3}{\mu^4 + a_1\mu^2 + a_2 - a_1^2}$	
$q_3 = \mu^4\left[\dfrac{-\mu^6 + a_1^2\mu^2 + a_3 - a_1a_2}{\mu^4 + a_1\mu^2 + a_2 - a_1^2}\right]$	
$q_4 = \mu^8$	

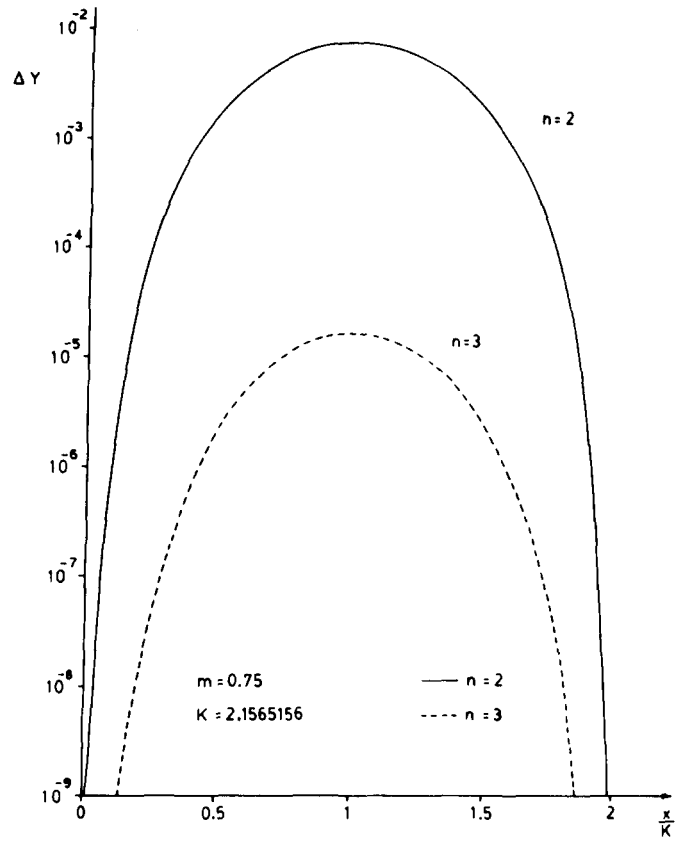


FIG. 1. Semilog plot of absolute error $\Delta Y = y - \hat{y}$ for second- and third-degree approximations for $m = 0.75$ as a function of x/K in the interval (0, 2).

The maximum error, as a function of the module m, is shown in Fig. 2 (for $n = 1, 2$) and Fig. 3 (for $n = 3, 4$). The maximum error increases with m, and the approximation is very poor near $m = 1$. The main problem for $m = 1$ is that the solution is $\tanh(x)$, which is not periodic. For $m = 1$, $\mu = 0$ and $b_0 = \infty$ and the approximation is not well defined.

The accuracy increases quickly with n. For instance, if $n = 2$ the maximum error is less than 10^{-4} for $m < 0.3$. Meanwhile, for $m < 0.3$ and $n = 3$ the maximum error is smaller than 10^{-8}. For $n = 4$ the accuracy is better than 10^{-8} for $m < 0.8$.

Our nth-order approximation can be compared with the nth-order main diagonal Padé approximation M_n for $\mathrm{sn}(x/m)$ given in Table 10.1, p. 91 of Ref. 5. Since Padé approximations are not periodic we compare the maximum errors in the first quarter of period, that is, for x in the interval $(0, K(m))$, and in particular for values $m = 0.4$ and $m = 0.6$.

Considering $m = 0.4$ our approximations give at least three, six, and ten exact digits for $n = 2$, 3, and 4, respectively, compared with two, four, and five exact digits for the Padés M_2, M_3, and M_4. Considering $m = 0.6$, the maximum error for $n = 2$ is about the same (0.0019 in our case com-

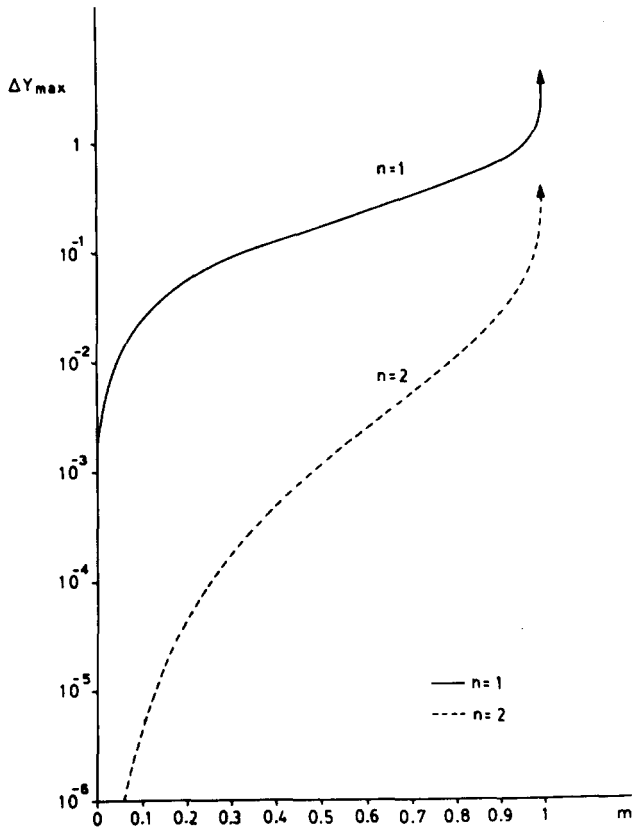


FIG. 2. Semilog plot of maximum absolute error ΔY_{max} for first- and second-degree approximations as a function of the parameter m.

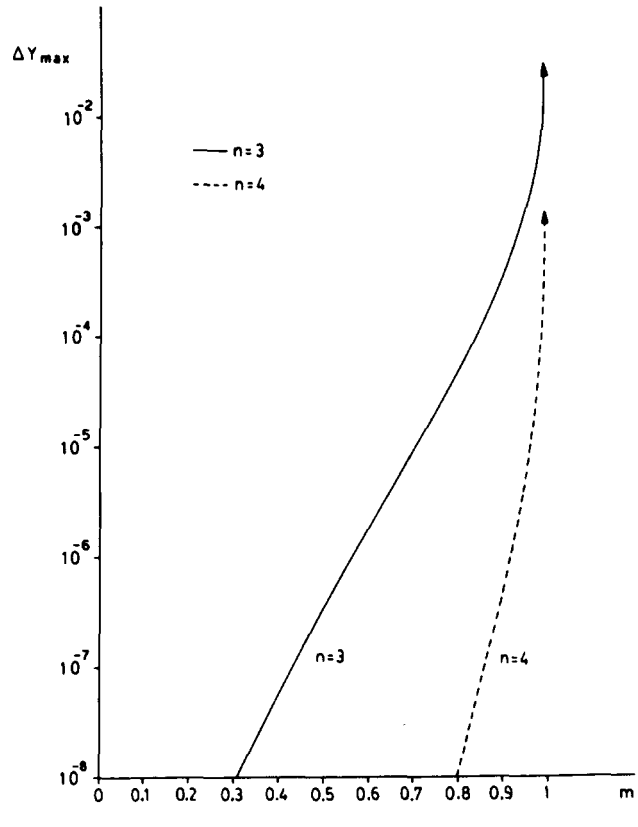


FIG. 3. Semilog plot of maximum absolute error ΔY_{max} for third- and fourth-degree approximations as a function of the parameter m.

pared with 0.0013). For $n = 3$ and 4 we get six and nine exact digits, respectively, compared with three and five exact digits in the Padé case. Therefore the accuracy of our approximations is much better than in the Padé case, and in addition we also obtain the characteristic periodic behavior of the Jacobi function.

V. CONCLUSION

We have shown how the recent methods of fractional approximations can be extended to some nonlinear differential equations by using a suitable change of variable. The adequate change of variable is determined by an auxiliary differential equation derived from the original one. For Jacobi functions the suitable variable is $\tan(\mu x)/\mu$ and the auxiliary differential equation is of the Riccati type. The Jacobi function $\mathrm{sn}(x/m)$ has been approximated by a quotient of polynomials of degrees 1 to 4. The accuracy is in general very good for any value of the module m not too close to unity ($m < 0.92$). For $m = 0$ the approximation reproduces the exact Jacobi function ($\mathrm{sn}(x/0) = \sin(x)$). For $m = 1$ the Jacobi function loses its periodicity and the suitable variable will be different. The accuracy in the fourth-degree case is at least of six digits exacts for any value of x if $m < 0.92$. In most of the cases the accuracy of our approximation is much better in several orders of magnitude than those using the Padé method.

ACKNOWLEDGMENT

We thank Professor Antonio L. Guerrero for his useful assistance in the computation involved in this work.

[1]P. Martín and A. Guerrero, J. Math. Phys. **26**, 705 (1985).
[2]E. Chalbaud and P. Martín, J. Math. Phys. **25**, 1268 (1984).
[3]P. Martín and A. L. Guerrero, Scientia XLVII, 123 (1982); Proc. Int. Conf. Plasma Phys. **1**, 325 (1982).
[4]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[5]Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. II, pp. 90–91.

Quadratures for self-dual GL(2,C) Yang–Mills fields

S. Chakravarty
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

E. T. Newman
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260 and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

(Received 31 July 1986; accepted for publication 24 September 1986)

It is the purpose of this paper to show that the GL(2,C) Yang–Mills equations can be solved in terms of integrals over the characteristic initial data. The method is based on showing that enough gauge freedom exists in the choice of characteristic initial data so that the data can always be put into either upper or lower triangular form. With triangular form data the Sparling equation (a linear first-order equation equivalent to the self-dual Yang–Mills equations) can be solved by explicit quadratures.

I. INTRODUCTION

The self-dual (or anti-self-dual) Yang–Mills equations have for a variety of reasons[1–5] been extensively studied over the past several years. Of the many solution generating techniques, two methods seem to stand out; namely (1) the use of the Sparling equation, a first-order linear matrix-valued differential equation, equivalent to the self-dual Yang–Mills equations and (2) the use of twistor theory via the solution of a Riemann–Hilbert problem. Recently[6] it was shown that the solution of the Sparling equation was identical to the solution of the Riemann–Hilbert (RH) problem.

It is the purpose of this work to show that (at least) in the case of GL(2,C), the Sparling equation (or the RH problem) can be solved explicitly in terms of quadratures over the characteristic data. (It has been pointed out to us by L. Mason and M. Hickman that our method probably does not allow solutions for arbitrary initial data, although our solutions probably do form a dense set and an arbitrary solution may be approximated by one of our solutions.) The method presented here is applicable only to globally regular fields. It, however, can be generalized to local fields.

In Sec. II we give a brief discussion of the Sparling equation and its connection with the twistor formulation of the RH problem. We also show how to obtain the self-dual Yang–Mills connection and field from the solution to the Sparling equation. Section III deals with the solution of the Sparling equation for the cases when the characteristic data (which "drives" the Sparling equation) is in either upper or lower triangular form for 2×2 matrices. In Sec. IV we discuss the "gauge" transformations on the Sparling equation and its solutions which leave the Yang–Mills connection invariant and further show how to exploit this gauge freedom to triangularize any 2×2 matrix-valued characteristic data. This then implies that any arbitrary, (2×2), characteristic data can be transformed to equivalent triangular data—whose associated Sparling equation was discussed in Sec. III. The triangularization procedure requires that the transformation be regular and certain ratios of components satisfy the Riccati equation. The remainder of the section is devoted to showing the existence of appropriate solutions to the Riccati equation with the correct singularity structure so that the required regularity conditions on the transformation are satisfied.

Finally in Sec. IV we illustrate our triangularization method to a special class of data that is nilpotent and trace-free. The importance of this class lies in the fact that it illustrates why the single instanton solutions are so easily found.

II. DISCUSSION OF SPARLING'S EQUATION

A. Derivation and meaning

Consider a GL(2,C) bundle $V\times M$ with the connection $\gamma_a(x^a)$ over Minkowski space M. Let C_x be the future null cone of a point $x^a\in M$. Let ζ and $\bar\zeta$ label the null generators of the null cone at future null infinity, i.e., $\mathscr{I}^+$. A null geodesic on C_x is labeled by the same complex generators ζ and $\bar\zeta$ obtained by its intersection with $\mathscr{I}$ and is denoted by $l_x(\zeta,\bar\zeta)$. Let s be a normalized affine parameter along $l_x(\zeta,\bar\zeta)$ so that $l\equiv\partial/\partial s = l^a(\zeta,\bar\zeta)\partial/\partial x^a$ is tangent to $l_x(\zeta,\bar\zeta)$. Then one can define the GL(n,C) matrix-valued function $G(x^a,\zeta,\bar\zeta)$ to be the linear map that propagates in a parallel manner an arbitrary vector in the fiber over x^a to $\mathscr{I}^+$ along $l_x(\zeta,\bar\zeta)$,

$$G(x^a,\zeta,\bar\zeta) = O\exp\left(\int_{l_x(\zeta,\bar\zeta)}\gamma_a\,dx^a\right) = O\exp\left(\int_{l_x}\gamma_a l^a\,ds\right). \tag{2.1}$$

Now consider an infinitesimal loop formed by two neighboring null geodesics $l_x(\zeta,\bar\zeta)$ and $l_x(\zeta+d\zeta,\bar\zeta)$ with common origin x^a, and a connecting vector $M^a\,d\zeta$ at $\mathscr{I}^+$. This loop lies in an anti-self-dual two-blade. Parallel transport of a self-dual Yang–Mills field F_{ab}, around this loop will then give the identity. This follows from the fact that the projection of a self-dual field on an anti-self-dual blade vanishes. Expressed in terms of G, one obtains

$$\mathrm{I} = G^{-1}(\zeta,\bar\zeta)G(\zeta+d\zeta,\bar\zeta)(\mathrm{I}+A\,d\zeta/(1+\zeta\bar\zeta)), \tag{2.2}$$

where $\mathrm{I}+A\,d\zeta/(1+\zeta\bar\zeta)$ is the infinitesimal parallel propagator along the connecting vector on $\mathscr{I}^+$, $A=\gamma_a M^a = \gamma_a\,dx^a/d\zeta$ is the asymptotic component (along the connecting vector) of the connection form. Hence A, which is defined on $\mathscr{I}^+$, is a function of three-variables, i.e., $A\equiv A(u,\zeta,\bar\zeta)$. We are interested in the restriction of A to the intersection of C_x with $\mathscr{I}^+$ which is given by $u=u(x^a,\zeta,\bar\zeta) = x^a l_a(\zeta,\bar\zeta)$. We thus have $A_R(x^a,\zeta,\bar\zeta)\equiv A(x^a l_a,\zeta,\bar\zeta)$. By

 0022-2488/87/020334-05$02.50 © 1987 American Institute of Physics

expanding Eq. (2.2) one obtains $(1+\zeta\bar{\zeta})\partial G/\partial\zeta + GA_R = 0$; defining $(1+\zeta\bar{\zeta})\partial G/\partial\zeta = \delta G$, we finally obtain

$$\delta G = -GA_R, \tag{2.3}$$

which is Sparling's equation for self-dual fields.

B. Connection of Sparling's equation to twistor theory and the RH problem

The material in this section is not essential to the remainder of this work and can hence be omitted. It is nevertheless closely related. Since the details have been given elsewhere we only sketch the ideas here.

The (dual) twistor approach[1] to self-dual Yang–Mills theory begins with an arbitrary matrix-valued function of the three variables

$$\begin{aligned} w_0 &= (x^0 - x^3) + (x^1 + ix^2)\tilde{\zeta}, \\ w_1 &= (x^1 - ix^2) + (x^0 + x^3)\tilde{\zeta}, \quad \tilde{\zeta}, \end{aligned} \tag{2.4}$$

i.e.,

$$a(\tilde{\zeta}, w_0, w_1),$$

which for fixed x^a is to be holomophic in $\tilde{\zeta}$ on an annular region in the neighborhood of the equator on the Riemann sphere ($\tilde{\zeta}$) or extended complex plane ($\tilde{\zeta}$). The idea then is to find two matrix-valued functions $G_N(x^a,\tilde{\zeta})$ and $G_S(x^a,\tilde{\zeta})$ which are both holomorphic in the annular region and also, respectively, in the northern and southern hemispheres of the Riemann sphere, such that in the annular region

$$G_N G_S^{-1} = a(\tilde{\zeta}, w_0, w_1), \tag{2.5}$$

with (w_0,w_1) given by (2.4). This defines a classical RH "splitting" problem. From the knowledge of G_N or G_S one can construct the self-dual Yang–Mills field.

The above RH problem is related to the Sparling equation in the following manner.

The twistor function $a(\zeta,w_0,w_1)$ can be constructed from A_R in the following way:

$$a(\tilde{\zeta}, w_0, w_1) = P\exp\left(\int_0^\infty A_R \frac{d\zeta}{(1+\zeta\bar{\zeta})}\right), \tag{2.6}$$

with P denoting path order integration.

If we now solve the Sparling equation (2.3) with the condition that $G(x,\zeta,\bar{\zeta})$ be an analytic function in both ζ and $\bar{\zeta}$ in the neighborhood of $\tilde{\zeta} = \bar{\zeta}$, then the RH problem is solved by

$$G_N(x^a,\tilde{\zeta}) = G(x^a,0,\tilde{\zeta}), \quad G_S(x^a,\tilde{\zeta}) = G(x^a,\infty,\tilde{\zeta}). \tag{2.7}$$

Solving the Sparling equation with the regularity conditions is thus equivalent to "splitting" the twistor function $a(\tilde{\zeta},w_0,w_1)$ and thus solving the RH problem.

C. The connection and field from G

From the regular solutions of the Sparling equation, $G(x^a,\zeta,\bar{\zeta})$, it is easy to construct the connection one-form $\gamma_a(x)$ and hence the field $F_{ab}(x)$.

One has immediately from (2.1) that

$$l^a\nabla_a GG^{-1} = l^a\gamma_a(x). \tag{2.8}$$

[It follows from the identity

$$\delta^2(l^a\nabla_a GG^{-1}) = 0,$$

obtained from the Sparling equation, that the γ_a in (2.8) is independent of $(\zeta,\bar{\zeta})$.] By applying δ to (2.8) (again using the Sparling equation) we obtain

$$m^a\nabla_a GG^{-1} = m^a\gamma_a(x). \tag{2.9}$$

[We are using the null tetrad defined by l^a, $m^a = \delta l^a$, $\bar{m}^a = \bar{\delta} l^a$, $n^a = l^a + \delta\bar{\delta} l^a$, with

$$l^a = [1/\sqrt{2}(1+\zeta\bar{\zeta})](1+\zeta\bar{\zeta}, \zeta+\bar{\zeta}, i(\bar{\zeta}-\zeta), -1+\zeta\bar{\zeta}),$$

so that $l\cdot n = -m\cdot\bar{m} = 1$, other products vanishing.]

Equations (2.8) and (2.9) imply that

$$\gamma_a(x) = \nabla_a GG^{-1} + jl_a - hm_a, \tag{2.10}$$

with j and h determined applying $\bar{\delta}$ to (2.10) and multiplying by m^a and l^a, respectively. This yields

$$h = l^a\bar{\delta}(\nabla_a GG^{-1}), \quad j = m^a\bar{\delta}(\nabla_a GG^{-1}) = \delta h. \tag{2.11}$$

The connection is thus

$$\gamma_a(x) = \nabla_a GG^{-1} + \delta h l_a - hm_a. \tag{2.12}$$

Note that if G is a regular solution to the Sparling equation, then so is $G' = g(x)G$ for g a nonsingular matrix function of x^a. This generates an ordinary gauge transformation on γ, i.e.,

$$\gamma'(x) = g\gamma g^{-1} + dgg^{-1}. \tag{2.13}$$

The Yang–Mills field is then obtained from (2.12) in the usual way

$$F = d\gamma - \gamma\wedge\gamma. \tag{2.14}$$

III. SPARLING'S EQUATION AND TRIANGULAR DATA

Before we discuss the integration of the Sparling equation with upper triangular data, we first investigate the Abelian version of the same equation, namely

$$\delta F = A_R, \quad F = -\log G, \tag{3.1}$$

with A_R the restriction of $A(u,\zeta,\bar{\zeta})$, the scalar null data for a Maxwell field, to the intersection of C_x with $\mathscr{I}^+$.

The general regular solution to (3.1) is

$$F(x^a,\zeta,\tilde{\zeta}) = \int_{S^2} K(\zeta,\bar{\zeta},\eta,\bar{\eta})A_R(x,\eta,\bar{\eta})dS_\eta + f(x^a), \tag{3.2}$$

with

$$dS_\eta \equiv d\eta\wedge d\bar{\eta}/(1+\eta\bar{\eta})^2,$$

and the kernel

$$K(\zeta,\bar{\zeta},\eta,\bar{\eta}) = \frac{1}{4\pi}\frac{l^a(\zeta,\bar{\zeta})\bar{m}_a(\eta,\bar{\eta})}{l^a(\zeta,\bar{\zeta})l_a(\eta,\bar{\eta})} = \frac{1}{4\pi}\frac{1+\eta\bar{\zeta}}{\bar{\zeta}-\bar{\eta}}. \tag{3.3}$$

One thus has in the Abelian case the simple result

$$G(x,\zeta,\bar{\zeta}) = g(x)\exp\left(-\int KA_R\, dS_\eta\right). \tag{3.4}$$

With the gauge freedom $g(x)$ can be made into 1.

Unfortunately it is not easy to generalize (3.4) to the non-Abelian case where G and A_R are matrix valued. However in the case of A_R being 2×2, upper triangular, a generalization does exist as we now show. (Though we are only concerned here with 2×2 matrices, this generalization

seems to work for larger upper triangular matrices.)

The Sparling equation written explicitly becomes

$$\eth\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = -\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

or

$$\eth G_{11} = -G_{11}A_{11}, \tag{3.5}$$

$$\eth G_{21} = -G_{21}A_{11}, \tag{3.6}$$

$$\eth G_{12} = -G_{12}A_{22} - G_{11}A_{12}, \tag{3.7}$$

$$\eth G_{22} = -G_{22}A_{22} - G_{21}A_{12}. \tag{3.8}$$

Equations (3.5) and (3.6) are the same as the Abelian case just discussed and integrate to

$$G_{11} = g_{11}(x)\exp\left(-\int KA_{11}\,dS_\eta\right), \tag{3.9}$$

$$G_{21} = g_{21}(x)\exp\left(-\int KA_{11}\,dS_\eta\right). \tag{3.10}$$

Equations (3.7) and (3.8) are the inhomogeneous versions of the same equation with the inhomogeneous terms given by (3.9) and (3.10). Their solutions are

$$G_{12} = \mathring{G}\left(g_{12}(x) - \int K\mathring{G}^{-1}G_{11}A_{12}\,dS_\eta\right), \tag{3.11}$$

$$G_{22} = \mathring{G}\left(g_{22}(x) - \int K\mathring{G}^{-1}G_{21}A_{12}\,dS_\eta\right), \tag{3.12}$$

with

$$\mathring{G} = \exp\left(-\int KA_{22}\,dS_\eta\right), \tag{3.13}$$

a solution to the homogeneous equation.

By a gauge transformation, i.e., multiplication of G on the left by a matrix-valued function of x^a, (3.9)–(3.12) can be put into the following simple form:

$$G(x,\zeta,\bar{\zeta}) = \begin{pmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{pmatrix}, \tag{3.14}$$

with

$$G_{11} = \exp\left(-\int KA_{11}\,dS_\eta\right),$$

$$G_{22} = \exp\left(-\int KA_{22}\,dS_\eta\right), \tag{3.15}$$

$$G_{12} = -G_{22}\int KG_{11}^{-1}G_{22}A_{12}\,dS_\eta.$$

Using (2.10) and (2.14) it is a simple task to express the connection and field in terms of the G's.

The above solution generating technique (for upper triangular A_R) applies if the elements of A_R are spin weight 1 functions. However, in a more general situation the diagonal elements of A_R, i.e., A_{11} and A_{22} can be spin weight 1 functions, and the off-diagonal element A_{12} can be a $-2s+1$ spin weighted function for $s \geqslant 0$. Equations (3.5)–(3.8) can still be solved, producing a regular G. We have also studied this situation and plan to communicate the details in a future paper.

IV. GAUGE FREEDOM IN THE CHOICE OF DATA

In this section we would like to show that there exist equivalence classes of data, where all data in the same class yields the same connection and Yang–Mills field. Furthermore, we will show that in each equivalence class there will be (at least) one choice of data that is in upper triangular form. We will thus have shown that all 2×2 self-dual Yang–Mills fields can be obtained from upper triangular data and they can be explicitly given by the method of the previous section.

The basic idea is to begin with the Sparling equation and a solution G and look for transformations

$$G \to Gg = G',$$

so that the new G' yields, via (2.10), the same connection and field as did G. [Note that here the g is different than in (2.13) and is multiplied on the right.] We then find a new Sparling equation for G' with a new A'; a transform of A. Finally to complete the argument we seek a specialization of the g so that A' is upper triangular.

We claim that G's related by

$$G' = Gg, \tag{4.1}$$

for a regular g of the form

$$g = g(l,m,\zeta,\bar{\zeta}), \tag{4.2}$$

with

$$l = x^a l_a, \quad m = x^a m_a = \eth l, \tag{4.3}$$

yield the identical connection. We now sketch the proof.

From (4.1) we have

$$\begin{aligned}\nabla_a G'G'^{-1} &= \nabla_a GG^{-1} + G\nabla_a gg^{-1}G^{-1} \\ &= \nabla_a GG^{-1} + G(g_{,l}l_a + g_{,m}m_a)g^{-1}G^{-1},\end{aligned} \tag{4.4}$$

where we have used (4.2) and (4.3). Substituting (4.4) into

$$\begin{aligned}\gamma'_a = {} & \nabla_a G'G'^{-1} + m^b\bar{\eth}(\nabla_b G'G'^{-1})l_a \\ & - l^b\bar{\eth}\nabla_b(G'G'^{-1})m_a,\end{aligned} \tag{4.5}$$

the primed version of (2.12), we find that all terms involving g cancel, leaving

$$\gamma'_a = \gamma_a \tag{4.6}$$

as claimed.

Since both G and G' satisfy Sparling equations for data A and A', respectively, i.e.,

$$\eth G = -GA, \quad \eth G' = G'A',$$

one immediately calculates the relationship between the A's, i.e.,

$$A' = g^{-1}Ag - g^{-1}\eth g. \tag{4.7}$$

We can now ask for the equation on the g so that for an arbitrary 2×2 A, we have A' in upper triangular form. Equation (4.7) becomes

$$\begin{aligned}&\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} A'_{11} & A'_{12} \\ 0 & A'_{22} \end{pmatrix} \\ &\quad = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} - \eth\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}.\end{aligned} \tag{4.8}$$

The two relevant components of (4.8) are

$$g_{11}A'_{11} = A_{11}g_{11} + A_{12}g_{21} - \eth g_{11},$$
$$g_{21}A'_{11} = A_{21}g_{11} + A_{22}g_{21} - \eth g_{21}, \tag{4.9}$$

which, when A'_{11} is eliminated, yields after some manipulation, the differential equation

$$\eth\lambda + A_{12}\lambda^2 + \Delta\lambda - A_{21} = 0, \tag{4.10}$$

where

$$\lambda = g_{21}/g_{11}, \quad \Delta = A_{11} - A_{22}. \tag{4.11}$$

Equation (4.10) is our required condition on the g for the triangularization of A. Note that aside from the condition that the components of g be regular functions and $|g| \neq 0$ there are no other restrictions on the g.

The main problem now is to show that (4.10) has solutions λ such that (a) g_{21} and g_{11} are appropriately regular or similarly that λ have an appropriate singularity structure and (b) that g_{21} and g_{11} (or λ) be functions only of $l,m,\zeta,\bar{\zeta}$. (The space-time points x^a enter into g only via its dependence in l and m.)

We have tacitly assumed here that g_{21} and g_{11} are holomorphic, spin weight 0 functions. In the course of further study, it has been discovered (with L. Mason) that, in general, the g_{21} and g_{11} can be spin weight s functions for $s \geqslant 0$. Their ratio (λ), however, still remains a spin weight 0 function with appropriate singularity structure. The upper triangularization in this case leads to the type of data described in the last paragraph of Sec. III.

Before showing that (a) and (b) can be satisfied we first discuss the meaning of the term regularity that we have been using. By assumption the characteristic data $A(u,\zeta,\bar{\zeta})$ was to be an analytic function of $u,\zeta,\bar{\zeta}$ in the region u in the neighborhood of the real line and $(\zeta,\bar{\zeta})$ in the thickened S^2 defined by $\bar{\zeta}$ near $\bar{\zeta}$. We refer to this latter region as CS^2. When A is restricted to $C_x \cap \mathscr{I}^+$, $A_R(x^a,\zeta,\bar{\zeta})$ is holomorphic in the neighborhood of real x^a and CS^2. We demand that after transformation, $A \to A' = g^{-1}Ag - g^{-1}\eth g$, $A'(l,m,\zeta,\bar{\zeta})$ also be holomorphic in the same region which implies that g have the same holomorphic behavior. Since a holomorphic function can have zeros of only finite order we have immediately the condition on λ that in our holomorphic region λ should have singularities no worse than finite order poles. This is what was meant by "appropriate singularity structure" for λ.

Returning to the questions (a) and (b), we first note that there is an alternate way to write the independent variables l and m using the explicit form of l_a and m_a, namely

$$l = l_a x^a = [1/\sqrt{2}(1+\zeta\bar{\zeta})]\{[x^0 - x^3) + (x^1 + ix^2)\bar{\zeta}] + \zeta[(x^1 - ix^2) + (x^0 + x^3)\bar{\zeta}]\},$$

or

$$l = (w_0 + \zeta w_1)/2\sqrt{2}P, \tag{4.12}$$

with

$$w_0 = (x^0 - x^3) + (x^1 + ix^2)\bar{\zeta},$$
$$w_1 = (x^1 - ix^2) + (x^0 + x^3)\bar{\zeta}, \tag{4.13}$$
$$P = \tfrac{1}{2}(1 + \zeta\bar{\zeta}),$$

and

$$m = m_a x^a = \eth l = (w_1 - \bar{\zeta}w_0)/2\sqrt{2}P. \tag{4.14}$$

The transformation, (4.12) and (4.14), from (l,m) to (w_0,w_1) has the advantage of explicitly displaying all the ζ dependence when we consider functions of $l,m,\zeta,\bar{\zeta}$ becoming functions of $w_0,w_1,\zeta,\bar{\zeta}$. [Note that $(w_0,w_1,\bar{\zeta})$ is a dual twistor.]

In this spirit Eq. (4.10) becomes

$$\frac{\partial\lambda}{\partial\zeta} + a\lambda^2 + b\lambda + c = 0, \tag{4.15}$$

with the holomorphic coefficients

$$a = a(w_0,w_1,\zeta,\bar{\zeta}) = A_{12}/(1+\zeta\bar{\zeta}),$$
$$b = b(w_0,w_1,\zeta,\bar{\zeta}) = \Delta/(1+\zeta\bar{\zeta}), \tag{4.16}$$
$$c = c(w_0,w_1,\zeta,\bar{\zeta}) = -A_{21}/(1+\zeta\bar{\zeta}).$$

Equation (4.15) is the Riccati equation for λ as a function of ζ with external parameters $\bar{\zeta},w_0$, and w_1. From the theory of solutions to the Riccati equation[7] we see that if at some point $\zeta = \zeta_0$ one is given the initial holomorphic data

$$\lambda_0 = \lambda(\zeta_0,\bar{\zeta},w_0,w_1),$$

there exists a unique holomorphic solution in the neighborhood of ζ_0, $\lambda = \lambda(\zeta,\bar{\zeta},w_0,w_1)$. Using (4.12) and (4.14) we see that λ is a function of only $(l,m,\zeta,\bar{\zeta})$ as required by condition (b).

To show that condition (a) is satisfied, i.e., that the appropriate singularity structure exists, we use a powerful theorem[7] concerning the fixed and movable singularities of solutions to the Riccati equation. The fixed singularities, which arise only at the singular points of the coefficients a,b,c in the Ricatti equation, are essential singularities or branch points while the movable singularities, arising from the choice of initial data, are only finite poles. Furthermore,[7] the solutions are holomorphic in the external parameters $(\bar{\zeta},w_0,w_1)$.

The first point to be noted is that the fixed singularities are of no concern to us since in our region of concern (real x^a and CS^2), the coefficients are all holomorphic. We thus can start with $\lambda = \lambda_0$ at ζ_0; then by analytic extension first in ζ and then in $\bar{\zeta}$ (near $\bar{\zeta}$) for fixed x^a (and perhaps repeating this process of extension in ζ and then $\bar{\zeta}$), we have λ holomorphic everywhere on CS^2 except at a finite number of points where there are just finite poles. If we adapt our $(\zeta,\bar{\zeta})$ coordinate system so that there is no pole in λ at the north pole (or at ∞) of $(\zeta,\bar{\zeta})$ we can write λ as the ratio of two functions which are both regular, with zeros on CS^2, i.e., we can write

$$\lambda = g_{21}/g_{11},$$

thus satisfying condition (a).

V. A SIMPLE EXAMPLE

We describe here a simple but important example of this triangularization process. The Yang–Mills field associated with this example is the single instanton field. Consider 2×2 data of the form

$$A = A_0(\zeta,\bar{\zeta})/l^2, \tag{5.1}$$

with

$$A_0^2 = 0, \quad \mathrm{tr}\, A_0 = 0, \tag{5.2}$$

$$\delta A_0 = 0. \tag{5.3}$$

From (5.2) A_0 has the form

$$A_0 = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}, \quad \alpha^2 + \gamma\beta = 0. \tag{5.4}$$

Substituting (5.1) and (5.4) into (4.10) we have

$$\delta\lambda = -(\beta/l^2)(\lambda + \alpha/\beta))^2, \tag{5.5}$$

which, since $\delta\alpha = \delta\beta = 0$, can be integrated immediately as

$$\lambda = -(\alpha + ml)/\beta, \tag{5.6}$$

i.e.,

$$g_{11} = \beta, \quad g_{21} = -(\alpha + ml). \tag{5.7}$$

Note that g_{11} and g_{21} are spin weight $s = 1$ functions and are of the type mentioned in Sec. IV.

VI. DISCUSSION

We have shown here now all global, on M, GL(2,C) self-dual Yang–Mills fields can be obtained in terms of explicit integrals over the characteristic data. Similar methods can be used to obtain local fields.

With these results comes a series of related questions and problems that are being pursued. An immediate question is, can this triangularization process be generalized to higher dimensional groups, e.g., GL(3,C). Another class of problems is how to give triangular data with specific symmetries so that the Yang–Mills solutions reduce to the known special cases,[5] e.g., axial-symmetric, stationary Einstein equations, sine–Gordon equations, etc.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under Grant No. PHY82–17853, supplemented by funds from the National Aeronautics and Space Administration, at the University of California at Santa Barbara.

[1] R. S. Ward, Phys. Lett. A **61**, 81 (1977).
[2] M. F. Atiyah, "Classical geometry of Yang–Mills fields," in *Fermi Lectures* (Scuola Normale, Pisa, 1980).
[3] A. A. Belarin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[4] R. Penrose, Gen. Relativ. Gravit. **7**, 31 (1976).
[5] R. Ward, Philos. Trans. R. Soc. London **315**, 451 (1985).
[6] E. Newman, J. Math. Phys. **27**, 2797 (1986).
[7] E. Hille, *Ordinary Differential Equations in the Complex Domain* (Wiley, New York, 1976).

Comment on a paper by Z-Z. Zhong [J. Math. Phys. 26, 404 (1985)]

K. Demys
Mathematics & Morphology Research Center, Editorial and Research Offices, 1052 Santa Fe Avenue, Albany, California 94706

(Received 6 June 1986; accepted for publication 15 October 1986)

The paper of Zhong [J. Math. Phys. **26**, 404 (1985)], though very interesting, did contain errors of both scientific and historical fact that should be corrected for the record.

I. SCIENTIFIC ERROR

In the paper by Zhong,[1] p. 404, Sec. II, first paragraph, lines 5–7 are in error, since z^{-1} always exists, even for numbers in H of zero norm. Thus the numbers $z^{-1} \equiv [k(1 \pm \epsilon)]^{-1}$ always exist, where k is real or complex, and $\epsilon^2 = 1, \epsilon \neq \pm 1$, even though they all have zero norms. But a distinction, easily verifiable, must be made here between z^{-1} and $1/z$ since we are now dealing with zero divisors. Thus $z^{-1} = (1 \pm \epsilon)/4k$ whereas $1/z = 1/k(1 \pm \epsilon)$, a class of numbers which, despite their closed representations, are divisors of infinity and have infinite norms. Also, which the author of the article in question also failed to observe, $[k(1 \pm \epsilon)]^0 = \frac{1}{2}(1 \pm \epsilon)$, which is of course idempotent; as it should, $z^0 z^n = z^n$. Also, $z^{-1}z = z^0$, and indeed

$$[(1 \pm \epsilon)/4k][k(1 \pm \epsilon)] = \tfrac{1}{2}(1 \pm \epsilon).$$

But note well that, since $z^0 \neq 1$, $z(1/z) \neq z(z^{-1})$; but $z(1/z) = 1$. Already in papers from 1977–1980 Musès[2] had pointed out these facts (see Sec. II).

Although the existence of z^{-1} does not affect Zhong's results *per se* and immediately, the fact that his lines on p. 404, above noted, are in error may well mislead others trying to extend his results, and this comment is thus offered in addition to instrinsic interest. It is worthwhile noting that these numbers lead to an important timesaving tool: the "countercomplex" form of de Moivre's theorem for complex numbers; namely

$$e^{k\theta\epsilon} = (\cosh\theta + \epsilon \sinh\theta)^k = \cosh k\theta + \epsilon \sinh k\theta,$$

where θ may be real, complex, or countercomplex (i.e., in the form $a + b\epsilon$, where a and b are real). Also,

$$\epsilon^k = \cosh^2 \tfrac{1}{2}\pi k + \epsilon \sinh^2 \tfrac{1}{2}\pi k - \tfrac{1}{2}i(1 - \epsilon)\sin \pi k.$$

Thus $\sqrt{\epsilon} = \pm\frac{1}{2}(1 + \epsilon - i + \epsilon i)$, where $\epsilon i = i\epsilon$ and hence $(\epsilon i)^2 = -1$. All the Pauli and Dirac spinors are susceptible of countercomplex hypernumber representation, often more convenient than matrices.

II. HISTORICAL ERROR

Again see p. 404, lines 3–5. It was not Kunstatter, Moffat, and Malzan[3] who first applied these numbers in physics. Before them Musès[2] was the first to point out these hypernumbers' physical usefulness in 1980 and considerably earlier in an invited lecture, in 1970.

ACKNOWLEDGMENTS

The writer is grateful to the editor and the referee for helpful suggestions.

[1] Z-Z. Zhong, J. Math. Phys. **26**, 404 (1985).
[2] C. Musès, J. Appl. Math. Comput. **6**, 63 (1980); invited lecture, Ames Research Center (NASA), Moffett Field, CA, 1970.
[3] G. Kunstatter, J. W. Moffat, and R. Malzan, J. Math. Phys. **24**, 886 (1983).

 0022-2488/87/020339-01$02.50 © 1987 American Institute of Physics

Multiplicative stochastic processes involving the time derivative of a Markov process

Henk F. Arnoldus and Thomas F. George
Departments of Physics and Astronomy and Chemistry, 239 Fronczak Hall, State University of New York at Buffalo, Buffalo, New York 14260

(Received 2 May 1986; accepted for publication 1 October 1986)

The characteristic functional of the derivative $\dot{\phi}(t)$ of a Markov process $\phi(t)$ and the related multiplicative process $\sigma(t)$, which obeys the stochastic differential equation $i\dot{\sigma}(t) = (A + \dot{\phi}(t)B)\sigma(t)$, have been studied. Exact equations for the marginal characteristic functional and the marginal average of $\sigma(t)$ are derived. The first equation is applied to obtain a set of equations for the marginal moments of $\dot{\phi}(t)$ in terms of the prescribed properties of $\phi(t)$. It is illustrated by an example how these equations can be solved, and it is shown in general that $\dot{\phi}(t)$ is delta correlated, with a smooth background. The equation of motion for the marginal average of $\sigma(t)$ is solved for various cases, and it is shown how closed-form analytical expressions for the average $\langle\sigma(t)\rangle$ can be obtained.

I. INTRODUCTION

The equation of motion for the density operator of an atom in a finite-bandwidth laser field or the equation for the regression of the atomic dipole correlations assumes the general form[1,2]

$$i\frac{d\sigma}{dt} = (A + \dot{\phi}(t)B)\sigma\,, \tag{1.1}$$

where A and B are linear operators in Liouville space, which act on the Liouville vector $\sigma(t)$. Here $\phi(t)$ represents the laser phase, which is considered to be a real-valued stochastic process. The fluctuating phase broadens the laser line, but the atom responds to the instantaneous frequency shift $\dot{\phi}(t)$, which is the time derivative of the laser phase.[3] The process $\dot{\phi}(t)$ is again a stochastic process, and via Eq. (1.1) the state of the atom or the correlation functions $\sigma(t)$ become stochastic quantities. The issue in quantum optics is then to solve the multiplicative stochastic differential equation (1.1) for the average $\langle\sigma(t)\rangle$. The first solution was obtained by Fox,[4] who assumed the process $\dot{\phi}(t)$ to be Gaussian white noise, which corresponds to a diffusive Gaussian phase $\phi(t)$ (the Wiener–Lévy process). This result was generalized to a Gaussian process $\dot{\phi}(t)$ with a finite correlation time and an exponentially decaying correlation function[5–7] (the Ornstein–Uhlenbeck process), and to a process $\phi(t)$, which is again diffusive, but not Gaussian[8,9] (the independent-increment process). Furthermore, Eq. (1.1) can be solved for $\langle\sigma(t)\rangle$ if we have $\dot{\phi}(t)$ as a Markov random-jump process,[10–13] which models a multimode laser.[14,15]

In these examples the solvability of the problem relies on the Gaussian property of $\dot{\phi}(t)$, or hinges on the prescribed stochastics of $\dot{\phi}(t)$. This implies that the process $\dot{\phi}(t)$ is actually considered to be the driving process. For a single-mode laser in general, however, the phase fluctuations $\phi(t)$ are specified rather than the derivative $\dot{\phi}(t)$ of this process. A prime example would be the atomic response to phase-locked radiation,[16] as it is generated for instance by some ring lasers.[17] In this paper we shall develop a general method to solve Eq. (1.1) for the case that $\phi(t)$ is a given Markov process. The formal theory will be exemplified by a specific choice for $\phi(t)$, which models phase-locked radiation. Furthermore, we shall study the time derivative of $\phi(t)$ itself and extract the stochastics of $\dot{\phi}(t)$ from the properties of $\phi(t)$.

II. THE STOCHASTICS OF $\phi(t)$

Let us define the phase $\phi(t)$ as a homogeneous Markov process.[18] Then its stochastics is fixed by the probability distribution $P(\phi,t)$ and the conditional probability distribution $P_\tau(\phi_2|\phi_1)$ $(\tau\geqslant 0)$, which has the significance of the probability density for the occurrence of $\phi(t+\tau) = \phi_2$ if $\phi(t) = \phi_1$. For a homogeneous process this is independent of t by definition. The higher-order statistics is now determined by the Markov property.[19] From the obvious relation

$$\int d\phi'\, P_{t-t_0}(\phi|\phi')P(\phi',t_0) = P(\phi,t), \quad t\geqslant t_0\,, \tag{2.1}$$

it follows that it is sufficient to prescribe the probability distribution $P(\phi,t)$ for a single time point t_0 only. The time evolution towards $t > t_0$ can then be found from Eq. (2.1) and $P_{t-t_0}(\phi|\phi')$.

The conditional probability distribution obeys the Master equation[18]

$$\frac{\partial}{\partial\tau}P_\tau(\phi_3|\phi_1) = \int d\phi_2\{W(\phi_3|\phi_2) - a(\phi_2)\delta(\phi_3-\phi_2)\}P_\tau(\phi_2|\phi_1)\,, \tag{2.2}$$

with $W(\phi'|\phi)\geqslant 0$ as the transition rate of the process from ϕ to ϕ' and

$$a(\phi) = \int d\phi' W(\phi'|\phi)\,, \tag{2.3}$$

which is the loss rate of ϕ, independent of the final value ϕ'. The initial condition for Eq. (2.2) reads

$$P_0(\phi_3|\phi_1) = \delta(\phi_3-\phi_1)\,, \tag{2.4}$$

so a given $W(\phi'|\phi)$ determines $P_\tau(\phi_3|\phi_1)$ for every $\tau\geqslant 0$.

 0022-2488/87/020340-07$02.50 © 1987 American Institute of Physics

Hence the stochastics of a homogeneous Markov process $\phi(t)$ is fixed as soon as $P(\phi,t_0)$ and $W(\phi'|\phi)$ are prescribed. These functions will from now on be assumed to be given.

III. THE CHARACTERISTIC FUNCTIONAL

A convenient way to represent the stochastic properties of a stochastic process is by means of its characteristic functional.[9,20] Since we are concerned with the process $\dot{\phi}(t)$, we define

$$Z_t[k] = \left\langle \exp\left(-i \int_{t_0}^{t} ds\, \dot{\phi}(s) k(s) \right) \right\rangle , \quad t \geqslant t_0 , \tag{3.1}$$

which is a functional of the test function $k(t)$. Here the angle brackets denote an average over the stochastic process $\phi(t)$ or $\dot{\phi}(t)$, whatever is prescribed. A general method to evaluate $Z_t[k]$ for the case where $\dot{\phi}(t)$ is a homogeneous Markov process has been given by van Kampen.[21]

Knowledge of the characteristic functional $Z_t[k]$ determines completely the stochastics of $\dot{\phi}(t)$, which can be seen as follows. Choose $k(s)$ as the sequence of δ functions

$$k(s) = -\sum_{l=1}^{n} \delta(s - t_l) k_l , \quad t_l > t_0 , \tag{3.2}$$

and take $t = \infty$ in (3.1). Then we find

$$Z_\infty[k] = \langle \exp(ik_n \dot{\phi}(t_n) + \cdots + ik_1 \dot{\phi}(t_1)) \rangle , \tag{3.3}$$

which is the moment-generating function of $\dot{\phi}(t)$. If we write $z_n(k_n,t_n;\ldots;k_1,t_1)$, then we can obtain the moments of $\psi(t) \equiv \dot{\phi}(t)$ according to

$$\begin{aligned} &\langle \psi(t_n) \cdots \psi(t_1) \rangle \\ &\quad = (-i)^n \frac{\partial}{\partial k_n} \cdots \frac{\partial}{\partial k_1} \\ &\qquad \times z_n(k_n,t_n;\ldots;k_1,t_1) \,\big|_{k_n = \cdots = k_1 = 0} , \end{aligned} \tag{3.4}$$

and the probability distributions by

$$\begin{aligned} &\bar{P}_n(\psi_n,t_n;\ldots;\psi_1,t_1) \\ &\quad = \frac{1}{(2\pi)^n} \int dk_n \cdots dk_1 \\ &\qquad \times e^{-ik_n\psi_n - \cdots - ik_1\psi_1} z_n(k_n,t_n;\ldots;k_1,t_1) , \end{aligned} \tag{3.5}$$

where we have introduced $\bar{P}_n$ in order to distinguish from the probability distributions for $\phi(t)$ itself.

IV. THE MARGINAL AVERAGE

A. General

The exponential in Eq. (3.1) is a functional of both $k(t)$ and $\dot{\phi}(t)$, so it depends on the values of $\dot{\phi}(t)$ in the complete interval $[t_0,t]$. After the average has been taken it will be only a functional of $k(t)$. The general attempt to evaluate averages of a functional is to derive an equation for the average. For subsequently solving this equation for functionals which involve Markov processes, this scheme is most conveniently carried out by an intermediate introduction of Burshtein's marginal averages.[22] Since in our problem the stochastics of $\phi(t)$ is assumed to be given, the appropriate marginal characteristic functional, which is related to $Z_t[k]$, should be defined as

$$Q_t[\phi_0,k] = \left\langle \delta(\phi(t) - \phi_0) \exp\left(-i \int_{t_0}^{t} ds\, \dot{\phi}(s) k(s) \right) \right\rangle , \tag{4.1}$$

for $t \geqslant t_0$. The initial value is then

$$Q_{t_0}[\phi_0,k] = \langle \delta(\phi(t_0) - \phi_0) \rangle = P(\phi_0,t_0) , \tag{4.2}$$

and $Z_t[k]$ follows from $Q_t[\phi_0,k]$ according to

$$Z_t[k] = \int d\phi_0\, Q_t[\phi_0,k] . \tag{4.3}$$

For $t = t_0$ we find with Eq. (4.2)

$$Z_{t_0}[k] = \int d\phi_0\, P(\phi_0,t_0) = 1 ,$$

in agreement with Eq. (3.1).

In order to derive an equation for the time evolution of $Q_t[\phi_0,k]$, we first increase t by a small amount $\Delta t > 0$. This gives

$$\begin{aligned} Q_{t+\Delta t}[\phi_0,k] = \Big\langle &\delta(\phi(t+\Delta t) - \phi_0) \\ &\times \exp\{ -i(\phi(t+\Delta t) - \phi(t)) k(t) \} \\ &\times \exp\left(-i \int_{t_0}^{t} ds\, \dot{\phi}(s) k(s) \right) \Big\rangle . \end{aligned} \tag{4.4}$$

Subsequently, we expand the exponential functional of $\dot{\phi}(s)$ in a series, and we take the average in (4.4) term by term. Hereafter, we apply the Master equation (2.2) for $P_{t+\Delta t}(\phi|\phi_0)$ and take the limit $\Delta t \to 0$. This yields an equation for the marginal average, and explicitly we find

$$\begin{aligned} \frac{\partial}{\partial t} Q_t[\phi_0,k] = \int d\phi \{ &W(\phi_0|\phi) - a(\phi)\delta(\phi_0 - \phi) \} \\ &\times e^{-i(\phi_0 - \phi)k(t)} Q_t[\phi,k] . \end{aligned} \tag{4.5}$$

The Markov process $\phi(t)$ is characterized by $P(\phi_0,t_0)$ and $W(\phi_0|\phi)$, which, respectively, determine the initial value and the time evolution of $Q_t[\phi_0,k]$. For a specific choice of $W(\phi_0|\phi)$, we have to solve Eq. (4.5), after which the characteristic functional $Z_t[k]$ can be obtained from Eq. (4.3).

Notice the resemblance between the result (4.5) and the Master equation (2.2). If we multiply Eq. (2.2) by $P(\phi_1,t_0)$, take $\tau = t - t_0$ and apply the relation (2.1), we find

$$\frac{\partial}{\partial t} P(\phi_0,t) = \int d\phi \{ W(\phi_0|\phi) - a(\phi)\delta(\phi_0 - \phi) \} P(\phi,t) , \tag{4.6}$$

which is the Master equation for $P(\phi_0,t)$. This equation is identical to Eq. (4.5), including the initial condition (4.2), if we set $k(t) \equiv 0$. On the other hand, it follows from Eq. (4.1) that $Q_t[\phi_0,k] = \langle \delta(\phi(t) - \phi_0) \rangle = P(\phi_0,t)$ if we take $k(t) = 0$, so that in this case Eq. (4.5) should indeed reduce to Eq. (4.6).

B. Independent Increments

In order to display the usefulness and applicability of the marginal-functional approach, we consider an example. Let us specify the transition rate by

$$W(\phi_0|\phi) = \gamma w(\phi_0 - \phi) , \quad \gamma > 0 , \tag{4.7}$$

where the function $w(\eta)$ is normalized as

$$\int d\eta\, w(\eta) = 1\,. \tag{4.8}$$

The stochastic process $\phi(t)$ will be defined on the real axis, with $-\infty < \phi < \infty$. The assertion (4.7) states that the probability for a transition $\phi \to \phi_0$ depends only on the phase difference $\phi_0 - \phi$, and from Eq. (2.3) we find that $a(\phi) = \gamma$, so that the total loss rate for ϕ is independent of ϕ. This is a diffusion process, and it is commonly referred to as the independent-increment process. As an initial condition for the probability distribution, we take

$$P(\phi,t_0) = \delta(\phi)\,. \tag{4.9}$$

Comparison of the Master equations for $P_\tau(\phi|\phi')$ and $P(\phi,t)$ then shows that the probability distribution and the conditional probability distribution are related according to

$$P_{t-t_0}(\phi|\phi_0) = P(\phi-\phi_0,t)\,. \tag{4.10}$$

The Master equation (4.6) for $P(\phi,t)$ can be solved by Fourier transformation with respect to ϕ. If we write

$$\hat{P}(\rho,t) = \langle e^{i\rho\phi(t)}\rangle = \int_{-\infty}^{\infty} d\phi\, e^{i\rho\phi} P(\phi,t)\,, \tag{4.11}$$

which has $\hat{P}(\rho,t_0) = 1$ as the initial condition, then the solution of Eq. (4.6) is immediately seen to be

$$\hat{P}(\rho,t) = e^{\gamma(\hat{w}(\rho)-1)(t-t_0)}\,, \quad t \geqslant t_0\,, \tag{4.12}$$

in terms of the Fourier transform $\hat{w}(\rho)$ of $w(\phi)$. Note that $\hat{w}(0) = 1$, as a result of the normalization (4.7). Along the very same lines we can solve Eq. (4.55) for the Fourier transform $\hat{Q}_t[\rho,k]$. We obtain

$$\hat{Q}_t[\rho,k] = \exp\left(-\gamma\int_{t_0}^{t} ds \int_{-\infty}^{\infty} d\phi\,(1 - e^{i\phi(\rho - k(s))})w(\phi)\right), \tag{4.13}$$

after which the characteristic functional follows from

$$Z_t[k] = \hat{Q}_t[0,k]\,, \tag{4.14}$$

which yields the familiar result.[8]

V. THE MARGINAL MOMENTS

A. General

If we take $k(s)$ as the sequence of delta functions (3.2) in the definition (4.1) of the marginal characteristic functional, it assumes the form

$$Q_t[\phi_0,k] = \left\langle \delta(\phi(t)-\phi_0)\exp\left(i\sum_{l=1}^{n} k_l \psi(t_l)\Theta(t-t_l)\right)\right\rangle, \tag{5.1}$$

with $\psi(t) = \dot{\phi}(t)$ and $\Theta(t)$ the unit-step function. Just as we can find the moments $\langle \psi(t_n)\cdots\psi(t_1)\rangle$ of $\psi(t)$ from $Z_\infty[k]$, we can obtain the marginal moments $\langle \delta(\phi(t)-\phi_0)\psi(t_n)\cdots\psi(t_1)\rangle$ from $Q_t[\phi_0,k]$. Obviously the integral over ϕ_0 of the marginal moments yields the moments. The characteristic functional $Z_t[k]$ becomes independent of t if $t > t_l$ for all l, but $Q_t[\phi_0,k]$ remains time dependent. This is due to the appearance of $\delta(\phi(t)-\phi_0)$. Furthermore, the time t is a dynamical variable in Eq. (4.55), so that care should be exercised in the time ordering. The marginal moments follow from $Q_t[\phi_0,k]$ by differentiation, according to

$$\langle \delta(\phi(t)-\phi_0)\psi(t_n)\cdots\psi(t_1)\rangle\Theta(t-t_n)\cdots\Theta(t-t_1) = (-i)^n \frac{\partial}{\partial k_n}\cdots\frac{\partial}{\partial k_1} Q_t[\phi_0,k]\Big|_{k_n=\cdots=k_1=0}\,. \tag{5.2}$$

Equation (4.5) for $Q_t[\phi_0,k]$ implies an equation for the marginal moments. First, we note that

$$\exp\{-i(\phi_0-\phi)k(t)\}Q_t[\phi,k] = \left\langle \delta(\phi(t)-\phi)\exp\left\{i\sum_{l=1}^{n} k_l\{(\phi_0-\phi)\delta(t-t_l) + \psi(t_l)\Theta(t-t_l)\}\right\}\right\rangle. \tag{5.3}$$

After substituting this expression in the right-hand side of Eq. (4.5), differentiating with respect to $k_n,\ldots,k_1$, setting $k_n = \cdots = k_1 = 0$, and integrating over time, we obtain

$$\begin{aligned}\langle \delta(\phi(t)-\phi_0)\psi(t_n)\cdots\psi(t_1)\rangle\Theta(t-t_n)\cdots\Theta(t-t_1) &= \int d\phi\{W(\phi_0|\phi) - a(\phi)\delta(\phi_0-\phi)\} \\ &\times \int_{t_0}^{t} dt' \langle \delta(\phi(t')-\phi)\{(\phi_0-\phi)\delta(t'-t_n) \\ &+ \psi(t_n)\Theta(t'-t_n)\} \\ &\cdots\{(\phi_0-\phi)\delta(t'-t_1) + \psi(t_1)\Theta(t'-t_1)\}\rangle\,.\end{aligned} \tag{5.4}$$

When we set $t > t_l$ for all l, we have a Master-like equation for $\langle \delta(\phi(t)-\phi_0)\psi(t_n)\cdots\psi(t_1)\rangle$, and the lower-order marginal moments $\langle \delta(\phi(t)-\phi_0)\psi(t_m)\cdots\psi(t_1)\rangle$ with $m < n$ appear as inhomogeneous terms. Hence Eq. (5.4) should be solved successively for $n = 1$, $n = 2,\ldots$. We note that Eq. (5.4) provides an explicit expression for $\langle \psi(t_n)\cdots\psi(t_1)\rangle$ in terms of the lower-order marginal moments after an integration over ϕ_0. Indeed, from the property

$$\int d\phi_0\{W(\phi_0|\phi) - a(\phi)\delta(\phi_0-\phi)\} = 0\,, \tag{5.5}$$

the term with $\langle \delta(\phi(t')-\phi)\psi(t_n)\cdots\psi(t_1)\rangle$ on the right-hand side of Eq. (5.4) vanishes after an integration over ϕ_0.

B. Lowest orders

In order to exhibit clearly the structure of the equation for the marginal moments, we consider the cases $n = 1$ and $n = 2$ in some more detail. After a slight rearrangement, Eq. (5.4) for $n = 1$ can be written as

$$\begin{aligned}\langle \delta(\phi(t)-\phi_0)\psi(t_1)\rangle &= \int d\phi\{W(\phi_0|\phi) - a(\phi)\delta(\phi_0-\phi)\} \\ &\times\left\{(\phi_0-\phi)P(\phi,t_1) + \int_{t_1}^{t} dt' \langle \delta(\phi(t')-\phi)\psi(t_1)\rangle\right\},\end{aligned} \tag{5.6}$$

for $t \geqslant t_1$. This integral equation in time is equivalent to the differential equation

$$\frac{\partial}{\partial t}\langle\delta(\phi(t)-\phi_0)\psi(t_1)\rangle$$
$$=\int d\phi\{W(\phi_0|\phi)-a(\phi)\delta(\phi_0-\phi)\}$$
$$\times\langle\delta(\phi(t)-\phi)\psi(t_1)\rangle\,, \tag{5.7}$$

together with the initial condition

$$\langle\delta(\phi(t_1)-\phi_0)\psi(t_1)\rangle$$
$$=\int d\phi\{W(\phi_0|\phi)-a(\phi)\delta(\phi_0-\phi)\}$$
$$\times(\phi_0-\phi)P(\phi,t_1)\,. \tag{5.8}$$

The equation for the first marginal average $\langle\delta(\phi(t)-\phi_0)\psi(t_1)\rangle$ is identical to the Master equation (2.2), but with a different initial value.

Integration of (5.8) over ϕ_0 yields

$$\langle\psi(t_1)\rangle=\int d\phi\int d\phi_0\, W(\phi_0|\phi)(\phi_0-\phi)P(\phi,t_1)\,, \tag{5.9}$$

which expresses explicitly the average of $\langle\psi(t_1)\rangle$ in the given functions $W(\phi_0|\phi)$ and $P(\phi,t_1)$. With the aid of the Master equation, we can cast (5.9) in the form

$$\langle\psi(t_1)\rangle=\int d\phi\,\phi\frac{\partial}{\partial t_1}P(\phi,t_1)=\frac{d}{dt_1}\langle\phi(t_1)\rangle\,, \tag{5.10}$$

as it should be.

The solution of Eq. (5.6) for $\langle\delta(\phi(t)-\phi_0)\psi(t_1)\rangle$ provides the input for the explicit expression for the two-time correlation function, which becomes

$$\langle\psi(t_2)\psi(t_1)\rangle$$
$$=\int d\phi\int d\phi_0\{W(\phi_0|\phi)-a(\phi)\delta(\phi_0-\phi)\}$$
$$\times\{(\phi_0-\phi)^2\delta(t_2-t_1)P(\phi,t_2)$$
$$+(\phi_0-\phi)(\langle\delta(\phi(t_2)-\phi)\psi(t_1)\rangle\Theta(t_2-t_1)$$
$$+\langle\delta(\phi(t_1)-\phi)\psi(t_2)\rangle\Theta(t_1-t_2))\}\,. \tag{5.11}$$

The appearance of $\delta(t_2-t_1)$ shows that the time derivative of any Markov process is δ correlated with a continuous background.

C. Random jumps

Equation (5.11) for instance might seem awkward, but it is really straightforward in its application. Let us illustrate this with an example. Consider the random-jump process $\phi(t)$, defined as a stationary process with transition rate

$$W(\phi|\phi')=\gamma P(\phi)\,,\quad \gamma>0\,, \tag{5.12}$$

in terms of an arbitrary probability distribution $P(\phi)$. Equation (5.12) is equivalent to the statement that the probability for a transition $\phi'\to\phi$ is independent of the initial value ϕ' (see Ref. 13). From Eq. (5.9) we immediately derive

$$\langle\psi(t_1)\rangle=0\,, \tag{5.13}$$

which is, in view of (5.10), necessary for a stationary process. From (2.3) we obtain $a(\phi)=\gamma$, and the solution of Eq. (5.7), with initial value (5.8), is readily found to be

$$\langle\delta(\phi(t)-\phi_0)\psi(t_1)\rangle=\gamma P(\phi_0)(\phi_0-b_1)e^{-\gamma(t-t_1)}\,,\quad t\geqslant t_1\,. \tag{5.14}$$

Here we have introduced the moments of $P(\phi)$ as

$$b_n=\int d\phi\,\phi^n P(\phi)\,, \tag{5.15}$$

which are parameters of the process $\phi(t)$. Solution (5.14) can be substituted into Eq. (5.11), which gives the correlation function

$$\langle\psi(t_1)\psi(t_2)\rangle=\gamma(b_2-b_1^2)\{2\delta(t_1-t_2)-\gamma e^{-\gamma|t_2-t_1|}\}\,, \tag{5.16}$$

for all t_1,t_2. From (5.15) it follows that

$$b_2-b_1^2\geqslant 0\,, \tag{5.17}$$

so that for $t_1\neq t_2$ the correlation (5.16) is negative. For $t_1=t_2$ the δ function dominates the negative term, so that $\langle\psi(t_1)^2\rangle$ is positive, as it should be.

VI. THE MULTIPLICATIVE PROCESS

So far we have considered the stochastics of $\dot\phi(t)$ itself, and its characteristic functional. In this section we shall generalize the method, in order to solve the multiplicative equation (1.1). To this end we write the formal solution of (1.1) for the stochastic vector $\sigma(t)$ as

$$\sigma(t)=e^{-iA(t-t_0)}T\exp\left[-i\int_{t_0}^{t}ds\,\dot\phi(s)\tilde B(s)\right]\sigma(t_0)\,, \tag{6.1}$$

where T is the time-ordering operator and $\tilde B(t)$ is defined as

$$\tilde B(t)=e^{iA(t-t_0)}Be^{-iA(t-t_0)}\,. \tag{6.2}$$

In close analogy to the definition of $Q_t[\phi_0,k]$ in Eq. (4.1), we now introduce the marginal average of $\sigma(t)$ by

$$\zeta(\phi_0,t)=\langle\delta(\phi(t)-\phi_0)\sigma(t)\rangle\,. \tag{6.3}$$

Then we substitute the expression (6.1) for $\sigma(t)$ and replace t by $t+\Delta t$, which gives a formula similar to Eq. (4.4). That this can also be done for the time-ordered exponential is sometimes referred to as the semigroup property of the evolution operator. Along the same lines that led to Eq. (4.5) we now find

$$i\frac{\partial}{\partial t}\zeta(\phi_0,t)=A\zeta(\phi_0,t)+i\int d\phi\{W(\phi_0|\phi)$$
$$-a(\phi)\delta(\phi_0-\phi)\}e^{-i(\phi_0-\phi)B}\zeta(\phi,t)\,, \tag{6.4}$$

or equivalently

$$\left(i\frac{\partial}{\partial t}-A+ia(\phi_0)\right)\zeta(\phi_0,t)$$
$$=i\int d\phi\, W(\phi_0|\phi)e^{-i(\phi_0-\phi)B}\zeta(\phi,t)\,. \tag{6.5}$$

Notice that the operator B appears in the exponential, rather than $\tilde B(t)$, as could be expected by analogy with the characteristic functional. For a given stochastic process $\phi(t)$, e.g., a given $W(\phi|\phi')$ and $P(\phi,t_0)$, we have to solve Eq. (6.4) with the initial condition

$$\zeta(\phi_0,t_0) = \langle \delta(\phi(t_0) - \phi_0)\sigma(t_0)\rangle , \tag{6.6}$$

after which $\langle\sigma(t)\rangle$ follows from

$$\langle\sigma(t)\rangle = \int d\phi_0\, \zeta(\phi_0,t) . \tag{6.7}$$

For a given nonstochastic state $\sigma(t_0)$, the initial condition reduces to

$$\zeta(\phi_0,t_0) = P(\phi_0,t_0)\sigma(t_0) , \tag{6.8}$$

which differs from (6.6) by the fact that there are no initial correlations. This means that the process $\sigma(t)$ has no memory to times smaller than t_0, and consequently its evolution for $t \geqslant t_0$ is completely determined by its initial state $\sigma(t_0)$. It was emphasized by Arnoldus and Nienhuis[13] that the common choice $\zeta(\phi_0,t_0) = P(\phi_0,t_0)\langle\sigma(t_0)\rangle$ is merely an approximation which only holds for small correlation times of $\dot{\phi}(t)$.

VII. SOLUTIONS

A. Independent increments

Equation (6.5) for the marginal average of $\sigma(t)$ can be solved for the independent-increment process with the same procedure as in Sec. IV, where we obtained the characteristic functional. If we adopt the Fourier transform

$$\hat{\zeta}(\rho,t) = \int_{-\infty}^{\infty} d\phi\, e^{i\rho\phi}\zeta(\phi,t) = \langle e^{i\rho\phi(t)}\sigma(t)\rangle , \tag{7.1}$$

where the second equality follows after application of Eq. (6.3), then $\langle\sigma(t)\rangle$ can be found from

$$\langle\sigma(t)\rangle = \hat{\zeta}(0,t) . \tag{7.2}$$

With the technique of Sec. IV we can find $\hat{\zeta}(\rho,t)$, and if we differentiate the result with respect to time, we find

$$i\frac{\partial}{\partial t}\hat{\zeta}(\rho,t) = (A - i\hat{W}(\rho))\hat{\zeta}(\rho,t) , \tag{7.3}$$

with

$$\hat{W}(\rho) = \gamma\int_{-\infty}^{\infty} d\eta(1 - e^{i\eta(\rho - B)})w(\eta) . \tag{7.4}$$

The operator $\hat{W}(\rho)$ accounts for the phase fluctuations. If we set $\rho = 0$ in Eq. (7.3), we achieve the equation for $\langle\sigma(t)\rangle$, with solution

$$\langle\sigma(t)\rangle = e^{-i(A - i\hat{W}(0))(t - t_0)}\langle\sigma(t_0)\rangle , \tag{7.5}$$

for $t \geqslant t_0$. We note that $\langle\sigma(t)\rangle$ can be expressed in terms of $\langle\sigma(t_0)\rangle$ for this process, so that there are no initial correlations for the diffusion process. The process $\dot{\phi}(t)$ has no memory, and with the results of Sec. V it can be shown that $\dot{\phi}(t)$ is indeed delta correlated. This means that $\langle\dot{\phi}(t_n)\cdots\dot{\phi}(t_1)\rangle$ for all n contains only δ functions, which implies the factorization in (7.5).

A special case arises if we take

$$\gamma w(\eta) = \gamma\delta(\eta) + \lambda\delta''(\eta) , \quad \lambda > 0 , \tag{7.6}$$

where the primes on the δ function denote differentiation with respect to its argument. It is easy to check that this process is the Wiener–Lévy process, or the phase-diffusion process. If we substitute (7.6) into (4.12), we find that $P(\phi,t)$ is Gaussian, and obviously this is the only Gaussian limit of the diffusion process. The operator $\hat{W}(\rho)$ in (7.4) reduces to

$$\hat{W}(\rho) = \lambda(\rho - B)^2 , \tag{7.7}$$

and the equation for $\langle\sigma(t)\rangle$ becomes

$$i\frac{d}{dt}\langle\sigma(t)\rangle = (A - i\lambda B^2)\langle\sigma(t)\rangle , \tag{7.8}$$

which is the result of Fox.[4]

B. Ornstein–Uhlenbeck process

The diffusion process has no memory and is essentially nonstationary. The initial distribution $P(\phi,t_0) = \delta(\phi)$ diffuses over the whole ϕ axis, $-\infty < \phi < \infty$. The inclusion of a finite memory time can stabilize this process. Let us define the transition probability as

$$\begin{aligned} W(\phi_0|\phi) - a(\phi)\delta(\phi_0 - \phi) &= \lambda\delta''(\phi_0 - \phi) \\ &\quad + \gamma\phi\delta'(\phi_0 - \phi) , \ \lambda > 0 , \ \gamma > 0 . \end{aligned} \tag{7.9}$$

Then the Master equation (4.6) for $P(\phi,t)$ becomes the Fokker–Planck equation[18]

$$\frac{\partial}{\partial t}P(\phi,t) = \left(\lambda\frac{\partial^2}{\partial\phi^2} + \gamma\frac{\partial}{\partial\phi}\phi\right)P(\phi,t) , \tag{7.10}$$

which has the solution, for $t \to \infty$,

$$P(\phi) = (2\pi\sigma^2)^{-1/2}e^{-\phi^2/2\sigma^2} , \quad \sigma^2 = \lambda/\gamma . \tag{7.11}$$

This $P(\phi)$, together with $W(\phi_0|\phi)$ from (7.9), defines a stationary Gaussian–Markov process, the Ornstein–Uhlenbeck process. In the limit $\gamma \to 0$ and λ finite (so $\sigma^2 \to \infty$), the process $\phi(t)$ reduces to the Wiener–Lévy process from Sec. VII A. From (7.11) we see that $\phi(t)$ is centered around $\phi = 0$. The distribution is Gaussian with a variance σ^2 around the average $\phi = 0$. The preference for $\phi = 0$ expresses that this process can be considered as a model for phase-locked radiation.

With the specific choice (7.9) for the transition rate, the Master equation (6.5) assumes the form of a second-order partial differential equation. We obtain

$$\begin{aligned} &\left(i\frac{\partial}{\partial t} - A + i\lambda B^2\right)\zeta(\phi,t) \\ &\quad = i\gamma\left\{\sigma^2\left(2iB + \frac{\partial}{\partial\phi}\right)\frac{\partial}{\partial\phi} + \left(iB + \frac{\partial}{\partial\phi}\right)\phi\right\}\zeta(\phi,t) . \end{aligned} \tag{7.12}$$

In the limit $\gamma \to 0$ and $\lambda = \gamma\sigma^2$ finite, we recover (the Fourier inverse of) Eq. (7.3) with $\hat{W}(\rho)$ from Eq. (7.7).

In order to obtain a solution of Eq. (7.12), we start with a Fourier transform with respect to ϕ. The transformed equation then reads

$$\begin{aligned} &\left(i\frac{\partial}{\partial t} - A + i\lambda B^2\right)\hat{\zeta}(\rho,t) \\ &\quad = -i\gamma\left\{\sigma^2(\rho - 2B)\rho + (\rho - B)\frac{\partial}{\partial\rho}\right\}\hat{\zeta}(\rho,t) , \end{aligned} \tag{7.13}$$

which is still a partial differential equation. Since we are interested in $\hat{\zeta}(0,t) = \langle\sigma(t)\rangle$, the obvious approach[23] would

be a Taylor expansion around $\rho = 0$. This yields however a cumbersome inhomogeneous four–term recurrence relation for the Taylor coefficients. This can be avoided by the transformation[15]

$$\hat{\xi}(\rho,t) = \hat{P}(\rho)\hat{g}(\rho,t) , \tag{7.14}$$

which defines $\hat{g}(\rho,t)$. The Fourier transform of the probability distribution is explicitly

$$\hat{P}(\rho) = \langle e^{i\rho\phi(t)} \rangle = e^{-(1/2)\sigma^2\rho^2} , \tag{7.15}$$

and in particular we have $\hat{P}(0) = 1$. The equation for $\hat{g}(\rho,t)$ becomes

$$\left(i\frac{\partial}{\partial t} - A + i\lambda B^2\right)\hat{g}(\rho,t) = -i\gamma\left\{\rho\frac{\partial}{\partial\rho} - B\left(\sigma^2\rho + \frac{\partial}{\partial\rho}\right)\right\}\hat{g}(\rho,t) , \tag{7.16}$$

and it has to be solved for

$$\hat{g}(0,t) = \langle \sigma(t) \rangle . \tag{7.17}$$

Let us define the Taylor coefficients $\pi_n(t)$ by the expansion

$$\hat{g}(\rho,t) = \sum_{n=0}^{\infty} \frac{(i\rho)^n}{n!}\pi_n(t) , \tag{7.18}$$

which can be inverted as

$$\pi_n(t) = \left\langle \sigma(t)\left(\frac{\partial}{\partial i\rho}\right)^n e^{i\rho\phi(t) - \frac{1}{2}(i\rho\sigma)^2} \right\rangle_{\rho=0} . \tag{7.19}$$

Substitution of (7.18) into (7.16) then gives the equation for the Taylor coefficients

$$\left(i\frac{\partial}{\partial t} - A + i\lambda B^2 + i\gamma n\right)\pi_n(t) = \gamma B\{n\sigma^2\pi_{n-1}(t) - \pi_{n+1}(t)\} , \tag{7.20}$$

which has to be solved for

$$\pi_0(t) = \langle \sigma(t) \rangle . \tag{7.21}$$

Equation (7.20) looks like a homogeneous three-term recurrence relation, but it will be shown below that the time derivative $\partial/\partial t$ gives rise to an inhomogeneous contribution. Notice that for $n = 0$ Eq. (7.20) reduces to a two-term relation between $\pi_0(t)$ and $\pi_1(t)$ only.

Equation (7.20) is most easily solved in the Laplace domain. If we introduce

$$\tilde{\pi}_n(\omega) = \int_{t_0}^{\infty} dt\, e^{i\omega(t-t_0)}\pi_n(t) , \tag{7.22}$$

then (7.20) attains the form

$$(\omega - A + i\lambda B^2 + i\gamma n)\tilde{\pi}(\omega) - \gamma B\{n\sigma^2\tilde{\pi}_{n-1}(\omega) - \tilde{\pi}_{n+1}(\omega)\} = i\pi_n(t_0) . \tag{7.23}$$

Here the initial values $\pi_n(t_0)$, for $n = 0, 1, 2,...$, appear as inhomogeneous terms. The set $\pi_n(t_0)$ for all n represents the initial correlations of $\sigma(t)$ on $t = t_0$, and they connect the time evolution of $\langle \sigma(t) \rangle$ for $t > t_0$ to its recent past.[13] In other words, Eq. (7.23) relates the set $\pi_n(t)$ for $t > t_0$ to the initial set $\pi_n(t_0)$.

Equation (7.23) can be solved for an arbitrary initial set $\pi_n(t_0)$ by standard techniques,[24] but the solution is not transparent. In order to elucidate the structure of the solution, we assume a nonstochastic initial state $\sigma(t_0)$. From Eq. (7.1) we then find at $t = t_0$

$$\hat{\xi}(\rho,t_0) = \hat{P}(\rho)\sigma(t_0) , \tag{7.24}$$

and from Eq. (7.14) we obtain

$$\hat{g}(\rho,t_0) = \sigma(t_0) . \tag{7.25}$$

Hence at $t = t_0$ the vector $\hat{g}(\rho,t_0)$ is independent of ρ, and therefore the expansion coefficients are simply

$$\pi_n(t_0) = \delta_{n,0}\sigma(t_0) . \tag{7.26}$$

Then only the recurrence relation for $n = 0$ is inhomogeneous, and the solution of (7.23) for $\tilde{\pi}_0(\omega) = \langle \tilde{\sigma}(\omega) \rangle$ is readily found to be

$$\langle \tilde{\sigma}(\omega) \rangle = \{i/[\omega - A + i\lambda B^2 + \tilde{K}(\omega)]\}\sigma(t_0) . \tag{7.27}$$

The effect of the finite correlation time, e.g., the deviation from the Wiener–Lévy limit, is accounted for by the operator

$$\tilde{K}(\omega) = \gamma B \cfrac{1\sigma^2}{\omega - A + i\lambda B^2 + 1i\gamma + \gamma B \cfrac{2\sigma^2}{\omega - A + i\lambda B^2 + 2i\gamma + \gamma B \cfrac{3\sigma^2}{\ddots}\gamma B}\gamma B} \tag{7.28}$$

which indeed vanishes for $\gamma \to 0$, λ finite. In this limit, Eq. (7.27) is the Laplace transform of Eq. (7.8).

The explicit expression (7.27) provides the exact solution for situations where the initial state is nonstochastic and for cases where the solution is independent of the initial state. As an example from quantum optics, we mention that Eq. (7.27) with $\sigma(t_0) = 1$, $A = 0$, and $B = 1$ represents the laser spectral profile. Another example pertains to the long-time behavior of the solution $\langle \sigma(t) \rangle$. If there is any damping in the system, which might be caused by the stochastic fluctuations itself, then the solution for $t \gg t_0$ will become independent of the initial state. If we indicate by $\bar{\sigma}$ the solution $\langle \sigma(t) \rangle$ for $t \to \infty$, then $\bar{\sigma}$ obviously obeys the equation

$$\{A - i\lambda B^2 - \tilde{K}(0)\}\bar{\sigma} = 0 . \tag{7.29}$$

For the problem of atomic fluorescence in a strong laser field, this is the equation for the atomic steady-state density matrix, which determines the fluorescence yield. There, the solution $\bar{\sigma}$ of Eq. (7.29) is unique.

VIII. CONCLUSIONS

Solving the multiplicative stochastic process $\sigma(t)$ for its average is rarely feasible by analytical methods. This is mainly due to the finite correlation time of the driving process $\phi(t)$, which prohibits the factorization of the average of a product into the product of the averages. Averages of a functional of $\phi(t)$ might factorize if the process is delta correlated. For Markov processes, however, we can simulate a δ correlation by the introduction of the marginal average

$\zeta(\phi_0,t) = \langle \delta(\phi(t) - \phi_0)\sigma(t)\rangle$. The combination of the multiplication by $\delta(\phi(t) - \phi_0)$ and the Markov property of the probability distributions of $\phi(t)$ then gives rise to a factorizationlike result for the formal expression for the average. Along the same lines as in a factorization assumption, we can now derive exact equations for $\zeta(\phi_0,t)$. In this paper we have studied Eq. (1.1), where we considered the stochastics of $\phi(t)$ to be given. The equation of motion for the marginal average turned out to be Eq. (6.4). The applicability of this equation was illustrated by some examples.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under Grant No. CHE–8519053, the Air Force Office of Scientific Research (AFSC), United States Air Force, under Contract No. F49620–86–C–0009, and the Office of Naval Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

[1]G. S. Agarwal, Phys. Rev. A **18**, 1490 (1978).
[2]A. T. George and S. N. Dixit, Phys. Rev. A **23**, 2580 (1981).
[3]H. F. Arnoldus and G. Nienhuis, J. Phys. B **19**, 2421 (1986).
[4]R. F. Fox, J. Math. Phys. **13**, 1196 (1972).
[5]S. N. Dixit, P. Zoller, and P. Lambropoulos, Phys. Rev. A **21**, 1289 (1980).
[6]P. Zoller, G. Alber and R. Salvador, Phys. Rev. A **24**, 398 (1981).
[7]R. I. Jackson and S. Swain, J. Phys. B **15**, 4375 (1982).
[8]H. F. Arnoldus and G. Nienhuis, J. Phys. B **16**, 2325 (1983).
[9]K. Wódkiewicz and J. H. Eberly, Phys. Rev. A **31**, 2314 (1985).
[10]A. Brissaud and U. Frisch, J. Quant. Spectrosc. Radiat. Transfer **11**, 1767 (1971).
[11]A. Brissaud and U. Frisch, J. Math. Phys. **15**, 524 (1974).
[12]V. E. Shapiro and V. M. Loginov, Physica A **91**, 563 (1978).
[13]H. F. Arnoldus and G. Nienhuis, J. Phys. A **19**, 1629 (1986).
[14]B. W. Shore, J. Opt. Soc. Am. B **1**, 183 (1984).
[15]K. Wódkiewicz, B. W. Shore, and J. H. Eberly, Phys. Rev. A **30**, 2390 (1984).
[16]P. Zoller and J. Cooper, Phys. Rev. A **28**, 2310 (1983).
[17]J. D. Cresser, W. H. Louisell, P. Meystre, W. Schleich, and M. O. Scully, Phys. Rev. A **25**, 2214 (1982).
[18]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[19]B. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963).
[20]R. F. Fox, Phys. Rev. A **33**, 467 (1986).
[21]N. G. van Kampen, Phys. Lett. A **76**, 104 (1980).
[22]A. Burshtein, Sov. Phys. JETP **21**, 567 (1965).
[23]J. J. Yeh and J. H. Eberly, Phys. Rev. A **24**, 888 (1981).
[24]H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications* (Springer, Berlin, in press).

On the integrability of multidimensional nonlinear evolution equations

H. H. Chen
Laboratory for Plasma and Fusion Energy Studies, University of Maryland, College Park, Maryland 20742
J. E. Lin
Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030

(Received 18 July 1986; accepted for publication 15 October 1986)

The integrability-test scheme of Chen, Lee, and Liu [H. H. Chen, Y. C. Lee, and C. S. Liu, Phys. Scr. **20**, 490 (1979)] from one-space dimension to multispace dimensions is generalized. The temporal equation of the Lax pair is still the linearized perturbed equation that defines the symmetries. But the spectral operator in the Lax pair is no longer the linear recursion operator for symmetries. The absence of the linear recursion operator for symmetries in higher spatial dimensions therefore presents no direct obstacle to the Chen–Lee–Liu test scheme. The Kadomtsev–Petviashvili equation is shown as an example.

I. INTRODUCTION

In the past two decades, the discovery of the soliton solutions for certain nonlinear evolution equations with physical applications has aroused great interest and attention among mathematicians and physicists.[1–5] The initial-value problems of these equations $u_t = K(u)$ can be formally solved by the inverse scattering method[1–5]: In general, we can associate a pair (Lax pair) of operators $A = A(u)$ and $L = L(u)$ to each of these equations such that these two operators satisfy the Lax condition,[2]

$$L_t = AL - LA, \tag{1}$$

hence, the eigenvalue problem

$$L\phi = \lambda\phi \tag{2a}$$

has constant eigenvalues, that is, $\lambda_t = 0$, if ϕ also satisfies

$$\phi_t = A\phi. \tag{2b}$$

Therefore, an important problem is to find a way to construct the pair of Lax operators for a given equation, if it possesses such a pair, to establish its integrability.

In the work of Chen, Lee, and Liu,[6] they pointed out a very important fact about an integrable nonlinear evolution equation $u_t = K(u)$ with one spatial dimension: The Lax-pair operators A and L in (1) can be identified as the Gateaux derivative K' of $K(u)$ (see Definition 1 in Sec. II) and a recursion operator that maps a symmetry to another symmetry of the equation, respectively. In other words, the existence of the recursion operator of the symmetries of an equation with one spatial dimension is equivalent to its integrability. In fact, for an integrable equation with one spatial dimension, the authors together with Lee[7] recently have found a method to construct the recursion operator of the symmetries from the knowledge of merely one symmetry propagator (see Definition 4 in Sec. II) of the equation. Take the Korteweg–de Vries equation as an example,

$$u_t = K(u) = 6uu_x + u_{xxx},$$

the recursion operator L of its symmetries can be constructed from the symmetry propagator[8] with the lowest "rank"

$$\tau_2 = x(u_{xxx} + 6uu_x) + 3t(u_{xxxxx} + 10uu_{xxx} + 20u_xu_{xx} + 30u^2u_x) + 4u_{xx} + 8u^2 + 2u_x\partial_x^{-1}u,$$

where ∂_x^{-1} denotes $\int_{-\infty}^{x} dx$. The relation is that

$$L = \tfrac{1}{4}(\partial_x\tau_2'^{*}\partial_x^{-1} + \tau_2') = \partial_x^2 + 4u + 2u_x\partial_x^{-1}.$$

One can easily verify that $L_t = AL - LA$, where $A = K' = 6u\partial_x + 6u_x + \partial_x^3$. The hierarchies of the symmetry propagators τ_n for the integrable nonlinear evolution equations were first discovered by the authors with Lee.[8–10] However, the above scheme does not work for the integrable equations with higher spatial dimensions. Take the Kadomtsev–Petviashvili equation as an example,

$$u_t = \partial_x^{-1}u_{yy} - 6uu_x - u_{xxx}; \tag{3}$$

it can be verified that $\partial_x\tau_3'^{*}\partial_x^{-1} + \tau_3' = 0$, where τ_3 is the symmetry propagator[10] with the lowest rank for the Kadomtsev–Petviashvili equation. Therefore, the recursion operator constructed from the above scheme is the zero operator. It is worth mentioning that no recursion operator has been found for any integrable equation with higher spatial dimensions.[11]

In this work, we want to point out that a condition more general than the Lax condition (1) can be used to establish the integrability of a nonlinear evolution equation with higher spatial dimensions,

$$L_t = BL - LA. \tag{4}$$

More precisely, given a nonlinear evolution equation with higher spatial dimensions, $u_t = K(u)$. Let A be K', the Gateaux derivative of K. If we can find an operator $L = L(u)$ such that $L_t = BL - LA$ for some operator $B = B(u)$, then the equation is integrable, a Miura transformation[12] can be found, and the corresponding inverse scattering problem is now given by the system

$$L\phi = 0, \quad \phi_t = A\phi. \tag{5}$$

Note the absence of $\lambda\phi$ on the right-hand side of the first

 0022-2488/87/020347-04$02.50 © 1987 American Institute of Physics

equation of the system (5). This is the difference between (5) and (2a), (2b). However, the gain is that we do not have to look for the recursion operator of an equation with higher spatial dimensions to determine its integrability. We can also derive infinite hierarchies of the symmetries of the equation through (5).

In Sec. II below, we shall give the definitions of the terminologies which appear in this paper. Our main result is presented in Sec. III. The Kadomtsev–Petviashvili equation (3) is used as a prototype in our work. The equivalence between our formulation (5) with (4) and the Lax formulation (2a) and (2b) with (1) is also demonstrated in Sec. III (Remarks 4 and 5). The derivation of the infinite hierarchies of the symmetries of the Kadomtsev–Petviashvili equation (3) is given in Sec. IV.

II. DEFINITIONS OF TERMINOLOGIES

Definition 1: Given a function $F(u)=F(x,...,t,u,u_x,...)$ which depends on u and its partial derivatives and possibly on the variables $x,...$, and t,

$$F'[v]\equiv\left.\frac{\partial F(u+\epsilon v)}{\partial\epsilon}\right|_{\epsilon=0}$$

is the Gateaux derivative of F in the direction v with respect to u.

Definition 2: Given a nonlinear evolution equation $u_t=K(u)$, where $K(u)$ depends on u and its partial derivatives and possibly on the variables $x,...$, and t, a function $s(x,...,t,u,u_x,...)$ is called a symmetry (generator) of the equation if s satisfies the linearized equation of $u_t=K(u)$,

$$\frac{\partial s}{\partial t}=K'[s],$$

where $K'[s]$ is defined as in Definition 1. These symmetries are the infinitesimal generators of one-parameter groups of invariant transformations of the equation.

Definition 3: The Lie product of F and G, $[F,G]$, is defined by $[F,G]\equiv F'[G]-G'[F]$, where $F'[G]$ and $G'[F]$ are defined as in Definition 1.

Definition 4: A symmetry of an equation is called a symmetry propagator of the equation if the Lie product of this symmetry with another symmetry of this equation gives a new symmetry of this equation, where the Lie product is defined in Definition 3.

III. THE KADOMTSEV–PETVIASHVILI EQUATION

Consider now the Kadomtsev-Petviashvili (KP) equation (3). Let $K(u)=\partial_x^{-1}u_{yy}-6uu_x-u_{xxx}$. We have

$$A=K'=\partial_x^{-1}\partial_y^2-6u\,\partial_x-6u_x-\partial_x^3. \tag{6}$$

Let

$$L=\partial_x-v_x\,\partial_x^{-1}-v-(1/\sqrt{3})i\,\partial_x^{-1}\partial_y, \tag{7}$$

then

$$AL-LA-L_t=f+g\partial_x^{-1}+h\partial_x+NL,$$

where

$$f=v_t+6uv_x+v_{xxx}-\sqrt{3}iv_{xy}-\sqrt{3}ivv_y+3v_x^2-2\sqrt{3}iu_y+6p_x-6vp,$$

$$p=u_x+\tfrac{1}{2}v_{xx}+\tfrac{1}{2}vv_x+(1/2\sqrt{3})iv_y,$$

$$g=f_x-6p_{xx}+2\sqrt{3}ip_y+6vp_x,\quad h=12p,$$

and

$$N=-6p-\sqrt{3}iv_y+\sqrt{3}iv_x\,\partial_x^{-1}\partial_y+3v_{xx}+3v_x\,\partial_x.$$

Therefore, in order to have (4), $L_t=BL-LA$, it follows that $f=0=p=g=h$. Hence, from $p=0$ and $f=0$,

$$u=-\tfrac{1}{2}v_x-\tfrac{1}{4}v^2-(1/2\sqrt{3})i\,\partial_x^{-1}v_y+\lambda, \tag{8}$$

where λ is a constant, and v satisfies

$$v_t-\tfrac{3}{2}v^2v_x-\sqrt{3}iv_x\,\partial_x^{-1}v_y+v_{xxx}-\partial_x^{-1}v_{yy}+6\lambda v_x=0. \tag{9}$$

Remark 1: Note that we can actually show that (8) implies

$$u_t+6uu_x+u_{xxx}-\partial_x^{-1}u_{yy}=M(v_t-\tfrac{3}{2}v^2v_x-\sqrt{3}iv_x\,\partial_x^{-1}v_y+v_{xxx}-\partial_x^{-1}v_{yy}+6\lambda v_x),$$

where

$$M=-\tfrac{1}{2}\partial_x-\tfrac{1}{2}v-(1/2\sqrt{3})i\,\partial_x^{-1}\partial_y.$$

Thus (8) and (9) imply that u satisfies the KP equation (3). Therefore, (8) is a Miura transformation and (9) is a modified KP equation.

Remark 2: The corresponding inverse scattering problem to (4) is

$$L\phi=(\partial_x-v_x\,\partial_x^{-1}-v-(1/\sqrt{3})i\,\partial_x^{-1}\partial_y)\phi=0,$$

$$\phi_t=A\phi=(\partial_x^{-1}\partial_y^2-6u\,\partial_x-6u_x-\partial_x^3)\phi. \tag{10}$$

Remark 3: Let $v=2\psi_x/\psi$. We can get

$$\psi_{xx}+(1/\sqrt{3})i\psi_y+u\psi=\lambda\psi$$

from (8) and

$$P(\psi_t+4\psi_{xxx}+6u\psi_x+3u_x\psi-\sqrt{3}i(\partial_x^{-1}u_y)\psi)=0,$$

where $P=-\psi_x+\psi\,\partial_x$, from (9). Note that

$$\begin{aligned}&\psi_{xx}+(1/\sqrt{3})i\psi_y+u\psi=\lambda\psi,\\&\psi_t+4\psi_{xxx}+6u\psi_x+3u_x\psi-\sqrt{3}i(\partial_x^{-1}u_y)\psi=0\end{aligned} \tag{11}$$

is the pair of Lax operators found by Dryuma.[13]

Remark 4: We can show by a direct computation that $\phi=(\psi\bar{\psi})_x$, where ϕ is from (10), ψ is from (11), and $\bar{\psi}$ is the complex conjugate of ψ. This justifies the equivalence between our formulation (10) with (4) and the Lax-pair formulation (2a) and (2b) [i.e., (11)] with (1).

Remark 5: As to the modified KP equation (9) with $\lambda=0$,

$$v_t=K_1(v)=\tfrac{3}{2}v^2v_x+\sqrt{3}iv_x\,\partial_x^{-1}v_y-v_{xxx}+\partial_x^{-1}v_{yy},$$

we can show that $(L_1)_t=B_1L_1-L_1A_1$ for some operator B_1, where $L_1=\partial_x-v-(1/\sqrt{3})i\,\partial_x^{-1}\partial_y$, and $A_1=K_1'$. Thus, the corresponding inverse scattering problem is

$$\begin{aligned}&L_1\mu=(\partial_x-v-(1/\sqrt{3})i\,\partial_x^{-1}\partial_y)\mu=0,\\&\mu_t=A_1\mu=(3vv_x+\tfrac{3}{2}v^2\,\partial_x+\sqrt{3}i(\partial_x^{-1}v_y)\partial_x\\&\qquad+\sqrt{3}iv_x\,\partial_x^{-1}\partial_y-\partial_x^3+\partial_x^{-1}\partial_y^2)\mu.\end{aligned} \tag{12}$$

Note that the relation between (12) and the Lax-pair formulation for the modified KP equation[14]

$$(\partial_x^2 - v\,\partial_x - (1/\sqrt{3})i\,\partial_y)\eta = \rho\eta,$$
$$\eta_t = (-4\partial_x^3 - \tfrac{3}{2}v^2\partial_x + \sqrt{3}i(\partial_x^{-1}v_y)\partial_x + 3v_x\,\partial_x + 6v\,\partial_x^2)\eta, \tag{13}$$

can be shown to be $\mu = \exp(-\sqrt{3}i\rho y)\eta_x$. This justifies the equivalence between our formulation (12) and the Lax-pair formulation (13).

IV. INFINITE HIERARCHIES OF THE SYMMETRIES FOR THE KP EQUATION

Let

$$v = \sum_{n=1}^{\infty} k^{-n}v_n - 2k, \quad \phi = \sum_{n=0}^{\infty} k^{-n}\phi_n,$$
$$L = \sum_{n=-1}^{\infty} k^{-n}L_n,$$

where $k = \sqrt{\lambda}$. Then, from (8),

$$u = -\frac{1}{2}\sum_{n=1}^{\infty} k^{-n}(v_n)_x - \frac{1}{4}\left(\sum_{n=1}^{\infty} k^{-n}v_n - 2k\right)^2 - \frac{1}{2\sqrt{3}}i\sum_{n=1}^{\infty} k^{-n}\partial_x^{-1}\partial_y v_n + k^2.$$

Hence,

$$v_1 = u, \quad v_2 = \tfrac{1}{2}u_x + (1/2\sqrt{3})i\,\partial_x^{-1}u_y,$$
$$v_3 = \tfrac{1}{4}u_{xx} + (1/2\sqrt{3})iu_y + \tfrac{1}{4}u^2 - \tfrac{1}{12}\partial_x^{-2}u_{yy},$$

and, in general,

$$v_{n+1} = \frac{1}{2}(v_n)_x + \frac{1}{4}\sum_{m=1}^{n-1} v_m v_{n-m} + \frac{1}{2\sqrt{3}}i\,\partial_x^{-1}(v_n)_y, \quad \text{for } n \geqslant 2.$$

From (7) and (10), we have

$$\sum_{n=-1}^{\infty} k^{-n}L_n = \partial_x - \sum_{n=1}^{\infty} k^{-n}(v_n)_x\,\partial_x^{-1} - \left(\sum_{n=1}^{\infty} k^{-n}v_n - 2k\right) - \frac{1}{\sqrt{3}}i\,\partial_x^{-1}\partial_y,$$
$$\left(\sum_{n=-1}^{\infty} k^{-n}L_n\right)\left(\sum_{n=0}^{\infty} k^{-n}\phi_n\right) = 0, \tag{14}$$

and

$$(\phi_n)_t = A\phi_n. \tag{15}$$

Hence,

$$L_{-1} = 2, \quad L_0 = \partial_x - (1/\sqrt{3})i\,\partial_x^{-1}\partial_y,$$
$$L_1 = -(v_1)_x\,\partial_x^{-1} - v_1,$$

and in general,

$$L_n = -(v_n)_x\partial_x^{-1} - v_n, \quad \text{for } n \geqslant 1.$$

From (14) we have

$$\sum_{m=-1}^{\infty} L_m\phi_{n-m} = 0. \tag{16}$$

Hence,

$$\phi_{n+1} = -\tfrac{1}{2}(L_0\phi_n + L_1\phi_{n-1} + \cdots + L_n\phi_0) = -\frac{1}{2}\left(\partial_x\phi_n - \frac{1}{\sqrt{3}}i\,\partial_y\xi_n - \sum_{m=1}^{n}(v_m)_x\xi_{n-m} - \sum_{m=1}^{n} v_m\phi_{n-m}\right), \tag{17}$$

for $n \geqslant 1$, where $(\xi_m)_x = \phi_m$.

Remark 6: Suppose that we have found ϕ_l for $0 \leqslant l \leqslant n$ and ξ_l for $0 \leqslant l \leqslant n-1$. In order to find ϕ_{n+1}, we have to find ξ_n. To find ξ_n, since $(\xi_n)_x = \phi_n$, we must use Eq. (15) for ϕ_{n+1}, $(\phi_{n+1})_t = A\phi_{n+1}$, to determine the function which does not depend on x and enters into the integration of ϕ_n with respect to x. This function is a polynomial in t and y in the following work.

Remark 7: Since ϕ_n's satisfy the linearized equation of the KP equation, Eq. (15), they are symmetries of the KP equation (cf. Definition 2 in Sec. II).

We are now ready to derive the infinite hierarchies of symmetries ϕ_n for the KP equation. The ξ_n's needed for deriving ϕ_{n+1}'s are also given for the sake of the completeness of the presentation.

From (16), we have $\phi_0 = 0$ and $(\phi_1)_x = 0$, hence, $\phi_1 = 0$ in order for ϕ_1 to be a symmetry. We then choose $\xi_0 = t^m$ and $\xi_1 = (m/2\sqrt{3})it^{m-1}y$ so that ϕ_2 is a symmetry, where m is a non-negative integer. Hierarchies (σ_n^m, $m,n = 0,1,2,3,\ldots$): From (17), we have

$$\sigma_0^m = \phi_2 = \tfrac{1}{2}t^m u_x - (m/12)t^{m-1},$$
$$\sigma_1^m = \phi_3 = (1/2\sqrt{3})i(t^m u_y + (m/2)t^{m-1}yu_x - (m(m-1)/12)t^{m-2}y),$$
$$\sigma_2^m = \phi_4 = -\tfrac{1}{8}(t^m K + (2m/3)t^{m-1}yu_y + (m/3)t^{m-1}xu_x + (2m/3)t^{m-1}u - (m(m-1)/18)t^{m-2}x + (m(m-1)/6)t^{m-2}y^2u_x - (m(m-1)(m-2)/36)t^{m-3}y^2),$$

where

$$K = K(u) = \partial_x^{-1}u_{yy} - 6uu_x - u_{xxx},$$
$$\sigma_3^m = \phi_5 = (-1/4\sqrt{3})it^m(\tfrac{1}{3}\partial_x^{-2}u_{yyy} - u_{xxy} - 4uu_y - 2u_x\,\partial_x^{-1}u_y) - (m/16\sqrt{3})it^{m-1}yK - (m/24\sqrt{3})it^{m-1}(2\,\partial_x^{-1}u_y + xu_y) - (m(m-1)/48\sqrt{3})it^{m-2}(2yu + y^2u_y + xyu_x) - (m(m-1)(m-2)/288\sqrt{3})it^{m-3}(y^3u_x - xy) + (m(m-1)(m-2)(m-3)/1728\sqrt{3})it^{m-4}y^3,$$

etc. Note that in the above derivation, ξ_2, ξ_3, and ξ_4 were found to be

$$\xi_2 = \tfrac{1}{2} t^m u - (m/12) t^{m-1} x - (m(m-1)/24) t^{m-2} y^2,$$

$$\xi_3 = (1/2\sqrt{3}) i (t^m \partial_x^{-1} u_y + (m/2) t^{m-1} yu - (m(m-1)/12) t^{m-2} xy - (m(m-1)(m-2)/72) t^{m-3} y^3),$$

and

$$\begin{aligned}\xi_4 = & -\tfrac{1}{8}(t^m \partial_x^{-1} K + (2m/3) t^{m-1} y \partial_x^{-1} u_y + (m/3) t^{m-1} xu + (m/3) t^{m-1} \partial_x^{-1} u \\ & - (m(m-1)/36) t^{m-2} x^2 + (m(m-1)/6) t^{m-2} y^2 u - (m(m-1)/6\sqrt{3}) i t^{m-2} y \\ & - (m(m-1)(m-2)/36) t^{m-3} xy^2 - (m(m-1)(m-2)(m-3)/432) t^{m-4} y^4),\end{aligned}$$

respectively.

Remark 8: σ_n^0's are the well-known symmetries[15] which can be derived from the conserved quantities[16,17] which do not depend explicitly on the variables x, y, and t and σ_n^1's are the new symmetries[8,10] that depend explicitly and linearly on the variables x, y, and t.

Remark 9: The general hierarchies σ_n^m's were also derived in the work of Chen, Lee, and Zhu[18] by a different approach. In that work, they also derived a relation among these infinite hierarchies of symmetries:

$$[\sigma_n^m, \sigma_s^r] = ((m(s+1) - r(n+1))/16)\, \sigma_{n+s-2}^{m+r-1},$$

$$\text{for } m + r \geqslant 1 \text{ and } n + s \geqslant 2.$$

Therefore, we need only *three* elements, namely, σ_0^0, σ_2^2, and σ_3^1, to generate the whole set of symmetries. Note that there are conserved quantities[18] corresponding to the σ_n^m's. Also the σ_n^m's are symmetry propagators for $m \geqslant 2$ or $n \geqslant 3$.

In summary, the integrability-test scheme proposed by Chen, Lee, and Liu,[6] namely, to identify the temporal equation and the spectral operator of the Lax pair as the linearized perturbed equation that defines the symmetries and the linear recursion operator of the symmetries, respectively, for an integrable equation with one spatial dimension, can be generalized to integrable equations with higher spatial dimensions. The temporal equation is still the linearized perturbed equation that defines the symmetries but the spectral operator is no longer the linear recursion operator of the symmetries and the Lax condition has to be modified in our formulation [cf. (1), (2a), (2b), and (4) and (5)].

ACKNOWLEDGMENTS

The work of the first author is supported in part by NSF Grant No. PHYS-80-14189 AOL. The second author is supported by the George Mason University 1986 Summer Stipend Grant.

[1] C. Gardner, J. Greene, M. Kruskal, and R. Miura, Commun. Pure Appl. Math. **27**, 97 (1974).
[2] P. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).
[3] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[4] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[5] S. Novikov, S. Manakov, L. Pitaevskii, and V. E. Zakharov, *Theory of Solitons* (Consultants Bureau, New York, 1984).
[6] H. H. Chen, Y. C. Lee, and C. S. Liu, Phys. Scr. **20**, 490 (1979).
[7] H. H. Chen, J. E. Lin, and Y. C. Lee, "On the direct construction of the inverse scattering operators of nonlinear Hamiltonian systems," University of Maryland preprint, 1985.
[8] H. H. Chen, Y. C. Lee, and J. E. Lin, in *Advances in Nonlinear Waves*, edited by L. Debnath (Pitman, New York, 1985), Vol. 2, p. 233.
[9] H. H. Chen, Y. C. Lee, and J. E. Lin, Phys. Lett. A **91**, 381 (1982).
[10] H. H. Chen, Y. C. Lee, and J. E. Lin, Physica D **9**, 439 (1983).
[11] V. E. Zakharov and B. G. Konopelchenko, Commun. Math. Phys. **94**, 483 (1984).
[12] R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[13] V. Dryuma, Sov. Phys. JETP Lett. **19**, 387 (1974).
[14] W.-H. Steeb, M. Kloke, B.-M. Spieker, and D. Grensing, Prog. Theor. Phys. **73**, 344 (1985).
[15] C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[16] V. E. Zakharov and E. I. Schulman, Physica D **1**, 192 (1980).
[17] J. E. Lin and H. H. Chen, Phys. Lett. A **89**, 163 (1982).
[18] H. H. Chen, Y. C. Lee, and G. C. Zhu, "Symmetries and Lie algebra for the Kadomtsev–Petviashvili equation," University of Maryland Plasma Preprint UMLPF No. 85-016, 1984.

Continuity of scattering data for particles on the line with directed repulsive interactions

Eugene Gutkin
Department of Mathematics, DRB-306, University of Southern California, University Park, Los Angeles, California 90089-1113

(Received 5 March 1986; accepted for publication 1 October 1986)

The mapping into the remote future for particles on the line interacting by repulsive directed forces is considered. Under suitable assumptions on the forces, it is proved that the mapping (which assigns the scattering data to the initial data) is continuous on a nonempty open set in the phase space.

I. INTRODUCTION

Let us consider the classical mechanical system of N point masses in the space $\mathbf{R}^d$ of d dimensions interacting by repulsive forces. It has been shown in Refs. 1 and 2 (under additional technical assumptions) that the asymptotic velocities of particles $v_i = \lim_{t\to\infty} v_i(t)$ exist for any trajectory of the system. The same is true for the dynamics determined by Hamiltonians with the cone potentials (Ref. 3, see also Sec. II).

Let X^{2n} be the phase space of the Hamiltonian system with n degrees of freedom with the property above, that is for any $x\in X$ the motion along the trajectory $x(t)$ with $x(0) = x$ has the maximal number of asymptotic velocities as $t\to\infty$. Coordinates of these velocities in some basis determine n functions $f_1(x),...,f_n(x)$ on X, which, as we pointed out in Ref. 3, are functionally independent and Poisson commute, $\{f_i,f_j\} = 0$ for all i and j. In other words, if the asymptotic velocities exist for any trajectory of a Hamiltonian system then the system is completely integrable.

The proof of this given in Ref. 3 in the case of cone potentials assumes that the functions $f_1,...,f_n$ on X are continuously differentiable. As F. Calogero and M. Kruskal (private communications) pointed out, the differentiability of asymptotic velocities can not be taken for granted. In fact, the question turns out to be quite difficult.

In this paper we investigate the case of particles on the line, i.e., $d = 1$, interacting by repulsive directed forces (see Sec. II). Assuming that the forces of interaction are continuous and decay sufficiently fast at infinity we prove that there exists an invariant (under dynamics) nonempty open set Y_+ of initial data such that the asymptotic velocities and the asymptotic phases (at $t = +\infty$) are continuous functions on Y_+. Naturally, the same is true for the asymptotic velocities and phases at $t = -\infty$.

If the forces have finite range then we show in Theorem 1 that the scattering data are smooth on Y_+.

The main techniques of the paper are the Hamiltonians with cone potential (cf. Refs. 3 and 4). Theorem 7 which is the main result of the paper follows from a more general assertion (Theorem 6) about the continuity of scattering data for the Hamiltonian dynamics with cone potentials. We will continue the study of regularity of the scattering data in forthcoming publications.

The material of the paper is new if the number of interacting particles is greater than 2. For two particles the problem reduces to the motion of one particle in an external field. Scattering in this case has been treated in detail in Ref. 5.

II. THE SETTING. THE CASE OF FINITE RANGE DIRECTED INTERACTIONS

Let for $1\leqslant i<j\leqslant n$ the functions p_{ij} be defined on (L_{ij},∞), where $-\infty\leqslant L_{ij}$. We assume that for any $i<j$ the function p_{ij} is either identically zero or non-negative, continuously differentiable, and monotonically decreases from ∞ to 0 when x runs form L_{ij} to infinity. We also assume that the derivatives p'_{ij} monotonically increase from minus infinity to zero. Let $m_1,...,m_n$ be the mass of particles. The Hamiltonian of our system of n particles with pair potentials p_{ij} is given by

$$H = \frac{1}{2}(m_1\dot{x}_1^2 + \cdots + m_n\dot{x}_n^2) + \sum_{i<j} p_{ij}(x_j - x_i)\,, \qquad (1)$$

where $\dot{x}_i = v_i$ are the particle velocities. We say that the particles with Hamiltonian (1) interact via pairwise directed potentials p_{ij}. The configuration space of the system (1) consists of n-tuples $(x_1,...,x_n)$ such that $L_{ij} < x_j - x_i$ for all $i<j$.

Let now W be a non-negative C^1 function on the configuration space. Assume that $\partial W/\partial x_i \leqslant 0$ for all i and consider the Hamiltonian

$$H = \frac{1}{2}\sum_{i=1}^{n} m_i v_i^2 + \sum_{i<j} p_{ij}(x_j - x_i) + W(x_1,...,x_n)\,. \qquad (2)$$

We say that the total potential

$$V = \sum_{i<j} p_{ij}(x_j - x_i) + W(x_1,...,x_n) \qquad (3)$$

is the sum of the internal potential $\Sigma_{i<j} p_{ij}(x_j - x_i)$ and the external potential W. An important special case of (3) is

$$W(x_1,...,x_n) = w_1(x_1) + \cdots + w_n(x_n)\,, \qquad (4)$$

with $w_i(x)\geqslant 0$ and $w_i'(x)\leqslant 0$.

We denote by C the proper (i.e., without straight lines) cone in $\mathbf{R}^n$ spanned by $e_j - e_i$ for $i<j$ and by e_i, where $e_1,...,e_n$ is the standard basis of $\mathbf{R}^n$. Then the total potential V is a cone potential in the sense of Ref. 3 with respect to the cone $-C$. The vector $F(x) = -\nabla V(x)$ is the force at the configuration x. Denote $-p'_{ij}(x)$ by $f_{ij}(x)$ and $-\nabla W(x)$ by $G(x)$. Then the force at $x = (x_1,...,x_n)$ is given by

 0022-2488/87/020351-09$02.50 © 1987 American Institute of Physics

$$F(x)=\sum_{i<j} f_{ij}(x_j-x_i)(e_j-e_i)+G(x)\,, \tag{5}$$

where, if the external potential is given by (4), $G(x)=\Sigma_i g_i(x_i)\,e_i$ with $g_i(x)=-w_i'(x)$. The term $f_{ij}\times(x_j-x_i)(e_j-e_i)$ accounts for the interaction of the ith and the jth particles and the total force F is directed with respect to the cone C (see Ref. 4).

Let $\ddot{x}_i$ be the acceleration of the ith particle. The dynamics $x(t)$ of our many-body problem is given by the Newton equations

$$(m_1\ddot{x}_1,...,m_n\ddot{x}_n)=F(x)\,, \tag{6}$$

and we are interested in the asymptotics of $x(t)$ as t goes to infinity.

Under our assumptions on the pair potentials p_{ij} and the external potential W the limits $\lim_{t\to\pm\infty} v_i(t)=v_i(\pm\infty)$ exist (cf. Ref. 3) for any trajectory $\{x(t)\}$ of the dynamics (6). Denote the vectors $(v_1(-\infty),...,v_n(-\infty))$ and $(v_1(+\infty),...,v_n(+\infty))$ by v_- and v_+, respectively. They are called the asymptotic velocities and the dynamics (6) such that the asymptotic velocities exist for every trajectory is called asymptotically free. Under certain decay assumptions on the potentials the asymptotics of almost all trajectories $\{x(t)\}$ satisfies

$$x(t)=v_\pm t+a_\pm+o(1),\quad \dot{x}(t)=v(t)=v_\pm+o(1)\,, \tag{7}$$

as $t\to\pm\infty$, respectively (cf. Ref. 6). The motions satisfying (7) are called asymptotically uniform and the vectors $a_\pm$ are called the asymptotic phases. The set of pairs (a_-,v_-) obtained from the asymptotics (7) is the domain of the scattering transformation (see, e.g., Ref. 5) S: $(a_-,v_-)\to(a_+,v_+)$ and for this reason (a_-,v_-) or (a_+,v_+) are called the scattering data. In what follows we denote by Y the space of positions and velocities and call Y the phase space (this is a slight abuse of terminology). Let (x,v) be a point in Y and let $\{x(t)\}$ by the trajectory such that $x(0)=x$, $\dot{x}(0)=v$. Assigning to (x,v) the scattering data (a_+,v_+) [resp. (a_-,v_-)] on the trajectory $\{x(t)\}$ we obtain the mappings W_+ [resp. W_-] defined almost everywhere on Y and we have $S=W_+W_-^{-1}$.

Since $W_\pm$ contain all the information about the scattering and for reasons outlined in the introduction, we are interested in the smoothness of these transformations. It suffices to investigate W_+, which can be called the mapping into the remote future.

In the rest of this section we consider the case of finite range interactions.

Definition 1: Let potentials p_{ij}, $1\leqslant i<j\leqslant n$, and W be as above. Assume that there exists $A>0$ such that all $p_{ij}(x)=0$ if $x>A$ and $W(x_1,...,x_n)=0$ if all $|x_k|>A$.

In this situation we say that the n-body system with the Hamiltonian (2) has finite range directed interactions.

Theorem 1: For a system of particles with directed finite range interactions the mappings $W_\pm$ are defined everywhere. There exist nonempty open sets $Y_\pm$ in the phase space such that the restrictions of $W_\pm$ on $Y_\pm$, respectively, are differentiable at least as many times as the functions p_{ij} and W.

Proof: Let $\{x(t)\}$ be the trajectory with the initial data (x,v) and let $v_+=(v_1(\infty),...,v_n(\infty))$ be the asymptotic velocity. If all potentials are zero then the particles are free and there is nothing to prove. Assume that not all p_{ij} are equal to zero. Let $i<j$ be such that $p_{ij}\neq 0$. Then $v_i(+\infty)\leqslant v_j(\infty)$. Indeed, assuming the opposite we obtain that $x_j(t)-x_i(t)$ goes to minus infinity as $t\to\infty$, thus, $p_{ij}(x_j(t)-x_i(t))$ goes to infinity with t which contradicts the conservation of energy.

Assume, for simplicity of exposition, that $p_{i,i+1}\neq 0$ for all i, i.e., that all nearest neighbors interact. Then, by argument above, $v_1(\infty)\leqslant\cdots\leqslant v_n(\infty)$. Denote by Y_+ the set of initial data for which the strict inequalities hold. Consider first the case when the external potential is zero. For $(x^0,v^0)\in Y_+$ there exist $3\epsilon>0$ such that $v^0_{i+1}(\infty)-v^0_i(\infty)>3\epsilon$ and there exists $T_1>0$ such that $|v^0_i(t)-v^0_i(\infty)|<\epsilon$ for $t>T_1$. Therefore $v^0_{i+1}(t)-v^0_i(t)>\epsilon$ for $t>T_1$, thus there exists $T>0$ such that $x^0_{i+1}(t)-x^0_i(t)>2A$ for $t\geqslant T$. Since particles which are more than A apart are free, we have $v^0_i(t)=v^0_i(\infty)$ and $x^0_i(t)=x^0_i(T)+v^0_i(\infty)\,(t-T)$ for $t\geqslant T$. By standard theorems about the smooth dependence of solutions of ordinary differential equations on initial data (see, e.g., Ref. 7), there exists for any $\delta>0$ a neighborhood Ω of (x^0,v^0) in Y such that for any (x,v) in Ω we have $|v_i(T)-v^0_i(T)|<3\epsilon/2$ and $|x_i(T)-x^0_i(T)|<\delta$. From the latter inequality we obtain that $x_{i+1}(T)-x_i(T)>A$ if δ is small enough and the former one implies that $v_{i+1}(T)-v_i(T)>\epsilon/2$. Therefore for any initial conditions (x,v) in Ω, the particles at time T are so far apart that they do not interact and, besides, they are moving further away from each other. Thus for any (x,v) in Ω, $x(t)=x(T)+v(T)(t-T)$ for $t\geqslant T$. Therefore the restriction of W_+ on Ω is given by

$$v_i(\infty)=v_i(T),\quad a_i=x_i(T)-v_i(T)T\,, \tag{8}$$

for $i=1,...,n$. By (8), $W_+|_\Omega$ is essentially the mapping W^T of translation by T along the trajectories of Eq. (6). By the smooth dependence of solutions on initial data, the mappings W^t are differentiable as many times as the right-hand side of (6), which proves the assertion for $W_+|_\Omega$. The argument also shows that Y_+ is open. It is obviously nonempty since for initial conditions (x,v) such that $x_{i+1}-x_i>A$ and $v_{i+1}-v_i>0$ for all i, we have $v(\infty)=v$.

Let now the external potential W be different from zero. Assume, for simplicity, that there is $L\geqslant-\infty$ such that $W(x)\to\infty$ if any coordinate x_i goes to L. Then the same argument as above shows that for the asymptotic velocities we have $0\leqslant v_1(\infty)\leqslant\cdots\leqslant v_n(\infty)$. Repeating the previous argument with obvious modifications we obtain that W_+ is differentiable on the nonempty open set Y_+ of initial data for which the strict inequalities $0<v_1(\infty)<\cdots<v_n(\infty)$ hold. The theorem is proved.

III. PARTICLES ON THE LINE WITH DIRECTED INTERACTIONS. CONTINUITY OF ASYMPTOTIC VELOCITIES

We consider a system of n particles on the line interacting by pairwise directed potentials p_{ij} and assume that the external potential is zero.

Theorem 2: Consider the n-body problem given by the

Hamiltonian (1) where the pair potentials p_{ij} satisfy the assumption of Sec. II. Let $v_1(\infty),\ldots,v_n(\infty)$ be the asymptotic velocities at plus infinity which are well defined functions on the phase space Y. Then there exists a nonempty invariant open set Y_+ in Y such that $v_1(\infty),\ldots,v_n(\infty)$ are continuous on Y_+.

We will prove the theorem after some preparation. Recall that, by our assumptions, $p_{ij}(x)\to\infty$ when $x\to L_{ij}$ from the right. The values of L_{ij} do not matter in the proof and we set for simplicity $L_{ij}=-\infty$ for all $i<j$.

First we consider the system of three particles with pair potentials p_{13}, p_{12}, and p_{23} and denote by b_1, b_2, b_3 the asymptotic velocities. By Ref. 3, they exist for any initial data (x,v) of the three-body problem and satisfy $b_1\leqslant b_2\leqslant b_3$. We write the Newtonian equations of motion explicitly as

$$\begin{aligned} m_1\ddot{x}_1 &= -f_{12}(x_2-x_1)-f_{13}(x_3-x_1)\,,\\ m_2\ddot{x}_2 &= f_{12}(x_2-x_1)-f_{23}(x_3-x_2)\,,\\ m_3\ddot{x}_3 &= f_{23}(x_3-x_2)+f_{13}(x_3-x_1)\,. \end{aligned} \tag{9}$$

We denote by Y the phase space of our three-body problem and for any A and B we denote by $Y_{A,B}$ the part of Y consisting of pairs (x,v) such that

$$x_2-x_1, x_3-x_2>A;\quad v_2-v_1, v_3-v_2>B\,. \tag{10}$$

Consider the system of equations

$$\begin{aligned} m_1\ddot{x}_1 &= -f_{12}(x_2-x_1)\,,\\ m_2\ddot{x}_2 &= f_{12}(x_2-x_1)\,,\\ m_3\ddot{x}_3 &= f_{23}(x_3-x_2)+f_{13}(x_3-x_1)\,, \end{aligned} \tag{11}$$

which has the Newtonian form $m\ddot{x}=\tilde{F}(x)$ where the force $\tilde{F}$ is no longer conservative. We want to compare solutions of (9) and (11) with the same initial date (x,v).

First of all, for any x the vector $\tilde{F}(x)$ belongs to the span of e_2-e_1 and e_3, i.e., the force $\tilde{F}$ is directed in the sense of Ref. 4. Thus, by results of Ref. 4, for any trajectory $\{\tilde{x}(t)\}$ of (11) the asymptotic velocities $\tilde{b}_1, \tilde{b}_2, \tilde{b}_3$ exist.

Comparing (11) with (9) we see that when the particles are moving according to (11), the first and the second particles are pushing the third to the right with the forces $f_{13}(x_3-x_1)$ and $f_{23}(x_3-x_2)$, respectively, but the third particle does not push back. Let the initial data (x,v) belong to $Y_{A,B}$ with A and B greater than zero, so that the particles at $t=0$ are located in the right order and are moving away from each other. Denote by $(x(t),v(t))$ and by $(\tilde{x}(t),\tilde{v}(t))$ the position and velocity at time $t>0$ for the "real motion" (9) and, respectively, for the "fake motion" (11). Since in the "fake motion" the second particle is not pushed back by the third, we have $\tilde{v}_2(t)\geqslant v_2(t)$, therefore $\tilde{x}_2(t)\geqslant x_2(t)$. The same argument applied to the first and the third particles shows that $\tilde{v}_1(t)\geqslant v_1(t)$ and $\tilde{x}_1(t)\geqslant x_1(t)$. Since, in the motion (11) the first and the second particles are further to the right than they are when moving according to (9), they push the third particle harder, thus, $\tilde{v}_3(t)\geqslant v_3(t)$ and $\tilde{x}_3(t)\geqslant x_3(t)$. Therefore for the asymptotic velocities we have

$$b_1\leqslant\tilde{b}_1,\quad b_2\leqslant\tilde{b}_2,\quad b_3\leqslant\tilde{b}_3\,. \tag{12}$$

In the preceeding argument we assume that the particles moving by (11) do not change their initial order, i.e., the first particle does not catch up with the second and the second particle does not catch up with the third. This will certainly be the case if $\tilde{b}_2\leqslant v_3$. We will now estimate $\tilde{b}_1, \tilde{b}_2, \tilde{b}_3$, thus providing estimates for the asymptotic velocities b_1, b_2, b_3.

The first two equations of (11) describe the Hamiltonian system of two particles with the potential of interaction $p_{12}(x_2-x_1)$. The asymptotic velocities $\tilde{b}_1$ and $\tilde{b}_2$ can be found from the conservation of total momentum

$$m_1v_1+m_2v_2=m_1\tilde{b}_1+m_2\tilde{b}_2\,, \tag{13}$$

and the conservation of energy

$$\tfrac{1}{2}m_1v_1^2+\tfrac{1}{2}m_2v_2^2+p_{12}(x_2-x_1)=\tfrac{1}{2}m_1\tilde{b}_1^2+\tfrac{1}{2}m_2\tilde{b}_2^2\,. \tag{14}$$

After elementary computations we obtain from (13) and (14)

$$\tilde{b}_1=\frac{m_1v_1+m_2v_2}{m_1+m_2}-\left(\frac{m_1m_2}{m_1+m_2}\right)\frac{\{(v_2-v_1)^2+[2(m_1+m_2)/(m_1m_2)]p_{12}(x_2-x_1)\}^{1/2}}{m_1} \tag{15}$$

and

$$\tilde{b}_2=\frac{m_1v_1+m_2v_2}{m_1+m_2}+\left(\frac{m_1m_2}{m_1+m_2}\right)\frac{\{(v_2-v_1)^2+[2(m_1+m_2)/(m_1m_2)]p_{12}(x_2-x_1)\}^{1/2}}{m_2}\,. \tag{16}$$

We could compute $\tilde{b}_3$ from (11), but it is too complicated and, fortunately, we can obtain an upper bound on $\tilde{b}_3$ in a much simpler way. Since the velocity of the second particle in the motion (11) is monotonically increasing approaching the limit $\tilde{b}_2$ as $t\to\infty$, the position of the second particle satisfies $\tilde{x}_2(t)\leqslant x_2+b_2t$ and, of course, $\tilde{x}_1(t)\leqslant\tilde{x}_2(t)\leqslant x_2+\tilde{b}_2t$. If we had the equality, the first and the second particles would have been closer to the third, thus, pushing it stronger to the right. Denote by $\tilde{\tilde{b}}_3$ the limit of $\dot{x}_3(t)$ as $t\to\infty$ from the equation

$$m_3\ddot{x}_3=f_{23}(x_3-x_2-\tilde{b}_2t)+f_{13}(x_3-x_2-\tilde{b}_2t)\,. \tag{17}$$

By (12) and the argument above, we have $b_3\leqslant\tilde{b}_3\leqslant\tilde{\tilde{b}}_3$. From now on we denote $\tilde{\tilde{b}}_3$ by $\tilde{b}_3$ and forget about the old $\tilde{b}_3$.

Now we calculate $\tilde{b}_3$ from (17). Denote $x_3-x_2-\tilde{b}_2t$ by x. Then (17) becomes

$$m_3\ddot{x}=f_{23}(x)+f_{13}(x)\,, \tag{18}$$

with the initial conditions $x(0)=x_3-x_2$, $\dot{x}(0)=v_3-\tilde{b}_2$. Since (18) describes the motion in the field of potential $p_{13}(x)+p_{23}(x)$, the energy $\tfrac{1}{2}m_3\dot{x}^2+p_{23}(x)+p_{13}(x)$ is conserved. Thus, $\lim_{t\to\infty}\dot{x}(t)=\tilde{b}_3-\tilde{b}_2$ satisfies

$$\tfrac{1}{2}m_3(\tilde{b}_3-\tilde{b}_2)^2=\tfrac{1}{2}m_3(v_3-\tilde{b}_2)^2+p_{13}(x_3-x_2)+p_{23}(x_3-x_2)\,. \tag{19}$$

Solving for $\tilde{b}_3$ we obtain

$$\tilde{b}_3=\tilde{b}_2+[(v_3-\tilde{b}_2)^2+(2/m_3)(p_{13}+p_{23})(x_3-x_2)]^{1/2}\,. \tag{20}$$

Now we have formulas (15), (16), and (20) for $\tilde{b}_1$, $\tilde{b}_2$, and $\tilde{b}_3$, respectively, but they are too complicated. Fortunately, by elementary computations that we leave to the reader, we obtain from these formulas the estimates

$$v_1 - [p_{12}(x_2 - x_1)]/m_1(v_2 - v_1) \leqslant \tilde{b}_1 \leqslant v_1 , \tag{21}$$

$$v_2 \leqslant \tilde{b}_2 \leqslant v_2 + [p_{12}(x_2 - x_1)]/m_2(v_2 - v_1) , \tag{22}$$

and

$$v_3 \leqslant \tilde{b}_3 \leqslant v_3 + \frac{m_2(v_2 - v_1)(p_{13} + p_{23})(x_3 - x_2)}{m_3[m_2(v_2 - v_1)(v_3 - v_2) - p_{12}(x_2 - x_1)]} . \tag{23}$$

The estimates above were obtained using the "fake system of particles" where the third particle does not push back on the first and the second Analogously, we can consider another fake system of particles where the first one does not push back on the second and the third. Following the previous argument (we leave details to the reader) we obtain for the corresponding asymptotic velocities $\bar{b}_1$, $\bar{b}_2$, $\bar{b}_3$

$$v_1 - \frac{m_2(v_3 - v_2)(p_{12} + p_{13})(x_2 - x_1)}{m_1[m_2(v_2 - v_1)(v_3 - v_2) - p_{23}(x_3 - x_2)]} \leqslant \bar{b}_1 \leqslant v_1 , \tag{24}$$

and

$$v_2 - p_{23}(x_3 - x_2)/m_2(v_3 - v_2) \leqslant \bar{b}_2 . \tag{25}$$

Parallel to (12) we have

$$\bar{b}_1 \leqslant b_1, \quad \bar{b}_2 \leqslant b_2, \quad \bar{b}_3 \leqslant b_3 . \tag{26}$$

Now we are ready to prove the following.

Lemma 1: Consider the system of three particles on the line with masses m_1, m_2, m_3 interacting via directed potentials p_{12}, p_{23}, p_{13} satisfying assumptions of Sec. II. Let b_1, b_2, b_3 be the asymptotic velocities (at $+\infty$) of the particles corresponding to an initial data (x,v) such that $x_1 < x_2 < x_3$, $v_1 < v_2 < v_3$, and

$$m_2(v_2 - v_1)(v_3 - v_2) > \max(p_{12}(x_2 - x_1), p_{23}(x_3 - x_2)) . \tag{27}$$

Then the following inequalities are satisfied:

$$\begin{aligned} v_1 - \frac{m_2(v_3 - v_2)(p_{12} + p_{13})(x_2 - x_1)}{m_1[m_2(v_2 - v_1)(v_3 - v_2) - p_{23}(x_3 - x_2)]} \\ \leqslant b_1 \leqslant v_1 \leqslant v_2 - \frac{p_{23}(x_3 - x_2)}{m_2(v_3 - v_2)} \leqslant b_2 \leqslant v_2 + \frac{p_{12}(x_2 - x_1)}{m_2(v_2 - v_1)} \\ \leqslant v_3 \leqslant b_3 \leqslant v_3 \\ + \frac{m_2(v_2 - v_1)(p_{13} + p_{23})(x_3 - x_2)}{m_3[m_2(v_2 - v_1)(v_3 - v_2) - p_{12}(x_2 - x_1)]} . \end{aligned} \tag{28}$$

Besides,

$$b_2 - b_1 \geqslant (v_2 - v_1) - \frac{p_{23}(x_3 - x_2)}{m_2(v_3 - v_2)} > 0 , \tag{29}$$

and

$$b_3 - b_2 \geqslant (v_3 - v_2) - \frac{p_{12}(x_2 - x_1)}{m_2(v_2 - v_1)} > 0 . \tag{30}$$

Proof: Recall that the argument with the "fake motions," which we used to obtain the inequalities (21)–(26), is valid if the particles do not overtake each other. This will certainly be the case if $\tilde{b}_2 < v_3$ and $v_1 < \bar{b}_2$. By (22) and (25), it suffices to have

$$v_2 + [p_{12}(x_2 - x_1)]/m_2(v_2 - v_1) < v_3 ,$$

and

$$v_1 < v_2 - [p_{23}(x_3 - x_2)]/m_2(v_3 - v_2) ,$$

which follow from (27). Putting the estimates (12) and (21)–(26) together we obtain the chain of inequalities (28), which implies (29) and (30).

In order to prove Theorem 2 we need the following.

Proposition 1: Consider the system of n particles on the line with masses $m_1,...,m_n$ interacting via directed potentials p_{ij} satisfying assumptions of Sec. II. Denote by $b_1,...,b_n$ the asymptotic velocities (at $+\infty$) of particles along the trajectory starting at a point (x,v) in the phase space Y. Then for any $\epsilon > 0$ there exist numbers A and B such that if $(x,v) \in Y_{A,B}$ then $|b_i - v_i| < \epsilon$ for $i = 1,...,n$.

Proof: We will prove the assertion by induction on n. The case $n = 2$ is obvious and the case $n = 3$ follows immediately from Lemma 1. Assume that the assertion is proved up to $n - 1$. Following the argument of Lemma 1 consider the "fake motion" where the nth particle does not push back on the other particles. Denote by $\tilde{b}_1,...,\tilde{b}_{n-1}$ the asymptotic velocities of the first $n - 1$ particles and by b'_n the asymptotic velocity of the nth particle in this "fake motion." By argument preceeding Lemma 1, we have

$$b_1 \leqslant \tilde{b}_1,...,b_{n-1} \leqslant \tilde{b}_{n-1}, b_n \leqslant b'_n . \tag{31}$$

Assume that $\tilde{b}_{n-1} \leqslant v_n$ and consider the "double fake motion" where the particles $1,...,n-1$ move as before and the motion of the nth particle is given by

$$\begin{aligned} m_n \ddot{x}_n = & f_{1,n}(x_n(t) - x_{n-1} - \tilde{b}_{n-1} t) \\ & + \cdots + f_{n-1,n}(x_n(t) - x_{n-1} - \tilde{b}_{n-1} t) . \end{aligned} \tag{32}$$

Denote by $\tilde{b}_n$ the asymptotic velocity of the nth particle in the motion (32). By our argument preceeding Lemma 1, $b_n \leqslant b'_n \leqslant \tilde{b}_n$. Setting $x(t) = x_n(t) - x_{n-1} - \tilde{b}_{n-1} t$ we transform (32) into

$$m_n \ddot{x} = f_{1n}(x) + \cdots + f_{n-1,n}(x) , \tag{33}$$

which is the equation of motion of a particle of mass m_n in the external potential $p_{1n}(x) + \cdots + p_{n-1,n}$. Using the conservation of energy for (33) and the initial conditions which are $x(0) = x_n - x_{n-1}$, $\dot{x}(0) = v_n - \tilde{b}_{n-1}$ we obtain

$$\begin{aligned} \tilde{b}_n = \tilde{b}_{n-1} + [(v_n - \tilde{b}_{n-1})^2 + (2/m_n)(p_{1n}(x_n - x_{n-1}) \\ + \cdots + p_{n-1,n}(x_n - x_{n-1}))]^{1/2} , \end{aligned} \tag{34}$$

which immediately implies the inequality

$$\tilde{b}_n \leqslant v_n + \frac{p_{1n}(x_n - x_n) + \cdots + p_{n-1,n}(x_n - x_{n-1})}{m_n(v_n - \tilde{b}_{n-1})} . \tag{35}$$

By induction hypothesis, we can find A_1 and B_1 such that for

$$x_2 - x_1,...,x_{n-1} - x_{n-2} > A_1 , \tag{36}$$

and

$$v_2 - v_1,\ldots,v_{n-1} - v_{n-2} > B_1\,, \tag{37}$$

we have

$$\tilde{b}_1 < v_1 + \epsilon,\ldots,\tilde{b}_{n-1} < v_{n-1} + \epsilon\,. \tag{38}$$

By (35)

$$\tilde{b}_n \leqslant v_n + \frac{p_{1n}(x_n - x_{n-1}) + \cdots + p_{n-1,n}(x_n - x_{n-1})}{m_n(v_n - v_{n-1} - \epsilon)}\,. \tag{39}$$

Find A_2 such that

$$\frac{p_{1n}(A_2) + \cdots + p_{n-1,n}(A_2)}{m_n B_1} < \epsilon\,, \tag{40}$$

and set $B' = B_1 + \epsilon$, $A' = \max(A_1, A_2)$. Then for $(x,v) \in Y_{A',B'}$ we have

$$\tilde{b}_1 \leqslant v_1 + \epsilon, \cdots, \tilde{b}_n \leqslant v_n + \epsilon\,. \tag{41}$$

Repeating this argument with the other "fake motion" where the first particle does not push back, we obtain the "fake asymptotic velocities" $\bar{b}_1,\ldots,\bar{b}_n$ such that

$$\bar{b}_1 \leqslant b_1,\ldots,\bar{b}_n \leqslant b_n\,, \tag{42}$$

and the numbers A'', B'' such that for $(x,v) \in Y_{A'',B''}$ we have

$$v_1 - \epsilon < \bar{b}_1,\ldots,v_n - \epsilon < \bar{b}_n\,. \tag{43}$$

Setting $A = \max(A', A'')$, $B = \max(B', B'')$ and putting (31), (41), (42), and (43) together we obtain the assertion of the proposition.

Corollary 1 (of the proof): We keep notation and the assumptions of Proposition 1. Then for any $\epsilon > 0$ and $\delta > 0$ one can find A such that the asymptotic velocities b_i satisfy $|b_i - v_i| < \epsilon$ if $(x,v) \in Y_{A,\delta}$.

Proof: In view of (21) and (22), the assertion is obvious if the number of particles is equal to 2. Assume, by induction, that the assertion holds up to $n-1$. In particular, it holds for $n-1$ particles with $\epsilon_1 = \min(\epsilon, \delta/2)$ and $\delta_1 = \delta$. This means that we can find A_1 such that $\tilde{b}_1 < v_1 + \delta/2,\ldots,\tilde{b}_{n-1} < v_{n-1} + \delta/2$ if $x_2 - x_1,\ldots,x_{n-1} - x_{n-2} > A_1$ and $v_2 - v_1,\ldots,v_{n-1} - v_{n-2} > \delta$. Then $v_n - \tilde{b}_{n-1} > v_n - v_{n-1} - \delta/2 > \delta/2$ and, by (35),

$$\tilde{b}_n - v_n < \frac{p_{1n}(x_n - x_{n-1}) + \cdots + p_{n-1,n}(x_n - x_{n-1})}{m_n \delta/2}\,. \tag{44}$$

We can find $A' \geqslant A_1$ such that $2(p_{1n}(A') + \cdots + p_{n-1,n}(A'))/\delta m_n < \epsilon$ insuring that the inequalities $x_{i+1} - x_i > A'$ and $v_{i+1} - v_i > \delta$ for all i imply $\tilde{b}_1 < v_1 + \epsilon,\ldots,\tilde{b}_n < v_n + \epsilon$. Repeating the argument for the other set of "fake asymptotic velocities" $\bar{b}_i$ we obtain A'' and the inequalities $\bar{b}_1 > v_1 - \epsilon,\ldots,\bar{b}_n > v_n - \epsilon$. In view of (32) and (42), it suffices to set $A = \max(A', A'')$.

Denote by G^t the group of time translations naturally acting on Y. For $(x,v) \in Y$ we simply have $G^t(x,v) = (x(t), v(t))$. By basic theorems on the continuous dependence of solutions on initial data (see, e.g., Ref. 7), the transformations G^t are homeomorphisms of Y. We call a set X in Y invariant if $G^t(X) = X$ for all t. Any subset X of Y defined by asymptotic properties of trajectories $\{x(t)\}$ starting in X is obviously invariant. Now we are able to prove Theorem 2.

Proof of Theorem 2: Recall that, by our original notation, the asymptotic velocities at $t \to +\infty$ are denoted by $v_1(\infty),\ldots,v_n(\infty)$ and let Y_+ be the subset of the phase space Y defined by the conditions $v_1(\infty) < \cdots < v_n(\infty)$. We will show that the functions $v_i(\infty)$ are continuous on Y_+. For any $(x^0,v^0) \in Y_+$ there exists $d > 0$ and $T_1 > 0$ such that for $t > T_1$ we have $v^0_{i+1}(t) > v^0_i(t) + d$ for $i = 1,\ldots,n-1$. It suffices to find for any $\epsilon > 0$ a neighborhood D_ϵ of (x^0,v^0) such that for any (x,v) in D_ϵ the corresponding asymptotic velocities $v_i(\infty)$ satisfy $|v_i(\infty) - v_i| < \epsilon$. We can assume that $\epsilon < d$. Consider the open set Y_{d,T_1} in Y defined by the conditions $|v_i - v^0_i(T_1)| < \epsilon/e$ for all i. Then Y_{d,T_1} contains $G^{T_1}(x^0,v^0)$ and for any (x,v) in Y_{d,T_1} we have $v_{i+1} - v_i > d/3$ for all i. By Corollary 1, there is $A > 0$ such that the inequalities $v_{i+1} - v_i > d/3$ and $x_{i+1} - x_i > A$ for all i imply that $|v_i(\infty) - v_i| < \epsilon$. Obviously, we can find $T > T_1$ such that $x^0_{i+1}(t) - x^0_i(t) > A + 2\epsilon/3$ for $t \geqslant T$. Consider the neighborhood $D_{\epsilon,T}$ of $(x^0(T), v^0(T))$ defined by inequalities $|x_i - x^0_i(T)| < \epsilon/2$, $|v_i - v^0_i(T)| < \epsilon/3$ for $i = 1,\ldots,n$. Then for any (x,v) in $D_{\epsilon,T}$ we have $x_{i+1} - x_i > A$, $v_{i+1} - v_i > d/3$ for all i. Therefore for the vector of asymptotic velocities $(v_1(\infty),\ldots,v_n(\infty))$ corresponding to (x,v) from $D_{\epsilon,T}$ we have $|v_i(\infty) - v_i| < \epsilon$ for $i = 1,\ldots,n$. Thus for $(x,v) \in D_{\epsilon,T}$

$$|v_i(\infty) - v^0_i(\infty)| < \epsilon + \epsilon/3 + \epsilon = 7\epsilon/3\,. \tag{45}$$

Consider $D_\epsilon = G^{-T} D_{\epsilon,T}$. Then D_ϵ is an open neighborhood of (x^0,v^0) and for any $(x,v) \in D_\epsilon$ the inequalities (45) hold. Since ϵ is arbitrary, we conclude that the mapping $(x,v) \to v_+$ is continuous at $(x^0,v^0) \in Y_+$. Since (x^0,v^0) is arbitrary, we proved the continuity of W_+ on Y_+. If ϵ is small enough, D_ϵ belongs to Y_+, thus Y_+ is open. It is nonempty, by Proposition 1, and invariant, by definition. The theorem is proved.

Now we treat the case of purely external potential.

Theorem 3: Consider the n-body Hamiltonian

$$H = \frac{1}{2}\sum_{i=1}^{n} m_i v_i^2 + W(x_1,\ldots,x_n)\,, \tag{46}$$

with potential W satisfying the assumptions of Sec. II. Then the asymptotic velocities $(v_1(\infty),\ldots,v_n(\infty))$ exist for any point (x,v) in the phase space Y and we denote by Y_+ the subset of Y given by inequalities $v_i(\infty) > 0$ for all i.

Then Y_+ in an invariant open (nonempty) set and the mapping $(x,v) \to v_+$ is continuous on Y_+.

Recall that, by our assumptions, $W \geqslant 0$ and there is $L \geqslant -\infty$ such that $W(x_1,\ldots,x_n) \to \infty$ if for at least one i, $x_i \to L$ from the right. For simplicity of exposition we assume that $L = -\infty$. Besides, $W(x_1,\ldots,x_n) \to 0$ when all $x_i \to \infty$. For any A and B we denote by $Y_{A,B}$ the subset of Y given by inequalities $x_i > A$, $v_i > B$ for all i. We need the following.

Lemma 2: For any $\epsilon > 0$ we can find A such that on $Y_{A,0}$ the inequalities $|v_i(\infty) - v_i| < \epsilon$ are satisfied for all i.

Proof: Let $\delta > 0$ be arbitrary and choose A such that $W(x_1,\ldots,x_n) < \delta$ if $x_i > A$ for all i. Let $(x,v) \in Y_{A,0}$ and let $(x(t),v(t))$, $t \geqslant 0$, be the corresponding trajectory. The functions $v_i(t)$ monotonically increase approaching their limits $v_i(\infty)$. Therefore we have $v_i(t) \geqslant v_i > 0$ for all i and, by conservation of energy, for any $t \geqslant 0$,

$$\frac{1}{2}\sum_{i=1}^{n} m_i v_i^2 + W(x_1,...,x_n)$$

$$= \frac{1}{2}\sum_{i=1}^{n} m_i v_i(t)^2 + W(x_1(t),...,x_n(t)) . \quad (47)$$

Denote $v_i(\infty)$ by u_i and take the limit $t\to\infty$ in (47). We obtain

$$\frac{1}{2}\sum_{i=1}^{n} m_i u_i^2 - \frac{1}{2}\sum_{i=1}^{n} m_i v_i^2 = W(x_1,...,x_n) < \delta . \quad (48)$$

For any $0\leqslant v\leqslant u$ we have $(u-v)^2 = (u-v)(u-v) \leqslant (u-v)(u+v) = u^2 - v^2$, which implies

$$u - v \leqslant (u^2 - v^2)^{1/2} . \quad (49)$$

Applying (49) to (48) we get for any i

$$u_i - v_i \leqslant [m_i(u_i^2 - v_i^2)]^{1/2}(m_i)^{-1/2}$$

$$\leqslant \left[\sum_{j=1}^{n} m_j u_j^2 - \sum_{j=1}^{n} m_j v_j^2\right](m_i)^{-1/2} \leqslant (2\delta/m_i)^{1/2} . \quad (50)$$

Thus to satisfy the inequalities $|v_i(\infty) - v_i| < \epsilon$ it suffices to take δ such that for any i, $(2\delta/m_i)^{1/2} < \epsilon$, i.e., $\delta < \frac{1}{2}\epsilon^2 \min(m_1,...,m_n)$ and find A such that $W(x_1,...,x_n) < \delta$ if $x_1,...,x_n > A$.

Proof of Theorem 3: Let $(x^0,v^0)\in Y_+$ and let $\epsilon > 0$ be arbitrary. It suffices to find a neighborhood D_ϵ of (x^0,v^0) such that $D_\epsilon \subset Y_+$ and for any (x,v) in D_ϵ we have $|v_i(\infty) - v_i^0(\infty)| < \epsilon$ for all i. Let $d > 0$ be such that $v_i^0(\infty) > d$ for all i and choose $T_1 > 0$ such that $v_i^0(t) > d/2$ for $t > T_1$. Since $x_i^0(t) \to \infty$ for all i when $t\to\infty$ we can choose $T_2 > T_1$ such that for $t \geqslant T_2$ we have $x_i^0(t) > 2A$ where $Y_{A,0}$ is the set defined in Lemma 2. Define the set D_{ϵ,T_2} by inequalities $|x_i - x_i^0(T_2)| < A$, $|v_i - v_i^0(T_2)| < \epsilon$ for all i. Then D_{ϵ,T_2} is an open neighborhood of $G^{T_2}(v^0,x^0)$ which is contained in $Y_{A,0}$ if ϵ is less than $d/2$ which can assume without loss of generality. For any (x,v) from D_{ϵ,T_2} we have for all i

$$|v_i(\infty) - v_i^0(\infty)| = |(v_i(\infty) - v_i) + (v_i - v_i^0(T_2)) + (v_i^0(T_2) - v_i^0(\infty))| < 3\epsilon . \quad (51)$$

Then $D_\epsilon = G^{-T_2} D_{\epsilon,T_2}$ is an open neighborhood of (x^0,v^0) (see the proof of Theorem 2) and for any $(x,v)\in D_\epsilon$ we have, by (51)

$$|v_i(\infty) - v_i^0(\infty)| < 3\epsilon, \quad i = 1,...,n .$$

Since ϵ is arbitrarily small, this proves the theorem.

Now we can treat the general many-body directed Hamiltonians.

Theorem 4: Consider the n-body problem on the line with the Hamiltonian (2) where the pair potentials p_{ij} and the external potential W satisfy the assumptions of Sec. II. For a point (x,v) in the phase space Y we denote by $v(\infty)$ the vector of asymptotic velocities at $t = +\infty$ and let $Y_+ \subset Y$ be given by

$$0 < v_1(\infty) < v_2(\infty) < \cdots < v_n(\infty) .$$

Then Y_+ is a nonempty open invariant set and the mapping $(x,v)\to v(\infty)$ is continuous on Y_+.

Proof: By remarks in Sec. II, the potential V of the n-body problem is a cone potential and the force $F(x) = -\nabla V(x)$ is directed with respect to the cone C spanned by vectors $e_j - e_i$ for $i<j$ and e_k, $k = 1,...,n$. The dual cone C^* consists of vectors $\Sigma_{i=1}^n b_i e_i$ such that

$$0 \leqslant b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n . \quad (52)$$

By results of Refs. 3 and 6, the mapping $(x,v)\to v(\infty)$ is defined everywhere and $v(\infty)$ belongs to C^* for any (x,v). The set Y_+ is defined by the condition $v(\infty)\in \mathrm{Int}\, C^*$. For any A and B we denote by $Y_{A,B}$ the subset of Y given by the inequalities

$$\begin{aligned} &x_1 > A, \quad x_2 - x_1 > A, ..., x_n - x_{n-1} > A ; \\ &v_1 > B, \quad v_2 - v_1 > B, ..., v_n - v_{n-1} > B . \end{aligned} \quad (53)$$

The following lemma is crucial.

Lemma 3: For any $\epsilon > 0$, $\delta > 0$ there exists A such that $(x,v)\in Y_{A,\delta}$ implies $|v_i - v_i(\infty)| < \epsilon$ for all i.

The proof is a combination of proofs of Corollary 1 and Lemma 2 and we leave it to the reader. Assertion of the theorem follows from the lemma the same way Theorem 2 follows from Corollary 1 or Theorem 3 follows from Lemma 2. We spare the details.

IV. CONTINUITY OF ASYMPTOTIC PHASES FOR CONE POTENTIALS

Recall that a potential V on $\mathbb{R}^n$ (which is allowed to take value $+\infty$) is called a cone potential with respect to a closed proper cone C if V is differentiable on the open set X in $\mathbb{R}^n$ where $V(x) < \infty$ and the force $F(x) = -\nabla V(x)$ is contained in C for any x in X. In what follows we assume that V is continuously differentiable on X and that Assumptions 1 and 2 of Ref. 6 are satisfied. That is, $V \geqslant 0$,

$$\inf_X V(x) = 0 ,$$

and that

$$V(x) \to \infty \text{ if } \langle x,c\rangle \to -\infty$$

for at least one c from C and that $V(x)\to 0$ if $\langle x,c\rangle \to \infty$ for all $c\in \mathrm{Int}\, C$.

Let $Y = X\times\mathbb{R}^n$ be the phase space and recall that a trajectory $x(t)$ defined by $(x,v)\in Y$ is called asymptotically uniform (at $t = +\infty$) if

$$x(t) = a + bt + o(1) , \quad (54)$$

as $t\to\infty$. Here a and b are the asymptotic phase and velocity, respectively. By Refs. 3 and 6, the mapping $(x,v)\to b$ is defined everywhere, b belongs to the dual cone C^*, and we denote by Y_+ the subset of Y given by the condition $b\in \mathrm{Int}\, C^*$.

We call Y_+ the set of regular points in Y.

Theorem 5: (See Ref. 6.) Let the potential V satisfy

$$\int_0^\infty V(ct)dt < \infty , \quad (55)$$

for any $c\in \mathrm{Int}\, C^*$. Then the mapping W_+: $(x,v)\to(a,b)$ is defined on Y_+.

At this point we have to digress a little bit. Let $f(y,t)$ be a continuous vector function on $Y_0\times[0,\infty)$ and let the limit $f(y) = \lim_{t\to\infty} f(y,t)$ exist for any y. We say that the convergence $f(y,t)\to f(y)$ is locally uniform on Y_0 if for every y_0 from Y_0 and any $\epsilon > 0$ there exists a neighborhood D of y_0 in Y_0 (Y_0 is assumed to be open) and $T > 0$ such that

$\|f(y,t) - f(y)\| < \epsilon$ for any $y \in D$ and $t > T$. By a standard theorem of analysis (see, e.g., Ref. 8), the limit function $f(y)$ is continuous if the convergence

$$f(y,t) \underset{t\to\infty}{\to} f(y)$$

is locally uniform.

The following theorem will be applied to the many-body problems with directed interactions.

Theorem 6: Let the assumptions be as above and let Y_0 be an open invariant subset of Y_+ such that the convergence $v(t) = v(y,t) \to v(\infty)$, as $t \to \infty$, is locally uniform. Then W_+: $(x,v) \to (a,b)$ is continuous on Y_0.

Proof: By Theorem 5, the mapping W_+: $(x,v) \to (a,b) = (a, v(\infty))$ is well defined on Y_0 and, by discussion preceeding the theorem, b depends continuously on (x,v). A straightforward computation (cf. Ref. 6) gives

$$a = x - \int\int_{\sigma>\tau>0} d\sigma\, d\tau\, F(x(\sigma)), \tag{56}$$

where the improper integral in (56) is the limit of

$$\sigma > \tau, \quad \int\int_{t>\tau>0} d\sigma\, d\tau\, F(x(\sigma)) \text{ as } t \to \infty .$$

Since $F(x(\sigma))$ belongs to the proper cone C, convergence of the integral in (56) is equivalent to its absolute convergence which is equivalent to the convergence of integrals $\int\int_{\sigma>\tau>0} d\sigma\, d\tau \langle c^*, F(x(\sigma))\rangle$ for all $c^* \in C^*$. We want to show the continuity of the mapping

$$(x,v) \to \int\int_{\sigma>\tau>0} d\sigma\, d\tau\, F(x(\sigma))$$

on Y_0. Because of the absolute convergence we can change the order of integration in (56) and obtain

$$\int\int_{\sigma>\tau>0} d\sigma\, d\tau\, F(x(\sigma)) = \int_0^\infty d\sigma\, F(x(\sigma)) \int_0^\sigma d\tau = \int_0^\infty \sigma F(x(\sigma))\, d\sigma . \tag{57}$$

The improper integral $\int_0^\infty \sigma F(x(\sigma))\, d\sigma$ is the limit function $f(y) = \lim f(y,t)$ where $f(y,t) = \int_0^t \sigma F(x(\sigma))\, d\sigma$ and $y = (x,v)$. In view of the discussion preceeding the theorem, it should be clear what we mean by saying that the mapping $y \to \int_0^\infty \sigma F(x(\sigma))\, d\sigma$ is continuous if the integral $\int_0^\infty \sigma F(x(\sigma))\, d\sigma$ converges locally uniformly on Y_0.

We claim that $\int_0^\infty \sigma\, F(x(\sigma))\, d\sigma$ converges locally uniformly if and only if the scalar integral $\int_0^\infty \sigma \langle F(x(\sigma)), v(\sigma)\rangle\, d\sigma$ does. Indeed, by assumption, the vector function $(y,\sigma) \to v(\sigma)$ converges to $y \to v(\infty) = b$ locally uniformly on Y_0, which proves the implication from the locally uniform convergence of $\int_0^\infty \sigma F(x(\sigma))\, d\sigma$ to that of $\int_0^\infty \sigma\, \langle F(x(\sigma)), v(\sigma)\rangle\, d\sigma$. To do the opposite implication, choose a basis $c_1,\ldots,c_n \in C$ of $\mathbb{R}^n$ and let $c_1^*,\ldots,c_n^* \in C^*$ be the dual basis. Then for any $\sigma \geqslant 0$

$$\sigma\langle F(x(\sigma)), v(\sigma)\rangle = \sum_i \sigma \langle F(x(\sigma)), c_i^*\rangle\, \langle c_i, v(\sigma)\rangle . \tag{58}$$

Let $y_0 \in Y_0$ and let $b_0 = v_0(\infty)$ be the corresponding asymptotic velocity. Since $b_0 \in \text{Int } C^*$ and since the convergence $v(t) \to v(\infty)$ is locally uniform, we can choose a neighborhood D of y_0 in Y_0 and $T > 0$ such that for any $y \in D$ and $t \geqslant T$ we have $v(t) \in \text{Int } C^*$. Since the continuity of a at y_0 and $G^T y_0$ are equivalent, we can assume without loss of generality that $T = 0$. Then all the factors in the right-hand side of (58) are non-negative, which implies for any $i = 1,\ldots,n$,

$$\sigma\langle F(x(\sigma)), c_i^*\rangle\, \langle \sigma\langle F(x(\sigma)), v(\sigma)\rangle\, \langle c_i, v(\sigma)\rangle^{-1} . \tag{59}$$

Using the locally uniform convergence of $v(\sigma)$ to $v(\infty)$ again, we can choose the neighborhood D of y_0 small enough so that for $i = 1,\ldots,n$ and $\sigma \geqslant 0$ we have $\langle c_i, v(\sigma)\rangle > d > 0$. Then the functions $(y,\sigma) \to \langle c_i, v(\sigma)\rangle^{-1}$ converge to $\langle c_i, v(\infty)\rangle^{-1}$ uniformly on D which implies the locally uniform convergence of the integral $\int_0^\infty \sigma\langle F(x(\sigma)), v(\sigma)\rangle \langle c_i, v(\sigma)\rangle^{-1}\, d\sigma$ and, by (59), the same holds for $\int_0^\infty \sigma\langle F(x(\sigma)), c_i^*\rangle d\sigma$. Since c_i^* form a basis of $\mathbb{R}^n$ we conclude that $\int_0^\infty \sigma F(x(\sigma)) d\sigma$ converges locally uniformly.

We have

$$\begin{aligned}\int_0^\infty \sigma\langle F(x(\sigma)), v(\sigma)\rangle d\sigma &= \int\int_{\sigma>\tau>0} \langle F(x(\sigma)), v(\sigma)\rangle d\sigma\, d\tau \\ &= \int_0^\infty d\tau \int_\tau^\infty \langle F(x(\sigma)), v(\sigma)\rangle d\sigma \\ &= -\int_0^\infty d\tau \int_{x(\tau)}^\infty \nabla V(\xi) = \int_0^\infty d\tau\, V(x(\tau)) .\end{aligned} \tag{60}$$

By (60) and the previous argument, the integral in (56) converges locally uniformly on Y_0 if and only if the integral $\int_0^\infty d\tau\, V(x(\tau))$ does.

It remains to show that the convergence of $\int_0^\infty V(c^* t) dt$ for any c^* from Int C^* implies the local uniform convergence of $\int_0^\infty d\tau\, V(x(\tau))$ on Y_0. Let $y_0 = (x_0, v_0) \in Y_0$ and let $b_0 \in \text{Int } C^*$ be the corresponding asymptotic velocity. For any $\epsilon > 0$ we can find a neighborhood D of y_0 and $T \geqslant 0$ such that for $y \in D$ and $t \geqslant T$, $\|v(t) - b_0\| < \epsilon$. Since the uniform convergence of $\int_0^\infty V(x(\tau)) d\tau$ and $\int_T^\infty V(x(t)) dt$ are equivalent, we can assume without loss of generality that $\|v(t) - b_0\| < \epsilon$ for $t \geqslant 0$. Since b_0 belongs to the open cone Int C^*, for ϵ small enough we can find $b_1 \epsilon$ Int C^* such that $\|b - b_0\| < \epsilon$ implies $b - b_0 \in C^*$. Thus for any $y = (x,v)$ from our neighborhood D of y_0 we have $x(t) - x - b_1 t \in C^*$ for $t \geqslant 0$. Therefore (cf. Ref. 6), $V(x(t)) \leqslant V(x + b_1 t)$ for all $t \geqslant 0$ and for any $T \geqslant 0$

$$\int_T^\infty V(x(t)) dt \leqslant \int_T^\infty V(x + b_1 t) dt . \tag{61}$$

For any two vectors x, y in $\mathbb{R}^n$ we write $x \geqslant y$ if $x - y \in C^*$. By Ref. 6, $x \geqslant y$ implies $V(y) \geqslant V(x)$. Since the neighborhood $D \subset \mathbb{R}^n$ of x_0 is bounded, there exists $x_1 \in \mathbb{R}^n$ such that $x \geqslant x_1$ for any x in $D \subset \mathbb{R}^n$. Therefore for $t \geqslant 0$ and any $y = (x,v)$ in D we have

$$x(t) \geqslant x + b_1 t \geqslant x_1 + b_1 t .$$

Since $b_1 \in \text{Int } C^*$, there exists $t_1 \geqslant 0$ such that $x_1 + b_1 t_1 \in C^*$. Thus for $t \geqslant t_1$

$$x(t) \geqslant (x_1 + b_1 t_1) + b_1 (t - t_1) \geqslant b_1 (t - t_1) . \tag{62}$$

In view of (61) and (62), for any $T \geqslant t_1$

$$\int_T^\infty V(x(t)) dt \leqslant \int_T^\infty V(b_1 (t - t_1)) dt = \int_{T - t_1}^\infty V(b_1 t) dt . \tag{63}$$

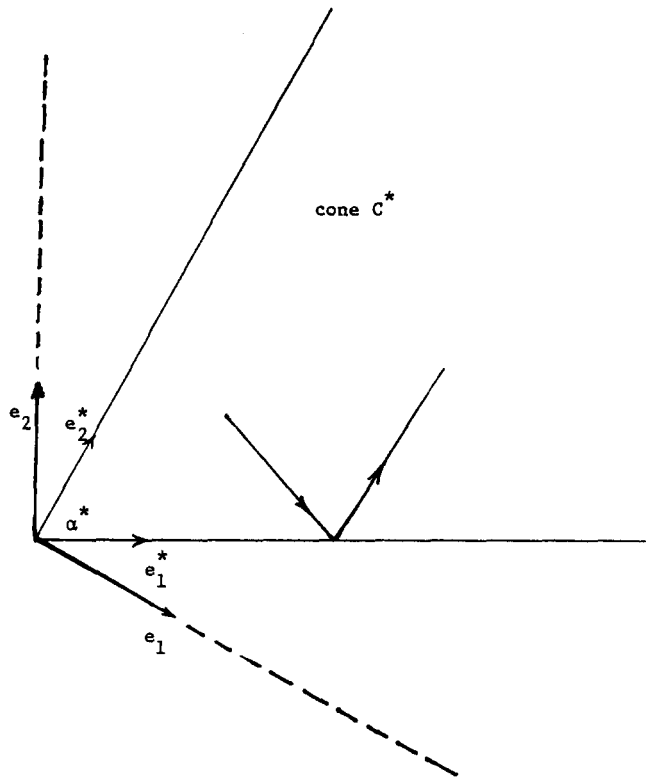


FIG. 1. Dual cones C^* and C.

Letting $T\to\infty$, we obtain, by (63), that the integral $\int_0^\infty V(x(t))dt$ converges uniformly on D, which proves the theorem.

Now we prove the main result of this paper.

Theorem 7: Let

$$H=\frac{1}{2}\sum_{i=1}^{n} m_i v_i^2+\sum_{i<j} p_{ij}(x_j x_i)+W(x_1,\ldots,x_n)$$

be the Hamiltonian of a n-body problem with directed interaction where the pair potentials p_{ij} and the external potential W satisfy the assumptions of Sec. II. By previous results, the vector b of asymptotic velocity (at $+\infty$) exists for any initial data y and satisfies $0<b_1\leqslant\cdots\leqslant b_n$ if $W\neq 0$ and $b_1\leqslant\cdots\leqslant b_n$ if $W=0$. Denote by Y_+ the set in Y where the strict inequalities hold. Then Y_+ is a nonempty open invariant set and the mapping W_+: $y\to(a,b)$ is continuous on Y_+.

Proof: Proving the results of Sec. III we have shown that for $y=(x,v)$ in Y_+ the vector function $v(t)$ converges to $v(\infty)=b$ locally uniformly (see Proposition 1, Lemma 2, and Lemma 3). The total potential V is a cone potential and we have shown in Sec. III that Y_+ is open and nonempty. Applying Theorem 6 with $Y_0=Y_+$ we obtain the assertion.

V. EXAMPLES AND DISCUSSION

It is not hard to give concrete examples of potentials satisfying assumptions of Sec. II.

A. Example 1: Exponential potentials

Let $n\geqslant 2$ be the number of particles and let the masses $m_1,\ldots,m_n$ be arbitrary. Let for any $1\leqslant i<j\leqslant n$ the numbers $a_{ij}>0$ and $c_{ij}\geqslant 0$ be given. Assume also that we have numbers $d_k\geqslant 0$, $1\geqslant k\leqslant N$, and $b_{ik}\geqslant 0$ for $1\leqslant i\leqslant n$, $1\leqslant k\leqslant N$, where $N\geqslant 1$ is some integer. Then the pair potentials $p_{ij}(x)=c_{ij}e^{-a_{ij}x}$ and the external potential

$$W(x_1,\ldots,x_n)=\sum_{k=1}^{N} d_k\exp\Big(-\sum_{i=1}^{n} b_{ik}x_i\Big)$$

satisfy the assumptions of Sec. II. Thus, Theorem 7 holds for the n-body problem on the line with the Hamiltonian

$$H=\frac{1}{2}\sum_{i=1}^{n} m_i v_i^2+\sum_{i<j} c_{ij}e^{-a_{ij}(x_j-x_i)}\sum_{k=1}^{N} d_k \times\exp\Big(-\sum_{i=1}^{n} b_{ik}x_i\Big). \tag{64}$$

When $m_1=\cdots=m_n=1$ and the only nonzero parameters in the potential are $c_{i,i+1}=a_{i,i+1}=1$, $i=1,\ldots,n-1$, we get the classical Toda lattice.

It is also easy to give examples of potentials for which the conclusions (and therefore the assumptions) of Theorem 7 are violated. We will briefly discuss a simple example of this type.

B. Example 2: Billiard in a wedge

Consider a closed cone C^* with angle $\alpha^*>0$ on the plane. Choose coordinates x,y so that the vertex of C^* is the origin and one of the sides of C^* is the positive x axis (see Fig. 1). The cone C^* is proper if $\alpha^*<\pi$ and we set $V=0$ inside C^* and $V=\infty$ outside C^*. Then V is a (degenerate) cone potential and the mechanical system with two degrees of freedom corresponding to the Hamiltonian

$$H=\tfrac{1}{2}(u^2+v^2)+V(x,y) \tag{65}$$

is the billiard inside C^*. Let e_1^*, e_2^* be the unit vectors spanning the walls of C^* (Fig. 1). Then the force F of the potential V is nontrivial only on the walls of C^* where it is proportional to e_1 and e_2, respectively (Fig. 1). Thus, F spans the dual cone C which has angle $\pi-\alpha$.

The billiard ball moving inside C^* with velocity (u,v) hits the walls of C^* a few times where it bounces off by the usual law of reflection until its velocity becomes a vector in C^*. Then it stops hitting the walls and goes away to infinity inside C^*.

Denote by s_i the reflections in the walls of C^* (s_i are linear orthogonal transformations of $\mathbb{R}^2$). Using the method of reflecting the billiard table instead of reflecting the ball (see, e.g., Ref. 9) we come to the following.

Proposition 2: Let $x\in C^*$ and $v\in\mathbb{R}^2$ be the initial data. Assume that the ray $x+tv$, $t\geqslant 0$ does not hit the vertex of C^*. Then there is a sequence $s_1,\ldots,s_k$ of reflections that depends only on the ray $\{x+tv,\ t\geqslant 0\}$ and $T\geqslant 0$ such that for $t\geqslant T$, $x(t)=s_k\cdots s_1(x+tv)$. The sequence $s_1,\ldots,s_k$ is locally constant when (x,v) varies and it is not defined if $x+tv$ hits the vertex of C^*.

The proof of this proposition is straightforward and is left to the reader. Figure 2 illustrates it. Now we deduce from it the mapping W_+.

Corollary 2: (i) Let $\alpha\neq\pi/n$. The mapping W_+: $(x,v)\to(a,b)$ is defined on all initial data $y=(x,v)$, $x\in C^*$ such that $x\neq -tv$ for $t\geqslant 0$. Denote by $Y_+\subset Y$ the dense open set of such initial data. For every y in Y_+ there is a sequence

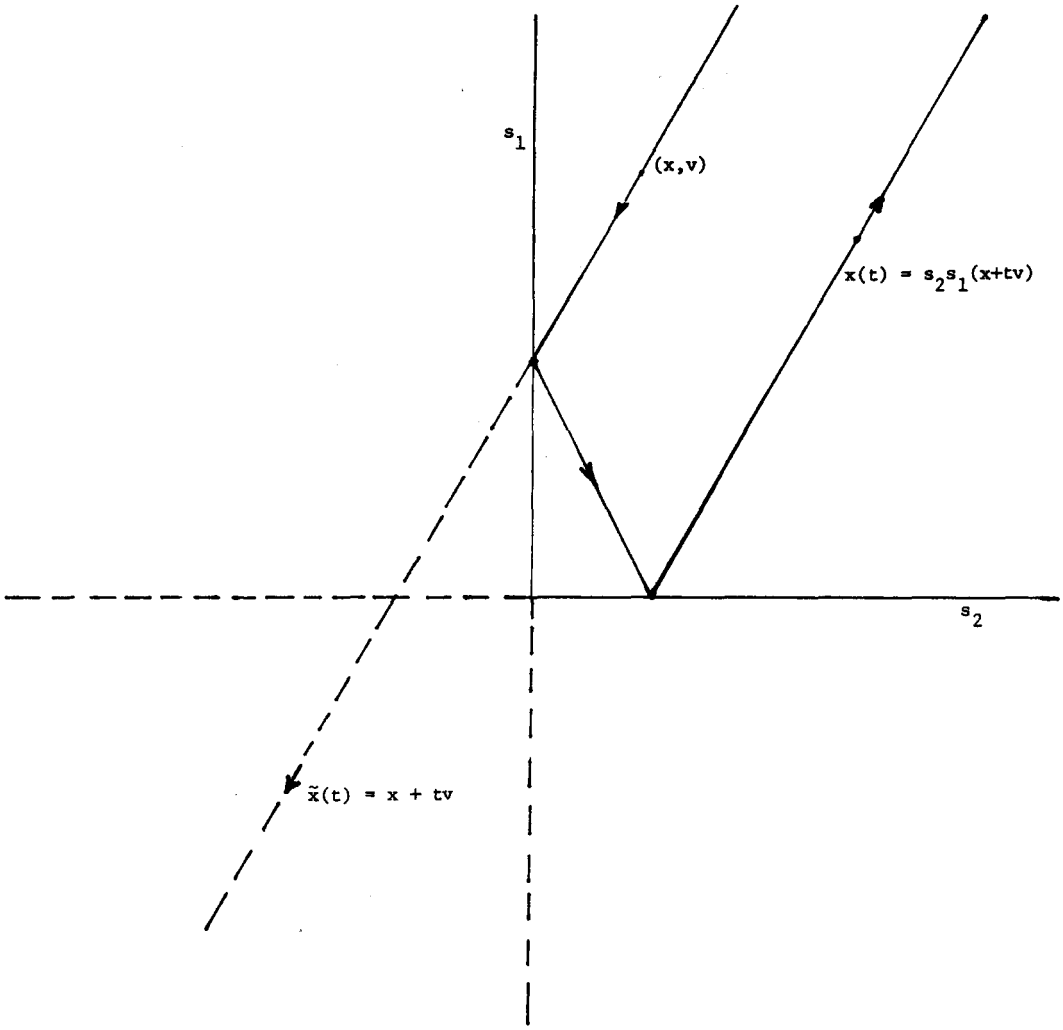


FIG. 2. Billiard trajectory in a wedge.

$s_1,\ldots,s_k$ of reflections about the walls of C^* and we denote $s_k \cdots s_1$ by $w_1(y)$. Then the mapping W_+ is given by $a = w_+(y)x$, $b = w_+(y)v$. The function $w_+(y)$ is locally constant on Y_+ and W_+ does not extend by continuity from Y_+ to Y.

(ii) Let $\alpha = \pi/n$ where $n \geqslant 2$ is an integer. Then all the assertions of (i) remain true except the last one. The mapping W_+ uniquely extends by continuity to all of Y.

The proof consists in a careful consideration of possible reflection patterns (see Fig. 2) and we leave it to the reader.

ACKNOWLEDGMENT

This paper was partially supported by NSF Grant No. DMS 84-03238.

[1]G. Galperin, "Asymptotic behavior of particle motion under repulsive forces," Commun. Math. Phys. **84**, 547 (1982).
[2]L. N. Wasserstein, "On system of particles with finite range and/or repulsive interactions," Commun. Math. Phys. **69**, 31 (1979).
[3]E. Gutkin, "Integrable Hamiltonians with exponential potential," Physica D **16**, 398 (1985).
[4]E. Gutkin, "Asymptotics of trajectories for Newtonian dynamics with directed force," to appear in Physica D.
[5]J. Moser, "The scattering problem for some particle systems on the line," Lecture Notes Math. **597**, 441 (1978).
[6]E. Gutkin, "Asymptotics of trajectories for cone potentials," Physica D **17**, 235 (1985).
[7]E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[8]J. Marsden, *Elementary Classical Analysis* (Freeman, San Francisco, 1974).
[9]E. Gutkin, "Billiards in polygons," Physica D **19**, 311 (1986).

Direct and inverse scattering in the time domain for a dissipative wave equation. III. Scattering operators in the presence of a phase velocity mismatch

G. Kristensson
Division of Electromagnetic Theory, Royal Institute of Technology, S-100 44 Stockholm, Sweden
R. J. Krueger
Applied Mathematical Sciences, Ames Laboratory—USDOE, Iowa State University, Ames, Iowa 50011

(Received 25 March 1986; accepted for publication 27 August 1986)

The direct scattering problem for an inhomogeneous lossy medium is examined for the one-dimensional case in which the phase velocity profile is discontinuous at the boundaries of the medium. Scattering operators (or impulse responses) and propagation operators are defined and equations that govern their behavior are developed. Knowledge of the scattering kernels for one round trip in the medium implies that the scattering kernels can be determined on any time interval. Numerical examples are presented. It is also shown that this scattering problem is reducible to one in which there are no phase velocity mismatches. This reduction provides considerable numerical advantage in the solution of the direct scattering problem. The inverse problem is examined in a companion paper.

I. INTRODUCTION

This paper deals with transient wave propagation in one-dimensional dissipative media. Various aspects of this problem have been studied in two previous papers,[1,2] hereafter referred to as Parts I and II. The present paper is nevertheless fairly self-contained, although the first two sections of Part I should be consulted for an overview of the problem and a discussion of the relevant literature. Some additional discussion of the literature is contained in the introductory section of Part II.

The model problem being considered here involves one-dimensional electromagnetic wave propagation in a medium characterized by spatially varying permittivity and conductivity profiles. The distinction between this model and that considered in Parts I and II is that phase velocity mismatches at the boundaries of the scattering medium are now allowed. Thus, the results presented here are generalizations of those derived in Part I.

The goal of this paper is to derive equations for the scattering and propagation operators for the model problem described above and to examine certain properties which they exhibit. These operators, as well as a concise statement of the problem, are given in Sec. II. Integrodifferential equations for these operators are then derived in Sec. III. These equations permit one to numerically construct scattering and propagation operators, and this is done at the end of Sec. III for a particular set of examples. In Sec. IV it is shown that finite time traces of these operators can be extended in a straightforward fashion to time traces over longer periods of time. The basic idea here is that if the scattering operators are known for one round trip in the medium, then for longer times nothing new is occurring in terms of scattering phenomena.

Throughout this paper it is assumed that the operators under discussion have certain simplified forms. In Appendix A, a justification for this simplification is given. The ideas in Appendix A then lead one to (correctly) surmise that a second simplification of the scattering problem under consideration would reduce the present problem to that considered in Part I. In Appendix B this further simplification is carried out. While this simplification presents a distinct computational advantage for the direct scattering problem, it is interesting that it has the opposite effect for the inverse scattering problem. Thus, the primary purpose for the analysis in Secs. III and IV lies in its usefulness in the inverse problem, which is the subject of a companion paper,[3] hereafter referred to as Part IV.

II. STATEMENT OF THE PROBLEM

As mentioned in the Introduction, the problem considered here is a generalization of that considered in Parts I and II. Thus, the reader should consult Sec. II of Part I for a full explanation of the underlying ideas introduced here. An inhomogeneous slab occupies the region $0 \leqslant z \leqslant L$. The permittivity ϵ and conductivity σ of the slab are functions of depth z only. On either side of the slab there is a homogeneous, lossless medium. Unlike the problem considered in Parts I and II, it is not assumed that the permittivity is continuous across the interfaces at $z = 0$ and $z = L$.

In the region $z < 0$ an incident electromagnetic plane wave propagates along the z axis, impinging on the slab at time $t = 0$. This produces an electric field $E(z,t)$ satisfying

$$E_{zz}(z,t) - c^{-2}(z)E_{tt}(z,t) - b(z)E_t(z,t) = 0, \quad (2.1)$$

where

$$c^{-2}(z) = \epsilon(z)\mu_0, \qquad b(z) = \sigma(z)\mu_0, \quad (2.2)$$

and μ_0 is the permeability in vacuum. The phase velocity c is assumed to be continuously differentiable within the slab and the dissipation b is continuous within the slab.

Converting to travel time coordinates as in Eq. (2.13) of Part I yields the transformed problem

$$u_{xx} - u_{ss} + A(x)u_x + B(x)u_s = 0, \quad (2.3)$$

where

 0022-2488/87/020360-11$02.50 © 1987 American Institute of Physics

$$A(x) = -\frac{d}{dx}\ln c(z(x)), \tag{2.4}$$

$$B(x) = -lb(z(x))c^2(z(x)), \tag{2.5}$$

$$l = \int_0^L c^{-1}(z)dz,$$

and ln denotes the natural logarithm function. The functions A and B vanish outside of the interval $[0,1]$. However, the discontinuities in c at the boundaries of the slab imply that in the transformed problem (2.3), the spatial derivative of u at $x = 0, x = 1$ is discontinuous. This discontinuity is given by

$$c_0 u_x(0^-,s) = u_x(0^+,s), \tag{2.6}$$

$$c_1 u_x(1^+,s) = u_x(1^-,s), \tag{2.7}$$

where

$$c_0 = [\epsilon(0^-)/\epsilon(0^+)]^{1/2} = c(0^+)/c(0^-),$$

$$c_1 = [\epsilon(L^+)/\epsilon(L^-)]^{1/2} = c(L^-)/c(L^+).$$

It is shown in Appendix A that the effects of the discontinuity in u_x at $x = 0$, Eq. (2.6), can be removed from both the direct and inverse problem in a straightforward manner. Therefore, it is assumed in the remainder of this and the next paper that this discontinuity is not present. In other words, the quantity c_0 in Eq. (2.6) is assumed to be equal to 1; i.e., the phase velocity will be taken to be continuous at $z = 0$. It is also possible to remove the effects of the discontinuity in Eq. (2.7). However, for reasons discussed in Appendix B and in Part IV, the discontinuity at $x = 1$ is retained.

As in Parts I and II, the solution of Eq. (2.3) in the regions exterior to the slab reduces to right and left moving waves related to each other through scattering and propagation operators. These are given by

$$u^r_+(s) = [\mathscr{R}^+(0)u^i_+(\cdot)](s) = \rho(0)u^i_+(s-2) + \int_0^s R^+(0,s-s')u^i_+(s')ds', \quad s>0, \tag{2.8}$$

$$u^t_+(s) = [\mathscr{T}^+(0)u^i_+(\cdot)](s) = \tau(0)\left[u^i_+(s) + \int_0^s T(0,s-s')u^i_+(s')ds'\right], \quad s>0, \tag{2.9}$$

$$u^i_+(s) = [\mathscr{W}^+(0)u^t_+(\cdot)](s) = \tau(0)^{-1}\left[u^t_+(s) + \int_0^s W(0,s-s')u^t_+(s')ds'\right], \quad s>0, \tag{2.10}$$

$$u^r_+(s) = [\mathscr{V}^+(0)u^t_+(\cdot)](s) = \nu(0)u^t_+(s-2) + \tau(0)^{-1} \times \int_0^s V^+(0,s-s')u^t_+(s')ds', \quad s>0, \tag{2.11}$$

where

$$\rho(0) = r\exp\left[\int_0^1 B(x)dx\right], \quad \tau(0) = 2t^+(0,1)/(c_1+1),$$

$$\nu(0) = (1-c_1)t^-(0,1)/2, \quad r = (1-c_1)/(1+c_1),$$

$$t^\pm(0,1) = \exp\left\{\mp\tfrac{1}{2}\int_0^1 [A(x) \mp B(x)]dx\right\}.$$

The form of the operators given in Eqs. (2.8)–(2.11) can be ascertained in a number of ways; see, for example, Appendix A. Notice that when $c_1 = 1$ these operators reduce to those used in Parts I and II.

The direct scattering problem studied in this paper is that of determining the scattering operators $\mathscr{R}^+$ and $\mathscr{T}^+$ given that $A(x)$ and $B(x)$ [or equivalently, $\epsilon(z)$ and $\sigma(z)$] are known. This will be done by deriving integrodifferential equations for the scattering kernels R^+ and T in Eqs. (2.8) and (2.9) which relate these kernels to the permittivity and conductivity of the medium.

The inverse problem studied in Part IV is that of determining $\epsilon(z)$ and $\sigma(z)$ and the total depth of the medium, L, using scattering data from the slab. In terms of the transformed problem Eq. (2.3) this means that $A(x)$ and $B(x)$ are to be constructed from finite time traces of the scattering kernels R^+ and T given in Eqs. (2.8) and (2.9). In the derivation in Part IV it is shown that two different sets of scattering data can be used for this purpose. One set involves only reflection data, while the other utilizes a smaller set of reflection data in conjunction with transmission data.

The formulations of the problems given here are asymmetric in the sense that scattering data from only one incident field are considered. This is in contrast to the problems studied in Parts I and II, in which incident fields from both sides of the slab are used. It is the discontinuity in ϵ at the interface $z = L$ that motivates this distinction. The notation in Eqs. (2.8)–(2.11) also displays this asymmetry in that the arguments of the kernels are $(0,s)$ instead of $(0,1,s)$ as in Part I, Sec. II. This reflects the fact that the right edge of the medium will always be fixed at $x = 1$. Differential changes in the operators only with respect to changes in the left edge of the medium will be considered. Thus, scattering and propagator kernels for subregions $[x,1]$ of the original medium will be denoted by $R^+(x,s)$, $T(x,s)$, $W(x,s)$, and $V^+(x,s)$.

III. EQUATIONS FOR THE SCATTERING AND PROPAGATOR KERNELS

In this section, relations between the scattering properties of the medium (given by A and B) and the scattering and propagator kernels are derived. The first portion of this analysis is similar to that given in Part I, Appendix C, Eqs. (C1)–(C3). Thus, replace the independent variable x in Eq. (2.3) with the dummy variable z [not to be confused with the variable appearing in Eq. (2.1)], since x is used to denote the end point of a subregion $[x,1]$ of the slab $[0,1]$. Now introduce the change of basis from $(u,u_z)^T$ to $(u^+,u^-)^T$ via

$$u^\pm(z,s) = \tfrac{1}{2}[u(z,s) \mp \partial_s^{-1}u_z(z,s)], \tag{3.1}$$

where

$$\partial_s^{-1}u_z(z,s) = \int_{-\infty}^s u_z(z,s')ds'.$$

Notice that $u^\pm$ are right and left moving waves in a homogeneous medium. In this new basis, Eq. (2.3) becomes

$$\frac{\partial}{\partial z}\begin{pmatrix}u^+(z,s)\\ u^-(z,s)\end{pmatrix} = \begin{pmatrix}\alpha(z) & \beta(z)\\ \gamma(z) & \delta(z)\end{pmatrix}\begin{pmatrix}u^+(z,s)\\ u^-(z,s)\end{pmatrix}, \tag{3.2}$$

where

$$\alpha(z) = -\frac{1}{2}[A(z) - B(z)] - \frac{\partial}{\partial s},$$
$$\beta(z) = \tfrac{1}{2}[A(z) + B(z)],$$
$$\gamma(z) = \tfrac{1}{2}[A(z) - B(z)], \tag{3.3}$$
$$\delta(z) = -\frac{1}{2}[A(z) + B(z)] + \frac{\partial}{\partial s}.$$

Now consider a subregion $[x,1]$ of the original slab $[0,1]$. Incident and scattered fields for this subregion take the form

$$u^+(z,s) = \begin{cases} u^i_+(s - z + x), & z \leqslant x, \\ u^r_+(s - z + x), & z \geqslant 1, \end{cases} \tag{3.4}$$

$$u^-(z,s) = u^r_+(s + z - x), \quad z \leqslant x. \tag{3.5}$$

The scattering and propagation operators for the subregion $[x,1]$ are the generalizations of the operator relations in Eqs. (2.8)–(2.11) and are written as

$$u^-(x,s) = \rho(x)u^+(x,s - 2(1-x)) + \int_0^s R^+(x,s-s')u^+(x,s')ds', \quad s > 0, \tag{3.6}$$

$$u^+(1,s + 1 - x) = \tau(x)\left[u^+(x,s) + \int_0^s T(x,s-s')u^+(x,s')ds'\right], \quad s > 0, \tag{3.7}$$

$$u^+(x,s) = \tau^{-1}(x)\left[u^+(1,s + 1 - x) + \int_0^s W(x,s-s')u^+(1,s' + 1 - x)ds'\right], \quad s > 0, \tag{3.8}$$

$$u^-(x,s) = \nu(x)u^+(1,s - 1 + x) + \tau^{-1}(x) \times \int_0^s V^+(x,s-s')u^+(1,s' + 1 - x)ds', \quad s > 0, \tag{3.9}$$

where

$$\rho(x) = r\exp\left[\int_x^1 B(x')dx'\right],$$
$$\tau(x) = 2t^+(x,1)/(c_1 + 1),$$
$$\nu(x) = (1 - c_1)t^-(x,1)/2, \tag{3.10}$$
$$r = (1 - c_1)/(1 + c_1),$$
$$t^\pm(x,1) = \exp\left\{\mp\frac{1}{2}\int_x^1 [A(x') \mp B(x')]dx'\right\},$$

and

$$u^+(x,s) = 0, \quad s < 0.$$

The imbedding equation for the reflection kernel $R^+(x,s)$ is derived by differentiating Eq. (3.6) with respect to x and then using Eqs. (3.2) and (3.6) to obtain

$$\gamma(x)u^+(x,s) - \beta(x)\left[\rho(x)u^+(x,s - 2(x-1)) + \int_0^s R^+(x,s-s')u^+(x,s')ds'\right] + u_s^-(x,s)$$
$$= \rho'(x)u^+(x,s - 2(x-1)) + \rho(x) \times u_x^+(x,s - 2(x-1)) + 2\rho(x)u_s^+(x,s - 2(x-1)) + \frac{\partial}{\partial x}\int_0^s R^+(x,s-s')u^+(x,s')ds'. \tag{3.11}$$

The kernel $R^+(x,s)$ is a piecewise continuous function in the domain $0 \leqslant x \leqslant 1$, $s > 0$ and to proceed with the derivation the discontinuities of $R^+(x,s)$ have to be treated with some care. Thus, assume that there are two discontinuities present in $R^+(x,s)$ along the curves

$$s = d_i(x), \quad i = 1,2,$$
$$0 < d_1(x) < d_2(x), \quad 0 < x < 1,$$

in the (x,s) plane. More general assumptions involving an arbitrary number of discontinuities would proceed analogously but the analysis below shows that two discontinuities are sufficient.

Straightforward but lengthy calculations now show that Eq. (3.11) implies

$$R_x^+(x,s) = 2R_s^+(x,s) - B(x)R^+(x,s) - \tfrac{1}{2}[A(x) + B(x)] \times \int_0^s R^+(x,s-s')R^+(x,s')ds' - H(s - 2(1-x))\rho(x)[A(x) + B(x)] \times R^+(x,s - 2(1-x)), \quad s > 0, \quad s \neq 4(1-x), \tag{3.12}$$

$$R^+(x,0^+) = -\tfrac{1}{4}[A(x) - B(x)], \quad 0 < x < 1, \tag{3.13}$$

$$R^+(1,s) = 0, \quad s > 0,$$

where

$$H(s) = \text{Heaviside function} = \begin{cases} 0, & s < 0, \\ 1, & s > 0. \end{cases}$$

The same calculation shows that the discontinuities $d_i(x)$, $i = 1,2$, satisfy

$$d_1'(x) = -2, \tag{3.14}$$
$$d_2(x) = 4(1-x), \tag{3.15}$$

with a jump

$$[R^+(x,s)]_{s=4(1-x)^-}^{s=4(1-x)^+} = -\tfrac{1}{4}\rho^2(x)[A(x) + B(x)]. \tag{3.16}$$

The jump along the characteristic curve $d_1(x)$ has to be determined using standard propagation of singularities arguments.[4] The jump is

$$[R^+(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+} = \tfrac{1}{4}\exp\left[\int_x^1 B(x')dx'\right]\left\{r^2[A(1) + B(1)] + A(1) - B(1) - r\int_x^1 [A^2(x') - B^2(x')]dx'\right\}. \tag{3.17}$$

Equation (3.12) is the imbedding equation for the reflection kernel for the slab with a right edge discontinuity in ϵ. Unlike the continuous case, there is a second jump in $R^+(x,s)$ along $s = 4(1-x)$. The imbedding equation now also contains an additional term due to the presence of a hard

back wall. Notice that for times less than one round trip, i.e., $0<s<2(1-x)$, the imbedding equation for $R^+(x,s)$ and the corresponding one in Eq. (3.1) of Part I are identical. Domain of dependence arguments show that the two kernels have to be identical for $0<s<2(1-x)$. In the absence of the right edge discontinuity (i.e., $c_1=1$), Eqs. (3.12)–(3.17) agree with the previous results in Part I for the continuous profile.

The imbedding equation for the transmission kernel is derived in an analogous way. The starting point is the transmission operator for the subregion $[x,1]$ given by Eq. (3.7). Differentiation with respect to x and use of Eqs. (3.2) and (3.6) gives

$$\begin{aligned}
&-u_s^+(1,1+s-x)\\
&\quad=\tau'(x)\left[u^+(x,s)+\int_0^s T(x,s-s')u^+(x,s')ds'\right]\\
&\qquad+\tau(x)\Big\{-\gamma(x)u^+(x,s)-u_s^+(x,s)\\
&\qquad+\beta(x)\left[\rho(x)u^+(x,s-2(1-x))\right.\\
&\qquad\left.+\int_0^s R^+(x,s-s')u^+(x,s')ds'\right]\\
&\qquad+\frac{\partial}{\partial x}\int_0^s T(x,s-s')u^+(x,s')ds'\Big\}. \qquad (3.18)
\end{aligned}$$

Similarly, a differentiation with respect to s gives

$$\begin{aligned}
&u_s^+(1,1+s-x)\\
&\quad=\tau(x)\left[u_s^+(x,s)+\frac{\partial}{\partial s}\int_0^s T(x,s-s')u^+(x,s')ds'\right]. \qquad (3.19)
\end{aligned}$$

As in the derivation of the equation for the reflection kernel, special consideration has to be taken to the presumed discontinuities of $T(x,s)$. This time it is enough to assume one discontinuity along a curve $s=d(x)$.

After some lengthy calculations with Eqs. (3.18) and (3.19), the final result is

$$\begin{aligned}
T_x(x,s)=&-\tfrac{1}{2}[A(x)+B(x)]\Big\{R^+(x,s)\\
&+\int_0^s T(x,s-s')R^+(x,s')ds'\\
&+H(s-2(1-x))\rho(x)T(x,s-2(1-x))\Big\},\\
&s>0,\quad s\neq 2(1-x), \qquad (3.20)
\end{aligned}$$

$$T(1,s)=0,\quad s>0,$$

and

$$d(x)=2(1-x),$$

where the jump is

$$[T(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+}=-\tfrac{1}{4}\rho(x)[A(x)+B(x)]. \qquad (3.21)$$

The early time behavior of T is obtained by integrating Eq. (3.20),

$$\begin{aligned}
T(x,0^+)=&-\tfrac{1}{8}\int_x^1[A^2(x')-B^2(x')]dx'\\
&+\tfrac{1}{4}r[A(1)+B(1)]. \qquad (3.22)
\end{aligned}$$

This imbedding equation also has an additional term not present in the continuous case and, furthermore, the kernel T is discontinous along the line $s=2(1-x)$. These are effects due to the jump in ϵ. Notice again, that as $c_1\to 1$, the results for the transmission kernel above reduce to the results in Part I.

The imbedding equation for the W kernel can be obtained similarly to the derivation presented above. However, there is another way of deriving the desired result by employing the resolvent equation of T. A combination of Eqs. (3.7) and (3.8) gives the resolvent equation for the transmission kernel T

$$T(x,s)+W(x,s)+\int_0^s W(x,s-s')T(x,s')ds'=0,\quad s>0. \qquad (3.23)$$

This equation is now differentiated with respect to x. The jump discontinuity of W is the negative of that for T. Using the imbedding equation for the transmission operator, Eq. (3.20), and repeated usage of the resolvent equation finally gives the result

$$\begin{aligned}
W_x(x,s)=&\tfrac{1}{2}[A(x)+B(x)]\Big\{R^+(x,s)\\
&+\int_0^s W(x,s-s')R^+(x,s')ds'\\
&+H(s-2(1-x))\rho(x)W(x,s-2(1-x))\Big\},\\
&s>0,\quad s\neq 2(1-x), \qquad (3.24)
\end{aligned}$$

and the jump is

$$[W(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+}=\tfrac{1}{4}\rho(x)[A(x)+B(x)], \qquad (3.25)$$

and

$$\begin{aligned}
&W(x,0^+)\\
&\quad=\frac{1}{8}\int_x^1[A^2(x')-B^2(x')]dx'-\frac{1}{4}r[A(1)+B(1)]. \qquad (3.26)
\end{aligned}$$

In Part I, Appendix A, it is shown that as a function of s the kernel $W(x,s)$ has compact support in $[0,2(1-x)]$. The jump condition in Eq. (3.25) therefore gives the value of $W(x,s)$ at $s=2(1-x)^-$, namely

$$W(x,2(1-x)^-)=-\tfrac{1}{4}\rho(x)[A(x)+B(x)]. \qquad (3.27)$$

The final imbedding equation for a variation of the left end point of the slab is the equation for the propagator kernel V^+. Its relation to R^+ and T is obtained by inserting Eqs. (3.6) and (3.7) into Eq. (3.9). This yields

$$\begin{aligned}
&R^+(x,s)\\
&\quad=V^+(x,s)+\int_0^s T(x,s-s')V^+(x,s')ds'\\
&\qquad+H(s-2(1-x))\rho(x)T(x,s-2(1-x)),\quad s>0, \qquad (3.28)
\end{aligned}$$

or by solving for $V^+(x,s)$ using Eq. (3.23),

$$V^+(x,s)$$
$$= R^+(x,s) + \int_0^s W(x,s-s')R^+(x,s')ds'$$
$$+ H(s-2(1-x))\rho(x)W(x,s-2(1-x)), \quad s>0. \tag{3.29}$$

(Notice that an immediate comparison between Eqs. (3.24) and (3.29) gives

$$W_x(x,s) = \tfrac{1}{2}[A(x)+B(x)]V^+(x,s), \quad s\neq 2(1-x).) \tag{3.30}$$

The desired equation for $V^+(x,s)$ now follows as above. Differentiate Eq. (3.29) once with respect to x and once with respect to s and use the imbedding equations for $R^+(x,s)$ and $W(x,s)$ together with repeated usage of Eq. (3.29). The final result is

$$V_x^+(x,s) = 2V_s^+(x,s) - B(x)V^+(x,s)$$
$$+ \tfrac{1}{2}[A(x)-B(x)]W(x,s), \quad s>0, \tag{3.31}$$
$$V^+(x,0^+) = -\tfrac{1}{4}[A(x)-B(x)], \quad 0<x<1. \tag{3.32}$$

The same calculation also shows the jump discontinuity in $V^+(x,s)$ at $s=2(1-x)$. But it is known from Appendix A of Part I that as a function of s the kernel $V^+(x,s)$ has its support in $[0,2(1-x)]$. Therefore, the jump discontinuity implies that

$$V^+(x,2(1-x)^-)$$
$$= -\frac{1}{4}\exp\left[\int_x^1 B(x')dx'\right]\left\{A(1)-B(1)\right.$$
$$\left.-\frac{1}{2}r\int_x^1 [A^2(x')-B^2(x')]dx'\right\}. \tag{3.33}$$

A series of numerical computations illustrates the new features of the kernels derived above. The ϵ and σ profiles as functions of the depth z are depicted in Fig. 1. Figures 2–5 show the properties of the reflection kernel $R^+(0,s)$, the transmission kernel $T(0,s)$, and the propagator kernel $W(0,s)$ for times up to three round trips for various values of the parameter c_1.

In Fig. 2 the value of the parameter c_1 is $\sqrt{1/6}$. This particular choice of c_1 corresponds to a case with vacuum on the right-hand side of the slab. The first and second jumps (at $s=2$ and $s=4$) in the reflection kernel $R^+(0,s)$ are clearly seen. Notice also the jumps in $T(0,s)$ and $W(0,s)$ at one round trip ($s=2$). In Fig. 3, $c_1=1$ and the transition in ϵ to the homogeneous background is continuous. The second jump in $R^+(0,s)$ at two round trips ($s=4$) vanishes and $T(0,s)$ and $W(0,s)$ are now continuous functions of time s. In general, the amplitude of the kernels in the second round trip is smaller in this continuous case compared to the kernels in Fig. 2. This is due to the lack of the hard echo from the back wall. The kernels for a slab backed up with a homogeneous medium of large relative permittivity ($c_1=2$) are shown in Fig. 4. The limit value $c_1=\infty$ corresponds to a perfectly conducting back wall at $z=L$. The kernels for this limit of c_1 are shown in Fig. 5. For the case $c_1=\infty$, the kernels $T(0,s)$ and $W(0,s)$ have no physical meaning and they are obtained as limits as $c_1\to\infty$. However, these kernels have mathematical meaning in the sense that they can be used in the solution of the inverse problem. In fact, it is shown in Part IV that even in the case $c_1=\infty$, the kernels $T(0,s)$ and $W(0,s)$ can be obtained uniquely from reflection data. Notice that the reflection kernels in all the examples presented above are identical for times less than one round trip. Notice also that the amplitude of the reflection kernel in the limit case $c_1=\infty$ remains large even after one and two round trips.

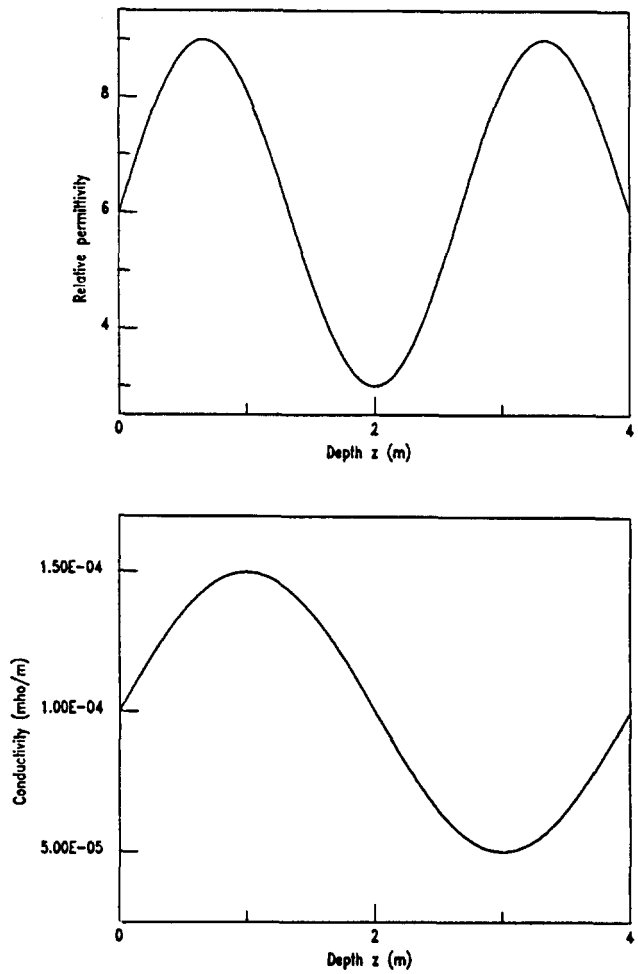


FIG. 1. The relative permittivity and conductivity profiles in the numerical examples.

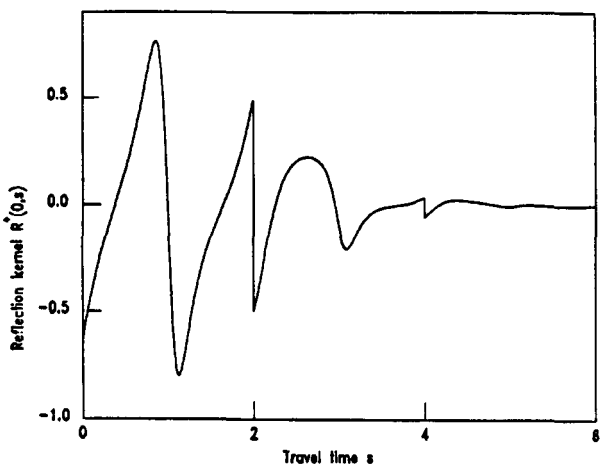
IV. THE EXTENSION OF DATA

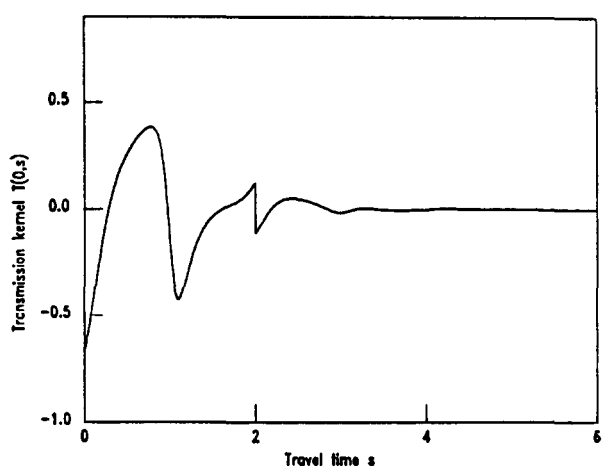
The concept of extension of data, which was developed in Part I, is here generalized to the case with a hard back wall. In this section the variable x is assumed to be fixed.

The important fact that both $W(x,s)$ and $V^+(x,s)$ vanish for times greater than one round trip in the subregion $[x,1]$ is the key to the extension of the scattering kernels. The compact support of W and V^+ is

$$W(x,s)=0, \quad s>2(1-x), \tag{4.1}$$
$$V^+(x,s)=0, \quad s>2(1-x). \tag{4.2}$$

The extension of the transmission kernel follows the same derivation as the continuous case, since the resolvent equation, Eq. (3.23), is the same in both cases. Thus, rewrite Eq. (3.23) for $s>2(1-x)$ and use Eq. (4.1) to obtain





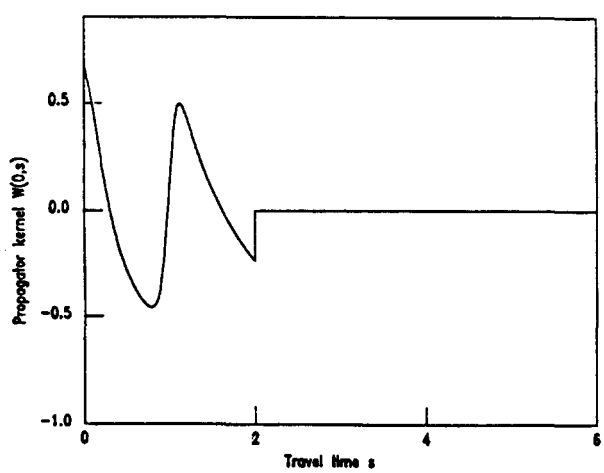
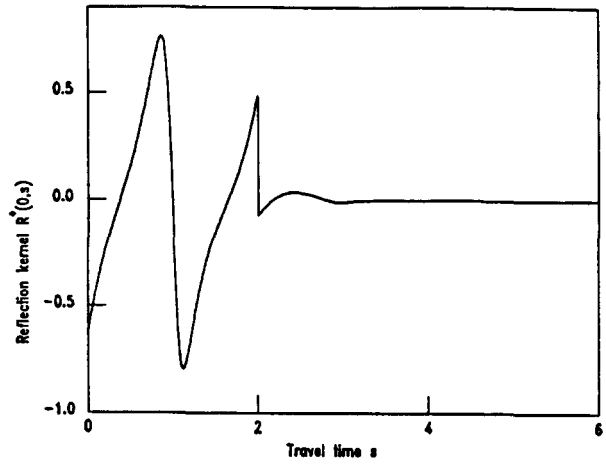
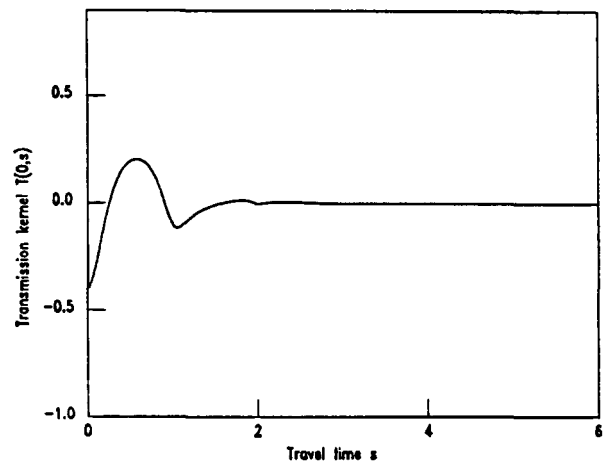


FIG. 2. The physical reflection, transmission, and propagator kernels $R^+(0,s)$, $T(0,s)$, and $W(0,s)$ for three round trips in the medium. The jump discontinuity at the back edge of the slab is $c_1 = \sqrt{1/6}$.





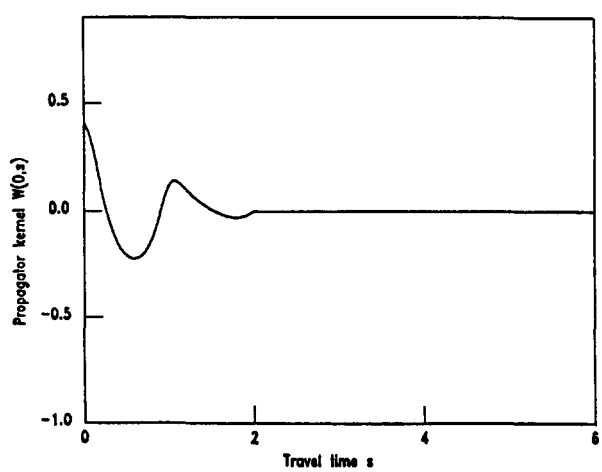
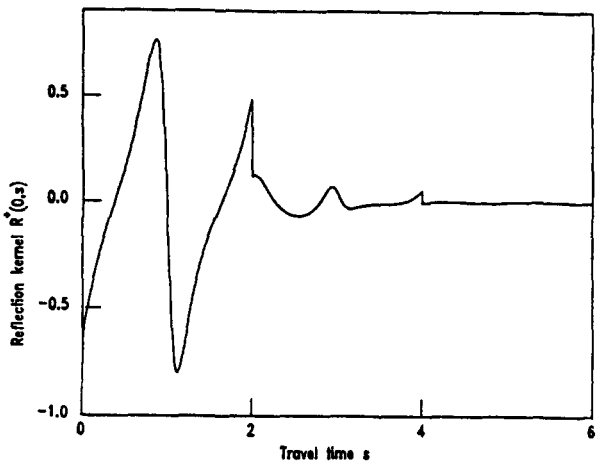
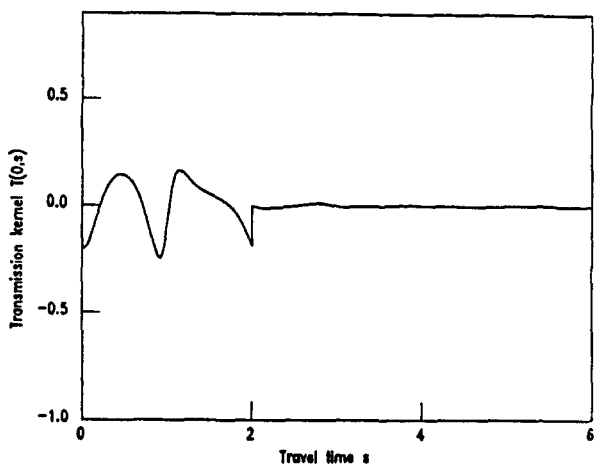


FIG. 3. The same as in Fig. 2 but the permittivity profile is continuous at the back edge, i.e., $c_1 = 1$.





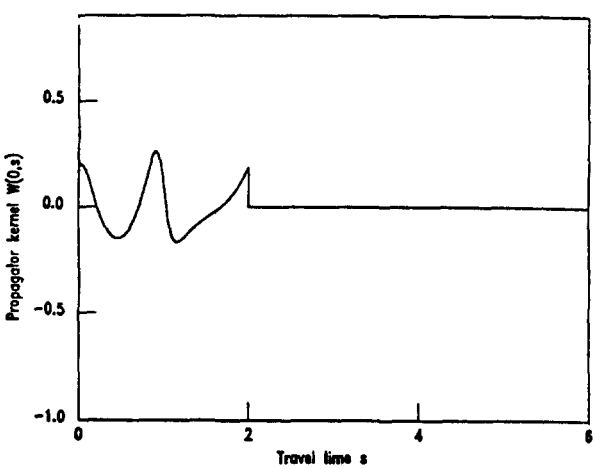
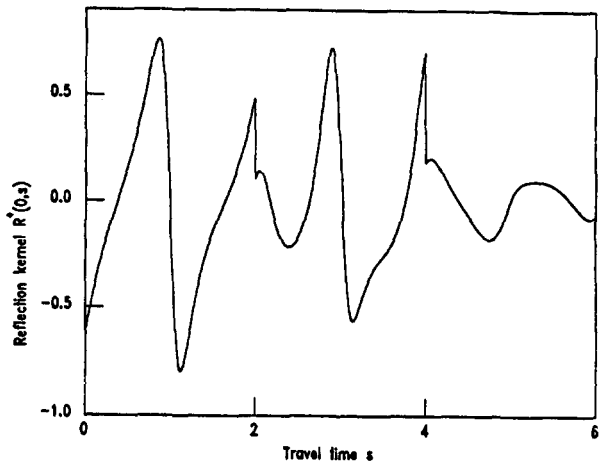
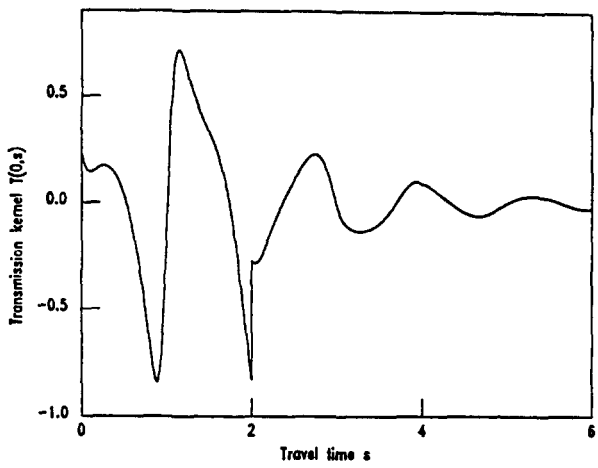


FIG. 4. The same as in Fig. 2 but the jump discontinuity at the back edge of the slab is $c_1 = 2$.





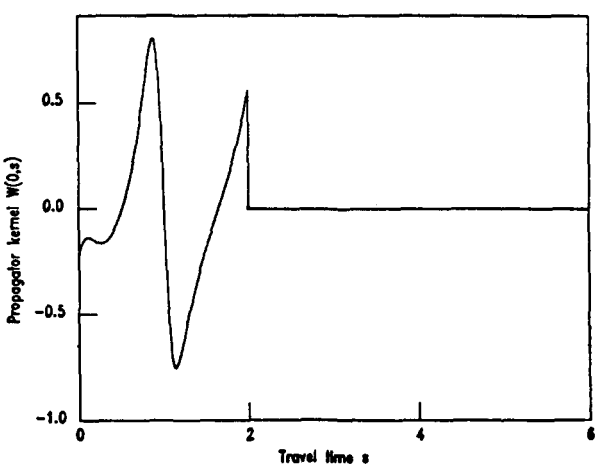


FIG. 5. The same as in Fig. 2 but the slab is backed up with a perfectly conducting medium.

$$T(x,s)+\int_{2(1-x)}^{s} W(x,s-s')T(x,s')ds' = G(x,s) = \begin{cases} -\int_{s-2(1-x)}^{2(1-x)} W(x,s-s')T(x,s')ds', & 2(1-x)<s<4(1-x), \\ 0, & s>4(1-x). \end{cases} \tag{4.3}$$

Suppose the transmission kernel $T(x,s)$ is known for one round trip, i.e., $0<s<2(1-x)$. The resolvent kernel $W(x,s)$ is then completely known for all $s>0$ by Eq. (4.1) and so is the function $G(x,s)$ for $s>2(1-x)$. Equation (4.3), which is a Volterra equation of the second kind for $T(x,s)$ for $s>2(1-x)$, then determines the transmission kernel uniquely for $s>2(1-x)$. Thus, data at a fixed x for one round trip completely determine the transmission kernel $T(x,s)$ for all s (at the fixed value of x). In particular, the jump in $T(x,s)$ at $s=2(1-x)$ is

$$\begin{aligned}[T(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+} &= -\int_0^{2(1-x)} W(x,2(1-x)-s')T(x,s')ds' - T(x,2(1-x)^-) \\ &= W(x,2(1-x)^-),\end{aligned} \tag{4.4}$$

which agrees with the results in Sec. III.

The extension of the reflection kernel for times beyond one round trip is also quite similar to the derivation in Part I. However, one additional term is now present due to the jump discontinuity in ϵ at the right edge. Straightforward calculations with Eqs. (3.28) and (3.29) for $s>2(1-x)$ and the compact support of $V^+(x,s)$ in Eq. (4.2) give

$$R^+(x,s) = \rho(x)T(x,s-2(1-x)) + \int_0^{2(1-x)} T(x,s-s')\left[R^+(x,s') + \int_0^{s'} R^+(x,s'-s'')W(x,s'')ds''\right]ds', \quad s>2(1-x). \tag{4.5}$$

Notice that only R^+ data for $0<s<2(1-x)$ enter in the integral on the right-hand side. Notice also that to be able to extend the reflection data beyond one round trip the transmission data $T(x,s)$ have to be known for $0<s<2(1-x)$ so that the resolvent kernel $W(x,s)$ and the extension of $T(x,s)$ can be obtained.

Another approach, more analogous to the extension of transmission data in Eq. (4.3), is to rewrite Eq. (3.29) for $s>2(1-x)$ and use Eq. (4.2) to obtain

$$R^+(x,s)+\int_{2(1-x)}^{s} W(x,s-s')R^+(x,s')ds' = g(x,s) - \rho(x)W(x,s-2(1-x)), \quad s>2(1-x), \tag{4.6}$$

where

$$g(x,s) = \begin{cases} -\int_{s-2(1-x)}^{2(1-x)} W(x,s-s')R^+(x,s')ds', & 2(1-x)<s<4(1-x), \\ 0, & s>4(1-x). \end{cases} \tag{4.7}$$

Equation (4.6), which is a Volterra equation of the second kind, determines the reflection kernel for $s>2(1-x)$. Once again notice that transmission data for one round trip are necessary for the extention of reflection data.

The jump in $R^+(x,s)$ at $s=2(1-x)$ is in both cases

$$\begin{aligned}[R^+(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+} &= \rho(x)T(x,0^+) + \int_0^{2(1-x)} T(x,2(1-x)-s')\left[R^+(x,s') + \int_0^{s'} W(x,s'-s'')R^+(x,s'')ds''\right]ds' - R^+(x,2(1-x)^-) \\ &= -\rho(x)W(x,0^+) - \int_0^{2(1-x)} W(x,2(1-x)-s')R^+(x,s')ds' - R^+(x,2(1-x)^-).\end{aligned} \tag{4.8}$$

The jump at $s=4(1-x)$ is just the identity

$$[R^+(x,s)]_{s=4(1-x)^-}^{s=4(1-x)^+} = \rho(x)[T(x,s)]_{s=2(1-x)^-}^{s=2(1-x)^+}. \tag{4.9}$$

V. SUMMARY AND CONCLUSIONS

The scattering problem treated in this paper is a generalization of the problem analyzed in Part I in that the permittivity of the slab is no longer assumed to be continuous across the interfaces $z=0$ and $z=L$. Thus, in this more general problem there are phase velocity mismatches at the interfaces of the slab which give rise to hard echoes from the front and back walls.

It is shown by employing a Redheffer star product technique that the effects of the leading edge discontinuity in the permittivity can be removed (Appendix A), and similarly for the effects from the back edge discontinuity (Appendix B). Thus, the direct scattering problem for the resulting continuous permittivity profile can be solved with the technique developed in Part I. However, in the solution of the inverse problem (Part IV) it is not desirable to remove the back edge discontinuity. In fact, it is shown in Part IV that by retaining

the back wall it is possible to reduce the amount of data required for simultaneous reconstruction of both the permittivity and the conductivity. It is the importance of the back edge discontinuity and the properties of this scattering problem that are developed in this paper.

With this idea in mind, equations for the scattering and propagator kernels for a slab with a back edge jump discontinuity are developed in Sec. III. These kernels show several new features in comparison with the results in Part I, the most notable being the presence of additional jump discontinuities in the kernels due to the hard reflector at the back wall.

The equations developed in Sec. III can be used to construct the scattering kernels $R^+(0,s)$ and $T(0,s)$ for arbitrarily large values of s. However, in Sec. IV it is shown that it suffices to compute the kernels for one round trip only $(0<s<2)$. For larger values of s the scattering kernels can be computed via extension of data, which amounts to solving a Volterra equation of the second kind. In terms of speed and accuracy, this is computationally superior to using the equations in Sec. III for $s>2$. Similarly, the technique of Appendix B is best used to compute scattering and propagator kernels for one round trip only, after which extension of data is used for larger values of s.

The model problem used in this paper is the same as that in Refs. 5 and 6. However, the approach used in the present series of papers is entirely different and more intuitive than in those earlier articles. The reason this approach is more intuitive is because the scattering and propagation operators are built up using invariant imbedding ideas, in which the physical scattering medium is envisioned to be one element in a set of media whose scattering operators are easily related to one another. The present approach is also considerably more general than that used in Refs. 5 and 6, as it has already been shown to be a viable technique in a variety of other problems, such as scattering in dispersive media[7] and viscoelastic media[8] as well as analysis of noise and bandlimiting effects in inverse problems.[9]

ACKNOWLEDGMENTS

This work was performed at Ames Laboratory, U.S. D.O.E., Iowa State University, and was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract No. W-7405-ENG-82, and by the Office of Naval Research Contract No. N0014-83-K-0038. One of the authors (G.K.) would like to thank Ames Laboratory for their hospitality during his visit in the United States. He is also grateful to the National Swedish Board for Technical Development (STUF) for partial support.

APPENDIX A: SIMPLIFIED EXPRESSIONS FOR THE SCATTERING OPERATORS

The form of the scattering operators $\mathscr{R}^+$ and $\mathscr{T}^+$ in Eqs. (2.8) and (2.9) and the propagation operators $\mathscr{W}^+$ and $\mathscr{V}^+$ in Eqs. (2.10) and (2.11) are the result of a simplification introduced in Sec. II, namely, the assumption that without loss of generality, the permittivity $\epsilon(z)$ can be assumed to be continuous at $z=0$. It is now shown that (1) if the permittivity is continuous at $z=0$, then the Eqs. (2.8)–(2.11) result, and (2) if the permittivity is not continuous at $z=0$, then suitable transformations of the physical scattering operators permit Eqs. (2.8) and (2.9) to still be used as the starting point for the direct and inverse problems.

Turning to the first point, assume that $c_0=1$ in Eq. (2.6). It follows from Theorem 2 of Ref. 5 (with suitable change of notation) that Eqs. (2.10) and (2.11) are valid for all $s>0$. With $T(0,s)$ defined to be the resolvent kernel of $W(0,s)$ [see Eq. (3.23) with $x=0$], Eq. (2.9) is then valid for all $s>0$. Finally, Eq. (2.8) results upon substituting Eq. (2.9) into Eq. (2.11).

Addressing the second point, assume that the permittivity is not continuous at $z=0$, so that $c_0\neq 1$. Then the operators given in Eqs. (2.8) and (2.9) are not the physical scattering operators in the sense that they do not correctly relate the physical fields, u^i_+, u^r_+, and u^t_+, to each other. In order to obtain the proper relations between these fields, star products of scattering operators will be used. This is done by assuming that the scattering medium is composed of two portions, the first consisting of the leading edge jump discontinuity in ϵ and the second consisting of the remainder of the medium as shown in Fig. 6. The scattering operators for the first portion are the familiar operators for medium 1 illustrated in Fig. 6. Denoting these operators with a subscript 1 results in

$$\begin{aligned} u^r_+(s) &= [\mathscr{R}_1^+ u^i_+(\cdot)](s) = r_1 u^i_+(s), \\ u^r_-(s) &= [\mathscr{R}_1^- u^i_-(\cdot)](s) = -r_1 u^i_-(s), \\ u^t_+(s) &= [\mathscr{T}_1^+ u^i_+(\cdot)](s) = t_1^+ u^i_+(s), \\ u^t_-(s) &= [\mathscr{T}_1^- u^i_-(\cdot)](s) = t_1^- u^i_-(s), \end{aligned} \tag{A1}$$

where

$$r_1=(c_0-1)/(c_0+1),$$
$$t_1^+=2c_0/(c_0+1), \quad t_1^-=2/(c_0+1).$$

The scattering operators for the remainder of the medium are those given in Eqs. (2.8) and (2.9). These are not denoted with a subscript as they are the operators considered

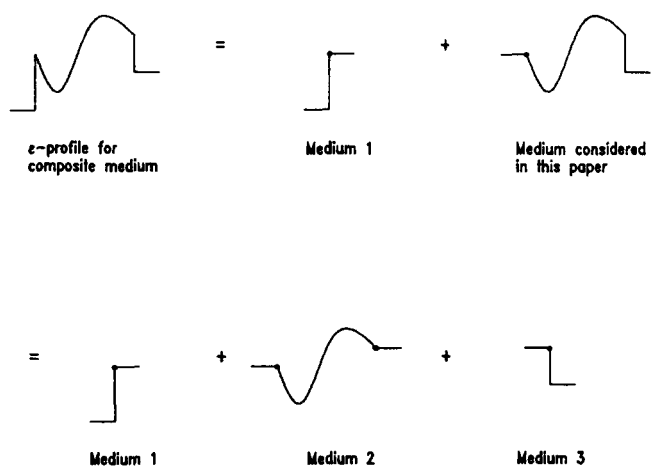


FIG. 6. Decomposition of the discontinuous permittivity profile. The conductivity profile does not need to be decomposed.

throughout this paper. The physical reflection and transmission operators can be built up using the composition of operators from these two subregions. Denote these physical operators with a subscript c (for composite). Using the Redheffer star product,[10] it follows that

$$\mathscr{R}_c^+ = \mathscr{R}_1^+ + \mathscr{T}_1^- \mathscr{R}^+ (1 - \mathscr{R}_1^- \mathscr{R}^+)^{-1} \mathscr{T}_1^+, \tag{A2}$$

$$\mathscr{T}_c^+ = \mathscr{T}^+ (1 - \mathscr{R}_1^- \mathscr{R}^+)^{-1} \mathscr{T}_1^+, \tag{A3}$$

where 1 is the identity operator.

Since the operator $1 - \mathscr{R}_1^- \mathscr{R}^+$ is a delay Volterra operator of the type studied in Ref. 11 it can be inverted. Write the inverse as

$$(1 - \mathscr{R}_1^- \mathscr{R}^+)^{-1} = 1 + \mathscr{R}_1^- \mathscr{R}^+ + (\mathscr{R}_1^- \mathscr{R}^+)^2 + \cdots .$$

Now substitute Eqs. (2.8), (2.9), and (A1) into (A2) and (A3) to obtain the general form of the composite operators. These are seen to be

$$\begin{aligned}[\mathscr{R}_c^+ u_+^i (\cdot)](s) &= r_1 u_+^i (s) + t_1^+ t_1^- \\ &\quad \times \sum_{n=1}^{\infty} (-r_1)^{n-1} (\rho(0))^n u_+^i (s-2n) \\ &\quad + \int_0^s R_c^+ (s-s') u_+^i (s') ds', \end{aligned} \tag{A4}$$

$$\begin{aligned}[\mathscr{T}_c^+ u_+^i (\cdot)](s) &= \tau(0) t_1^+ \left[\sum_{n=0}^{\infty} (-r_1 \rho(0))^n u_+^i (s-2n) \right. \\ &\quad \left. + \int_0^s T_c (s-s') u_+^i (s') ds' \right]. \end{aligned} \tag{A5}$$

Now a relation between the kernel R_c^+ above and the kernel R^+ can be derived. Rewrite Eq. (A2) by operating with $1 - \mathscr{R}_1^- \mathscr{R}^+$ and use the fact that $\mathscr{R}^+$ and $\mathscr{R}_1^-$ commute with $\mathscr{T}_1^+$ to obtain

$$(\mathscr{R}_c^+ - \mathscr{R}_1^+)(1 - \mathscr{R}_1^- \mathscr{R}^+) = \mathscr{T}_1^- \mathscr{R}^+ \mathscr{T}_1^+ .$$

Expanding this using Eq. (A4) produces

$$\begin{aligned} R^+(0,s) - R_c^+ (s) - H(s-2) r_1 &\times [\rho(0) R^+(0,s-2) + R_c^+ (s-2)] \\ &+ H(s-4)(r_1 \rho(0))^2 R^+(0,s-4) - r_1 (t_1^+ t_1^-)^{-1} \\ &\times \int_0^s R^+(0,s') R_c^+ (s-s') ds' = 0, \quad 0 < s < 6, \end{aligned}$$

where $H(s)$ is the Heaviside function. This is again a Volterra equation of delay type and consequently R^+ can be uniquely determined from R_c^+ and conversely.

Next, a relation between T_c and T is derived. Begin by writing Eq. (A3) as

$$\mathscr{T}_c^+ (1 - \mathscr{R}_1^- \mathscr{R}^+) = \mathscr{T}^+ \mathscr{T}_1^+ .$$

Expanding this using Eq. (A5) yields

$$\begin{aligned} T(0,s) = T_c (s) + r_1 R^+(0,s) + H(s-2) r_1 \rho(0) &\times [T_c (s-2) - r_1 R^+(0,s-2)] \\ &+ H(s-4) r_1^3 (\rho(0))^2 R^+(0,s-4) \\ &+ r_1 \int_0^s T_c (s') R^+(0,s-s') ds', \quad 0 < s < 6. \end{aligned}$$

Thus, T can be found from T_c and R^+ and conversely T_c can be found from T and R^+.

APPENDIX B: AN ALTERNATE APPROACH TO THE DIRECT PROBLEM

It was shown in Appendix A that the effect of the discontinuity in permittivity at $z = 0$ can be removed from both the direct and inverse problem. This produced a much simpler formulation of those problems than would have been otherwise possible. In the same manner, it is possible to remove the effect of the discontinuity in permittivity at $z = L$. For the inverse problem, removing this discontinuity is not desirable. It will be shown in Part IV that the discontinuity simplifies the implementation of the inversion algorithm and reduces the amount of data needed as compared to the continuous case. On the other hand, removing the discontinuity in the direct problem provides a distinct computational advantage. This point is now examined.

In Appendix A the physical scattering medium was decomposed into two portions, the first being just the leading edge discontinuity in ϵ and the other being the remaining portion of the medium. That remaining portion will not be further subdivided into two simpler media, one consisting of the back edge jump discontinuity in ϵ and the other consisting of the remainder of the medium. As shown in Fig. 6, the operators for these media will be denoted with subscripts 3 and 2, respectively. The entire conductivity profile can be incorporated into medium 2 as the discontinuities in σ do not produce hard reflections. The scattering operators for medium 3 are given by

$$u_+^r (s) = [\mathscr{R}_3^+ u_+^i (\cdot)](s) = r u_+^i (s),$$

$$u_+^t (s) = [\mathscr{T}_3^+ u_+^i (\cdot)](s) = t_3^+ u_+^i (s),$$

where

$$r = (1 - c_1)/(1 + c_1),$$

$$t_3^+ = 2/(1 + c_1).$$

The scattering operators for medium 2 are the same as those used in Parts I and II:

$$\begin{aligned} u_+^r (s) &= [\mathscr{R}_2^+ u_+^i (\cdot)](s) \\ &= \int_0^s R_2^+ (0,s-s') u_+^i (s') ds', \end{aligned}$$

$$\begin{aligned} u_-^r (s) &= [\mathscr{R}_2^- u_-^i (\cdot)](s) \\ &= \int_0^s R_2^- (0,s-s') u_-^i (s') ds', \end{aligned}$$

$$\begin{aligned} u_\pm^t (s) &= [\mathscr{T}_2^\pm u_\pm^i (\cdot)](s) \\ &= t^\pm (0,1) \left[u_\pm^i (s) + \int_0^s T_2 (0,s-s') u_\pm^i (s') ds' \right]. \end{aligned}$$

The goal is now to express the $\mathscr{R}^+$, $\mathscr{T}^+$, and $\mathscr{W}^+$ operators of Sec. II in terms of the operators given above. Consider the transmission operator first. It follows from using the star product that

$$\mathscr{T}^+ = \mathscr{T}_3^+ (1 - \mathscr{R}_2^- \mathscr{R}_3^+)^{-1} \mathscr{T}_2^+$$

or

$$(1 - \mathscr{R}_2^- \mathscr{R}_3^+) \mathscr{T}^+ = \mathscr{T}_3^+ \mathscr{T}_2^+ .$$

Expressing these operators in terms of their kernels results in the identity

$$T_2(0,s) = T(0,s) - rR_2^-(0,s) - r\int_0^s R_2^-(0,s')T(0,s-s')ds', \quad s>0. \tag{B1}$$

It follows that if R_2^- and T_2 are known, then T can be found as it is the solution of a Volterra second kind equation. This completes the process of finding T in terms of simpler kernels.

It is possible to use Eq. (B1) to obtain the propagator kernel $W(0,s)$ in terms of kernels for the continuous medium 2. To do this, use the fact that $W(0,s)$ is the resolvent kernel for $T(0,s)$ and solve Eq. (B1) for $rR_2^-(0,s)$ to obtain (after some simplification)

$$rR_2^-(0,s) + T_2(0,s) + W(0,s) + \int_0^s W(0,s')T_2(0,s-s')ds' = 0, \quad s>0. \tag{B2}$$

Finally, solve Eq. (B2) for $W(0,s)$ using the fact that $W_2(0,s)$ is the resolvent kernel for $T_2(0,s)$. The results in

$$W(0,s) = W_2(0,s) - rR_2^-(0,s) - r\int_0^s R_2^-(0,s')W_2(0,s-s')ds', \quad s>0. \tag{B3}$$

Notice that this is an explicit expression for $W(0,s)$. This equation can be expressed in terms of the propagator kernel $V_2^-(0,s)$, by using Eq. (3.20) in Part I. The result is

$$W(0,s) = W_2(0,s) - rV_2^-(0,s), \quad s>0. \tag{B4}$$

Finally, consider the reflection kernel R^+. Using the star product, it is seen that

$$\mathscr{R}^+ = \mathscr{R}_2^+ + \mathscr{T}_2^-\mathscr{R}_3^+(1-\mathscr{R}_2^-\mathscr{R}_3^+)^{-1}\mathscr{T}_2^+,$$

or, since the Volterra operators $(1-\mathscr{R}_2^-\mathscr{R}_3^+)$ and $\mathscr{T}_2^-\mathscr{R}_3^+$ commute,

$$(1-\mathscr{R}_2^-\mathscr{R}_3^+)(\mathscr{R}^+ - \mathscr{R}_2^+) = \mathscr{T}_2^-\mathscr{R}_3^+\mathscr{T}_2^+.$$

Expressing these operators in terms of their kernels results in

$$R^+(0,s) - R_2^+(0,s) - r\int_0^s R_2^-(0,s-s') \times [R^+(0,s') - R_2^+(0,s')]ds'$$

$$= H(s-2)\rho(0)\Big[2T_2(0,s-2) + rR_2^-(0,s-2) + \int_0^{s-2} T_2(0,s-2-s')T_2(0,s')ds'\Big], \quad s>0, \tag{B5}$$

where $H(s)$ is again the Heaviside function. Equation (B5) is a Volterra equation of the second kind for $R^+ - R_2^+$. In particular, for $0<s<2$ the right-hand side of Eq. (B5) vanishes which implies that

$$R^+(0,s) = R_2^+(0,s), \quad 0<s<2$$

as was mentioned in Sec. III. For $s>2$, the kernel R^+ can be uniquely determined from Eq. (B5) provided R_2^+, R_2^-, and T_2 are known.

[1] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. I: Scattering operators," J. Math. Phys. **27**, 1667 (1986).

[2] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. II. Simultaneous reconstruction of dissipation and phase velocity profiles," J. Math. Phys. **27**, 1683 (1986).

[3] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. IV. Use of phase velocity mismatches to simplify inversions," preprint.

[4] I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1968), Vol. II, pp. 44-46.

[5] R. J. Kureger, "An inverse problem for a dissipative hyperbolic equation with discontinuous coefficients," Q. Appl. Math. **34**, 129 (1976).

[6] R. J. Krueger, "Numerical aspects of a dissipative inverse problem," IEEE Trans. Antennas Propag. **AP-29**, 253 (1981).

[7] R. S. Beezley and R. J. Krueger, "An electromagnetic inverse problem for dispersive media," J. Math. Phys. **26**, 317 (1985).

[8] E. Ammicht, J. P. Corones, and R. J. Krueger, "Direct and inverse scattering for viscoelastic media," Ames Laboratory preprint.

[9] J. P. Corones, R. J. Krueger, and C. R. Vogel, "The effects of noise and bandlimiting on a one dimensional time dependent inverse scattering technique," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1985), Vol. 4, pp. 551-558.

[10] R. Redheffer, "On the relation of transmission-line theory to scattering and transfer," J. Math. Phys. (Cambridge, Mass.) **41**, 1 (1962).

[11] R. J. Krueger, "Some results on linear delay integral equations," J. Math. Anal. Appl. **67**, 232 (1979).

Implementability of gauge transformations and quantization of fermions in external fields

H. Grosse and G. Karner[a)]
Institut für Theoretische Physik, Universität Wien, A-1090 Wien, Boltzmanngasse 5, Austria

(Received 3 July 1986; accepted for publication 15 October 1986)

Quantization of fermions in an external soliton field, leading to a representation of the canonical anticommutation relation (CAR), which is inequivalent to the representation connected to the massive Dirac operator, is studied. Classes of gauge and axial gauge transformations that can be unitarily implemented are determined. In the latter case quantization conditions for gauge functions are obtained; integers entering can be interpreted as winding numbers.

I. INTRODUCTION

Although extensive treatments of the quantization of fermions in external fields exist in the literature,[1] one usually deals with potentials that vanish at infinity. New problems arise if one takes potentials with nontrivial asymptotics: the simplest situation is given by a one-dimensional Dirac operator with external soliton potential. In this paper we shall treat the second quantization of the one-particle operator

$$(H\psi)(x) = (\alpha p + \beta \tanh x)\psi(x) = E\psi(x), \tag{1.1}$$

with $\psi \in L^2(\mathbb{R}) \otimes C^2 = \mathscr{H}$; α and β denote two σ matrices, which we choose to be $\alpha = -\sigma_2$ and $\beta = \sigma_1$. Equation (1.1) is of the special form

$$H = \begin{pmatrix} 0 & A \\ A^\dagger & 0 \end{pmatrix}, \quad A = \frac{d}{dx} + \tanh x, \tag{1.2}$$

which leads to two Schrödinger operators:

$$\begin{aligned} AA^\dagger\psi_1 &= E^2\psi_1, \quad AA^\dagger = -\frac{d^2}{dx^2} + 1, \\ A^\dagger A\psi_2 &= E^2\psi_2, \quad A^\dagger A = -\frac{d^2}{dx^2} + 1 - \frac{2}{\cosh^2 x}; \end{aligned} \tag{1.3}$$

$AA^\dagger$ and $A^\dagger A$ are "almost" isospectral, which means that their spectra agree except for a zero mode of the latter. Solving (1.3) now is trivial and leads to generalized eigensolutions corresponding to scattering from the left and right and to positive (+) and negative (−) energies:

$$\begin{aligned} \psi^{(1)}_\pm(k,x) &= \Theta(k)\frac{e^{ikx}}{\sqrt{4\pi}}\begin{pmatrix} 1 \\ \pm\dfrac{-ik + \tanh x}{\sqrt{k^2+1}} \end{pmatrix}, \\ \psi^{(2)}_\pm(k,x) &= \psi^{*(1)}_\pm(k,x). \end{aligned} \tag{1.4}$$

In addition there exists one bound state for energy zero with wave function

$$\psi_s(x) = \begin{pmatrix} 0 \\ \phi_s(x) \end{pmatrix}, \quad \phi_s(x) = \frac{1}{\sqrt{2}\cosh x}. \tag{1.5}$$

The spectrum consists therefore of a single point and two continua from $-\infty$ to $-m$ and from m to ∞, which are twofold degenerate.

Let us remark that the potential treated here is the lowest soliton solution of the modified Korteweg–de Vries equation. We take it for simplicity reasons, although certain results apply to a more general class of potentials. Any potential with nontrivial asymptotics

$$\lim_{x \to \pm\infty} V(x) = \pm m,$$

has exactly one zero energy bound state; in addition we should remark that a suitable defined charge quantum number assigns charge $-\frac{1}{2}$ to the ground state representation which we shall study below.[2] Recently we have shown[3] that this representation of the canonical anticommutation relations (CAR) is inequivalent to the representation connected to the Dirac operator with constant mass. This indicates that the calculation of a ground state charge has to be done carefully and a regularization procedure has to be used. A possible approach uses the resolvent regularization for an index which may be evaluated using scattering theory.[4]

Here we shall start with the representation of the CAR connected to the one-particle Hamiltonian H of Eq. (1.1) and study the question of unitary implementability of gauge transformations. Let $\mathscr{A}$ denote the C^*-algebra generated by operators $a(f)$ and $a^\dagger(g)$, where $a(f)$ depends antilinear on f and $a^\dagger(g)$ linear on g, and f and g belong to $\mathscr{H}$; assume that the operators obey the CAR

$$\{a(f), a^\dagger(g)\} = \langle f, g\rangle, \quad \{a(f), a(g)\} = 0, \quad f,g \in \mathscr{H}, \tag{1.6}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathscr{H}$.

Since the spectrum of H splits into a positive energy part, a negative energy part, and a zero energy part, we may split our Hilbert space into

$$\mathscr{H} = P_+\mathscr{H} \oplus P_s\mathscr{H} \oplus P_-\mathscr{H}, \tag{1.7}$$

where $P_\pm$ and P_s denote projection operators onto the appropriate spectral parts of H. Note that due to the special structure of H, there exists an antiunitary operator C satisfying $C^2 = 1$ such that

$$CH = -HC, \quad CP_+ = P_-C, \quad CP_s = P_sC, \tag{1.8}$$

which means charge conjugation. For the chosen representation, C is given explicitly by

[a)] Present address: Fakultät für Physik, Universität Bielefeld, Federal Republic of Germany.

 0022-2488/87/020371-05$02.50 © 1987 American Institute of Physics

$$(C\phi)(x) = -\sigma_3\phi^*(x), \quad \phi\in\mathscr{H} \tag{1.9}$$

(C will be used later on in order to define a conjugation in $\mathscr{H}$).

In the following we shall study a quasifree state over the algebra $\mathscr{A}$ determined with the help of the projection operator P_+:

$$\omega_s(a(f_n)\cdots a(f_1)a^\dagger(g_1)\cdots a^\dagger(g_m)) = \delta_{nm}\det\langle f_i, P_+ g_j\rangle. \tag{1.10}$$

Choosing P_+ on the rhs amounts to the intuitive idea of filling the negative Dirac sea and will lead to a lower bounded second quantized Hamiltonian. There are actually two pure states we may treat. Replacing P_+ in (1.10) by $P_+ + P_s$ means filing the zero mode, too. This leads to a state $\omega_{\bar{s}}$

$$\omega_{\bar{s}}(a(f_n)\cdots a(f_1)a^\dagger(g_1)\cdots a^\dagger(g_m)) = \delta_{nm}\det\langle f_i,(P_+ + P_s)g_j\rangle. \tag{1.11}$$

Both states ω_s and $\omega_{\bar{s}}$ are pure states.[5] With the help of the Gel'fand, Naimark, Segal (GNS) construction we associate to ω_s as well as to $\omega_{\bar{s}}$ cyclic representations

$$(\mathscr{H}_{\omega_s},\Pi_{\omega_s},\Omega_{\omega_s}) \quad \text{and} \quad (\mathscr{H}_{\omega_{\bar{s}}},\Pi_{\omega_{\bar{s}}},\Omega_{\omega_{\bar{s}}})$$

of $\mathscr{A}$ such that

$$\begin{aligned}\omega_s(x) &= (\Omega_{\omega_s},\Pi_{\omega_s}(x)\Omega_{\omega_s}),\\ \omega_{\bar{s}}(x) &= (\Omega_{\omega_{\bar{s}}},\Pi_{\omega_{\bar{s}}}(x)\Omega_{\omega_{\bar{s}}}), \quad \forall x\in\mathscr{A}.\end{aligned} \tag{1.12}$$

Later on it will be convenient to denote $\Pi_{\omega_s}(a(f))$ for $f\in P_+\mathscr{H}$, $f\in P_s\mathscr{H}$, or $f\in P_-\mathscr{H}$ differently [see (2.4)]:

$$\begin{aligned}&\Pi_{\omega_s}(a^\dagger(P_+f)) = B^\dagger(f),\\ &\Pi_{\omega_s}(a^\dagger(P_sf)) = C(f), \quad \Pi_{\omega_s}(a^\dagger(P_-f)) = D(f).\end{aligned} \tag{1.13}$$

With the help of the charge conjugation we shall construct the physical Hilbert space in Sec. II.

In order to study gauge transformations we start with a group G and a unitary representation of operators $\{V_\alpha,\alpha\in G\}$ acting on $\mathscr{H}$. To any V_α there exists a unique *-automorphism τ_α of $\mathscr{A}$ such that τ_α reduces to V_α on the operators $a(f)$:

$$\tau_\alpha(a(f)) = a(V_\alpha f), \quad \forall f\in\mathscr{H}. \tag{1.14}$$

Examples of automorphism are the time translation automorphism

$$\tau_t(a(f)) = a(e^{itH}f), \quad t\in\mathbb{R}, \tag{1.15}$$

and the charge conjugation automorphism

$$\tau_c(a(f)) = a^\dagger(Cf). \tag{1.16}$$

Our main interest concerns gauge, axial gauge, and chiral transformations of the type

$$\begin{aligned}&\tau_\Theta(a(f)) = a(V_\Theta f), \quad V_\Theta = e^{i\Theta(x)+i\sigma_2\Theta_5(x)},\\ &\Theta,\Theta_5\in C^\infty(\mathbb{R}), \quad \Theta',\Theta_5'\in C_0^\infty(\mathbb{R}),\end{aligned} \tag{1.17}$$

where we made restrictions on Θ and Θ_5 to simplify the proofs. We will require that the automorphism is unitarily implementable; in Sec. III we shall see that this leads to the quantization condition on Θ_5:

$$\lim_{x\to-\infty}\Theta_5(x) = m\pi, \quad \lim_{x\to+\infty}\Theta_5(x) = n\pi,$$

where the integer $n-m$ can be interpreted as a winding number.[6] No restriction is implied on Θ, which reflects the fact that global gauge transformations leave H invariant for any Θ while $\Theta_5 = n\pi$ is required for invariance of axial gauge transformations.

Consequences of this result, an elaboration of the current algebra connected to our problem as well as a discussion of the projective representation obtained from the second quantized form of V_Θ with an appropriate Schwinger term will be published separately.

Let us finally make remarks on the literature. The general discussion of quasifree states over the CAR-algebra has been done by Araki and Wyss,[7] Powers and Størmer,[5] and Lundberg.[8] The representation of the CAR related to the massless as well as massive Dirac operator has been studied extensively.[6,9,10] Various other states have been studied by the Streater group.[11] More recently the current algebra for fermion currents in one plus one dimensions[12] as well as the boson–fermion correspondence[13] and the connection of these problems to infinite-dimensional groups has been studied.[14]

II. CONSTRUCTION OF THE REPRESENTATION $(\mathscr{H}_{\omega_s},\Pi_{\omega_s},\Omega_{\omega_s})$

We start from the splitting of our Hilbert space $\mathscr{H}$ into $P_+\mathscr{H}\oplus P_s\mathscr{H}\oplus P_-\mathscr{H}$, follow Ref. 15, and define first the "physical" Hilbert space $\mathscr{H}^s$ by

$$\mathscr{H}^s = \mathscr{H}_p\oplus\mathscr{H}_s\oplus\mathscr{H}_a, \tag{2.1}$$

where $\mathscr{H}_p$ and $\mathscr{H}_a$ are copies of $\mathscr{H}_+ = P_+\mathscr{H}$ and $\mathscr{H}_s = P_s\mathscr{H}$. We define projection operators $\mathscr{P}_p,\mathscr{P}_s$, and $\mathscr{P}_a$ projecting onto $\mathscr{H}_p,\mathscr{H}_s$, and $\mathscr{H}_a$, respectively, and denote by I_p, I_s, and I_a the identification maps

$$I_p\colon \mathscr{H}_+\to\mathscr{H}_p, \quad I_s\colon \mathscr{H}_s\to\mathscr{H}_s, \quad I_a\colon \mathscr{H}_+\to\mathscr{H}_a. \tag{2.2}$$

From the above information we conclude that there is a bijection γ from $\mathscr{H}\to\mathscr{H}^s$ such that

$$\begin{aligned}\gamma &= I_pP_+ + I_sP_s + I_aCP_-,\\ \gamma^{-1} &= I_p^{-1}\mathscr{P}_p + I_s^{-1}\mathscr{P}_s + CI_a^{-1}\mathscr{P}_a,\end{aligned} \tag{2.3}$$

where C denotes the charge conjugation operator (1.8). With the help of these operators we may identify

$$\begin{aligned}\Pi_{\omega_s}(a^\dagger(f)) &= \begin{cases}b^\dagger(\gamma P_+f) = B^\dagger(f),\\ c(\gamma P_sf) = C(f),\\ d(\gamma P_-f) = D(f),\end{cases}\\ \Pi_{\omega_{\bar{s}}}(a^\dagger(f)) &= \begin{cases}\bar{b}^\dagger(\gamma P_+f) = \bar{B}^\dagger(f),\\ \bar{c}^\dagger(\gamma P_sf) = \bar{C}^\dagger(f),\\ \bar{d}(\gamma P_-f) = \bar{D}(f),\end{cases}\end{aligned} \tag{2.4}$$

for all $f\in\mathscr{H}$. The ground state vectors are now determined by

$$\begin{aligned}&B(f)\Omega_{\omega_s} = C(f)\Omega_{\omega_s} = D(f)\Omega_{\omega_s} = 0,\\ &\bar{B}(f)\Omega_{\omega_{\bar{s}}} = \bar{C}(f)\Omega_{\omega_{\bar{s}}} = \bar{D}(f)\Omega_{\omega_{\bar{s}}} = 0,\end{aligned} \tag{2.5}$$

and identified with the cyclic vectors of the representations

$(\mathscr{H}_{\omega_s},\Pi_{\omega_s},\Omega_{\omega_s})$ and $(\mathscr{H}_{\omega_{\bar{s}}},\Pi_{\omega_{\bar{s}}},\Omega_{\omega_{\bar{s}}})$. The relation between both ground state vectors is trivially given by

$$C^\dagger(f)\Omega_{\omega_s}=\Omega_{\omega_{\bar{s}}},\quad \bar{C}^\dagger(f)\Omega_{\omega_{\bar{s}}}=\Omega_{\omega_s}. \tag{2.6}$$

The fermion field operator can now be decomposed as

$$\psi(f)=B(f)+C(f)+D^\dagger(f), \tag{2.7}$$

and acts on the representation space $\mathscr{H}^s$, which is the closure of the linear hull of vectors of the form

$$\prod_{i,j=1}^{n,m} B^\dagger(f_i)D^\dagger(g_j)\Omega_{\omega_s}$$

and (2.8)

$$\prod_{i,j=1}^{k,l} B^\dagger(h_i)D^\dagger(k_j)C^\dagger(f)\Omega_{\omega_s};$$

thus $\mathscr{H}^s$ can be identified with $\mathscr{H}_{\omega_s}$.

In order to discuss the representation of the time evolution automorphism we first define the Hamiltonian operator H_s acting in $\mathscr{H}^s$ by starting from H [Eq. (1.1)] and mapping the negative part of the spectrum of H onto the postive half-line. More precisely H_s is defined by

$$\gamma e^{iHt}\gamma^{-1}=e^{iH_st}. \tag{2.9}$$

Thus

$$\begin{aligned}&\gamma e^{iHt}\gamma^{-1}\\&\quad=I_pP_+e^{iHt}I_p^{-1}\mathscr{P}_p+I_sP_sI_s^{-1}\mathscr{P}_s\\&\qquad+I_aCP_-e^{iHt}CI_a^{-1}\mathscr{P}_a\\&\quad=e^{itH_s}=\exp it\{I_pP_+HI_p^{-1}\mathscr{P}_p+I_aP_+HI_a^{-1}\mathscr{P}_a\},\end{aligned} \tag{2.10}$$

where obviously $H_s\geqslant 0$ is positive definite and the contribution from the zero mode drops out.

From the above we get for the second quantized operators

$$d\Gamma(H_s)\Omega_{\omega_s}=0,\quad \Gamma(e^{iH_st})=e^{id\Gamma(H_s)t}. \tag{2.11}$$

A similar procedure works obviously for the representation $(\mathscr{H}_{\omega_{\bar{s}}},\Pi_{\omega_{\bar{s}}},\Omega_{\omega_{\bar{s}}})$.

III. UNITARY IMPLEMENTABILITY OF GAUGE TRANSFORMATIONS

We shall consider gauge transformations and axial gauge transformations of the general type

$$V_{\underline{\Theta}}(x)=e^{i\Theta(x)+i\sigma_2\Theta_5(x)}=\begin{pmatrix}c(x) & s(x)\\ -s(x) & c(x)\end{pmatrix}, \tag{3.1}$$
$$c(x)=e^{i\Theta(x)}\cos\Theta_5(x),\quad s(x)=e^{i\Theta(x)}\sin\Theta_5(x).$$

Starting from a state ω_s defined in Eq. (1.10) by the projection P_+, one goes over, with the help of the transformation (3.1), to a new state defined by $V_{\underline{\Theta}}P_+V_{\underline{\Theta}}^{-1}$. We shall require that this transformation can be implemented unitarily in the Hilbert space $\mathscr{H}_{\omega_s}$ so that there exists $\Gamma(V_{\underline{\Theta}})$ with

$$\Pi_{\omega_s}(a(V_{\underline{\Theta}}f))=\Gamma(V_{\underline{\Theta}})\Pi_{\omega_s}(a(f))\Gamma(V_{\underline{\Theta}})^{-1}. \tag{3.2}$$

The two states lead to equivalent representations iff

$$X_{\underline{\Theta}}=V_{\underline{\Theta}}P_+V_{\underline{\Theta}}^{-1}-P_+\in\mathscr{B}_2(\mathscr{H}), \tag{3.3}$$

where $\mathscr{B}_2(\mathscr{H})$ denotes the class of Hilbert–Schmidt operators over $\mathscr{H}$. Note that $X_{\underline{\Theta}}$ obeys a cocycle condition

$$X_{\underline{\Theta}\cdot\underline{\chi}}=V_{\underline{\Theta}}X_{\underline{\chi}}V_{\underline{\Theta}}^{-1}+X_{\underline{\Theta}}, \tag{3.4}$$

and a Hilbert–Schmidt operator valued cohomology can be associated with the above problem.[11]

In order to figure out conditions on $\Theta(x)$ and $\Theta_5(x)$ such that (3.3) is fulfilled it is convenient to replace this condition by the equivalent ones

$$P_+V_{\underline{\Theta}}P_-\in\mathscr{B}_2\mathscr{H}\quad\text{and}\quad P_-V_{\underline{\Theta}}P_+\in\mathscr{B}_2(\mathscr{H}); \tag{3.5}$$

the equivalence can be easily seen from the identities

$$\begin{aligned}&\|P_\pm-VP_\pm V^{-1}\|_{\mathrm{HS}}\\&\quad=\mathrm{Tr}(VP_\mp V^{-1}P_\pm+P_\mp VP_\pm V^{-1}),\\&\|P_\pm VP_\mp\|_{\mathrm{HS}}=\mathrm{Tr}\,P_\pm VP_\mp V^{-1}P_\pm\end{aligned} \tag{3.6}$$

which holds for any orthogonal projection P_- (with $P_+=1-P_-$) and unitary V.

In order to check (3.5), an explicit representation of the projection operators is useful; from (1.4) we get immediately

$$P_+=\frac{1}{2}\begin{pmatrix}1 & \pm\Pi\\ \pm\Pi^* & 1-\Pi_s\end{pmatrix},\quad P_s=\begin{pmatrix}0 & 0\\ 0 & \pi_s\end{pmatrix}, \tag{3.7}$$

where Π and Π_s have the kernel representation

$$\begin{aligned}&\Pi(x,y)=\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ik(x-y)}\frac{(-ik+\tanh y)}{\sqrt{k^2+1}},\\&\Pi_s(x,y)=\phi_s(x)\phi_s(y).\end{aligned} \tag{3.8}$$

A simple calculation yields the matrix elements of $4P_+V_{\underline{\Theta}}P_-=M$:

$$\begin{aligned}&M_{11}=C_1-S_1,\quad M_{12}=C_2+S_2-s\cdot\Pi_s-\Pi\cdot c\cdot\Pi_s,\\&M_{21}=-C_3-S_3+\Pi_s\cdot s+\Pi_s\cdot c\cdot\Pi^*,\\&M_{22}=C_4+S_4-\Pi^*\cdot s\cdot\Pi_s\\&\qquad-\Pi_s\cdot s\cdot\Pi-\Pi_s\cdot c-c\cdot\Pi_s+\Pi_s\cdot c\cdot\Pi_s,\end{aligned} \tag{3.9}$$

where C_i and S_i are given by

$$\begin{aligned}&C_1=c-\Pi\cdot c\cdot\Pi^*,\quad C_2=\Pi\cdot c-c\cdot\Pi,\\&C_3=c\cdot\Pi^*-\Pi^*\cdot c,\quad C_4=c-\Pi^*\cdot c\cdot\Pi,\\&S_1=\Pi\cdot s+s\cdot\Pi^*,\quad S_2=s+\Pi\cdot s\cdot\Pi,\\&S_3=s+\Pi^*\cdot s\cdot\Pi^*,\quad S_4+s\cdot\Pi+\Pi^*\cdot s.\end{aligned} \tag{3.10}$$

Since a matrix operator is Hilbert–Schmidt iff all matrix elements have this property, we require that $M_{ij}\in\mathscr{B}_2(\tilde{\mathscr{H}})$ for $i,j=1,2$, with $\tilde{\mathscr{H}}=L^2(\mathbb{R},dx)$. It turns out that $C_i\in\mathscr{B}_2(\tilde{\mathscr{H}})$ without putting any restrictions onto $c(x)$, but $S_i\in\mathscr{B}_2(\tilde{\mathscr{H}})$ iff $s(x)$ fulfills special boundary conditions. This we formulate as the following theorem.

Theorem: The gauge transformation $V_{\underline{\Theta}}$ with $\underline{\Theta}=(\Theta(x),\Theta_5(x))$, $\Theta,\Theta_5\in C^\infty$, and $\Theta',\Theta_5'\in C_0^\infty$ is unitarily implementable in the representation $(\mathscr{H}_{\omega_s},\Pi_{\omega_s},\Omega_{\omega_s})$ iff $\lim_{x\to-\infty}\Theta_5(x)=m\pi$ and $\lim_{x\to+\infty}\Theta_5(x)=n\pi$ with $n,m\in Z$.

Proof: We remark first that Π_s is a one-dimensional projection operator; all terms of M_{ij} where Π_s enters have, therefore, finite Hilbert–Schmidt (HS) norm and are irrelevant.

Next we remark that finiteness of $\|C_1\|_{\mathrm{HS}}$ implies finiteness of $\|C_i\|_{\mathrm{HS}}$ for $i=2,3,4$; this is seen by writing out the

Hilbert–Schmidt norms, using cyclicity of the trace, the properties that $\Pi\Pi^* = 1$ and $\Pi^*\Pi = 1 - \Pi_s$ and observing that terms involving Π_s are finite.

It is therefore enough to require that $\|S_1\|_{\mathrm{HS}} < \infty$ and $\|C_1\|_{\mathrm{HS}} < \infty$.

(i) From (3.8) we first obtain the kernel of S_1 in coordinate space:

$$S_1(x,y) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{e^{ik(x-y)}}{\sqrt{k^2+1}} \{(-ik + \tanh y)s(y) + (ik + \tanh x)s(x)\}, \tag{3.11}$$

which transformed to momentum space may be split into two parts

$$\begin{aligned} \hat{S}_1(p,q) &= U_s(p,q) + W_s(p,q), \\ U_s(p,q) &= \hat{s}'(p-q)\cdot\kappa_u(p,q), \\ W_s(p,q) &= \hat{t}_s(p-q)\kappa_w(p,q), \end{aligned} \tag{3.12}$$

where $\hat{s}'$ denotes the Fourier transform of $s'(x)$ and $\hat{t}_s$ the Fourier transform of $\tanh x \cdot s(x)$. The kernels κ_u and κ_w are explicitly given by

$$\begin{aligned} \kappa_u(p,q) &= \left(\frac{p}{\sqrt{p^2+1}} - \frac{q}{\sqrt{q^2+1}}\right)\frac{1}{p-q}, \\ \kappa_w(p,q) &= \left(\frac{1}{\sqrt{p^2+1}} + \frac{1}{\sqrt{q^2+1}}\right). \end{aligned} \tag{3.13}$$

Note that we have done a partial integration in the first contribution to $\hat{S}_1$; $\kappa_u(p,q)$ is defined by continuity at $p = q$. We next need an estimate on the kernel; since $p/\sqrt{p^2+1}$ is monotonous in p we get

$$|\kappa_u(p,q)|^2 \leqslant [1/(q^2+1)] + [1/(p^2+1)]. \tag{3.14}$$

Since $\hat{s}'$ is $\in S(\mathbb{R})$ we observe that the HS norm of U_s is finite, we get [$S(\mathbb{R})$ being the Schwartz space]

$$\|U_s\|_{\mathrm{HS}} \leqslant \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq |\hat{s}'(p-q)|^2 \times \left(\frac{1}{q^2+1} + \frac{1}{p^2+1}\right) < \infty. \tag{3.15}$$

As for the second contribution from $W_s(p,q)$, we first note that

$$\frac{1}{p^2+1} \leqslant |\kappa_w(p,q)|^2 \leqslant 2\left(\frac{1}{p^2+1} + \frac{1}{q^2+1}\right), \tag{3.16}$$

from which we deduce the bounds

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{dp}{p^2+1} \int_{-\infty}^{\infty} dq |\hat{t}_s(p-q)|^2 \\ &\quad \leqslant \|W_s\|_{\mathrm{HS}} \leqslant 2\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq |\hat{t}_s(p-q)|^2 \\ &\qquad \times \left(\frac{1}{p^2+1} + \frac{1}{q^2+1}\right). \end{aligned} \tag{3.17}$$

From (3.17) we find that the necessary and sufficient condition for $W_s \in \mathscr{B}_2(\tilde{\mathscr{H}})$ is given by $\hat{t}(p) \in L^2(\mathbb{R}, dp)$ or equivalently $t(x) \in L^2(\mathbb{R}, dx)$. Therefore $\sin^2 \Theta_5(x)$ has to be integrable for large $|x|$ or, since $\Theta_5' \in C_0^{\infty}(\mathbb{R})$ we get

$$\int_C^{+\infty} dx \sin^2 \Theta_5(x) < \infty \Leftrightarrow \lim_{x \to \infty} \Theta_5 = n\pi$$

and

$$\lim_{x \to -\infty} \Theta_5(x) = m\pi, \quad n,m \in Z. \tag{3.18}$$

(ii) Using (3.8) again, we may first write the kernel of C_1 in coordinate space,

$$\begin{aligned} C_1(x,y) = c(x)\delta(x-y) &- \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{e^{ik(x-z)}}{\sqrt{k^2+1}} \\ &\times (-ik + \tanh z)c(z) \\ &\times \int_{-\infty}^{\infty} \frac{dl}{2\pi} \frac{e^{-il(z-y)}}{\sqrt{l^2+1}} (il + \tanh z), \end{aligned} \tag{3.19}$$

and transform to momentum space

$$\begin{aligned} \hat{C}_1(p,q) = \hat{c}(p-q)\cdot\kappa_c(p,q) &+ \hat{t}'(p-q)\cdot\kappa_u(p,q) \\ &+ \hat{\mathrm{ch}}(p-q)\cdot\kappa_{\hat{\mathrm{ch}}}(p,q), \end{aligned} \tag{3.20}$$

where we have been splitting the $\tanh^2 z$ contribution and $\hat{c}$, $\hat{t}'$, and $\hat{\mathrm{ch}}$ denote the Fourier transform of $c(x)$, $(d/dx)(\tanh x \cdot c(x))$ and of $c(x)/\cosh^2(x)$, respectively. The two new kernels are given explicitly by

$$\begin{aligned} \kappa_c(p,q) &= 1 - \frac{pq+1}{\sqrt{p^2+1}\sqrt{q^2+1}}, \\ \kappa_{\hat{\mathrm{ch}}}(p,q) &= \frac{1}{\sqrt{p^2+1}\sqrt{q^2+1}}, \end{aligned} \tag{3.21}$$

while κ_u was already entering in (3.13). It is now simple to see that the reasoning leading to finiteness of the HS norm in (i) applies here as well. For the first term in (3.20) one does the "partial integration trick" and uses $\hat{c}' \in S(\mathbb{R})$, and the last two terms also give finite contributions. Note that we need no restrictions on the asymptotic values of $c(x)$.

IV. REMARK

In this paper we have started the investigation of an external field problem which "determines" certain quantum numbers by themselves." We observe that the gauge transformations are implementable iff the gauge functions take asymptotic values, which also give invariance of H for rigid transformations. This fact resembles the situation of the massive Dirac operator.[9] The obtained quantization condition is obviously connected to the nonconservation of the axial charge for our external field problem. It is tempting to suggest that the integer involved $(n - m)$ may correspond to the Fredholm index of the operator $P_+ V_\Theta P_+$ entering into a Bogoliubov transformation, where P_+ denotes the projector onto the positive energy part of H. This together with a study of the current algebra connected to our problem is under investigation.

ACKNOWLEDGMENTS

One of us (G.K.) would like to thank Professor W. Thirring for the hospitality extended to him in Vienna.

This paper was supported in part by "Fonds zur Förderung der wissenschaftlichen Forschung in Österreich," No. P5588.

[1]M. Klaus and G. Scharf, Helv. Phys. Acta **50**, 779 (1977).
[2]A. J. Niemi and G. W. Semenoff, Phys. Rep. **135**, 99 (1986).
[3]H. Grosse and G. Karner, "Quantization in presence of external soliton fields," Phys. Lett. **172**, 231 (1986).
[4]D. Bollé, F. Gesztesy, H. Grosse, and B. Simon, "Kreins spectral shift function and Fredholm determinants as efficient methods to study supersymmetric quantum mechanics," University of Leuven preprint, 1986.
[5]R. T. Powers and E. Størmer, Commun. Math. Phys. **16**, 1 (1970).
[6]A. L. Carey, C. A. Hurst, and D. M. O'Brien, J. Funct. Anal. **48**, 360 (1982).
[7]H. Araki and W. Wyss, Helv. Phys. Acta **37**, 136 (1964).
[8]L. E. Lundberg, Commun. Math. Phys. **50**, 103 (1976).
[9]F. Gallone, A. Sparzani, G. Ubertone, and R. F. Streater, J. Phys. A: Math. Gen. **19**, 241 (1986).
[10]A. K. Raina and G. Wanders, Ann. Phys. (NY) **132**, 404 (1981).
[11]K. Kraus and R. F. Streater, J. Phys. A: Math. Gen. **14**, 2467 (1981).
[12]A. L. Carey and C. A. Hurst, and D. M. O'Brien, J. Math. Phys. **24**, 2212 (1983).
[13]A. L. Carey and C. A. Hurst, Commun. Math. Phys. **98**, 435 (1985).
[14]A. L. Carey and S. N. M. Ruijsenaars, "On fermion gauge groups, current algebras and Kac–Moody algebras," University of Canberra preprint, 1986.
[15]P. J. M. Bongaarts, Ann. Phys. (NY) **56**, 108 (1970).

Transition amplitude spaces

Stanley Gudder
Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208

Sylvia Pulmannová
Mathematics Institute, Slovak Academy of Sciences, 814 73 Bratislava, Czechoslovakia

(Received 9 June 1986; accepted for publication 1 October 1986)

Motivated by physically plausible axioms, the concept of a transition amplitude space (tas) is defined. Various connections between this framework and other axiomatic approaches to the foundations of quantum mechanics are derived. In particular, transition probability spaces, the algebraic approach, the operational statistics and quantum logic approaches, and traditional Hilbert space quantum mechanics are considered. It is shown that a tas always admits a Hilbert space representation. Results are obtained concerning isomorphisms and automorphisms for tas's. The concept of an A-form is introduced and the relationship between certain A-forms and bounded linear operators on a Hilbert space are studied. Finally, it is shown that sums and tensor products of tas's can be formulated in a natural way.

I. INTRODUCTION

In the traditional Hilbert space formulation of quantum mechanics, transition amplitudes play a secondary role. Their definition and properties follow from the Hilbert space structure of the pure states of the physical system. It is our view that transition amplitudes should play the central role in an axiomatic foundation for quantum mechanics. Thus we should begin with the transition amplitude as the basic undefined axiomatic element. The properties of these elements should be delineated and the axiomatic structure should be built upon these properties. This idea is basic to the early work of Feynman[1-3] in his path integral formalism, although he was more interested in computational matters than foundational ones. Moreover, many investigators have studied axiomatic systems based on transition probabilities,[4-8] and this seems to be a step in the right direction. However, we feel that transition amplitudes are more fundamental and that transition probabilities are easily derived in terms of them. The fact that the transition probability is the modulus squared of the transition amplitude is the best way to explain interference phenomena that are characteristic of quantum systems.[1,2] If this fact is ignored, then the interference phenomena are obscured. As we shall show, once a few simple properties of transition amplitudes are given, Hilbert space representations follow quickly. This is a great advantage, since a Hilbert space structure for quantum mechanics is physically unmotivated and is only the result of fairly restrictive *ad hoc* assumptions.

At the basic level, our main problem is to find the correct properties of transition amplitudes. To this end, let S be the set of pure states for a quantum system. If A is to be a transition amplitude, then A should be a map from $S \times S$ into $\mathbb{C}$, where we interpret $A(x,y)$ as the transition amplitude from state x to state y. To understand what this means, it is important to distinguish between two types of transition amplitudes, a static transition amplitude A_0 and a dynamic transition amplitude A_t, $t>0$. Dynamic transition amplitudes are physically more transparent so we begin with them. It is clear that $A_t(x,y)$ should be interpreted as follows. If the system is initially in the state x, then the amplitude that the system is in the state y at time $t>0$ is $A_t(x,y)$. We would then define the probability $P_t(x,y)$ of a transition from x to y in time t to be $|A_t(x,y)|^2$.

There are strong physical reasons that our system acts like a Markov process at the amplitude level. If the system is in the state x at time t_0, then future states for $t>t_0$ depend only on x and not on states prior to t_0. In other words, quantum state evolutions do not appear to possess memory. An important property of Markov processes is that they satisfy the Chapman–Kolmogorov equation. In our framework this may be stated as follows. There exists a set M of intermediate states such that

$$A_{t_1+t_2}(x,y) = \sum_{z\in M} A_{t_1}(x,z)A_{t_2}(z,y)\,, \tag{1.1}$$

for all $t_1,t_2>0$ and for every $x,y\in S$. As far as we are concerned, (1.1) is the most important basic property of A_t.

Now we may view the static transition amplitude $A_0(x,y)$ as the limit $A_0(x,y) = \lim_{t\to 0} A_t(x,y)$. Taking the limit of (1.1) gives

$$A_0(x,y) = \sum_{z\in M} A_0(x,z)A_0(z,y)\,, \tag{1.2}$$

for all $x,y\in S$. It is convenient and useful to extend the process A_t, $t>0$, to negative t. This is done by defining $A_{-t}(x,y) = \overline{A}_t(y,x)$, $t>0$, where $\overline{A}_t$ denotes the complex conjugate of A_t. Thus if the initial state is x and $t>0$, we interpret $\overline{A}_t(y,x)$ to be the amplitude that the state was y at t units of time previously. If we let $t\to 0$, we obtain $A_0(x,y) = \overline{A}_0(y,x)$ and we can write (1.2) in the form

$$A_0(x,y) = \sum_{z\in M} A_0(x,z)\overline{A}_0(y,z)\,, \tag{1.3}$$

for all $x,y\in S$. Equation (1.3) will be used as the main defining relation for A_0 in Sec. II. The other defining relation will be

$$A_0(x,x) = 1\,. \tag{1.4}$$

This has a natural physical interpretation. We could also derive (1.4) from the plausible assumption that a state x transforms into the intermediate states in M with certainty.

 0022-2488/87/020376-10$02.50 © 1987 American Institute of Physics

Then (1.3) gives

$$1 = \sum_{z\in M} P_0(x,z) = \sum_{z\in M} A_0(x,z)\overline{A}_0(x,z) = A_0(x,x) .$$

One can give other interpretations for A_0, and these give alternative ways of justifying (1.2) [and subsequently (1.3)]. We can view A_0 as a transmission amplitude. In this viewpoint,[5,6] we represent the physical system as a particle. In fact, since we are making a statistical analysis, it is better to represent the system as an ensemble or beam of noninteracting particles. A pure state x can be thought of as a filter which transmits only particles with certain specified properties. We call this an x filter. If an x filter is placed in the path of the particle beam, then only particles in the x state are transmitted. Suppose we now place an x filter preceding a y filter in the beam path. Then $A_0(x,y)$ is interpreted as the amplitude of transmission for the beam through the y filter given that it is transmitted through the x filter. The transmission probability $P_0(x,y) = |A_0(x,y)|^2$ is the ratio of the number of particles that pass through the y filter to the number of particles that pass through the x filter in the two-filter experiment.

Now suppose we place a z filter between the x and y filters. Let $A_0(x,z,y)$ be the transmission amplitude for this set of three filters and let $P_0(x,z,y) = |A_0(x,z,y)|^2$ be the corresponding transmission probability. It is then clear that

$$P_0(x,z,y) = P_0(x,z)P_0(z,y) . \tag{1.5}$$

However, at a deeper level we propose that

$$A_0(x,z,y) = A_0(x,z)A_0(z,y) . \tag{1.6}$$

Property (1.6) has been emphasized by Feynman and others[2,3] and is called the product rule for transition amplitudes. Of course, (1.5) easily follows from (1.6). We next propose the existence of a complete set M of z filters. Such a set M has the property that an individual particle of the beam is transmitted by precisely one of the filters in M. That is, M is a selection process that classifies particles into distinct categories. There may be many such selection processes but the important point is that there exists at least one. It is now natural to assume that $A_0(x,y)$ is the sum of the transmission amplitudes over the various particle categories in M. That is,

$$A_0(x,y) = \sum_{z\in M} A_0(x,z,y) . \tag{1.7}$$

Then (1.2) follows from (1.6) and (1.7).

It should be mentioned that (1.2) does not hold for transition probabilities. For these, only the much weaker relation $\Sigma_{z\in M} P_0(x,z) = 1$, holds. Because of this less restrictive condition, there are pathological examples of transition probabilities that do not come from transition amplitudes. Although some of these examples may have physical significance,[5,6] the situation is not entirely clear. In any case, (1.2) is not only physically plausible, it has strong mathematical consequences.

Still another interpretation is that $A_0(x,y)$ gives an amplitude for the information that y has in common with x (see Refs. 4 and 9). An analysis similar to that given previously can be used to justify (1.2) in this situation.

In the Hilbert space formulation of quantum mechanics, (1.1)–(1.4) follow from the Hilbert space structure. In this case, a pure state is represented by a unit vector x in a complex Hilbert space H. We then define $A_0(x,y)$ to be the inner product $\langle x,y\rangle$. Then (1.4) is just the relation $\|x\| = 1$. If M is an orthonormal basis for H, then (1.2) [or (1.3)] is Parseval's equality. The time evolution for pure states is given by a one-parameter unitary group U_t and the dynamic transition amplitude is defined as $A_t(x,y) = \langle U_t\, x,y\rangle$. Then (1.1) follows from

$$\begin{aligned} A_{t_1+t_2}(x,y) = \langle U_{t_1+t_2}\, x,y\rangle &= \langle U_{t_1}\, x, U_{t_2}^*\, y\rangle \\ &= \sum_{z\in M} \langle U_{t_1}\, x,z\rangle\langle z, U_{t_2}^*\, y\rangle \\ &= \sum_{z\in M} \langle U_{t_1}\, x,z\rangle\langle U_{t_2}\, z,y\rangle \\ &= \sum_{z\in M} A_{t_1}(x,z)A_{t_2}(z,y) . \end{aligned}$$

II. DEFINITIONS AND EXAMPLES

Let S be a nonempty set and let $A: S\times S\to\mathbb{C}$. We say that $x,y\in S$ are *orthogonal* ($x\perp y$) if $x\neq y$ and $A(x,y) = 0$. It follows from Zorn's lemma that the collection of maximal orthogonal sets $\mathcal{M}_A$ covers S; that is, $S = \cup\mathcal{M}_A$. We call a set $M\subseteq S$ an *A-set* if for every $x,y\in S$, we have $\Sigma_{z\in M}|A(x,z)\times\overline{A}(y,z)| < \infty$ and

$$A(x,y) = \sum_{z\in M} A(x,z)\overline{A}(y,z) . \tag{2.1}$$

Denote the collection of A-sets by $\mathcal{N}_A$. We call $A: S\times S\to\mathbb{C}$ a *transition amplitude* if (i) $\mathcal{N}_A\neq 0$, and (ii) $A(x,x) = 1$ for all $x\in S$. Property (ii) is a mild normalization condition since it follows from (i) and (2.1) that $A(x,x)\geqslant 0$. Notice that if A is a transition amplitude on S, then from (2.1) we have $\overline{A}(x,y) = A(y,x)$ for all $x,y\in S$. If A is a transition amplitude on S we call (S,A) a *transition amplitude space* (tas). A *strong* (*ultrastrong*) tas is a tas (S,A) which satisfies (iii) $A(x,y) = 1$ [(iv) $|A(x,y)| = 1$] implies $x = y$. Of course, an ultrastrong tas is strong. A tas (S,A) is *total* if $\mathcal{M}_A = \mathcal{N}_A$. The following lemma shows that (S,A) is total if $\mathcal{M}_A\subseteq\mathcal{N}_A$.

Lemma 2.1: If (S,A) is a tas, then $\mathcal{N}_A\subseteq\mathcal{M}_A$.

Proof: Let $M\in\mathcal{N}_A$ and suppose $x, y\in M$ with $x\neq y$. Denoting $M\setminus\{x\}$ by M', we have

$$1 = A(x,x) = \sum_{z\in M} |A(x,z)|^2 = 1 + \sum_{z\in M'} |A(x,z)|^2 .$$

Hence $A(x,y) = 0$ so $x\perp y$. To show that $M\in\mathcal{M}_A$ suppose $x\perp M$. Since $A(x,z) = 0$ for all $z\in M$ we have

$$1 = A(x,x) = \sum_{z\in M}, |A(x,z)|^2 = 0 .$$

This is a contradiction. □

If a tas (S,A) is not strong, we can easily construct a strong tas closely associated with (S,A). Define a relation $\approx$ on S by $x\approx y$ if $A(x,y) = 1$. Then clearly $\approx$ is reflexive and symmetric. It will follow from Lemma 3.1(b) that $\approx$ is transitive so $\approx$ is an equivalence relation. Denote the corresponding equivalence classes by $\mathring{x},x\in S$, and let $\mathring{S} = S/\approx$. Then the function $\mathring{A}: \mathring{S}\times\mathring{S}\to\mathbb{C}$ defined by $\mathring{A}(\mathring{x},\mathring{y}) = A(x,y)$ is well defined [again by Lemma 3.1(b)] and $(\mathring{S},\mathring{A})$ becomes a strong tas. We call $(\mathring{S},\mathring{A})$ the *associated* strong tas.

In a similar way, we define a relation $\sim$ on S by $x\sim y$ if $|A(x,y)| = 1$. It will follow from Corollary 3.3(a) that $\sim$ is transitive and hence is an equivalence relation. Denote the corresponding equivalence classes by $\hat{x}, x\in S$, and let $\hat{S} = S/\sim$. Unfortunately, A cannot be transferred to S in the usual way. However, if $S_1\subseteq S$ consists of one representative from each $\hat{x}, x\in S$, then (S_1,A) becomes an ultrastrong tas, which we call an *associated* ultrastrong tas.

Example 1: The unit sphere $S(H)$ of a complex Hilbert space H with $A(x,y) = \langle x,y\rangle$ is a total, strong tas. In this case we use the notation $\mathscr{N}_H = \mathscr{N}_A = \mathscr{M}_A$. Notice that $(S(H),A)$ is not ultrastrong since $|A(\alpha x,x)| = 1$ whenever $|\alpha| = 1$. An equivalence class $\hat{x}$ has the form

$$\hat{x} = \{\alpha x\colon\ \alpha\in\mathbb{C},\ |\alpha| = 1\}\ .$$

If we select a representative $x_1\in\hat{x}$ from each equivalence class and let $S_1 = \{x_1\colon\ x\in S(H)\}$, then (S_1,A) is a total, ultrastrong tas. Of course, $\mathscr{N}_H$ is the set of orthonormal bases for H.

Example 2: Let S be a nonempty set and let $A\colon S\times S\to\mathbb{C}$ be defined by $A(x,y) = 1$ if $x = y$ and $A(x,y) = 0$ if $x\neq y$. Then S itself is the only A-set. Indeed, for every $x,y\in S$ we have

$$\sum_{z\in S} A(x,z)A(z,y) = A(x,y)\ ,$$

so S is an A-set. Moreover, if M is a proper subset of S and $x\notin M$, then

$$A(x,x) = 1\neq 0 = \sum_{z\in M} A(x,z)A(z,x)\ ,$$

so $M\notin\mathscr{N}_A$. It follows that (S,A) is a total, ultrastrong tas. We call (S,A) a *trivial* tas.

Example 3: Let H be a complex Hilbert space and let $S = \{x\in H\colon\ x\neq 0\}$. Define $A\colon S\times S\to\mathbb{C}$ by $A(x,y) = \langle x,y\rangle/\|x\|\ \|y\|$. Then $A(x,x) = 1$ and for any $M\in\mathscr{M}_A$ we have

$$\sum_{z\in M} A(x,z)A(z,y) = \sum_{z\in M}\frac{\langle x,z\rangle}{\|x\|\ \|z\|}\frac{\langle z,y\rangle}{\|z\|\ \|y\|} = \frac{\langle x,y\rangle}{\|x\|\ \|y\|} = A(x,y)\ ,$$

for every $x,y\in S$. Hence (S,A) is a total tas. However, (S,A) is not strong since $A(x,y) = 1$ if and only if $x/\|x\| = y/\|y\|$ so, for example, $A(x,2x) = 1$.

Example 4: Let $A(x,y) = \langle x,y\rangle$ be the usual inner product on $\mathbb{C}^2$. Let $x = (1,0)$, $y = (0,1)$, $z = (1,1)/\sqrt{2}$, be elements of $\mathbb{C}^2$ and let $S = \{x,y,z\}$. Then (S,A) is an ultrastrong tas with $\mathscr{N}_A = \{\{x,y\}\}$. Moreover, $\{z\}\in\mathscr{M}_A$ but $\{z\}\notin\mathscr{N}_A$ so (S,A) is not total.

We now consider the relationship between a tas and a transition probability space.[5] Let S be a nonempty set and let $T\colon S\times S\to[0,1]$. We call (S,T) a *transition probability space* if (i) $T(x,y) = 1$ if and only if $x = y$, (ii) $T(x,y) = T(y,x)$ for every $x,y\in S$, (iii) for any $x\in S$ and $M\in\mathscr{M}_T$ we have $\Sigma_{y\in M}\ T(x,y) = 1$.

Theorem 2.2: Let (S,A) be a total tas. Then the pair $(\hat{S},T)$ where $T(\hat{x},\hat{y}) = |A(x,y)|^2$ is a transition probability space.

Proof: If $\hat{x} = \hat{y}$, then we shall show in Corollary 3.3(a) that $A(x,z) = A(x,y)A(y,z)$ for all $z\in S$. It follows that $|A(x,z)|^2 = |A(y,z)|^2$ for all $z\in S$. This implies that $T\colon\hat{S}\times\hat{S}\to\mathbb{R}$ is well defined. It will also be shown in Corollary 3.3(a) that $|A(x,y)|\leqslant 1$ for all $x,y\in S$ so $T\colon\hat{S}\times\hat{S}\to[0,1]$. Now conditions (i) and (ii) clearly hold. Since $A(x,y) = 0$ if and only if $T(\hat{x},\hat{y}) = 0$ we have $\mathscr{M}_T = \hat{\mathscr{M}}_S$. Since (S,A) is total, for any $\hat{M}\in\mathscr{M}_T$ and $\hat{x}\in\hat{S}$ we have

$$\sum_{y\in M} T(\hat{x},\hat{y}) = \sum_{y\in M}|A(x,y)|^2 = A(x,x) = 1\ .$$

Hence condition (iii) holds. □

III. REPRESENTATIONS

A *representation* of a tas (S,A) into a Hilbert space H is a map $\phi\colon S\to H$ such that $A(x,y) = \langle\phi(x),\phi(y)\rangle$ for all $x,y\in S$ and $\phi(M)\in\mathscr{N}_H$ for some $M\in\mathscr{N}_A$. Notice that if ϕ is a representation, then $\|\phi(x)\| = 1$ for all $x\in S$ and $x\perp y$ if and only if $\phi(x)\perp\phi(y)$. It follows that ϕ is injective on sets of mutually orthogonal elements. However, ϕ need not be injective on S. For instance, in Example 3 if we define $\phi\colon S\to H$ by $\phi(x) = x/\|x\|$, then ϕ is a representation that is not injective.

Lemma 3.1: Let (S,A) be a tas.

(a) If $\phi\colon S\to H$ is a representation, then $\phi(M)\in\mathscr{N}_H$ for all $M\in\mathscr{N}_A$. Conversely, if $M\subseteq S$ satisfies $\phi|M$ is injective and $\phi(M)\in\mathscr{N}_H$, then $M\in\mathscr{N}_A$.

(b) The following statements are equivalent.

(i) There exists an injective representation $\phi\colon S\to H$.

(ii) (S,A) is strong.

(iii) $A(x,z) = A(y,z)$ for every $z\in S$ implies $x = y$.

(iv) Every representation of (S,A) is injective.

Proof: (a) Since ϕ is a representation, there is a $M_1\in\mathscr{N}_A$ such that $\phi(M_1)\in\mathscr{N}_H$. Now let $M\in\mathscr{N}_A$. By Lemma 2.1, $\phi(M)$ is an orthonormal set in H. Moreover, for any $x\in M_1$ we have

$$\|\phi(x)\|^2 = A(x,x) = \sum_{z\in M}|A(x,z)|^2 = \sum_{z\in M}|\langle\phi(x),\phi(z)\rangle|^2\ .$$

It follows that

$$\phi(x) = \sum_{z\in M}\langle\phi(x),\phi(z)\rangle\,\phi(z)\ ,$$

for all $x\in M_1$. For $f\in H$, if $f\perp\phi(z)$ for every $z\in M$, then $f\perp\phi(x)$ for every $x\in M_1$. Since $\phi(M_1)\in\mathscr{N}_H$, $f = 0$. Thus $\phi(M)$ is a maximal orthonormal set so $\phi(M)\in\mathscr{N}_H$. Conversely, suppose $\phi|M$ is injective and $\phi(M)\in\mathscr{N}_H$. Then for any $x,y\in S$ we have

$$A(x,y) = \langle\phi(x),\phi(y)\rangle = \sum_{z\in M}\langle\phi(x),\phi(z)\rangle\langle\phi(z),\phi(y)\rangle = \sum_{z\in M}A(x,z)A(z,y)\ .$$

Hence $M\in\mathscr{N}_A$.

(b) (i) $\Rightarrow$ (ii) Suppose (i) holds and $A(x,y) = 1$. Since $\langle\phi(x),\phi(y)\rangle = \|\phi(x)\|\ \|\phi(y)\|$ it follows from Schwarz's inequality that $\phi(x) = \alpha\phi(y)$ for some $\alpha\in\mathbb{C}$. But then

$$\alpha = \alpha\langle \phi(y), \phi(y)\rangle = \langle \phi(x), \phi(y)\rangle = 1 ,$$

so $\phi(x) = \phi(y)$. Since ϕ is injective $x = y$.

(ii) $\Rightarrow$ (iii). Suppose (ii) holds and $A(x,z) = A(y,z)$ for every $z \in S$. Then $A(x,y) = A(y,y) = 1$ so $x = y$.

(iii) $\Rightarrow$ (iv). Suppose (iii) holds and $\phi: S \to H$ is a representation. If $\phi(x) = \phi(y)$, then for every $z \in S$ we have

$$A(x,z) = \langle \phi(x), \phi(z)\rangle = \langle \phi(y), \phi(z)\rangle = A(y,z) .$$

Hence $x = y$ so ϕ is injective.

(iv) $\Rightarrow$ (i) will follow from Theorem 3.2. □

If (S,A) is a tas and $x \in S$ we call the function $A_x: S \to \mathbb{C}$ defined by $A_x(y) = A(x,y)$ an *amplitude function*. Amplitude functions correspond to the "wave functions" of traditional quantum mechanics. The next theorem shows that these wave functions are elements of a Hilbert space and they may be used to construct a representation of (S,A).

Theorem 3.2: Every tas admits a representation. In particular, if (S,A) is a tas, then there exists a Hilbert space containing $\{A_x : x \in S\}$ such that $x \mapsto A_x$ is a representation.

Proof: Fix an $M \in \mathcal{N}_A$ and let H be the set of functions $f: S \to \mathbb{C}$ such that

$$\sum_{z \in M} |f(z)|^2 < \infty \tag{3.1}$$

and

$$f(x) = \sum_{z \in M} f(z)A(z,x) , \tag{3.2}$$

for all $x \in S$. It follows from (3.1) that the sum in (3.2) always converges absolutely since by Schwarz's inequality we have

$$\sum_{z \in M} |f(z)A(z,x)| \leqslant \left[\sum |f(z)|^2\right]^{1/2} \left[\sum |A(z,x)|^2\right]^{1/2} = \left[\sum |f(z)|^2\right]^{1/2} .$$

It is straightforward to show that H is a complex linear space. We define $\langle f, g\rangle = \Sigma_{z \in M} f(z)\,\bar{g}(z)$ for $f, g \in H$ and by Schwarz's inequality this sum converges absolutely. It is clear that $\langle \cdot, \cdot \rangle$ is sesquilinear, Hermitian, and positive semidefinite. If $\langle f, f\rangle = 0$, then $f(z) = 0$ for all $z \in M$. Applying (3.2), we have $f(x) = 0$ for all $x \in S$ so $f = 0$. Thus $(H, \langle \cdot, \cdot \rangle)$ is an inner product space.

To show that H is complete, let $f_n \in H$ be a Cauchy sequence. If $x \in S$, then by (3.2) we have

$$|f_i(x) - f_j(x)| = \left|\sum_{z \in M} [f_i(z) - f_j(z)]A(z,x)\right| \leqslant \sum |f_i(z) - f_j(z)|\,|A(z,x)| \leqslant \left[\sum |f_i(z) - f_j(z)|^2\right]^{1/2} = \|f_i - f_j\| .$$

Hence $f_n(x)$ is a Cauchy sequence in $\mathbb{C}$ so it converges to a number $f(x) \in \mathbb{C}$. We now show that $f \in H$. For any $\epsilon > 0$ there exists an n such that $i, j > n$ implies $\|f_i - f_j\|^2 < \epsilon$. Then for any finite $M_0 \subseteq M$ we have

$$\sum_{z \in M_0} |f_i(z) - f_j(z)|^2 \leqslant \|f_i - f_j\|^2 < \epsilon ,$$

if $i, j > n$. Hence for $j > n$ we obtain

$$\sum_{z \in M_0} |f(z) - f_j(z)|^2 = \lim_{i \to \infty} \sum_{z \in M_0} |f_i(z) - f_j(z)|^2 < \epsilon .$$

It follows that for $j > n$

$$\sum_{z \in M} |f(z) - f_j(z)|^2 < \epsilon , \tag{3.3}$$

and hence by Minkowski's inequality

$$\left[\sum_{z \in M} |f(z)|^2\right]^{1/2} \leqslant \left[\sum_{z \in M} |f(z) - f_j(z)|^2\right]^{1/2} + \left[\sum_{z \in M} |f_j(z)|^2\right]^{1/2} < \infty .$$

Thus (3.1) holds. Moreover, (3.2) holds since

$$f(x) = \lim_{i \to \infty} f_i(x) = \sum_{z \in M} \lim_{i \to \infty} f_i(z)A(z,x) = \sum_{z \in M} f(z)A(z,x) .$$

Hence $f \in H$ and $f_i \to f$ follows from (3.3).

Now $A_y \in H$ for any $y \in S$ since $\Sigma_{z \in M} |A_y(z)|^2 = 1$, and for any $x \in S$

$$A_y(x) = A(y,x) = \sum_{z \in M} A(y,z)A(z,x) = \sum_{z \in M} A_y(z)A(z,x) .$$

Define $\phi: S \to H$ by $\phi(x) = A_x$. Then

$$\langle \phi(x), \phi(y)\rangle = \sum_{z \in M} A_x(z)\bar{A}_y(z) = \sum_{z \in M} A(x,z)\bar{A}(y,z) = A(x,y) .$$

It follows from Lemma 2.1 that $\phi(M)$ is an orthonormal set in H. If $f \in H$, $z \in M$, it follows from (3.2) that

$$\langle f, \phi(z)\rangle = \left\langle \sum_{z' \in M} f(z')\,\phi(z'), \phi(z)\right\rangle = f(z) .$$

Hence

$$\|f\|^2 = \sum_{z \in M} |f(z)|^2 = \sum_{z \in M} |\langle f, \phi(z)\rangle|^2 .$$

It follows that $\phi(M) \in \mathcal{N}_H$. □

The *null set* $n(A)$ for a transition amplitude A is the set of all finite sequences (α_i, x_i), $i = 1,...,n$, $\alpha_i \in \mathbb{C}$, $x_i \in S$, such that $\Sigma\, \alpha_i A(x_i, y) = 0$ for all $y \in S$.

Corollary 3.3: If (S,A) is a tas, then the following conditions hold.

(a) $|A(x,y)| \leqslant 1$ and $|A(x,y)| = 1$ if and only if $A(y,z) = A(y,x)A(x,z)$ for all $z \in S$.

(b) For any $\alpha_1,...,\alpha_n \in \mathbb{C}$, $x_1,...,x_n \in S$ we have

$$\sum_{i,j} \alpha_i \bar{\alpha}_j A(x_i, x_j) \geqslant 0 ,$$

and equality holds if and only if $(\alpha_i, x_i)_{i=1}^n \in n(A)$.

(c) Every $M \in \mathcal{N}_A$ has the same cardinality, and if (S,A) is total, every $M \in \mathcal{M}_A$ has the same cardinality.

Proof: Let $\phi: S \to H$ be a representation.

(a) $|A(x,y)| = |\langle \phi(x), \phi(y)\rangle| \leqslant \|\phi(x)\|\,\|\phi(y)\| = 1.$

If $|A(x,y)| = 1$, then there is an $\alpha \in \mathbb{C}$ such that $\phi(x) = \alpha\phi(y)$. Now

$$1 = \langle \phi(x), \phi(x)\rangle = \alpha\langle \phi(x), \phi(y)\rangle = \alpha A(x,y) .$$

Hence for any $z \in S$

$$A(y,z) = \alpha^{-1} A(x,z) = A(x,y)A(x,z) .$$

$$\text{(b)} \sum_{i,j} \alpha_i \bar{\alpha}_j A(x_i,x_j) = \sum_{i,j} \alpha_i \bar{\alpha}_j \langle \phi(x_i), \phi(x_j) \rangle$$

$$= \left\langle \sum \alpha_i \phi(x_i), \sum \alpha_i \phi(x_i) \right\rangle$$

$$\geqslant 0 .$$

If equality holds, then for any $y \in S$ we have

$$\sum_i \alpha_i A(x_i,y) = \left\langle \sum_i \alpha_i \phi(x_i), \phi(y) \right\rangle = 0 .$$

Hence $(\alpha_i,x_i)_{i=1}^n \in n(A)$. The converse follows from the fact that $\{\phi(y): y \in S\}$ is total in H.

(c) If $M \in \mathcal{N}_A$, then $\phi(M)$ is an orthonormal basis for H and all orthonormal bases have the same cardinality. □

The next result characterizes transition amplitudes which coincide.

Corollary 3.4: If A and B are transition amplitudes on S satisfying (a) $n(B) \subseteq n(A)$, (b) $\mathcal{N}_B \cap \mathcal{N}_A \neq 0$, then $B = A$.

Proof: Let ϕ and ψ be representations of (S,A) and (S,B) on Hilbert spaces H_A and H_B, respectively. Let K_A be the linear hull of $\{\phi(x): x \in S\}$ and define the map $U: K_A \to H_B$ by

$$U\left(\sum \alpha_i \phi(x_i)\right) = \sum \alpha_i \psi(x_i) .$$

To show that U is well defined, suppose

$$\sum \alpha_i \phi(x_i) = \sum \beta_j \phi(y_j) .$$

Then for any $y \in S$, we have

$$\sum \alpha_i A(x_i,y) = \sum \alpha_i \langle \phi(x_i), \phi(y) \rangle$$

$$= \sum \beta_j \langle \phi(y_j), \phi(y) \rangle = \sum \beta_j A(y_j,y) .$$

Applying (a) we have

$$\sum \alpha_i B(x_i,y) = \sum \beta_j B(y_j,y) .$$

Hence

$$\sum \alpha_i \psi(x_i) = \sum \beta_j \psi(y_j) .$$

Applying (b) there exists an $M \in \mathcal{N}_B \cap \mathcal{N}_A$ and by Lemma 3.1(a), $\phi(M) \in \mathcal{N}_{H_A}$, $\psi(M) \in \mathcal{N}_{H_B}$. Since $U\phi(M) = \psi(M)$, U has a unique extension to a unitary transformation from H_A onto H_B. Then for any $x,y \in S$ we have

$$B(x,y) = \langle \psi(x), \psi(y) \rangle_B$$

$$= \langle U\phi(x), U\phi(y) \rangle_A = \langle \phi(x), \phi(y) \rangle_A$$

$$= A(x,y) . \qquad \square$$

The *dimension* $\dim(S,A)$ of a tas (S,A) is the cardinality of any $N \in \mathcal{N}_A$.

Theorem 3.5: Let (S,A) and (T,B) be tas's. Then $\dim(S,A) = \dim(T,B)$ if and only if (S,A) and (T,B) admit representations in the same Hilbert space.

Proof: If (S,A) and (T,B) admit representations in H, then

$$\dim(S,A) = \dim H = \dim(T,B).$$

Conversely, suppose $\dim(S,A) = \dim(T,B)$ and let $\phi: S \to H$ and $\psi: T \to K$ be representations. Since $\dim H = \dim K$, there exists a unitary transformation $U: K \to H$. Define $\psi': T \to H$ by $\psi' = U\psi$. Then ψ' is a representation of T since $\psi'(M) \in \mathcal{N}_H$ for all $M \in \mathcal{N}_B$ and for all $x,y \in T$ we have

$$B(x,y) = \langle \psi(x), \psi(y) \rangle_K = \langle U\psi(x), U\psi(y) \rangle_H$$

$$= \langle \psi'(x), \psi'(y) \rangle_H . \qquad \square$$

IV. ISOMORPHISMS AND A-FORMS

Two tas's (S,A) and (T,B) are *isomorphic* if there exists a bijection $J: S \to T$ such that $A(x,y) = B(Jx,Jy)$ for all $x,y \in S$. We then call J an *isomorphism*. An isomorphism from S to S is called an *automorphism*. Notice that if $J: S \to T$ is an isomorphism, then $J(\mathcal{N}_A) = \mathcal{N}_B$.

Theorem 4.1: Let (S,A) and (T,B) be tas's.

(a) if $J: S \to T$ is an isomorphism, then for any representations $\phi: S \to H$, $\psi: T \to K$ there exists a unique unitary transformation $U: H \to K$ such that $\psi J = U\phi$.

(b) Conversely, if (S,A), (T,B) are strong, $\phi: S \to H$, $\psi: T \to K$ are representations and there exists a unitary transformation $U: H \to K$ such that $U\phi(S) = \psi(T)$, then (S,A) and (T,B) are isomorphic.

Proof: (a) If $M \in \mathcal{N}_A$, then $\phi(M) \in \mathcal{N}_H$. Now $J(M) \in \mathcal{N}_B$ and $\psi[J(M)] \in \mathcal{N}_K$. Define the unitary transformation $U: H \to K$ by defining $U\phi(z) = \psi(Jz)$ for all $z \in M$ and extend by linearity and closure. Now for every $x \in S$, we have

$$\phi(x) = \sum_{z \in M} A(x,z) \phi(z) ,$$

and hence

$$U\phi(x) = \sum_{z \in M} A(x,z) \psi(Jz) .$$

Moreover,

$$\psi(Jx) = \sum_{z \in M} B(Jx,Jz) \psi(Jz) = \sum_{z \in M} A(x,z) \psi(Jz) .$$

Hence $U\phi = \psi J$. To show that U is unique, let $V: H \to K$ be unitary and suppose $V\phi = \psi J$. Then $U \mid \phi(S) = V \mid \phi(S)$ and in particular U and V agree on $\phi(M) \in \mathcal{N}_H$. Hence $U = V$.

(b) Define $J: S \to T$ by $Jx = \psi^{-1} U\phi(x)$. Then J is a bijection and for all $x,y \in S$ we have

$$A(x,y) = \langle \phi(x), \phi(y) \rangle_H = \langle U\phi(x), U\phi(y) \rangle_K$$

$$= B(\psi^{-1} U\phi(x), \psi^{-1} U\phi(y))$$

$$= B(Jx,Jy). \qquad \square$$

We say that two representations $\phi: S \to H$ and $\psi: S \to K$ are *unitarily equivalent* if there exists a unitary transformation $U: H \to K$ such that $\psi = U\phi$.

Corollary 4.2: Any two representations of a tas are unitarily equivalent.

Proof: Let J be the identity map I in Theorem 4.1(a). □

Corollary 4.3: Let (S,A) be a tas and let $\phi: S \to H$ be a

representation. If $J: S \to S$ is an automorphism, then there exists a unique unitary operator U on H such that $\phi J = U\phi$. Conversely, if (S,A) is strong, $U: H \to H$ is a unitary operator and $U\phi(S) = \phi(S)$, the $J = \phi^{-1}U\phi$ is an automorphism.

A *one-parameter group of automorphisms* on a tas (S,A) is a map J from $\mathbb{R}$ to the automorphism group of S such that $J(0) = I$ and $J(s+t) = J(s)\,J(t)$ for all $s, t \in \mathbb{R}$. A one-parameter group of automorphisms J is *continuous* if $t \mapsto A(J(t)x,y)$ is continuous for all $x,y \in S$.

Theorem 4.4: Let $\phi: S \to H$ be a representation of the tas (S,A). If J is a continuous one-parameter group of automorphisms on (S,A), then there exists a unique self-adjoint operator T on H such that $\phi J(t) = e^{-itT}\phi$ for all $t \in \mathbb{R}$.

Proof: By Corollary 4.3, there exist unique unitary operators $U(t)$ on H such that $\phi J(t) = U(t)\phi$ for all $t \in \mathbb{R}$. Since $U(0)\phi = \phi$ and

$$U(s+t)\,\phi = \phi J(s+t) = \phi J(s)\,J(t) = U(s)\,\phi J(t) = U(s)U(t)\,\phi\,,$$

we see that $U(0) = I$ and $U(s+t) = U(s)U(t)$. Let $M \in \mathcal{N}_A$ and hence $\phi(M) \in \mathcal{N}_H$. If $f, g \in H$, we have

$$\begin{aligned}
&\langle U(t)f,g\rangle \\
&\quad = \left\langle U(t) \sum_{z\in M} \langle f,\phi(z)\rangle\,\phi(z), \sum_{z\in M} \langle g,\phi(z)\rangle\,\phi(z)\right\rangle \\
&\quad = \sum_{z,z'} \langle f,\phi(z)\rangle\langle \phi(z'),g\rangle\langle U(t)\,\phi(z), \phi(z')\rangle \\
&\quad = \sum_{z,z'} \langle f,\phi(z)\rangle\langle \phi(z'),g\rangle\langle \phi[J(t)z], \phi(z')\rangle \\
&\quad = \sum_{z,z'} \langle f,\phi(z)\rangle\langle \phi(z'),g\rangle A(J(t)z,z')\,.
\end{aligned}$$

It follows that $t \mapsto \langle U(t)f,g\rangle$ is continuous so $U(t)$ is a weakly continuous one-parameter group of unitary operators on H. By Stone's theorem there exists a unique self-adjoint operator T on H such that $U(t) = e^{-itT}$ for all $t \in \mathbb{R}$. □

If (S,A) is a tas, a map $B: S \times S \to \mathbb{C}$ is called a *form*. A sequence $(\alpha_i,x_i) \in \mathbb{C}\times S$ is called a *null sequence* if for every $y \in S$ we have

$$\sum_i |\alpha_i\,A(x_i,y)| < \infty \quad \text{and} \quad \sum_i \alpha_i\,A(x_i,y) = 0\,. \tag{4.1}$$

We denote the set of null sequences by $n_0(A)$. Notice that $n(A) \subseteq n_0(A)$. We call a form B an *A-form* if for all $y \in S$ and $(\alpha_i,x_i) \in n_0(A)$ we have

$$\sum_i \alpha_i\,B(x_i,y) = \sum_i \alpha_i\,B(y,x_i) = 0\,. \tag{4.2}$$

Notice that (4.2) is a type of absolute continuity condition. We now prove a variant of the Radon–Nikodym theorem.

Lemma 4.5: A form B is a A-form if and only if for all $M \in \mathcal{N}_A$ and $x, y \in S$ we have

$$B(x,y) = \sum_{z\in M} B(x,z)A(z,y) = \sum_{z\in M} A(x,z)B(z,y)\,. \tag{4.3}$$

Proof: Suppose B is an A-form and let $M \in \mathcal{N}_A$, $x,y \in S$. Since $\Sigma_{z\in M}\,|A(x,z)|^2 < \infty$, $A(x,z) = 0$ except for a countable set $z_2,z_3,\ldots$. Let $x_1 = x$, $x_i = z_i$, $i = 2,3,\ldots$, and let $\alpha_1 = 1$, $a_i = -A(x,z_i)$, $i = 2,3,\ldots$. We now show that $(\alpha_i,x_i) \in n_0(A)$. Indeed,

$$\begin{aligned}
&\sum_i |\alpha_i\,A(x_i,y)| \\
&\quad = |A(x,y)| + \sum_i |A(x,z_i)A(z_i,y)| \\
&\quad \leqslant |A(x,y)| + \left[\sum |A(x,z_i)|^2\right]^{1/2}\left[\sum |A(z_i,y)|^2\right]^{1/2} \\
&\quad = |A(x,y)| + 1\,,
\end{aligned}$$

and

$$\sum \alpha_i\,A(x_i,y) = A(x,y) - \sum A(x,z_i)A(z_i,y) = 0\,.$$

It follows from (4.2) that

$$B(x,y) - \sum A(x,z_i)B(z_i,y) = 0\,.$$

A similar method gives the other equality in (4.3). Conversely, suppose (4.3) holds and $(\alpha_i,x_i) \in n_0(A)$. Then for $M \in \mathcal{N}_A$ we have

$$\begin{aligned}
\sum_i \alpha_i B(x_i,y) &= \sum_i \alpha_i \sum_{z\in M} A(x_i,z)B(z,y) \\
&= \sum_{z\in M} B(z,y) \sum_i \alpha_i\,A(x_i,z) = 0\,.
\end{aligned}$$

Again, the other equality in (4.3) is similar. □

A form B is *bounded* if there exists a $b \geqslant 0$ such that for any $M \in \mathcal{N}_A$ and any $(\alpha_i,z_i) \in \mathbb{C}\times M$, $i = 1,\ldots,n$, we have

$$\sum_{z\in M} \left|\sum_i \alpha_i\,B(z_i,z)\right|^2 \leqslant b^2 \sum_i |\alpha_i|^2\,. \tag{4.4}$$

Notice that A itself is bounded since

$$\sum_{z\in M} \left|\sum_i \alpha_i\,A(z_i,z)\right|^2 = \sum_i |\alpha_i|^2\,.$$

Theorem 4.6: Let (S,A) be a tas and let $\phi: S \to H$ be a representation. There exists a bijection $B \to \hat{B}$ from the set of bounded A-forms onto the set of bounded linear operators on H such that $B(x,y) = \langle \hat{B}\phi(x), \phi(y)\rangle$ for every $x,y \in S$.

Proof: Let B be a bounded A-form and let $M \in \mathcal{N}_A$. Let $H_0 \subseteq H$ be the linear hull of $\{\phi(z): z \in M\}$. On H_0 define

$$\hat{B}\left(\sum_{i=1}^n \alpha_i\,\phi(z_i)\right) = \sum_{z\in M}\sum_i \alpha_i B(z_i,z)\,\phi(z)\,.$$

Then $\hat{B}$ is a bounded linear operator on H_0 since

$$\begin{aligned}
\left|\left|\hat{B}\left(\sum \alpha_i\,\phi(z_i)\right)\right|\right|^2 &= \sum_{z\in M}\left|\sum \alpha_i\,B(z_i,z)\right|^2 \leqslant b^2 \sum |\alpha_i|^2 \\
&= b^2\left|\left|\sum \alpha_i\,\phi(z_i)\right|\right|^2.
\end{aligned}$$

Since H_0 is dense in H, $\hat{B}$ has a unique bounded linear extension to H, which we also denote by $\hat{B}$. If $x,y \in S$, $z' \in M$, since

$$\hat{B}\phi(z') = \sum_{z\in M} B(z',z)\,\phi(z)\,,$$

we have from Lemma 4.5,

$$\langle \hat{B}\phi(x), \phi(y)\rangle$$
$$= \left\langle \hat{B} \sum_{z'\in M} A(x,z')\,\phi(z'), \sum_{z\in M} A(y,z)\,\phi(z)\right\rangle$$
$$= \left\langle \sum_{z,z'\in M} A(x,z')B(z',z)\,\phi(z), \sum_{z\in M} A(y,z)\,\phi(z)\right\rangle$$
$$= \left\langle \sum_{z\in M} B(x,z)\,\phi(z), \sum_{z\in M} A(y,z)\,\phi(z)\right\rangle$$
$$= \sum_{z\in M} B(x,z)A(z,y) = B(x,y)\,.$$

It is clear that $B\to\hat{B}$ is injective. To show surjectivity, suppose L is a bounded linear operator on H and define a form B by $B(x,y) = \langle L\phi(x), \phi(y)\rangle$. To show that B is bounded, let $M\in\mathcal{N}_A$ and $(\alpha_i,z_i)\in\mathbb{C}\times M$, $i = 1,\ldots,n$. Then

$$\sum_{z\in M}\left|\sum \alpha_i\, B(z_i,z)\right|^2$$
$$= \sum_{z\in M}\left|\sum \alpha_i\langle L\phi(z_i), \phi(z)\rangle\right|^2$$
$$= \sum_{z\in M}\left|\left\langle L\sum \alpha_i\,\phi(z_i), \phi(z)\right\rangle\right|^2$$
$$= \left|\left| L\sum \alpha_i\,\phi(z_i)\right|\right|^2 \leqslant \|L\|^2\left|\left|\sum \alpha_i\,\phi(z_i)\right|\right|^2$$
$$= \|L\|^2\sum|\alpha_i|^2\,.$$

To show that B is an A-form, we have for $M\in\mathcal{N}_A$

$$\sum_{z\in M} B(x,z)A(z,y) = \sum_{z\in M}\langle L\phi(x), \phi(z)\rangle\langle \phi(z), \phi(y)\rangle$$
$$= \langle L\phi(x), \phi(y)\rangle = B(x,y)\,,$$

and

$$\sum_{z\in M} A(x,z)B(z,y) = \sum_{z\in M}\langle \phi(x), \phi(z)\rangle\langle L\phi(z), \phi(y)\rangle$$
$$= \sum_{z\in M}\langle \phi(x), \phi(z)\rangle\langle \phi(z), L^*\phi(y)\rangle$$
$$= \langle \phi(x), L^*\phi(y)\rangle$$
$$= \langle L\phi(x), \phi(y)\rangle = B(x,y)\,.$$

We conclude from Lemma 4.5 that B is an A-form. □

Let $B(S,A)$ be the set of bounded A-forms on the tas (S,A). It is easy to check that $B(S,A)$ is a complex linear space under pointwise addition and scalar multiplication. For $B\in B(S,A)$, define $B^*(x,y) = \overline{B}(y,x)$. Then $B^*\in(S,A)$. Moreover, for $B\in B(S,A)$ define $\|B\|$ to be the infimum of the b's in (4.4). Finally, for $B, C\in B(S,A)$, $M\in\mathcal{N}_A$, define the form BC by

$$BC(x,y) = \sum_{z\in M} B(z,y)C(x,z)\,.$$

It is not hard to show that $BC\in B(S,A)$ and is independent of the $M\in\mathcal{N}_A$. The next result is a straightforward corollary of Theorem 4.6.

Corollary 4.7: $B(S,A)$ is a C^*-algebra under the above operations with identity A, and $B\to\hat{B}$ is an isometric *-isomorphism.

A *state* on $B(S,A)$ is a positive linear functional μ: $B(S,A)\to\mathbb{R}$ such that $\mu(A) = 1$. An important class of states are given as follows. For $x\in S$, define the state μ_x by $\mu_x(B) = B(x,x)$. If $\phi: S\to H$ is a representation, we see that $\mu_x(B) = \langle\hat{B}\phi(x), \phi(x)\rangle$. This observation together with Corollary 4.7 gives a connection between the present framework and the algebraic (or C^*-algebra) approach to quantum mechanics.[10–13]

An A-form B is *automorphic* if for every $M\in\mathcal{N}_A$, $x,y\in S$, we have

$$\sum_{z\in M} B(x,z)\overline{B}(y,z) = \sum_{z\in M} B(z,y)\overline{B}(z,x) = A(x,y)\,. \tag{4.5}$$

If $J: S\to S$ is an automorphism, we define the form B_J by $B_J(x,y) = A(Jx,y)$ for all $x,y\in S$.

Lemma 4.8: (a) If $J: S\to S$ is an automorphism, then B_J is automorphic.

(b) If B is automorphic, then B is bounded.

Proof: (a) Let $\phi: S\to H$ be a representation. By Corollary 4.3 there is a unitary operator U on H such that $\phi J = U\phi$. Hence

$$B_J(x,y) = A(Jx,y)$$
$$= \langle \phi(Jx), \phi(y)\rangle = \langle U\phi(x), \phi(y)\rangle\,.$$

By Theorem 4.6, B_J is an A-form. Moreover (4.5) follows immediately.

(b) If $M\in\mathcal{N}_A$, $(\alpha_i, z_i)\in\mathbb{C}\times M$, $i = 1,\ldots,n$, then we have

$$\sum_{z\in M}\left|\sum_i \alpha_i\, B(z_i,z)\right|^2 = \sum_{z\in M}\sum_{i,j}\alpha_i\overline{\alpha}_j\, B(z_i,z)\overline{B}(z_j,z)$$
$$= \sum_{i,j}\alpha_i\overline{\alpha}_j\sum_{z\in M} B(z_i,z)\overline{B}(z_j,z)$$
$$= \sum_{i,j}\alpha_i\overline{\alpha}_j A(z_i,z_j) = \sum|\alpha_i|^2.$$ □

It follows from Lemma 4.8, that if B is automorphic, then $B\in B(S,A)$. It is easy to show that the set of automorphic forms is a group under multiplication in $B(S,A)$, which we call the *automorphic group*. If $B = B_J$ for some automorphism J we say that B is *implemented* by J.

Corollary 4.9: Let (S,A) be a tas and let $\phi: S\to H$ be a representation.

(a) The map $B\to\hat{B}$ is a group isomorphism from the automorphic group onto the group of unitary operators on H.

(b) If (S,A) is strong, an automorphic form B is implemented if and only if $\hat{B}\phi(S) = \phi(S)$.

Proof: (a) Equation (4.5) is equivalent to $BB^* = B^*B = A$. Hence $\hat{B}\hat{B}^* = \hat{B}^*\hat{B} = I$, so $\hat{B}$ is unitary. The rest is straightforward.

(b) This follows from Corollary 4.3. □

We call $E\subseteq S$ an *event* if there exists an $M\in\mathcal{N}_A$ such that $E\subseteq M$. Two events E, F are *orthogonal* if $E\cap F = \varnothing$ and $E\cup F\subseteq M$ for some $M\in\mathcal{N}_A$. If E is an event, we call the form A_E defined by

$$A_E(x,y) = \sum_{z\in E} A(x,z)A(z,y)$$

the *A transition amplitude conditioned by E*. For an event E and an $x \in S$ we call $\mu_x(E) = A_E(x,x)$ the *probability* of E in the state x. Notice that $0 \leqslant \mu_x(E) \leqslant 1$, $\mu(M) = 1$ for all $M \in \mathcal{N}_A$, and if E, F are orthogonal, then $\mu_x(E \cup F) = \mu_x(E) + \mu_x(F)$. Let $\phi: S \to H$ be a representation, let $f \in H$ be a unit vector and let P_f be the one-dimensional projection onto f. If E is an event, form the projection operator $P_E = \Sigma_{z \in E} P_{\phi(z)}$. We then have

$$A_E(x,y) = \sum_{z \in E} \langle \phi(x), \phi(z) \rangle \langle \phi(z), \phi(y) \rangle$$
$$= \langle P_E \phi(z), \phi(y) \rangle .$$

It follows from Theorem 4.6 that $A_E \in B(S,A)$ and $P_E = \hat{A}_E$. Let $\mathcal{E} = \{A_E : E \text{ is an event}\}$. For $A_E, A_F \in \mathcal{E}$, define $A_E \leqslant A_F$ if $A_E A_F = A_E$ and define $A'_E = A - A_E$. The proof of the next theorem is straightforward.

Theorem 4.10: If (S,A) is a total tas, then $(\mathcal{E}, \leqslant, ')$ is an atomistic, σ orthocomplete, orthomodular poset.

A bounded A-form B is an *A-conditional transition amplitude* if for every $M \in \mathcal{N}_A$, $x,y \in S$ we have

$$\sum_{z \in M} B(x,z)\overline{B}(y,z) = B(x,y) .$$

Denote the set of A-conditional transition amplitudes by L. Notice that $\mathcal{E} \subseteq L$. For $B_1, B_2 \in L$, define $B_1 \leqslant B_2$ if $B_1 B_2 = B_1$ and $B'_1 = A - B_1$. Again, the proof of the next result is straightforward.

Theorem 4.11: Let (S,A) be a tas and let $\phi: S \to H$ be a representation. Then L is a complete, atomic, weakly modular, orthomodular lattice, and the map $B \to \hat{B}$ is an isomorphism from L onto the lattice of all closed subspaces of H.

The last two theorems give a connection between the present framework and the operational statistics[14,15] and quantum logic approaches[9,16–21] to quantum mechanics.

V. SUMS AND TENSOR PRODUCTS

Direct sums and tensor products are two important constructions in the Hilbert space formulation of quantum mechanics. Direct sums are necessary in describing systems that have superselection rules while tensor products are used in describing combined systems. As we shall see, both of these constructions proceed quite naturally in the present framework.

Let (S_1,A_1) and (S_2,A_2) be tas's where $S_1 \cap S_2 = \emptyset$. Let $S = S_1 \cup S_2$ and define $A: S \times S \to \mathbb{C}$ by

$$A(x,y) = \begin{cases} A_i(x,y) & \text{if } x,y \in S_i, \quad i = 1,2, \\ 0, & \text{otherwise.} \end{cases}$$

We use the notation $S = S_1 \oplus S_2$, $A = A_1 \oplus A_2$, and call $(S_1 \oplus S_2, A_1 \oplus A_2)$ the *sum* of (S_1,A_1) and (S_2,A_2).

Lemma 5.1: $(S_1 \oplus S_2, A_1 \oplus A_2)$ is a tas and $M \in \mathcal{N}_{A_1 \oplus A_2}$ if and only if $M = M_1 \cup M_2$, where $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$.

Proof: Let $S = S_1 \oplus S_2$ and $A = A_1 \oplus A_2$. Clearly $A(x,x) = 1$ for all $x \in S$. Let $M = M_1 \cup M_2$, where $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$. Then

$$\sum_{z \in M} A(x,z)\overline{A}(y,z)$$
$$= \sum_{z \in M_1} A(x,z)\overline{A}(y,z) + \sum_{z \in M_2} A(x,z)\overline{A}(y,z) . \qquad (5.1)$$

If x and y are not in the same S_i, then both sums on the right side of (5.1) vanish so we obtain

$$\sum_{z \in M} A(x,z)\overline{A}(y,z) = 0 = A(x,y).$$

If both x and y are in the same S_i, then one of the sums on the right side of (5.1) vanishes and the other equals $A_i(x,y) = A(x,y)$. We conclude that A is a transition amplitude on S and that $M_1 \cup M_2 \in \mathcal{N}_A$. Finally, suppose $M \in \mathcal{N}_A$ and let $M_i = M \cap S_i$, $i = 1,2$. Then for all $x,y \in S$ we have

$$A(x,y) = \sum_{z \in M_1} A(x,z)\overline{A}(y,z) + \sum_{z \in M_2} A(x,z)\overline{A}(y,z) . \qquad (5.2)$$

If $x,y \in S_i$, $i = 1$ or 2, then we obtain from (5.2)

$$A_i(x,y) = \sum_{z \in M_i} A_i(x,z)\overline{A}_i(y,z) .$$

Hence $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$, and $M = M_1 \cup M_2$. □

Lemma 5.2: A tas (S,A) is isomorphic to a sum of two tas's if and only if there is a proper, nonempty subset S_1 of S such that $A(x,y) = 0$ whenever $x \in S_1$, $y \notin S_1$.

Proof: Let S_1 be a proper, nonempty subset of S such that $A(x,y) = 0$ whenever $x \in S_1$, $y \notin S_1$. Define $A_1: S_1 \times S_1 \to \mathbb{C}$ by $A_1 = A \mid S_1 \times S_1$, let $S_2 = S \setminus S_1$ and define $A_2: S_2 \times S_2 \to \mathbb{C}$ by $A_2 = A \mid S_2 \times S_2$. We now show that (S_1,A_1) is a tas [the proof for (S_2,A_2) being similar]. Clearly, $A_1(x,x) = 1$ for all $x \in S_1$. Let $M \in \mathcal{N}_A$ and let $M_1 = M \cap S_1$. For $x,y \in S_1$ we have

$$A_1(x,y) = A(x,y) = \sum_{z \in M} A(x,z)A(z,y)$$
$$= \sum_{z \in M_1} A(x,z)A(z,y)$$
$$= \sum_{z \in M_1} A_1(x,z)A_1(z,y) .$$

Hence $M_1 \in \mathcal{N}_{S_1}$ and (S_1,A_1) is a tas. We now show that the identity map $I: S \to S_1 \oplus S_2$ is an isomorphism from (S,A) to $(S_1 \oplus S_2, A_1 \oplus A_2)$. Indeed, if $x,y \in S_i$, $i = 1$ or 2, then

$$A(x,y) = A_i(x,y) = A_1 \oplus A_2(x,y) ,$$

and otherwise

$$A(x,y) = 0 = A_1 \oplus A_2(x,y) .$$

Conversely, suppose there is an isomorphism $J: S \to S_1 \oplus S_2$, where (S,A), (S_i,A_i), $i = 1,2$, are tas's. Then $J^{-1}(S_1)$ is a proper, nonempty subset of S. Moreover, if $x \in J^{-1}(S_1)$, $y \notin J^{-1}(S_1)$ we have

$$A(x,y) = A_1 \oplus A_2(Jx,Jy) = 0 . \qquad \square$$

If S_1 is a proper, nonempty subset of S satisfying the condition in Lemma 5.2, we call $(S_1, A \mid S_1)$ a *sub-tas* of (S,A).

Corollary 5.3: (a) $\dim(S_1 \oplus S_2, A_1 \oplus A_2) = \dim(S_1,A_1) + \dim(S_2,A_2)$.

(b) If $(S_1, A \mid S_1)$ is a sub-tas of (S,A), then $\dim(S,A) = \dim(S_1, A \mid S_1) + \dim(S \setminus S_1, A \mid (S \setminus S_1))$.

If H_1 and H_2 are Hilbert spaces, we denote the usual Hilbert space direct sum by $H_1 \oplus H_2$.

Theorem 5.4: A map $\phi: S_1 \oplus S_2 \to H$ is a representation of

$(S_1 \oplus S_2, A_1 \oplus A_2)$ if and only if there are Hilbert spaces H_1, H_2 such that $H = H_1 \oplus H_2$ and representations $\phi_i: S_i \to H_i$, $i = 1,2$ such that $\phi(x) = \phi_i(x)$ for all $x \in S_i$, $i = 1,2$.

Proof: Let $\phi: S_1 \oplus S_2 \to H$ be a representation. If H_i is the closed span of $\phi(S_i)$, $i = 1,2$, then H_i is a Hilbert space and is a closed subspace of H. If $x \in S_1$, $y \in S_2$, then

$$\langle \phi(x), \phi(y) \rangle = A(x,y) = 0 ,$$

so $H_1 \perp H_2$. If $M \in \mathcal{N}_A$, then by Lemma 5.1, $M = M_1 \cup M_2$, where $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$. Since $\phi(M)$ is an orthonormal basis for H we have for any $f \in H$,

$$f = \sum_{z \in M} \langle f, \phi(z) \rangle \phi(z)$$

$$= \sum_{z \in M_1} \langle f, \phi(z) \rangle \phi(z) + \sum_{z \in M_2} \langle f, \phi(z) \rangle \phi(z) .$$

It follows that $H = H_1 \oplus H_2$. Define $\phi_i: S_i \to H_i$ as $\phi_i = \phi | S_i$, $i = 1,2$. We now show that ϕ_i is a representation of (S_i, A_i), $i = 1,2$. Indeed, if $x, y \in S_i$, then for $i = 1$ or 2 we have

$$A_i(x,y) = A(x,y) = \langle \phi(x), \phi(y) \rangle = \langle \phi_i(x), \phi_i(y) \rangle .$$

If $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$, then by Theorem 5.1, $M_1 \cup M_2 \in \mathcal{N}_A$. Since $\phi(M_1 \cup M_2) \in \mathcal{N}_H$, it follows that $\phi(M_i) \in \mathcal{N}_{H_i}$, $i = 1,2$. It is clear that $\phi(x) = \phi_i(x)$ for all $x \in S_i$, $i = 1,2$.

Conversely, suppose $\phi_i: S_i = H_i$, $i = 1,2$, are representations and $\phi: S_1 \oplus S_2 \to H_1 \oplus H_2$ satisfies $\phi(x) = \phi_i(x)$ for all $x \in S_i$, $i = 1,2$. If $x, y \in S_i$, $i = 1$ or 2, then

$$A(x,y) = A_i(x,y) = \langle \phi_i(x), \phi_i(y) \rangle = \langle \phi(x), \phi(y) \rangle ,$$

and otherwise $A(x,y) = 0 = \langle \phi(x), \phi(y) \rangle$. If $M \in \mathcal{N}_A$, then by Lemma 5.1, $M = M_1 \cup M_2$, $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$. Since $\phi(M_i) \in \mathcal{N}_H$, $i = 1,2$, we have $\phi(M) \in \mathcal{N}_H$. □

If we define the sum of an arbitrary collection of tas's in the natural way, then it is straightforward to generalize the above theorems to this situation. Moreover, it is easy to show that $(S_1 \oplus S_2, A_1 \oplus A_2)$ is strong (ultrastrong) if and only if (S_1, A_1), (S_2, A_2) are strong (ultrastrong).

If (S_1, A_1) and (S_2, A_2) are tas's, define

$$S_1 \otimes S_2 = S_1 \times S_2 = \{(x_1, x_2): x_1 \in S_1, x_2 \in S_2\} ,$$

and define $A_1 \otimes A_2: S_1 \otimes S_2 \to \mathbb{C}$ by

$$A_1 \otimes A_2[(x_1,x_2),(y_1,y_2)] = A_1(x_1,y_1) A_2(x_2,y_2) .$$

We call $(S_1 \otimes S_2, A_1 \otimes A_2)$ the *tensor product* of (S_1, A_1) and (S_2, A_2).

Lemma 5.5: If (S_1, A_1), (S_2, A_2) are tas's then $(S_1 \otimes S_2, A_1 \otimes A_2)$ is a tas and $M \in \mathcal{N}_{A_1 \otimes A_2}$ if $M = M_1 \times M_2$, where $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$.

Proof: Let $S = S_1 \otimes S_2$ and $A = A_1 \otimes A_2$. It is clear that $A[(x_1,x_2),(x_1,x_2)] = 1$. Now let $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$, and let $M = M_1 \times M_2$. Then

$$\sum_{(z_1,z_2) \in M} A[(x_1,x_2),(z_1,z_2)] \overline{A}[(y_1,y_2),(z_1,z_2)]$$

$$= \sum_{(z_1,z_2) \in M} A_1(x_1,z_1) A_2(x_2,z_2) \overline{A}(y_1,z_1) \overline{A}(y_2,z_2)$$

$$= \sum_{z_1 \in M_1} A_1(x_1,z_1) \overline{A}(y_1,z_1) \sum_{z_2 \in M_2} A(x_2,z_2) \overline{A}(y_2,z_2)$$

$$= A_1(x_1,y_1) A_2(x_2,y_2) = A[(x_1,x_2),(y_1,y_2)] .$$

We conclude that (S,A) is a tas and $M_1 \times M_2 \in \mathcal{N}_A$. □

Corollary 5.6:

$$\dim(S_1 \otimes S_2, A_1 \otimes A_2) = \dim(S_1, A_1) \dim(S_2, A_2) .$$

We denote the usual tensor product of two Hilbert spaces H_1, H_2 by $H_1 \otimes H_2$.

Theorem 5.7: Let (S_1, A_1), (S_2, A_2) be tas's. A map $\phi: S_1 \otimes S_2 \to H$ is a representation of $(S_1 \otimes S_2, A_1 \otimes A_2)$ if and only if there are Hilbert spaces H_1, H_2 such that $H = H_1 \otimes H_2$ and representations $\phi_i: S_i \to H_i$, $i = 1,2$, such that $\phi(x_1,x_2) = \phi_1(x_1) \otimes \phi_2(x_2)$.

Proof: Let $\phi: S_1 \otimes S_2 \to H$ be a representation. Let $M_i \in \mathcal{N}_{A_i}$, $i = 1,2$, $u \in M_1$, $v \in M_2$ be fixed, let $H_1 \subseteq H$ be the closed span of $\{\phi(x_1,v): x_1 \in S_1\}$ and let $H_2 \subseteq H$ be the closed span of $\{\phi(u,x_2): x_2 \in S_2\}$. Define the maps $\phi_i: S_i \to H_i$, $i = 1,2$, by $\phi_1(x_1) = \phi(x_1,v)$, $\phi_2(x_2) = \phi(u,x_2)$. We now show that ϕ_1 is a representation (the proof for ϕ_2 is similar). For $x_1, y_1 \in S_1$ we have

$$\langle \phi_1(x_1), \phi_1(y_1) \rangle = \langle \phi(x_1,v), \phi(y_1,v) \rangle$$

$$= A_1 \otimes A_2[(x_1,v),(y_1,v)]$$

$$= A_1(x_1,y_1) .$$

If $z, z' \in M_1$, $z \neq z'$, then $\phi_1(z) \perp \phi_1(z')$ so $\{\phi_1(z): z \in M_1\}$ is an orthonormal set in H_1. Moreover, for any $x_1 \in S_1$, since $M_1 \times M_2 \in \mathcal{N}_{S_1 \otimes S_2}$, we have

$$\phi(x_1,v) = \sum_{(z,z') \in M_1 \times M_2} \langle \phi(x_1,v), \phi(z,z') \rangle \phi(z,z')$$

$$= \sum_{(z,z') \in M_1 \times M_2} A_1(x_1,z) A_2(v,z') \phi(z,z')$$

$$= \sum_{z \in M_1} A_1(x_1,z) \phi(z,v) = \sum_{z \in M_1} A_1(x_1,z) \phi_1(z) .$$

Hence if $f \in H_1$ satisfies $f \perp \phi_1(z)$ for all $z \in M_1$, we have $\langle f, \phi(x_1,v) \rangle = 0$ for all $x_1 \in S_1$. Since $\{\phi(x_1,v): x_1 \in S_1\}$ is dense in H_1, we conclude that $f = 0$, and hence $\phi_1(M_1) \in \mathcal{N}_{H_1}$. Define the unitary transformation $U: H \to H_1 \otimes H_2$ by $U\phi(z_1,z_2) = \phi_1(z_1) \otimes \phi_2(z_2)$, $z_i \in M_i$, $i = 1,2$, and extend by linearity and closure. We can thus identify H with $H_1 \otimes H_2$. Moreover, for any $(x_1,x_2) \in S_1 \otimes S_2$ we have

$$U\phi(x,x_2) = \sum_{(z,z') \in M_1 \times M_2} \langle \phi(x_1,x_2), \phi(z,z') \rangle \phi_1(z) \otimes \phi_2(z')$$

$$= \sum_{(z,z') \in M_1 \times M_2} \langle \phi_1(x_1), \phi_1(z) \rangle \times \langle \phi_2(x_2), \phi_2(z') \rangle \phi_1(z) \otimes \phi_2(z')$$

$$= \sum_{z \in M_1} \langle \phi_1(x_1), \phi_1(z) \rangle \phi_1(z) \otimes \sum_{z \in M_2} \langle \phi_2(x_2), \phi_2(z') \rangle \phi_2(z')$$

$$= \phi_1(x_1) \otimes \phi_2(x_2) .$$

Conversely, suppose $\phi_i: S_i = H_i$, $i = 1,2$, are representations and $\phi: S_1 \otimes S_2 \to H = H_1 \otimes H_2$ satisfies $\phi(x_1,x_2) = \phi_1(x_1) \otimes \phi_2(x_2)$. We now show that ϕ is a representa-

tion. For (x_1,x_2), $(y_1,y_2)\in S_1\otimes S_2$ we have

$$\begin{aligned}\langle \phi(x_1,x_2),\phi(y_1,y_2)\rangle &= \langle \phi_1(x_1)\otimes\phi_2(x_2),\phi_1(y_1)\otimes\phi_2(y_2)\rangle\\ &= \langle \phi_1(x_1),\phi_1(y_1)\rangle\langle \phi_2(x_2),\phi_2(y_2)\rangle\\ &= A_1(x_1,y_1)A_2(x_2,y_2)\\ &= A_1\otimes A_2((x_1,x_2),(y_1,y_2))\,.\end{aligned}$$

Finally, if $M_i\in\mathcal{N}_{A_i}$, $i=1,2$, then by Lemma 5.5, $M_1\times M_2\in\mathcal{N}_{A_1\otimes A_2}$. For $(x_1,x_2)\in S_1\otimes S_2$ we have

$$\begin{aligned}&\sum_{(z,z')\in M_1\times M_2}|\langle \phi(x_1,x_2),\phi(z,z')\rangle|^2\\ &\quad=\sum|A_1(x_1,z)|^2\sum|A_2(x_2,z')|^2=1=\|\phi(x_1,x_2)\|^2\,.\end{aligned}$$

Hence

$$\phi(x_1,x_2)=\sum_{(z,z')\in M_1\times M_2}\langle \phi(x_1,x_2),\phi(z,z')\rangle\,\phi(z,z')\,.$$

If $f\in H$ satisfies, $f\perp\phi(z,z')$ for all $(z,z')\in M_1\times M_2$, then $\langle f,\phi(x_1,x_2)\rangle=0$ for all $(x_1,x_2)\in S_1\otimes S_2$. Since the elements $\phi(x_1,x_2)=\phi_1(x_1)\otimes\phi_2(x_2)$, $(x_1,x_2)\in S_1\otimes S_2$ are total in H we conclude that $f=0$. Hence, $\phi(M_1\times M_2)\in\mathcal{N}_H$ and ϕ is a representation. □

If we define the tensor product of a finite number of tas's in the natural way, then it is straightforward to generalize the above theorems to this situation. Moreover, it is easy to show that $(S_1\otimes S_2,A_1\otimes A_2)$ is ultrastrong if and only if $(S_1,A_1),(S_2,A_2)$ are ultrastrong. One can give simple examples which show that the converse of Lemma 5.5 does not hold. That is, if $M\in\mathcal{N}_{A_1\otimes A_2}$, then M need not have the form $M=M_1\times M_2$ for $M_i\in\mathcal{N}_{A_i}$, $i=1,2$.

[1]R. Feynman, "The space-time approach to nonrelativistic quantum mechanics," Rev. Mod. Phys. **20**, 367 (1948).
[2]R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[3]L. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[4]V. Cantoni, "Generalized 'transition probability,'" Commun. Math. Phys. **44**, 125 (1975).
[5]B. Mielnik, "Geometry of quantum states," Commun. Math. Phys. **9**, 55 (1968).
[6]B. Mielnik, "Theory of filters," Commun. Math. Phys. **15**, 1 (1969).
[7]S. Pulmannová, "Transition probability spaces," J. Math. Phys. **27**, 1791 (1986).
[8]A. Uhlmann, "The transition probability in the state space of a *-algebra," Rep. Math. Phys. **9**, 273 (1976).
[9]S. Gudder, *Stochastic Methods in Quantum Mechanics* (Elsevier–North-Holland, New York, 1979).
[10]G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[11]R. Haag and D. Kastler, "An algebraic approach to quantum field theory," J. Math. Phys. **5**, 848 (1964).
[12]P. Jordan, J. von Neumann, and E. Wigner, "On an algebraic generalization of the quantum mechanical formalism," Ann. Math. **32**, 29 (1934).
[13]I. Segal, "Postulates for general quantum mechanics," Ann. Math. **48**, 930 (1947).
[14]D. Foulis and C. Randall, "Operational statistics I. Basic concepts," J. Math. Phys. **13**, 1167 (1972).
[15]C. Randall and D. Foulis, "Operational statistics. II. Manuals of operations and their logics," J. Math. Phys. **14**, 1472 (1972).
[16]E. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1981).
[17]G. Birkhoff and J. von Neumann, "The logic of quantum mechanics," Ann. Math. **37**, 832 (1936).
[18]J. Jauch, *Foundations of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1968).
[19]G. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[20]C. Piron, *Foundations of Quantum Mechanics* (Benjamin, New York, 1976).
[21]V. Varadarajan, *Geometry of Quantum Theory* (Springer, New York, 1985).

Quantum motion on a half-line connected to a plane

P. Exner[a] and P. Šeba[b]
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, USSR

(Received 24 April 1986; accepted for publication 17 September 1986)

In this paper, the free motion of a particle on a manifold that consists of a one-dimensional and a two-dimensional part connected in one point is discussed. The class of admissible Hamiltonians is found using the theory of self-adjoint extensions. Particular attention is paid to those Hamiltonians that allow the particle to pass through the point singularity; the reflection coefficient and other quantities characterizing scattering on the connection point are calculated. A possible application is also discussed.

I. INTRODUCTION

The theory of self-adjoint extensions represents a powerful heuristic way to construct Hamiltonians of quantum systems in cases when the correspondence principle yields only incomplete information. For instance, Hamiltonians describing the point-interaction phenomena in quantum mechanics are obtained as self-adjoint extensions of the corresponding free Hamiltonians with the interaction points removed.[1,2]

As another illustration, one can consider Schrödinger operators with singular potentials.[3,4] When the potential is singular enough, the correspondence principle provides us with a differential operator that is not essentially self-adjoint (e.s.a.). In this case, it is natural to approach the problem by constructing all self-adjoint extensions of this operator. After that one must select the appropriate one among them; it requires, of course, additional physical information. There are other quantum-mechanical problems to which the theory of self-adjoint extensions can be applied, e.g., a one-dimensional model of three-particle collisions.[5]

Particularly interesting are the situations when a quantum particle moves on a spatial manifold that consists of several more simple parts. As an example, let us recall the free-electron (or metallic) model of organic molecules in which one assumes that the π-electrons move only along the graph Γ representing the molecule (cf., e.g., Ref. 6 or Chap. 6 of Ref. 7 for the one-dimensional case). Suppose that the motion along the line Γ_j is described by the Hamiltonian

$$H_j = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_j(x)$$

with a suitably chosen domain in $L^2(\Gamma_j, dx)$ where the coordinate x parametrizes $\Gamma_j \sim [0, l_j]$. The full Hamiltonian H of the model is now obtained as an appropriate self-adjoint extension of the operator constructed by "gluing together" the line Hamiltonians H_j.

The distinct parts of such a "configuration space" are not necessarily of the same dimension. In the present paper, we are going to discuss the most simple situation of this kind, where the manifold consists of a half-line attached to a plane, i.e., the dimensions are 1 and 2, respectively. The motion in either part is assumed to be free. Notice that one may regard the sketched situation also as a motion in $\mathbb{R}^2$ subjected to a point interaction with some internal structure (compare to Ref. 8). Such an interpretation, however, does not suit to the model situation we are going to discusss, for which analysis of the motion on the half-line is essential.

Let us resume briefly contents of the following sections. First of all, we construct the class of admissible Hamiltonians as self-adjoint extensions of the operator obtained by "gluing together" the free Hamiltonians for the motion on the half-line and on the plane (Sec. II). Since the direct characterization of these extensions obtained from the von Neumann's theory is not very suitable for practical calculations, we deduce in Sec. III an alternative classification of them using singular boundary conditions. In Sec. IV we analyze scattering on the point singularity, with particular attention paid to the reflection coefficient for the particle traveling initially along the half-line. In conclusion, we discuss a possible application of the present analysis to modeling the quantum point-contact spectroscopy.

II. ADMISSIBLE HAMILTONIANS

Let us consider a particle, an electron for definiteness, moving on the manifold G that consists of two parts—the plane $\mathbb{R}^2$ and the half-line $\mathbb{R}^- = (-\infty, 0]$—which are connected at a point P as sketched in Fig. 2. The state Hilbert space of such a system is therefore the orthogonal sum of the state spaces referring to the plane and to the half-line. If we neglect the possible internal degrees of freedom (spin of the electron, for instance), we have

$$\mathscr{H} = L^2(G) := L^2(\mathbb{R}^2) \oplus L^2(\mathbb{R}^-). \tag{1}$$

Since the electron motion is supposed to be free except at the point P, we start the construction of Hamiltonian with the operator

$$H_0 = H_{0,1} \oplus H_{0,2}, \tag{2}$$

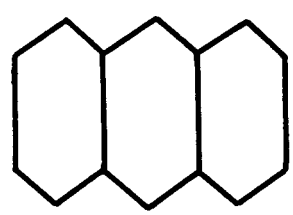


FIG. 1. The graph Γ for an anthracene molecule.

[a] On leave of absence from Nuclear Physics Institute, Czechoslovak Academy of Sciences, 25068 Řež near Prague, Czechoslovakia.
[b] On leave of absence from Nuclear Centre, Charles University, V Holešovičkách 2, 18000 Prague 8, Czechoslovakia.

 0022-2488/87/020386-06$02.50 © 1987 American Institute of Physics

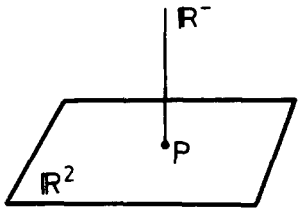


FIG. 2. The manifold G.

where the $H_{0,j}$ are restrictions of the respective free Hamiltonians on the two parts of G, namely,

$$H_{0,1} = -\frac{d^2}{dx^2}, \qquad D(H_{0,1}) = C_0^\infty(\mathbb{R}^-\setminus\{0\}), \tag{3a}$$

and

$$H_{0,2} = -\Delta, \qquad D(H_{0,2}) = C_0^\infty(\mathbb{R}^2\setminus\{P\}). \tag{3b}$$

The symbol $C_0^\infty(\Omega)$ denotes conventionally the set of all infinitely differentiable functions with a compact support contained in Ω.

The operator H_0 is not self-adjoint. It is well known[9] that the deficiency indices of $H_{0,1}$ are (1, 1), and the same is true for $H_{0,2}$ as we shall show a little later. Consequently, the deficiency indices of H_0 are (2, 2) so there is a four-parameter family of self-adjoint extensions. Let us construct them explicitly.

We use the polar coordinates with the center at P and decompose the space $L^2(\mathbb{R}^2)$ in the following way:

$$L^2(\mathbb{R}^2) = L^2(\mathbb{R}^+, r\,dr)\otimes L^2(0,2\pi) = \bigoplus_{m=-\infty}^{\infty} L^2(\mathbb{R}^+, r\,dr)\otimes\{Y_m\}_{\text{lin}}, \tag{4}$$

where the functions

$$Y_m\colon\ Y_m(\varphi) = (2\pi)^{-1/2}e^{im\varphi}, \quad \varphi\in[0,2\pi),$$

form the "trigonometric" orthonormal basis in $L^2(0,2\pi)$. Using the standard procedure (Ref. 9, appendix to Sec. X.1), one obtains the decomposition

$$H_{0,2}\restriction D_{\min} = \bigoplus_{m=-\infty}^{\infty} h_m\otimes I, \tag{5}$$

where

$$h_m = -\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{m^2}{r^2}, \qquad D(h_m) = C_0(\mathbb{R}^+\setminus\{0\}), \tag{6}$$

and the domain $D_{\min}$ consists of all finite linear combinations of the functions ψ: $\psi(r,\varphi) = f(r)Y_m(\varphi)$ with $f\in D(h_m)$. The deficiency indices of the operators h_m on $L^2(\mathbb{R}^+, r\,dr)$ are easily found. The latter are unitarily equivalent to

$$\tilde{h}_m = -\frac{d^2}{dr^2} + \frac{m^2-\frac{1}{4}}{r^2}$$

on $D(\tilde{h}_m) = D(h_m)$, so we have

$$n(h_0) = 1, \qquad n(h_m) = 0, \text{ for } m\neq 0. \tag{7}$$

The second relation follows from Theorem X.10 of Ref. 9, while the first one is verified directly (a solution to the deficiency equations will be presented below). Relations (5) and (7) yield

$$n(H_{0,2}) \leqslant \sum_{m=-\infty}^{\infty} n(h_m) = 1. \tag{8a}$$

We want to show that the equality holds in the last relation. To this end, one has to check that the functions $\psi_\pm\colon\ \psi_\pm(r,\varphi) = H_0^{(\alpha)}(\pm\sqrt{i}r)$, with α equal to 1,2 for the plus and minus sign, respectively, which span the deficiency subspaces [cf. (11) below], belong to $D(H_{0,2}^*)$. This can be performed in a straightforward manner using integration by parts and properties of the Hankel functions. Relation (2) then gives

$$n(H_0) = n(H_{0,1}) + n(H_{0,2}) = 2. \tag{8b}$$

Any self-adjoint extension of H_0 is therefore of the form

$$H = K\oplus\bar{h}, \tag{9a}$$

where K is a self-adjoint extension of the operator K_0 on $L^2(\mathbb{R}^-)\oplus(L^2(\mathbb{R}^+, r\,dr)\otimes\{Y_0\})$ defined by

$$K_0 := H_{0,1}\oplus(h_0\otimes I), \tag{9b}$$

and $\bar{h}$ denotes closure of the operator

$$h := \bigoplus_{\substack{m\in\mathbb{Z}\\ m\neq 0}} h_m, \tag{9c}$$

which is e.s.a. due to (7).

Now we must choose suitable bases in the deficiency subspaces $\mathscr{K}_\pm = \operatorname{Ker}(K_0^*\mp iI)$. It is easy to find that $\mathscr{K}_+$ is spanned by the functions

$$\varphi_1^{(+)} = (f_1, 0), \quad \varphi_2^{(+)} = (0, f_2), \tag{10}$$

$$f_1(x) := e^{\bar{\epsilon}x}, \quad \epsilon = e^{\pi i/4}, \qquad f_2(r) := (2\pi)^{1/2}H_0^{(1)}(\epsilon r). \tag{11}$$

In the same way, the functions

$$\varphi_1^{(-)} = (\bar{f}_1, 0), \quad \varphi_2^{(-)} = (0, \bar{f}_2) \tag{10'}$$

form a basis in $\mathscr{K}_-$. The self-adjoint extensions of K_0 are now specified by isometries $\mathscr{K}_+\to\mathscr{K}_-$, i.e., by 2×2 unitary matrices U. The von Neumann's theory gives a prescription how the extensions K_U can be constructed for an arbitrary U. Its domain $D(K_U)$ consists of all functions of the form

$$f = \psi + c_1(\varphi_1^{(+)} + u_{11}\varphi_1^{(-)} + u_{12}\varphi_2^{(-)}) + c_2(\varphi_2^{(+)} + u_{21}\varphi_1^{(-)} + u_{22}\varphi_2^{(-)}), \tag{12}$$

with $\psi\in D(\bar{K}_0)$ and $c_1, c_2\in\mathbb{C}$, where u_{jk} are elements of U. One might write an expression for $K_U f$ as a linear combination of $\bar{K}_0\psi$ and the deficiency functions. Instead, we are going to derive a more transparent expression for the action of K_U.

III. BOUNDARY CONDITIONS

For practical calculations, it is more convenient to characterize extensions K_U by appropriate boundary conditions. In this way, we are able to describe K_U completely since $K_U\subset K_0^*$, and it is easy to see that

$$K_0^* f = \left\{ -\frac{d^2\varphi_1}{dx^2}, -\frac{d^2\varphi_2}{dr^2} - \frac{1}{r}\frac{d\varphi_2}{dr} \right\}, \tag{13}$$

for each $f = \{\varphi_1, \varphi_2\}$ from $D(K_0^*)$. The deficiency functions $\varphi_2^{(\pm)}$ are, however, singular around P, but we can eliminate this difficulty by defining the regularized boundary values[10]

$$L_0(\varphi) := \lim_{r\to 0} \frac{\varphi(r)}{\ln r},$$

$$L_1(\varphi) := \lim_{r\to 0} [\varphi(r) - L_0(\varphi)\ln r], \tag{14}$$

which will be used together with

$$\varphi_1(0_-) := \lim_{x\to 0_-} \varphi_1(x), \quad \varphi_1'(0_-) := \lim_{x\to 0_-} \left(\frac{d}{dx}\varphi_1\right)(x).$$

In particular, the standard expansion of Hankel functions[11] yields

$$\begin{aligned} L_0(f_2) &= -L_0(\bar{f}_2) = 2i/\pi, \\ L_1(f_2) &= \tfrac{1}{2} + (2i/\pi)(\gamma - \ln 2), \\ L_1(\bar{f}_2) &= \tfrac{1}{2} - (2i/\pi)(\gamma - \ln 2), \end{aligned} \tag{15}$$

where $\gamma = 0.577\,216...$ is Euler's constant.

Before proceeding further, we shall split the set of matrices U characterizing the extensions into five disjoint classes.

Class I contains all U such that

$$1 + u_{11} - u_{22} - \det U \neq 0.$$

Class II contains all nondiagonal U such that

$$1 + u_{11} - u_{22} - \det U = 0.$$

Class III consists of the matrices

$$\begin{pmatrix} -1 & 0 \\ 0 & e^{i\omega} \end{pmatrix}$$

with $\omega \in (0, 2\pi)$.

Class IV consists of the matrices

$$\begin{pmatrix} e^{i\omega} & 0 \\ 0 & 1 \end{pmatrix}$$

with $\omega \in (-\pi, \pi)$.

Class V contains the matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now we can formulate the mentioned result.

Theorem: Every self-adjoint extension of the operator H_0 is of the form $H_U := K_U \oplus \tilde{h}$, where the operator K_U is specified uniquely by the following boundary conditions. If $f = \{\varphi_1, \varphi_2\}$ belongs to $D(K_U)$, then we have the following.

(i) For U of class I, we have

$$\begin{aligned} \varphi_1'(0_-) &= A\varphi_1(0_-) + BL_0(\varphi_2), \\ L_1(\varphi_2) &= C\varphi_1(0_-) + DL_0(\varphi_2), \end{aligned} \tag{16a}$$

where the coefficients are related to the matrix elements of U by

$$A = \frac{\bar{\epsilon}(1 - u_{22}) + \epsilon(u_{11} - \det U)}{1 + u_{11} - u_{22} - \det U}, \tag{16b}$$

$$B = \frac{\pi}{\sqrt{2}} \frac{u_{21}}{1 + u_{11} - u_{22} - \det U}. \tag{16c}$$

$$C = \frac{u_{12}}{1 + u_{11} - u_{22} - \det U}, \tag{16d}$$

$$D = \gamma - \ln 2 + \frac{\pi}{4i} \frac{1 + u_{11} + u_{22} + \det U}{1 + u_{11} - u_{22} - \det U}. \tag{16e}$$

(ii) If U belongs to class II, then

$$\begin{aligned} L_0(\varphi_2) &= E\varphi_1(0_-), \\ L_1(\varphi_2) &= F\varphi_1(0_-) + G\varphi_1'(0_-), \end{aligned} \tag{17a}$$

where

$$E = \frac{2i}{\pi} \frac{1 - u_{22}}{u_{21}} = -\frac{2i}{\pi} \frac{u_{12}}{1 + u_{11}}, \tag{17b}$$

$$F = \frac{i}{\sqrt{2}\, u_{21}} \{\bar{\epsilon} + \epsilon u_{11} + i\sqrt{2}(1 - u_{22}) L_1(\bar{f}_2)\}, \tag{17c}$$

$$G = -\frac{i}{\sqrt{2}} \frac{1 + u_{11}}{u_{21}}. \tag{17d}$$

(iii) For U of class III, the boundary conditions read

$$\begin{aligned} &\varphi_1(0_-) = 0, \\ &L_1(\varphi_2) = (\pi/4)(\cot(\omega/2)) L_0(\varphi_2). \end{aligned} \tag{18}$$

(iv) If U belongs to class IV, we have

$$\begin{aligned} &\varphi_1'(0_-) = (1/\sqrt{2})(1 - \tan(\omega/2))\varphi_1(0_-), \\ &L_0(\varphi_2) = 0. \end{aligned} \tag{19}$$

(v) Finally, if

$$U = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

then

$$\varphi_1(0_-) = L_0(\varphi_2) = 0. \tag{20}$$

Proof: Suppose first that U belongs to class I. We express φ_1, φ_2 from (12) and insert it into (16a). It yields the following equations:

$$\begin{aligned} &\bar{\epsilon} + \epsilon u_{11} = A(1 + u_{11}) + Bu_{12} L_0(\bar{f}_2), \\ &u_{12}\epsilon = Au_{21} + B[L_0(f_2) + u_{22} L_0(\bar{f}_2)], \\ &u_{12} L_1(\bar{f}_2) = C(1 + u_{11}) + Du_{12} L_0(\bar{f}_2), \\ &L_1(f_2) + u_{22} L_1(\bar{f}_2) = Cu_{21} + D[L_0(f_2) + u_{22} L_0(\bar{f}_2)]. \end{aligned}$$

Now we substitute from (15), then the solution for $1 + u_{11} - u_{22} - \det U \neq 0$ is given by (16b)–(16e). In a similar way, one obtains relations (17)–(20).

Next one has to check that the mapping from the set of matrices U to the set of boundary conditions is injective. This is easy for Classes III–V. Assume further that there are U, U' of class II, both leading to conditions (17), i.e., $E = E'$, etc. Then we have $G/E = G'/E'$ and $F/E = F'/E'$ so

$$(1 + u_{11})/(1 - u_{22}) = (1 + u'_{11})/(1 - u'_{22}),$$

and

$$(\bar{\epsilon} + \epsilon u_{11})/(1 - u_{22}) = (\bar{\epsilon} + \epsilon u'_{11})/(1 - u'_{22}).$$

Now one has to multiply the second equation by $\bar{\epsilon}$ and subtract it from the first one. It yields $u_{22} = u'_{22}$. Substituting this back to the first equation, we get $u_{11} = u'_{11}$. Finally, the relations $u_{jk} = u'_{jk}$ for $jk = 12, 21$ follow from (17b). One has to notice that $|u_{jj}| < 1$ since U is unitary and nondiagonal.

The argument is most complicated for class I. Assume again that there are U, U' which yield the same values of the coefficients (16b)–(16e). Then we have

$$\frac{u_{jk}}{1+u_{11}-u_{22}-\det U}=\frac{u'_{jk}}{1+u'_{11}-u'_{22}-\det U'}, \tag{21}$$

for $jk = 12$ and 21. Moreover, the relations (16e) and (16b) after multiplication by ϵ give

$$\frac{1+\epsilon^{1\pm 1}u_{11}\pm u_{22}+\epsilon^{-1\pm 1}\det U}{1+u_{11}-u_{22}-\det U}=\frac{1+\epsilon^{1\pm 1}u'_{11}\pm u'_{22}+\epsilon^{-1\pm 1}\det U'}{1+u'_{11}-u'_{22}-\det U'}.$$

It further imples

$$\begin{aligned}1+u_{11}&=\alpha(1+u'_{11}),\\ 1-u_{22}&=\alpha(1-u'_{22}),\end{aligned} \tag{22}$$

where we have denoted

$$\alpha=\frac{1+u_{11}-u_{22}-\det U}{1+u'_{11}-u'_{22}-\det U'}. \tag{23}$$

Relations (21) can be similarly rewritten as

$$u_{12}=\alpha u'_{12},\ u_{21}=\alpha u'_{21}. \tag{22'}$$

Hence

$$(1+u_{11})(1-u_{22})+u_{12}u_{21}=\alpha^2[(1+u'_{11})(1-u'_{22})+u'_{12}u'_{21}],$$

and combining this relation with (23), we get $\alpha^2=\alpha$. Since α is nonzero by assumption, we obtain $\alpha=1$. Then $U=U'$ follows from (22). ■

In this way, we have been able to characterize the operators H_U by means of the boundary conditions. The relations (16b)–(16e) and (17b)–(17d) do not show explicitly which values the coefficients may assume. It becomes more clear if one uses a suitable parametrization of the matrix U, such as (29) below.

IV. SCATTERING ON THE SINGULARITY

Now we are going to discuss the extensions H_U, with a particular attention paid to scattering on the singular point P. We shall distinguish two cases.

(a) U diagonal: Going through the boundary conditions (16), (18)–(20), one finds easily that they separate. We can express them in the form

$$\varphi'_1(0_-)=A\varphi_1(0_-), \tag{24a}$$

$$L_1(\varphi_2)=DL_0(\varphi_2), \tag{24b}$$

where the coefficients are with the usual license written as

$$\text{class I: } A=\frac{\overline{\epsilon}+\epsilon u_{11}}{1+u_{11}},$$

$$D=\gamma-\ln 2+\frac{\pi}{4i}\frac{1+u_{22}}{1-u_{22}}.$$

[Notice that both A and D are real; this remains true for nondiagonal matrices U—cf. (30a) and (30d) below.]

$$\text{class III: } A=\infty,\quad D=(\pi/4)\cot(\omega/2);$$

$$\text{class IV: } A=(1/\sqrt{2})(1/\tan(\omega/2)),\quad D=\infty;$$

$$\text{class V: } A=D=\infty.$$

Hence the system separates for a diagonal U into two independent subsystems and its Hamiltonian is of the form

$$H_U=H^{(A)}_{0,1}\oplus H^{(D)}_{0,2} \tag{25}$$

where $H^{(A)}_{0,1}$ is the half-line Hamiltonian (cf. Ref. 9, Sec. X.1) specified by the boundary condition (24a) and $H^{(D)}_{0,2}=h^{(D)}_0\oplus\bar{h}$ is the two-dimensional point-interaction Hamiltonian.[12] Scattering by this point interaction as well as reflection on the half-line should be considered separately. Passage of the electron between the two parts of the configuration manifold is impossible.

(b) U nondiagonal: Now the transitions from $\mathbb{R}^-$ to $\mathbb{R}^2$ and vice versa become possible. First of all, we shall discuss in detail the situation when U belongs to class I. Let us consider reflection of the electron moving initially along the half-line towards P. Using boundary conditions (16), it is easy to see that the function $f_U=(\varphi^U_1,\varphi^U_2)$ with

$$\varphi^U_1(x)=e^{ikx}+a_Ue^{-ikx}, \tag{26a}$$

$$\varphi^U_2(r)=\tilde{b}_uH^{(1)}_0(kr), \tag{26b}$$

for a given $k>0$ belongs locally to $D(H_U)$ if

$$a_U=-\left\{(A-ik)\left[1+\frac{2i}{\pi}\left(\gamma-D+\ln\frac{k}{2}\right)\right]+\frac{2i}{\pi}BC\right\}\times\left\{(A+ik)\left[1+\frac{2i}{\pi}\left(\gamma-D+\ln\frac{k}{2}\right)\right]+\frac{2i}{\pi}BC\right\}^{-1}, \tag{27a}$$

$$\tilde{b}_U=2iCk\left\{(A+ik)\left[1+\frac{2i}{\pi}\left(\gamma-D+\ln\frac{k}{2}\right)\right]+\frac{2i}{\pi}BC\right\}^{-1}. \tag{27b}$$

Moreover, it holds that

$$\left(-\frac{d^2}{dx^2}-k^2\right)\varphi^U_1=0,$$

$$(-\Delta-k^2)\varphi^U_2=0,$$

so f_U given by the relations (26) and (27) is a generalized eigenvector of H_U, and $|a_U|^2$ is therefore nothing but the reflection coefficient at the point singularity.

Relation (27a) shows that the reflection coefficient depends on the chosen Hamiltonian H_U. In particular, $|a_U|=1$ holds if U becomes diagonal so $BC=0$. Then the electron can be only reflected at P. On the other hand, occurrence of the transitions from $\mathbb{R}^-$ to $\mathbb{R}^2$ means

$$|a_U|<1. \tag{28}$$

We shall check directly that this inequality holds once U is nondiagonal. To this end, we shall use the following explicit parametrization of a unitary 2×2 matrix:

$$U=e^{i\xi}\begin{pmatrix}e^{i(\alpha+\delta)}\cos\beta & e^{i(\delta-\alpha)}\sin\beta\\ -e^{i(\alpha-\delta)}\sin\beta & e^{-i(\alpha+\delta)}\cos\beta\end{pmatrix}, \tag{29}$$

where $\alpha, \beta, \delta, \xi$ are real parameters (the first three of them are nothing but doubled Euler angles). It yields the following expressions for the coefficients (16b)–(16e):

$$A = \mathscr{S}/\mathscr{D}, \tag{30a}$$

$$B = (i/2\sqrt{2})(e^{i(\alpha-\delta)}\sin\beta)/\mathscr{D}, \tag{30b}$$

$$C = -(i/\sqrt{2})(e^{i(\delta-\alpha)}\sin\beta)/\mathscr{D}, \tag{30c}$$

$$D = \gamma - \ln 2 - (\pi/4)(\mathscr{C}/\mathscr{D}), \tag{30d}$$

where

$$\mathscr{S} = \sin(\alpha+\delta+(\pi/4))\cos\beta - \sin(\xi+(\pi/4)),$$

$$\mathscr{D} = \sin(\alpha+\delta)\cos\beta - \sin\xi,$$

$$\mathscr{C} = \cos(\alpha+\delta)\cos\beta + \cos\xi.$$

Notice that $\mathscr{D}$ is nonzero for Class I matrices. As we have remarked, the relations (30) show that the "diagonal" coefficients A, D are real valued, while the "nondiagonal" ones are complex conjugated up to a real multiplicative constant. Using these expressions, we find

$$a_U = -\left\{(\mathscr{S}-ik\mathscr{D})\left[\left(1+\frac{2i}{\pi}\ln k\right)\mathscr{D}+\frac{1}{2}\mathscr{C}\right]+\frac{i}{2\sqrt{2}}\sin^2\beta\right\}\left\{(\mathscr{S}+ik\mathscr{D})\times\left[\left(1+\frac{2i}{\pi}\ln k\right)\mathscr{D}+\frac{1}{2}\mathscr{C}\right]+\frac{i}{2\sqrt{2}}\sin^2\beta\right\}^{-1}, \tag{31}$$

so after a short calculation we arrive at the relation

$$1-|a_U|^2 = \sqrt{2}\mathscr{D}^2 k\sin^2\beta \times\left[\mathscr{D}^2\left(\mathscr{S}-\frac{2}{\pi}\mathscr{D}k\ln k-\frac{1}{2}\mathscr{C}k\right)^2+\left(\frac{1}{2}\mathscr{S}\mathscr{C}+\frac{2}{\pi}\mathscr{S}\mathscr{D}\ln k+\mathscr{D}^2k+\frac{1}{2\sqrt{2}}\sin^2\beta\right)^2\right]^{-1}>0, \tag{28'}$$

which proves (28).

Notice that the squared modulus of (27b) is not the transition coefficient, since it is not properly normalized. Relation (28') shows that it is $b_U = (\sqrt{2}/k)^{1/2}\tilde{b}_U$ which fulfills $|a_U|^2 + |b_U|^2 = 1$.

The fact that the singularity is penetrable for a nondiagonal U can be seen also when one considers scattering of the electron moving in the plane on the point singularity P. The corresponding generalized eigenfunction is $f_U = (\psi_1^U, \psi_2^U)$ with

$$\varphi_1^U(x) = \tilde{c}_U e^{-ikx}, \tag{32a}$$

$$\varphi_2^U(r) = J_0(kr) + d_U H_0^{(1)}(kr), \tag{32b}$$

where

$$\tilde{c}_U = \frac{2i}{\pi}B\left\{(A+ik)\left[1+\frac{2i}{\pi}\left(\gamma-D+\ln\frac{k}{2}\right)\right]+\frac{2i}{\pi}BC\right\}^{-1}, \tag{33a}$$

$$d_U = -\left[1+\frac{2i}{\pi}\left(\gamma-D+\ln\frac{k}{2}+\frac{BC}{A+ik}\right)\right]^{-1}, \tag{33b}$$

if we require f_U to belong locally to $D(H_U)$. The asymptotics for large r can be found easily,

$$\psi_2^U(r) = (2/\pi kr)^{1/2}\{e^{i\delta_0(k)}\sin(kr+(\pi/4)+\delta_0(k)) + O(r^{-1})\}, \tag{34}$$

where $\delta_0(k)$ is given by

$$S_0(k) = e^{2i\delta_0(k)} = \left(\gamma-D+\ln\frac{k}{2}+\frac{\pi i}{2}+\frac{BC}{A+ik}\right)\times\left(\gamma-D+\ln\frac{k}{2}-\frac{\pi i}{2}+\frac{BC}{A+ik}\right)^{-1}. \tag{35a}$$

Here $\delta_0(k)$ represents the s-wave scattering phase shift and $S_0(k)$ is the on-shell s-wave scattering matrix. For the higher partial waves, we get

$$\delta_m(k) = 0, \quad m = \pm 1, \pm 2, \ldots . \tag{35b}$$

In general, the scattering matrix is not unitary. This is not surprising because the electron can continue its motion in $\mathbf{R}^-$ after the scattering, vanishing thus from the plane. In order to demonstrate it explicitly, one has to express $S_0(k)$ using the parametrization (29). A short calculation then gives

$$1-|S_0(k)|^2 = 1-|a_U|^2, \tag{36}$$

where the rhs is given by (28'). Hence S is nonunitary iff U is nondiagonal.

Let us turn now to matrices U of class II. In this case, too, the electron is able to pass through the singular point. The analysis is essentially the same as above. We restrict ourselves with presenting the results. For the generalized eigenvector (26), we find now

$$a_U = -\left[1+\frac{2i}{\pi}\left(\gamma-\frac{F+ikG}{E}+\ln\frac{k}{2}\right)\right]\times\left[1+\frac{2i}{\pi}\left(\gamma-\frac{F-ikG}{E}+\ln\frac{k}{2}\right)\right]^{-1}, \tag{37a}$$

$$\tilde{b}_U = 2ikG\left[1+\frac{2i}{\pi}\left(\gamma-\frac{F-ikG}{E}+\ln\frac{k}{2}\right)\right]^{-1}. \tag{37b}$$

On the other hand, for the scattering in the plane corresponding to (32), one can find the coefficients $\tilde{c}_U$, d_U, which give

$$S_0(k) = e^{2i\delta_0(k)} = \left(\gamma-\frac{F-ikG}{E}+\frac{\pi i}{2}+\ln\frac{k}{2}\right)\times\left(\gamma-\frac{F-ikG}{E}-\frac{\pi i}{2}+\ln\frac{k}{2}\right)^{-1}. \tag{38}$$

V. A POSSIBLE APPLICATION

The problem treated in the preceding sections may seem somewhat bizarre. Nevertheless, it can have a quite reasonable physical application as a model of the quantum point-contact spectroscopy.

For a metallic contact, one usually expects a linear relation between the applied voltage and the current according to Ohm's law. This is true if the size of the contact is large enough. On the other hand, if its linear dimensions becomes comparable with mean free path of the electrons in metal, then interesting nonlinear effects in the current-voltage characteristics can be observed.[13] In this case, the electrons are scattered at the orifice giving rise to a backward flow, which adds a negative and voltage-dependent contribution to the current.

The results of the present analysis can be used for modeling of such a contact whose linear dimension tends to zero. In order to calculate the current through the contact, one has to know the electron-gas density and the transmission coefficient through the singular point.[14] In the simplest case, when the electrons are supposed to be free, the latter is given by (28') (or an analogous expression for U of class II). If we add a potential to H_U which should describe the metallic structure of the system (a wire connected to a thin plate), then the transmission coefficient must be calculated anew. It remains possible, however, to characterize the admissible Hamiltonians by the boundary conditions listed in the theorem of Sec. III as far as the potential is bounded.

We are going to discusss the model, which we have sketched briefly here, in a subsequent paper.

ACKNOWLEDGMENTS

The authors are indebted to Professor A. Uhlmann and Professor G. Vojta who suggested the problem.

[1]S. Albeverio and R. Høegh-Krohn, J. Operator Theory **6**, 313 (1981).
[2]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and W. Kirsch, J. Operator Theory **12**, 101 (1984).
[3]J. Dittrich and P. Exner, J. Math. Phys. **26**, 2000 (1985).
[4]P. Šeba, Czech. J. Phys. B **36**, 455, 559 (1986).
[5]Yu. A. Kuperin, K. A. Makarov, and B. S. Pavlov, Teor. Mat. Fiz. **63**, 78 (1985) (in Russian).
[6]B. S. Pavlov and M. D. Faddeev, Teor. Mat. Fiz. **55**, 257 (1983) (in Russian).
[7]Yu. N. Demkov and V. N. Ostrovsky, *The Method of Zero-Range Potentials in Nuclear Physics* (Leningrad U.P., Leningrad, 1975) (in Russian).
[8]B. S. Pavlov, Teor. Mat. Fiz. **59**, 345 (1984) (in Russian).
[9]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II. Fourier Analysis. Self-Adjointness* (Academic, New York, 1975).
[10]W. Bulla and F. Gesztesy, J. Math. Phys. **26**, 2520 (1985).
[11]I. S. Gradshtein and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products* (Nauka, Moscow, 1971) (in Russian).
[12]S. Albeverio, E. Gesztesy, R. Høegh-Krohn, and H. Holden, "Solvable models in quantum mechanics," in *Texts and Monographs in Physics* (Springer, Berlin, to appear).
[13]For a review, see A. G. M. Jansen, A. P. van Gelder, and P. Wyder, J. Phys. C **13**, 6073 (1980).
[14]K. C. Kao and W. Huang, *Electrical Transport in Solids* (Pergamon, 1981).

The Jacobi-matrix method in parabolic coordinates: Expansion of Coulomb functions in parabolic Sturmians

P. C. Ojha
Department of Chemistry, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637

(Received 7 July 1986; accepted for publication 8 October 1986)

Previous analysis of the Jacobi-matrix method based on the underlying SO(2,1) Lie algebra is extended to the Coulomb Hamiltonian in parabolic coordinates. The general solution of the generic SO(2,1) eigenvalue equation is constructed and special cases, which furnish expansions of the Coulomb functions $\psi_{\mathbf{k}}^{(\pm)}(\mathbf{r})$ in a complete set of parabolic Sturmians, are discussed.

I. INTRODUCTION

New possibilities for computing the scattering wave function are afforded by expansion in a set of square-integrable functions. There are two closely related but complementary techniques; the first is suitable when the focus is on the overall behavior of a cross section over a broad energy range, and the second is suitable for bringing out finer detail over a narrow range of energy.

The first technique (Stieltjes imaging[1]) relies on the fact that given any Hamiltonian H [or a functional $A(H)$], one can recursively generate a basis set $\{|\phi_\nu\rangle, \nu = 0,1,2,...,\infty\}$ of square-integrable functions in which the Hamiltonian H is tridiagonal. This Jacobi-matrix representation of the Hamiltonian immediately permits application of the powerful techniques for solving the classical "problem of moments" to the problem at hand, usually a direct computation of a matrix element for transition to the scattering state.

The alternative technique, christened the Jacobi-matrix method,[2] is applicable when the Hamiltonian is of a special form, $H = H_0 + V$, where H_0 is tridiagonal in a known, analytically defined, square-integrable basis set and the resulting Jacobi matrix can be diagonalized analytically. Furthermore, the potential V is supposed to have a short range in the sense that its matrix representation in this basis set contains, at most, a finite nonzero submatrix. Under these conditions, the scattering states of H may be constructed by expansion in the set $\{|\phi_\nu\rangle, \nu = 0,1,2,...,\infty\}$. Problems that commonly arise in atomic physics are amenable to this treatment.

An enumeration of the Hamiltonians H_0 and the corresponding basis sets $\{|\phi_\nu\rangle, \nu = 0,1,2,...,\infty\}$ is thus central to the practical application of the Jacobi-matrix method. Heller, Yamani, and Fishman[2] identified two distinct cases: (i) the kinetic energy operator $T(\mathbf{r})$ diagonalized in the set of harmonic oscillator functions, and (ii) the Coulomb Hamiltonian, $H(\mathbf{r}) = T(\mathbf{r}) - Z/r$, diagonalized in a set of Sturmian functions.

The unifying feature of these two seemingly distinct cases has been identified in a recent paper.[3] It so happens that both these basis sets constitute infinite-dimensional, unitary, irreducible representations of different realizations of the SO(2,1) Lie algebra. Moreover, in both cases, solving the Schrödinger equation for scattering states amounts to diagonalizing a linear combination of the compact and noncompact generators of the algebra in a basis in which the compact generator is diagonal.

The analysis of Refs. 2 and 3 was limited to spherical geometry, i.e., to instances when L^2 and L_z are simultaneously diagonalized. The task of expanding continuum functions of a particular symmetry in square-integrable basis functions of the same symmetry was accomplished there.

The analysis of the Coulomb potential, thus begun within the framework of partial-wave expansion, is completed in this paper. The final result is a set of coefficients for the expansion of the Coulomb functions,

$$\psi_{\mathbf{k}}^{(\pm)}(\mathbf{r}) = (2\pi)^{-3/2}\exp\left(\frac{\pi}{2}\frac{Z}{k}\right)\Gamma\left(1\mp\frac{iZ}{k}\right) \times\exp\left[i\frac{k}{2}(\eta-\xi)\right]{}_1F_1\left[\pm\frac{iZ}{k};1;\pm ik\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right], \tag{1.1}$$

in a set of parabolic Sturmian functions,[4]

$$\begin{aligned}\psi_{\nu_1,\nu_2,m}(\xi,\eta,\phi) &= \left[\frac{\zeta^2\Gamma(\nu_1+1)\Gamma(\nu_2+1)}{\pi\Gamma(\nu_1+|m|+1)\Gamma(\nu_2+|m|+1)}\right]^{1/2}\exp(im\phi)\\ &\times\left[\exp\left(-\frac{1}{2}\zeta\xi\right)(\zeta\xi)^{|m|/2}L_{\nu_1}^{|m|}(\zeta\xi)\right]\\ &\times\left[\exp\left(-\frac{1}{2}\zeta\eta\right)(\zeta\eta)^{|m|/2}L_{\nu_2}^{|m|}(\zeta\eta)\right],\end{aligned} \tag{1.2}$$

where $\xi = r - \hat{k}\cdot\mathbf{r}$, $\eta = r + \hat{k}\cdot\mathbf{r}$ are the two parabolic coordinates, ϕ is the azimuthal angle, ζ is a common Sturmian exponent, the parabolic quantum numbers ν_1 and ν_2 range from 0 to ∞, and m is the magnetic quantum number. This basis set is orthonormal with the inner product defined with a $1/r$ weight:

$$\int d\mathbf{r}\,\psi^*_{\nu_1,\nu_2,m}(\mathbf{r})\frac{1}{r}\psi_{\nu_1',\nu_2',m'}(\mathbf{r}) = \delta_{\nu_1\nu_1'}\delta_{\nu_2\nu_2'}\delta_{mm'}.$$

Of course, only $m = 0$ states appear in the expansion of $\psi_{\mathbf{k}}^{(\pm)}(\mathbf{r})$.

Central to this expansion is the fact that the parabolic bound states of the hydrogen atom are also generated by an SO(2,1) Lie algebra.[5] Conversely, the basis functions of Eq. (1.2), for a fixed value of m, constitute a unitary, irreducible representation of the SO(2,1) ⊗ SO(2,1) Lie algebra.

The generators of this algebra are specified in Sec. II. The generic eigenvalue problem for the SO(2,1) algebra that arises when the Schrödinger equation is rewritten in terms of these generators (Sec. IV) is solved in Sec. III. Specific solu-

 0022-2488/87/020392-05$02.50 © 1987 American Institute of Physics

tions that give the expansion coefficients of $\psi_k^{(\pm)}(\mathbf{r})$ in the Sturmian functions are constructed in Sec. IV and possible application to the Stark spectrum of hydrogenic atoms is pointed out in the concluding section (Sec. V).

II. PARABOLIC GENERATORS OF THE HYDROGENIC SO(2,1) LIE ALGEBRA

One may define the following parabolic generators of the SO(2,1) Lie algebra:

$$N_1 = \tfrac{1}{2}(T_1 + \hat{k}\cdot\mathbf{B}), \quad N_2 = \tfrac{1}{2}(T_2 - \hat{k}\cdot\mathbf{\Gamma}), \quad N_3 = \tfrac{1}{2}(T_3 + \hat{k}\cdot\mathbf{A}) \tag{2.1a}$$

and

$$M_1 = \tfrac{1}{2}(T_1 - \hat{k}\cdot\mathbf{B}), \quad M_2 = \tfrac{1}{2}(T_2 + \hat{k}\cdot\mathbf{\Gamma}), \quad M_3 = \tfrac{1}{2}(T_3 - \hat{k}\cdot\mathbf{A}), \tag{2.1b}$$

in terms of the standard generators of the SO(4,2) noninvariance algebra,[5,6]

$$T_1 = \frac{1}{2\zeta} r(p^2 - \zeta^2), \quad T_2 = \mathbf{r}\cdot\mathbf{p} - i,$$

$$T_3 = \frac{1}{2\zeta} r(p^2 + \zeta^2), \tag{2.2a}$$

$$\mathbf{A} = \frac{1}{2\zeta}\mathbf{r}p^2 - \frac{1}{\zeta}\mathbf{p}(\mathbf{r}\cdot\mathbf{p}) - \frac{\zeta}{2}\mathbf{r}, \tag{2.2b}$$

$$\mathbf{B} = \frac{1}{2\zeta}\mathbf{r}p^2 - \frac{1}{\zeta}\mathbf{p}(\mathbf{r}\cdot\mathbf{p}) + \frac{\zeta}{2}\mathbf{r}, \tag{2.2c}$$

$$\mathbf{\Gamma} = r\mathbf{p}, \tag{2.2d}$$

and the unit vector $\hat{k}$ which may be taken to lie along the z axis. The raising and lowering operators are defined, as usual, by $N_\pm = N_1 \pm iN_2$ and $M_\pm = M_1 \pm iM_2$.

It is easily verified from the known commutation relations of the generators of (2.2) that the generators of (2.1) do, indeed, satisfy the canonical SO(2,1) commutation relations:

$$[N_1,N_2] = -iN_3, \quad [N_2,N_3] = iN_1, \quad [N_3,N_1] = iN_2, \tag{2.3a}$$

and similarly for M_1, M_2, and M_3. It may also be verified that the corresponding Casimir invariants are equal for the two realizations:

$$M^2 = N^2 = N_3^2 - N_1^2 - N_2^2 = \tfrac{1}{4}(L_k^2 - 1), \tag{2.3b}$$

where L_k is the projection of the angular momentum on the vector $\hat{k}$.

The label "parabolic generators" is justified by the explicit form of the operators of (2.1) in parabolic coordinates:

$$\eta = r + \hat{k}\cdot\mathbf{r} \quad \text{and} \quad \xi = r - \hat{k}\cdot\mathbf{r}. \tag{2.4}$$

Then,

$$N_1 = -\frac{1}{\zeta}\left[\frac{\partial}{\partial\xi}\left(\xi\frac{\partial}{\partial\xi}\right) + \frac{1}{4\xi}\left(\frac{\partial^2}{\partial\phi^2}\right) + \frac{\zeta^2}{4}\xi\right], \tag{2.5a}$$

$$N_2 = -i\left[\xi\frac{\partial}{\partial\xi} + \frac{1}{2}\right], \tag{2.5b}$$

$$N_3 = -\frac{1}{\zeta}\left[\frac{\partial}{\partial\xi}\left(\xi\frac{\partial}{\partial\xi}\right) + \frac{1}{4\xi}\left(\frac{\partial^2}{\partial\phi^2}\right) - \frac{\zeta^2}{4}\xi\right], \tag{2.5c}$$

$$M_1 = -\frac{1}{\zeta}\left[\frac{\partial}{\partial\eta}\left(\eta\frac{\partial}{\partial\eta}\right) + \frac{1}{4\eta}\left(\frac{\partial^2}{\partial\phi^2}\right) + \frac{\zeta^2}{4}\eta\right], \tag{2.6a}$$

$$M_2 = -i\left[\eta\frac{\partial}{\partial\eta} + \frac{1}{2}\right], \tag{2.6b}$$

$$M_3 = -\frac{1}{\zeta}\left[\frac{\partial}{\partial\eta}\left(\eta\frac{\partial}{\partial\eta}\right) + \frac{1}{4\eta}\left(\frac{\partial^2}{\partial\phi^2}\right) - \frac{\zeta^2}{4}\eta\right]. \tag{2.6c}$$

The action of the operators N_1 to M_3 on the basis functions of (1.2) can be deduced from the properties of associated Laguerre polynomials.[7] It may be summarized succinctly by replacing the quantum numbers (ν_1,ν_2,m) by the set (q_1,q_2,t), where $q_1 = \nu_1 + \frac{1}{2} + \frac{1}{2}|m|$, $q_2 = \nu_2 + \frac{1}{2} + \frac{1}{2}|m|$, and $t = -\frac{1}{2} + \frac{1}{2}|m|$, and relabeling the basis functions, i.e., $|q_1,q_2,t\rangle \equiv |\nu_1,\nu_2,m\rangle$. Then,

$$N^2|q_1,q_2,t\rangle = t(t+1)|q_1,q_2,t\rangle \tag{2.7a}$$

$$N_3|q_1,q_2,t\rangle = q_1|q_1,q_2,t\rangle, \tag{2.7b}$$

$$N_\pm|q_1,q_2,t\rangle = [(q_1 \mp t)(q_1 \pm t \pm 1)]^{1/2}|q_1 \pm 1,q_2,t\rangle. \tag{2.7c}$$

Similarly the action of M^2, M_3, and $M_\pm$ is given by replacing $q_1 \to q_2$ in Eq. (2.7); i.e., $|q_1,q_2,t\rangle$ is also an eigenfunction of M^2 and M_3 with eigenvalues $t(t+1)$ and q_2, respectively, and $M_\pm$ are raising and lowering operators in the index q_2.

This is, of course, the standard action of SO(2,1) generators on the basis functions of the $\mathscr{D}^+(t)$ representation.[5,6] Thus the set $\{|q_1,q_2,t\rangle, q_1 = t+1, t+2, \ldots, \infty\}$ constitutes a $\mathscr{D}^+(t)$ representation of the SO(2,1) Lie algebra [Eq. (2.1a)]; similarly the set $\{|q_1,q_2,t\rangle, q_2 = t+1, t+2, \ldots, \infty\}$ constitutes a $\mathscr{D}^+(t)$ representation of another realization of the algebra [Eq. (2.1b)]. Taken together, the set $\{|q_1,q_2,t\rangle, q_1,q_2 = t+1, t+2, \ldots, \infty\}$ constitutes a $\mathscr{D}^+(t) \otimes \mathscr{D}^+(t)$ representation of the corresponding algebra SO(2,1) $\otimes$ SO(2,1).

Transcribing this statement in terms of the original quantum numbers, the basis set of Eq. (1.2), for a fixed value of the magnetic quantum number m, constitutes a $\mathscr{D}^+(-\frac{1}{2}+\frac{1}{2}|m|) \otimes \mathscr{D}^+(-\frac{1}{2}+\frac{1}{2}|m|)$ representation of the SO(2,1) $\otimes$ SO(2,1) algebra defined by the generators of (2.1).

III. SOLUTION OF THE BASIC EIGENVALUE EQUATION

In anticipation of the problem encountered in the next section, I will now summarize the relevant solutions of the basic eigenvalue problem,

$$(T_1 + \eta_1 T_3 - \eta_2)|\psi\rangle = 0, \tag{3.1}$$

where T_1 and T_3 are generators of an SO(2,1) Lie algebra, $-1 < \eta_1 < 1$, and η_2 is a *complex* number. The corresponding problem for *real* values of η_2 was solved in Ref. 3. The extension to complex η_2 is straightforward, so all proofs will be omitted.

Expansion of $|\psi\rangle$ in the basis functions of the $\mathscr{D}^+(t)$ representation of the algebra,

$$|\psi\rangle = \sum_{\nu=0}^{\infty} a_\nu |t+1+\nu,t\rangle,$$

leads to the following three-term recursion for a_ν:

$$[\nu(\nu+2t+1)]^{1/2}a_{\nu-1} - 2[(\nu+t+1)\cos\phi + \gamma\sin\phi]a_\nu + [(\nu+1)(\nu+2t+2)]^{1/2}a_{\nu+1} = 0, \tag{3.2}$$

where $\sin\phi = \sqrt{1-\eta_1^2}$, $\cos\phi = -\eta_1$, $0 < \phi < \pi$, and $\gamma \equiv \gamma_1 + i\gamma_2 = \eta_2/\sin\phi$.

It can be easily verified[8] that

$$a_\nu^{\rm I}(\gamma)=\frac{[\Gamma(\nu+1)\Gamma(\nu+2t+2)]^{1/2}}{\Gamma(\nu+t+2+i\gamma)}\exp(i\nu\phi)\times{}_2F_1(t+1+i\gamma,\nu+2t+2;\nu+t+2+i\gamma;e^{i2\phi}) \tag{3.3a}$$

and

$$a_\nu^{\rm II}(\gamma)=\frac{[\Gamma(\nu+1)\Gamma(\nu+2t+2)]^{1/2}}{\Gamma(\nu+t+2-i\gamma)}\exp(-i\nu\phi)\times{}_2F_1(t+1-i\gamma,\nu+2t+2;\nu+t+2-i\gamma;e^{-i2\phi}) \tag{3.3b}$$

are two linearly independent solutions of the recursion (3.2). [Here ${}_2F_1(a,b;c;z)$ is the Gauss hypergeometric function.] The second of these solutions is obtained from the first by replacing $\gamma\to-\gamma$, $\phi\to-\phi$.

One must exercise some care in writing the asymptotic forms of these solutions for complex $\gamma\equiv\gamma_1+i\gamma_2$:

$$\lim_{\nu\to\infty} a_\nu^{\rm I}(\gamma)=\frac{\exp[\gamma_1(\phi-\pi/2)]}{(2\sin\phi)^{t+1/2}}[2(\nu+t+1)\sin\phi]^{-1/2+\gamma_2}\times e^{i\nu\phi}\exp[-i(t+1-\gamma_2)(\phi-\pi/2)-i\gamma_1\ln(2(\nu+t+1)\sin\phi)], \tag{3.4a}$$

and

$$\lim_{\nu\to\infty} a_\nu^{\rm II}(\gamma)=\frac{\exp[-\gamma_1(\phi-\pi/2)]}{(2\sin\phi)^{t+1/2}}[2(\nu+t+1)\sin\phi]^{-1/2+\gamma_2}\times e^{-i\nu\phi}\exp[i(t+1+\gamma_2)(\phi-\pi/2)+i\gamma_1\ln(2(\nu+t+1)\sin\phi)]. \tag{3.4b}$$

The linear combination

$$\mathscr{S}_\nu(\gamma)=\Gamma(t+1+i\gamma)[1-\exp(i2\phi)]^{2t+1}e^{i\phi}a_\nu^{\rm I}(\gamma)-\Gamma(t+1-i\gamma)[1-\exp(-i2\phi)]^{2t+1}e^{-i\phi}a_\nu^{\rm II}(\gamma) \tag{3.5a}$$

satisfies the initial condition

$$\sqrt{(2t+2)}\,\mathscr{S}_1(\gamma)-2[(t+1)\cos\phi+\gamma\sin\phi]\mathscr{S}_0(\gamma)=0, \tag{3.5b}$$

and hence defines the regular solution of the recursion. It may be written compactly in the following form:

$$\mathscr{S}_\nu(\gamma)=\exp[(2\phi-\pi)(\gamma-i\tfrac{1}{2})](2\sin\phi)^{2t+1}\frac{\Gamma(t+1+i\gamma)\Gamma(t+1-i\gamma)}{\Gamma(2t+2)}\times\left[\frac{\Gamma(\nu+2t+2)}{\Gamma(\nu+1)}\right]^{1/2}e^{i(\nu+1)\phi}{}_2F_1(-\nu,t+1+i\gamma;2t+2;1-e^{-i2\phi}), \tag{3.6}$$

which brings out the finite polynomial nature of the solution. Asymptotically,

$$\lim_{\nu\to\infty}\mathscr{S}_\nu(\gamma)=\frac{\exp[\gamma_1(\phi-\pi/2)](2\sin\phi)^{t+1/2}}{[2(\nu+t+1)\sin\phi]^{1/2-\gamma_2}}\left\{\Gamma(t+1+i\gamma)\exp\left[i(\nu+1)\phi+i(t+\gamma_2)\left(\phi-\frac{\pi}{2}\right)-i\gamma_1\ln(2(\nu+t+1)\sin\phi)\right]-\frac{\Gamma(t+1-i\gamma)}{[2(\nu+t+1)\sin\phi]^{2\gamma_2}}\times\exp\left[-i(\nu+1)\phi-i(t-\gamma_2)\left(\phi-\frac{\pi}{2}\right)+i\gamma_1\ln(2(\nu+t+1)\sin\phi)\right]\right\}. \tag{3.7}$$

One may define another solution of the recursion by the linear combination

$$\mathscr{C}_\nu(\gamma)=-i\Gamma(t+1+i\gamma)[1-\exp(i2\phi)]^{2t+1}e^{i\phi}a_\nu^{\rm I}(\gamma)+i\Gamma(t+1-i\gamma)[1-\exp(-i2\phi)]^{2t+1}e^{-i\phi}a_\nu^{\rm II}(\gamma).$$

This satisfies the initial condition

$$\sqrt{(2t+2)}\,\mathscr{C}_1(\gamma)-2[(t+1)\cos\phi+\gamma\sin\phi]\mathscr{C}_0(\gamma)\neq0,$$

and it is therefore an irregular solution of the recursion. It may be written more compactly as

$$\mathscr{C}_\nu(\gamma)+i\mathscr{S}_\nu(\gamma)=-2i\frac{\Gamma(t+1-i\gamma)}{\Gamma(\nu+t+2-i\gamma)}[\Gamma(\nu+1)\Gamma(\nu+2t+2)]^{1/2}e^{-i(\nu+1)\phi}{}_2F_1(\nu+1,-t-i\gamma;\nu+t+2-i\gamma;e^{-i2\phi}). \tag{3.8}$$

Asymptotically,

$$\lim_{\nu\to\infty}\mathscr{C}_\nu(\gamma)=-i\frac{\exp[\gamma_1(\phi-\pi/2)](2\sin\phi)^{t+1/2}}{[2(\nu+t+1)\sin\phi]^{1/2-\gamma_2}}\left\{\Gamma(t+1+i\gamma)\exp\left[i(\nu+1)\phi+i(t+\gamma_2)\left(\phi-\frac{\pi}{2}\right)-i\gamma_1\ln(2(\nu+t+1)\sin\phi)\right]+\frac{\Gamma(t+1-i\gamma)}{[2(\nu+t+1)\sin\phi]^{2\gamma_2}}\times\exp\left[-i(\nu+1)\phi-i(t+\gamma_2)\left(\phi-\frac{\pi}{2}\right)+i\gamma_1\ln(2(\nu+t+1)\sin\phi)-i\pi\gamma_2\right]\right\}. \tag{3.9}$$

This irregular solution is defined so that for real γ ($\gamma_2 = 0$), in the asymptotic limit, it has the same amplitude as the regular solution and its phase leads that of the regular solution by $\pi/2$.

The Wronskian of these solutions, defined by

$$W[\mathscr{S},\mathscr{C}] \equiv [(\nu+1)(\nu+2t+2)]^{1/2} \times [\mathscr{S}_{\nu+1}\mathscr{C}_{\nu} - \mathscr{S}_{\nu}\mathscr{C}_{\nu+1}],$$

may be computed most easily for $\nu = 0$:

$$W[\mathscr{S},\mathscr{C}] = 2\exp[\gamma(2\phi-\pi)](2\sin\phi)^{2t+1} \times \Gamma(t+1+i\gamma)\,\Gamma(t+1-i\gamma). \quad (3.10)$$

IV. SCATTERING STATES OF THE COULOMB POTENTIAL

The Schrödinger equation

$$(\tfrac{1}{2}p^2 - Z/r - E)|\psi\rangle = 0 \quad (4.1)$$

is rewritten in terms of the operators of (2.1) after multiplying on the left by r:

$$[(N_1+M_1) + \eta_1(N_3+M_3) - \eta_2]|\psi\rangle = 0, \quad (4.2)$$

where

$$\eta_1 = \frac{1-2E/\zeta^2}{1+2E/\zeta^2} \quad \text{and} \quad \eta_2 = \frac{2Z/\zeta}{1+2E/\zeta^2}.$$

Note that for scattering states ($E>0$), $-1<\eta_1<1$.

Equation (4.2) is solved by expanding $|\psi\rangle$ in the basis set $\{|\nu_1,\nu_2,m\rangle,\ \nu_1,\nu_2 = 0,1,2,\ldots\}$ of (1.2) and requiring the expansion coefficients to be separable in the indices ν_1 and ν_2;

$$|\psi\rangle = \sum_{\nu_1,\nu_2=0}^{\infty} a_{\nu_1} b_{\nu_2} |\nu_1,\nu_2,m\rangle. \quad (4.3)$$

Next introduce separation constants η_{2N} and η_{2M} such that $\eta_{2N} + \eta_{2M} = \eta_2$. (This implies a separation of charge, $Z_M + Z_N = Z$.) Then (4.2) separates into two equations

$$[N_1 + \eta_1 N_3 - \eta_{2N}]\left[\sum_{\nu_1=0}^{\infty} a_{\nu_1}|\nu_1,\nu_2,m\rangle\right] = 0 \quad (4.4a)$$

and

$$[M_1 + \eta_1 M_3 - \eta_{2M}]\left[\sum_{\nu_2=0}^{\infty} b_{\nu_2}|\nu_1,\nu_2,m\rangle\right] = 0. \quad (4.4b)$$

Note that η_{2N} and η_{2M} may be complex and recall that $t = \frac{1}{2} + \frac{1}{2}|m|$ (Sec. II). Indeed, in the following special cases, which correspond to the Coulomb function $\psi_k^{\pm}(\mathbf{r})$ solutions of the Schrödinger equation, both η_{2N} and η_{2M} will be complex.

A. $m=0, Z_N = Z + ik/2, Z_M = -ik/2$

This choice of the separation constants, being identical to the choice made in obtaining $\psi_k^{(+)}(\mathbf{r})$ as a separable solution of the Schrödinger equation in parabolic coordinates, implies that the corresponding coefficients $a_{\nu_1} b_{\nu_2}$ in (4.3) are the expansion coefficients of $\psi_k^{(+)}(\mathbf{r})$ in the Sturmian set of (1.2). Note that in the following, the dependence of a_{ν_1} and b_{ν_2} on the Sturmian exponent ζ is only through the angle ϕ $(0<\phi<\pi)$ defined by $\sin\phi = \sqrt{1-\eta_1^2} = 2\zeta k/(\zeta^2+k^2)$. The parameter $\gamma = \eta_2/\sqrt{1-\eta_1^2} = Z/k$ does not depend on ζ.

Corresponding to the parameters defined in the heading, $t = \frac{1}{2}$, $\gamma_N = Z/k + i\frac{1}{2}$, and $\gamma_M = -i\frac{1}{2}$. The corresponding regular solutions of the recursion are

$$a_{\nu_1} = \frac{2^{1/4}}{\sqrt{\pi\zeta}}\sin\left(\frac{\phi}{2}\right) \times \exp\left[i\frac{1}{2}(\phi-\pi)\right]\exp\left[\frac{Z}{k}\left(\phi-\frac{\pi}{2}\right)\right]\Gamma\left(1-\frac{iZ}{k}\right) \times e^{i\nu_1\phi}\,{}_2F_1(-\nu_1, iZ/k; 1; 1-e^{-i2\phi}) \quad (4.5a)$$

and

$$b_{\nu_2} = \frac{2^{1/4}}{\sqrt{\pi\zeta}}\sin\left(\frac{\phi}{2}\right)\exp\left[-i\frac{1}{2}(\phi-\pi)\right]e^{-i\nu_2\phi}. \quad (4.5b)$$

Note the close correspondence between the asymptotic form of a_{ν_1},

$$\lim_{\nu_1\to\infty} a_{\nu_1} = \frac{2^{1/4}}{\sqrt{\pi\zeta}}\sin\left(\frac{\phi}{2}\right)(-1)^{\nu}\left\{\exp\left[-i\left(\nu+\frac{1}{2}\right)(\pi-\phi) - i\frac{Z}{k}\ln\left(2\left(\nu+\frac{1}{2}\right)\sin(\pi-\phi)\right)\right] + \frac{1}{i2(\nu+1/2)\sin\phi}\frac{\Gamma(1-iZ/k)}{\Gamma(iZ/k)}\exp\left[i\left(\nu+\frac{1}{2}\right)(\pi-\phi) + i\frac{Z}{k}\ln(2(\nu+1)\sin(\pi-\phi))\right]\right\}, \quad (4.6a)$$

and the asymptotic form of the ξ component of $\psi_k^{(+)}(\mathbf{r})$,

$$\lim_{\xi\to\infty}\psi_k^{(+)}(\xi) = \frac{1}{2^{1/4}\sqrt{2\pi}}\left\{\exp\left[-i\frac{k}{2}\xi - i\frac{Z}{k}\ln(k\xi)\right] + \frac{1}{ik\xi}\frac{\Gamma(1-iZ/k)}{\Gamma(iZ/k)}\exp\left[i\frac{k}{2}\xi + i\frac{Z}{k}\ln(k\xi)\right]\right\}. \quad (4.6b)$$

The normalization of a_{ν_1} and b_{ν_2} is verified by computing a_0 and b_0 by direct integration.

B. $m=0, Z_N = ik/2, Z_M = Z - ik/2$

For this choice of the separation constants, the coefficients a_{ν_1} and b_{ν_2} are the coefficients for expansion of $\psi_k^{(-)}(\mathbf{r})$ in the Sturmian basis set. Corresponding to these parameters, $t = \frac{1}{2}$, $\gamma_N = i\frac{1}{2}$, and $\gamma_M = Z/k - i\frac{1}{2}$. The corresponding regular solutions of the recursion are

$$a_{\nu_1} = \frac{2^{1/4}}{\sqrt{\pi\zeta}}\sin\left(\frac{\phi}{2}\right)\exp\left[i\frac{1}{2}(\phi-\pi)\right]e^{i\nu_1\phi} \quad (4.7a)$$

and

$$b_{\nu_2}=\frac{2^{1/4}}{\sqrt{\pi\xi}}\sin\left(\frac{\phi}{2}\right)\exp\left[-i\frac{1}{2}(\phi-\pi)\right]\exp\left[\frac{Z}{k}\left(\phi-\frac{\pi}{2}\right)\right]\Gamma(1+iZ/k)\exp(-i\nu_2\phi)\,{}_2F_1(-\nu_2,-iZ/k;1;1-e^{-i2\phi}). \tag{4.7b}$$

Once again, note the close correspondence between the asymptotic form of b_{ν_2},

$$\lim_{\nu_2\to\infty} b_{\nu_2}=\frac{2^{1/4}}{\sqrt{\pi\xi}}\sin\left(\frac{\phi}{2}\right)(-1)^{\nu}\left\{\exp\left[i\left(\nu+\frac{1}{2}\right)(\pi-\phi)+i\frac{Z}{k}\ln\left(2\left(\nu+\frac{1}{2}\right)\sin(\pi-\phi)\right)\right]\right.$$
$$\left.-\frac{1}{i2(\nu+1/2)\sin\phi}\frac{\Gamma(1+iZ/k)}{\Gamma(-iZ/k)}\exp\left[-i\left(\nu+\frac{1}{2}\right)(\pi-\phi)-i\frac{Z}{k}\ln\left(2\left(\nu+\frac{1}{2}\right)\sin(\pi-\phi)\right)\right]\right\}, \tag{4.8a}$$

and that of the η component of $\psi_{\mathbf{k}}^{(-)}(\mathbf{r})$,

$$\lim_{\eta\to\infty}\psi_{\mathbf{k}}^{(-)}(\eta)=\frac{1}{2^{1/4}\sqrt{2\pi}}\left\{\exp\left[i\frac{k}{2}\eta+i\frac{Z}{k}\ln(k\eta)\right]-\frac{1}{ik\eta}\frac{\Gamma(1+iZ/k)}{\Gamma(-iZ/k)}\exp\left[-i\frac{k}{2}\eta-i\frac{Z}{k}\ln(k\eta)\right]\right\}. \tag{4.8b}$$

As before, the normalization of a_{ν_1} and b_{ν_2} is verified by direct calculation of a_0 and b_0.

V. CONCLUSIONS

The preceding parabolic formulation of the Jacobi-matrix method may be used to study the Rydberg spectrum of hydrogenic atoms in a Stark field. This problem is of much current theoretical as well as experimental interest.[9] Classically, the motion is bounded in the η coordinate; at a given electric field $\mathcal{F}$, and at energies less than a critical energy $E_c(\mathcal{F})$, the motion in the ξ coordinate is also bounded. However, the quantum-mechanical spectrum is continuous at all energies due to tunneling. At energies below the classical critical energy for field ionization, the effective Stark potential in the ξ coordinate may also be approximated by a short-range potential, i.e., its matrix representation in the Sturmian basis may be truncated, and the Jacobi-matrix method is applicable. The resulting matrices being banded, one can use large basis sets and ensure effective completeness. At higher energies, further generalization would be necessary.

ACKNOWLEDGMENTS

A portion of this work was completed during a visit to the Queen's University of Belfast. I am most grateful to that institution, and to Professor P. G. Burke, for making available the necessary facilities.

[1] P. W. Langhoff, in *Methods in Computational Molecular Physics*, edited by G. H. F. Diercksen and S. Wilson (Riedel, New York, 1983).
[2] E. J. Heller and H. A. Yamani, Phys. Rev. A **9**, 1201 (1974); H. A. Yamani and L. Fishman, J. Math. Phys. **16**, 410 (1975).
[3] P. C. Ojha, Phys. Rev. A **34**, 969 (1986).
[4] This is a straightforward generalization of the Sturmian functions discussed by M. Rotenberg, Adv. At. Mol. Phys. **6**, 233 (1970).
[5] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[6] B. G. Adams, J. Čížek, and J. Paldus, Int. J. Quant. Chem. **21**, 153 (1982).
[7] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Natl. Bur. Stand., Washington, D.C., 1964); A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.
[8] This follows from the properties of the Γ functions and the three-term recursion relations for ${}_2F_1(a,b;c;z)$.
[9] S. Feneuille and P. Jacquinot, Adv. At. Mol. Phys. **17**, 99 (1981); C. W. Clark, K. T. Lu, and A. F. Starace, in *Progress in Atomic Spectroscopy*, edited by H. J. Beyer and H. Kleinpoppen (Plenum, New York, 1984), Part C.

Coherent angular momentum states for the two-dimensional oscillator

A. J. Bracken, D. S. McAnally, and O. A. Òdúndún[a)]
Department of Mathematics, University of Queensland, St. Lucia 4067, Brisbane, Australia

(Received 3 April 1986; accepted for publication 24 September 1986)

Coherent angular momentum states are defined for the two-dimensional isotropic harmonic oscillator. They share many attractive properties with the familiar (Cartesian) coherent states, but are in general distinct from those states. The probabilities of obtaining particular values for the radial and angular momentum quantum numbers follow independent Poisson distributions in the new states, but not in the old. In a quasiclassical description of the oscillator, corresponding to a given classical trajectory, the uncertainty in the angular momentum of the system is smaller if the new states are used rather than the old. The new states are the natural analogs of the coherent angular momentum states introduced for the three-dimensional oscillator by Bracken and Leemon [A. J. Bracken and H. I. Leemon, J. Math. Phys. **22**, 719 (1981)].

I. INTRODUCTION

Coherent angular momentum (CAM) states have been defined for the three-dimensional isotropic harmonic oscillator in quantum mechanics by Bracken and Leemon[1,2] (henceforth referred to as BL1 and BL2). Previously such states had been defined in various ways by various authors,[3] for systems with purely rotational degrees of freedom.

The Hamiltonian operator for the r-dimensional oscillator with mass μ and angular frequency ω is

$$H = (\mathbf{p}^2/2\mu) + \tfrac{1}{2}\mu\omega^2\mathbf{x}^2 = \hbar\omega(N + \tfrac{1}{2}r), \tag{1.1}$$

where

$$N = N_1 + N_2 + \cdots + N_r, \tag{1.2}$$

$$N_i = a_i^\dagger a_i \quad (i = 1,2,...,r, \text{ no sum}), \tag{1.3}$$

$$a_i = (2\mu\hbar\omega)^{-1/2}(ip_i + \mu\omega x_i). \tag{1.4}$$

The operators a_i and their Hermitian conjugates $a_i^\dagger$ satisfy the usual boson relations and so are lowering and raising operators for the operators N_i, which have non-negative integer eigenvalues n_i. For $r = 3$, it was shown in BL1 that we can write, as well as (1.2),

$$N = 2K + L, \tag{1.5}$$

where K and L have non-negative integer eigenvalues k and l, the radial and total angular momentum quantum numbers. Commuting lowering operators ν and λ for K and L were introduced, and CAM states defined as their common eigenvectors (with complex eigenvalues).

This procedure mirrors that used for the usual Cartesian coherent (CC) states which are eigenvectors of the lowering operators a_i, again with complex eigenvalues. The CAM states and CC states have many attractive properties in common, but they form quite distinct sets. A major difference is that, in a CAM state, the probabilities of obtaining particular k and l values follow independent Poisson distributions, whereas in a CC state, such a property holds instead for the values of n_1, n_2, and n_3. These properties are only shared by both sorts of states in the special cases corresponding to circular classical orbits, when in fact the CAM states and CC states can be identified with one another.

It was remarked in BL2 that the treatment given there should be capable of generalization to $r > 3$ dimensions. In fact it is clear from the solution of the eigenvalue problem for H in a spherical basis, that (1.5) holds for all $r \geqslant 2$, with K and L taking non-negative integral eigenvalues k and l in each case. Then k is in each case the radial quantum number while, for the r-dimensional oscillator, $l(l + r - 2)$ is the eigenvalue of the SO(r) Casimir operator $\frac{1}{2}L_{ij}L_{ij}$, where

$$L_{ij} = (x_i p_j - x_j p_i)/\hbar = i(a_i a_j^\dagger - a_j a_i^\dagger), \tag{1.6}$$

are the generators of SO(r). Thus

$$\tfrac{1}{2}L_{ij}L_{ij} = L(L + r - 2). \tag{1.7}$$

As remarked in BL2, it should be possible to define CAM states for $r > 3$ by identifying suitable commuting lowering operators for K and L in much the same way as for $r = 3$, and by then finding their common eigenvectors.

The case $r = 2$ is special, and in BL2 it was observed that, if CAM states are defined in this case as common eigenvectors of the familiar "angular" boson operators[4] for the two-dimensional oscillator,

$$\rho = 2^{-1/2}(a_1 - ia_2), \quad \sigma = 2^{-1/2}(a_1 + ia_2), \tag{1.8}$$

which are shift operators for the SO(2) generator/invariant $M(= L_{12})$, then the CC states and CAM states can be identified with each other. The common eigenvectors of ρ and σ are also eigenvectors of a_1 and a_2, as (1.8) shows. However, following the work of Òdúndún[5] (henceforth referred to as OD), it can be seen that the situation is not quite so simple.

In particular, the operator M has eigenvalues m running over all the integers while, as already mentioned, the eigenvalues l of L are non-negative. Indeed M, although an SO(2) scalar, is not the analog in two dimensions of the operator L when $r > 2$, since that L is not only an SO(r) scalar but also an O(r) scalar, like H, N, and K. In contrast, M is an O(2) pseudoscalar, changing sign when, for example, x_1 and p_1 are replaced by their negatives. In the two-dimensional case we

[a)] Permanent address: Department of Physics, University of Ife, Ile-Ife, Nigeria.

 0022-2488/87/020397-07$02.50 © 1987 American Institute of Physics

must in fact define L as an O(2) scalar by setting

$$L = |M|, \tag{1.9}$$

consistently with (1.7). Then we may ask if it is possible to define CAM states for $r = 2$ which are analogs of those for $r = 3$, as eigenvectors of lowering operators for K and L rather than eigenvectors of ρ and σ, which are shift operators for M and N.

Despite the close relationship betweek K, L, M, and N indicated above, and the simple algebraic structure of the two-dimensional oscillator, we shall show that CAM states can indeed be defined in this way, and that they differ fundamentally from the CC states, much as in the three-dimensional case. It is surprising to find new coherent states for so well-known and simple a system, and their study is of interest not only because of the importance of the two-dimensional oscillator in applications, but also because the simplicity of the algebraic structure makes their properties easier to appreciate than in higher-dimensional cases.

Lowering operators ν and λ for K and L were defined in OD for $r = 2$, with λ a two-vector operator. However, the algebraic structure is so simple in the two-dimensional case that it is convenient in what follows to do away with the SO(2) tensor rotation entirely and work always with one component entities.

II. SHIFT OPERATORS FOR K AND L

For the Hilbert space $\mathfrak{H}$ of the two-dimensional oscillator, let P_+, P_0, P_- be projectors onto the mutually orthogonal subspaces $\mathfrak{H}_+$, $\mathfrak{H}_0$, $\mathfrak{H}_-$ on which M has positive, zero, and negative eigenvalues, respectively. Evidently these projectors sum to unity and commute with K, L, M, and N. Since

$$\begin{aligned} &LP_0 = MP_0 = 0, \\ &LP_+ = MP_+, \quad LP_- = -MP_-, \end{aligned} \tag{2.1}$$

it follows from (1.9) that

$$L = M(P_+ - P_-), \tag{2.2}$$

and then, from (1.5), that

$$\begin{aligned} K &= \tfrac{1}{2}N - \tfrac{1}{2}M(P_+ - P_-) \\ &= \tfrac{1}{2}(N - M)P_+ + \tfrac{1}{2}(N + M)P_- + \tfrac{1}{2}NP_0. \end{aligned} \tag{2.3}$$

The operator σ of (1.8) lowers N and raises M by one unit, that is to say,

$$N\sigma = \sigma(N - 1), \quad M\sigma = \sigma(M + 1). \tag{2.4}$$

It follows from the second of these equations and the definitions of $P_\pm$, P_0 that

$$\sigma(P_+ + P_0) = P_+\sigma, \quad \sigma P_- = (P_0 + P_-)\sigma. \tag{2.5}$$

Similarly,

$$\rho(P_0 + P_-) = P_-\rho, \quad \rho P_+ = (P_+ + P_0)\rho. \tag{2.6}$$

The relations Hermitian conjugate to these must also hold:

$$\begin{aligned} &(P_+ + P_0)\sigma^\dagger = \sigma^\dagger P_+, \quad \sigma^\dagger(P_0 + P_-) = P_-\sigma^\dagger, \\ &(P_0 + P_-)\rho^\dagger = \rho^\dagger P_-, \quad \rho^\dagger(P_+ + P_0) = P_+\rho^\dagger. \end{aligned} \tag{2.7}$$

Because σ lowers N and raises M, and because $K = \frac{1}{2}(N + M)$ and $L = -M$ on $\mathfrak{H}_-$, it follows that, when applied to vectors in $\mathfrak{H}_-$, σ lowers L by one unit and commutes with K. Therefore

$$L\sigma P_- = \sigma P_-(L - 1), \quad K\sigma P_- = \sigma P_- K, \tag{2.8}$$

which can be verified directly from (2.1) and (2.3)–(2.5). Similarly,

$$L\rho P_+ = \rho P_+(L - 1), \quad K\rho P_+ = \rho P_+ K. \tag{2.9}$$

Noting further than any lowering operator for L must vanish on $\mathfrak{H}_0$, where L has its least value, we therefore identify

$$\lambda = f(K,L)\,\rho P_+ + g(K,L)\sigma P_- \tag{2.10}$$

as the general form of the operator we seek which lowers L by one unit while commuting with K. In (2.10), the functions f and g are arbitrary at this stage. The Hermitian conjugate operator $\lambda^\dagger$, which raises L by one unit and commutes with K, is then

$$\begin{aligned} \lambda^\dagger &= P_+\rho^\dagger f(K,L)^\dagger + P_-\sigma^\dagger g(K,L)^\dagger \\ &= f(K,L-1)^\dagger\rho^\dagger(P_+ + P_0) \\ &\quad + g(K,L-1)^\dagger\sigma^\dagger(P_0 + P_-), \end{aligned} \tag{2.11}$$

using the shifting properties of $P_+\rho^\dagger$ and $P_-\sigma^\dagger$, and the relations (2.7).

The O(2) scalar $(a_1)^2 + (a_2)^2 (= 2\rho\sigma)$ commutes with M (and hence with L) and lowers N by two units; it therefore lowers K by one unit, according to (1.5). The most general operator with these shifting properties for K and L is then

$$\nu = h(K,L)\,\rho\sigma, \tag{2.12}$$

where h is arbitrary. The conjugate operator, which raises K by one unit and commutes with L, is

$$\nu^\dagger = \sigma^\dagger\rho^\dagger h(K,L)^\dagger = h(K-1,L)^\dagger\rho^\dagger\sigma^\dagger. \tag{2.13}$$

We now choose the functions f, g, and h so as to make as simple as possible the commutation and other relations amongst λ, $\lambda^\dagger$, ν, and $\nu^\dagger$, and arrive at the expressions

$$\begin{aligned} &\nu = (K + L + 1)^{-1/2}\rho\sigma, \quad \nu^\dagger = \rho^\dagger\sigma^\dagger(K + L + 1)^{-1/2}, \\ &\lambda = [(L + 1)/(K + L + 1)]^{1/2}(\rho P_+ + \sigma P_-), \\ &\lambda^\dagger = (P_+\rho^\dagger + P_-\sigma^\dagger)[(L + 1)/(K + L + 1)]^{1/2}. \end{aligned} \tag{2.14}$$

It can be checked that these operators satisfy not only

$$\begin{aligned} &K\nu = \nu(K - 1), \quad K\nu^\dagger = \nu^\dagger(K + 1), \\ &[L,\nu] = 0 = [L,\nu^\dagger], \\ &L\lambda = \lambda(L - 1), \quad L\lambda^\dagger = \lambda^\dagger(L + 1), \\ &[K,\lambda] = 0 = [K,\lambda^\dagger], \end{aligned} \tag{2.15}$$

but also

$$\begin{aligned} &[\nu,\lambda] = [\nu,\lambda^\dagger] = [\nu^\dagger,\lambda] = [\nu^\dagger,\lambda^\dagger] = 0, \\ &\nu^\dagger\nu = K, \quad \nu\nu^\dagger = K + 1, \quad [\nu,\nu^\dagger] = 1, \\ &\lambda^\dagger\lambda = L + Q, \quad \lambda\lambda^\dagger = L + 1 + P_0, \\ &[\lambda,\lambda^\dagger] = 1 + P_0 - Q, \end{aligned} \tag{2.16}$$

where

$$Q = (K + 1)^{-1}(P_+\rho^\dagger\sigma P_- + P_-\sigma^\dagger\rho P_+). \tag{2.17}$$

This operator Q is nonzero only on vectors that are eigenvectors of L with $l = 1$. (It interchanges any vector correspond-

ing to $m = \pm 1$ with a vector corresponding to $m = \mp 1$.) It follows that λ and λ^+ satisfy boson relations on that part of $\mathfrak{H}$ spanned by eigenvectors of K and L with $l > 1$: that they do not satisfy such relations on *all* of $\mathfrak{H}$ is associated with the fact that the common eigenvectors of K and L are nondegenerate for $l = 0$, but doubly degenerate for $l > 0$. The expressions (2.14) are determined from (2.10)–(2.13), up to unitary transformations, by the relations (2.16).

The structure of the operators (2.14) is better appreciated if one considers their action on a familiar basis for $\mathfrak{H}$, provided by the common eigenvectors $|r,s\rangle$ of the operators $\rho^\dagger\rho[=\frac{1}{2}(N+M)]$ and $\sigma^\dagger\sigma[=\frac{1}{2}(N-M)]$. These vectors are given by

$$|r,s\rangle = [r!s!]^{-1/2}(\rho^\dagger)^r(\sigma^\dagger)^s|0\rangle, \quad r,s = 0,1,2,..., \tag{2.18}$$

where $|0\rangle$ is the normalized vector corresponding to the oscillator ground state, satisfying

$$\rho|0\rangle = \sigma|0\rangle = 0. \tag{2.19}$$

The vectors $|r,s\rangle$ with $r > s$ span $\mathfrak{H}_+$, those with $r = s$ span $\mathfrak{H}_0$, and those with $r < s$ span $\mathfrak{H}_-$. It follows from the definitions given above that, if $r > s$,

$$\begin{aligned}
&N|r,s\rangle = (r+s)|r,s\rangle, \quad K|r,s\rangle = s|r,s\rangle,\\
&L|r,s\rangle = M|r,s\rangle = (r-s)|r,s\rangle,\\
&P_+|r,s\rangle = |r,s\rangle, \quad P_-|r,s\rangle = P_0|r,s\rangle = 0,\\
&\nu|r,s\rangle = s^{1/2}|r-1,s-1\rangle,\\
&\nu^\dagger|r,s\rangle = (s+1)^{1/2}|r+1,s+1\rangle,\\
&\lambda|r,s\rangle = (r-s)^{1/2}|r-1,s\rangle,\\
&\lambda^\dagger|r,s\rangle = (r-s+1)^{1/2}|r+1,s\rangle,
\end{aligned} \tag{2.20}$$

while, if $r < s$,

$$\begin{aligned}
&N|r,s\rangle = (r+s)|r,s\rangle, \quad K|r,s\rangle = r|r,s\rangle,\\
&L|r,s\rangle = -M|r,s\rangle = (s-r)|r,s\rangle,\\
&P_-|r,s\rangle = |r,s\rangle, \quad P_+|r,s\rangle = P_0|r,s\rangle = 0,\\
&\nu|r,s\rangle = r^{1/2}|r-1,s-1\rangle,\\
&\nu^\dagger|r,s\rangle = (r+1)^{1/2}|r+1,s+1\rangle,\\
&\lambda|r,s\rangle = (s-r)^{1/2}|r,s-1\rangle,\\
&\lambda^\dagger|r,s\rangle = (s-r+1)^{1/2}|r,s+1\rangle.
\end{aligned} \tag{2.21}$$

Finally, if $r = s$,

$$\begin{aligned}
&N|r,r\rangle = 2r|r,r\rangle, \quad K|r,r\rangle = r|r,r\rangle,\\
&L|r,r\rangle = M|r,r\rangle = 0,\\
&P_0|r,r\rangle = |r,r\rangle, \quad P_+|r,r\rangle = P_-|r,r\rangle = 0,\\
&\nu|r,r\rangle = r^{1/2}|r-1,r-1\rangle,\\
&\nu^\dagger|r,r\rangle = (r+1)^{1/2}|r+1,r+1\rangle,\\
&\lambda|r,r\rangle = 0, \quad \lambda^\dagger|r,r\rangle = |r+1,r\rangle + |r,r+1\rangle.
\end{aligned} \tag{2.22}$$

III. EIGENVECTORS OF THE LOWERING OPERATORS

We now define CAM states as common eigenvectors $|z,\zeta\rangle$ of the lowering operators ν and λ,

$$\nu|z,\zeta\rangle = z|z,\zeta\rangle, \quad \lambda|z,\zeta\rangle = \zeta|z,\zeta\rangle, \tag{3.1}$$

and we seek each such vector in the form

$$\begin{aligned}
|z,\zeta\rangle &= \sum_{r,s=0}^{\infty} c_{rs}|r,s\rangle\\
&= \left\{\sum_{0\le s<r} + \sum_{r=s} + \sum_{0\le r<s}\right\} c_{rs}|r,s\rangle.
\end{aligned} \tag{3.2}$$

Then, given the action of ν as in (2.20)–(2.22) and the orthogonality of the vectors $|r,s\rangle$, the first of Eqs. (3.1) yields

$$\begin{aligned}
zc_{rs} &= (s+1)^{1/2}c_{r+1s+1}, \quad 0\le s<r,\\
zc_{rr} &= (r+1)^{1/2}c_{r+1r+1}, \quad 0\le r,\\
zc_{rs} &= (r+1)^{1/2}c_{r+1s+1}, \quad 0\le r<s.
\end{aligned} \tag{3.3}$$

Similarly, the second of Eqs. (3.1) yields

$$\begin{aligned}
\zeta c_{rs} &= (r-s+1)^{1/2}c_{r+1s}, \quad 0\le s<r,\\
\zeta c_{rr} &= c_{r+1r} + c_{rr+1}, \quad 0\le r,\\
\zeta c_{rs} &= (s-r+1)^{1/2}c_{rs+1}, \quad 0\le r<s.
\end{aligned} \tag{3.4}$$

Equations (3.3) and (3.4) are straightforward to solve and give, if $\zeta \neq 0$ and z are any complex numbers

$$\begin{aligned}
c_{rs} &= \alpha z^s\zeta^{r-s}[s!(r-s)!]^{-1/2}, \quad 0\le s<r<\infty,\\
c_{rs} &= (\alpha+\beta)z^r[r!]^{-1/2}, \quad 0\le r<\infty,\\
c_{rs} &= \beta z^r\zeta^{s-r}[r!(s-r)!]^{-1/2}, \quad 0\le r<s<\infty,
\end{aligned} \tag{3.5}$$

where α and β may vary arbitrarily with z and ζ, but are independent of r and s. The special case $\zeta = 0$ gives, for any complex z,

$$\begin{aligned}
c_{r+1r} &= -c_{rr+1} = \gamma z^r[r!]^{-1,2}, \quad 0\le r<\infty,\\
c_{rr} &= \delta z^r[r!]^{-1/2}, \quad 0\le r<\infty,
\end{aligned} \tag{3.6}$$

with γ and δ arbitrary. In this case

$$c_{rs} = 0, \quad |r-s| > 1. \tag{3.7}$$

Thus we have, if $\zeta \neq 0$,

$$\begin{aligned}
|z,\zeta\rangle = {}&\alpha\sum_{0\le s<r} z^s\zeta^{r-s}[s!(r-s)!]^{-1/2}|r,s\rangle\\
&+\beta\sum_{0\le r<s} z^r\zeta^{s-r}[r!(s-r)!]^{-1/2}|r,s\rangle,
\end{aligned} \tag{3.8}$$

and, if $\zeta = 0$,

$$\begin{aligned}
|z,0\rangle = {}&\gamma\sum_{r=0}^{\infty} z^r[r!]^{-1/2}(|r+1,r\rangle - |r,r+1\rangle)\\
&+\delta\sum_{r=0}^{\infty} z^r[r!]^{-1/2}|r,r\rangle.
\end{aligned} \tag{3.9}$$

The appearance of two arbitrary constants in the expression for $|z,\zeta\rangle$, whether or not $\zeta = 0$, indicates that some operator other than ν and λ can be diagonalized on these vectors in order to complete their specification. Setting first β and γ, and then α and γ, equal to zero we may define

$$\begin{aligned}
|z,\zeta,+\rangle &= A\sum_{0\le s<r} z^s\zeta^{r-s}[s!(r-s)!]^{-1/2}|r,s\rangle,\\
|z,\zeta,-\rangle &= B\sum_{0\le r<s} z^r\zeta^{s-r}[r!(s-r)!]^{-1/2}|r,s\rangle,
\end{aligned} \tag{3.10}$$

whether or not $\zeta = 0$. (Then $|z,0,+\rangle = |z,0,-\rangle = |z,0\rangle$.) It can then be seen from Eqs. (2.20)–(2.22) that, in addition to Eqs. (3.1), these vectors satisfy

$$P_\mp|z,\zeta,\pm\rangle = 0. \tag{3.11}$$

Despite the fact that no vector of the form (3.9) with $\gamma \neq 0$ appears in the set (3.10), we shall see that this set is in fact overcomplete, and that the states $|z,\zeta \pm\rangle$ have properties which justify their identification as the natural CAM states for the two-dimensional oscillator. The values of A and B in (3.10) are determined (up to unimportant phases) by normalization of $|z,\zeta,\pm\rangle$ to unit length, as

$$A = B = \exp[-\tfrac{1}{2}(|z|^2 + |\zeta|^2)]. \tag{3.12}$$

(*Aside*: As a possible procedure alternative to that leading from (3.8)–(3.9) to the vectors (3.10), we could have first set $\alpha = \beta$ in (3.8) and $\gamma = 0$ in (3.9), and then $\alpha = -\beta$ and $\delta = 0$, in order to obtain, if $\zeta \neq 0$,

$$|z,\zeta\rangle_{\text{even}} = C \sum_{0<s<r} z^s \zeta^{r-s} [s!(r-s)!]^{-1/2} (|r,s\rangle + |s,r\rangle),$$
$$|z,\zeta\rangle_{\text{odd}} = D \sum_{0<s<r} z^s \zeta^{r-s} [s!(r-s)!]^{-1/2} (|r,s\rangle - |s,r\rangle), \tag{3.13}$$

and, if $\zeta = 0$,

$$|z,0\rangle_{\text{even}} = E \sum_{r=0}^{\infty} z^r [r!]^{-1/2} |r,r\rangle,$$
$$|z,0\rangle_{\text{odd}} = F \sum_{r=0}^{\infty} z^r [r!]^{-1/2} (|r+1,r\rangle - |r,r+1\rangle). \tag{3.14}$$

These vectors satisfy, in addition to Eqs. (3.1) and whether or not $\zeta = 0$,

$$T|z,\zeta\rangle_{\text{even}} = |z,\zeta\rangle_{\text{even}},$$
$$T|z,\zeta\rangle_{\text{odd}} = -|z,\zeta\rangle_{\text{odd}}, \tag{3.15}$$

where T is the Hermitian operator, defined by

$$T|r,s\rangle = |s,r\rangle, \quad r,s = 0,1,2,..., \tag{3.16}$$

which interchanges the operators ρ and σ of (1.8)

$$T\rho T = \sigma, \quad T = T^{-1}, \tag{3.17}$$

and changes the sign of M

$$TMT = -M. \tag{3.18}$$

[In a treatment of the two-dimensional oscillator using a Hilbert space of functions f of polar variables (r,θ), the operator T acts as

$$Tf(r,\theta) = f(r,-\theta). \tag{3.19}$$

The vectors $|z,\zeta\rangle_{\text{even}}$ and $|z,\zeta\rangle_{\text{odd}}$ would then be represented by functions even and odd, respectively, in θ.] The constants C, D, E, and F in (3.13)–(3.14) can be fixed by normalizing the corresponding vectors, as

$$C = [2\exp(|z|^2)(\exp|\zeta|^2 + 1)]^{-1/2},$$
$$D = [2\exp(|z|^2)(\exp|\zeta|^2 - 1)]^{-1/2}, \tag{3.20}$$
$$E = \exp[-\tfrac{1}{2}|z|^2] = 2^{1/2}F.$$

This alternative set of vectors, defined by (3.13), (3.14), and (3.20), also has interesting mathematical properties, and could also be considered as a set of candidate CAM states. However, since the angular momentum operator $\hbar M$ has zero expectation value for all such states, as is easily checked, it seems that they could not provide a quasiclassical description of an oscillator with a nonzero angular momentum. We do not consider these vectors further.)

IV. PROPERTIES OF CAM STATES

We shall now list some properties of the CAM states $|z,\zeta,\epsilon\rangle$ ($\epsilon = \pm$) defined by (3.10) and (3.12). They may be compared with properties[6] of the familiar CC states $|u,v\rangle$ defined by

$$|u,v\rangle = \exp[-\tfrac{1}{2}(|u|^2 + |v|^2)] \sum_{r,s=0}^{\infty} u^r v^s [r!s!]^{-1/2} |r,s\rangle, \tag{4.1}$$

for arbitrary complex u and v. These vectors $|u,v\rangle$ satisfy

$$\rho|u,v\rangle = u|u,v\rangle, \quad \sigma|u,v\rangle = v|u,v\rangle, \tag{4.2}$$

and hence, by (1.8),

$$a_1|u,v\rangle = 2^{-1/2}(u+iv)|u,v\rangle = z_1|u,v\rangle,$$
$$a_2|u,v\rangle = 2^{-1/2}(u-iv)|u,v\rangle = z_2|u,v\rangle. \tag{4.3}$$

Note that the vectors (4.1) and (3.10) are quite distinct, except when $z = 0$. In that special case,

$$|z=0,\zeta,+\rangle = |u=\zeta, v=0\rangle,$$
$$|z=0,\zeta,-\rangle = |u=0, v=\zeta\rangle. \tag{4.4}$$

Because the derivations of the properties of the CAM states are quite similar to those for the three-dimensional case, as given in BL2, we shall omit such details.

A. Expectation values in the state $|z,\zeta,\epsilon\rangle$

$$\langle\lambda\rangle = \zeta, \quad \langle\lambda^\dagger\rangle = \zeta^*, \quad \langle L\rangle = |\zeta|^2, \quad \langle M\rangle = \epsilon|\zeta|^2,$$
$$\langle\nu\rangle = z, \quad \langle\nu^\dagger\rangle = z^*, \quad \langle K\rangle = |z|^2, \tag{4.5}$$
$$\langle H\rangle = \hbar\omega(2|z|^2 + |\zeta|^2 + 1).$$

[In the CC state $|u,v\rangle$ we have

$$\langle N_1\rangle = |z_1|^2 = \tfrac{1}{2}(|u|^2 + |v|^2 - iuv^* + iu^*v),$$
$$\langle N_2\rangle = |z_2|^2 = \tfrac{1}{2}(|u|^2 + |v|^2 + iuv^* - iu^*v), \tag{4.6}$$
$$\langle H\rangle = \hbar\omega(|u|^2 + |v|^2 + 1).]$$

B. Poisson distributions

The probability $p(k,l)$ of obtaining simultaneously the values k and l for K and L in the state $|z,\zeta,\epsilon\rangle$ is given by

$$p(k,l) = (|z|^{2k}/k!)e^{-|z|^2}(|z|^{2l}/l!)e^{-|\zeta|^2}. \tag{4.7}$$

The unconditional probabilities $p(k)$ and $p(l)$ of obtaining values k for K or l for L therefore follow independent Poisson distributions with means $|z|^2$ and $|\zeta|^2$, respectively. (In constrast, for the CC state $|u,v\rangle$, a result analogous to (4.7) holds instead for the probability $p(n_1,n_2)$ of obtaining values n_1, n_2 on measuring N_1 and N_2, with z_1, z_2 [as in (4.3)] replacing z,ζ, and n_1, n_2 replacing k,l.)

C. Minimum uncertainty

Let

$$\nu = 2^{-1/2}(a+ib), \quad \lambda = 2^{-1/2}(\alpha + i\beta), \tag{4.8}$$

where a, b, α, and β are Hermitian. In a general state, let $\Delta a = (\langle a^2\rangle - \langle a\rangle^2)^{1/2}$, etc. Then it follows from (2.16) that

$$\Delta a \Delta b \geqslant \tfrac{1}{2}, \quad \Delta\alpha\Delta\beta \geqslant \tfrac{1}{2}(1 + \langle P_0\rangle - \langle Q\rangle). \tag{4.9}$$

However, in the CAM state $|z,\zeta,\epsilon\rangle$,

$$\Delta a \Delta b = \tfrac{1}{2}, \qquad \Delta\alpha\Delta\beta = \tfrac{1}{2}(1 + \langle P_0\rangle - \langle Q\rangle) \ [= \tfrac{1}{2}(1 + e^{-|\zeta|^2})]. \tag{4.10}$$

The CAM states are therefore minimum uncertainty states for a, b, α, and β. The equalities (4.10) do *not* hold in the CC state $|u,v\rangle$, which is therefore not a minimum uncertainty state for these variables (unless $u = 0$ or $v = 0$, when the CC state is also a CAM state). [On the other hand, the equalities

$$\Delta x_1 \Delta p_1 = \tfrac{1}{2}\hbar = \Delta x_2 \Delta p_2 \tag{4.11}$$

hold in a CC state but not (unless $z = 0$) in a CAM state. The latter are not in general minimum uncertainty states for the Cartesian variables $\mathbf{x}$ and $\mathbf{p}$.]

D. Evolution in the Schrödinger picture

If the state vector $|\psi(t)\rangle$ of the oscillator is a CAM state at one time, then it is so at all subsequent times t. Thus, if

$$|\psi(0)\rangle = |z_0,\zeta_0,\epsilon\rangle, \tag{4.12}$$

then

$$|\psi(t)\rangle = e^{-i\omega t}|z(t),\zeta(t),\epsilon\rangle, \quad t>0, \qquad z(t) = z_0 e^{-2i\omega t}, \quad \zeta(t) = \zeta_0 e^{-i\omega t}. \tag{4.13}$$

The expectation values $z(t), \zeta(t)$ of ν, λ then follow a trajectory of the corresponding classical variables $\hat{\nu}, \hat{\lambda}$ in a "complex phase space" $(C \times C, \epsilon)$. We can associate a volume of uncertainty $\Delta a \Delta b \Delta\alpha\Delta\beta$ with the representative point in this space, subject to inequalities (4.9). When the motion proceeds through a succession of CAM states as in (4.13), this volume is maintained at a constant minimal value.

E. Uncertainty in angular momentum

Consider a typical classical trajectory of the oscillator, for which the classical analogs $\hat{\mathbf{x}}, \hat{\mathbf{p}}$ of the operators $\mathbf{x}$ and $\mathbf{p}$ are given by

$$\hat{\mathbf{x}} = (A\cos\omega t,\ B\sin\omega t), \qquad \hat{\mathbf{p}} = \mu\omega(-A\sin\omega t,\ B\cos\omega t), \tag{4.14}$$

with A, B real and $A \geqslant |B| \geqslant 0$. The angular momentum $\hat{M}$ corresponding to the operator $\hbar M$) is, for this trajectory,

$$\hat{M}(= \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1) = \mu\omega AB, \tag{4.15}$$

and the classical analogs of ν and λ take the values

$$\hat{\nu} = \tfrac{1}{2}(\mu\omega)^{1/2}(A - |B|)e^{-2i\omega t}, \qquad \hat{\lambda} = (\mu\omega A|B|)^{1/2}e^{-i\omega t}. \tag{4.16}$$

Comparison with (4.5) shows that a quasiclassical description of this motion is given in terms of CAM states if we take

$$z_0 = \tfrac{1}{2}(\mu\omega/\hbar)^{1/2}(A - |B|), \quad \zeta_0 = (\mu\omega A|B|/\hbar)^{1/2}, \tag{4.17}$$

and choose ϵ to match the sign of the angular momentum, or equivalently, from (4.15), the sign of B. (If $B = 0$, we must take $\zeta_0 = 0$ and ϵ is then meaningless; recall that $|z,0,+\rangle = |z,0,-\rangle$.) A measure of the (constant) uncertainty in the angular momentum of the quantum oscillator is then given by $(\hbar\Delta M)^2$, and we find that

$$(\hbar\Delta M)^2 = \hbar\mu\omega A|B|. \tag{4.18}$$

An alternative quasiclassical description of the motion (4.14) using CC states is given in terms of the state vector

$$|\chi(t)\rangle = e^{i\omega t}|u(t),v(t)\rangle, \qquad u(t) = u_0 e^{-i\omega t}, \quad v(t) = v_0 e^{-i\omega t}, \qquad u_0 = \tfrac{1}{2}(\mu\omega/\hbar)^{1/2}(A - B), \quad v_0 = \tfrac{1}{2}(\mu\omega/\hbar)^{1/2}(A + B). \tag{4.19}$$

In this case we find

$$(\hbar\Delta M)^2 = \tfrac{1}{2}\hbar\mu\omega(A^2 + B^2). \tag{4.20}$$

We see that the (squared) uncertainty in the angular momentum is greater in general for the description using CC states than for the one using CAM states, by an amount $\frac{1}{2}\hbar\mu\omega(A - |B|)^2$. (In the case of a circular orbit, $A = |B|$, the CAM states and CC states coincide and the two descriptions become one.)

F. Classical limit

This can be treated very simply by considering at each time t a succession of states $|z,\zeta,\epsilon\rangle$, with $|z| \to \infty$, $|\zeta| \to \infty$ and $\hbar \to 0$ in such a way that $(\hbar)^{1/2}z$ and $(\hbar)^{1/2}\zeta$ remain finite [and equal to $\frac{1}{2}(\mu\omega)^{1/2}(A - |B|)e^{-2i\omega t}$, $(\mu\omega A|B|)^{1/2}e^{-i\omega t}$, respectively, for the case of the typical orbit described by (4.14)]. The sign of ϵ must be chosen to match that of the classical angular momentum. The case of a circular orbit is special and corresponds always to $z = 0$. It is not difficult to check that, as the limit is approached, the relative uncertainties in H, K, L, M, a, b, α, and β all go to zero. (In a treatment in terms of CC states, one considers instead a succession of states $|u,v\rangle$ with $|u| \to \infty$, $|v| \to \infty$, $\hbar \to 0$, and $(\hbar)^{1/2}u$, $(\hbar)^{1/2}v$ finite [and equal to $\frac{1}{2}(\mu\omega)^{1/2}(A - B)e^{-i\omega t}$, $\frac{1}{2}(\mu\omega)^{1/2}(A + B)e^{-i\omega t}$ for the orbit (4.14)].)

G. Nonorthogonality

The CAM states are not mutually orthogonal. Instead,

$$|\langle z',\zeta',\epsilon|z,\zeta,\epsilon\rangle|^2 = \exp(-|z' - z|^2 - |\zeta' - \zeta|^2), \tag{4.21}$$

$$|\langle z',\zeta',-\epsilon|z,\zeta,\epsilon\rangle|^2 = \exp(-|z' - z|^2 - |\zeta'|^2 - |\zeta|^2),$$

so that orthogonality is approached as $|z' - z|$ and $|\zeta' - \zeta|$ become large. [Similarly, for the CC states,

$$|\langle u',v'|u,v\rangle|^2 = \exp(-|u' - u|^2 - |v' - v|^2).] \tag{4.22}$$

H. Overcompleteness

The CAM states are overcomplete in $\mathfrak{H}$. Completeness is most conveniently expressed in terms of the nonunit vectors

$$|z,\zeta,\epsilon) = \exp(\tfrac{1}{2}|z|^2 + \tfrac{1}{2}|\zeta|^2)|z^*,\zeta^*,\epsilon\rangle, \tag{4.23}$$

as

$$\int (|z,\zeta,+)(z,\zeta,+| + |z,\zeta,-)(z,\zeta,-| - |z,0)(z,0|)d\rho(z,\zeta) = I, \tag{4.24}$$

where I is the unit operator,

$$d\rho(z,\zeta) = (1/\pi^2)\exp(-|z|^2 - |\zeta|^2)d^2z\,d^2\zeta, \tag{4.25}$$

and the integrals run over all complex values of z and ζ.

Let $|\phi\rangle$ be an arbitrary vector in $\mathfrak{H}$, with expansion in terms of the vectors $|r,s\rangle$ of (2.18) as

$$|\phi\rangle = \sum_{r,s=0}^{\infty} \phi_{rs}|r,s\rangle. \tag{4.26}$$

Then, let

$$\phi^{(+)}(z,\zeta) = (z,\zeta,+|\phi\rangle = \sum_{0\le s\le r} \phi_{rs}[s!(r-s)!]^{-1/2} z^{\,s}\zeta^{\,r-s},$$
$$\phi^{(-)}(z,\zeta) = (z,\zeta,-|\phi\rangle \tag{4.27}$$
$$= \sum_{0\le r\le s} \phi_{rs}[r!(s-r)!]^{-1/2} z^r \zeta^{\,s-r},$$

and note that

$$\phi^{(+)}(z,0) = \phi^{(-)}(z,0) = \phi_0(z). \tag{4.28}$$

The function $\phi^{(\epsilon)}(z,\zeta)$ is entire on $C\times C$, and satisfies

$$|\phi^{(\epsilon)}(z,\zeta)| \le (\langle\phi|\phi\rangle)^{1/2} \exp\tfrac{1}{2}(|z|^2 + |\zeta|^2), \tag{4.29}$$

and

$$\int |\phi^{(\epsilon)}(z,\zeta)|^2\, d\rho < \infty. \tag{4.30}$$

Furthermore,

$$\langle\phi|\phi\rangle = \int (|\phi^{(+)}(z,\zeta)|^2 + |\phi^{(-)}(z,\zeta)|^2 - |\phi_0(z)|^2) d\rho, \tag{4.31}$$

and

$$|\phi\rangle = \int (\phi^{(+)}(z,\zeta)|z,\zeta,+\rangle + \phi^{(-)}(z,\zeta)|z,\zeta,-\rangle - \phi_0(z)|z,0\rangle) d\rho. \tag{4.32}$$

Conversely, given any pair of entire functions $\phi^{(\epsilon)}(z,\zeta)$, $\epsilon = \pm$, which satisfy (4.28)–(4.30), we can define a vector $|\phi\rangle$ in $\mathfrak{H}$ by (4.32), check that its squared length is given by (4.31), and that

$$\phi^{(\epsilon)}(z,\zeta) = (z,\zeta,\epsilon|\phi\rangle. \tag{4.33}$$

I. Reproducing kernel Hilbert space

Equations (4.27)–(4.32) establish a 1–1 correspondence between vectors in $\mathfrak{H}$ and pairs of functions $\phi^{(\epsilon)}(z,\zeta)$. Accordingly, a realization of $\mathfrak{H}$ is provided by taking the pair $\Phi(z,\zeta) = \{\phi^{(+)}(z,\zeta),\phi^{(-)}(z,\zeta)\}$ as the representative of the abstract vector $|\phi\rangle$, and identifying the scalar product of two pairs Φ, Ψ as

$$(\Phi,\Psi) = \int (\phi^{(+)}(z,\zeta)^*\psi^{(+)}(z,\zeta) + \phi^{(-)}(z,\zeta)^*\psi^{(-)}(z,\zeta) - \phi_0(z)^*\psi_0(z)) d\rho \tag{4.34}$$

(equal in fact to $\langle\phi|\psi\rangle$). In this realization of $\mathfrak{H}$, we find

$$\nu = \frac{\partial}{\partial z}, \quad \nu^\dagger = z, \quad K = z\frac{\partial}{\partial z},$$
$$\lambda^\dagger = \zeta, \quad L = \zeta\frac{\partial}{\partial\zeta}, \tag{4.35}$$

while λ and M act as

$$\lambda\Phi = \left\{\frac{\partial\phi^{(+)}}{\partial\zeta} + \frac{\partial\phi^{(-)}}{\partial\zeta}\bigg|_{\zeta=0}, \frac{\partial\phi^{(-)}}{\partial\zeta} + \frac{\partial\phi^{(+)}}{\partial\zeta}\bigg|_{\zeta=0}\right\},$$
$$M\Phi = \left\{\zeta\frac{\partial\phi^{(+)}}{\partial\zeta}, -\zeta\frac{\partial\phi^{(-)}}{\partial\zeta}\right\}. \tag{4.36}$$

This realization of $\mathfrak{H}$ has a reproducing kernel

$$K(z',\zeta';z,\zeta) = \{k(z',\zeta';z,\zeta), k(z',\zeta';z,\zeta)\},$$
$$k(z',\zeta';z,\zeta) = \exp(z'^*z + \zeta'^*\zeta), \tag{4.37}$$

since

$$(K(z',\zeta';\cdot,\cdot),\phi) = \phi(z',\zeta'). \tag{4.38}$$

V. CONCLUDING REMARKS

We have identified CAM states for the two-dimensional oscillator. They have many properties in common with the familiar CC states, but are quite distinct from those states in general. The expressions (3.10) and (3.12) for the CAM states are of a simplicity comparable with that for the CC states, as in (4.1). Perhaps the most important distinguishing features of the CAM states, *vis-à-vis* the CC states, are those described in IV B and IV E. We can summarize by saying that CAM states have special properties in relation to the radial and angular momentum operators K and L for the oscillator while CC states have similar properties in relation to the number operators N_1 and N_2.

The CAM states could also have been called O(2) coherent states for the oscillator, in that they are eigenvectors of a lowering operator for the O(2) invariant L. [In contrast, the CC states could be called SO(2) coherent states, being eigenvectors of a lowering operator for the SO(2) invariant M.] However, the names "O(2) coherent states" and "CAM states" are both deficient to the extent that they do not reflect the fact that a lowering operator for the radial quantum number K is also diagonalized on these states. This is an important feature, associated with the fact that the two-dimensional oscillator has a radial as well as a rotational degree of freedom. Even in the two-dimensional case it is not a trivial matter to find suitable commuting lowering operators for the radial and angular momentum quantum numbers k and l, as in Sec. II. If this feature is overlooked, then the CAM states defined here (and those for the three-dimensional oscillator, defined in BL2) can easily be confused with those defined by many others[3] for systems, such as the rigid rotor, which do have only rotational degrees of freedom. This potential for confusion is compounded by the fact that the Schwinger boson calculus is often used to provide a convenient realization of the SO(2) < SO(3) group for such systems.

Bhaumik *et al.*[7] have defined "charged" coherent states, also different from CC states, for the two-dimensional oscillator. Their states are, however, also quite distinct from the CAM states defined here. In our notation (with their a and b replaced by our ρ and σ), their states are eigenvectors of M and $\rho\sigma$ ($= [K+L+1]^{1/2}\nu$). They are therefore labeled by a definite charge q (an integer, equal to our m) and a complex number ξ (essentially) equivalent to our z in (3.1) and (3.10), whereas a CAM state is labeled by two complex numbers. From the point of view of Ref. 7, a CAM state does not have a definite charge (though it does have a definite charge sign); rather, the probability of obtaining a given charge in such a state follows a Poisson distribution.

ACKNOWLEDGMENTS

A. J. B. thanks Dr. H. I. Leemon for helpful discussions. O. A. O. thanks Dr. V. G. Hart and members of his department for their kind hospitality.

[1]A. J. Bracken and H. I. Leemon, J. Math. Phys. **21**, 2170 (1980).
[2]A. J. Bracken and H. I. Leemon, J. Math. Phys. **22**, 719 (1981). See also A. J. Bracken and H. I. Leemon, *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 135, edited by K. B. Wolf (Springer, Berlin, 1980), p. 67.
[3]P. Bonifacio, D. M. Kim, and M. O. Scully, Phys. Rev. **187**, 441 (1969); P. W. Atkins and J. C. Dobson, Proc. R. Soc. London Ser. A **321**, 321 (1971); V. V. Mikhailov, Theor. Math. Phys. **15**, 584 (1973); D. Bhaumik, T. Nag, and B. Dutta-Roy, J. Phys. A: Math., Nucl. Gen. **8**, 1868 (1975); T. M. Makhviladze and L. A. Shelepin, *Proceedings of the P. N. Lebedev Physics Institute*, Vol. 70, edited by D. V. Skobeltsyn (Consultants Bureau, New York, 1975); J. Mostowski, Phys. Lett. A **56**, 369 (1976); R. Delbourgo, J. Phys. A: Math., Nucl. Gen. **10**, 1837 (1977); T. Santhanam, "Angular momentum coherent states," preprint, Australian National University, 1978; P. Gulshani, Can. J. Phys. **57**, 998 (1979). For a recent comprehensive review of coherent states, see J. R. Klauder and B.-S. Skagerstam, *Coherent States: Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).
[4]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1965), Vol. 1, p. 454.
[5]O. Òdúndún, J. Math. Phys. **24**, 2340 (1983); **27**, 1492(E) (1986).
[6]H. M. Nussenzveig, *Introduction to Quantum Optics* (Gordon and Breach, New York, 1973), Chap. 3.
[7]D. Bhaumik, K. Bhaumik, and B. Dutta-Roy, J. Phys. A: Math. Nucl. Gen. **9**, 1507 (1976). We are indebted to a referee for drawing this paper to our attention.

Comments on exact solutions for a one-dimensional periodic solid

G. Dattoli, J. C. Gallardo,[a] and A. Torre[b]
Comitato Nazionale per la Ricerca e per lo Sviluppo della Energia Nucleare e delle Energie Alternative, Centro Ricerche Energia Frascati, CP-65, 00044 Frascati, Italy

(Received 21 March 1986; accepted for publication 24 September 1986)

An alternative method to get exact solutions of the single-band time-dependent Schrödinger equation for an electron in a spatially periodic one-dimensional potential in the presence of a uniform electric field is presented.

I. INTRODUCTION

In a recent paper Luban[1] has presented an exact solution for a single-band time-dependent Schrödinger equation for an electron in a one-dimensional periodic solid in the presence of a constant uniform electric field. The technique of solution used by this author consists of mapping the Schrödinger equation of the problem to the exact solvable problem, in the eikonal approximation, of a quantum planar rotor subject to an arbitrary periodic time-dependent external potential.

In this paper we will propose an alternative method by treating the problem in a more direct way. In fact, in Ref. 1 it has been shown that the time-dependent Schrödinger equation of the relevant problem can be reduced to a differential difference equation. We will use a solution technique proposed by us in Ref. 2 to give an exact solution in terms of a combination of Bessel functions.

The Hamiltonian considered in Ref. 1 is

$$\hat{H}(x,t) = \hat{H}_0(x,t) + e\xi\hat{x}, \tag{1}$$

where $\hat{H}_0$ is an Hermitian operator, spatially periodic with period $a(\hat{H}_0(x+a,t) = \hat{H}_0(x,t))$. The electron charge is e and ξ is the magnitude of the electric field. The wave function can be expanded in terms of orthonormal Wannier functions; according to Ref. 1 we write ψ in the single band approximation as

$$\psi(x,t) = \sum_{n=-\infty}^{\infty} f_n(t)\langle x|n,l\rangle, \tag{2}$$

where l is the band index and $\langle x|n,l\rangle = \phi_l(x-na)$ is a set of Wannier functions satisfying the condition

$$\int_{-\infty}^{+\infty} dx\, \phi_l^*(x-na)\phi_l(x-n'a) = \delta_{l,l'}\delta_{n,n'}. \tag{3}$$

In the single-band approximation we replace the Hamiltonian Eq. (1) with the new operator

$$\widetilde{\hat{H}}_l = \sum_{n,n'=-\infty}^{+\infty} \langle n,l|\hat{H}|n',l\rangle |n,l\rangle\langle n',l|. \tag{4}$$

All the temporal behavior of the ψ-function is contained in the time-dependent coefficients $f_n(t)$ which have been shown to satisfy the following set of coupled homogeneous equations (see Ref. 1 for further comments):

$$i\frac{df_n}{d\tau} = nf_n + \sum_{n'=1}^{\infty}\{V_{n'}f_{n+n'} + V_{-n'}f_{n-n'}\}, \tag{5}$$

where

$$\tau = t/\tau_0, \quad \tau_0 = \hbar/(ea\xi), \quad V_{n'} = \langle 0|\hat{H}|n'\rangle\tau_0/\hbar. \tag{6}$$

II. GENERAL PROCEDURES

We will deal with Eq. (5) following a rather direct method which is an immediate generalization of the technique developed in Ref. 2.

As a first step, it is convenient to introduce the new function $C_n(\tau) = f_n(\tau)e^{-in\tau}$ thus getting

$$i\frac{dC_n}{d\tau} = \sum_{n'=1}^{\infty}(\Omega_{n'}C_{n+n'} + \Omega_{-n'}C_{n-n'}), \tag{7}$$

with $\Omega_{n'} = V_{n'}e^{-in'\tau}$.

The solution of Eq. (7) can be found straightforwardly. We introduce the "Hamiltonian" operator

$$\hat{T}(\tau) = \sum_{n'=1}^{\infty}[\Omega_{n'}(\hat{E}^-)^{n'} + \Omega_{-n'}(\hat{E}^+)^{n'}], \tag{8}$$

where $(\hat{E}^{\pm})^{n'}$ is a shifting operator defined as $(\hat{E}^{\pm})^{n'}|n\rangle = |n\pm n'\rangle$. The time evolution of the states driven by Eq. (8) can be found by solving the equation for the evolution operator, namely

$$i\frac{d\hat{U}}{d\tau} = \hat{T}(\tau)\hat{U}, \quad \hat{U}(0) = \hat{1}. \tag{9}$$

Finally, $C_n(\tau)$ can be easily obtained as

$$C_n(\tau) = \langle n|\hat{U}|0\rangle. \tag{10}$$

The solution of Eq. (9) can be immediately obtained since the operators $\hat{E}^{\pm}$ are commuting and therefore there is no problem with time ordering. We find

$$\hat{U}(\tau) = \prod_{n'=1}^{\infty} e^{-iL_{n'}(\tau)(\hat{E}^-)^{n'}}e^{-iL_{-n'}(\tau)(\hat{E}^+)^{n'}}, \tag{11}$$

where $L_{n'}(\tau) = \int_0^\tau d\tau' V_{n'}(\tau')e^{-in'\tau'}$. The evaluation of the scalar product in Eq. (10) is straightforward, indeed we easily get

$$C_n(\tau) = \prod_{n'=1}^{\infty}\sum_{s=0}^{\infty}\sum_{r=0}^{\infty}(-i)^{r+s}\frac{L_{n'}^r}{r!}\frac{L_{-n'}^s}{s!}\delta_{n,n'(s-r)}. \tag{12}$$

We can now write $f_n(\tau)$ in a closed form, recalling the series expansion of the Bessel function of first kind, i.e.,

$$J_n(x) = \left(\frac{x}{2}\right)^n\sum_{k=0}^{\infty}\frac{(-)^k(x/2)^{2k}}{k!(n+k)!}, \tag{13}$$

thus we get

[a] Permanent address: Quantum Institute, University of California, Santa Barbara, California 93106.
[b] Permanent address: Instituto Nazionale di Fisica Nucleare, Napoli, Italy and Dipartimento di Fisica, Universitá di Salerno, Salerno, Italy.

 0022-2488/87/020404-02$02.50 © 1987 American Institute of Physics

$$f_n(\tau) = e^{-in\tau} \prod_{n'=1}^{\infty} \left(\frac{L_{-n'}}{L_{n'}} \right)^{n/2n'} \times (-i)^{n/n'} J_{n/n'}(2(L_{n'}L_{-n'})^{1/2}). \quad (14)$$

We must note that Eq. (14) is relevant to the initial conditions $f_n(0) = \delta_{n,0}$. For the more general case $\bar{f}_n(0)$ we find

$$\bar{f}_n(\tau) = \sum_{l=-\infty}^{+\infty} \bar{f}_l(0) f_{n-l}(\tau). \quad (15)$$

If the Hamiltonian Eq. (1) only contains near-neighbors interaction,[3] then

$$\langle 0|\hat{H}|n\rangle = H_1(\delta_{n,1} + \delta_{n,-1}). \quad (16)$$

We easily get for f_n

$$f_n(\tau) = (-i)^n e^{-in\tau 1/2} J_n(2V_1([\sin(\tau/2)]/\tfrac{1}{2})). \quad (17)$$

In this paper we have presented an alternative method to get exact solutions of the single-band time-dependent Schrödinger equation in the presence of a uniform electric field. The method is based on the technique developed by the authors to get solutions of the Raman–Nath-type equations.[2]

ACKNOWLEDGMENTS

J.G. wishes to thank ENEA, Frascati, Italy for the warm hospitality.

This work was partially supported by a travel grant R.G. 85/0366 extended to us by NATO (North Atlantic Treaty Organization) Collaborative Research Program.

[1] M. Luban, J. Math. Phys. **26**, 2386 (1985).
[2] G. Dattoli, J. C. Gallardo, and A. Torre, J. Math. Phys. **27**, 772 (1986).
[3] J. Gallinar and D. Mattis, J. Phys. A **18**, 2583 (1985).

Stationary axisymmetric Kaluza–Klein black hole

P. O. Mazur
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106 and Department of Physics, Syracuse University, Syracuse, New York 13244

Luca Bombelli
Department of Physics, Syracuse University, Syracuse, New York 13244

(Received 2 July 1986; accepted for publication 15 October 1986)

Stationary, rotating Kaluza–Klein black hole solutions are studied in the case of vanishing electric charge. It is shown that in this case the field equations decouple and consist of the vacuum Ernst equations and the Laplace equation for the scalar (dilaton) field. The regular, electrically neutral, rotating Kaluza–Klein black hole is described uniquely by the trivial embedding of the Kerr metric in the five-dimensional space-time.

I. INTRODUCTION

The five-dimensional Einstein gravity according to Kaluza and Klein[1] leads to a unified theory of gravity, electromagnetism, and a scalar field when the compact extra dimension representing the U(1) gauge group is not visible. The vacuum four-dimensional Einstein and Einstein–Maxwell equations possess a number of physically important solutions like the stationary, rotating Kerr and Kerr–Newman black holes which were proved to be the unique stationary equilibrium states of black holes.[2,3] In the Kaluza–Klein theory new solutions are arising like the topologically nontrivial Kaluza–Klein magnetic monopoles[4,5] corresponding to the twisted U(1) bundles representing five-dimensional space-times. One should expect therefore that the presence of a scalar field will modify the four-dimensional black hole solutions. Black hole solutions for the Kaluza–Klein theory are known only for the spherically symmetric case.[6–10] It seems interesting to explore the stationary, rotating Kaluza–Klein black hole solutions. One may also ask if the no-hair conjecture is true for the Kaluza–Klein black holes. One should expect that the Kaluza–Klein black holes are characterized completely by four parameters, the mass M, the angular momentum J, the electric charge Q, and the scalar charge Σ, which seems not to be independent from the other three parameters.[10] Before attempting to prove the uniqueness theorem (or no-hair lemma) for Kaluza–Klein black holes one should first study the exact solutions describing them.

We will study the uncharged black holes, i.e., $Q=0$ case, where the electromagnetic field A_a is assumed to be zero. The five-dimensional metric is "static" with respect to the Killing vector $\partial/\partial x^5 = \partial/\partial\psi$. It is also a stationary metric. The stationary Einstein–Maxwell equations possess a hidden symmetry group[3] which is represented nonlinearly and is isomorphic to the pseudounitary group SU(2,1). As it is well known, this property of the field equations is related to the fact that the Einstein–Maxwell action for stationary fields can be dimensionally reduced to the action describing three-dimensional gravity coupled to the nonlinear sigma model on the hyperbolic Kähler symmetric space $SU(2,1)/S(U(2)\times U(1))$ (see Ref. 3). The fact that the dimensionally reduced action contains a part describing a sigma model turned out to be crucial in proving black hole uniqueness theorems.[3] Of course, this is the case because during dimensional reduction of the Einstein–Hilbert Lagrangian the sigma model Lagrangian emerges as a rule, with the scalar fields parametrizing a coset symmetric space.[11] The scalar fields correspond to degrees of freedom related to invisible extra dimensions (forming a torus). The hidden symmetry group of the dimensionally reduced action is a noncompact one and the coset space is always hyperbolic. Maison[12] has shown that the stationary Kaluza–Klein fields possess the hidden symmetry group $SL(3,R)$. The field equations in this case are equivalent to a sigma model on the hyperbolic symmetric space $SL(3,R)/SO(3)$. The five scalar fields parametrizing the $SL(3,R)/SO(3)$ coset space correspond to two complex Ernst potentials, the gravitational and the electromagnetic ones, and to a real scalar field. Looking for the Kaluza–Klein black hole solutions we will use the sigma model form of the field equations.

In Sec. II of this paper we will discuss the Geroch formulation of the field equations for the stationary Kaluza–Klein theory. We will show that the field equations in the case when the electromagnetic field vanishes decouple and they consist of the vacuum Ernst equations and the Laplace equation for the scalar field. This fact considerably simplifies the analysis and allows one to find the rotating Kaluza–Klein black hole solution. In Sec. III we will construct the metric describing the rotating Kaluza–Klein black hole solution. We will show that the solution is nonsingular only when the scalar charge vanishes. This result could be anticipated on the basis of the fact that black holes cannot have scalar "hair." The regular, electrically neutral, rotating Kaluza–Klein black hole is described by the trivial embedding of the four-dimensional Kerr solution in the five-dimensional space-time, i.e., it is a trivial product metric of a circle with the Kerr space-time. Using this observation it is easy to show that this is the unique electrically neutral, rotating Kaluza–Klein black hole.

II. STATIONARY KALUZA–KLEIN FIELDS

In this section we shall present the Geroch formulation[12,13] of the stationary Kaluza–Klein theory. We will also discuss the axisymmetric fields. We will assume that the metric g_{ab}, $a,b=1,\ldots,5$ of signature $(-,+,+,+,+)$ satisfies the five-dimensional Einstein equations and admits two commuting Killing vectors ξ^a_A, $A=1,2$: $\xi_1=\partial/\partial t$, $\xi_2=\partial/\partial x^5=\partial/\partial\psi$, and that the projection from the five-dimensional space-time manifold M onto the space S of orbits of the Killing vectors induces a smooth manifold structure on S. This is always true unless the isometry group has

 0022-2488/87/020406-04$02.50 © 1987 American Institute of Physics

fixed points. Introducing the matrix λ_{AB} defined by

$$\lambda_{AB} = g_{ab}\xi^a_A\xi^b_B, \tag{2.1}$$

and its inverse λ^{AB}, we can define the projection operator on the tangent space to S,

$$h^a{}_b = \delta^a{}_b - \lambda^{AB}\xi^a_A\xi_{Bb}, \tag{2.2}$$

and the metric on S,

$$h_{ab} = h_a{}^c h_b{}^d g_{cd}. \tag{2.3}$$

The metric g_{ab} is described completely by the geometrical objects on S: h_{ab}, λ_{AB} and the "twists" ω_{Aa}

$$\omega_{Aa} = \epsilon_{abcde}\xi^b_1\xi^c_2\nabla^d\xi^e_A, \tag{2.4}$$

which are curl-free $\nabla_{[a}\omega_{Ab]} = 0$ because of the vacuum Einstein's equations $R_{ab} = 0$, and can be locally expressed as $\omega_{Aa} = \nabla_a\omega_A$. One can show that $R_{ab} = 0$ is equivalent to the following equations on S (Ref. 12):

$$^{(3)}R_{ab}(\gamma) = \tfrac{1}{4}\mathrm{Tr}(\lambda^{-1}D_a\lambda\lambda^{-1}D_b\lambda) + \tfrac{1}{4}\tau^{-2}D_a\tau D_b\tau + \tfrac{1}{2}\tau^{-1}D_a\omega^T\lambda^{-1}D_b\omega, \tag{2.5}$$

$$D^aD_a\lambda = D_a\lambda\lambda^{-1}D^a\lambda - \tau^{-1}D_a\omega D^a\omega^T, \tag{2.6}$$

$$D^aD_a\omega = D_a\lambda\lambda^{-1}D^a\omega + \tau^{-1}D_a\tau D^a\omega, \tag{2.7}$$

where λ is the 2×2 matrix λ_{AB} with nonvanishing determinant $\tau = \det\lambda$ and ω is a column two-vector $\omega = (\omega_A)$. The covariant derivative operator D_a is the one compatible with the conformally rescaled three-metric $\gamma_{ab} = \tau h_{ab}$. Equations (2.6) and (2.7) can be written as the sigma model equations for a 3×3 matrix χ (Ref. 12),

$$\chi = \begin{pmatrix} \tau^{-1} & -\tau^{-1}\omega T \\ -\tau^{-1}\omega & \lambda + \tau^{-1}\omega\omega^T \end{pmatrix}, \tag{2.8}$$

with the properties

$$\chi^T = \chi, \quad \det\chi = 1. \tag{2.9}$$

The equations for a sigma model are

$$D^aJ_a = 0, \tag{2.10}$$

where

$$J_a = D_a\chi\chi^{-1}.$$

From conditions (2.8) and (2.9) we can see that χ transforms as a covariant, second-rank tensor under SL(3,R). Moreover a matrix satisfying (2.9) can be decomposed as

$$\chi = A^TA, \quad \det A = 1, \quad A\in\mathrm{SL}(3,R). \tag{2.11}$$

It is easy to see that the left SL(3,R) translation $A\mapsto A' = hA$ leaves χ unchanged if $h^Th = I$, or $h\in$SO(3). In other words there is a one-to-one correspondence beteween χ and elements of the coset symmetric space SL(3,R)/SO(3). Equation (2.10) describes the nonlinear sigma model on the symmetric space SL(3,R)/SO(3).

In the case of the vanishing electromagnetic field $\omega_2 = 0$, so we set $\omega_1 = \omega$ and the field equations (2.6) and (2.7) simplify considerably. The five-dimensional metric can be written then in the form

$$ds^2 = e^{\Phi}\,d\psi^2 + \lambda(dt - \omega\,d\phi)^2 - \lambda^{-1}e^{-\Phi}[e^{2\mu_2}\,dx_2^2 + e^{2\mu_3}\,dx_3^2 + \rho^2\,d\phi^2]. \tag{2.12}$$

From the form (2.12) of the five-dimensional metric we have $\lambda_{11} = \lambda, \lambda_{22} = e^{\Phi}, \lambda_{12} = 0$, and the matrix λ_{AB} is diagonal. In the axisymmetric case $\partial/\partial\phi$ is a Killing vector and $-\rho^2$ is the determinant of the matrix of scalar products of the three Killing vectors $(\partial/\partial\psi, \partial/\partial t, \partial/\partial\phi)$. It can be seen from (2.5) that ρ satisfies the Laplace equation on the two-dimensional space of orbits of three Killing vectors. Equations (2.6) and (2.7) reduce to

$$\lambda D^aD_a\lambda = D^a\lambda D_a\lambda - e^{-\Phi}D^a\omega D_a\omega, \tag{2.13}$$

$$\lambda D^aD_a\omega = 2D^a\lambda D_a\omega + \lambda D^a\Phi D_a\omega, \tag{2.14}$$

$$D^aD_a\Phi = 0. \tag{2.15}$$

These equations are similar to the vacuum Ernst equations[14] for which we know many solutions. It turns out that by rescaling λ one can reduce (2.13) and (2.14) to the form of the Ernst equations. Let $\lambda = Ff$, where F is a solution of

$$FD^aD_aF = D^aFD_aF. \tag{2.16}$$

At this point we observe that $F' = \beta F^{\alpha}$ is also a solution of (2.16) for arbitrary $\alpha,\beta\neq0$: this corresponds to the symmetry of the Laplace equation satisfied by $\ln F$: $\ln F\to\alpha\ln F + \ln\beta$. Using (2.16) we get from (2.13) and (2.14)

$$F^2(fD^aD_af - D^afD_af) = -e^{-\Phi}D^a\omega D_a\omega, \tag{2.17}$$

$$F(fD^aD_a\omega - 2D^afD_a\omega) = F(2F^{-1}D^aF + D^a\Phi)D_a\omega. \tag{2.18}$$

One can easily see that when $F = e^{-\Phi/2}$ then f and ω satisfy the Ernst equation for a complex potential $\epsilon = f + i\omega$,

$$fD^aD_a\epsilon = D^a\epsilon D_a\epsilon. \tag{2.19}$$

Now we observe that when $\Phi = 0$, or $F = 1$, then the field equations reduce to the vacuum four-dimensional Einstein equations. Equation (2.19) has a very simple solution describing the Kerr black hole. Therefore if we take for ϵ the Ernst potential of the Kerr solution and for Φ the simplest monopole solution of the Laplace equation then we will automatically obtain a Kaluza–Klein solution which in the limit of vanishing scalar charge describes the vacuum Kerr black hole. We expect this solution to describe a genuine rotating Kaluza–Klein black hole.

The five-dimensional metric can be written in a new form,

$$\begin{aligned} ds^2 &= e^{\Phi}\,d\psi^2 + e^{-\Phi/2}[f(dt - \omega\,d\phi)^2 \\ &\quad - f^{-1}(e^{2\mu_2}\,dx_2^2 + e^{2\mu_3}\,dx_3^2 + \rho^2\,d\phi^2] \\ &= e^{\Phi}\,d\psi^2 + e^{-\Phi/2}\,{}^{(4)}g_{\alpha\beta}\,dx^{\alpha}\,dx^{\beta}, \end{aligned} \tag{2.20}$$

where the four-dimensional metric $^{(4)}g_{\alpha\beta}$ is regular on and outside the event horizon of a black hole solution. In order to reconstruct the metric (2.20) from the Ernst potential ϵ and the scalar field Φ one has to solve Eq. (2.5), which in the present case has the form

$$^{(3)}R_{ab}(\gamma) = \tfrac{1}{2}f^{-2}(f_{,a}f_{,b} + \omega_{,a}\omega_{,b}) + \tfrac{3}{8}\Phi_{,a}\Phi_{,b}. \tag{2.21}$$

We also need the relation between the twist ω and the metric coefficient w:

$$\omega_{,2} = \rho^{-1}e^{\mu_2-\mu_3}\lambda^2e^{\Phi}w_{,3}, \tag{2.22a}$$

$$\omega_{,3} = -\rho^{-1}e^{\mu_3-\mu_2}\lambda^2e^{\Phi}w_{,2}, \tag{2.22b}$$

which after rescaling $\lambda = e^{-\Phi/2}f$ has the same form as in the

4-D vacuum case,

$$\omega_{,2} = \rho^{-1} e^{\mu_2 - \mu_3} f^2 w_{,3}, \tag{2.23a}$$

$$\omega_{,3} = -\rho^{-1} e^{\mu_3 - \mu_2} f^2 w_{,2}. \tag{2.23b}$$

When we take the Ernst potential ϵ of the Kerr solution then w will be the same as in the Kerr metric. The presence of the scalar field Φ has the effect of changing the four-dimensional metric ${}^{(4)}g_{\alpha\beta}$ only through the change in μ_2, μ_3 with respect to their vacuum form.

III. ROTATING KALUZA–KLEIN BLACK HOLE SOLUTION

We are looking for the axisymmetric black hole solution to Eqs. (2.15), (2.19), and (2.21), where we define the Kaluza–Klein black hole solution in the standard way, i.e., we assume that the metric ${}^{(4)}g_{\alpha\beta}$ has a regular event horizon and symmetry axis, and it is asymptotically flat. We assume that the event horizon is the Killing horizon spanned by the Killing vectors $\partial/\partial t$ and $\partial/\partial\phi$. The event horizon is a null hypersurface given by the equation[15]

$$N(x^2, x^3) = 0, \quad {}^{(4)}g^{\alpha\beta} N_{,\alpha} N_{,\beta} = 0. \tag{3.1}$$

This leads to the condition on N

$$e^{2(\mu_3 - \mu_2)} (N_{,r})^2 + (N_{,\theta})^2 = 0, \tag{3.2}$$

where $x^2 = r, x^3 = \theta$. From Eq. (3.2) we obtain the equation of the event horizon

$$e^{2(\mu_3 - \mu_2)} = \Delta(r) = 0. \tag{3.3}$$

The function ρ vanishes on the event horizon $\mathscr{H}$ when $\Delta = 0$ because the horizon is a fixed set for the vector field $\partial/\partial t + \Omega_H \partial/\partial\phi$, where Ω_H is the angular velocity of the horizon. We may put

$$\rho = \Delta^{1/2} H(\theta). \tag{3.4}$$

Then the Laplace equation for ρ gives

$$\tfrac{1}{2}\Delta_{,rr} + H^{-1} H_{,\theta\theta} = 0, \tag{3.5}$$

with the solution

$$\Delta(r) = r^2 - 2mr + a^2, \quad H(\theta) = \sin\theta, \tag{3.6}$$

for some constants m and a. The harmonic conjugate functions z and ρ have the form

$$\rho = \Delta^{1/2} \sin\theta, \; z = (r - m)\cos\theta. \tag{3.7}$$

It is convenient to introduce the ellipsoidal coordinates η, μ because the solutions describing black holes have a very simple form in terms of these coordinates

$$\eta = \kappa^{-1}(r - m), \quad \mu = \cos\theta, \quad \kappa^2 = m^2 - a^2. \tag{3.8}$$

The coordinate range for η and μ is $\eta \in [1, \infty], \mu \in [-1, 1]$. In this coordinate system, Δ, ρ, and z have the form

$$\Delta = \kappa^2(\eta^2 - 1), \quad \rho = \kappa(\eta^2 - 1)^{1/2}(1 - \mu^2)^{1/2},$$
$$z = \kappa\eta\mu. \tag{3.9}$$

The horizon position is now at $\eta = 1$, the axis at $\mu = \pm 1$, and spatial infinity at $\eta = \infty$. In ellipsodial coordinates, the field equations become

$$f[((\eta^2 - 1)\epsilon_{,\eta})_{,\eta} + ((1 - \mu^2)\epsilon_{,\mu})_{,\mu}]$$
$$= (\eta^2 - 1)\epsilon_{,\eta}^2 + (1 - \mu^2)\epsilon_{,\mu}^2, \tag{3.10}$$

$$((\eta^2 - 1)\Phi_{,\eta})_{,\eta} + ((1 - \mu^2)\Phi_{,\mu})_{,\mu} = 0, \tag{3.11}$$

The boundary condition on ϵ as $\eta \to \infty$ is $\epsilon \to -1$, which corresponds to asymptotic flatness of the metric. This suggests the change of variables

$$\xi = (1 - \epsilon)/(1 + \epsilon), \quad \epsilon = (1 - \xi)/(1 + \xi), \tag{3.12}$$

where ξ satisfies the equation

$$(\xi\bar{\xi} - 1) D^a D_a \xi = 2\bar{\xi} D^a \xi D_a \xi. \tag{3.13}$$

The Kerr solution to the Ernst equation (3.13) has a surprisingly simple form

$$\xi = p\eta + iq\mu, \tag{3.14}$$

where $p^2 + q^2 = 1$.

Now we have to find an appropriate solution for the Kaluza–Klein scalar field Φ with the boundary condition $\Phi \to 0$ as $\eta \to \infty$. If Φ depends only on one arbitrary constant, i.e., the scalar charge Σ, then Φ must be a function of η only. The solution of Eq. (3.11) with the correct boundary condition has the form

$$\Phi = \frac{c}{2} \ln\left(\frac{\eta - 1}{\eta + 1}\right). \tag{3.15}$$

Then Φ has the asymptotic form as $r \to \infty$, $\Phi \sim c\kappa/r$, where $c\kappa$ is proportional to the scalar charge Σ. The presence of the scalar field Φ will have only an effect on the two-metric on the space of orbits of the Killing vectors $\partial/\partial t$ and $\partial/\partial\phi$. In order to calculate the two-metric on the space of orbits of the Killing vectors it is convenient to work in a "conformal gauge"

$$ds_{(2)}^2 = e^{2\mu_2} dx_2^2 + e^{2\mu_3} dx_2^2 = e^{2\gamma}(d\rho^2 + dz^2). \tag{3.16}$$

The remaining field equations (2.21) simplify considerably in the conformal gauge

$$\gamma_{,\rho} = \rho(G_{\rho\rho} - G_{zz}), \tag{3.17a}$$

$$\gamma_{,z} = 2\rho R_{\rho z}, \tag{3.17b}$$

where

$$G_{\rho\rho} - G_{zz} = \tfrac{1}{2} f^{-2}(f_{,\rho}^2 - f_{,z}^2 + \omega_{,\rho}^2 - \omega_{,z}^2)$$
$$+ \tfrac{3}{8}(\Phi_{,\rho}^2 - \Phi_{,z}^2), \tag{3.18a}$$

$$R_{\rho z} = \tfrac{1}{2} f^{-2}(f_{,\rho} f_{,z} + \omega_{,\rho}\omega_{,z}) + \tfrac{3}{8}\Phi_{,\rho}\Phi_{,z}. \tag{3.18b}$$

Using Eqs. (3.12), (3.14), (3.15), (3.17), and (3.18) one obtains

$$e^{2\gamma}(d\rho^2 + dz^2) = \kappa^2\left(\frac{\eta^2 - 1}{\eta^2 - \mu^2}\right)^{3c^2/16} (p^2\eta^2 + q^2\mu^2 - 1)$$
$$\times\left(\frac{d\eta^2}{\eta^2 - 1} + \frac{d\mu^2}{1 - \mu^2}\right). \tag{3.19}$$

From Eqs. (2.23a) and (2.23b) one can also calculate ω,

$$w = 2\kappa q p^{-1}(1 - \mu^2)(p\mu + 1)(p^2\eta^2 + q^2\mu^2 - 1)^{-1}. \tag{3.20}$$

Observing that the constraint $p^2 + q^2 = 1$ is easily solved by $p = \kappa/m, q = a/m$ one has $\kappa q p^{-1} = a$. The four-dimensional metric ${}^{(4)}g_{\alpha\beta}$ looks like the slightly modified Kerr metric. The only difference is that the scale factor $e^{2\gamma}$ of the two-dimensional metric on the space of orbits of $\partial/\partial t$ and $\partial/\partial\phi$ is

modified by the presence of the scalar field

$$e^{2\gamma} = \left(\frac{\eta^2 - 1}{\eta^2 - \mu^2}\right)^{3c^2/16} e^{2\gamma_k}, \tag{3.21}$$

where $e^{2\gamma_k}$ stand for a scale factor of the Kerr metric. The multiplicative factor vanishes on the event horizon when the scalar charge is nonvanishing. The effect of the scalar charge is to produce a curvature singularity on the even horizon $\mathscr{H}$. The regularity of the event horizon $\mathscr{H}$ and the boundedness of Φ on $\mathscr{H}$ implies that the scalar charge Σ (or c) must vanish identically. We conclude therefore that the only acceptable regular electrically neutral Kaluza–Klein black hole is the trivially embedded Kerr black hole in five dimensions. The metric describing such a black hole is a product metric on $S^1 \times M_k$, where M_k is the Kerr space-time. It is quite easy to extend the standard argument of uniqueness (and no hair) theorems to the case of electrically neutral Kaluza–Klein black holes, and show that the product metric of the Kerr metric with the metric on the circle is the unique electrically neutral rotating Kaluza–Klein black hole solution. The regular Kaluza–Klein black hole solution with nonzero scalar charge is necessarily electrically charged. The proof of uniqueness, once the boundary conditions for the fields Φ, f, and ω on the space of orbits of the Killing vectors are given, uses the ordinary Green's identity to show that $\Phi \equiv 0$ is the only possible solution for the scalar field, and the generalized Green's identity[3] to show uniqueness for f and ω, which in fact reduces to the analogous question for ordinary four-dimensional general relativity. Details of the proof and a generalization to charged rotating black holes will be given elsewhere.[16]

To conclude, we will remark that all our arguments can be trivially extended to a particular class of higher-dimensional Kaluza–Klein theories, where the internal manifold is a d-dimensional torus T^d. If we assume again the presence of a stationary Killing vector $\xi_1 = \partial/\partial t$ and a Killing vector $\xi_\alpha = \partial/\partial\psi^\alpha$ for each of the internal dimensions, and the metric ansatz

$$ds^2 = \sum_{\alpha=1}^{d} e^{\Phi_{(\alpha)}} (d\psi^\alpha)^2 + \lambda (dt - w\, d\varphi)^2 - \lambda^{-1} \prod_{\alpha=1}^{d} e^{-\Phi_{(\alpha)}} (e^{2\mu_2} dx_2^2 + e^{2\mu_3} dx_3^2 + \rho^2 d\varphi^2)$$

(where for simplicity we are taking a unit radius for all internal dimensions, but this does not affect our conclusions), then each scalar field $\Phi_{(\alpha)}$ will satisfy an uncoupled Laplace equation and, rescaling $\lambda = Ff$, with

$$F = \prod_{\alpha=1}^{d} e^{-\Phi_{(\alpha)}/2},$$

$\epsilon = f + i\omega$ will satisfy the Ernst equation (2.19). The remaining equations for μ_2 and μ_3, solved in the conformal gauge, again tell us that, to get a regular solution on the horizon, we must impose that all scalar charges associated with the extra dimensions vanish, and the only solutions we obtain are trivial metric products of Kerr space-time with a d-dimensional torus T^d.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under Grants No. PHY82-17853 and PHY83-18350, supplemented by funds from the National Aeronautics and Space Administration, at the University of California at Santa Barbara.

[1]Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1**, 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926).
[2]B. Carter, in *Black Holes*, edited by B. DeWitt and C. DeWitt (Les Houches, 1972) (Gordon and Breach, New York, 1973); D. C. Robinson, Phys. Rev. Lett. **34**, 905 (1975).
[3]P. O. Mazur, J. Phys. A **15**, 3173 (1982); Phys. Lett A **100**, 341 (1984).
[4]R. D. Sorkin, Phys. Rev. Lett. **51**, 87 (1983).
[5]D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983).
[6]H. Leutwyler, Arch. Sci. **13**, 549 (1960).
[7]P. Dobiasch and D. Maison, Gen. Relativ. Gravit. **14**, 231 (1982).
[8]A. Chodos and S. Detweiler, Gen. Relativ. Gravit. **14**, 879 (1982).
[9]D. Pollard, J. Phys. A **16**, 565 (1983).
[10]G. W. Gibbons and D. L. Wiltshire, Ann. Phys. (NY) **167**, 201 (1986).
[11]J. Scherk, Phys. Lett. B **88**, 265 (1979).
[12]D. Maison, Gen. Relativ. Gravit. **10**, 717 (1979).
[13]R. Geroch, J. Math. Phys. **12**, 918 (1971).
[14]F. Ernst, Phys. Rev. **167**, 1175 (1968).
[15]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., New York, 1983).
[16]L. Bombelli and P. O. Mazur, in preparation.

Imperfect fluids and repulsive gravitation

J. Ponce de León[a)]
Universidad Simon Bolivar, Division de Fisica y Matematicas, Departamento de Fisica, Apdo. 80659, Caracas 1081-A, Venezuela and Departamento de Fisica, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela

(Received 15 July 1986; accepted for publication 15 October 1986)

Imperfect fluid sources to the Schwarzschild exterior solution are studied under the assumption that the metric coefficients g_{00} and g_{11} of the interior solution satisfy the relation $g_{00}g_{11} = -1$. It was found that the core of such a distribution is gravitationally repulsive provided the energy density is positive.

I. INTRODUCTION

The coefficients g_{00} and g_{11} of the Schwarzschild exterior solution have the interesting property that in curvature coordinates they satisfy the relation $g_{00}g_{11} = -1$. This fact motivates one to study the assumption that this relation also is valid within the matter distribution. Tiwari, Rao, and Kanakamedala[1] showed (among other things) that in the case in which the interior is filled with perfect fluid, this assumption leads to the vanishing of the density and pressure identically, i.e., there exists no interior solution. This means that the assumption $g_{00}g_{11} = -1$ is incompatible with the assumption of perfect fluid.

In this work we investigate the relation in the case in which the interior is filled with imperfect fluid. First, we will show that in this case there exist sources to the Schwarzschild exterior metric. Second, we will prove that the gravitational mass is negative in the central region of such sources. Third, we will exhibit a specific model which illustrates the results.

II. THEORY

Let us then consider a static and spherically symmetric gravitational field. In curvature coordinates, the associated line element reads

$$ds^2 = e^{\nu}dt^2 - e^{\lambda}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{1}$$

where ν and λ are functions of r only.

The Einstein field equations corresponding to this line element can be written as

$$\frac{dT_1^1}{dr} = \frac{\nu'}{2}(T_0^0 - T_1^1) + \frac{2}{r}(T_2^2 - T_1^1), \tag{2}$$

$$4\pi r T_1^1 - \frac{m(r)}{r^2} + \frac{\nu'}{2}e^{-\lambda} = 0, \tag{3}$$

$$m(r) = 4\pi\int_0^r r^2 T_0^0\, dr \equiv \frac{1 - e^{-\lambda}}{2}r, \tag{4}$$

$$T_2^2 = T_3^3. \tag{5}$$

The gravitational mass inside a sphere of "radius" r is given by the Tolman–Whittaker formula, viz.,

$$M_G(r) = 4\pi\int_0^r (T_0^0 - T_1^1 - T_2^2 - T_3^3)r^2 e^{(\nu+\lambda)/2}\, dr. \tag{6}$$

We assume that the matter distribution extends to radius r_0. Hence the exterior space-time is described by the Schwarzschild metric

$$ds^2 = (1 - 2M/r)dt^2 - (1 - 2M/r)^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{7}$$

where M is the total gravitational mass inside the sphere, viz., $M = M_G(r_0)$.

Necessary and sufficient conditions for matching the metrics (1) and (7) are given by the continuity of the first and second fundamental form across r_0, viz.,

$$e^{\nu(r_0)} = e^{-\lambda(r_0)} = 1 - 2M/r_0, \tag{8}$$

$$r_0\nu'(r_0) = (2M/r_0)(1 - 2M/r_0)^{-1}, \tag{9}$$

Eq. (3) then shows that on the boundary $T_1^1 = 0$.

We now assume the relation $g_{00}g_{11} = -e^{(\nu+\lambda)} = -1$ to be valid within the distribution. Then from Eqs. (3) and (4) we find

$$T_0^0 = T_1^1. \tag{10}$$

The converse is also true, viz., if $T_0^0 = T_1^1$ then from Eqs. (3), (4), (8), (9) it follows that $g_{00}g_{11} = -1$.

Now substituting Eq. (10) into Eq. (2) and using the boundary conditions we find

$$T_0^0 = 2\int_r^{r_0} \frac{(T_1^1 - T_2^2)}{r}\, dr. \tag{11}$$

This equation shows that if a specific relation between the stresses is given *a priori*, then the source will be fully determined. In particular, the energy momentum tensor vanishes everywhere only when $T_1^1 = T_2^2$ (perfect fluid). This equation also indicates that the energy density T_0^0 will decrease monotonically outward provided $T_0^0 > T_2^2$.

For the configurations under study the gravitational mass inside r, as given by Eq. (6), is

$$M_G = -8\pi\int_0^r T_2^2 r^2\, dr. \tag{12}$$

Since at the center of the distribution all physical quantities must be finite, it follows from Eq. (11) that $(T_1^1 - T_2^2)$ should vanish at least as rapidly as r when $r \to 0$. Hence $T_0^0 = T_1^1 \approx T_2^2$ near the center. Equation (12) then shows that if $T_0^0 > 0$ then the gravitational mass is negative in the central region. Consequently, for $T_0^0 > 0$ the core of the

[a)] Postal address: Apartado 2816, Caracas 1010A, Venezuela.

 0022-2488/87/020410-02$02.50 © 1987 American Institute of Physics

sphere is gravitationally repulsive in the sense that a free particle will be accelerated away from this region.[2]

Let us now give an explicit solution to the set of Eqs. (2)–(5), (10), and (11). With this aim we assume the following relation between the stresses:

$$T_1^1 - T_2^2 = K^2 r^2 (1 - r^2/r_0^2), \tag{13}$$

where K^2 is a constant. Using the field equations and the boundary conditions we obtain the solution as follows:

$$e^{\nu} = e^{-\lambda} = 1 - (\epsilon/4)v^2(35 + 15v^4 - 42v^2), \tag{14}$$

$$T_0^0 = T_1^1 = (105\epsilon/32\pi r_0^2)(1 - v^2)^2, \tag{15}$$

$$T_1^1 - T_2^2 = (105\epsilon v^2/16\pi r_0^2)(1 - v^2), \tag{16}$$

where $v \equiv r/r_0$ and $\epsilon \equiv M/r_0$.

The above solution has the following properties.

(i) The energy density T_0^0 is positive and decreases monotonically outward. Furthermore, $T_0^0 = T_1^1$ and $T_0^0 \geqslant T_2^2 = T_3^3$.

(ii) The metric coefficients e^{ν} and e^{λ} are positive throughout the matter for $M \lesssim 0.438 r_0$.

(iii) At the center $T_0^0 = T_1^1 = T_2^2 = T_3^3 > 0$, and at the boundary $T_0^0 = T_1^1 = T_2^2 = T_3^3 = 0$.

(iv) The red shift (from a point in the sphere to infinity) is maximum at $r \approx 0.792 r_0$ rather than at the center.

(v) The gravitational mass inside r

$$M_G(r) = (M/4)v^3(84v^2 - 45v^4 - 35), \tag{17}$$

is negative for $r \lesssim 0.792 r_0$.

(vi) The mass function $m(r)$ defined by Eq. (4) is positive at all points within the distribution.

III. CONCLUSION

We conclude from this work that (a) the necessary and sufficient condition for $g_{00} g_{11} = -1$ is $T_1^1 = T_0^0$; (b) if $T_1^1 = T_0^0$, then all physical quantities depend on the "degree of imperfection" of the fluid and vanish for perfect fluid; (c) if $T_1^1 = T_0^0$ and $T_0^0 > 0$, then the central region of the sphere is gravitationally repulsive.

Concerning the solution presented here it could serve as initial (or final) configuration in the evolution scenario and it could be interesting to investigate the time evolution of such models where there exist gravitational repulsion.

We would like to finish with the following remarks.

(1) The energy momentum tensor of a viscous fluid satisfies the relation $g_{00} g_{11} = -1$ (and consequently $T_0^0 = T_1^1$). In fact, for a viscous fluid

$$T_{\nu}^{\mu} = (\rho + p)u^{\mu}u_{\nu} - p\delta_{\nu}^{\mu} + 2\eta\sigma_{\nu}^{\mu}, \tag{18}$$

where ρ and p are the density and pressure of the fluid, respectively, u^{μ} is the four velocity ($u^{\mu}u_{\mu} = 1$), η is the coefficient of viscosity, and σ_{ν}^{μ} is the shear tensor which satisfies

$$\sigma_{\mu\nu}u^{\mu} = 0, \quad \sigma_{\mu\nu}g^{\mu\nu} = 0, \tag{19}$$

from Eqs. (18) and (19) and using $T_0^1 = T_1^0 = 0$ we obtain (we recall that $u^1 \neq 0$ otherwise $\sigma_{\mu\nu} = 0$)

$$T_0^0 = T_1^1 = \rho. \tag{20}$$

Thus we see that the central region of any viscous static fluid sphere which satisfies the regularity conditions is gravitationally repulsive.

(2) The relation $g_{00} g_{11} = -1$ has already been used for perfect fluid spheres with charge, and it was shown by Grøn[2] and Gautreau[3] that such spheres also give rise to gravitational repulsion.

(3) The gravitational repulsion in these models[1–3] is a consequence of the violation of the "strong energy condition" $[(T_{\mu\nu} - g_{\mu\nu}T/2)W^{\mu}W^{\nu} \geqslant 0$ where W^{μ} is any timelike vector] which requires that gravity is always an attractive force (for details see Hawking and Ellis[4]). As it is known, this condition is violated in a number of situations (see, for example, Refs. 5–10).

[1] R. N. Tiwari, J. R. Rao, and R. R. Kanakamedala, Phys. Rev. D **30**, 489 (1984).
[2] Ø. Grøn, Phys. Rev. D **31**, 2129 (1985).
[3] R. Gautreau, Phys. Rev. D **31**, 1860 (1985).
[4] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., London, 1973).
[5] G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[6] P. N. Baaklini, Phys. Lett. A **66**, 357 (1978).
[7] L. Parker and S. A. Fulling, Phys. Rev. D **7**, 2357 (1973).
[8] F. J. Tipler, Phys. Rev. D **17**, 2521 (1978).
[9] Yu. A. Beletsky, Preprint ITP-83-148E of Academy of Sciences of the Ukrainian S. S. R., Institute for Theoretical Physics, 1983.
[10] Yu. A. Beletsky, Preprint ITF-84-7P, Academy of Sciences of the Ukrainian S. S. R., Institute for Theoretical Physics 1984 (in Russian).

The Green's function of a slab of a diatomic Montroll–Potts crystal

Paul Mazur
Rutgers University, Camden, New Jersey 08102
Robert H. Barron
Stockton State College, Pomona, New Jersey 08240

(Received 18 June 1986; accepted for publication 15 October 1986)

Using the Dyson equation repeatedly, starting with the Green's function for the infinite medium, the Green's function for the semi-infinite medium and finally the Green's function for a slab of a diatomic NaCl-type crystal using the Montroll–Potts model of nearest-neighbor central and noncentral forces are obtained.

I. INTRODUCTION

The model of a crystal popularized by Montroll and Potts[1] has had a long and useful history. There seem to be prevalent philosophies among theoretical physicists when solving real physical problems. One view is that one should start with the exact equations of motion and then make approximations as required as one works analytically towards the final solution. One disadvantage of this method is that the character of the solution may be related very intimately, but in an unknown manner, to the approximations made in the course of arriving at the solution. The other view is that one tries to approximately model the real problem in the beginning in such a way that the analysis carries all the way through without any further approximations. The solution may then reveal characteristics one might deduce must carry over to the solution of the real problem and, thereby, open up new vistas. This second attitude or philosophy seemed to be the one that guided Montroll throughout his illustrious scientific career. An outstanding example of this attitude was Montroll's exact calculation of the frequency distribution for a simple model of a two-dimensional harmonic lattice, for which he obtained log singularities.[2] van Hove[3] later showed that the log singularities were not due to the special model that Montroll had chosen but were due to the fact that the frequency versus wave vector relationship had mathematical critical points which, in turn, were due to the fact that the frequency was a periodic function of the wave vector K.

The model, now called the Montroll–Potts model, seems to be still alive. Recently, Dobrzynski *et al.*[4] have used it to calculate the exact Green's function for a superlattice.

One knows that the model is not useful for the calculation of some properties of interest. For example, the model does not yield Rayleigh (surface) waves.

Our plan is to first calculate the Green's function for the infinite lattice and then use the Dyson equation twice to eventually obtain the slab Green's function. A detailed description of the model will be found in Sec. II.

In Sec. II, as an application of the infinite lattice Green's function to the calculation of the mean-square displacements of each of the two kinds of atoms in an NaCl-type lattice, we show that both mean-square displacements approach the same value as the temperature increases. To show this we had to use a symmetry property of the frequency distribution function for our particular mode. But it is generally known that the high-temperature limit is the classical limit and in this limit two different atoms, irrespective of their masses, will have the same mean-square displacements if the potential energy functions connecting each atom to its neighbors are identical. It is a little peculiar that we had to use a symmetry property of the distribution function to prove this. In the Appendix, we show this classical limit in the case of a two-atom crystal as an example.

II. CALCULATION

To establish notation we first determine the Green's function for the infinite lattice. Our crystal is diatomic of the NaCl-type structure and we assume central and noncentral forces between nearest neighbors only. The central and noncentral force constants are taken to be equal ($=\beta$). The atomic positions are designated by three integers, l, m, and n. The atoms of mass M_1 are located at the positions where the sum $l+m+n$ is even. Atoms of mass M_2 are located at positions where this sum is odd.

The equations of motion of the two kinds of atoms are given by

$$M_1\ddot{u}_{l,m,n} = \beta\,[u_{l,m,n+1} + u_{l,m,n-1} + u_{l,m+1,n} + u_{l,m-1,n} + u_{l+1,m,n} + u_{l-1,m,n} - 6u_{l,m,n}], \quad l+m+n \text{ even}, \tag{1}$$

$$M_2\ddot{u}_{l,m,n} = \beta\,[u_{l,m,n+1} + u_{l,m,n-1} + u_{l,m+1,n} + u_{l,m-1,n} + u_{l+1,m,n} + u_{l-1,m,n} - 6u_{l,m,n}], \quad l+m+n \text{ odd}. \tag{2}$$

The quantities $u_{l,m,n}$ are the displacements of the l,m,nth atoms in the x direction from their equilibrium positions. There are similar sets of equations for displacements in the y and z directions. The matrix whose eigenvalues are the frequencies can be made Hermitian if we renormalize the displacements right at the beginning.

We set

$$u_{l,m,n} = V_{l,m,n}/\sqrt{M_1} \quad \text{for } l+m+n \text{ even}, \tag{3}$$

and

$$u_{l,m,n} = V_{l,m,n}/\sqrt{M_2} \quad \text{for } l+m+n \text{ odd}. \tag{4}$$

To obtain the frequencies of vibration of the lattice we assume the following form for $V_{l,m,n}$:

$$V_{l,m,n} = Ae^{i\omega t}e^{i\phi_l l}e^{i\phi_k m}e^{i\phi_l n}, \quad l+m+n \text{ even}, \tag{5}$$

$$= Be^{i\omega t}e^{i\phi_j l}e^{i\phi_k m}e^{i\phi_l n}, \quad l+m+n \text{ odd}. \tag{6}$$

Here A and B are constants to be determined.

 0022-2488/87/020412-13$02.50 © 1987 American Institute of Physics

Substituting Eqs. (3)–(6) into (1) and (2), we obtain, after a slight simplification, two equations for A, B, and ω^2:

$$\left[\frac{6\beta}{M_1}-\omega^2\right]A + \left[\frac{2\beta}{\sqrt{M_1M_2}}(\cos\phi_1+\cos\phi_2+\cos\phi_3)\right]B=0, \quad (7)$$

$$\left[\frac{2\beta}{\sqrt{M_1M_2}}(\cos\phi_1+\cos\phi_2+\cos\phi_3)\right]A + \left[\frac{6\beta}{M_2}-\omega^2\right]B=0. \quad (8)$$

The solutions of (7) and (8) are

$$\omega^2_{\pm} = 3\beta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) \pm \sqrt{9\beta^2\left(\frac{1}{M_1}-\frac{1}{M_2}\right)^2 + \frac{4\beta^2}{M_1M_2}(\cos\phi_j+\cos\phi_k+\cos\phi_l)^2},$$

$$A_{\pm} = \frac{2\beta(\cos\phi_j+\cos\phi_k+\cos\phi_l)}{\sqrt{M_1M_2}(+\omega^2_{\pm}-6\beta/M_1)}B. \quad (9)$$

For convenience, we can set $B=1$.

Our normalized eigenvectors for the infinite lattice can now be written as

$$\{V_{\pm}(\mathbf{l},\boldsymbol{\phi})\} = \sqrt{\frac{2}{N^3}}\frac{1}{\sqrt{1+C^2_{\pm}}}\{\alpha_{\pm}(\mathbf{l},\boldsymbol{\phi})e^{i\mathbf{l}\cdot\boldsymbol{\phi}}\}. \quad (10)$$

where

$$\mathbf{l}\equiv(l,m,n), \quad \boldsymbol{\phi}\equiv(\phi_j,\phi_k,\phi_l),$$

and

$$\alpha_{\pm}(\mathbf{l},\boldsymbol{\phi}) = 1, \quad \text{if } l+m+n \text{ is odd,}$$

$$= C_{\pm} = \frac{2\beta(\cos\phi_j+\cos\phi_k+\cos\phi_l)}{\sqrt{M_1M_2}[-\omega^2_{\pm}+(6\gamma/M_1)]}, \quad \text{if } l+m+n \text{ is even.}$$

We now form the sum

$$G^{\mathbf{l},\mathbf{l}'}(E) = \sum_{\boldsymbol{\phi}}\frac{V(\mathbf{l},\boldsymbol{\phi})V^*(\mathbf{l}',\boldsymbol{\phi})}{E-\omega^2_{\pm}(\boldsymbol{\phi})}. \quad (11)$$

The sum in Eq. (11) takes different forms depending on whether $\mathbf{l},\mathbf{l}'$ refer to M_1 or M_2 atoms. We write out explicitly the three possible cases.

Case I: ($\mathbf{l}$ and $\mathbf{l}'$ both refer to M_1 atoms; $l+m+n$, $l'+m'+n'$ both even)

$$G_0^{\mathbf{l},\mathbf{l}'} = G_0^{M_1M_1} + \sum_{\boldsymbol{\phi}}\frac{2}{N^3}\left\{\frac{C^2_+e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_+)(E-\omega^2_+(\boldsymbol{\phi}))} + \frac{C^2_-e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_-)(E-\omega^2_-(\boldsymbol{\phi}))}\right\}. \quad (12)$$

Case II: ($\mathbf{l}$ and $\mathbf{l}'$ both refer to M_2 atoms; $l+m+n$, $l'+m'+n'$ both odd)

$$G_0^{\mathbf{l},\mathbf{l}'} = G_0^{M_2M_2} = \frac{2}{N^3}\sum_{\boldsymbol{\phi}}\left\{\frac{e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_+)(E-\omega^2_+(\boldsymbol{\phi}))} + \frac{e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_-)(E-\omega^2_-(\boldsymbol{\phi}))}\right\}. \quad (13)$$

Case III: ($\mathbf{l}$ corresponds to M_1 atom, $\mathbf{l}'$ to M_2 atom)

$$G_0^{\mathbf{l},\mathbf{l}'} = G_0^{M_1M_2} = \frac{2}{N^3}\sum_{\boldsymbol{\phi}}\left\{\frac{C_+e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_+)(E-\omega^2_+)} + \frac{C_-e^{i(\mathbf{l}-\mathbf{l}')\cdot\boldsymbol{\phi}}}{(1+C^2_-)(E-\omega^2_-)}\right\}. \quad (14)$$

Let us rewrite the sums again using the following notations:

$$a_2 = 3\beta\left(\frac{1}{M_1}-\frac{1}{M_2}\right),$$

$$b_1 = \frac{2\beta}{\sqrt{M_1M_2}}(\cos\phi_j+\cos\phi_k+\cos\phi_l),$$

$$a_1 = E-3\beta\left(\frac{1}{M_1}+\frac{1}{M_2}\right),$$

$$S = \sqrt{a_2^2+b_1^2}.$$

In the above notations, the three sums in Eqs. (12)–(14) simplify to

$$G_0^{M_1M_1} = \frac{2}{N^3}\sum_{\boldsymbol{\phi}}e^{i\boldsymbol{\phi}\cdot(\mathbf{l}-\mathbf{l}')}\left\{\frac{b_1}{-a_1^2+a_2^2+b_1^2}\right\}, \quad (15)$$

$$G_0^{M_2M_2} = \frac{2}{N^3}\sum_{\boldsymbol{\phi}}e^{i\boldsymbol{\phi}\cdot(\mathbf{l}-\mathbf{l}')}\left\{\frac{a_1+a_2}{a_1^2-a_2^2-b_1^2}\right\}, \quad (16)$$

$$G_0^{M_1M_2} = \frac{2}{N^3}\sum_{\boldsymbol{\phi}}e^{i\boldsymbol{\phi}\cdot(\mathbf{l}-\mathbf{l}')}\left\{\frac{a_1-a_2}{a_1^2-a_2^2-b_1^2}\right\}. \quad (17)$$

As a check on these Green's functions, we have used Eq. (8.6.49) in the book by Maradudin *et al.*[5] to calculate the mean square displacement of an M_1 and M_2 atom. The results are

$$\langle u_x^2\rangle_{M_1} = \frac{\hbar}{2(N+1)^3}\int_0^{\omega_L^2}\frac{g(\omega^2)[M_2\omega^2-6\beta]d\omega^2}{[M_1M_2\omega^2-3\beta(M_1+M_2)]\omega\tanh(\hbar\omega/2kT)}, \quad (18)$$

and

$$\langle u_x^2\rangle_{M_2} = \frac{\hbar}{2(N+1)^3}\int_0^{\omega_L^2}\frac{g(\omega^2)[M_1\omega^2-6\beta]d\omega^2}{[M_1M_2\omega^2-3\beta(M_1+M_2)]\omega\tanh(\hbar\omega/2kT)}. \quad (19)$$

In Eqs. (18) and (19), $g(\omega^2)d\omega^2$ is the number of normal modes between ω^2 and $\omega^2+d\omega^2$. Equations (18) and (19) were first derived by Mazur.[6] In the classical limit ($T\to\infty$),we should have that

$$\langle u_x^2\rangle_{M_1}=\langle u_x^2\rangle_{M_2}. \tag{20}$$

Explicitly, we have for $T\to\infty$ in Eqs. (18) and (19) that

$$\langle u_x^2\rangle_{M_1}\approx\frac{kT}{(N+1)^3}\int_0^{\omega_L^2}\frac{g(\omega^2)[M_1\omega^2-6\beta]d\omega^2}{[M_1M_2\omega^2-3\beta(M_1+M_2)]\omega^2}, \tag{21}$$

and

$$\langle u_x^2\rangle_{M_2}\approx\frac{kT}{(N+1)^3}\int_0^{\omega_L^2}\frac{g(\omega^2)[M_2\omega^2-6\beta]d\omega^2}{[M_1M_2\omega^2-3\beta(M_1+M_2)]\omega^2}. \tag{22}$$

It is not obvious by inspection that the integrals in Eqs. (21) and (22) are equal to each other. However, by using a symmetry property of $g(\omega^2)$, we can demonstrate the equality of the two expressions. It is well known that the normal modes of a diatomic crystal separate themselves into two groups or bands separated by a gap. [It is easy to show in our case that $g(\omega^2)$ is (symmetrical about the middle of the gap when) plotted as a function of ω^2.] We now subtract Eq. (21) from Eq. (22) to obtain

$$\langle u_x^2\rangle_{M_2}-\langle u_x^2\rangle_{M_1}=\frac{kT}{(N+1)^3}\int_0^{\omega_L^2}\frac{g(\omega^2)(M_2-M_1)d\omega^2}{[M_2M_2\omega^2-3\beta(M_2+M_1)]}. \tag{23}$$

Since also the widths of the acoustical and optical bands are equal in units of ω^2, we have that

$$g(\omega^2)=g(\omega_L^2-\omega^2), \tag{24}$$

where ω_L is the highest frequency in the optical band. Changing the variable of integration from ω^2 to $(\omega')^2=\omega_L^2-\omega^2$, Eq. (23) becomes

$$\langle u_x^2\rangle_{M_2}-\langle u_x^2\rangle_{M_1}=\frac{kT}{(N+1)^3}\times\int_{\omega_L^2}^{0}\frac{g(\omega_L^2-(\omega')^2)(-d(\omega')^2)(M_2-M_1)}{[M_2M_1(\omega_L^2-(\omega')^2)-3\beta(M_2+M_1)]}. \tag{25}$$

Using Eq. (24) and the fact that $\omega_L^2=6\beta(1/M_1+1/M_2)$, Eq. (25) becomes

$$\begin{aligned}\langle u_x^2\rangle_{M_2}-\langle u_x^2\rangle_{M_1}&=\frac{+kT}{(N+1)^3}\times\int_{\omega_L^2}^{0}\frac{g(\omega^2)(-d\omega^2)(M_2-M_1)}{[-M_2M_1\omega^2+3\beta(M_2+M_1)]}\\&=-\frac{kT}{(N+1)^3}\times\int_0^{\omega_L^2}\frac{g(\omega^2)d\omega^2(M_2-M_1)}{[M_2M_1\omega^2-3\beta(M_2+M_1)]}.\end{aligned} \tag{26}$$

Comparing Eq. (26) with Eq. (23), we see immediately that

$$\langle u_x^2\rangle_{M_2}-\langle u_x^2\rangle_{M_1}=0. \tag{27}$$

The three-dimensional sums in Eqs. (15)–(17) can be replaced by 3-D integrals and one of the integrations can be carried out explicitly. We now do this for each of the sums. First, we note that $\phi=(2\pi j/N,2\pi k/N,2\pi s/N)$. Any two of the l, m, and n can be taken to go from 1 to N. The other goes from 1 to $N/2$. Each band contains $N^3/2$ frequencies and both bands together contain N^3 frequencies, which is correct. Eventually we want to consider a slab with l numbering the layers. We let j go from 1 to $N/2$. We sum over j in Eqs. (15)–(17). We let $\theta=2\pi j/N$ and θ goes from 0 to π and $\Sigma_{j=1}^{N/2}=(N/2\pi)\int_0^\pi d\theta$. First we calculate $G^{M_2M_2}$,

$$G^{M_2M_2}=\frac{-1}{\pi N^2}\sum_{k,s}e^{i\phi_k(m-m')}e^{i\phi_s(n-n')}g^{M_2M_2}(\phi_k,\phi_s), \tag{28}$$

with

$$\begin{aligned}g^{M_2M_2}(\phi_k,\phi_s)&=\frac{-(a_1+a_2)}{2\sqrt{a_1^2-a_2^2}}\left\{\int_0^\pi d\phi\left[\frac{1}{b_1-\sqrt{a_1^2-a_2^2}}-\frac{1}{b_1+\sqrt{a_1^2-a_2^2}}\right]e^{i\phi(l-l')}\right\},\\&=\frac{-(a_1+a_2)}{2\sqrt{a_1^2-a_2^2}\beta}\left[\frac{a_+\pi}{\sqrt{1-a_+^2}}\left(\frac{\sqrt{1-a_+^2}-1}{a_+}\right)^{|l-l'|}-\frac{a_-\pi}{\sqrt{1-a_-^2}}\left(\frac{\sqrt{1-a_-^2}-1}{a_-}\right)^{|l-l'|}\right],\end{aligned} \tag{29}$$

with

$$a_+=\beta/\left[\beta(\cos\phi_k+\cos\phi_s)-\sqrt{a_1^2-a_2^2}\right],$$

and

$$a_-=\beta/\left[\beta(\cos\phi_k+\cos\phi_s)+\sqrt{a_1^2-a_2^2}\right].$$

It is clear from Eqs. (16) and (17) that

$$G^{M_1M_2}=\left(\frac{a_1-a_2}{a_1+a_2}\right)G^{M_2M_2}. \tag{30}$$

Note that when $M_1=M_2$, we have $a_2=0$ and we obtain the obvious result.

Finally, for $G^{M_1M_1}$, we have

$$G^{M_1M_1} = \frac{1}{\pi N^2}\sum_{k,s} e^{i\phi_k(m-m')}e^{i\phi_s(n-n')}\int_0^{\pi}\frac{b_1 e^{i\theta(l-l')}d\theta}{-a_1^2+a_2^2+b_1^2}$$

$$= \frac{1}{2\pi N^2}\sum_{k,s} e^{i\phi_k(m-m')}e^{i\phi_s(n-n')}\int_0^{\pi} d\theta\, e^{i\theta(l-l')}\left[\frac{1}{b_1+\sqrt{a_1^2-a_2^2}}+\frac{1}{b_1-\sqrt{a_1^2-a_2^2}}\right]$$

$$= \frac{1}{2\pi N^2}\sum_{k,s} e^{i\phi_k(m-m')}e^{i\phi_s(n-n')}g^{M_1M_2}(\phi_k,\phi_s), \tag{31}$$

with

$$g^{M_1M_1}(\phi_k,\phi_s) = \left[\frac{a_+\pi}{\sqrt{1-a_+^2}}\left(\frac{\sqrt{1-a_+^2}-1}{a_+}\right)^{|l-l'|}+\frac{a_-\pi}{\sqrt{1-a_-^2}}\left(\frac{\sqrt{1-a_-^2}-1}{a_-}\right)^{|l-l'|}\right].$$

It is not immediately clear that $G^{M_1M_2} = G^{M_1M_1}$ when $M_1 = M_2$. In fact, it is not true. To see this, note that $l+m+n$ is even and $l'+m'+n'$ is odd in $G^{M_1M_2}$, while these sums are both even or both odd for $G^{M_1M_1}$ and $G^{M_2M_2}$, respectively. Now l and l' appear in the G's through the combinations $l-l'$, $m-m'$, and $n-n'$. But these quantities in $G^{M_1M_2}$ can never have the same values of these quantities when they occur in the expressions for $G^{M_1M_2}$ and $G^{M_1M_1}$. So simply setting $M_2 = M_1$ in $G^{M_1M_2}$ does not give $G^{M_1M_1}$.

We now proceed to obtain the Green's functions for the semi-infinite medium. The equations of motion, (1) and (2), can be written in terms of the reduced displacements, $V(l,m,n)$, as a matrix equation:

$$\sum_{l',m',n'} D_0(l,m,n,l',m',n')v(l'm'n') = 0,$$

with

$$\mathsf{D}_0(l,m,n,l'm'n')$$
$$= \left(\frac{6\beta}{M_1}-\omega^2\right)\delta_{ll'}\delta_{mm'}\delta_{nn'} - \frac{\beta}{\sqrt{M_1M_2}}(\delta_{l+1,l'}\delta_{mm'}\delta_{nn'}$$
$$+ \delta_{l-1,l'}\delta_{mm'}\delta_{nn'} + \delta_{ll'}\delta_{m+1,m'}\delta_{nn'}$$
$$+ \delta_{ll'}\delta_{m-1,m'}\delta_{nn'} + \delta_{ll'}\delta_{mm'}\delta_{m+1,m'}$$
$$+ \delta_{ll'}\delta_{mm'}\delta_{n-1,n'}), \quad l+m+n \text{ even}. \tag{32}$$

Substitute M_1 for M_2 and M_2 for M_1 in Eq. (32) to obtain the matrix D_0, for $l+m+n$ odd.

We now break the bonds between the layers labeled $n=0$ and $n=1$. The new dynamical matrix, D_s, is given by

$$D_s(l,m,n,l',m',n')$$
$$= D_0(lm,n,l',m',n') + V_s(lm,n,l',m',n'), \tag{33}$$

where

$$V_s(l,m,n,l',m',n')$$
$$= \frac{-\beta}{M_1}\delta_{0n}\delta_{l,l'}\delta_{mm'}\delta_{0n'} - \frac{\beta}{M_2}\delta_{ll'}\delta_{mm'}\delta_{1n'}\delta_{1n}$$
$$+ \frac{\beta}{\sqrt{M_1M_2}}\{\delta_{ll'}\delta_{mm'}\delta_{0n'}\delta_{1n}$$
$$+ \delta_{ll'}\delta_{mm'}\delta_{1n'}\delta_{0n}\}, \quad l+m \text{ is even}, \tag{34}$$

and

$$V_s(l,m,n,l',m',n')$$
$$= -\frac{\beta}{M_2}\delta_{ll'}\delta_{mm'}\delta_{0n}\delta_{0n'} - \frac{\beta}{M_1}\delta_{ll'}\delta_{mm'}\delta_{1n}\delta_{1n'}$$
$$+ \frac{\beta}{\sqrt{M_1M_2}}\{\delta_{ll}\delta_{mm'}\delta_{1n}\delta_{0n'}$$
$$+ \delta_{ll'}\delta_{mm'}\delta_{0n}\delta_{1n'}\}, \quad l+m \text{ is odd}. \tag{35}$$

The Green's function for our semi-infinite medium, G_s, is now determined through Dysons' equation,

$$\mathsf{G}_s = \mathsf{G}_0 - \mathsf{G}_0\mathsf{V}_s\mathsf{G}_s. \tag{36}$$

Writing out Eq. (36) explicitly, we obtain

$$G_s(l,m,n,l',m',n')$$
$$= G_0(l,m,n,l',m',n') - \sum_{\substack{l''m''n''\\ l'''m'''n'''}} G_0(l,m,n,l''m''n'')$$
$$\times V_s(l''m''n'''lmn)G_s(lmnl'm'n'). \tag{37}$$

The form of V_s changes depending on whether $l''+m''$ is even or odd. Substituting for V_s its two forms we expand Eq. (37) to obtain

$$G_s(l,m,n,l'm'n')$$
$$= G_0(l,m,n,l',m',n') - \sum_{\substack{l''m''n''\\ l'''m'''n'''\\ l''+m'' \text{ even}}} G_0(l,m,n,l'',m'',n'')\left[\frac{-\beta}{M_1}\delta_{0n''}\delta_{l''l'''}\delta_{m''m'''}\delta_{0n} - \frac{\beta}{M_2}\delta_{l''l'''}\delta_{m''m'''}\delta_{1n'''}\delta_{1n''}\right.$$
$$\left. + \frac{\beta}{\sqrt{M_1M_2}}\{\delta_{l''l'''}\delta_{m''m'''}\delta_{1n''}\delta_{0n'''} + \delta_{l''l'''}\delta_{m''m'''}\delta_{0n''}\delta_{1n'''}\}\right]G_s(lmnl'm'n')\}$$
$$- \sum_{\substack{l''m''n''\\ l'''m'''n'''\\ l''+m'' \text{ odd}}} G_0(l,m,n,l'',m'',n'')\left[\frac{-\beta}{M_2}\delta_{l''l'''}\delta_{m''m'''}\delta_{0n''}\delta_{0n'''} - \frac{\beta}{M_1}\delta_{l''l'''}\delta_{m''m'''}\delta_{1n'''}\delta_{1n''}\right.$$
$$\left. + \frac{\beta}{\sqrt{M_1M_2}}\{\delta_{l''l'''}\delta_{m''m'''}\delta_{1n''}\delta_{0n'''} + \delta_{l''l'''}\delta_{m''m'''}\delta_{0n''}\delta_{1n'''}\}\right]G_s(l'''m'''n'''l'm'n'). \tag{38}$$

In Eq. (38), we can sum over l''', m''', n''', and n'' immediately to give

$$
\begin{aligned}
G_s(l,m,n,l',m',n') = {} & G_0(l,m,n,l,m',n') \\
& - \sum_{\substack{l'',m'' \\ l''+m'' \text{ even}}} \Big[G_0(l,m,n,l'',m'',0) G_s(l'',m'',0,l',m',n') \left(\frac{-\beta}{M_1}\right) \\
& - \left(\frac{\beta}{M_2}\right) G_0(l,m,n,l'',m'',n'',1) G_s(l'',m'',1,l'm'n') \\
& + \frac{\beta}{\sqrt{M_1 M_2}} \{ G_0(l,m,n,l'',n'',1) G_s(l'',m'',0,l',m',n') + G_0(l,m,n,l'',m'',0) G_s(l'',m'',1,l',m',n') \} \Big] \\
& - \sum_{\substack{l'',m'' \\ l''+m'' \text{ odd}}} \Big[-\frac{\beta}{M_2} G_0(l,m,n,l'',m'',0) G_s(l'',m'',0',m',n') \\
& - \frac{\beta}{M_1} G_0(l,m,n,l'',m'',1) G_s(l'',m'',1,l',m',n') + \frac{\beta}{\sqrt{M_1 M_2}} \{ G_0(l,m,n,l'',m'',1) G_s(l'',m'',0,l',m',n') \\
& + G_0(l,m,n,l'',m'',0) G_s(l'',m'',1,l',m',n') \} \Big] . \qquad (39)
\end{aligned}
$$

The G_0's appearing in Eq. (39) are either $G^{M_1M_1}$, $G^{M_2M_2}$, or $G^{M_1M_2}$ depending on the oddness or evenness of $l+m+n$ and $l'+m'+n'$. First, we take $l+m+n$ and $l'+m'+n'$ both even. Since we still have periodicity in the x and y directions after breaking of the bond we can Fourier analyze $G_s(l,m,n,l',m',n')$. Let

$$
\begin{aligned}
G_s(l,m,n,l',m',n') &= \frac{1}{N^2} \sum_{\phi_x \phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_s^{M_1M_1}(\phi_x,\phi_y,n,n',E), \\
G_0(l,m,n,l',m',n') &= \frac{1}{N^2} \sum_{\phi_x \phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_0^{M_1M_1}(\phi_x,\phi_y,n,n',E). \qquad (40)
\end{aligned}
$$

We also need the following factors which appear in Eq. (39), with $l''+m''$:

$$
G_0(l,m,n,l'',m'',0) = \frac{1}{N^2} \sum_{\phi_x \phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_0^{M_1M_1}(n,0), \qquad (41a)
$$

$$
G_s(l'',m'',0,l',m',n') = \frac{1}{N^2} \sum_{\phi_x,\phi_y} e^{i\phi_x(l''-l)} e^{i\phi_y(m''-m')} g_s^{M_1M_1}(0,n'), \qquad (41b)
$$

$$
G_0(l,m,n,l'',m'',1) = \frac{1}{N^2} \sum_{\phi_x,\phi_y} e^{i\phi_x(l-l'')} e^{i\phi_y(m-m'')} g_0^{M_1M_2}(n,1), \qquad (41c)
$$

$$
G_s(l'',m'',1,l',m',n') = \frac{1}{N^2} \sum_{\phi_x,\phi_y} e^{i\phi_x(l''-l')} e^{i\phi_y(m''-m')} g_s^{M_2M_1}(1,n'). \qquad (41d)
$$

Substituting Eqs. (40) and (41a)–(41d) into Eq. (39), we obtain

$$
\begin{aligned}
& \frac{1}{N^2} \sum_{\phi_x,\phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_s^{M_1M_1}(n,n') \\
& = \frac{1}{N^2} \sum_{\phi_x,\phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_0^{M_1M_1}(n,n') + \frac{\beta}{N^4} \sum_{\substack{\phi_x,\phi_y \\ \phi_x',\phi_y' \\ l'',m'' \\ l''+m'' \text{ even}}} \Big[e^{i\phi_x(l-l')} e^{i\phi_y(m-m'')} e^{i\phi_x'(l''-l')} e^{i\phi_y'(m''-m')} \\
& \times \left\{ \frac{g_0^{M_1M_1}(n,0) g_s^{M_1M_1}(0,n')}{M_1} + \frac{g_0^{M_1M_1}(n,1) g_s^{M_2M_1}(1,n')}{M_2} - \frac{1}{\sqrt{M_1M_2}} g_0^{M_1M_2}(n,1) g_s^{M_1M_1}(0,n') \right. \\
& \left. - \frac{1}{\sqrt{M_1M_2}} g_0^{M_1M_1}(n,0) g_s^{M_2M_1}(1,n') \right\} \Big] + \frac{\beta}{N^4} \sum_{\substack{\phi_x,\phi_y \\ \phi_x',\phi_y' \\ l'',m'' \\ l''+m'' \text{ odd}}} \Big[e^{i\phi_x(l-l')} e^{i\phi_y(m-m'')} e^{i\phi_x'(l''-l')} e^{i\phi_y'(m''-m')} \\
& \times \left\{ \frac{g_0^{M_1M_2}(n,0) g_s^{M_2M_1}(0,n')}{M_2} + \frac{g_0^{M_1M_1}(n,1) g_s^{M_1M_1}(1,n')}{M_1} - \frac{g_0^{M_1M_1}(n,1) g_s^{M_2M_1}(0,n')}{\sqrt{M_1M_2}} - \frac{g_0^{M_1M_2}(n,0) g_s^{M_1M_1}(1,n')}{\sqrt{M_1M_2}} \right\} \Big] . \qquad (42)
\end{aligned}
$$

We now show that the sums over l'' and m'' demand that $\phi_x = \phi_x'$ and $\phi_y = \phi_y'$. We must show that

$$\sum_{\substack{l'' \\ l''+m'' \text{ is even}}} e^{il''(\phi_x - \phi_x')} = 0 \quad \text{unless } \phi_x = \phi_x'.$$

Note that as we do the sum over l'', m'' is considered to be a constant (odd integer or even integer). If we choose m'' as an odd integer, then l'' must go over the odd integers from 1 to N. On the other hand, if we choose m'' as an even integer, then l'' goes over the even integers from 1 to N.

First, let N be even and m'' be even. Then

$$\sum_{\substack{l'' \text{ even} \\ N \text{ even}}} e^{i2\pi jl''/N} = \sum_{n=1}^{N/2} e^{i2\pi j2n/N} = \frac{e^{i4\pi j/N} - e^{i(4\pi j/N)(N/2+1)}}{1 - e^{i4\pi j/N}} = 0,$$

unless $j = 0$ or $\phi_x = \phi_x'$. Since $\Sigma_{l=1}^{N} e^{i2\pi jl''/N} = 0$, when l'' goes over all integers from 1 to N we can conclude that

$$\sum_{\substack{l'' \text{ odd} \\ N \text{ even}}} e^{i2\pi jl''/N} = 0 \text{ also} \quad \text{unless } \phi_x = \phi_x'.$$

Setting $\phi_x = \phi_x'$ and $\phi_y = \phi_y'$ and equating Fourier coefficients on both sides of Eq. (42), we get

$$\begin{aligned} g_s^{M_1M_1}(n,n') = {} & g_0^{M_1M_1}(n,n') + \frac{\beta}{M_1}\frac{1}{2} g_0^{M_1M_1}(n,0) g_s^{M_1M_1}(0,n') \\ & + \frac{\beta}{M_2}\frac{1}{2} g_0^{M_1M_2}(n,1) g_s^{M_2M_1}(1,n') - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(n,1) g_s^{M_1M_1}(0,n') \\ & - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(n,0) g_s^{M_2M_1}(1,n') + \frac{\beta}{M_2}\frac{1}{2} g_0^{M_1M_2}(n,0) g_s^{M_2M_1}(0,n') \\ & + \frac{\beta}{M_1}\frac{1}{2} g_0^{M_1M_1}(n,1) g_s^{M_1M_1}(1,n') - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(n,1) g_s^{M_2M_1}(0,n') \\ & - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(n,0) g_s^{M_1M_1}(1,n'). \end{aligned} \tag{43}$$

In g_s, we want both n and n' to be greater than or equal to 1. Because of this criterion, four terms in the above equation are dropped. The above equation, therefore, simplifies to

$$\begin{aligned} g_s^{M_1M_1}(n,n') = {} & g_0^{M_1M_1}(n,n') + \frac{\beta}{2M_2} g_0^{M_1M_2}(n,1) g_s^{M_2M_1}(1,n') + \frac{\beta}{2M_1} g_0^{M_1M_1}(n,1) g_s^{M_1M_1}(1,n') \\ & - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(n,0) g_s^{M_2M_1}(1,n') - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(n,0) g_s^{M_1M_1}(1,n'). \end{aligned} \tag{44}$$

Remember, n,n' can be any integers greater than or equal to 1. What we would like to do is to find $G_s^{M_1M_1}(n,n')$ in terms of the G_0's. But we need other equations similar to Eq. (42). We now derive the equation for $G_s^{M_2M_1}(n,n')$. We start out with $G_s(l,m,n,l',m',n')$ with $l+m+n$ odd and $l'+m'+n'$ even. We start with Eq. (38) again. We need to Fourier analyze the Green's functions again.

For $l+m+n$ odd and $l'+m'+n'$ even, we get

$$G_s(l,m,n,l',m',n') = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l-l')} e^{i\phi_y(m-m')} g_s^{M_2M_1}(n,n'), \tag{45a}$$

$$G_0(l,m,n,l'',m'',0) = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')} e^{i\phi_y(m-m'')} g_0^{M_2M_1}(n,0), \quad \text{for } l''+m'' \text{ even}, \tag{45b}$$

$$G_0(l,m,n,l'',m'',0) = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')} e^{i\phi_y(m-m'')} g_0^{M_2M_2}(n,0), \quad \text{for } l''+m'' \text{ odd}, \tag{45c}$$

$$G_s(l'',m'',0,l',m',n') = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l''-l')} e^{i\phi_y(m''-m')} g_s^{M_1M_1}(0,n'), \quad \text{for } l''+m'' \text{ even}, \tag{45d}$$

$$G_s(l'',m'',0,l',m',n') = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l''-l')} e^{i\phi_y(m''-m')} g_s^{M_2M_1}(0,n'), \quad \text{for } l''+m'' \text{ odd}, \tag{45e}$$

$$G_0(l,m,n,l'',m'',1) = \frac{1}{N^2} \sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')} e^{i\phi_y(m-m'')} g_0^{M_2M_2}(n,1), \quad \text{for } l''+m'' \text{ even}, \tag{45f}$$

$$G_0(l,m,n,l'',m'',1)=\frac{1}{N^2}\sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')}e^{i\phi_y(m-m'')}g_0^{M_2M_1}(n,1), \quad \text{for } l''+m'' \text{ odd}, \tag{45g}$$

$$G_s(l'',m'',1,l',m',n')=\frac{1}{N^2}\sum_{\phi_x\phi_y} e^{i\phi_x(l''-l')}e^{i\phi_y(m''-m')}g_s^{M_2M_1}(1,n'), \quad \text{for } l''+m'' \text{ even}, \tag{45h}$$

$$G_s(l'',m'',1,l',m',n')=\frac{1}{N^2}\sum_{\phi_x\phi_y} e^{i\phi_x(l''-l')}e^{i\phi_y(m''-m')}g_s^{M_1M_1}(1,n'), \quad \text{for } l''+m'' \text{ odd}, \tag{45i}$$

$$G_0(l,m,n,l'',m'',1)=\frac{1}{N^2}\sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')}e^{i\phi_y(m-m'')}g_0^{M_2M_2}(n,1), \quad \text{for } l''+m'' \text{ even}, \tag{45j}$$

$$G_0(l,m,n,l'',m'',1)=\frac{1}{N^2}\sum_{\phi_x\phi_y} e^{i\phi_x(l-l'')}e^{i\phi_y(m-m'')}g_0^{M_2M_1}(n,1), \quad \text{for } l''+m'' \text{odd}.$$

Substituting the above expressions into Eq. (39), we get an equation similar to Eq. (42) and then Eq. (43),

$$\begin{aligned} g_s^{M_2M_1}(n,n') &= g_0^{M_2M_1}(n,n')+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,0)g_s^{M_1M_1}(0,n')+\frac{\beta}{2M_2}g_0^{M_2M_2}(n,1)g_s^{M_2M_1}(1,n') \\ &\quad -\frac{\beta}{\sqrt{M_1M_2}}g_0^{M_2M_2}(n,1)g_s^{M_2M_1}(0,n')-\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_1}(n,0)g_s^{M_2M_1}(1,n') \\ &\quad +\frac{\beta}{2M_2}g_0^{M_2M_2}(n,0)g_s^{M_1M_1}(0,n')+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,1)g_s^{M_1M_1}(1,n') \\ &\quad -\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_2M_1}(n,1)g_s^{M_2M_1}(0,n')+g_0^{M_2M_2}(n,0)g_s^{M_1M_1}(1,n')\}. \end{aligned} \tag{46}$$

Again, since G_s is zero unless $n,n'\geqslant 1$, we strike out four terms [Eq. (46)] and obtain

$$\begin{aligned} g_s^{M_2M_1}(n,n') &= g_0^{M_2M_1}(n,n')+\frac{\beta}{2M_2}g_0^{M_2M_2}(n,1)g_s^{M_2M_1}(1,n')+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,1)g_s^{M_1M_1}(1,n') \\ &\quad -\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_1}(n,0)g_s^{M_2M_1}(1,n')-\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_2}(n,0)g_s^{M_1M_1}(1,n'). \end{aligned} \tag{47}$$

In Eqs. (44) and (47), we can have $n=1$. When we do this we get two equations that we have to solve simultaneously for $g_s^{M_2M_1}(1,n')$ and $g_s^{M_1M_1}(1,n')$. The equations are

$$\begin{aligned} &\left(1-\frac{\beta}{2M_1}g_0^{M_1M_1}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_1M_2}(1,0)\right)g_s^{M_1M_1}(1,n') \\ &\quad +\left(-\frac{\beta}{2M_2}g_0^{M_1M_2}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_1M_1}(1,0)\right)g_s^{M_2M_1}(1,n')=g_0^{M_1M_1}(1,n'), \end{aligned} \tag{48a}$$

$$\begin{aligned} &\left(-\frac{\beta}{2M_1}g_0^{M_2M_1}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_2}(1,0)\right)g_s^{M_1M_1}(1,n') \\ &\quad +\left(1-\frac{\beta}{2M_2}g_0^{M_2M_2}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_1}(1,0)\right)g_s^{M_2M_1}(1,n')=g_0^{M_2M_1}(1,n'). \end{aligned} \tag{48b}$$

One can solve Eqs. (48) simultaneously for $g_s^{M_1M_1}(1,n')$ and $g_s^{M_2M_1}(1,n')$ and substitute back into Eqs. (44) and (47). When $M_1=M_2=M$, both Eqs. (48) give

$$g_s(1,n')=\frac{g_0(1,n')}{1-(\beta/M)\{g_0(1,1)-g_0(1,0)\}}, \tag{49}$$

which is the monatomic case.

We now repeat the arguments leading to Eqs. (48) to find determining equations for $g_s^{M_2M_2}(n,n')$. We again start with Eq. (39) but for the case that both $l+m+n$ and $l'+m'+n'$ are odd. In this case, the Fourier transformed Green's functions needed are the following.

For $l''+m''$ even,

$$\begin{aligned} &G_s(l,m,n,l',m',n')\to g_s^{M_2M_2}(n,n'), \quad G_0(l,m,n,l'',m'',0)\to g_0^{M_2M_1}(n,0), \\ &G_s(l'',m'',0,l',m',n')\to g_s^{M_1M_2}(0,n'), \quad G_0(l,m,n,l'',m'',1)\to g_0^{M_2M_2}(n,1), \\ &G_s(l'',m'',1,l',m',n')\to g_s^{M_2M_2}(1,n'). \end{aligned}$$

For $l''+m''$ odd,

$$G_s(l,m,n,l'm',n')\to g_s^{M_2M_2}(n,n'),\quad G_0(l,m,n,l'',m'',0)\to g_0^{M_2M_2}(n,0),$$
$$G_s(l'',m'',0,l',m',n')\to g_s^{M_2M_2}(n,0),\quad G_0(l,m,n,l'',m'',1)\to g_0^{M_2M_1}(n,1),$$
$$G_s(l'',m'',1,l',m',n')\to g_s^{M_1M_2}(1,n').$$

Equation (39) then gives, after simplification,

$$\begin{aligned}g_s^{M_2M_2}(n,n') &= g_0^{M_2M_2}(n,n')+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,0)g_s^{M_1M_2}(0,n')+\frac{\beta}{2M_2}g_0^{M_2M_2}(n,1)g_s^{M_2M_2}(1,n')\\&\quad-\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_2}(n,1)g_s^{M_2M_2}(0,n')-\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_1}(n,0)g_s^{M_2M_2}(1,n')+\frac{\beta}{2M_2}g_0^{M_2M_2}(n,0)g_s^{M_2M_2}(n,0)\\&\quad+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,1)g_s^{M_1M_2}(1,n')-\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_2M_1}(n,1)g_s^{M_2M_2}(n,0)+g_0^{M_2M_2}(n,0)g_s^{M_1M_2}(1,n')\}.\end{aligned}\tag{50}$$

After eliminating the four terms which connect g_s with the other layer we get

$$\begin{aligned}g_s^{M_2M_2}(n,n') &= g_0^{M_2M_2}(n,n')+\frac{\beta}{2M_2}g_0^{M_2M_2}(n,1)g_s^{M_2M_2}(1,n')+\frac{\beta}{2M_1}g_0^{M_2M_1}(n,1)g_s^{M_1M_2}(1,n')\\&\quad-\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_2M_1}(n,0)g_s^{M_2M_2}(1,n')+g_0^{M_2M_2}(n,0)g_s^{M_1M_2}(1,n')\}.\end{aligned}\tag{51}$$

We now do the case for $l+m+n$ even and $l'+m'+n'$ odd to obtain a second equation which can be solved simultaneously with Eq. (51). We now give the Green's functions needed.

For $l''+m''$ even,

$$G_s(l,m,n,l',m',n')\to g_s^{M_1M_2}(n,n'),\quad G_0(l,m,n,l'',m'',0)\to g_0^{M_1M_1}(n,0),$$
$$G_s(l'',m'',0,l',m',n')\to g_s^{M_1M_2}(0,n'),\quad G_0(l,m,n,l'',m''1)\to g_0^{M_1M_2}(n,1),$$
$$G_s(l'',m'',1,l',m',n')\to g_s^{M_2M_2}(1,n).$$

For $l''+m''$ odd,

$$G_s(l,m,n,l',m',n')\to g_s^{M_1M_2}(n,n'),\quad G_0(l,m,n,l''m'',0)\to g_0^{M_1M_2}(n,0),$$
$$G_s(l'',m'',0,l',m',n')\to g_s^{M_2M_2}(n,0),\quad G_0(l,m,n,l'',m'',1)\to g_0^{M_1M_1}(n,1),$$
$$G_s(l'',m'',1,l',m',n')\to g_s^{M_1M_2}(1,n').$$

Equation (39) then gives, after simplification,

$$\begin{aligned}g_s^{M_1M_2}(n,n') &= g_0^{M_1M_2}(n,n')+\frac{\beta}{2M_1}g_0^{M_1M_1}(n,0)g_s^{M_1M_2}(0,n')+\frac{\beta}{2M_2}g_0^{M_1M_2}(n,1)g_s^{M_2M_2}(1,n')\\&\quad-\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_1M_2}(n,1)g_s^{M_1M_2}(0,n')+g_0^{M_1M_1}(n,0)g_s^{M_2M_2}(1,n')\}+\frac{\beta}{2M_2}g_0^{M_1M_2}(n,0)g_s^{M_2M_2}(0,n')\\&\quad-\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_1M_1}(n,1)g_s^{M_2M_2}(n,0)+g_0^{M_1M_2}(n,0)g_s^{M_1M_2}(1,n')\}+\frac{\beta}{2M_1}g_0^{M_1M_1}(n,1)g_s^{M_1M_2}(1,n'),\end{aligned}\tag{52}$$

and after striking out terms like $g_s(0,n')$ we obtain

$$\begin{aligned}g_s^{M_1M_2}(n,n') &= g_0^{M_1M_2}(n,n')+\frac{\beta}{2M_2}g_0^{M_1M_2}(n,1)g_s^{M_2M_2}(1,n')+\frac{\beta}{2M_1}g_0^{M_1M_1}(n,1)g_s^{M_1M_2}(1,n')\\&\quad-\frac{\beta}{2\sqrt{M_1M_2}}\{g_0^{M_1M_2}(n,0)g_s^{M_1M_2}(1,n')+g_0^{M_1M_1}(n,0)g_s^{M_2M_2}(1,n')\}.\end{aligned}\tag{53}$$

By letting $n=1$ in Eqs. (51) and (53), we can solve them simultaneously for $g_s^{M_1M_2}(1,n')$ and $g_s^{M_2M_2}(1,n')$. Rearranging terms we rewrite (51) and (53) as follows (after setting $n=1$):

$$\begin{aligned}&\left[1-\frac{\beta}{2M_2}g_0^{M_2M_2}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_1}(1,0)\right]g_s^{M_2M_2}(1,n')\\&+\left[-\frac{\beta}{2M_1}g_0^{M_2M_1}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_2M_2}(1,0)\right]g_s^{M_1M_2}(1,n')=g_0^{M_2M_2}(1,n'),\end{aligned}\tag{54}$$

$$\left[-\frac{\beta}{2M_2}g_0^{M_1M_2}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_1M_1}(1,0)\right]g_s^{M_2M_2}(1,n')$$
$$+\left[1-\frac{\beta}{2M_1}g_0^{M_1M_1}(1,1)+\frac{\beta}{2\sqrt{M_1M_2}}g_0^{M_1M_2}(1,0)\right]g_s^{M_1M_2}(1,n')=g_0^{M_1M_2}(1,n'). \tag{55}$$

If we let $M_1=M_2$ in Eqs. (52) and (53), either of the equations give

$$g_s(1,n')=\frac{g_0(1,n')}{1-(\beta/M)g_0(1,1)+(\beta/M)g_0(1,0)},$$

which is correct. We now have effectively all of the Green's functions for the semi-infinite medium written in terms of the g_0's.

To finally form the slab we break the bonds between the L th layer and the $L+1$st layer and use the Dyson equation again. As in the semi-infinite medium case, we obtain four equations relating the semi-infinite Green's functions to the slab Green's functions,

$$U^{M_1M_2}(n,n')=g_s^{M_1M_2}(n,n')+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_2}(n,L+1)+\frac{\beta}{2M_1}g_s^{M_1M_1}(n,L)\right\}U^{M_1M_2}(L,n')$$
$$+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_1}(n,L+1)+\frac{\beta}{2M_2}g_s^{M_1M_2}(n,L)\right\}U^{M_2M_2}(L,n'), \tag{56}$$

$$U^{M_2M_2}(n,n')=g_s^{M_2M_2}(n,n')+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_2}(n,L+1)+\frac{\beta}{2M_1}g_s^{M_2M_1}(n,L)\right\}U^{M_1M_2}(L,n')$$
$$+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_1}(n,L+1)+\frac{\beta}{2M_2}g_s^{M_2M_2}(n,L)\right\}U^{M_2M_2}(L,n'), \tag{57}$$

$$U^{M_2M_1}(n,n')=g_s^{M_2M_1}(n,n')+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_2}(n,L+1)+\frac{\beta}{2M_1}g_s^{M_2M_1}(n,L)\right\}U^{M_1M_1}(L,n')$$
$$+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_1}(n,L+1)+\frac{\beta}{2M_2}g_s^{M_2M_2}(n,L)\right\}U^{M_2M_1}(L,n'), \tag{58}$$

$$U^{M_1M_1}(n,n')=g_s^{M_1M_1}(n,n')+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_2}(n,L+1)+\frac{\beta}{2M_1}g_s^{M_1M_1}(n,L)\right\}U^{M_1M_1(L,n')}$$
$$+\left\{-\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_1}(n,L+1)+\frac{\beta}{2M_2}g_s^{M_1M_2}(n,L)\right\}U^{M_2M_1}(L,n'). \tag{59}$$

In Eqs. (56)–(59) we have that $1\leqslant n,n'\leqslant L$.

$U^{M_1M_2}(L,n')$, $U^{M_2M_2}(L,n')$, $U^{M_1M_1}(L,n')$, and $U^{M_2M_1}(L,n')$ on the right-hand side of Eqs. (56)–(59) are determined by setting $n=L$ in the same equations. We obtain

$$\left\{1-\frac{\beta}{2M_1}g_s^{M_1M_1}(L,L)+\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_2}(L,L+1)\right\}U^{M_1M_2}(L,n')$$
$$+\left\{\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_1}(L,L+1)-\frac{\beta}{2M_2}g_s^{M_1M_2}(L,L)\right\}U^{M_2M_2}(L,n')=g_s^{M_1M_2}(L,n'), \tag{60}$$

$$\left\{1-\frac{\beta}{2M_2}g_s^{M_2M_2}(L,L)+\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_1}(L,L+1)\right\}U^{M_2M_2}(L,n')$$
$$+\left\{\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_2}(L,L+1)-\frac{\beta}{2M_1}g_s^{M_2M_1}(L,L)\right\}U^{M_1M_2}(L,n')=g_s^{M_2M_2}(L,n'), \tag{61}$$

$$\left\{1-\frac{\beta}{2M_2}g_s^{M_2M_2}(L,L)+\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_1}(L,L+1)\right\}U^{M_2M_1}(L,n')$$
$$+\left\{\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_2M_2}(L,L+1)-\frac{\beta}{2M_1}g_s^{M_2M_1}(L,L)\right\}U^{M_1M_1}(L,n')=g_s^{M_2M_1}(L,n'), \tag{62}$$

$$\left\{1+\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_2}(L,L+1)-\frac{\beta}{2M_1}g_s^{M_1M_1}(L,L)\right\}U^{M_1M_1}(L,n')$$
$$+\left\{\frac{\beta}{2\sqrt{M_1M_2}}g_s^{M_1M_1}(L,L+1)-\frac{\beta}{2M_2}g_s^{M_1M_2}(L,L)\right\}U^{M_2M_1}(L,n')=g_s^{M_1M_1}(L,n'). \tag{63}$$

Equations (60) and (61) can be solved simultaneously for $U^{M_1M_2}(L,n')$ and $U^{M_2M_2}(L,n')$ while Eqs. (62) and (63) can

be solved for $U^{M_2M_1}(L,n')$ and $U^{M_1M_1}(L,n')$. Since all of the U's are now known in terms of the g_s's and the g_s's are known in terms of the g_0's, we have in hand explicit expressions for the slab Green's functions.

Solving Eqs. (60) and (61) we obtain

$$U^{M_2M_2}(L,n') = \frac{A_1F_1 - D_1E_1}{A_1C_1 - B_1D_1}; \quad U^{M_1M_2}(L,n') = \frac{C_1E_1 - B_1F_1}{A_1C_1 - B_1D_1},$$

where

$$A_1 = 1 - \frac{\beta}{2M_1} g_s^{M_1M_1}(L,L) + \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_1M_2}(L,L+1),$$

$$B_1 = \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_1M_1}(L,L+1) - \frac{\beta}{2M_2} g_s^{M_1M_2}(L,L),$$

$$C_1 = 1 - \frac{\beta}{2M_2} g_s^{M_2M_2}(L,L) + \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_2M_1}(L,L+1),$$

$$D_1 = \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_2M_2}(L,L+1) - \frac{\beta}{2M_1} g_s^{M_2M_1}(L,L),$$

$$E_1 = g_s^{M_1M_2}(L,n'),$$

$$F_1 = g_s^{M_2M_2}(L,n').$$

We need the following g_s's:

$$g_s^{M_1M_1}(L,L) = g_0^{M_1M_1}(L,L) + \frac{\beta}{2M_2} g_0^{M_1M_2}(L,1) g_s^{M_2M_2}(1,L) + \frac{\beta}{2M_1} g_0^{M_1M_1}(L,1) g_s^{M_1M_1}(1,L)$$
$$- \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(L,0) g_s^{M_2M_1}(1,L) - \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(L,0) g_s^{M_1M_1}(1,L),$$

$$g_s^{M_1M_1}(L,L+1) = g_0^{M_1M_1}(L,L+1) - \frac{\beta}{2M_2} g_0^{M_1M_2}(L,1) g_s^{M_2M_1}(1,L+1) - \frac{\beta}{2M_1} g_0^{M_1M_1}(L,1) g_s^{M_1M_1}(1,L+1)$$
$$+ \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(L,0) g_s^{M_2M_1}(1,L+1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(L,0) g_s^{M_1M_1}(1,L+1),$$

$$g_s^{M_1M_2}(L,L+1) = g_0^{M_1M_2}(L,L+1) - \frac{\beta}{2M_2} g_0^{M_1M_2}(L,1) g_s^{M_2M_2}(1,L+1) - \frac{\beta}{2M_1} g_0^{M_1M_1}(L,1) g_s^{M_1M_2}(1,L+1)$$
$$+ \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_1M_2}(L,0) g_s^{M_1M_2}(1,L) + g_0^{M_1M_1}(L,0) g_s^{M_2M_2}(1,L+1)\},$$

$$g_s^{M_1M_2}(L,L) = g_0^{M_1M_2}(L,L) + \frac{\beta}{2M_2} g_0^{M_1M_2}(L,1) g_s^{M_1M_2}(L,1) + \frac{\beta}{2M_1} g_0^{M_1M_1}(L,1) g_s^{M_1M_2}(1,L)$$
$$- \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_1M_2}(L,0) g_s^{M_1M_2}(1,L) + g_0^{M_1M_1}(L,0) g_s^{M_2M_2}(1,L)\},$$

$$g_s^{M_2M_2}(L,L) = g_0^{M_2M_2}(L,L) + \frac{\beta}{2M_2} g_0^{M_2M_2}(L,1) g_s^{M_2M_2}(1,L) + \frac{\beta}{2M_1} g_0^{M_2M_1}(L,1) g_s^{M_1M_2}(1,L)$$
$$- \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_2M_1}(L,0) g_s^{M_2M_2}(1,L) + g_0^{M_2M_2}(L,0) g_s^{M_1M_2}(1,L)\},$$

$$g_s^{M_2M_2}(L,L+1) = g_0^{M_2M_2}(L,L+1) - \frac{\beta}{2M_2} g_0^{M_2M_2}(L,1) g_s^{M_2M_2}(1,L+1) - \frac{\beta}{2M_1} g_0^{M_2M_1}(L,1) g_s^{M_1M_2}(1,L+1)$$
$$+ \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_2M_1}(L,0) g_s^{M_2M_2}(1,L+1) + g_0^{M_2M_2}(L,0) g_s^{M_1M_2}(1,L+1)\},$$

$$g_s^{M_2M_1}(L,L) = g_0^{M_2M_1}(L,L) - \frac{\beta}{2M_2} g_0^{M_2M_2}(L,1) g_s^{M_2M_1}(1,L) - \frac{\beta}{2M_1} g_0^{M_2M_1}(L,1) g_s^{M_1M_1}(1,L)$$
$$+ \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_2M_1}(L,0) g_s^{M_2M_1}(1,L) + g_0^{M_2M_1}(L,0) g_s^{M_1M_1}(1,L)\},$$

$$g_s^{M_2M_1}(L,L+1) = g_0^{M_2M_1}(L,L+1) + \frac{\beta}{2M_2} g_0^{M_2M_2}(L,1) g_s^{M_2M_1}(1,L+1) + \frac{\beta}{2M_1} g_0^{M_2M_1}(L,1) g_s^{M_1M_1}(1,L+1)$$
$$- \frac{\beta}{2\sqrt{M_1M_2}} \{g_0^{M_2M_1}(L,0) g_s^{M_2M_1}(1,L) + g_0^{M_2M_2}(L,0) g_s^{M_1M_1}(1,L)\}.$$

We now write down the expressions for $g_s^{M_1M_1}(1,n')$, $g_s^{M_2M_1}(1,n')$, $g_s^{M_2M_2}(1,n')$, and $g_s^{M_1M_2}(1,n')$:

$$g_s^{M_1M_1}(1,n') = \frac{E_2D_2 - F_2B_2}{A_2D_2 - B_2C_2};$$

$$g_s^{M_2M_1}(1,n') = \frac{A_2F_2 - E_2C_2}{A_2D_2 - B_2C_2},$$

$$A_2 = 1 - \frac{\beta}{2M_1} g_0^{M_1M_1}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(1,0),$$

$$B_2 = -\frac{\beta}{2M_2} g_0^{M_1M_2}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(1,0),$$

$$C_2 = -\frac{\beta}{2M_1} g_0^{M_2M_1}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_2M_2}(1,0),$$

$$D_2 = 1 - \frac{\beta}{2M_2} g_0^{M_2M_2}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_2M_1}(1,0),$$

$$E_2 = g_0^{M_1M_1}(1,n'),$$

$$F_2 = g_0^{M_2M_1}(1,n').$$

We also need the following g_s's:

$$g_s^{M_2M_1}(1,L) = \frac{A_2 g_0^{M_2M_1}(1,L) - C_2^{M_1M_1}(1,L)}{A_2D_2 - B_2C_2},$$

$$g_s^{M_1M_1}(1,L) = \frac{D_2 g_0^{M_2M_1}(1,L) - B_2 g_0^{M_2M_1}(1,L)}{A_2D_2 - B_2C_2},$$

$$g_s^{M_2M_2}(1,n') = \frac{E_3D_3 - B_3F_3}{A_3D_3 - B_3C_3};$$

$$g_s^{M_1M_2}(1,n') = \frac{A_3F_3 - C_3E_3}{A_3D_3 - B_3C_3},$$

$$A_3 = 1 - \frac{\beta}{2M_2} g_0^{M_2M_2}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_2M_1}(1,0),$$

$$B_3 = -\frac{\beta}{2M_1} g_0^{M_2M_1}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_2M_2}(1,0),$$

$$C_3 = -\frac{\beta}{2M_2} g_0^{M_1M_2}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_1}(1,0),$$

$$D_3 = 1 - \frac{\beta}{2M_1} g_0^{M_1M_1}(1,1) + \frac{\beta}{2\sqrt{M_1M_2}} g_0^{M_1M_2}(1,0),$$

$$E_3 = g_0^{M_2M_2}(1,n'),$$

$$F_3 = g_0^{M_1M_2}(1,n'),$$

$$g_s^{M_2M_2}(1,L) = \frac{D_3 g_0^{M_2M_2}(1,L) - B_3 g^{M_1M_2}(1,L)}{A_3D_3 - B_3C_3},$$

$$g_s^{M_1M_2}(1,L) = \frac{A_3 g_0^{M_1M_2}(1,L) - C_3 g^{M_2M_2}(1,L)}{A_3D_3 - B_3C_3}.$$

Solving Eqs. (62) and (63) for $U^{M_2M_1}(L,n')$ and $U^{M_1M_1}(L,n')$, we obtain

$$U^{M_2M_1}(L,n') = \frac{E_4D_4 - B_4F_4}{A_4D_4 - B_4C_4}, \tag{64}$$

$$U^{M_1M_1}(L,n') = \frac{A_4F_4 - C_4E_4}{A_4D_4 - B_4C_4}, \tag{65}$$

with

$$A_4 = 1 - \frac{\beta}{2M_2} g_s^{M_2M_2}(L,L) + \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_2M_1}(L,L+1),$$

$$B_4 = \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_2M_2}(L,L+1) - \frac{\beta}{2M_1} g_s^{M_2M_1}(L,L),$$

$$C_4 = \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_1M_1}(L,L+1) - \frac{\beta}{2M_2} g_s^{M_1M_2}(L,L),$$

$$D_4 = 1 + \frac{\beta}{2\sqrt{M_1M_2}} g_s^{M_1M_2}(L,L+1) - \frac{\beta}{2M_1} g_s^{M_1M_1}(L,L),$$

$$E_4 = g_s^{M_2M_1}(L,n'),$$

$$F_4 = g_s^{M_1M_1}(L,n').$$

We now have in hand explicitly all of the infinite lattice and semi-infinite lattice Green's functions to determine our slab Green's functions, U. Mazur and Maradudin[7] calculated analytically the root-mean square displacement of atoms in a slab for the monatomic case in the high-temperature limit. The mass dependence disappears in this limit so nothing new would be obtained using the present diatomic Green's function. Mazur[3] calculated analytically the low-temperature specific heat for the monatomic slab. Our intention is to use a computer to calculate the low-temperature specific heat of a diatomic slab and (as a by-product when $M_1 = M_2$) compare also the numerical calculation with Mazur's analytic result.

APPENDIX: AN EXAMPLE OF QUANTUM TO CLASSICAL LIMIT PERTINENT TO EQ. (27)

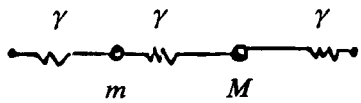
We present an elementary example illustrating a general principle of quantum mechanics that quantum mechanical results, such as, for example, the averages of dynamical quantities, must go to well-known classical results in certain limits. In the context of this paper we are thinking, in particular, of the mean-square displacements of atoms from their equilibrium positions in a crystal lattice. Classically, the mean-square displacement x_i of the ith atom from its equilibrium position in the x direction is given by

$$\langle x_i^2 \rangle = \frac{\int x_i^2 \exp[-E(\mathbf{x}_i,\mathbf{p}_i)/kT]\, d\mathbf{x}_1\, d\mathbf{x}_2 \cdots d\mathbf{x}_N\, d\mathbf{p}_1\, d\mathbf{p}_2 \cdots d\mathbf{p}_N}{\int \exp[-E(\mathbf{x}_i,\mathbf{p}_i)/kT]\, d\mathbf{x}_1\, d\mathbf{x}_2 \cdots d\mathbf{x}_N\, d\mathbf{p}_1\, d\mathbf{p}_2 \cdots d\mathbf{p}_N}, \tag{A1}$$

where T is the absolute temperature, $\mathbf{x}_i = (x_i, y_i, z_i)$, $\mathbf{p}_i = (p_{xi}, p_{yi}, p_{zi})$, and E is the total energy of the lattice. The corresponding quantum mechanical average should become equal to the above classical result when T becomes large.

Further, if the potential energy part of E is invariant to an exchange of two particular atoms, then the mean-square displacements of these two atoms are the same, independent of their individual masses. This can be seen immediately from

Eq. (A1) if the potential energy part of E does not contain the masses, which is the case for the most useful forms of the crystal potential. Our model consists of two atoms



connected by a spring. Each atom is also anchored by a similar spring to a fixed point as seen in the diagram. Each spring has the same force constant γ. Let u and v be the displacements from equilibrium of the two masses, m and M, respectively. The Lagrangian for the system is

$$L = \tfrac{1}{2}m\dot{u}^2 + \tfrac{1}{2}M\dot{v}^2 - \tfrac{1}{2}\gamma u^2 - \tfrac{1}{2}\gamma v^2 - \tfrac{1}{2}\gamma(u-v)^2. \quad \text{(A2)}$$

This Lagrangian yields the following equations of motion:

$$m\ddot{u} = -\gamma u + \gamma(u-v), \quad \text{(A3)}$$

$$M\ddot{v} = -\gamma v + \gamma(u-v), \quad \text{(A4)}$$

Transforming coordinates, the Lagrangian becomes

$$L = \frac{1}{2}\dot{x}_1^2 + \frac{1}{2}\dot{x}_2^2 - \frac{\gamma}{m}x_1^2 - \frac{\gamma}{M}x_2^2 + \frac{\gamma}{\sqrt{Mm}}x_1x_2, \quad \text{(A5)}$$

with $u = x_1\sqrt{m}$ and $v = x_2/\sqrt{M}$.

Letting $x_1 = A\exp(i\omega t)$ and $x_2 = B\exp(i\omega t)$, the equations of motion from Eq. (A5) are

$$\frac{2\gamma}{m}A - \frac{\gamma}{\sqrt{Mm}}B = \omega^2 A, \quad \text{(A6)}$$

$$\frac{-\gamma}{\sqrt{Mm}}A + \frac{2\gamma}{M}B = \omega^2 B. \quad \text{(A7)}$$

Equations (A6) and (A7) yield two solutions for ω^2, which are

$$\omega_+^2 = (\gamma/Mm)\{M+m+\sqrt{M^2+m^2-Mm}\} > 0, \quad \text{(A8)}$$

$$\omega_-^2 = (\gamma/Mm)\{M+m-\sqrt{M^2+m^2-Mm}\} > 0, \quad \text{(A9)}$$

with

$$A_+ = \frac{\gamma/\sqrt{Mm}}{2\gamma/m - \omega_+^2}B_+; \quad A_- = \frac{\gamma/\sqrt{Mm}}{2\gamma/m - \omega_-^2}B_-. \quad \text{(A10)}$$

Set $B_+ = B_- = 1$ and transforms to a new set of coordinates,

$$x_1 = \frac{A_+y_1}{\sqrt{A_+^2+1}} + \frac{A_-y_2}{\sqrt{A_-^2+1}}, \quad \text{(A11)}$$

$$x_2 = \frac{y_1}{\sqrt{A_+^2+1}} + \frac{y_2}{\sqrt{A_-^2+1}}. \quad \text{(A12)}$$

In these new coordinates, the Schrödinger Hamiltonian operator becomes

$$H = -\frac{\hbar^2}{2}\left\{b_1\frac{\partial^2\psi}{\partial y_1^2} + b_2\frac{\partial^2\psi}{\partial y_2^2}\right\} + \frac{a_1y_1^2}{2} + \frac{a_2y_2^2}{2}, \quad \text{(A13)}$$

where a_1, b_1, a_2, and b_2 are functions of m, M, and γ. The Schrödinger equation gives two harmonic oscillator-type equations,

$$-\frac{\hbar^2}{2}b_1\frac{\partial^2\psi_1}{\partial y_1^2} + \frac{a_1}{2}y_1^2\psi_1 = E_1\psi_1; \quad \text{(A14)}$$

$$E_{1n} = \left(n+\frac{1}{2}\right)\hbar\sqrt{a_1b_1},\ \psi_{1n}(y_1),$$

$$-\frac{\hbar^2}{2}b_2\frac{\partial^2\psi_2}{\partial y_2^2} + \frac{a_2}{2}y_2^2\psi_2 = E_2\psi_2 = E_2\psi_2; \quad \text{(A15)}$$

$$E_{2n} = \left(n+\frac{1}{2}\right)\hbar\sqrt{a_2b_2},\ \psi_{2n}(y_2).$$

Our complete eigenfunctions and energies for our system are given by

$$\psi_{mn} = \psi_{1m}(y_1)\psi_{2n}(y_2); \quad E_{mn} = E_{1m} + E_{2n}. \quad \text{(A16)}$$

The mean-square displacement of the m atom is given by

$$\langle x_1^2\rangle_{mn} = \iint dy_1\,dy_2\,\psi_{1m}^*(y_1)\psi_{2n}^*(y_2)x_1^2\psi_{1m}(y_1)\psi_{2n}(y_2)$$

$$= \frac{A_+^2\langle y_1^2\rangle_m}{A_+^2+1} + \frac{\langle y_2^2\rangle_n}{A_-^2+1}, \quad \text{(A17)}$$

with

$$\langle y_1^2\rangle_n = \left(n+\frac{1}{2}\right)\frac{\hbar h_1}{\sqrt{a_1b_1}};$$

$$\langle y_2^2\rangle_n = \left(n+\frac{1}{2}\right)\frac{\hbar b_2}{\sqrt{a_2b_2}}. \quad \text{(A18)}$$

The thermal average of $\langle x_1^2\rangle$ is given by

$$\langle\langle x_1^2\rangle\rangle = \frac{\Sigma_{mn}\langle x_1^2\rangle_{mn}\exp(-E_{mn}/kT)}{\Sigma_{mn}\exp(-E_{mn}kT)}$$

$$= \frac{A_+^2}{A_+^2+1}\frac{\hbar b_1}{\sqrt{a_1b_1}}\frac{1}{2}\coth\frac{\alpha_1}{2}$$

$$+ \frac{A_-^2}{A_-^2+1}\frac{\hbar b_2}{\sqrt{a_2b_2}}\frac{1}{2}\coth\frac{\alpha_2}{2}, \quad \text{(A19)}$$

with

$$\alpha_1 = \frac{\hbar\sqrt{a_1b_1}}{kT}; \quad \alpha_2 = \frac{\hbar\sqrt{a_2b_2}}{\tau}.$$

As the temperature, T, goes to infinity the mean-square displacement becomes

$$\langle\langle x_1^2\rangle\rangle_{T\to\infty} \to \left[\frac{A_+^2}{(A_+^2+1)a_1} + \frac{A_-^2}{(A_-^2+1)a_2}\right]kT, \quad \text{(A20)}$$

with

$$a_1 = \frac{\gamma}{m}\frac{A_+^2}{A_+^2+1} + \frac{\gamma}{M}\frac{1}{A_+^2+1} - \frac{\gamma}{\sqrt{Mm}}\frac{A_+}{A_+^2+1},$$

$$a_2 = \frac{\gamma}{m}\frac{A_-^2}{A_-^2+1} + \frac{\gamma}{M}\frac{1}{A_-^2+1} - \frac{\gamma}{\sqrt{Mm}}\frac{A_-}{A_-^2+1},$$

$$A_+ = \left[\sqrt{Mm}\left\{\frac{1}{m} - \frac{1}{M} - \frac{\sqrt{M^2+m^2-Mm}}{Mm}\right\}\right]^{-1},$$

$$A_- = \left[\sqrt{Mm}\left\{\frac{1}{m} - \frac{1}{M} + \frac{\sqrt{M^2+m^2-Mm}}{Mm}\right\}\right]^{-1}.$$

We then have that

$$\frac{1}{kT}\langle\langle u^2\rangle\rangle = \left[\frac{A^2_+}{(A^2_+ + 1)a_1 m} + \frac{A^2_-}{(A^2_- + 1)a_2 m}\right]. \tag{A21}$$

The bracket in Eq. (A21) can be shown to be equal to $4/3\gamma$ which is independent of m and M, which is the result we wanted to show.

[1]E. W. Montroll and R. B. Potts, Phys. Rev. **100**, 525 (1955).
[2]E. W. Montroll, J. Chem.Phys. **15**, 575 (1947).
[3]L. van Hove, Phys. Rev. **89**, 1189 (1953).
[4]L. Dobrzynski, B. Djafari-Rouhani, and O. Hardouin Duparc, Phys. Rev. B **29**, 3138 (1984).
[5]A. A. Maradudin *et al.*, *Theory of Lattice Dynamics in The Harmonic Approximation* (Academic, New York, 1971).
[6]P. Mazur, Ph.D. thesis, University of Maryland, 1957 (unpublished).
[7]P. Mazur and A. A. Maradudin, Phys. Rev. B **24**, 2996 (1981).
[8]P. Mazur, Phys. Rev. B. **23**, 6503 (1981).

Physical condition for elimination of ambiguity in conditionally convergent lattice sums

K. Young
Department of Physics, The Chinese University of Hong Kong, Hong Kong

(Received 25 March 1986; accepted for publication 24 September 1986)

The conditional convergence of the lattice sum defining the Madelung constant gives rise to an ambiguity in its value. It is shown that this ambiguity is related, through a simple and universal integral, to the average charge density on the crystal surface. The physically correct value is obtained by setting the charge density to zero. A simple and universally applicable formula for the Madelung constant is derived as a consequence. It consists of adding up dipole–dipole energies together with a nontrivial correction term.

I. INTRODUCTION

It is well known that the lattice sum defining the Madelung constant[1,2] is conditionally convergent, so that its value is not mathematically unique. Recently an unambiguous definition has been provided using analytic continuation[3] in which the Coulomb potential r^{-1} is replaced by r^{-s} and the corresponding lattice sum is investigated in the complex s plane, giving mathematical precision lacking in earlier works using direct summation.[4]

The selection of one out of infinitely many possible values must correspond to some *physical* assumption about the crystal and our purpose is to complement the work on analytic continuation by addressing the question in a physical manner. The first observation is that real crystals are large but finite, so the conditional convergence for the infinite sum translates into the possible dependence of the corresponding finite sum on the *shape* (and size) of the crystal. We shall show that the shape dependence can be isolated into a simple integral I over the surface of the crystal. More importantly, the integral is *universal*, being independent of the details of the crystal structure and *macroscopic*, being determined only by the averaged surface charge density. The analysis clarifies not only the question of convergence (which is widely discussed in the literature[1–5]) but also the question of convergence to the *physically correct* value (which is quite distinct and in general ignored). In fact much of the early work,[1–9] done in days of limited computational power, was concerned with *rate* of convergence, which is now increasingly irrelevant. In contrast, the algorithm developed here is extremely simple to implement.

II. FORMALISM

Consider a finite crystal made up of ions with charges $\pm q$ at positions $\mathbf{r}_i$. The energy required to remove the ion i is, in units of q^2,

$$\alpha = \sum_{j\neq i} \frac{s(i,j)}{r_{ij}}, \tag{1}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $s(i,j) = -1(+1)$ if the charges i and j have the same (opposite) charge.

In general it will be possible to organize the charges into pairs forming dipoles with separation $\mathbf{d}$ between the opposite charges. The case of NaCl-type crystals is shown in Fig. 1. Now

$$\alpha = 1/d + \beta/2, \tag{2}$$

where β is the energy required to remove one dipole from the crystal. Choose the center of one dipole as the origin and let $\mathbf{r}$ be the center of another dipole. The attractive energy between the pair is

$$U(\mathbf{r}) = \frac{-2}{r} + \frac{1}{|\mathbf{r}-\mathbf{d}|} + \frac{1}{|\mathbf{r}+\mathbf{d}|}. \tag{3}$$

Then

$$\beta = \sum_{\mathbf{r}} U(\mathbf{r}). \tag{4}$$

Next choose some $R \gg 1$ and break β into $\beta = \beta_1(R) + \beta_2(R)$, where the two terms correspond to the sum in (4) being restricted to $r < R$ and $r > R$, respectively. The first sum presents no problem and in the second sum we expand $|\mathbf{r} \pm \mathbf{d}|^{-1}$ in powers of r^{-1} to get

$$\beta_2(R) = 2 \sum_{\substack{l\neq 0 \\ l \text{ even}}} S_l(R), \tag{5}$$

where

$$S_l(R) = \sum_{r>R} \frac{d^l}{r^{l+1}} P_l(\cos\theta), \tag{6}$$

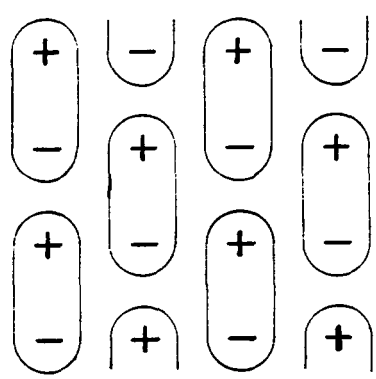


FIG. 1. Dipole pairs in NaCl.

 0022-2488/87/020425-03$02.50 © 1987 American Institute of Physics

and θ is the polar angle of $\mathbf{r}$, and $\mathbf{d}$ is assumed to be along $+z$. Each sum in (6) is *finite*, since r is restricted to within the surface of the crystal. However for $l \geqslant 4$, the sum (6) is absolutely convergent so the upper limit may be extended to infinity. Moreover, $S(R) = O(R^{-l+2})$, so the sum in (5) is rapidly convergent.

Any problem then lies in $S_2(R)$. To evaluate it first consider the sum between the surface $r = R$ and another spherical surface inside the crystal at $r = R'$, where R' is of macroscopic dimensions (Fig. 2). This sum will, in general, be zero by symmetry of the crystal and the region of summation (e.g., C_4 symmetry for a cubic crystal). Therefore the sum for S_2 may start at the macroscopic distance R', and hence equals the following integral between $r = R'$ and Σ:

$$S_2(R) = \rho d^2 \int dV \frac{P_2(\cos\theta)}{r^3}, \tag{7}$$

where ρ is the average density of dipoles. The average (i.e., coarse-grained) value ρ may be used because fluctuations in ρ give a contribution proportional to the gradient of the integrand, which is of order r^{-4}, integrating to $O(1/R')$, where R' is macroscopic. Use the identity

$$\frac{P_2(\cos\theta)}{r^3} = \frac{1}{2}\left(\frac{\partial}{\partial z}\right)^2 \frac{1}{r}$$

to convert (7) to surface integrals

$$S_2(R) = -\frac{1}{2}\rho d^2 \int_{R'} d\mathbf{S}\cdot\hat{n}_3 \frac{\partial}{\partial z}\frac{1}{r} + \frac{1}{2}\rho d^2 \int_{\Sigma} d\mathbf{S}\cdot\hat{n}_3 \frac{\partial}{\partial z}\frac{1}{r}, \tag{8}$$

where $d\mathbf{S}$ in each case points away from the origin. The first term is readily evaluated to be $(2\pi/3)\rho d^2$ and we denote the second term as I_1. However, it is not strictly true (as we have assumed so far) that all changes can be organized into dipoles. There are "leftover" charges at the surface (Fig. 3), with surface charge density σ' (which consists of two-dimensional δ-functions representing point charges). They contribute an extra electric field E' at the origin:

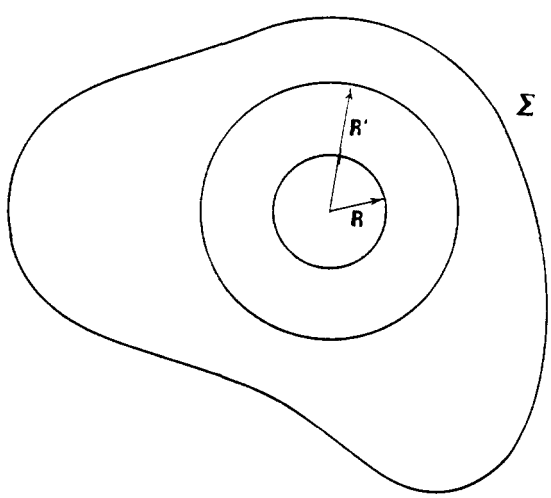


FIG. 2. The mathematical surfaces R and R' inside the crystal surface Σ.

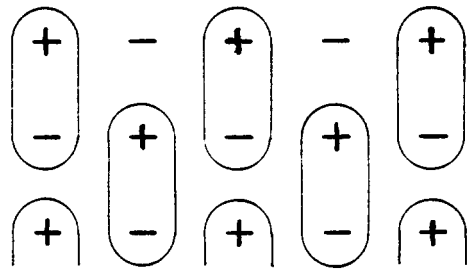


FIG. 3. Leftover single ions at the surface.

$$E'_z = \int_{\Sigma} dS\,\sigma' \frac{\partial}{\partial z}\frac{1}{r}.$$

This then contributes to the potential energy β of the dipole at the origin a term $\beta = E'_z d$. Thus α contains an extra term $I_2 = \frac{1}{2}E'_z d$, and

$$\alpha(\Sigma) = \frac{1}{d} + \frac{1}{2}\sum_{r<R} U(\mathbf{r}) + \sum_{l=4,6,\ldots} S_l(R) + \frac{2\pi}{3}\rho d^2 + I(\Sigma), \tag{9}$$

where

$$I(\Sigma) = I_1 + I_2 = \frac{d}{2}\int_{\Sigma} dS(\rho\mathbf{d}\cdot\hat{n} + \sigma')\frac{\partial}{\partial z}\frac{1}{r}, \tag{10}$$

in which $\hat{n}$ is the normal to Σ, and we emphasize that I and hence α may depend on the shape of the crystal.

First, σ' can be replaced by its average $\langle\sigma'\rangle$ over several lattice distances along the surface, since the difference (due to fluctuations in σ') contributes a term proportional to

$$\int_{\Sigma} dS\,\partial_\alpha\partial_\beta\frac{1}{r},$$

which is negligible since $dS \sim L^2$ and $\partial_\alpha\partial_\beta(1/r) \sim L^{-3}$, where $L \gg 1$ is the size of the crystal. Second, $\rho\mathbf{d}\cdot\hat{n}$ is the average surface density of (positive) dipole ends on Σ, while $\langle\sigma'\rangle$ is the average surface density of "leftover" charge. Their sum is the average surface charge σ on Σ. Therefore

$$I(\Sigma) = \frac{d}{2}\int_{\Sigma} dS\,\sigma\frac{\partial}{\partial z}\frac{1}{r}. \tag{11}$$

Thus I depends only on the *macroscopic* property of the crystal. Moreover, I is *universal* in that it does not depend on the crystal structure.

Finally, I is unchanged if Σ is scaled up keeping σ constant. In other words I is independent of size; it is at most dependent on shape.

The removal of the shape-dependent ambiguity is now achieved by the *physical* assumption that $\sigma = 0$. Incidentally this condition would not hold if the crystal is polarized.

In evaluating (9), any choice of $R > 1$ may be used, and some choice of R will optimize computational efficiency. But since computational time is now hardly a problem, we may as well take $R \to \infty$ to obtain the extremely simple algorithm

$$\alpha = \frac{1}{d} + \frac{1}{2}\sum_{\mathbf{r}}^{\infty} U(\mathbf{r}) + \frac{2\pi}{3}\rho d^2, \tag{12}$$

where the infinite sum is to be performed by expanding spheres of radii $R \to \infty$. Note that $U(\mathbf{r}) \propto r^{-4}$ at large r, therefore the sum is clearly convergent. Incidentally this justifies extending the sum to infinity. The rate of convergence is controlled by the first remainder $S_4 = O(R^{-2})$. The practical advantage of (12) is its extremely simple form and its universal applicability to all crystal types.

Physically this algorithm corresponds to adding up the dipole–dipole interaction energies. What we have demonstrated are two features that are nonintuitive: (a) this sum, which is nominally logarithmic divergent by power counting, is actually convergent; and (b) however, a correction term $(2\pi/3)\rho d^2$ is necessary. Note the difference between (12) and (4): the sum in (4) is a *finite* sum, whereas the one in (12) is an infinite sum performed by expanding spheres. The correction term is the difference between the two. Incidentally, this illustrates the difference between convergence [e.g., (12) without the correction term] and convergence to the correct value.

As a check, we have evaluated (12) for CsCl-type crystals, for which $\rho = \frac{1}{8}$, $\mathbf{d} = (1,1,1)$, and $\mathbf{r} = (x,y,z)$, where x,y,z are even integers (not all zero). We find $\Sigma U(\mathbf{r}) = -0.69013$, so that $\alpha = 1.0177 = 1.7627/d$, in agreement with the known value.[2]

III. DISCUSSION

We first make explicit the connection between shape dependence of the finite sum and the conditional convergence of the infinite sum. Consider an infinite crystal and perform the sum by the method of expanding Σ shapes,[3] i.e., evaluate α as

$$\alpha = \alpha_\infty \equiv \lim_{k\to\infty} \alpha(k\Sigma),$$

where $k\Sigma$ denotes the shape Σ scaled up k times. Different choices of Σ correspond to different ordering of terms, so α_∞ is dependent on the ordering of terms if and only if $\alpha(\Sigma)$ is dependent on Σ.

The above analysis shows that in the method of expanding shapes, only those Σ with $\sigma = 0$ will converge to the correct physical value. Thus for a NaCl-type crystal, the naive use of expanding spheres is *in*appropriate, since the average charge density on a spherical surface is nonzero. [However, the situation is salvaged if we count dipoles and include the correction term as in (12).] On the other hand, expanding cubes will give the correct value, since the average charge density on a plane is zero.

The comparison between different expanding shapes has been emphasized,[3] but there is one slight difference in detail in the present treatment. We use dipoles as units so σ interacts with the central dipole through the electric field $E \propto \sigma/r^2$ and the surface contribution is finite as $L \to \infty$ but possibly nonunique. If we use single ions as units,[3] σ interacts with the central charge through the potential $\Phi \propto 1/r$. The surface contribution therefore scales as L and would *diverge* unless $\sigma = 0$.

An obvious method of direct summation is to group the ions into cells in such a way as to remove the leading multipoles.[4,5] In this context it has been emphasized (at least for CsCl-type crystals) that *cells* must have surface neutrality in order to guarantee convergence,[5] which is of course closely related to the surface neutrality of the *crystal* emphasized here.

Another method of evaluating such lattice sums is to use a Fourier series,[6] assuming that the potential has the same periodicity as the lattice—which would not be true if there is any surface charge. Thus the correct value is always based on the same physical condition.

For lattice sums in two dimensions involving r^{-1}, the analog of S_2 would be absolutely convergent and there is no shape dependence. However, if the two-dimensional Coulomb potential $\ln r$ is used instead, then the situation would be essentially the same as that discussed here.

In conclusion we have shown how the ambiguity due to the conditional convergence is related to the macroscopic boundary condition on the crystal surface and emphasized how the latter singles out the *correct* value for such lattice sums. A simple and universal algorithm based on adding up dipole–dipole interaction energies—but with a nontrivial correction term—has been derived.

[1] von E. Madelung, Phys. Z. **19**, 524 (1918).
[2] Elementary accounts may be found in, e.g., C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1953).
[3] D. Borwein, J. M. Borwein, and K. F. Taylor, J. Math. Phys. **26**, 2999 (1985).
[4] H. Evjen, Phys. Rev. **39**, 675 (1932).
[5] S. K. Roy, Can. J. Phys. **32**, 509 (1954).
[6] P. P. Ewald, Ann. Phys. (Leipzig) **64**, 253 (1921).
[7] H. Kornfield, Z. Phys. **22**, 27 (1924).
[8] F. C. Frank, Philos. Mag. **41**, 1287 (1950).
[9] J. Kanamori, T. Moriya, K. Matizuki, and T. Nagamiya, J. Phys. Soc. Jpn. **10**, 93 (1955).

Yang–Mills theory and the Batalin–Fradkin–Vilkovisky formalism

David McMullan[a)]
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 19 November 1985; accepted for publication 15 October 1986)

In this paper a generalized dynamical description of classical Yang–Mills theory will be presented. As a consequence of this work, a firm dynamical underpinning to the Batalin–Fradkin–Vilkovisky (BFV) formalism will emerge. This in turn will supply a precise geometric characterization of the ghost and conjugate ghost fields.

I. INTRODUCTION

Local gauge invariance is used to great effect in the attempts to build realistic models of the (nongravitational) interactions observed in nature. However, there is a technical price to pay for the usefulness of these theories since a consequence of the gauge invariance is that the dynamical content is obscured. Thus, care must be taken in both the classical and quantum description of such theories.

In order to construct a covariant, unitary path integral for these theories, one has to use an effective Lagrangian[1] composed of the original Yang–Mills Lagrangian, a covariant gauge fixing part, and a term containing ghost and antighost fields (scalar fields with fermionic statistics). The validity of this construction can be shown in perturbation theory[2,3] and hence, from a pragmatic point of view, Yang–Mills theory is well understood.

It was noticed[4,5] that the effective action possessed a new global symmetry, called Becchi–Rouet–Stora–Tyutin (BRST) invariance, under which the ghost field had a very definite transformation property. This lead many people[6–13] to speculate on the possibility of a nonperturbative understanding of this construction. Over the last few years, two well developed descriptions of the ghost fields have emerged. In one,[8–11] the ghosts are identified with Maurer–Cartan forms on a suitable group, whereas in the other approach,[12,13] a superspace formalism is developed in which the ghost field is related to a special connection on a super principle bundle. The usefulness of these interpretations depends very much on the type of problem one wishes to solve. In particular, the superspace approach is the closest in spirit to the way ghosts are introduced into the path integral, whereas the Maurer–Cartan form interpretation leads to a powerful cohomology construction,[7,11] which in turn throws some light on the geometric origins of chiral anomalies (see also, Refs. 14 and 15). However, both approaches suffer from the drawbacks that they provide no greater understanding of the dynamical role ghosts play in Yang–Mills theory. Indeed, these constructions are carried out in a purely classical framework and yet they do not appear to have any role in our understanding of classical Yang–Mills theory. Also, both approaches have difficulty incorporating the antighosts in a natural way.

The classical dynamics of Yang–Mills theory is best described within the Hamiltonian formulation. One finds that the local gauge invariance introduces constraints into the phase space and hence a generalized dynamics[16,17] must be implemented.

Using phase space methods, Batalin, Fradkin, and Vilkovisky[18,19] (BFV) developed a formalism whereby the effective Lagrangian for Yang–Mills theory could be derived from an effective Hamiltonian construction on a super phase space. There are many intriguing aspects to this work, not the least of which is the claim that it can be applied to theories like gravity, where the constraints are not related to any group action on the phase space.

The aim of this paper is to understand the BFV construction for Yang–Mills theory in terms of the generalized dynamics formalism developed by Dirac.[16] Thus a fully classical understanding of the phase space ghosts and conjugate ghosts will emerge and, as a consequence of this construction, they will automatically be supplied with both a super space and geometric interpretation. The precise relationship between these constructions and the fields in the effective Lagrangian will be presented in a later paper.

The classical observables on the Yang–Mills configuration space correspond to the gauge invariant function, thus one can use homological methods in a trivial way to describe them. However, lifting this construction to the phase space is complicated by the fact that the constraints have the effect of imposing only weak equations among the smooth functions. Thus a weak cohomology theory must be developed and we shall find that this construction supplies the homological interpretation for the phase space version of the BRST transformation.

The motivation for this work is twofold: First, we want to understand the content of the BFV construction when applied to gravity. In this theory it is well known[20,21] that the general coordinate transformations of the space-time have a complicated relationship with the symmetries generated in the phase space. Thus, it is important to understand the analogous problem of relating the phase space BRST invariance with that constructed from the Hilbert action for gravity, especially in the light of the nontrivial measures occurring in the BFV approach. Obviously it is a good idea to start with the simpler problem found in Yang–Mills theory. Second, the BRST charge has been used by several authors to give useful insights into the quantization of constrained systems.[22,23] However, these discussions have been heuristic in nature and a more detailed analysis is called for. Hence the full content of the classical construction must be made clear. In particular, the weak homological interpretation seems well suited to the recent work on the deformation approach to quantization[24] and the global, nonlinear quantization scheme.[25]

The plan of this paper is as follows: In Sec. II we shall

[a)] Present address: Department of Physics and Astronomy, The University, Glasgow G12 8QQ, United Kingdom.

 0022-2488/87/020428-10$02.50 © 1987 American Institute of Physics

present the main ideal of constrained dynamics as it applies to Yang–Mills theory. This will be followed, in Sec. III, with the weak homology construction alluded to above where the associated superphase space structure will be made clear. In the final section the dynamical content of this construction is presented along with various general ideas on constrained systems.

II. YANG–MILLS THEORY AS A CONSTRAINED DYNAMICAL SYSTEM

Let M be a given space-time with a compact spacelike Cauchy surface Σ and G a compact, semisimple Lie group with (dual) Lie algebra $\mathfrak{g}$ ($\mathfrak{g}^*$). Then the Yang–Mills fields on M, with structure group G, are conveniently described[26] as the space of connections associated with a principle G bundle over M. For the sake of simplicity, we shall only consider the trivial principle bundle.

The Hamiltonian description of Yang–Mills theory is achieved by using the Cauchy surface Σ to foliate M. Then the space-time fields are projected onto this foliation in order to separate the dynamical quantities from the kinematical. As is well known, the resulting phase space description has constraints and a generalized dynamics is needed.

If we are given manifolds D and B, we denote by $\Omega^r(D,B)$ the space of B-valued r forms on D. Then, in this canonical approach to Yang–Mills theory, the configuration space is identified with $\Omega^1(\Sigma,\mathfrak{g})$, which we denote as $\mathscr{A}$. Thus, locally $A\in\mathscr{A}$ can be written as $A = A_i^a\, e_a\, dx^i$ where x^i are local coordinates on Σ ($i = 1,2,3$) and e_a are a basis of $\mathfrak{g}$ ($a = 1,\ldots,k = \dim G$).

Let $\mathscr{G} = \Omega^0(\Sigma,G)$, then this is a Lie group with Lie algebra $E := \Omega^0(\Sigma,\mathfrak{g})$. There is a $\mathscr{G}$ action on $\mathscr{A}$ which generates the spatial gauge transformations. Thus we have a Lie algebra morphism

$$\gamma\colon E\to \text{Vect}\,\mathscr{A}\ (\text{Vector fields on }\mathscr{A})$$
$$\lambda\to\gamma(\lambda)$$

such that acting on $A\in\mathscr{A}$

$$\gamma(\lambda)(A) = (\nabla_i\lambda^a)e_a\, dx^i = (\lambda^a_{\,|i} + C^a_{bc}\, A_i^b\,\lambda^c)e_a\, dx^i, \tag{2.1}$$

where $\lambda = \lambda^a(x)e_a$ and $[e_a,e_b] = C^c_{ab}\, e_c$ with C^c_{ab} the structure constants of $\mathfrak{g}$.

Since $\mathscr{G}$ acts as a gauge group on $\mathscr{A}$, we known that the true configuration space must be identified with the space of orbits $\tilde{\mathscr{A}} = \mathscr{A}/\mathscr{G}$. In order for $\tilde{\mathscr{A}}$ to be a manifold one has to be careful about the choice of functional spaces $\mathscr{A}$ and $\mathscr{G}$ belong to[26,27] and also (due to the compactness of Σ) one needs to restrict $\mathscr{A}$ to the irreducible connections and remove the center of $\mathscr{G}$.[27] We shall assume that all these technical constructions have been carried out. Thus, $\mathscr{G}$ is taken to have free action on $\mathscr{A}$ and $\tilde{\mathscr{A}}$ is a smooth Hilbert manifold.

The true phase space is thus $T^*\tilde{\mathscr{A}}$ and, in principle, the dynamics of Yang–Mills theory takes place on this space. There are, however, drawbacks to this construction since $\tilde{\mathscr{A}}$ is an infinite-dimensional manifold and hence $T^*\tilde{\mathscr{A}}$ is not uniquely defined. Also $\tilde{\mathscr{A}}$, and hence $T^*\tilde{\mathscr{A}}$, is topologically nontrivial[27] and is thus impossible to parametrize with any physically relevant coordinates. These problems have limited the usefulness of this approach.

The most common description of constrained systems takes place on the phase space associated with the extended configuration space. So for Yang–Mills theory, we need a generalized dynamics on $T^*\mathscr{A}$. Again, since $\mathscr{A}$ is infinite dimensional, the construction of $T^*\mathscr{A}$ is not unique. However, now there is a natural choice of cotangent bundle since we require the associated Cauchy problem to be well posed.[28] This leads to the use of the L_2 dual.[29] Thus, we define $(\Omega^p(\Sigma,\mathfrak{g}))^* := \Omega^{p*}(\Sigma,\mathfrak{g}^*)$ as the L_2 dual of $\Omega^p(\Sigma,\mathfrak{g})$. So $\Omega^{1*}(\Sigma,\mathfrak{g})$ is the space of smooth $\mathfrak{g}^*$-valued vector densities on Σ. The L_2-cotangent bundle of $\mathscr{A}$ is then $T^*\mathscr{A}\approx\mathscr{A}\times\mathscr{E}$, where $\mathscr{E} := \Omega^{1*}(\Sigma,\mathfrak{g}^*)$ can be regarded as the space of generalized electric fields. Locally $\pi\in\mathscr{E}$ can be written as $\pi = \pi^i_a\, e^a(\partial/\partial x^i)$, where e^a is the basis of $\mathfrak{g}^*$ dual to e_a. Here $T^*\mathscr{A}$ comes equipped with a (weak) symplectic form,[28] and we shall regard it as the extended phase space for Yang–Mills theory.

Using standard phase techniques,[30] the action of $\mathscr{G}$ on $\mathscr{A}$ can be lifted to a symplectic action on $T^*\mathscr{A}$, where now the lift of γ is a map of $\tilde{\gamma}$ such that $\tilde{\gamma}\colon E\to\text{Ham}(T^*\mathscr{A})$, where $\text{Ham}(T^*\mathscr{A})$ denotes the Hamiltonian vector fields on $T^*\mathscr{A}$. Since the $\mathscr{G}$ action on $T^*\mathscr{A}$ is the lift of an action on $\mathscr{A}$, an equivariant momentum mapping Φ can be constructed.[30] We recall that $\Phi\in\Omega^0(T^*\mathscr{A},E^*)$, where $E^* := \Omega^0(\Sigma,\mathfrak{g}^*)$ is the L_2 dual of E, is such that given λ_1 and $\lambda_2\in E$ we have that $\langle\Phi,\lambda_1\rangle\in\Omega^0(T^*\mathscr{A})$ generates the Hamiltonian vector field $\tilde{\gamma}(\lambda_1)$ and

$$\{\langle\Phi,\lambda_1\rangle,\langle\Phi,\lambda_2\rangle\}_{T^*\mathscr{A}} = \langle\Phi,[\lambda_1,\lambda_2]\rangle\,,$$

where

$$\Omega^0(T^*\mathscr{A}) := \Omega^0(T^*\mathscr{A},\mathbb{R}),$$

$\{\ ,\ \}_{T^*\mathscr{A}}$ is the Poisson bracket on $T^*\mathscr{A}$, $\langle\ ,\ \rangle$ denotes the L_2 pairing between E^* and E, and $[\ ,\ ]$ is the Lie bracket on E. Locally we have that

$$\langle\Phi(A,\pi),\lambda\rangle = \int_\Sigma \Phi_a\lambda^a\,,$$

where

$$\Phi_a(A,\pi) = \pi^i_{a|i} + C^c_{ab}\, A_j^b\,\pi^j_c\,. \tag{2.2}$$

The momentum map Φ is central to the description of the constrained dynamics on $T^*\mathscr{A}$. The physically allowed dynamics is restricted to $\Phi^{-1}(0)$ which (with the restrictions on $\mathscr{A}$ and $\mathscr{G}$ discussed above) is a first class[30] submanifold of $T^*\mathscr{A}$. The true degrees of freedom can thus be formally identified with $\Phi^{-1}(0)/\mathscr{G}$. As before, this reduction formalism has limited use and thus we must construct a generalized dynamics on $T^*\mathscr{A}$ that faithfully describes the true dynamics on $\Phi^{-1}(0)/\mathscr{G}$.

On the extended configuration space $\mathscr{A}$, the observables correspond to the gauge invariant functions. This is because such functions are compatible with the reduction procedure to the true configuration space $\tilde{\mathscr{A}}$. Likewise, the observables on $T^*\mathscr{A}$ are identified with the functions compatible with the projection to $\Phi^{-1}(0)/\mathscr{G}$, i.e., those functions that are gauge invariant when restricted to $\Phi^{-1}(0)/\mathscr{G}$. It is traditional to call such functions weakly gauge invariant.[16] We consider this characterization of observables as the minimal

conditions compatible with the constraint formalism. The physical imput will often impose a richer structure on the observables.

As it stands, the concept of weak invariance is too nebulous to be of any practical use in developing a generalized dynamics on $T^*\mathscr{A}$. We need a more concrete realization of it.

Let us consider a finite-dimensional analogy of the symplectic structure associated with Yang–Mills theory. So let P be a finite-dimensional phase space and F a Lie group with a free symplectic action on P, with momentum map J. Then $J^{-1}(0)/F$ is the analogy of the Yang–Mills true degrees of freedom. We can write $J=\phi_\alpha f^\alpha$, where f^α is a basis of $\mathscr{L}(F)^*$ (the dual to the Lie algebra of F) and $\phi_\alpha\in\Omega_0(P)$ are the constraint functions.

The constraints impose an equivalence class structure on $\Omega^0(P)$ via equality when restricted to $\phi_\alpha=0$, so if $g,h\in\Omega^0(P)$, then $g\sim h$ (g weakly equivalent to h) if $(g-h)_{|\phi_\alpha=0}=0$. In this finite dimension situation we have the following result: if $g\sim h$ then there exists $v^\alpha\in\Omega^0(P)$ such that $g=h+v^\alpha\phi_\alpha$ [sketch proof: F has free action on $P\Rightarrow 0$ is a regular value of J^{-1} (see Ref. 30) $\Rightarrow$ the constraints are a regular sequence (see Ref. 31) in the ring $\Omega^0(P)\Rightarrow$ they generate the ideal of functions vanishing on $J^{-1}(0)$ (see Ref. 32)]. Hence if g is weakly invariant then we can write $\{\phi_\alpha,g\}=v_\alpha^\beta\,\phi_\beta$, for some $v_\alpha^\beta\in\Omega^0(P)$.

This description of weak equivalence in terms of the constraints is central to the Dirac analysis of constrained systems. For the Yang–Mills case it is not known whether one can always replace a weak equality with the constraints, due to the infinite dimensions involved. However, it is always assumed to be the case[16,17] and thus we shall require the observable to be those $H\in\Omega^0(T^*\mathscr{A})$ such that

$$\{\Phi_a(x),H\}_{T^*\mathscr{A}}=V_a^b\,\Phi_b(x)\,,\qquad \text{for some } V_a^b\in\Omega^0(T^*\mathscr{A})\,. \tag{2.3}$$

If H_1 and H_2 are weakly invariant then $\{H_1,H_2\}_{T^*\mathscr{A}}$ is also weakly invariant and thus we have an algebra of observables. This algebra projects down to the Poisson algebra on $\Phi^{-1}(0)/\mathscr{G}$.

It must be kept in mind that we are trying to develop a generalized dynamics on $T^*\mathscr{A}$ and hence we do not actually want to implement the reduction to $\Phi^{-1}(0)/\mathscr{G}$. So the definition of observable is only useful if it can be made compatible with the dynamical description on $T^*\mathscr{A}$. In particular, we need to understand the allowed values an observable can have, that is, we must specify the states of the system. We shall return to a discussion of the states and how they effect the definition of observables in the final part of this paper.

III. OBSERVABLES AND COHOMOLOGY

A. The construction

We shall now develop an alternative description of the observables in Yang–Mills theory. We start by characterizing the gauge invariant functions on $\mathscr{A}$.

Let us first introduce some notation: Given a vector space V and a smooth manifold W, we define $\Gamma^p(V,W)$ as the space of p-linear, continuous, skew mappings from $\underbrace{V\times\cdots\times V}_{p\text{ times}}$ to $\Omega^0(W)$. We write $\Gamma^\cdot(V,W)=\Sigma_p\Gamma^p(V,W)$.

Consider $\Gamma^\cdot(E,\mathscr{A})$, we can define an operator $\delta'\colon\Gamma^p(E,\mathscr{A})\to\Gamma^{p+1}(E,\mathscr{A})$ by

$$\begin{aligned}(\delta'\omega)(\lambda_1,\ldots,\lambda_{p+1})&=\sum_{i=1}^{p+1}(-1)^{i+1}\gamma(\lambda_i)\omega(\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\lambda_{p+1})\\&\quad+\sum_{i<j}(-1)^{i+j}\omega([\lambda_i,\lambda_j]\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\hat{\lambda}_j,\ldots,\lambda_{p+1})\,,\end{aligned} \tag{3.1}$$

where $\omega\in\Gamma^p(E,\mathscr{A})$, $\lambda_1\ldots,\lambda_{p+1}\in E$ and $\hat{\lambda}_i$ indicates omission of λ_i.

It is straightforward to show that $\delta'^2=0$ and thus $(\Gamma^\cdot(E,\mathscr{A}),\delta')$ is a complex. The associated cohomology groups $H^\cdot(E,\mathscr{A})$ can be identified with the Lie algebra cohomology of E taking values on $\Omega^0(\mathscr{A})$.[33]

The gauge invariant functions can now be identified with $H^0(E,\mathscr{A})$ and we see that we have developed a cohomological description of the observables on $\mathscr{A}$. This is a rather trivial use of cohomology since the rest of the complex has no obvious dynamical significance (at least in the classical theory). However, let us push on and try to extend this analysis to the phase space observables.

The naive thing to do would be to replace $\mathscr{A}$ by $T^*\mathscr{A}$ and γ by $\tilde{\gamma}$ in the above construction. Thus we shall end up with $H^\cdot(E,T^*\mathscr{A})$, the cohomology of E taking values in $\Omega^0(T^*\mathscr{A})$. However, $H^0(E,T^*\mathscr{A})$ only characterizes the invariant functions on $T^*\mathscr{A}$ and thus misses the full set of phase space observables.

How should one lift this configuration space cohomology so as to get a useful construction on the phase space? We know that if there is a vector field on a manifold, then it can be lifted to a Hamiltonian vector field on the cotangent bundle. In other words, given a derivation on the ring of functions on a manifold, we can lift it to a derivation on the Poisson algebra of functions on the cotangent bundle. This construction is very simple to implement: Let x^α be local coordinates on our manifold and consider the vector field $X:=X^\alpha\,\partial/x^\alpha$. Then $X^\alpha P_\alpha$ is a function on the cotangent bundle (with P_α the conjugate momentum to x^α) whose Hamiltonian vector field generates the lift of X. We shall now show that it is possible to view δ' as a derivation in this sense and hence lift its action to a suitable phase space in a nontrivial way.

Given $\omega_1\in\Gamma^p(E,\mathscr{A})$ and $\omega_2\in\Gamma^q(E,\mathscr{A})$, we can define $\omega_1\cdot\omega_2\in\Gamma^{p+q}(E,\mathscr{A})$ by using the wedge product on ΛE^* and the ring structure on $\Omega^0(\mathscr{A})$. In particular, if we take the case $p=1$, then $\omega_1\cdot\omega_2\in\Gamma^{q+1}(E,\mathscr{A})$ with

$$(\omega_1\cdot\omega_2)(\lambda_1,\ldots,\lambda_{q+1})=\sum_{i=1}^{q+1}(-1)^{i+1}\omega_1(\lambda_i)\cdot\omega_2(\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\lambda_{q+1})\,. \tag{3.2}$$

Thus $\Gamma^\cdot(E,\mathscr{A})$ can be given the structure of a Grassman algebra over E^* with $\Omega^0(\mathscr{A})$ coefficients. Note that formula (3.2) also makes sense if $\omega_1(\lambda_i)$ is an element of Vect $\mathscr{A}$.

If $\tau\in\Lambda^kE^*\otimes E$, then we can define a derivation $i_\tau\colon\Gamma^p(E,\mathscr{A})\to\Gamma^{p+k-1}(E,\mathscr{A})$ $(p>0)$ as the composition

of the normal inner product between E and E^* plus the product structure on $\Gamma^{\cdot}(E,\mathscr{A})$. In particular, if $\tau\in\Lambda^2E^*\otimes E$, then i_τ: $\Gamma^p(E,\mathscr{A})\to\Gamma^{p+1}(E,\mathscr{A})$ is given by

$$(i_\tau\omega)(\lambda_1,\ldots,\lambda_{p+1})$$

$$=\sum_{i<j}(-1)^{i+j}\omega(\tau(\lambda_i,\lambda_j),\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\hat{\lambda}_j,\ldots,\lambda_{p+1}). \tag{3.3}$$

Formulas (3.2) and (3.3) suggest that we can construct the mapping δ' if we can choose an $\omega_1\in E^*\otimes\mathrm{Vect}\,\mathscr{A}$ and a $\tau\in\Lambda^2E^*\otimes E$ such that

(i) $\omega_1(\lambda)=\gamma(\lambda)$ for all $\lambda\in E$,

(ii) $\tau(\lambda_i,\lambda_j)=[\lambda_i,\lambda_j]$ for all $\lambda_i,\lambda_j\in E$.

Such objects are uniquely constructed by using the Maurer–Cartan form θ on $\mathscr{G}$.

Recall, $\theta\in E^*\otimes E$ is defined as the identity homomorphism from E to E, thus $\theta(\lambda)=\lambda$ for all $\lambda\in E$. Let us define $\bar{\gamma}$: $E^*\otimes E\to E^*\otimes\mathrm{Vect}\,\mathscr{A}$ by $\bar{\gamma}\cdot\beta(\lambda)=\gamma(\beta(\lambda))$, where $\beta\in E^*\otimes E$ and $\lambda\in E$. Thus if we put $\omega_1=\bar{\gamma}\cdot\theta$, then $\omega_1(\lambda)=\bar{\gamma}\cdot\theta(\lambda)=\gamma(\theta(\lambda))=\gamma(\lambda)$ as required.

The Lie algebra structure on E is such that

$$[\ ,\ ]:\Lambda^pE^*\otimes E\times\Lambda^qE^*\otimes E\to\Lambda^{p+q}E^*\otimes E.$$

Therefore $[\theta,\theta]\in\Lambda^2E^*\otimes E$ and we have the result that $\frac{1}{2}[\theta,\theta](\lambda_i,\lambda_j)=[\lambda_i,\lambda_j]$. Thus we take $\tau=\frac{1}{2}[\theta,\theta]$.

This analysis shows that we can consider δ' as a derivation on the Grassman algebra $\Gamma^{\cdot}(E,\mathscr{A})$ where δ' is represented by

$$\delta'=\bar{\gamma}\theta+i_{\frac{1}{2}[\theta,\theta]}. \tag{3.4}$$

It is useful to introduce a basis for E and E^*. Since E represents the space of all mappings from Σ (three-dimensional manifold) to $\mathscr{G}$ (k-dimensional vector space), we write a basis of E as $\rho_a(x)$ $(a=1,\ldots,k,x\in\Sigma)$. From Ref. 26, we know that the Lie algebra structure of E is pointwise that of $\mathfrak{g}$, thus we define Lie brackets by

$$[\rho_a(x),\rho_b(x')]=C^c_{ab}\,\rho_c(x)\delta(x,x'). \tag{3.5}$$

The L_2-dual basis of E^* is written $\eta^a(x)$, where

$$\langle\eta^a(x),\rho_b(x')\rangle=\delta^a_b\delta(x,x'). \tag{3.6}$$

In terms of this basis, we can write δ' in the suggestive form

$$\delta'=\int\left[\eta^a\gamma(\rho_a)-\frac{1}{2}C^c_{ab}\,\eta^a\eta^b\frac{\partial}{\partial\eta^c}\right], \tag{3.7}$$

where $\partial/\partial\eta^c$ is defined symbolically by its action on E and hence $\Lambda\,E^*$ via

$$\frac{\partial}{\partial\eta^c(x)}(\eta^a(x')):=\langle\eta^a(x'),\rho_c(x)\rangle=\delta^a_c\,\delta(x',x). \tag{3.8}$$

Thus, if $\omega=\omega_{ab}\eta^a\cdot\eta^b\in\Gamma^2(E,\mathscr{A})$,

$$\frac{\partial}{\partial\eta^c}\omega=\omega_{ab}\frac{\partial\eta^a}{\partial\eta^c}\cdot\eta^b-\omega_{ab}\eta^a\frac{\partial\eta^b}{\partial\eta^c}. \tag{3.9}$$

Although we shall not use this basis in the construction of the homological description of the Yang–Mills dynamics, it will be found useful for motivating some of the analysis. Also the connection with the BFV formalism is easiest when using a basis. We call η^a the ghost field and, for reasons that will become apparent later, ρ_a the conjugate ghost field.

Before we get too engrossed in the homological algebra, let us again recall the motivation for this description of δ'. We wanted to analyze the action of δ' on the (Grassmann) algebra of functions $\Gamma^{\cdot}(E,\mathscr{A})$ in order to construct an associated (graded) Poisson algebra and hence to lift the δ' action. So what is the Poisson algebra related to $\Gamma^{\cdot}(E,\mathscr{A})$? Normally, if we are given a ring of functions which we could identify with $\Omega^0(N)$, for some manifold N, then the Poisson algebra would be $(\Omega^0(T^*N),\{\ ,\ \})$, where $\{\ ,\ \}$ is the Poisson bracket defined by the canonical symplectic form on T^*N. However, in our case we need to take into account the Grassmann structure to $\Gamma^{\cdot}(E,\mathscr{A})$ and hence, the standard phase space methods do not, at first sight, seem appropriate.

We now present an algebraic construction of the required Poisson algebra. The symplectic underpinning to this will be discussed later in this paper.

The algebra $\Gamma^{\cdot}(E,\mathscr{A})$ is modeled on the exterior algebra of E with $\Omega^0(\mathscr{A})$ coefficients. Thus, using the L_2-duality philosophy we expect the associated Poisson algebra to be based on $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$, the Grassmann algebra modeled on the exterior algebra over T^*E with $\Omega^0(T^*\mathscr{A})$ coefficients. As usual, we can decompose this into homogeneous parts as $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})=\Sigma_r\Gamma^r(T^*E,T^*\mathscr{A})$ with multiplication defined by the wedge product. However, under the identification $T^*E\simeq E\times E^*$, we can induce a finer decomposition of $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$ via

$$\Gamma^r(T^*E,T^*\mathscr{A})\simeq\Gamma^r(E\times E^*,T^*\mathscr{A})$$

$$=\bigoplus_{p+q=r}\Gamma^{p,q}(E,E^*;T^*\mathscr{A}),$$

where $\Gamma^{p,q}(E,E^*;T^*\mathscr{A})$ is the space of $(p+q)$-linear, continuous, skew mappings from $\underbrace{E\times\cdots\times E}_{p\text{ times}}\times\underbrace{E^*\times\cdots\times E^*}_{q\text{ times}}$ to $\Omega^0(T^*\mathscr{A})$. We shall often write $\Gamma^{p,q}$ for $\Gamma^{p,q}(E,E^*;T^*\mathscr{A})$.

Given $\omega_1\in\Gamma^{p_1,q_1}$ and $\omega_2\in\Gamma^{p_2,q_2}$ we define the product $\omega_1\cdot\omega_2\in\Gamma^{p_1+p_2,q_1+q_2}$ as that derived from the product structure on $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$. Thus, given $\lambda_i\in E$ and $\mu^i\in E^*$ we have

$$(\omega_1\cdot\omega_2)(\lambda_1,\ldots,\lambda_{p_1},\mu_1^{p_1+1},\ldots,\mu^{p_1+q_1},\lambda_{p_1+q_1+1},\ldots,\lambda_{p_1+p_2+q_1},\mu^{p_1+p_2+q_1+1},\ldots,\mu^{p_1+p_2+q_1+q_2})$$

$$=\sum_{\sigma\in\mathscr{S}_{p_1+p_2+q_1+q_2}}(\operatorname{sgn}\sigma)\omega_1(\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(p_1)},\mu^{\sigma(p_1+1)},\ldots,\mu^{\sigma(p_1+q_1)})$$

$$\cdot\omega_2(\lambda_{\sigma(p_1+q_1+1)},\ldots,\lambda_{\sigma(p_1+p_2+q_1)},\mu^{\sigma(p_1+p_2+q_1+1)},\ldots,\mu^{\sigma(p_1+p_2+q_1+q_2)}), \tag{3.10}$$

where $\mathscr{S}_{p_1+p_2+q_1+q_2}$ is the subset of the permutation group $\mathscr{G}_{p_1+p_2+q_1+q_2}$ consisting of those permutations which leave invariant the sets

$$\{1,\ldots,p_1,p_1+q_1+1,\ldots,p_1+p_2+q_1\}$$

and

$\{p_1+1,\ldots,p_1+q_1,p_1+p_2+q_1+1,\ldots,p_1+p_2+q_1+q_2\}$, and are such that

$$\sigma(1)<\cdots<\sigma(p_1),$$
$$\sigma(p_1+q_1+1)<\cdots<\sigma(p_1+q_1+p_2),$$
$$\sigma(p_1+1)<\cdots<\sigma(p_1+q_1),$$
$$\sigma(p_1+p_2+q_1+1)<\cdots<\sigma(p_1+p_2+q_1+q_2).$$

From this definition one can easily show that

$$\omega_1\cdot\omega_2=(-1)^{(p_1+q_1)(p_2+q_2)}\omega_2\cdot\omega_1, \tag{3.11}$$

as expected from the product structure on $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$.

If $\omega_1\in\Gamma^{1,0}$ and $\omega_2\in\Gamma^{0,1}$, then we define $\omega_1\lrcorner\omega_2\in\Gamma^{0,0}$ by

$$\omega_1\lrcorner\omega_2:=\omega_1(\omega_2)=\langle\omega_1(\hat{\ }),\omega_2(\hat{\ })\rangle. \tag{3.12}$$

Then $\omega_1\lrcorner\omega_2=\omega_2\lrcorner\omega_1$. We now extend this construction: Given $\omega_1\in\Gamma^{p_1,q_1}$ and $\omega_2\in\Gamma^{p_2,q_2}$ (with p_1,p_2,q_1,q_2 not equal to zero) we define $\omega_1\lrcorner\omega_2\in\Gamma^{p_1+p_2-1,q_1+q_2-1}$ via

$$\begin{aligned}(\omega_1\lrcorner\omega_2)&(\lambda_1,\ldots,\lambda_{p_1},\mu^{p_1+1},\ldots,\mu^{p_1+q_1},\lambda_{p_1+q_1+1},\ldots,\lambda_{p_1+p_2+q_1-1},\mu^{p_1+p_2+q_1},\ldots,\mu^{p_1+p_2+q_1+q_2-2})\\
&=\sum_{\sigma\in\mathscr{S}_{(p_1-1)+p_2+q_1+(q_2-1)}}(\operatorname{sgn}\sigma)\langle\omega_1(\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(p_1-1)},\mu^{\sigma(p_1+1)},\ldots,\mu^{\sigma(p_1+q_1)},\hat{\ }),\\
&\quad\omega_2(\hat{\ },\lambda_{\sigma(p_1)},\lambda_{\sigma(p_1+q_1+1)},\ldots,\lambda_{\sigma(p_1+p_2+q_1-1)},\mu^{\sigma(p_1+p_2+q_1)},\ldots,\mu^{\sigma(p_1+p_2+q_1+q_2-2)})\rangle\\
&\quad+\sum_{\sigma\in\mathscr{S}_{p_1+(p_2-1)+(q_1-1)+q_2}}(\operatorname{sgn}\sigma)\langle\omega_1(\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(p_1)},\mu^{\sigma(p_1+1)},\mu^{\sigma(p_1+q_1-1)},\hat{\ }),\\
&\quad\omega_2(\hat{\ },\lambda_{\sigma(p_1+q_1+1)},\ldots,\lambda_{\sigma(p_1+p_2+q_1-1)},\mu^{\sigma(p_1+p_2)},\mu^{\sigma(p_1+p_2+q_1)},\ldots,\mu^{\sigma(p_1+p_2+q_1+q_2-2)})\rangle.\end{aligned} \tag{3.13}$$

If p_1 or q_2 is zero, then there is only the second term and similarly if p_2 or q_1 is zero then only the first term survives.

One can derive the following properties of $\lrcorner$: Consider $\omega_i\in\Gamma^{p_i,q_i}$ and define $r_i=p_i+q_i$, then

(i) $\omega_1\lrcorner\omega_2=-(-1)^{r_1r_2}\omega_2\lrcorner\omega_1$,

(ii) $(\omega_1+\omega_2)\lrcorner\omega_3=\omega_1\lrcorner\omega_3+\omega_2\lrcorner\omega_3$,

(iii) $(\omega_1\cdot\omega_2)\lrcorner\omega_3=\omega_1(\omega_2\lrcorner\omega_3)+(-1)^{r_2r_3}(\omega_1\lrcorner\omega_3)\cdot\omega_2$,

(iv) $(-1)^{r_1r_3}\omega_1\lrcorner(\omega_2\lrcorner\omega_3)+(-1)^{r_2r_1}\omega_2\lrcorner(\omega_3\lrcorner\omega_1)+(-1)^{r_2r_3}\omega_3\lrcorner(\omega_1\lrcorner\omega_2)=0$.

This last relation is called the super Jacobi identity. The proof of these results are not very illuminating and will be omitted. In terms of the basis $\eta^a(x),\rho_a(x)$ introduced above, we see that the fundamental relation is $\eta^a(x)\lrcorner\rho_b(x')=\rho_b(x')\lrcorner\eta^a(x)=\delta^a_b\delta(x,x')$, and all other expressions can be built up from repeated use of the properties (i) to (iv) stated above.

It is clear that $\lrcorner$ imposes a graded-Poisson algebra on $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$. However, it is only responding to the duality properties of E and E^*, and is trivial on the $\Omega^0(T^*\mathscr{A})$ coefficients. But on $\Omega^0(T^*\mathscr{A})$ we already have a Poisson bracket $\{\ ,\ \}_{T^*\mathscr{A}}$ derived from the symplectic structure on $T^*\mathscr{A}$. This bracket can easily be extended to $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$. Thus we define

$$\{\omega_1,\omega_2\}_{T^*\mathscr{A}}\in\Gamma^{p_1+p_2,q_1+q_2},$$

with

$$\{\omega_1,\omega_2\}_{T^*\mathscr{A}}=-(-1)^{r_1r_2}\{\omega_2,\omega_1\}_{T^*\mathscr{A}},$$

via

$$\begin{aligned}\{\omega_1,\omega_2\}_{T^*\mathscr{A}}&(\lambda_1,\ldots,\lambda_{p_1},\mu^{p_1+1},\ldots,\mu^{p_1+q_1},\lambda_{p_1+1+q_1},\ldots,\lambda_{p_1+p_2+q_1},\mu^{p_1+p_2+q_1+1},\ldots,\mu^{p_1+p_2+q_1+q_2})\\
&=\sum_{\sigma\in\mathscr{S}_{p_1+p_2+q_1+q_2}}(\operatorname{sgn}\sigma)\{\omega_1(\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(p_1)},\mu^{\sigma(p_1+1)},\ldots,\mu^{\sigma(p_1+q_1)}),\\
&\quad\omega_2(\lambda_{\sigma(p_1+1+q_1)},\ldots,\lambda_{\sigma(p_1+p_2+q_1)},\mu^{\sigma(p_1+p_2+q_1+1)},\ldots,\mu^{\sigma(p_1+p_2+q_1+q_2)})\}_{T^*\mathscr{A}}.\end{aligned} \tag{3.14}$$

Combining $\{\ \}_{T^*\mathscr{A}}$ and $\lrcorner$, we define a graded Poisson bracket $\{\ ,\ \}$ on $\Gamma^{\cdot}(T^*E,T^*\mathscr{A})$ via

$$\{\omega_1,\omega_2\}=\{\omega_1,\omega_2\}_{T^*\mathscr{A}}+\omega_1\lrcorner\omega_2. \tag{3.15}$$

We now consider some examples of this bracket action.

The Maurer–Cartan form θ is an element of $\Gamma^{1,1}$ and thus $[\theta,\theta]\in\Gamma^{2,1}$. With our conventions we find that

$$\tfrac{1}{2}[\theta,\theta](\lambda_1,\lambda_2,\mu)=-\langle\mu,[\lambda_1,\lambda_2]\rangle. \tag{3.16}$$

Let us investigate $\frac{1}{2}[\theta,\theta]\lrcorner\omega$, where $\omega\in\Gamma^{1,0}$,

$$\begin{aligned}(\tfrac{1}{2}[\theta,\theta]\lrcorner\omega)(\lambda_1,\lambda_2)&=\langle\omega,\tfrac{1}{2}[\theta,\theta](\hat{\ },\lambda_1,\lambda_2)\rangle\\
&=\tfrac{1}{2}[\theta,\theta](\lambda_1,\lambda_2,\omega)\\
&=-\langle\omega,[\lambda_1,\lambda_2]\rangle\ \text{by (3.16)}\\
&=-\omega([\lambda_1,\lambda_2])=-\omega(ad_{\lambda_1}\lambda_2).\end{aligned}$$

Similarly, if $\omega\in\Gamma^{p,0}$

$$(\tfrac{1}{2}[\theta,\theta]\lrcorner\omega)(\lambda_1,\ldots,\lambda_{p+1})=\sum_{i<j}(-1)^{i+j}\omega([\lambda_i,\lambda_j],\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\hat{\lambda}_j,\ldots,\lambda_{p+1}). \tag{3.17}$$

Acting on $\omega\in\Gamma^{0,1}$ we have

$$\begin{aligned}(\tfrac{1}{2}[\theta,\theta]\lrcorner\omega)(\lambda,\mu) &= \langle \tfrac{1}{2}[\theta,\theta]\omega(\lambda,\mu,\hat{\ }),\omega\rangle \\ &= -\tfrac{1}{2}[\theta,\theta](\lambda,\omega,\mu) \\ &= \langle\mu,[\lambda,\omega]\rangle \\ &= \omega(ad^*_\lambda\mu)\,. \end{aligned} \tag{3.18}$$

We note that $\{\frac{1}{2}[\theta,\theta],\}_{T^*\mathscr{A}}=0$. So, acting on $\omega\in\Gamma^{p,q}$, $\{\frac{1}{2}[\theta,\theta],\omega\}=\frac{1}{2}[\theta,\theta]\lrcorner\omega\in\Gamma^{p+1,q}$.

From the definition of the momentum mapping Φ, we see that it is an element of $\Gamma^{1,0}$. So given $\omega\in\Gamma^{p,q}$ we have

$$\{\Phi,\omega\}=\{\Phi,\omega\}_{T^*\mathscr{A}}+\Phi\lrcorner\omega\in\Gamma^{p+1,q}+\Gamma^{p,q-1}\,,$$

where

$$\begin{aligned}&\{\Phi,\omega\}_{T^*\mathscr{A}}(\lambda_1,\ldots,\lambda_{p+1},\mu^1,\ldots\mu^q) \\ &\quad=\sum_{i=1}^{p+1}(-1)^{i+1} \\ &\qquad\times\{\Phi(\lambda_i),\omega(\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\lambda_{p+1},\mu^1,\ldots,\mu^q)\}_{T^*\mathscr{A}} \\ &\quad=\sum_{i=1}^{p+1}(-1)^{i+1}\tilde{\gamma}(\lambda_i) \\ &\qquad\cdot\omega(\lambda_1,\ldots,\hat{\lambda}_i,\ldots,\lambda_{p+1},\mu^1,\ldots,\mu^q)\,, \end{aligned} \tag{3.19}$$

and

$$\begin{aligned}&(\Phi\lrcorner\omega)(\lambda_1,\ldots,\lambda_p,\mu^1,\ldots,\mu^{q-1}) \\ &\quad=\langle\Phi,\omega(\hat{\ },\lambda_1,\ldots,\lambda_p,\mu^1,\ldots,\mu^{q-1})\rangle \\ &\quad=\omega(\Phi,\lambda_1,\ldots,\lambda_p,\mu^1,\ldots,\mu^{q-1})\,. \end{aligned} \tag{3.20}$$

Thus we can define $\delta\omega\in\Gamma^{p+1,q}+\Gamma^{p,q-1}$ via

$$\delta\omega:=\{\Phi+\tfrac{1}{2}[\theta,\theta],\omega\}\,. \tag{3.21}$$

From this definition we see that δ can be decomposed into $\delta_0+\delta_1$, where δ_0: $\Gamma^{p,q}\to\Gamma^{p,q-1}$ is given by (3.20) and δ_1: $\Gamma^{p,q}\to\Gamma^{p+1,q}$ is given by $\{\Phi,\omega\}_{T^*\mathscr{A}}+\frac{1}{2}[\theta,\theta]\lrcorner\omega$.

We claim that δ is the desired lift of δ' to $\Gamma(T^*E,T^*\mathscr{A})$. Indeed, δ induces (odd) symplectic transformations on $\Gamma(T^*E,T^*\mathscr{A})$ with generator $\Phi+\frac{1}{2}[\theta,\theta]$. Written in terms of the basis $\eta^a(x)$ and $\rho_b(x')$ we see that this generator can be written as

$$\Phi+\tfrac{1}{2}[\theta,\theta]=\int_\Sigma[\Phi_a\eta^a-\tfrac{1}{2}C^c_{ab}\,\eta^a\eta^b\rho_c]\,. \tag{3.22}$$

This is as expected from (3.7) if we interpret ρ_c as the conjugate variable to η^c. Acting on $\Gamma^{p,0}$, we see from (3.17) and (3.19) that δ incorporates the natural extension of δ', (3.1), with $\tilde{\gamma}$ replacing γ. However, to be acceptable we must be able to develop a cohomology theory from δ, and hence it must be nilpotent. From the super Jacobi identity we see that it is sufficient to check the following:

$$\{\Phi+\tfrac{1}{2}[\theta,\theta],\Phi+\tfrac{1}{2}[\theta,\theta]\}=0\,. \tag{3.23}$$

Now

$$\begin{aligned}&\{\Phi,+\tfrac{1}{2}[\theta,\theta],\Phi+\tfrac{1}{2}[\theta,\theta]\} \\ &\quad=\{\Phi,\Phi\}_{T^*\mathscr{A}}\in\Gamma^{2,0} \\ &\qquad+\Phi\lrcorner[\theta,\theta]\in\Gamma^{2,0} \\ &\qquad+\tfrac{1}{2}[\theta,\theta]\lrcorner\tfrac{1}{2}[\theta,\theta]\in\Gamma^{3,1}\,, \end{aligned}$$

where we note from (3.20) that $\Phi\lrcorner\Phi=0$, i.e., $\delta_0^2=0$;

$$\begin{aligned}&(\tfrac{1}{2}[\theta,\theta]\lrcorner\tfrac{1}{2}[\theta,\theta])(\lambda_1,\lambda_2,\lambda_3,\mu) \\ &\quad=\langle\mu,[\lambda_1,[\lambda_2,\lambda_3]]+[\lambda_2,[\lambda_3,\lambda_1]]+[\lambda_3,[\lambda_1,\lambda_2]]\rangle \\ &\quad=0\quad\text{by the Jacobi identity on } E; \end{aligned}$$

$$\begin{aligned}\{\Phi,\Phi\}_{T^*\mathscr{A}}(\lambda_1,\lambda_2)&=2\{\Phi(\lambda_1),\Phi(\lambda_2)\} \\ &=2\Phi([\lambda_1,\lambda_2])\,, \end{aligned} \tag{3.24}$$

by equivariance of Φ;

$$\begin{aligned}(\Phi\lrcorner[\theta,\theta])(\lambda_1,\lambda_2)&=\langle\Phi,[\theta,\theta](\hat{\ },\lambda_1,\lambda_2)\rangle \\ &=[\theta,\theta](\lambda_1,\lambda_2,\Phi) \\ &=-2\Phi([\lambda_1,\lambda_2])\quad\text{from (3.16).} \end{aligned}$$

Combining these results we find that $\delta^2=0$. In terms of δ_0 and δ_1 this is

$$\delta_0^2=0\,, \tag{3.25a}$$

$$\delta_0\delta_1+\delta_1\delta_0=0\,, \tag{3.25b}$$

$$\delta_1^2=0\,. \tag{3.25c}$$

In order to discuss the cohomology associated with δ we need to construct the complex upon which it acts. Define the p-ghost number cochain Γ^p as

$$\Gamma^p=\bigoplus_{r-s=p}\Gamma^{r,s}\,, \tag{3.26}$$

then δ: $\Gamma^p\to\Gamma^{p+1}$ and hence we can define cohomology groups in the normal fashion. In the next section we shall analyze these objects and relate them to a dynamical description of Yang–Mills theory. We conclude this section with a brief discussion about the geometry of this construction.

The arguments leading up to formula (3.21) comprise a liberal use of the L_2-duality philosophy in conjunction with various ideas from a graded version of symplectic geometry. Thus one would like to know whether it is actually possible to give this construction a firm geometric underpinning by using the graded (or super) manifold techniques from the start. There are several reasons for doing this: First, we are ultimately going to want to quantize this classical system and in this context we expect the homological arguments to take a secondary role to the (super) phase space ones. Second, we will want to apply this type of analysis of constrained systems whose constraints are not related to the action of any group (e.g., gravity, supergravity, strings,...). In these cases the geometric input to the homological construction may be lost, yet we can still hope for a (super) manifold analysis.

There are various ways of extending the standard manifold theory to include graded elements, for a good review see Ref. 34. By far the most attractive approach is that of Rogers,[35] since it is closest in spirit to the familiar Banach manifold techniques and yet allows for the possibility of nontrivial structure in the odd directions. However, direct application of her construction to the case at hand is problematic. There are two reasons for this: (i) Our manifolds are infinite dimensional and thus one needs to extend the model space $B^{n,m}$ to some suitable graded version of the model space for $T^*\mathscr{A}$ and E. It has proven to be difficult to achieve this and retain a G^∞ structure as one would like. (ii) Even if one considers finite-dimensional situations, as we did

above, then one must ensure that the supermanifold version of the homology theory (as constructed by an odd vector field) is equivalent to the standard Lie algebra cohomology theory. There is a standard way of comparing such structures (introduce G^∞ chain maps, chain homotopies, etc.) but again, one finds that the full G^∞ structure can obstruct such constructions.

The conclusion of this is that one must use a much more restrictive class of supermanifolds, namely the Batchelor[36] supermanifolds. This might almost be expected since we know[34] that any such Batchelor supermanifold can be related to a Kostant[37] graded manifold, which in turn can be related to the sheaf of sections of an exterior vector bundle over a normal manifold. Their analysis can be extended to the situation found in this paper and we find that $\Gamma^\cdot(E,\mathscr{A})$ can be given a graded manifold structure. The L_2-cotangent bundle is then precisely the graded manifold based on $\Gamma^\cdot(T^*E,T^*\mathscr{A})$ with the Poisson bracket given by (3.15).[37]

Thus, our construction can be given a graded manifold structure. However, it is of a trivial nature and thus does not warrant any separate development, at least for the classical Yang–Mills situation.

B. Weak cohomology and the BFV formalism

In the previous subsection we developed a method for extending the Lie algebra cohomology theory into a phase space situation. We now show how this construction can be used in a generalized dynamics and thus in what sense it can be viewed as a weak cohomology theory.

Cohomology groups are notoriously difficult to calculate, especially when infinite-dimensional spaces are being used. Thus all we can hope to do in this section is point out various aspects of this theory and motivate some of the assumptions needed. We start by analyzing the δ_0 map introduced via Eq. (3.20).

From its definition, δ_0: $\Gamma^{p,q}\to\Gamma^{p,q-1}$ and, without loss of generality, let us put $p=0$. Now $\Gamma^{0,q}\approx\Omega^0(T^*\mathscr{A})\otimes\Lambda^q E$, and thus we have a complex

$$\to\Omega^0(T^*\mathscr{A})\otimes\Lambda^q E\to^{\delta_0}\Omega^0(T^*\mathscr{A})\otimes\Lambda^{q-1}E\to^{\delta_0}\cdots$$
$$\to^{\delta_0}\Omega^0(T^*\mathscr{A})\otimes E\to^{\delta_0}\Omega^0(T^*\mathscr{A})\,. \tag{3.27}$$

We recognize this as the Koszul complex[31] associated with the momentum map Φ: $T^*\mathscr{A}\to E^*$. Let us investigate the final term in this complex:

$$\Omega^0(T^*\mathscr{A})\otimes E\to^{\delta_0}\Omega^0(T^*\mathscr{A})\,,\qquad \omega\mapsto\langle\Phi,\omega\rangle\,. \tag{3.28}$$

So if

$$\omega=\int_\Sigma V^a\rho_a,\ \text{then}\ \delta_0\omega=\int_\Sigma V^a\Phi_a\,,\quad V^a\in\Omega^0(T^*\mathscr{A})\,.$$

Thus we see that the image of δ_0 is precisely those functions which we have assumed as weakly equivalent to the zero function.

Let us return to the finite-dimensional model introduced in the first section. Using the momentum map J: $P\to\mathscr{L}(F)^*$ one can construct a Koszul complex as in (3.7) and there we have the following result: If we know that $h\sim 0$ implies that $h=V^a\phi_a$, for some $V^a\in\Omega^0(P)$, then the Koszul complex is a resolution of $\Omega^0(p)/\sim$. That is, all the homology groups are zero (see Ref. 38). So if F acts freely on P then we have the result that the Koszul complex is a resolution.

How this result can be extended to the infinite-dimensional situation is not clear. One would like to be able to say that our original assumption that in Yang–Mills theory weak equivalence can be described in terms of the momentum map, ensures the resolution property of (3.7). However, until such a result can be shown we shall have to take it as an additional assumption on the class of observables we are allowing. Let us now investigate the consequences of this assumption.

Acting on $F\in\Gamma^{0,0}$, $\delta F=\delta_1 F\in\Phi^{1,0}$ is given by (3.19), i.e., $(\delta_1 F)(\lambda)=\{\Phi(\lambda),F\}_{T^*\mathscr{A}}$. Thus we can characterize the weakly invariant functions, (2.3), as those $F\in\Gamma^{0,0}$ such that there exists an $\omega_1\in\Gamma^{1,1}$ with

$$\delta_1 F+\delta_0\omega_1=0\,. \tag{3.29}$$

Suppose now that $\delta_1\omega_1=0$. Then $\mathscr{F}:=F+\omega_1\in\Gamma^0$ [i.e., has zero ghost number, cf. (3.26)] and $\delta\mathscr{F}=0$. So in this situation we have been able to relate a δ-closed zero ghost number function with the weakly invariant function F. What happens when $\delta_1\omega_1\neq 0$? From (3.25) we know that $\delta_0\delta_1\omega_1=-\delta_1\delta_0\omega_1=\delta_1^2 F=0$ and hence from our assumption above, there exists $\omega_2\in\Gamma^{2,2}$ such that $\delta_1\omega_1+\delta_0\omega_2=0$. If $\delta_1\omega_2=0$, then we can relate F with $\mathscr{F}'=F+\omega_1+\omega_2\in\Gamma^0$ such that $\delta\mathscr{F}'=0$. If $\delta_1\omega_2\neq 0$, then repeating the argument will produce an $\omega_3\in\Gamma^{3,3}$, etc.

Thus the ability to accommodate weakly invariant functions into the homology structure of δ requires that (3.27) be a resolution of $\Omega^0(T^*\mathscr{A})/\sim$. When this is the case we shall describe δ as producing a weak cohomology for the Lie algebra E acting on $T^*\mathscr{A}$.

It is important not to get too carried away with the general mathematical formalism at the expense of physical necessity. Indeed for the observables known in Yang–Mills theory (essentially just the Hamiltonian) one has that $\delta_1\omega_1=0$, and the Koszul complex (3.27) is not used. This may not be the situation for all useful observables; however, it does point out that we will always use a much richer structure for the observables than simply being elements of $\Omega^0(T^*\mathscr{A})$.

Suppose $F_1\in\Omega^0(T^*\mathscr{A})$ is weakly invariant and $\mathscr{F}_1$ the corresponding δ-closed element of Γ^0. If $F_2\in\Omega^0(T^*\mathscr{A})$ is such that $F_1\sim F_2$, then F_2 is weakly invariant and one can construct an $\mathscr{F}_2\in\Gamma^0$ as above. Now it is easy to show that there exists $\chi\in\Gamma^{-1}$ such that $\mathscr{F}_1=\mathscr{F}_2+\delta\chi$. Hence, if we require that $\mathscr{F}_1$ and $\mathscr{F}_2$ describe the same observables, then we must identify the observables with

$$H^0(\delta):=\frac{\ker\delta\colon\ \Gamma^0\to\Gamma^1}{\text{image}\,\delta\colon\ \Gamma^{-1}\to\Gamma^0}\,.$$

By using the resolution property of (3.27) one can see that each nontrivial element of $H^0(\delta)$ is constructed from a weakly gauge invariant function on $T^*\mathscr{A}$.

This relationship between weak invariance and δ-closed zero ghost number functions is the basis of the BFV formalism. In their approach to the construction of the S matrix for Yang–Mills theory they first had to build an odd, Abelian function Q out of the constraints and fermionic degrees of freedom. The weakly invariant Hamiltonian is then used to

construct a zero ghost number function $\mathscr{H}$, which is invariant under the (nilpotent) transformations generated by Q. They then consider the S matrix constructed from the effective Hamiltonian H_χ defined by

$$H_\chi := \mathscr{H} + \{Q,\chi\}, \tag{3.30}$$

where χ is an arbitrary odd function with ghost number -1. Their main theorem is that the S matrix is independent of χ (see Ref. 39 for a more detailed presentation).

The analysis of this paper is very close to the BFV formalism with $Q = \Phi + \frac{1}{2}[\theta,\theta]$. However, there are two important points where this presentation does not cover the BFV work.

(i) BFV wanted to relate the phase space construction with the effective Lagrangian for a constrained system. In order to build the Lagrangian one must first extend the phase space by the inclusion of the primary constraints[16,17] and their associated ghosts and conjugate ghosts. In this way one can relate the new Q with the phase space version of the BRST charge. In a separate paper we shall discuss the extension procedure in detail.

(ii) BFV justified their method by constructing a path integral expression which agreed with the accepted S matrix for Yang–Mills theory. However, we are not discussing the quantization of Yang–Mills theory in this paper. Rather, we have presented this work as an example of a generalized dynamics. Thus, in order to justify the characterization of observables presented here we must show that it can be related to the standard Dirac analysis. This will be discussed in the next section.

IV. CLASSICAL STATES AND OBSERVABLES

A full description of a classical dynamical system is comprised of an algebra of physical observables, including the Hamiltonian, plus a rule for determining the allowed values the observables can take. This involves knowledge of the states of the system. There are two ways of viewing the states; on the one hand we think of them geometrically as the points of the phase space, but more generally, they can be viewed as normalized, positive elements of a suitable dual space to the algebra of observables.[40]

The geometric characterization of states has limited value, when we have constraints, because it assumes the reduction to the true degrees of freedom has been implemented. So we shall take the view that states assign expectation values to observables. In order to understand how this is to be achieved in constrained systems we first discuss the standard phase space approach.[40]

On an unconstrained phase space P we identify the observables with $\Omega^0(P)$. Thus a state ψ is an element of $\Omega^0(P)'$, the dual space to $\Omega^0(P)$, such that (i) $\psi \geqslant 0$ and (ii) $(\psi,1) = 1$, where $(\ ,\ )$ implements the duality between $\Omega^0(P)'$ and $\Omega^0(P)$. Given a state ψ, the expectation value $\langle F\rangle$ of an observable $F \in \Omega^0(P)$ is given by

$$\langle F\rangle = (\psi,F) = \int_P \psi F\, d\mu\,. \tag{4.1}$$

The measured $d\mu$ on P is the Liouville measure. In the field theory situation where P is infinite dimensional, and possibly not a linear space, great care must be taken in defining the dual space and the measure involved[41] (see also Ref. 25). We shall not dwell on these important points here, but rather we shall investigate the formal extension of these ideas to constrained systems.

For Yang–Mills theory a state will give the allowed values of physical observables. We do not attach any significance to the pairing between a state and an unphysical observable. The constraints impose an equivalence class structure on the observable where we say two weakly invariant functions F_1 and F_2 are equivalent if they are weakly equal to each other, i.e., $F_1 \sim F_2$, if there exists $V^a \in \Omega^0(T^*\mathscr{A})$ such that

$$F_1 = F_2 + \int_\Sigma V^a \Phi_a\,. \tag{4.2}$$

For this equivalence relation to have any physical significance, we must require that $\langle F_1\rangle = \langle F_2\rangle$. In other words, if $F \sim 0$, then $\langle F\rangle = 0$. Thus we require that our states Ψ are normalized, positive elements of the dual to $\Omega^0(T^*\mathscr{A})$ such that

$$\left\langle \int_\Sigma \Phi_a V^a \right\rangle = 0\,, \tag{4.3}$$

for all $V^a \in \Omega^0(T^*\mathscr{A})$. We shall write this condition on states as

$$\langle \Phi_a\rangle = 0\,. \tag{4.4}$$

This definition of states can be thought of as the minimal requirement that we can impose. If we want to relate states on $T^*\mathscr{A}$ with states on the true degrees of freedom, then we must impose stronger conditions on ψ. In particular, we would require it to be gauge invariant or at least show that it can be decomposed into an invariant part. However, as a rule for assigning values to physical observables, formula (3.3) suffices. It is straightforward to show that this definition of states is compatible with the dynamical evolution of the system.

We now investigate the introduction of classical states into the BFV formalism. The observables are now elements of $H^0(\delta)$ so in order to introduce states we must first construct a dual space to Γ^0, and then pick out the positive, normalized elements compatible with the equivalence class structure found in $H^0(\delta)$.

Define $\Gamma'^{q,p}$ as the space of $(q+p)$-linear, continuous, skew mappings from $\underbrace{E \times \cdots \times E}_{q \text{ times}} \times \underbrace{E^* \times \cdots \times E^*}_{p \text{ times}}$ to $\Omega^0(T^*\mathscr{A})'$, the dual space to $\Omega^0(T^*\mathscr{A})$. Then $\Gamma'^{q,p}$ can be considered the dual to $\Gamma^{p,q}$; we shall denote this duality by the bracket $(\ ,\ )$. The dual to Γ^p is then Γ'^{-p}, where

$$\Gamma'^{-p} := \bigoplus_{r-s=-p} \Gamma'^{r,s}\,. \tag{4.5}$$

Given $\omega \in \Gamma^p$, then $\delta\omega \in \Gamma^{p+1}$ and we can define the adjoint δ^+: $\Gamma'^{-p-1} \to \Gamma'^{-p}$ by

$$(\zeta,\delta\omega) = (\delta^+\zeta,\omega)\,, \tag{4.6}$$

for all $\zeta \in \Gamma'^{-p-1}$. Since $\delta = \delta_0 + \delta_1$, we can decompose δ^+ into $\delta^+ = \delta_0^+ + \delta_{-1}$, where δ_0^+ and δ_{-1} are the adjoints to δ_0 and δ_1, respectively. The adjoint δ^+ also satisfies $(\delta^+)^2 = 0$ and hence δ_0^+ and δ_{-1} obey the adjoint version of the relations (3.25). We denote the cohomology groups associated with the complex (Γ'^p,δ^+) by $H^{\cdot}(\delta^+)$.

The states must be elements of Γ'^0, and we write $\psi\in\Gamma'^0$ as

$$\psi=\psi_0+\psi_1+\psi_2+\cdots, \tag{4.7}$$

where $\psi_i\in\Gamma'^{i,i}$. The equivalence class structure on $H^0(\delta)$ requires that the states must be such that

$$\langle\delta\chi\rangle=0, \tag{4.8}$$

for all $\chi\in\Gamma^{-1}$. In terms of the adjoint δ^+, we can write (4.8) as

$$\langle\delta\chi\rangle=(\delta\chi,\psi)=(\chi,\delta^+\psi)=0. \tag{4.9}$$

Since (4.9) must hold for all χ, we have that the states ψ must satisfy

$$\delta^+\psi=0. \tag{4.10}$$

If we put $\psi=\delta^+\zeta$, for some $\zeta\in\Gamma'^{-1}$, then (4.10) holds since $(\delta^+)^2=0$. However, if we use such a state then the observables $\mathscr{F}\in H^0(\delta)$ will all have zero expectation value since

$$\langle\mathscr{F}\rangle=(\mathscr{F},\delta^+\zeta)=(\delta\mathscr{F},\zeta)=0. \tag{4.11}$$

We shall consider such states as trivial ones, and thus we are led to the conclusion that the states are elements of $H^0(\delta^+)$. Not all elements of $H^0(\delta^+)$ will correspond to states though, since we must require that we can recover the standard description presented above.

It is not clear how much of the full structure of $H^0(\delta)$, for observables, and $H^0(\delta^+)$, for states, can be accommodated in a generalized dynamical description of Yang–Mills theory. What we shall do is look at a restricted class of observables and construct the states for these.

A δ-closed element $\mathscr{F}\in\Gamma^0$ will in general have terms containing ghosts and conjugate ghosts. Since these fermionic components have nothing to do with the true dynamics, we might expect that a representative f can be chosen from the equivalence class $[\mathscr{F}]\in H^0(\delta)$ such that $f\in\Gamma^{0,0}$, i.e., f has no ghosts and conjugate ghosts. If $f\in\Gamma^{0,0}$ and $\delta f=0$, then f is gauge invariant. So we are asking whether it is possible to construct a $\chi\in\Gamma^{-1}$ such that $\mathscr{F}=f+\delta\chi$. Equivalently, in terms of the Dirac analysis, we would like to know that if F is weakly gauge invariant, then F is weakly equivalent to a gauge invariant function. A local version of this theorem has become part of the folklore of constrained dynamics (at least in the situation where the constraints are associated with a group action on a phase space). However, we can see that in general there might be global obstructions. These would be measured by $H^r(E,\Omega^0(T^*\mathscr{A})\otimes\Lambda^r E)$ $(r>0)$, and it is difficult to determine when this is zero. All one can say is that for the known observables in Yang–Mills theory one does not hit any obstruction. We now take this as a property of the observables in our theory.

Given this decomposition of the δ-closed functions it now becomes simple to describe the states. Let $\psi\in\Gamma'^0$ be such that

(i) $\delta^+\psi=0$,

(ii) $\psi=\psi_0+\delta^+\zeta$, for some $\zeta\in\Gamma'^{-1}$.

If $\mathscr{F}\in\Gamma^0$ is such that

(i) $\delta\mathscr{F}=0$,

(ii) $\mathscr{F}=f+\delta\chi$, for some $\chi\in\Gamma^{-1}$,

then

$$\langle\mathscr{F}\rangle=\langle f+\delta\chi\rangle=\langle f\rangle=(f,\psi_0)$$

and $\delta^+\psi_0=0$ implies that $\delta_0^+\psi_0=0$, i.e.,

$$\int_{T^*\mathscr{A}}\Phi_a V^a\psi_0=0, \tag{4.12}$$

for all $V^a\in\Omega^0(T^*\mathscr{A})$.

Thus, as long as ψ_0 satisfies the positivity and normalization conditions introduced at the beginning of this section, we see that we have recovered the expected values the observables f can have and hence, we have shown that the BFV construction is a generalized dynamical description of Yang–Mills theory.

The decomposition of $\mathscr{F}\in\Gamma^0$ into an invariant part plus a coboundary is an important technical step in relating the BFV analysis with the work of Dirac. One would like to understand the associated decomposition of states as a consequence of this decomposition of observables, since then the states will simply correspond to the dual space of the observables.

The extension of these ideas to theories, like gravity, whose constraints are not described in terms of a momentum map is highly nontrivial. The main problem is that one cannot then decompose a generic observable into an invariant part, and thus extracting the dynamical content is much harder. The analysis needed for such a situation will be presented in a later publication.

V. CONCLUSIONS

In this paper a generalized dynamical description of classical Yang–Mills theory has ben presented. Various technical assumptions have been needed in order to carry out this construction: however, we have seen that these are closely related to the assumptions inherent in setting up the Dirac analysis of such a field theory. A consequence of this work is that it supplies a firm dynamical underpinning to the BFV formalism and thus gives a precise geometric characterization of the ghost and conjugate ghost fields.

In this classical analysis we have shown the equivalence between the BFV and the Dirac description of Yang–Mills theory. The BFV approach is distinguished by the use of homological methods both in the description of the observables and the states. This is very reminiscent of the important work by Kugo and Ojima[22] on the use of the BRST charge in the quantization of gauge theories. This suggests that many of the ideas developed here will have a counterpart in the quantum description of these constrained theories. Indeed, one finds that in many respects the BFV analysis of the quantum theory is much richer than the classical one presented here, and has several advantages over the Dirac quantization methods. In particular, the structure of the Hilbert space of states can encode global aspects of the true degrees of freedom.[42] This has important consequences for theories containing anomalies and, more generally, for a deeper understanding of the quantization of constrained systems. This will be discussed in a later publication.

ACKNOWLEDGMENTS

I wish to thank C. J. Isham, K. V. Kuchar, and M.

Henneaux for helpful discussions.

This work was partially supported by the Science and Engineering Research Council (United Kingdom) and under the National Science Foundation Grant No. PHY8503653.

[1]L. D. Faddeev and V. N. Popov, Phys. Lett. **25** B, 30 (1967).
[2]R. P. Feynmann, Acta Phys. Polo. **XXIV**, 697 (1963).
[3]B. S. DeWitt, Phys. Rev. **162**, 1195 (1967).
[4]C. Becchi, A. Rouet, and R. Stora, Ann. Phys. (NY) **98**, 287 (1976).
[5]I. V. Tyutin, Internal report FIAN **39** (1975).
[6]R. Stora, *Cargese Lectures 1976*, edited by M. Levy and P. Mitter (Plenum, New York, 1977).
[7]J. Dixon (unpublished).
[8]J. Thierry-Mieg, J. Math. Phys. **21**, 2834 (1980).
[9]J. M. Leinaas and K. Olaussen, Phys. Lett. B **108**, 199 (1982).
[10]M. E. Mayer, Phys. Lett. B **120**, 355 (1983).
[11]L. Bonora and P. Cotta-Ramusino, Commun. Math. Phys. **87**, 589 (1983).
[12]L. Bonora, P. Pasti, and M. Tonin, J. Math. Phys. **23**, 839 (1982).
[13]J. Hoyos, M. Quiros, J. Ramierz Mittelbrunn, and F. J. DeUrries, Nucl. Phys. B **128**, 159 (1983).
[14]L. Baulieu, Nucl. Phys. B **241**, 557 (1984).
[15]B. Zumino, in *Relativity, Groups and Topology II* (Proceedings of the 1984 Les Houches Summer School), edited by B. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[16]P. A. M. Dirac, "Lectures on Quantum Mechanics," Belfer Graduate School of Science, Yeshiva University, New York, 1964.
[17]A. J. Hanson, T. Regge, and C. Teiltelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976).
[18]E. S. Fradkin and G. A. Vilkovisky, CERN Report No. TH-2332 (1977).
[19]I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).
[20]C. J. Isham and K. V. Kuchar, Ann. Phys. (NY) **164**, 288 (1985).
[21]C. J. Isham and K. V. Kuchar, Ann. Phys. (NY) **164**, 316 (1985).
[22]T. Kugo and I. Ojima, Suppl. Prog. Theor. Phys. **66**, 1 (1970).
[23]S. Hwang, Phys. Rev. D **28**, 2614 (1983).
[24]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 61 (1978).
[25]C. J. Isham, in *Relativity, Groups and Topology II*, edited by B. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[26]M. Atiyah and R. Bott, Philos. Trans. R. Soc. London A **308**, 523 (1982).
[27]I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).
[28]J. M. Arms, J. Math. Phys. **20**, 443 (1979).
[29]A. E. Fischer and J. E. Marsden, in *General Relativity–An Einstein's Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[30]R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reading, MA, 1978).
[31]W. Fulton, *Intersection Theory* (Springer, Berlin, 1984).
[32]B. Malagrange, *Ideals of Differentiable Functions* (Oxford U. P., Oxford, 1966).
[33]N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[34]M. Batchelor, in *Mathematical Aspects of Superspace*, edited by H. J. Seifert, C. J. S. Clark, and A. Rosenblum (Reidel/NATO, Boston, 1984).
[35]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[36]M. Batchelor, Trans. Am. Math. Soc. **258**, 257 (1980).
[37]B. Kostant, in *Differential Geometric Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977).
[38]D. McMullan, Imperial preprint TP/83-84/21, 1984.
[39]M. Henneaux, Phys. Rep. **126**, 1 (1985).
[40]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[41]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis Manifolds and Physics*, (North-Holland, Amsterdam, 1982), 2nd Rev. ed.
[42]D. McMullan, Phys. Rev. D **33**, 2501 (1986).

The Batalin, Fradkin, and Vilkovisky formalism for higher-order theories

A. D. Browning
The Blackett Laboratory, Imperial College, London SW7 2BZ, United Kingdom
David McMullan[a)]
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 19 November 1985; accepted for publication 15 October 1986)

The generalized Batalin, Fradkin, and Vilkovisky (BFV) formalism was developed as a method for determining the ghost structure for theories, such as gravity and supergravity, whose Hamiltonian formalism has constraints not related to a Lie algebra action. Previously, the classical dynamical content of the BFV description of Yang–Mills theory was investigated. There it was found that this approach had a homological interpretation, derived from the Lie algebra cohomology of the gauge group, which allowed one to understand the construction in terms of the Dirac approach to constrained systems. In this paper the dynamical consequences of the generalized BFV formalism are investigated. It is found that even though one no longer has a Lie algebra structure associated with the constraints, one can still develop a homology theory that reproduces the Dirac analysis and from which the generalized BFV formalism can be derived. Some of the consequences of this approach are discussed.

I. INTRODUCTION

In Ref. 1, the BFV formulation of Yang–Mills theory was shown to have a useful cohomological interpretation. The main ingredient to that analysis was the ghost number complex (Ω,δ), where $\Omega = \Sigma_n \Omega^n$ with elements of Ω^n having ghost number n, and δ: $\Omega^n \to \Omega^{n+1}$ was such that $\delta^2 = 0$. Up to this point, the BFV approach seemed identical to the standard discussions of Becchi, Rouet, Stora, and Tyutin (BRST) symmetry; however, as a consequence of working in the canonical formalism, it was found that (Ω,δ) could be written as a double complex. This in turn supplied a rich geometric input which provided a dynamical understanding to the whole construction.

Let us recall the salient points of the above analysis. We started with a phase space $\mathcal{S}$ and a set of first class constraints ϕ_α. So for Yang–Mills theory $\mathcal{S} = \mathcal{A} \times \mathcal{E}$ where $\mathcal{A}$ is the space of connections, $\mathcal{E}$ the space of generalized electric fields, and the constraints are the non-Abelian version of Gauss' law. It was crucial for the whole construction that the constraints were of the first order, i.e., $\{\phi_\alpha,\phi_\beta\} = C^\gamma_{\alpha\beta}\,\phi_\gamma$, where the coefficients $C^\gamma_{\alpha\beta}$ are constants. This was the case for Yang–Mills theory since the constraints could be identified with the momentum map for the gauge group action on $\mathcal{S}$.

For each constraint ϕ_α, a ghost η^α, and conjugate ghost ρ_β were added to the phase space in order to construct a graded phase space $\mathcal{G}$. The ring of all smooth functions on $\mathcal{G}$ was identified with Ω and δ was the (weak) coboundary operator constructed from the constraints. Writing

$$\Omega = \sum_n \Omega^n = \sum_n \sum_{r-s=n} \Omega^{r,s},$$

where an $\omega \in \Omega^{r,s}$ can be written as

$$\omega = \omega^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_r}\eta^{\alpha_1}\cdots\eta^{\alpha_r}\rho_{\beta_1}\cdots\rho_{\beta_s}$$

with coefficients being smooth functions on $\mathcal{S}$, one found that $\delta = \delta_0 + \delta_1$, where

$$\delta_0\colon\ \Omega^{r,s} \to \Omega^{r,s-1} \tag{1.1}$$

and

$$\delta_1\colon\ \Omega^{r,s} \to \Omega^{r+1,s}. \tag{1.2}$$

This decomposition of δ allowed us to compare the cohomology associated with δ to the Dirac analysis of such a constrained system. The antiderivation δ_1 was related to the symplectic action of the constraints, whereas δ_0 induced a "weak equivalence" on this action. A basic result needed for this analysis was that δ_0 is a resolution of the δ_1 action, i.e., δ_0 is an exact operator. This requirement was equivalent to the statement that the constraints were independent, i.e., if we have k constraints ϕ_α and $\mathcal{S}$ has dimension $2n$ $(n > k)$, then $\phi_\alpha = 0$ is a $2n - k$ dimensional submanifold of $\mathcal{S}$.

So for Yang–Mills theory we found that δ produced a weak Lie-algebra cohomology for the gauge group action, and in particular, the zeroth cohomology group correctly characterized the equivalence classes of weakly gauge invariant functions one usually associates with the observables of the constrained theory.

The conclusion from this work was that one could understand the introduction of ghosts into gauge theories in a classical way: They are simply introduced in order to give an alternative characterization of the constrained dynamics. However, there is a sense in which such an approach seems at odds with the way we normally view constrained systems.

The standard classification of constraints is motivated from symplectic geometry and is in terms of first or second class constraints. Thus, a first class system is one for which the constrained surface is coisotropic and, for suitably chosen constraints ϕ_α, this means that

$$\{\phi_\alpha,\phi_\beta\} = U^\gamma_{\alpha\beta}\,\phi_\gamma, \tag{1.3}$$

where the coefficients $U^\gamma_{\alpha\beta}$ are now "structure functions." For a second class system the constrained surface is itself a phase space, and in terms of constraints ψ_α we have

$$\det(\{\psi_\alpha,\psi_\beta\}) \neq 0. \tag{1.4}$$

So, for a first class system the form of the structure functions in (1.3) plays no role in our understanding of the dynamics. Indeed, for Yang–Mills theory we could replace the structure constants $C^\gamma_{\alpha\beta}$ with $C^\gamma_{\alpha\beta} + \phi_\delta H^{\delta\gamma}_{\alpha\beta}$, where $H^{\delta\gamma}_{\alpha\beta}$

[a)] Present address: Department of Physics and Astronomy, The University of Glasgow, Glasgow G128QQ, Scotland.

 0022-2488/87/020438-07$02.50 © 1987 American Institute of Physics

$= H^{[\delta\gamma]}_{[\alpha\beta]}$, and not change the dynamical content of the theory at all. However, the cohomological analysis presented above only worked for the structure constant case.

We also know that there are many important systems where the constraints are not associated with a group action on a phase space, i.e., gravity, supergravity, and string theory. Thus, any new method for dealing with constraints is limited if it cannot cope with such theories.

In Ref. 2, the BFV formalism was formally extended to accommodate systems with structure functions. The basic idea of this extension was quite simple: For a system with structure constants, one could construct a charge Q which has ghost number one and is Abelian, i.e., $\{Q,Q\} = 0$. Explicitly one finds that

$$Q = \phi_\alpha \eta^\alpha - \tfrac{1}{2}\rho_\gamma U^\gamma_{\alpha\beta}\eta^\alpha\eta^\beta. \tag{1.5}$$

To generalize this to the structure function situation one writes the required charge as a power series in the conjugate ghosts, i.e.,

$$Q = \phi_\alpha \eta^\alpha + \rho_\gamma F^\gamma + \rho_\alpha \rho_\beta F^{\alpha\beta} + \cdots \tag{1.6}$$

and then impose the condition $\{Q,Q\} = 0$ in order to determine the coefficients $F^\alpha, F^{\alpha\beta}\cdots$. It is a remarkable fact that such a procedure works (for a careful discussion see Ref. 3) and a charge Q can be constructed. However, one finds that such a Q is not unique and has no obvious geometric significance. These facts combine to obscure any dynamical content to this construction. Indeed it is difficult to see whether this analysis has any physical justification at all, since the types of theories it is attempting to describe are notoriously hard to analyze and thus it has been virtually impossible to trace the consequences of such a formalism.

It is the aim of this paper to present a new approach to the extension of the BFV formalism outlined above. The physical content, and the effect of the various ambiguities, will be made clear. Again, we shall be able to relate this analysis to the standard Dirac approach to such systems. The homological aspects of this construction will be emphasized in this paper.

II. THE CONSTRUCTION

We recall that for a first-order theory, i.e., when the structure functions are constants, the operator δ could be written as $\delta = \delta_0 + \delta_1$. Then $\delta^2 = 0$ implied the following relations:

$$\delta_0^2 = 0, \tag{2.1a}$$

$$\delta_0\delta_1 + \delta_1\delta_0 = 0, \tag{2.1b}$$

$$\delta_1^2 = 0. \tag{2.1c}$$

The action of δ_0 and δ_1 on the basic variables were defined as follows:

$$\delta_1 f := \{\phi_\alpha, f\}\eta^\alpha, \quad f\in\Omega^{0,0}, \tag{2.2}$$

$$\delta_1\eta^\alpha = -\tfrac{1}{2}U^\alpha_{\beta\gamma}\,\eta^\beta\eta^\gamma, \tag{2.3}$$

$$\delta_1\rho_\alpha = -U^\gamma_{\alpha\beta}\eta^\beta\rho_\gamma, \tag{2.4}$$

and

$$\delta_0 f := 0, \quad \delta_0\eta^\alpha := 0, \quad \delta_0\rho_\alpha := \phi_\alpha. \tag{2.5}$$

It is straightforward to check that these definitions satisfy (2.1). Let us now investigate what happens when the structure functions are no longer constant.

From the definitions of δ_0 and δ_1, we find that we can still maintain relations (2.1a) and (2.1b). However, acting on $f\in\Omega^{0,0}$,

$$\delta_1^2 f = \tfrac{1}{2}\phi_\gamma\{U^\gamma_{\alpha\beta}, f\}\eta^\alpha\eta^\beta, \tag{2.6}$$

which is not identically zero.

Let us define $\delta_2\colon \Omega^{0,0}\to\Omega^{2,1}$ by

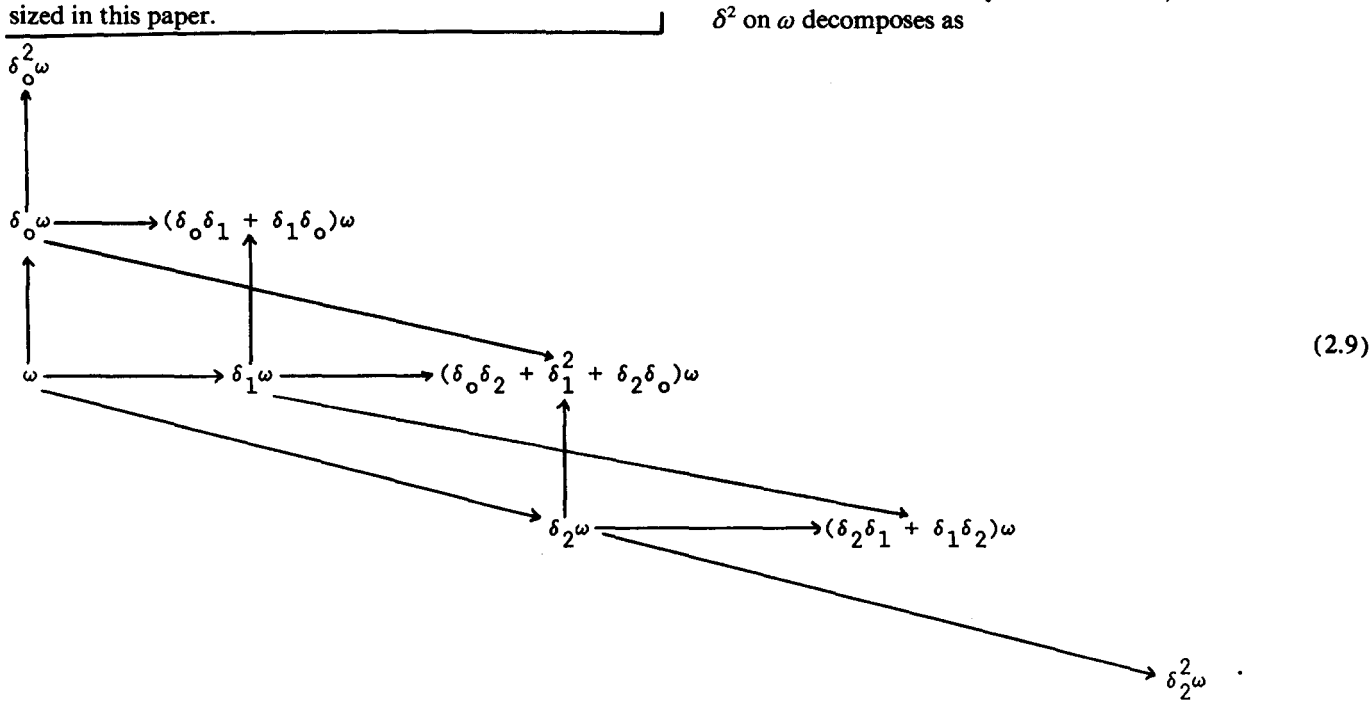
$$\delta_2 f = -\tfrac{1}{2}\{U^\gamma_{\alpha\beta}, f\}\eta^\alpha\eta^\beta\rho_\gamma, \tag{2.7}$$

then we see that

$$(\delta_1^2 + \delta_0\delta_2)f = 0. \tag{2.8}$$

This suggests that instead of writing $\delta = \delta_0 + \delta_1$, we should write $\delta = \delta_0 + \delta_1 + \delta_2$, where $\delta_2\colon \Omega^{r,s}\to\Omega^{r+2,s+1}$, and require that $\delta^2 = 0$. (Note that $\delta_2\eta$ and $\delta_2\rho$ are, as yet, undetermined.)

Consider an arbitrary element $\omega\in\Omega^{r,s}$; then the action of δ^2 on ω decomposes as



(2.9)

From this diagram we see that the required properties are

$$\delta_0^2 = 0, \tag{2.10a}$$

$$\delta_1\delta_0 + \delta_0\delta_1 = 0, \tag{2.10b}$$

$$\delta_2\delta_0 + \delta_1^2 + \delta_0\delta_2 = 0, \tag{2.10c}$$

$$\delta_1\delta_2 + \delta_2\delta_1 = 0, \tag{2.10d}$$

$$\delta_2^2 = 0. \tag{2.10e}$$

We have already seen that (2.10c) is true when acting on functions. Before verifying it on ghosts and conjugate ghosts we need to reconsider Jacobi's identity,

$$0 = \{\phi_{[\alpha}\{\phi_\beta\phi_{\gamma]}\}\} = \{\phi_{[\alpha}U^\delta_{\beta\gamma]}\phi_\delta\}$$

$$= (\{\phi_{[\alpha}U^\delta_{\beta\gamma]}\} - U^\delta_{\epsilon[\alpha}U^\epsilon_{\beta\gamma]})\phi_\delta. \tag{2.11}$$

As was the case in the first-order theory, we assume that δ_0 is exact, thus we can deduce that

$$\{\phi_{[\alpha}U^\delta_{\beta\gamma]}\} - U^\delta_{\epsilon[\alpha}U^\epsilon_{\beta\gamma]} = \phi_\epsilon U^{\epsilon\delta}_{\alpha\beta\gamma}. \tag{2.12}$$

Using (2.12), we find

$$\delta_1^2\eta^\alpha = -\tfrac{1}{2}\phi_\epsilon U^{\epsilon\alpha}_{\beta\gamma\delta}\eta^\beta\eta^\gamma\eta^\delta. \tag{2.13}$$

Thus we take

$$\delta_2\eta^\alpha = \tfrac{1}{2}U^{\alpha\epsilon}_{\beta\gamma\delta}\eta^\beta\eta^\gamma\eta^\delta\rho_\epsilon. \tag{2.14}$$

Similarly, one can show that

$$\delta_2\rho_\alpha = -\tfrac{2}{3}U^{\delta\epsilon}_{\alpha\beta\gamma}\eta^\beta\eta^\gamma\rho_\delta\rho_\epsilon. \tag{2.15}$$

If we define δ_2 by (2.7), (2.14), and (2.15) and extend it by requiring it to be an antiderivation, then it is clear that (2.10c) is true. With this example in mind, we proceed to the general case.

We are given a theory with first-class constraints ϕ_α satisfying (1.3), and we have extended the phase space by the addition of anticommuting variables $(\eta^\alpha,\rho_\alpha)$. (The index α is only used to keep track of algebraic manipulations and thus, unless we say otherwise, can be thought of as an abstract index labeling both discrete and continuous sets of constraints.)

As before we define an antiderivation δ_0: $\Omega^{r,s}\to\Omega^{r,s-1}$ by (2.5) and we assume the constraints are such that this operator is exact. We define the action of δ_1 on $f\in\Omega^{0,0}$ by (2.2). We then require the existence of antiderivations $\delta_1,\ldots,\delta_s,\ldots$, where δ_s: $\Omega^{i,j}\to\Omega^{i+s,j+s-1}$, which satisfy the following relationships:

$$\delta_0^2 = 0,$$

$$\delta_1\delta_0 + \delta_0\delta_1 = 0,$$

$$\delta_0\delta_2 + \delta_1^2 + \delta_2\delta_0 = 0, \tag{2.16}$$

$$\delta_0\delta_3 + \delta_1\delta_2 + \delta_2\delta_1 + \delta_3\delta_0 = 0,$$

$$\delta_0\delta_4 + \delta_1\delta_3 + \delta_2^2 + \delta_3\delta_1 + \delta_4\delta_0 = 0,$$

$$\vdots \quad .$$

In order to prove the existence of these antiderivations we need the followng lemma.

Lemma: Given a set of constraints $\{\phi_\alpha\}$, which are closed under the Poisson bracket, one can define the nth order ($n>0$) structure functions $Q^{\alpha_1\cdots\alpha_n}_{\beta_1\cdots\beta_{n+1}}$ by the formula

$$nQ^{\alpha_1\cdots\alpha_n}_{\beta_1\cdots\beta_{n+1}}\phi_{\alpha_n} = B^{[\alpha_1\cdots\alpha_{n-1}]}_{[\beta_1\cdots\beta_{n+1}]}, \tag{2.17}$$

where

$$B^{\alpha_1\cdots\alpha_n}_{\beta_1\cdots\beta_{n+2}} := \frac{1}{2}\sum_{p=0}^{n}(-1)^{np+1}\{Q^{\alpha_1\cdots\alpha_p}_{\beta_1\cdots\beta_{p+1}},Q^{\alpha_{p+1}\cdots\alpha_n}_{\beta_{p+2}\cdots\beta_{n+2}}\}$$

$$-\sum_{p=0}^{n-1}(-1)^{n(p+1)}(p+1)(n-p+1)$$

$$\times Q^{\alpha_1\cdots\alpha_p k}_{\beta_1\cdots\beta_{p+2}}Q^{\alpha_{p+1}\cdots\alpha_n}_{\beta_{p+3}\cdots\beta_{n+2},k}, \tag{2.18}$$

and we take $Q_\alpha = \phi_\alpha$.

Proof: See Ref. 3. Using the higher-order structure functions we assert that (2.16) is satisfied when

$$\delta_s f = \{Q^{\alpha_1\cdots\alpha_{s-1}}_{\beta_1\cdots\beta_s}, f\}\eta^{\beta_1}\cdots\eta^{\beta_s}\rho_{\alpha_1}\cdots\rho_{\alpha_{s-1}}, \tag{2.19}$$

$$\delta_s\eta^\alpha = (-1)^{s+1}sQ^{\alpha,\alpha_2\cdots\alpha_s}_{\beta_1\cdots\beta_{s+1}}\eta^{\beta_1}\cdots\eta^{\beta_{s+1}}\rho_{\alpha_2}\cdots\rho_{\alpha_s}, \tag{2.20}$$

$$\delta_s\rho_\alpha = (s+1)Q^{\alpha_1\cdots\alpha_s}_{\alpha,\beta_1\cdots\beta_s}\eta^{\beta_1}\cdots\eta^{\beta_s}\rho_{\alpha_1}\cdots\rho_{\alpha_s}. \tag{2.21}$$

We prove this by induction.

From the lemma we have $Q_\alpha = \phi_\alpha$ and $Q^\alpha_{\beta\gamma} = -\tfrac{1}{2}U^\alpha_{\beta\gamma}$. Therefore the definitions of δ_1 agree with (2.2)–(2.4), and hence satisfy $\delta_0^2 = 0$, $\delta_0\delta_1 + \delta_1\delta_0 = 0$. We assume now that we have δ_s defined according to (2.19)–(2.21) for $0\leqslant s\leqslant m$, and that identities

$$\delta_0^2 = 0,$$

$$\delta_1\delta_0 + \delta_0\delta_1 = 0,$$

$$\vdots$$

$$\delta_0\delta_m + \delta_1\delta_{m-1} + \cdots + \delta_{m-1}\delta_1 + \delta_m\delta_0 = 0$$

are true. We now define δ_{m+1} and show that

$$\delta_0\delta_{m+1} + \delta_1\delta_m + \cdots\delta_m\delta_1 + \delta_{m+1}\delta_0 = 0. \tag{2.22}$$

To do this we evaluate $(\delta_1\delta_m + \cdots + \delta_m\delta_1 + \delta_{m+1}\delta_0)$ on each of the functions, ghosts, and conjugate ghosts. Note that for functions and ghosts the term $\delta_{m+1}\delta_0$ drops out, so that there is no difficulty here:

$$\sum_{i=1}^{m}\delta_i\delta_{m+1-i}f = \sum_{i=1}^{m}\delta_i(\{Q^{\alpha_1\cdots\alpha_{m-i}}_{\beta_1\cdots\beta_{m-i+1}},f\}\eta^{\beta_1}\cdots\eta^{\beta_{m+1-i}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m-i}})$$

$$= \sum_{i=1}^{m}\delta_i(\{Q^{\alpha_1\cdots\alpha_{m-i}}_{\beta_1\cdots\beta_{m-i+1}},f\})\eta^{\beta_1}\cdots\eta^{\beta_{m+1-i}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m-i}}$$

$$+ \sum_{i=1}^{m}\{Q^{\alpha_1\cdots\alpha_{m-i}}_{\beta_1\cdots\beta_{m-i+1}},f\}(m-i+1)(\delta_i\eta^{\beta_1})\eta^{\beta_2}\cdots\eta^{\beta_{m+1-i}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m-i}}$$

$$+ \sum_{i=1}^{m}\{Q^{\alpha_1\cdots\alpha_{m-i}}_{\beta_1\cdots\beta_{m-i+1}},f\}\eta^{\beta_1}\cdots\eta^{\beta_{m+1-i}}(m-i)(-1)^{m-i+1}(\delta_i\rho_{\alpha_1})\rho_{\alpha_2}\cdots\rho_{\alpha_{m-i}}.$$

Where we have used the antisymmetry in $\alpha_1\cdots\alpha_{m-i}$ and $\beta_1\cdots\beta_{m-i+1}$, and the properties of the antiderivation. Substituting from (2.19)–(2.21) and simplifying gives

$$\sum_{i=1}^{m}\delta_i\delta_{m+1-i}f$$
$$=\left[\{-B^{\alpha_1\cdots\alpha_{m-1}}_{\beta_1\cdots\beta_{m+1}},f\}\right.$$
$$\left.+\{\phi_\tau,f\}m(-1)^{m+1}Q^{\tau,\alpha_1\cdots\alpha_{m-1}}_{\beta_1\cdots\beta_{m+1}}\right]$$
$$\times\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m-1}}.$$

Then, using the lemma we get

$$\sum_{i=1}^{m}\delta_i\delta_{m+1-i}f$$
$$=\delta_0(-\{Q^{\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_{m+1}},f\}\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_{\alpha_1}\cdots\rho_{\alpha_m}).$$

Hence, if we define

$$\delta_{m+1}f=\{Q^{\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_{m+1}},f\}\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_{\alpha_1}\cdots\rho_{\alpha_m} \quad (2.23)$$

then we have shown that (2.22) is true when applied to functions.

Similarly one can show

$$\sum_{i=1}^{m}\delta_i\delta_{m+1-i}\eta^\alpha$$
$$=(-1)^m mB^{\alpha,\alpha_1\cdots\alpha_{m-1}}_{\beta_1\cdots\beta_{m+2}}\eta^{\beta_1}\cdots\eta^{\beta_{m+2}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m-1}}$$
$$=-\delta_0((-1)^m(m+1)$$
$$\times Q^{\alpha,\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_{m+2}}\eta^{\beta_1}\cdots\eta^{\beta_{m+2}}\rho_{\alpha_1}\cdots\rho_{\alpha_m}).$$

Then we can take

$$\delta_{m+1}\eta^\alpha=(-1)^m(m+1)$$
$$\times Q^{\alpha,\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_{m+2}}\eta^{\beta_1}\cdots\eta^{\beta_{m+2}}\rho_{\alpha_1}\cdots\rho_{\alpha_m}. \quad (2.24)$$

When considering conjugate ghosts, we notice that the $\delta_{m+1}\delta_0$ term in (2.22) is not zero. However $\delta_{m+1}\delta_0\rho_\alpha=\delta_{m+1}\phi_\alpha$, for which we have an expression from (2.23). Thus

$$\sum_{i=1}^{m+1}\delta_i\delta_{m+1-i}\rho_\alpha=(m+2)B^{\alpha_1\cdots\alpha_m}_{\alpha,\beta_1\cdots\beta_{m+1}}$$
$$\times\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_{\alpha_1}\cdots\rho_{\alpha_m}$$
$$=-\delta_0((m+2)Q^{\alpha_1\cdots\alpha_{m+1}}_{\alpha,\beta_1\cdots\beta_{m+1}}$$
$$\times\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_\alpha\cdots\rho_{\alpha_{m+1}}).$$

Thus we choose

$$\delta_{m+1}\rho_\alpha=(m+2)Q^{\alpha_1\cdots\alpha_{m+1}}_{\alpha,\beta_1\cdots\beta_{m+1}}\eta^{\beta_1}\cdots\eta^{\beta_{m+1}}\rho_{\alpha_1}\cdots\rho_{\alpha_{m+1}}. \quad (2.25)$$

So, if we define δ_{m+1} by (2.23)–(2.25) and extend it as an antiderivation to all elements in Ω, then (2.22) will be true. This completes the induction step.

If there exists an integer k such that $\delta_m=0$ for all $m>k$, then we say that we have a k th-order theory and we can define $\delta\colon\Omega^p\to\Omega^{p+1}$ by

$$\delta=\delta_0+\delta_1+\cdots+\delta_k \quad (2.26)$$

and $\delta^2=0$. Thus Yang–Mills theory is of the first order, and one can show that gravity is of the second order.

It is possible that no finite k exists and hence we have a theory of infinite order. Indeed, this could well be an interpretation of the results in Ref. 4 where a gauge theory of self-interacting massless spin-3 particles was shown to be unobtainable. Such infinite-order theories will be excluded in what follows.

The order of a theory is a useful concept, but it is not well defined since in the above proof we have simply found a choice for the $\{\delta_m\}$, which satisfies (2.16): As mentioned for the first-order theories, there is an arbitrariness in the definitions of the structure functions and hence in the order of the theory. We now discuss what is well defined in this formalism.

III. THE AMBIGUITIES

In the Introduction we saw that one can always replace $U^\gamma_{\alpha\beta}$ by $U^\gamma_{\alpha\beta}+\phi_\delta H^{[\delta\gamma]}_{\alpha\beta}$ in (1.3) without changing the dynamics of the system. Clearly such a possibility can occur for each order of structure functions, and we need to investigate how the dynamical content of the generalized BFV formalism is effected by such variations.

Under the replacement $U^\gamma_{\alpha\beta}\to\hat{U}^\gamma_{\alpha\beta}=U^\gamma_{\alpha\beta}+\phi_\delta H^{[\delta\gamma]}_{\alpha\beta}$, the definition of δ_0 (2.5) and $\delta_1 f$ (2.2) are unchanged. However both $\delta_1\eta$ and $\delta_1\rho$ will pick up an additional $\delta_0(\ )$ term. Thus, given a derivation $\kappa_1\colon\Omega^{i,j}\to\Omega^{i+1,j+1}$ with $\kappa_1 f=0$, the most general change in δ_1, due to the above ambiguity in the structure functions, is to replace δ_1 by $\hat{\delta}_1$, where

$$\hat{\delta}_1=\delta_1-\delta_0\kappa_1. \quad (3.1)$$

However as it stands, (3.1) is not satisfactory since it is not necessarily an antiderivation and it no longer satisfies (2.1b). Both of these problems are resolved by writing

$$\hat{\delta}_1=\delta_1+\kappa_1\delta_0-\delta_0\kappa_1=\delta_1+[\kappa_1,\delta_0], \quad (3.2)$$

since the commutator of a derivation and an antiderivation is always an antiderivation and clearly (2.1) holds.

This argument can be repeated for $\delta_2,\delta_3,\ldots$ and one finds that the possible changes in such maps are given by

$$\hat{\delta}_2=\delta_2+[\kappa_2,\delta_0]+[\kappa_1,\delta_1]+\tfrac{1}{2}[\kappa_1,[\kappa_1,\delta_0]],$$
$$\hat{\delta}_3=\delta_3+[\kappa_3,\delta_0]+[\kappa_2,\delta_1]+[\kappa_1,\delta_2]$$
$$+\tfrac{1}{2}[\kappa_1,[\kappa_1,\delta_1]]+\tfrac{1}{2}[\kappa_2,[\kappa_1,\delta_0]] \quad (3.3)$$
$$+\tfrac{1}{2}[\kappa_1,[\kappa_2,\delta_0]]+(1/3!)[\kappa_1[\kappa_1,[\kappa_1,\delta_0]]],$$

where $\kappa_i\colon\Omega^{r,s}\to\Omega^{r+i,s+i}$ are derivations. While writing the general form of $\hat{\delta}_m$ like this is easy, we shall need to express them in a form more amenable to calculation.

Lemma: If we are given δ and $\hat{\delta}$ on Ω, both of which satisfy (2.16), and are such that

(i) $\hat{\delta}_0=\delta_0$,

(ii) $\hat{\delta}_1=\delta_1+\kappa_1\delta_0-\delta_0\kappa_1$.

Then the general form for $\hat{\delta}_m$ is

$$\hat{\delta}_m=\delta_m+\sum_{s>0,t>0}\frac{(-1)^t}{s!t!}\kappa_{i_1}\cdots\kappa_{i_s}\delta_a\kappa_{i_{s+1}}\cdots\kappa_{i_{s+t}}, \quad (3.4)$$

where the sum is over all s and t, and all indices that satisfy $\Sigma_{j=1}^{s+t}i_j+a=m$. In the expression the κ_i are derivations with $k_i\colon\Omega^{r,s}\to\Omega^{r+i,s+i}$.

Before sketching the proof of this lemma we first make a few comments.

(a) The statement of the lemma is slightly stronger than needed for the ambiguity discussed above, since we do not require that acting on $f\in\Omega^{0,0}$, $\kappa_1 f=0$.

(b) It can be shown directly that a $\hat{\delta}_m$ defined by (3.4) will satisfy (2.16).

(c) $\hat{\delta}_m$ constructed as shown can be written in the $[\kappa_i[\kappa_j\cdots[\kappa_1,\delta_a]]]$ form and hence it is an antiderivation.

Proof: We use induction, so by (ii), we can assume that, for all $m<s$, (3.4) holds. Acting on f (or η^α)

$$\delta_0\hat{\delta}_{s+1}f=\sum_{j=1}^{s}\hat{\delta}_j\hat{\delta}_{s+1-j}f. \tag{3.5}$$

However, we can now use the induction hypothesis to rewrite the right-hand side of (3.5) in terms of δ_a and κ_i,

$$\delta_0\hat{\delta}_{s+1}f=\delta_0\Bigg(\delta_{s+1}+\sum_{s>0,t>0}\frac{(-1)^t}{s!t!}\times k_{i_1}\cdots\kappa_{i_s}\delta_a\kappa_{i_{s+1}}\cdots\kappa_{i_{s+t}}+\sum_{u>0,v>0}\frac{(-1)^v}{u!v!}\kappa_{i_1}\cdots\kappa_{i_u}\delta_0\kappa_{i_{u+1}}\kappa_{i_{u+v}}\Bigg)f,$$

where in the first sum $a\neq 0$. Using the resolution property of δ_0, the $\delta_0\hat{\delta}_{s+1}$ action on ρ_α and the requirement that $\hat{\delta}_{s+1}$ is an antiderivation then gives the required result.

So we can now construct two ghost number complexes, (Ω,δ) and $(\Omega,\hat{\delta})$, and we need to understand the relationship between them. We know that it is the cohomology of the complex that has dynamical significance, thus what we want to show is that these complexes have the same cohomology.

There is a standard way to compare two complexes; we need chain mappings κ: $(\Omega,\delta)\to(\Omega,\hat{\delta})$ and κ': $(\Omega,\hat{\delta})\to(\Omega,\delta)$ such that

$$\kappa\delta=\hat{\delta}\kappa, \tag{3.6a}$$

$$\kappa'\hat{\delta}=\delta\kappa', \tag{3.6b}$$

and

$$\kappa\kappa'=id_\Omega=\kappa'\kappa. \tag{3.6c}$$

We now prove that

$$\kappa=1+\sum_{r>0}\frac{1}{r!}\kappa_{i_1}\cdots\kappa_{i_r} \tag{3.7}$$

and

$$\kappa'=1+\sum_{r>0}\frac{(-1)^r}{r!}\kappa_{i_1}\cdots\kappa_{i_r} \tag{3.8}$$

are the required mappings. In (3.7) and (3. 8) the sum is over all $r>0$, and all values of $i_1\cdots i_r$.

The typical term on the left-hand side of (3.6a) will be

$$\sum_{r>0}\frac{1}{r!}\kappa_{i_1}\cdots\kappa_{i_r}\delta_a. \tag{3.9}$$

The typical term on the right-hand side is

$$\sum_{s>0,t>0}C(s,t)\kappa_{i_1}\cdots\kappa_{i_s}\delta_a\kappa_{i_{s+1}}\cdots\kappa_{i_{s+t}}, \tag{3.10}$$

where $c(s,t)\in\mathbb{R}$. To prove the equality of these expressions we separate the cases: (i) $t\neq 0$, (ii) $t=0$.

We fix values of $i_1, i_2\cdots i_{s+t}$ and a. For case (i), $t\neq 0$,

$$C(s,t)=\sum_{u=0}^{t}\frac{(-1)^u}{s!u!}\frac{1}{(t-u)!}=\frac{1}{s!t!}\sum_{u=0}^{t}\binom{t}{u}(-1)^u=0.$$

To see this, one looks at the possible contributions to a particular term in (3.10),

$$\left(\frac{1}{s!}\kappa_{i_1}\cdots\kappa_{i_s}\delta_a\right)\left(\frac{1}{t!}\kappa_{i_{s+1}}\cdots\kappa_{i_{s+t}}\right)+\left(\frac{-1}{s!}\kappa_{i_1}\cdots\kappa_{i_s}\delta_a\kappa_{i_{s+1}}\right)\times\left(\frac{1}{(t-1)!}\kappa_{i_{s+2}}\cdots\kappa_{i_{s+t}}\right)+\text{etc.}$$

For case (ii), $t=0$, the contribution must come from $\hat{\delta}1$. Thus $C(s,0)=1/s!$.

This proves that κ is a chain mapping between the two complexes. Similarly one can show (3.6b) and (3.6c). Now, using standard homological methods[5] we can deduce that the two complexes above have isomorphic cohomology.

We already know that for a first-order theory $H^0(\delta)$, the zeroth cohomology group, characterizes the physical observables of the constrained system. Thus, we have now shown that for any higher-order formulation $\hat{\delta}$ of this theory, where $\hat{\delta}$ is related to δ via (3.4), $H^0(\hat{\delta})$ still describes the physical observables.

This result is all well and good, however, it falls short of giving us a complete understanding of the dynamical content of a general higher-order theory since it is not at all clear that one can always relate such a theory to a first-order one via relations of the form (3.4).

IV. THE DYNAMICAL CONTENT

We have shown that for a given set of first class constraints ϕ_α, we can construct a ghost number complex (Ω,δ). For a first-order theory we know that $H^0(\delta)$ characterizes the observables, and we now extend this result to higher-order theories.

Recall that a function $f\in\Omega^{0,0}$ is said to be weakly invariant if

$$\{\phi_\alpha,f\}=V_\alpha^\beta\phi_\beta, \tag{4.1}$$

for some $V_\alpha^\beta\in\Omega^{0,0}$. We now show that given such an f, we can construct $\mathcal{F}=f+\omega_1+\omega_2+\cdots\in\Omega^0$, with $\omega_i\in\Omega^{i,i}$, such that $\delta\mathcal{F}=0$. Rather than prove this in general for a theory of order n, we prove it only for $n=2$. The general case is similar.

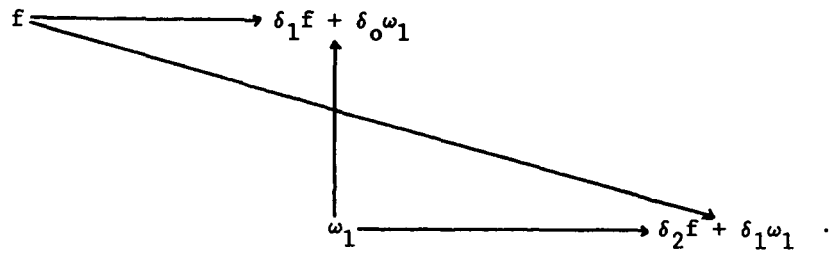
Define

$$\omega=V_\alpha^\beta\eta^\alpha\rho_\beta\in\Omega^{1,1}, \tag{4.2}$$

then

$$\delta_1 f+\delta_0\omega=0. \tag{4.3}$$

Consider the following diagram:



Then,

$$\begin{aligned}\delta_0(\delta_2 f+\delta_1\omega_1) &= \delta_0\delta_2 f-\delta_1\delta_0\omega_1\\ &= -\delta_1^2 f-\delta_1\delta_0\omega_1\\ &= 0, \quad \text{using (4.3).}\end{aligned}$$

So, by the exactness of δ_0, there exists an $\omega_2\in\Omega^{2,2}$ such that

$$\delta_2 f+\delta_1\omega_1+\delta_0\omega_2=0. \tag{4.4}$$

Now, let us look at the $\Omega^{3,1}$ terms in $\delta(f+\omega_1+\omega_2)$:

$$\begin{aligned}\delta_0(\delta_2\omega_1+\delta_1\omega_2) &= -\delta_2\delta_0\omega_1-\delta_1^2\omega_1-\delta_1\delta_0\omega_2\\ &= \delta_2\delta_1 f-\delta_1(\delta_1\omega_1+\delta_0\omega_2)\\ &= \delta_2\delta_1 f+\delta_1\delta_2 f, \quad \text{from (4.4)}\\ &= 0.\end{aligned}$$

Therefore, there exists $\omega_3\in\Omega^{3,3}$ such that

$$\delta_2\omega_1+\delta_1\omega_2+\delta_0\omega_3=0. \tag{4.5}$$

It is easy to see that this argument can be continued and $\delta(f+\omega_1+\omega_2+\cdots)=0$. For a generic field theory, one might have to worry about a finite termination to this process; however, we shall not concern ourselves with this point here.

Thus to each weakly invariant function we can associate an equivalence class in $H^0(\delta)$. For a first-order theory, one could show that if $\omega\in\Omega^0$ is such that $\delta\omega=0$, then there exists (at least locally in the phase space) a $\psi\in\Omega^{-1}$ such that $\omega-\delta\psi\in\Omega^{0,0}$, i.e., in each equivalence class there exists an invariant function. This allowed us to relate $H^0(\delta)$ with the Dirac prescription for observables. However, for a higher-order theory we do not expect such a decomposition of $\omega\in\Omega^0$ and thus it is not clear how $H^0(\delta)$ will relate to the standard class of observables. In order to analyze $H^0(\delta)$ we first discuss another amibiguity in the formulation of constrained dynamics.

The first-class characterization of a constrained system is a statement about the symplectic structure of the constrained surface. As such, the constraints used to describe this surface play a minor role in the dynamical understanding of the system. So consider a new set of constraints $\tilde{\phi}_\alpha$, where

$$\tilde{\phi}_\alpha=\Lambda_\alpha^\beta\phi_\beta, \tag{4.6}$$

and Λ_α^β is a (locally) invertible transformation acting on the old constraints. Then $\tilde{\phi}_\alpha$ are also first class with

$$\{\tilde{\phi}_\alpha,\tilde{\phi}_\beta\}=\overline{U}^\gamma_{\alpha\beta}\tilde{\phi}_\gamma,$$

where

$$\begin{aligned}\overline{U}^\gamma_{\alpha\beta} &= \Lambda^\tau_\alpha\Lambda^\delta_\beta U^\rho_{\tau\delta}(\Lambda^{-1})^\gamma_\rho+\{\Lambda^\tau_{[\alpha},\Lambda^\delta_{\beta]}\}\phi_\tau(\Lambda^{-1})^\gamma_\delta\\ &\quad+\Lambda^\tau_\alpha\{\phi_\tau,\Lambda^\rho_\beta\}(\Lambda^{-1})^\gamma_\rho-\Lambda^\tau_\beta\{\phi_\tau,\Lambda^\rho_\alpha\}(\Lambda^{-1})^\gamma_\rho.\end{aligned} \tag{4.7}$$

Thus we would expect that any formulation of the constrained dynamics should be invariant under transformations of the form (4.6). A problem with the Dirac analysis is that (4.6) is not a canonical transformation on the extended phase space $\mathcal{S}$, and thus one has to explicitly show the relationship between the two formulations of the same theory. This introduces several undesirable aspects to the Dirac formalism, some of which we shall discuss later.

One can show, at least for finite dimensional systems, that locally the $\tilde{\phi}_\alpha$ can be constructed such that $\overline{U}^\gamma_{\alpha\beta}=0$. Thus all first-class theories have a first-order formulation. This is a well known result, but has had little use in our understanding of constrained systems, since there are usually many other physical reasons for not changing the constraints, i.e., physical interpretation and locality. However, we shall find that (4.6) provides us with a method for analyzing $H^0(\delta)$ and will ultimately lead to a deeper understanding of the central role fermionic methods can play in constrained dynamics.

For the constraints ϕ_α, we introduce the ghosts η^α and conjugate ghosts ρ_α, and construct the ghost number complex (Ω,δ). Similarly, for $\tilde{\phi}_\alpha$ we introduce $\tilde{\eta}^\alpha$ and $\tilde{\rho}_\alpha$ and build the complex $(\tilde{\Omega},\tilde{\delta})$. In order to study the relationship between these objects, we need to understand the $\tilde{\delta}$ action on Ω.

We know that $\delta_0\rho_\alpha=\phi_\alpha$ and $\tilde{\delta}_0\tilde{\rho}_\alpha=\tilde{\phi}_\alpha$, therefore $\delta_0(\Lambda^\beta_\alpha\rho_\beta)=\tilde{\delta}_0\tilde{\rho}_\alpha$. This suggests that we should take

$$\tilde{\rho}_\alpha=\Lambda^\beta_\alpha\rho_\beta, \tag{4.8}$$

since then $\tilde{\delta}_0=\delta_0$.

Let us now investigate $\tilde{\delta}_1 f, f\in\Omega^{0,0}=\tilde{\Omega}^{0,0}$,

$$\tilde{\delta}_1 f:=\{\tilde{\phi}_\alpha,f\}\tilde{\eta}^\alpha=\{\phi_\alpha,f\}\Lambda^\alpha_\beta\tilde{\eta}^\beta+\phi_\beta\{\Lambda^\beta_\alpha,f\}\tilde{\eta}^\alpha.$$

If we let

$$\tilde{\eta}^\alpha=(\Lambda^{-1})^\alpha_\beta\eta^\beta, \tag{4.9}$$

then

$$\tilde{\delta}_1 f=\delta_1 f-\delta_0\kappa_1 f,$$

where

$$\begin{aligned}\kappa_1 f &= (\Lambda^{-1})^\alpha_\gamma\{\Lambda^\beta_\alpha,f\}\eta^\gamma\rho_\beta\in\Omega^{1,1}\\ &= (\Lambda^{-1})^\delta_\beta\{\Lambda^\beta_\alpha,f\}\tilde{\eta}^\alpha\tilde{\rho}_\delta\in\tilde{\Omega}^{1,1}.\end{aligned} \tag{4.10}$$

Similarly, one can show that

$$\tilde{\delta}_1\eta^\alpha = \tilde{\delta}_1(\Lambda^\alpha_\beta\tilde{\eta}^\beta)$$
$$= \delta_1\eta^\alpha - \delta_0\kappa_1\eta^\alpha,$$

where

$$\kappa_1\eta^\alpha = -\tfrac{1}{2}(\Lambda^{-1})^\gamma_\rho(\Lambda^{-1})^\beta_\tau\{\Lambda^\delta_\gamma,\Lambda^\alpha_\beta\}\eta^\rho\eta^\tau\rho_\delta\in\Omega^{2,1}. \tag{4.11}$$

Using (4.10), one finds that

$$\tilde{\delta}_1\rho_\alpha = \delta_1\rho_\alpha + \kappa_1\delta_0\rho_\alpha - \delta_0\kappa_1\rho_\alpha,$$

where

$$\kappa_1\rho_\alpha = -\tfrac{1}{2}(\Lambda^{-1})^\beta_\gamma(\Lambda^{-1})^\epsilon_\alpha\{\Lambda^\kappa_\epsilon,\Lambda^\tau_\beta\}\eta^\gamma\rho_\kappa\rho_\tau. \tag{4.12}$$

Thus, by using the transformations (4.6), (4.8), and (4.9), we have been able to show that $\tilde{\delta}_0 = \delta_0$ and $\tilde{\delta}_1 = \delta_1 + [\kappa_1,\delta_0]$. Hence, we can use the results of Sec. III to deduce that $H^0(\delta) = H^0(\tilde{\delta})$.

As we have already pointed out, one can choose a first-order $\tilde{\delta}$, where we already know that $H^0(\tilde{\delta})$ corresponds to the observables. So we can deduce that, at least locally in the phase space, $H^0(\delta)$ will correctly characterize the observable in a constrained theory of any order.

V. CONCLUSION

We have presented a homological description of first class constrained systems. The relationship between this construction and the generalized BFV formalism is accomplished by substituting the higher-order structure functions (smeared with the ghost fields) for the coefficients in (1.6), then $\delta\mathcal{F} := \{Q,\mathcal{F}\}$, where the super-Poisson bracket is used. These methods have been shown to supply the same dynamical information as the standard Dirac analysis of such systems.

It might be felt that, even if this is a novel and possibly unexpected result, all we have really achieved is an unnecessary complication of what was basically a simple way to understand constrained dynamics. In order to answer this criticism let us discuss the motivation for this work.

Ghosts were originally introduced into quantum field theory in order to construct a unitary expression for the S matrix of Yang–Mills theory. It was clear that the breakdown of unitarity was related to the existence of constraints in the classical formulation of these theories; however, the ghosts, and associated BRST symmetry, were taken to be purely quantum constructions, with no classical dynamical significance. In recent years, the homological background to these quantum techniques have taken on a central role in our understanding of such theories. One finds that even if the quantization procedure is incompatible with the gauge invariant structure of the system, it still respects the cohomological structure of the gauge theory.[6]

Thus, it is natural to ask whether the central role played by homological methods in the quantum theory is an artifact of quantization or a reflection of a more fundamental description of such constrained systems. This question takes on a new significance when one is dealing with theories like gravity, where a perturbative understanding of the quantum field theory does not exist, so that we would like to introduce the homological structure as a consequence of some classical analysis.

We have shown that the use of homological methods is not an artifact of quantization, and indeed, the use of ghosts and their symmetries translates directly into the classical vocabulary. However, there is a surprising bonus to this approach which we have not discussed in this paper. The additional odd variables in this construction introduce a graded symplectic structure into the classical formalism. This structure has an unexpected interplay with the homological methods introduced in this paper. One finds that δ induces (odd) canonical transformations on this space. This could be understood for the BFV formulation for Yang–Mills theory, since there the odd generator Q, (1.5), could be described as the momentum map for the lift of the configuration space Lie algebra cohomology.[1] However, that such a result holds for an arbitrary order theory is far from obvious and suggests that the full significance of the (graded) symplectic structure should be investigated.

One finds[7] that the additional odd degrees of freedom have subtly changed the structure of the allowed canonical transformations. In particular, one finds that the chain mappings κ and κ' (3.7) and (3.8) are actually (even) canonical transformations. So, in the formalism, the replacement of ϕ_α by $\tilde{\phi}_\alpha$ [cf. (4.6)] is an allowed invariance of the theory, in contrast to the Dirac approach. This allows us a greater flexibility in describing constrained systems, and in particular, allows for a much deeper insight into how one should construct polarizations for constrained systems.[7]

One can view the introduction of an odd variable into constrained dynamics as analogous to the introduction of a complex number into the study of polynomials. Indeed we propose that the methods presented in this paper are a natural and useful description of first class constrained systems. The full consequences of this approach are presently under investigation.

ACKNOWLEDGMENTS

We both wish to thank C. J. Isham for his interest and encouragement during the preparation of this paper. D. M. is grateful to K. Kuchar andM. Henneaux for useful discussions.

A. D. B. was supported by the Carnegee trust for the Universities of Scotland. The work of D. M. was partially supported by the S.E.R.C. and under NSF Grant No. PHY-8503653.

[1] D. McMullan, J. Math.Phys. **28**, 428 (1987).
[2] E. S. Fradkin and T. E. Fradkina, Phys. Lett. B **72**, 343 (1978); I. A. Batalin and G. A. Vilkovisky, J. Math. Phys. **26**, 172 (1985).
[3] M. Henneaux, Phys. Rep. **126**, 1 (1985).
[4] F. A. Berends, G. J. H. Burgers, and H. Van Dam, Nucl. Phys. B **260**, 295 (1985).
[5] S. Maclane, *Homology* (Springer, Berlin, 1975).
[6] L. Bonora and P. Cotta-Rumusino, Commun. Math. Phys. **87**, 589 (1983); B. Zumino, in *Relativity, Groups and Topology II* (Proceedings of the 1984 Les Houches Summer School), edited by B. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[7] D. McMullan, "The use of ghost variables in the description of constrained systems," in *Proceedings of the First International Conference on the Physics of Phase Space, Lecture Notes in Physics* (Springer, Berlin, to appear).

Geometrostatis and the current algebra of nonlinear sigma models on supergroup manifolds

K. Kobayashi
Department of Physics, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

(Received 12 March 1986; accepted for publication 8 October 1986)

The radiative correction of nonlinear sigma models on supermanifolds that have invertible metrics is investigated. It will be shown that the equation of motion for Riemannian supergravity (nonstandard supergravity) is derived from a consistency condition. This condition can be satisfied in the case of supergroup manifolds. We shall explicitly construct the model following the methods of Braaten, Curtright, and Zachos [E. Braaten, T. L. Curtright, and C. K. Zachos, Nucl. Phys. B **260**, 630 (1985)] and of Witten [E. Witten, Commun. Math. Phys. **92**, 455 (1984)]. Finally, super-Kac–Moody algebras of these models are derived.

I. INTRODUCTION

There has been much interest in two-dimensional nonlinear sigma models with the Wess–Zumino term.[1] It is shown that they have nontrivial infrared fixed points and, on these fixed points, its light-cone currents realize the Kac–Moody algebra.[2] Witten has used this algebra for bosonization, that is, to show the equivalence of free fermionic theories and nonlinear sigma models.

These models were extended to supersymmetric models, which have supersymmetry in two dimensions.[3] This is an interesting extension because they have the super-Kac–Moody algebra and the super-Virasoro algebra. In this paper we investigate another type of supersymmetric extension. For that purpose we introduce Grassmannian fields in addition to bosonic fields (which are often called coordinates of target space). This means that we construct supersymmetry in target space.

In the following we adopt a formulation of geometry of a superspace following Ref. 4. But here we shall use the word geometry in a restricted sense; we are not concerned with global or topological structure in this paper. The formulation presented below is only a formal description. The reader who is interested in a more rigorous treatment of supermanifolds is referred to Refs. 4 and 5, for example.

We shall investigate the nonlinear sigma model in that formalism, restricting ourselves to the case where the metric of target space is invertible. This is necessary for Riemannian geometry and background field expansion. This assumption leads to the equation of motion of nonstandard supergravity as a consistency condition. For that reason our model does not seem relevant to the string model. But we think that these models are interesting by themselves, because of their Riemannian structure and current algebra, which we shall see in the following.

In the bosonic case, the parallelism of group manifolds plays an important role.[6] We can use the parallelism of supergroup manifolds to show the one-loop finiteness and construct the super-Kac–Moody algebra as an extension of Refs. 6 and 2.

II. RIEMANNIAN GEOMETRY OF THE NONLINEAR SIGMA MODEL ON A SUPERMANIFOLD

We consider the nonlinear sigma model described by the following Lagrangian:

$$\mathscr{L} = \partial_\mu Z^a \, {}_a g_b \, \partial_\mu Z^b + \tfrac{2}{3} \epsilon^{\nu\mu} \, \partial_\mu Z^a \, {}_a e_b \, \partial_\nu Z^b, \tag{1}$$

where μ represents the suffix of the two-dimensional coordinate, Z^a is the coordinate of the target space, and ${}_a g_b$ and ${}_a e_b$ are the metric and antisymmetric tensors. [They have the following symmetries: ${}_b g_a = (-1)^{ab+a+b} \, {}_a g_b$, ${}_b e_a = -(-1)^{ab+a+b} \, {}_a e_b$.] We briefly summarize the geometry of the superspace in the Appendix. In this formulation, the equation of the motion from the above Lagrangian is given by

$$\partial_\mu \, \partial_\mu Z^a + \Gamma^a{}_{bc} \, \partial_\mu Z^c \, \partial_\mu Z^b - S^a{}_{bc} \epsilon_{\mu\nu} \, \partial_\mu Z^c \, \partial_\nu Z^b = 0,$$
$$S^a{}_{bc} = (-1)^{a\,gad} \, e_{(db,c)_s}, \tag{2}$$

where

$$[ABC]_s = \tfrac{1}{3}\{ABC + (-1)^{A(B+C)} BCA + (-1)^{C(A+B)} CAB\}.$$

and where $\Gamma^a{}_{bc}$ and $S^a{}_{bc}$ are the Christoffel symbol and the torsion.

We shall investigate the radiative correction to this model in the background field method. If we expand Z^a around the classical value $Z^a_{\rm cl}$ in terms of the normal coordinate ξ^a as

$$Z^a = Z^a_{\rm cl} = \xi^a + \tfrac{1}{2} \Gamma^a{}_{bc} \xi^c \xi^b + \cdots, \tag{3}$$

then the action is written as follows:

$$I = I^{(0)} + I^{(1)} + I^{(2)} + \cdots,$$
$$I^{(0)} = I(Z^a_{\rm cl}),$$
$$I^{(1)} = \int d^2x \{ -2(\partial_\mu Z^a \, \tilde{D}_\mu \, {}_a g_b \, \xi^b)\}, \tag{4}$$
$$I^{(2)} = \int d^2x \{(\xi^a \tilde{D}_\mu) \, {}_a g_b \, (\xi^b \tilde{D}_\mu) - \partial_\mu Z^a_{\rm cl} \, {}_a g_b \, \tilde{R}^b{}_{cde} \xi^e \, \partial_\nu Z^d_{\rm cl} \, \xi^c (\delta_{\mu\nu} - \epsilon_{\mu\nu})\},$$

where

 0022-2488/87/020445-03$02.50 © 1987 American Institute of Physics

$$V^a\tilde{D}_\mu = V^b\{\delta^a{}_b \overleftarrow{\partial}_\mu + (-1)^{b(a+1)} \times (\Gamma^a{}_{bc}\delta_{\mu\nu} - S^a{}_{bc}\epsilon_{\mu\nu})\partial_\nu Z^c_{\text{cl}}\}$$

and $\tilde{R}^a{}_{bcd}$ is the generalized Riemannian tensor given by the connection $F^a{}_{bc} = \Gamma^a{}_{bc} - S^a{}_{bc}$.

If we define the fluctuating quantum field on the tangent space using the vierbein $g_{ab} = V_a{}^i V_b{}^j \eta_{ij}(-1)^{ai}$ by

$$\zeta^i = V^i{}_a \xi^a, \tag{5}$$

then the vacuum function $\langle \zeta^i(x)\zeta^j(y)\rangle$ is proportional to η^{ij} [η^{ij} is the inverse of η_{ij} and satisfies $\eta^{ji} = (-1)^{ij}\eta^{ij}$, it is a generalization of δ^{ij} in the bosonic case]. Then the on-shell one-loop divergence in $\langle e^{iI}\rangle$ is proportional to

$$\tilde{R}^a{}_b = \tilde{R}^a{}_{cbd} g^{cd}(-1)^{bc}. \tag{6}$$

So the one-loop finiteness of this model is equivalent to the generalized Ricci flatness ($\tilde{R}^a{}_b = 0$) of the target space in the Riemannian formalism. This is the equation of motion of nonstandard supergravity, i.e., the Riemannian supergravity.[7] The Riemannian supergravity chooses the supergroup as a tangent group. This corresponds to our assumption of the invertibility of ${}_a g_b$. (Although we can interpret the new superstring action of Green and Schwarz as a nonlinear sigma model on a supermanifold,[8] its metric does not have the inverse.)

III. NONLINEAR SIGMA MODEL ON SUPERGROUP MANIFOLD

If we construct the nonlinear sigma model on a supergroup manifold, this model satisfies the generalized Ricci flatness condition, more precisely $\tilde{R}^a{}_{bcd} = 0$ in this case, when the coefficient of the Wess–Zumino term is properly chosen as in the bosonic case.

A super Lie algebra is characterized by the following commutation relation[9]:

$$[\omega_1{}^i\lambda_i, \omega_2{}^j\lambda_j] = i\omega_1{}^i\omega_2{}^j f_{ij}{}^k\lambda_k, \tag{7}$$

where the ω^i are parameters. Here we asume $i+j+k=0$ (mod 2) for a physical application. In the following we use matrix representation of this algebra, so we regard λ_i as matrices which satisfy

$$\text{str}\,\lambda_i\lambda_j = \eta_{ij} \quad [i+j=0 \text{ (mod 2)}], \tag{8}$$

where str is supertrace and we assume the invertibility of η_{ij} for further construction of g_{ab}. The f_{ijk} defined by $f_{ij}{}^l\eta_{lk}$ is fully antisymmetric in i,j,k with an extra factor, i.e.,

$$f_{ijk} = -(-1)^{ij} f_{jik} = -(-1)^{jk} f_{ikj}. \tag{9}$$

For a given superalgebra, we can parametrize the group element as

$$U = \exp(i\theta S)\exp(ixT), \tag{10}$$

where S_i and T_i are the fermionic and bosonic generators, respectively. Hence Cartan's left invariant one-form is given using the inverse of η_{ij}, by

$$DZ^i = -i\,\text{str}(U^{-1}dU\lambda_j)\eta^{ij}. \tag{11}$$

[These one-forms are invariant under right transformations $\theta\to\theta'$, $x\to x'$, where

$$gU = \exp(i\theta' S)\exp(ix'T),$$

and are transformed as adjoint representations under left transformations $\theta\to\theta'$, $x\to x'$, where

$$Ug = \exp(i\theta' S)\exp(ix'T).]$$

This one-form satisfies the following Maurer–Cartan equation, which is important in the following discussions:

$$d(DZ^i) = \tfrac{1}{2} DZ^l DZ^m f_{lm}{}^i. \tag{12}$$

The vierbein $V_a{}^i$, which is defined by $DZ^i = V^i_a\, dZ^a$, satisfies

$$V^i{}_{a;b} = -\tfrac{1}{2}(-1)^{am} V^l{}_a V^m{}_b f_{lm}{}^i \tag{13}$$

as a result of the Maurer–Cartan equation and the symmetry of $\Gamma^a{}_{bc}$ and covariant constantness of ${}_a g_b$. (See the Appendix.)

The kinematic term and the Wess-Zumino term are constructed as

$$I_0 = -\int d^2x\,\text{str}(U^{-1}\partial_\mu U\, U^{-1}\partial_\mu U),$$

$$I_{\text{WZ}} = \tfrac{2}{3}\eta\int_{M^3}\text{str}(U^{-1}dU)^3, \tag{14}$$

where we omitted the overall coefficient of the action for simplicity. (The discussion about quantization of coefficient of the WZ term is found in Ref. 10.) The I_{WZ} can be written as

$$I_{\text{WZ}} = \tfrac{2}{3}\eta\int d^2x\,\epsilon^{\nu\mu}\partial_\mu Z^a\,{}_a e_b\,\partial_\nu Z^b, \tag{15}$$

where

$$e_{[ab,c]_s} = \tfrac{1}{2} f_{ijk} V^i{}_c V^j{}_b V^k{}_a (-1)^{b(k+a)+c(j+k+a+b)}. \tag{16}$$

After these, the Riemannian tensor $\tilde{R}^a{}_{bcd}$ is easily calculated as follows. The curvature two-form

$$\tilde{R}^i{}_j = \tfrac{1}{2} V^i \tilde{R}^a{}_{bcd}\, dZ^d dZ^c V^b{}_j \qquad (V^a{}_i \text{ is the inverse of } V^i{}_a), \tag{17}$$

is written as

$$\tilde{R}^i{}_j = d\omega^i{}_j - \omega^i{}_k\omega^k{}_j,$$
$$\omega^i{}_j = V^i{}_a \tilde{D} V^a{}_j. \tag{18}$$
$$\omega^i{}_j = \tfrac{1}{2}(1-\eta) DZ^m f_{mj}{}^i,$$
$$\tilde{R}^i{}_j = \tfrac{1}{4}(1-\eta^2) DZ^l DZ^k f_{lm}{}^i f_{kj}{}^m. \tag{19}$$

Therefore $\tilde{R}^a{}_{bcd} = 0$ when $\eta = \pm 1$.

An example in the case of $U(1/1)$ is given in Ref. 11. (The notations of Ref. 11 are a little different from those of this paper.) The Lagrangian is given by

$$\mathcal{L} = 2\partial_\mu x_1\{\partial_\mu x_2 - (i/4)(\vartheta_1\partial_\mu\vartheta_1 + \vartheta_2\partial_\mu\vartheta_2)\} + i\,\partial_\mu\vartheta_2\,\partial_\mu\vartheta_1 \pm (i/2)\epsilon_{\mu\nu}\,\partial_\mu x_1(\vartheta_1\partial_\nu\vartheta_1 + \vartheta_2\partial_\nu\vartheta_2). \tag{20}$$

We notice that this model has the WZ term although the bosonic part is a trivial $U(1)\oplus U(1)$.

IV. LIGHT-CONE CURRENT ALGEBRA

It has been shown that the above model is (one-loop) finite. Next we investigate the current algebra following Ref. 2.

It is easy to show that these models have conserved currents (light-cone currents)

$$\partial_- J_+ = 0, \quad J_+ = U^{-1}\partial_+ U, \qquad \text{when } \eta = 1, \tag{21}$$
$$\partial_+ J_- = 0, \quad J_- = \partial_- U\, U^{-1},$$

where $\tau = x^+ = (1/\sqrt{2})(x^0 + x^1)$, $\sigma = x^- = (1/\sqrt{2}) \times (x^0 - x')$. Algebras of these currents are constructed by the canonical quantization model. In the following we fix $\eta = 1$.

Our Lagrangian is written as

$$I = \int d\tau\, A_a\, \partial_+ \phi^a, \tag{22}$$

where A_a is independent of $\partial_+ \phi^a$. The (super-) Poisson bracket is defined as

$$\{A,B\}_{\mathrm{PB}} = F^{ab}\left(\frac{\vec{\partial} A}{\partial \phi^b}\right)\left(\frac{\vec{\partial} B}{\partial \phi^a}\right), \tag{23}$$

where F^{ab} is the inverse of

$${}_aF_b = \frac{\vec{\partial}}{\partial \phi^a} A_b - {}_aA \frac{\overleftarrow{\partial}}{\partial \phi^b}.$$

If we choose target space vectors $i\delta Z^i = \mathrm{str}(U^{-1}\, dU\, \lambda_i)\eta^{ij}$ as variables instead of the coordinates Z^a, then ${}_iF_j$ is easily obtained from a variation of action I,

$$\delta I = K \int d\tau\, d\sigma\, \mathrm{str}\left(U^{-1}\delta U \frac{d}{d\sigma}\left(U^{-1}\frac{d}{d\tau} U\right)\right), \tag{24}$$

comparing with

$$\delta I = \int d\tau\, \delta\phi^i\, {}_iF_j \frac{d\phi^j}{d\tau}, \tag{25}$$

where K is a normalization constant. This leads to

$$\begin{aligned} {}_iF_j &= K\eta_{ij} \otimes \frac{d}{d\sigma}, \\ F^{ij} &= (1/K)\eta^{ij} \otimes \vartheta(\sigma - \sigma^1). \end{aligned} \tag{26}$$

As a result, the Poisson bracket of

$$X = \mathrm{str}\left(A \frac{dU}{d\sigma} U^{-1}\right) \text{ and } Y = \mathrm{str}\left(B \frac{dU}{d\sigma} U^{-1}\right)$$

is

$$\begin{aligned} \{X,Y\}_{\mathrm{PB}} = &-\frac{1}{K}\delta(\sigma - \sigma')\mathrm{str}\{A,B\}\frac{dU}{d\sigma} U^{-1} \\ &-\frac{1}{K}\delta'(\sigma - \sigma')\mathrm{str}\, AB, \end{aligned} \tag{27}$$

where A and B are elements of the super-Lie algebra. Then the quantum mechanical commutation relations of $J_{-i} = K\,\mathrm{str}(\partial_- U\, U^{-1}\lambda_i)$ are

$$\begin{aligned} [J_{-i}(\sigma), J_{-j}(\sigma')\} = &if_{ij}{}^k J_{-k}(\sigma)\delta(\sigma - \sigma') \\ &+ iK\eta_{ij}\delta'(\sigma - \sigma'). \end{aligned} \tag{28}$$

The above current algebra is a superversion of the Kac–Moody algebra. We must notice that this "super" corresponds to the supersymmetry of target space and not to that of two dimensions.

V. CONCLUDING REMARKS

We have investigated the nonlinear sigma model on a supermanifold. The finiteness condition of the nonlinear sigma model leads to the equation $\tilde{R}^a{}_b = 0$, which corresponds to the Riemannian supergravity. The nonlinear sigma model on the supergroup manifold can satisfy this condition and have light-cone currents that realize the Kac–Moody algebra.

ACKNOWLEDGMENTS

I would like to thank Professor H. Miyazawa and Dr. T. Uematsu for useful discussions and the reading of this manuscript. I also thank Y. Matsuo for useful discussions.

APPENDIX: CONVENTIONS

De Witt has introduced four types of vectors,

$$V^a, V_a, {}^aV, {}_aV,$$

where V^a and V_a are contravariant and covariant vectors, respectively, and ${}^aV$ and ${}_aV$ are defined as ${}^aV = V^a$, ${}_aV = (-1)^a V_a$. We use conventional $(-1)^a$, that is, ± 1 corresponds to bosonic or fermionic properties of index a. The above formulation is free from troublesome factor $(-1)^a$ in many cases. The contraction is written as

$$V^a\, {}_ag_b\, {}^bV = V^a\, {}_aV = V_a\, {}^aV,$$

where ${}_ag_b$ is the metric tensor. There are two types of derivatives, right and left ones. We define right derivatives as

$$f_{,a} = f\frac{\overleftarrow{d}}{dz^a}.$$

And covariant derivatives are defined with Christoffel symbols $\Gamma^a{}_{bc}$ as

$$\begin{aligned} V_{a;b} &= V_{a,b} - V_c \Gamma^c{}_{ab}, \\ V^a{}_{;b} &= V^a{}_{,b} + (-1)^{c(a+1)} V^c \Gamma^a{}_{cb}. \end{aligned}$$

The Christoffel symbols $\Gamma^a{}_{bc}$ are defined to have the symmetry $\Gamma^a{}_{bc} = (-1)^{bc}\Gamma^a{}_{cb}$, and are written in consequence of this symmetry and covariant constantness of g_{ab} as follows:

$$\begin{aligned} \Gamma^a{}_{bc} = \tfrac{1}{2}(-1)^a g^{da}(&g_{ab,c} + (-1)^{bc} g_{ac,b} \\ &- (-1)^{a(c+b)} g_{bc,a}). \end{aligned}$$

Here, g^{da} is the inverse of metric ${}_ag_b$, i.e.,

$${}_ag_c g^{cb} = \delta^b{}_a, \quad g^{ac}\, {}_cg_b = \delta^a{}_b.$$

[1] J. Wess and B. Zumino, Phys. Lett. B 37, 95 (1971).
[2] E. Witten, Commun. Math. Phys. 92, 455 (1984).
[3] T. L. Curtright and C. K. Zachos, Phys. Rev. Lett. 53, 1799 (1984); P. Di Vecchia, V. G. Knizhnik, J. L. Petersen, and P. Rossi, Nucl. Phys. B 253, 701 (1985).
[4] B. De Witt, *Supermanifolds* (Cambridge U. P., Cambridge, 1984).
[5] A. Rogers, J. Math. Phys. 21, 1352 (1980).
[6] E. Braaten, T. L. Curtright, and C. K. Zachos, Nucl. Phys. B 260, 630 (1985).
[7] R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B 56, 81 (1975); P. Nath, in *Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (MIT, Cambridge, MA, 1976), p. 281.
[8] M. Henneaux and L. Mezinescu, Phys. Lett. B 152, 367 (1985); T. L. Curtright, L. Mezincescu, and C. K. Zachos, *ibid.* 161, 79 (1985).
[9] I. Bars, in *Applications in Group Theory in Physics and Mathmatical Physics* (Am. Math. Soc., Providence, RI, 1985).
[10] J. M. Rabin, preprint Enrico Fermi Institute, EFI 85-61, 1985.
[11] K. Kobayashi, Z. Phys. C 30, 643 (1986).

Systematics of arbitrary-helicity Lagrangian wave equations

Graham P. Collins[a] and Noel A. Doughty
Department of Physics, University of Canterbury, Christchurch 1, New Zealand

(Received 22 July 1986; accepted for publication 24 September 1986)

The nontrivially gauge-invariant Lagrangian field equations of arbitrary helicity are constructed directly by integration of the corresponding non-Lagrangian Poincaré field strengths based on the Lorentz irreps of unmixed spin. All the details of the gauge properties of the Lagrangian formulation of the free massless Poincaré irrep fields are derived systematically and uniformly for arbitrary spin from the very simple field-strength formulation as a direct consequence of an integration process. The gauge transformations, including the tracelessness of the gauge parameters, the form of the fermionic and bosonic potential field equations (first-, and second-order, respectively), and the vanishing double trace of the higher-spin ($\geqslant\frac{7}{2}$) potentials are all seen to be natural consequences of the integration of the classical equations for the field strengths without need for appeal to other requirements such as those of quantum field theory. Thus the equivalence (modulo gauge freedom and derivability from a Lagrangian) of the field strength and standard potential formulations of classical free fields are demonstrated explicitly, and uniformly for arbitrary helicity.

I. INTRODUCTION

The study of the Lagrangian formulation for arbitrary spin, in which all field equations and subsidiary conditions should be derived from an action principle, was started in 1939 by Fierz and Pauli.[1] Their classical approach began the now well-established "Fierz–Pauli program" for the setting up of interacting field theories free of certain algebraic pathologies such as the loss or gain of degrees of freedom on interaction. However, the technique was exceedingly difficult to extend beyond spin-2 even for the free fields and involved increasingly large numbers of auxiliary parameters. The well-known Lagrangian forms of the Poincaré fields of spin-1 (Maxwell or Proca[2]) and spin-2 (linearized Einstein[1,3] or massive gravity[4]) were first supplemented to include the spin-$\frac{3}{2}$ fields by Rarita and Schwinger[5] in 1941. The spin-$\frac{5}{2}$ Lagrangian was formulated by Kawakami and Kamefuchi[6] in 1967 while Chang,[7] also in 1967, extended the Lagrangian formulation of bosonic fields to spins-3, -$\frac{7}{2}$, and -4. The culmination of efforts to extend the formulation to arbitrarily high spins came with the work of Singh and Hagen[8] in 1974, who completed the massive formulation for both bosonic and fermionic fields. The corresponding massless Lagrangian fields for arbitrary spin j were then obtained by Fronsdal and Fang[9,10] in 1978. Many of these developments in the theory of higher-spin fields, especially those up to spin-2, are reviewed in the early chapters of Wiltshire.[11]

A further major step in the analysis of the massless fields was the analysis by de Wit and Freedman[12] in 1980 where the equations of higher-spin massless fields were very elegantly based on a hierarchy of generalized Christoffel symbols with simple gauge properties. The highest-order Christoffel symbol is a gauge-invariant generalized Weyl (vacuum Riemann) tensor which plays the role of a higher-spin field strength.

A feature common to many treatments of massless higher-spin theories has been the extensive *a priori* use made of gauge principles. Appeal is also made to other techniques and results that may ultimately have their basis in physical principles arising either from quantum field theory or from the theory of interacting fields rather than solely from classical free field theory. Some of these gauge properties, in particular, have been derived, *ab initio*, only for certain lower-spin cases. Quite often the form of the gauge transformations of the spin-j Lagrangian potentials are assumed at the outset by plausible generalization of these well-known lower-spin cases. These forms are then validated by demonstrating, with the use, for example, of projection operator techniques, or by counting components, that the equations to which they apply are indeed those of Poincaré irreps of spin-j, free of lower-spin contributions. In the process of demonstrating that this is the case, one may also demand, as is well known to be necessary for lower spin, that bosonic field equations be of second order and the fermionic of first order. These requirements are, of course, closely related to energy positivity and the existence of a positive-definite probability density as demanded by the eventual desire to quantize the field theory being constructed.

It is our contention that some of these assumptions and techniques are unnecessary in the context of classical free field theory and that the form of the gauge-invariant Lagrangian fields of arbitrary spin may be obtained, by an essentially algebraic analysis, from the much simpler features of the field-strength Poincaré irreps. For this to be so, the analysis should supply, as a consequence of the field strengths and the equations they satisfy, all the details of the gauge properties of the standard completely symmetric Lagrangian potentials. We shall denote the latter by $\varphi = (\varphi_{\mu_1\cdots\mu_n})$ for bosonic fields and $\psi = (\psi_{\mu_1\cdots\mu_n}) = (\psi_{\mu_1\cdots\mu_n}{}^{\alpha})$ for fermionic fields, where α is a Dirac index (which we shall almost always suppress).

To achieve this goal in a systematic, uniform manner we shall, in essence, reverse the technique of de Wit and Freedman,[12] Burgers,[13] and Berends, Burgers, and van Dam,[14]

[a] Present address: Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3800.

 0022-2488/87/020448-09$02.50 © 1987 American Institute of Physics

who begin with Lagrangian potentials (satisfying relatively involved field equations), from which they define field strengths in terms of derivatives. We show that the field strengths of arbitrary spin, based on unmixed-spin Lorentz irreps which satisfy such simple differential equations that they can be considered almost trivial, may be systematically integrated to obtain correctly constrained Lagrangian potentials satisfying the appropriate gauge-invariant field equations. The nontrivial gauge invariance of the Lagrangian potential formulation arises naturally from the arbitrariness necessarily introduced with each integration. Much of that arbitrariness can be removed by taking advantage of the evident freedom to select the higher-spin potentials to be in their standard symmetric and, for higher spin, zero double trace form. There is, in fact, no freedom of choice in the symmetry type of the potentials for spins-1 and -$\frac{3}{2}$. Various choices of symmetries are possible for spin-2 and -$\frac{5}{2}$ but these ultimately lead to equivalent potential formalisms. For spin $\geqslant 3$ the choice of symmetries made here may be necessary in order to apply the Poincaré Lemma (Appendix B) sufficiently often (see Secs. IV and V) to reach the Lagrangian potentials from the field strengths.

One of the effects of our analysis is thus to demonstrate the equivalence, for arbitrary spin, between the non-Lagrangian Poincaré formulation based on the unmixed Lorentz irreps, $(j,0)\oplus(0,j)$, and the corresponding Lagrangian formulations based on the mixed-spin Lorentz irreps $(j/2, j/2)$, of Fierz and Pauli,[1] and $((n+1)/2,\ n/2)\oplus(n/2,\ (n+1)/2)$, where $n=j-\frac{1}{2}$, of Rarita and Schwinger.[5] The principal results of this paper have been reported briefly in Doughty and Collins.[15]

We shall introduce the arbitrary-spin field strengths in Sec. II, outline the lower-spin potential results in general terms in Sec. III, and establish the arbitrary integer-spin result in Sec. IV. The minor modifications required for the half-odd-integer case are provided in Sec. V. Section VI discusses the possibility of arriving at potentials with nonstandard symmetries. Our main conventions are set out briefly in Appendix A. The Poincaré Lemma is described in Appendix B and three extensions to it are established in Appendices C, D, and E for use in Secs. IV–VI.

II. ARBITRARY-SPIN FIELD STRENGTHS

The symmetries of the arbitrary-helicity tensor and tensor–spinor field strengths are discussed by Weinberg[16] and Rodriguez and Lorente[17] in Dirac notation. Wiltshire[11] and Doughty and Wiltshire[18] used a Weyl spinor analysis to systematically set out these and other symmetries.

For any spin-j we begin, as in Collins[19] and Doughty and Collins,[20] with a totally symmetric $2(2j+1)$-component Dirac multispinor $\psi=(\psi_{\alpha_1\cdots\alpha_{2j}})$ (where $\alpha_i=1,2,3,4$), obeying the very simple massless Bargmann–Wigner equation,[11,18,21] $\partial\psi=0$, and the γ_5 condition appropriate to zero mass. These provide the unique, linear, parity-covariant, zero-mass Poincaré irrep equations of lowest possible order. Using the Bargmann–Wigner equation as a starting point has the enormous advantage of providing the same almost trivial structure for all nonzero spins. For spin-$\frac{1}{2}$ this equation is simply the massless Dirac equation. For spin $j>1$ we define,[19,20] from ψ, a field strength F by

$$F_{\mu_1\nu_1\cdots\mu_n\nu_n}=(\tfrac{1}{4})^n\,\mathrm{Tr}(\cdots\mathrm{Tr}(\psi C^{-1}\gamma_{\mu_1\nu_1})\cdots C^{-1}\gamma_{\mu_n\nu_n}), \tag{1a}$$

where C is the charge-conjugation matrix, $\gamma_{\mu\nu}=\frac{1}{2}i(\gamma_\mu\gamma_\nu-\gamma_\nu\gamma_\mu)$ is the generator of Lorentz transformations in the Dirac irrep, and F is a tensor for integer spin $j=n$ and a tensor–spinor with one free Dirac index (suppressed) for half-odd-integer spin $j=n+\frac{1}{2}$. Construction of the inverse relation

$$\psi=(\tfrac{1}{2})^n F_{\mu_1\nu_1\cdots\mu_n\nu_n}(\gamma^{\mu_1\nu_1}C\otimes\cdots\otimes\gamma^{\mu_n\nu_n}C), \tag{1b}$$

shows that F and ψ contain the same information. A Majorana ψ corresponds[19,20] to a real tensor or a Majorana tensor–spinor F, and $F_{\mu_1\nu_1\cdots\mu_n\nu_n}$ is equivalent to the corresponding quantity of Burgers[13] and Berends *et al.*[14] The following symmetries and field equations[11–20] are deducible from the complete symmetry on all indices of ψ and from the properties of the Bargmann–Wigner equation it satisfies:

$$F_{\mu_1\nu_1\cdots\mu_i\nu_i\cdots\mu_n\nu_n}=F_{\mu_1\nu_1\cdots[\mu_i\nu_i]\cdots\mu_n\nu_n}, \tag{2}$$

$$F_{\mu_1\nu_1\cdots\mu_i\nu_i\cdots\mu_k\nu_k\cdots\mu_n\nu_n}=F_{\mu_1\nu_1\cdots\mu_k\nu_k\cdots\mu_i\nu_i\cdots\mu_n\nu_n}, \tag{3}$$

$$F_{\mu_1\nu_1\cdots\mu_{n-1}[\nu_{n-1}\mu_n\nu_n]}=0, \tag{4}$$

$$F^{\lambda}{}_{\nu_1\lambda\nu_2\cdots\mu_n\nu_n}=0, \tag{5}$$

$$F_{\mu_1\nu_1\cdots\mu_{n-1}\nu_{n-1}[\mu_n\nu_n\lambda]}=0, \tag{6}$$

$$\partial^\lambda F_{\lambda\nu_1\mu_2\nu_2\cdots\mu_n\nu_n}=0, \tag{7}$$

$$\Box F_{\mu_1\nu_1\cdots\mu_n\nu_n}=0. \tag{8}$$

For the fermionic case, F also satisfies

$$\gamma^\lambda F_{\lambda\nu_1\cdots\mu_n\nu_n}=0,\quad \partial F_{\mu_1\nu_1\cdots\mu_n\nu_n}=0. \tag{9}$$

For both cases these equations have no gauge freedom and ensure[19,20] that F describes a massless Poincaré field of spin j. For $n=1$, Eq. (3), (4), and (5) are clearly vacuous and the rest of this section is then trivial. We define an associated field strength, $\hat{F}\equiv\Gamma^{(n)}$, by

$$\hat{F}^{\rho_1\cdots\rho_n}{}_{\mu_1\cdots\mu_n}=F^{\rho_1}{}_{(\mu_1}{}^{\rho_2\cdots}{}_{\mu_2\cdots}{}^{\rho_n}{}_{\mu_n)}. \tag{10}$$

[The index levels used in (10) were chosen to ease display of the index ordering and symmetrizing.] The field strength $\hat{F}$ corresponds to the generalized Riemann tensor or tensor–spinor of de Wit and Freedman[12] and has the following symmetries:

$$\hat{F}_{\rho_1\cdots\rho_n\mu_1\cdots\mu_n}=\hat{F}_{(\rho_1\cdots\rho_n)(\mu_1\cdots\mu_n)}=(-1)^n\hat{F}_{\mu_1\cdots\mu_n\rho_1\cdots\rho_n}, \tag{11}$$

$$\hat{F}_{\rho_1\cdots\rho_{n-1}(\rho_n\mu_1\cdots\mu_n)}=0, \tag{12}$$

$$\hat{F}^{\lambda}{}_{\lambda\rho_3\cdots\rho_n\mu_1\cdots\mu_n}=0=\hat{F}^{\lambda}{}_{\rho_2\cdots\rho_n\lambda\mu_2\cdots\mu_n}, \tag{13}$$

$$\gamma^\lambda\hat{F}_{\lambda\rho_2\cdots\rho_n\mu_1\cdots\mu_n}=0\quad\text{(fermionic case)}. \tag{14}$$

Furthermore, F may be recovered from $\hat{F}$ via

$$F^{\rho_1\mu_1}{}_{\rho_2\mu_2\cdots\rho_n\mu_n}=[2^n/(n+1)]\,\hat{F}^{\rho_1[\rho_2}{}_{\cdots[\rho_n\mu_n]}{}^{\cdots\mu_2]\mu_1}, \tag{15}$$

showing that F and $\hat{F}$ contain the same information. We

apply the Poincaré Lemma repeatedly to $\hat{F}$ to obtain the same hierarchy of generalized Christoffel symbols $\Gamma^{(i)}$ $(i = n-1, n-2, \ldots, 1)$ used by de Wit and Freedman.[12] Ultimately we demonstrate that, associated with $\hat{F}$, there exists a nontrivially gauge-invariant Lagrangian potential $\Gamma^{(0)}$ ($\equiv\varphi$ or ψ for the bosonic or fermionic cases, respectively) which obeys the standard arbitrary-spin, classical free field equation.

III. LOWER-SPIN POTENTIALS

The analysis is well known, and almost trivial, for the Maxwell field and was demonstrated for the massless spin-2 (linearized Einstein) field by, for example, Pirani.[22] Wiltshire[11] and Doughty and Wiltshire[18] show that the spin-$\frac{3}{2}$ gauge-invariant Lagrangian equations may be constructed directly by integration from the field strengths paralleling the spin-1 and -2 results. Collins[19] and Doughty and Collins[20] note that the integration method may easily be extended to spin-$\frac{5}{2}$ which has many features in common with the spin-2 case. The general method presented below in Secs. IV and V is essentially uniform for all values of spin, apart from some of the steps or conditions being vacuous in lower-spin cases corresponding to a lack of sufficient indices. There are also, of course, minor differences between the integer (bosonic) and half-odd-integer (fermionic) spin cases.

For spin-1, $F_{\mu\nu}$ is simply the Maxwell field strength and one of its field equations, $\partial_{[\mu}F_{\nu\lambda]} = 0$, constitutes an integrability condition on $F_{\mu\nu}$ so that the Poincaré Lemma (Appendix B) guarantees the existence of the electromagnetic potential A_μ satisfying $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. It arises directly from this integration that $F_{\mu\nu}$ is invariant under $\delta A_\mu = \partial_\mu \xi$ (with ξ an arbitrary scalar field) and hence that the first-order field equation, $\partial_\mu F^{\mu\nu} = 0$ (which becomes Maxwell's equation for the potential, $\Box A^\mu - \partial^\mu \partial\cdot A = 0$) is nontrivially gauge invariant.

For spin-$\frac{3}{2}$ the steps are almost identical. The equations, $\partial_{[\mu}F_{\nu\lambda]} = 0$, satisfied by the tensor–spinor field strength F imply the existence of a vector–spinor potential $\psi = (\psi_\mu)$ (it should not be confused with the multispinor ψ) satisfying $F_{\mu\nu} = \partial_\mu \psi_\nu - \partial_\nu \psi_\mu$, which is invariant under $\delta\psi_\mu = \partial_\mu \epsilon$ (with ϵ an arbitrary spin-$\frac{1}{2}$ Dirac spinor). The vanishing γ trace of the field strength, $\gamma^\mu F_{\mu\nu} = 0$, yields the simplest form, $\not{\partial}\psi_\mu - \partial_\mu \gamma\cdot\psi = 0$, of the gauge-invariant Rarita–Schwinger equation,[5,11] while $\partial_\mu F^{\mu\nu} = 0$ yields differential identities on the potentials also derivable directly from the field equation for the potential. It should be noted that these procedures are automatically and naturally supplying equations that are already of the standard type for each statistics, namely equations of second order for bosonic fields and of first order for the fermionic. Any second-order differential equation may, of course, be rewritten in first-order form by the introduction of auxiliary variables. However, we disregard such a recasting as it obscures rather than clarifies the structures involved.

For spin-2, the Poincaré Lemma must be applied twice to obtain[3,11,18–20,22] the symmetric massless Fierz–Pauli or linearized gravity potential $h_{\mu\nu} \equiv \varphi_{\mu\nu}$. The first application of the Poincaré Lemma yields a tensor $\Gamma_{\rho\mu\nu}$ which has a freedom (partly gauge) that is largely (but not entirely) fixed by selecting specific symmetries for $\Gamma_{\rho\mu\nu}$. There are several alternative ways to do this but each choice of symmetries is essentially equivalent in that each leads ultimately to the same potential $h_{\mu\nu}$, and the differences amount only to redefinitions of the intermediate $\Gamma_{\rho\mu\nu}$. After setting out the general arbitrary-spin derivation in Secs. IV and V, we comment briefly on these alternatives in Sec. VI. An algebraic equation obeyed by $F_{\lambda\rho\mu\nu}$ induces the necessary $\Gamma_{\rho\mu\nu}$ differential equation, which permits the Poincaré Lemma to be applied the second time, guaranteeing the existence of the potential $h_{\mu\nu}$. The residual arbitrariness in $\Gamma_{\rho\mu\nu}$ leaves $h_{\mu\nu}$ with the gauge freedom, $\delta h_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$. The massless Fierz–Pauli or linearized gravity field equations (in non-Lagrangian Ricci form), $\Box h_{\mu\nu} - 2\partial^\lambda \partial_{(\mu} h_{\nu)\lambda} + \partial_\mu \partial_\nu h = 0$ (where $h = h^\lambda{}_\lambda$), arise from the zero trace, $F^\lambda{}_{\mu\lambda\nu} = 0$, of the field strength. Combining this equation with a multiple of its own trace gives the Einstein Lagrangian form, which is suitable, by its identically divergence-free left-hand side, for coupling to a conserved source.[16]

The spin-$\frac{5}{2}$ steps are essentially the same as for spin-2 except that it is possible to fix $\Gamma_{\rho\mu\nu}$ so that it obeys $\gamma^\lambda \Gamma_{\lambda\mu\nu} = 0$ and this algebraic condition supplies[19,20] the desired first-order differential field equation for the potential $\psi_{\mu\nu}$. For spin-3, the Poincaré Lemma is applied to the field strength $F_{\rho_1\rho_2\rho_3\mu_1\mu_2\mu_3}$ three times, yielding the two intermediate tensors of de Wit and Freedman,[12] $\Gamma^{(2)}{}_{\rho_1\rho_2\mu_1\mu_2\mu_3}$, $\Gamma^{(1)}{}_{\rho\mu_1\mu_2\mu_3}$, and the completely symmetric Lagrangian potential $\varphi_{\mu_1\mu_2\mu_3}$ ($\equiv\Gamma^{(0)}$). The Lemma also allows one to impose the condition, $\Gamma^{(2)\lambda}{}_{\lambda\mu_1\mu_2\mu_3} = 0$, which yields the appropriate second-order field equation.[12] Indeed, for arbitrary spin$\geqslant\frac{5}{2}$, second-order bosonic field equations are supplied by the zero trace of $\Gamma^{(2)}$ and first-order fermionic field equations are supplied by the vanishing γ trace of $\Gamma^{(1)}$.

The only remaining new feature for spin$\geqslant\frac{7}{2}$ is the zero double-trace condition on the higher-spin potentials, namely $\gamma^\lambda \psi_\lambda{}^\rho{}_{\rho\mu_4\cdots\mu_n} = 0$ for fermionic fields and $\varphi^\lambda{}_\lambda{}^\rho{}_{\rho\mu_5\cdots\mu_j} = 0$ for bosonic fields. As will be seen in the following section, this is closely related to the Bianchi identity of de Wit and Freedman[12] also referred to as a source constraint (Burgers[13] and Berends *et al.*[14]).

IV. ARBITRARY INTEGER-SPIN POTENTIALS

All the lower-spin results summarized in the previous section are implicitly included in our derivation here of the arbitrary-spin case although, in practice, each of the cases with spin-1, -$\frac{3}{2}$, -2, and -$\frac{5}{2}$ can be treated individually somewhat more simply. We shall omit explicit mention of the spin$\leqslant\frac{3}{2}$ cases in order to avoid repeated qualification.

Equation (6) and the Poincaré Lemma ensure the existence of a tensor field A such that

$$F_{\rho_1\mu_1\cdots\rho_{n-1}\mu_{n-1}\rho_n\mu_n} = A_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}[\mu_n,\rho_n]}. \tag{16}$$

This A is free to be varied according to

$$\delta A_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}\mu_n} = B_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1},\mu_n}, \tag{17}$$

with B arbitrary, without affecting the value of F. By (5) we could take A to be traceless across any pair of indices in $\{\rho_1,\ldots,\rho_{n-1},\mu_1,\ldots,\mu_{n-1}\}$. However, since it will turn out

to be possible to ensure zero trace only on the ρ indices of the generalized Christoffel tensor, $\Gamma^{(n-1)}$, this is all that we shall impose on A, namely

$$\eta^{\rho_i\rho_j}A_{\rho_1\cdots\rho_i\cdots\rho_j\cdots\rho_{n-1}\mu_1\cdots\mu_n}=0 \quad (n\geqslant 3). \tag{18}$$

This condition partially fixes the freedom (17). The tensor field A may still be freely varied according to (17) with B restricted to be traceless on its ρ indices. We define

$$C_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_n}=A_{(\rho_1\cdots\rho_{n-1})(\mu_1\cdots\mu_{n-1})\mu_n}. \tag{19}$$

Then by (10)

$$\hat{F}^{\rho_1\cdots\rho_{n-1}\rho_n}{}_{\mu_1\cdots\mu_{n-1}\mu_n}=C^{\rho_1\cdots\rho_{n-1}}{}_{(\mu_1\cdots\mu_{n-1}\mu_n)}{}^{,\rho_n}-C^{\rho_1\cdots\rho_{n-1}}{}_{(\mu_1\cdots\mu_{n-1}}{}^{,\rho_n}{}_{,\mu_n)}, \tag{20}$$

and

$$\delta C_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}\mu_n}=B_{(\rho_1\cdots\rho_{n-1})(\mu_1\cdots\mu_{n-1})\mu_n}, \tag{21}$$

is allowed. Thus C, like A, is traceless across any pair of ρ indices

$$C^{\lambda}{}_{\lambda\rho_3\cdots\rho_{n-1}\mu_1\cdots\mu_n}=0 \quad (n\geqslant 3), \tag{22}$$

and B in (21) must be similarly restricted to preserve this condition. Equation (4) implies

$$A_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-2}[\mu_{n-1}\mu_n,\rho_n]}=0, \tag{23}$$

and similarly

$$A^{\rho_1\cdots\rho_{n-1}\mu_1\cdots}{}_{[\mu_i}{}^{\cdots\mu_{n-1}}{}_{\mu_n,\rho_n]}=0 \quad (i=1,\ldots,n-1). \tag{24}$$

Hence

$$C_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-2}[\mu_{n-1}\mu_n,\lambda]}=0. \tag{25}$$

Results (19) and (25) allow an extension of the Poincaré Lemma (PLX1, Appendix C) to be applied, yielding the existence of a tensor B, as in (21), such that

$$\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}\mu_n}\equiv C_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}\mu_n}-B_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1},\mu_n} \tag{26}$$

satisfies

$$\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_n}=\Gamma^{(n-1)}_{(\rho_1\cdots\rho_{n-1})(\mu_1\cdots\mu_n)}, \tag{27}$$

and hence

$$\hat{F}_{\rho_1\cdots\rho_{n-1}\rho_n\mu_1\cdots\mu_{n-1}\mu_n}=\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}\mu_n,\rho_n}-\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\rho_n(\mu_1\cdots\mu_{n-1},\mu_n)}. \tag{28}$$

PLX1 allows B to be chosen to be traceless on its ρ indices. Hence one may restrict $\Gamma^{(n-1)}$ to satisfy

$$\Gamma^{(n-1)\lambda}{}_{\lambda\rho_3\cdots\rho_{n-1}\mu_1\cdots\mu_n}=0 \quad (n\geqslant 3). \tag{29}$$

Even if C were traceless across any pair in $\{\rho_1,\cdots\rho_{n-1},\mu_1,\cdots\mu_{n-1}\}$, Eq. (29) would still be the maximum restriction that could be imposed on $\Gamma^{(n-1)}$. The restrictions (27) and (29) eliminate all those freedoms in $\Gamma^{(n-1)}$ that may be fixed solely by choosing the algebraic symmetry of $\Gamma^{(n-1)}$.

Nevertheless, Eq. (22) does not exhaust all the freedom of (17). Consider the variation

$$\delta\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_n}=Y_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1},\mu_n}. \tag{30}$$

To preserve the symmetries of $\Gamma^{(n-1)}$ the tensor Y can, without loss of generality, be restricted to obey

$$Y_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}}=Y_{(\rho_1\cdots\rho_{n-1})(\mu_1\cdots\mu_{n-1})} \tag{31}$$

$$Y^{\lambda}{}_{\lambda\rho_3\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}}=0 \quad (n\geqslant 3), \tag{32}$$

and, furthermore, Y must obey

$$Y_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1},\mu_n}=Y_{\rho_1\cdots\rho_{n-1}(\mu_1\cdots\mu_{n-1},\mu_n)}. \tag{33}$$

This implies

$$Y_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-2}[\mu_{n-1},\mu_n]}=0, \tag{34}$$

and hence, by our second extension to the Poincaré Lemma (PLX2, Appendix D), there exists a field ξ such that

$$Y_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_{n-1}}=\xi_{\rho_1\cdots\rho_{n-1},\mu_1\cdots\mu_{n-1}}, \tag{35}$$

$$\xi_{\rho_1\cdots\rho_{n-1}}=\xi_{(\rho_1\cdots\rho_{n-1})}, \tag{36}$$

$$\xi^{\lambda}{}_{\lambda\rho_3\cdots\rho_{n-1}}=0 \quad (n\geqslant 3). \tag{37}$$

In other words, $\Gamma^{(n-1)}$ has the residual gauge freedom

$$\delta\Gamma^{(n-1)}_{\rho_1\cdots\rho_{n-1}\mu_1\cdots\mu_n}=\xi_{\rho_1\cdots\rho_{n-1},\mu_1\cdots\mu_n}, \tag{38}$$

with ξ satisfying (36) and (37). This freedom is the most general variation of $\Gamma^{(n-1)}$ that maintains the symmetries of $\Gamma^{(n-1)}$ and the value of $\hat{F}$ but it is insufficient to impose further covariant symmetries on $\Gamma^{(n-1)}$. In fact, (38) is the gauge freedom of $\Gamma^{(n-1)}$ corresponding to the usual gauge freedom[12] that will be deduced below for the spin n potential which corresponds to it.

A further $n-1$ essentially identical integrations with the Poincaré Lemma can now be used to recreate the hierarchy of generalized Christoffel symbols and potential, $\Gamma^{(n-2)}\cdots\Gamma^{(1)}$, $\Gamma^{(0)}\equiv\varphi$, of de Wit and Freedman.[12] These can be carried out by induction using the above derivation of $\Gamma^{(n-1)}$ from $\Gamma^{(n)}\equiv\hat{F}$ as the initial step. Given $m\geqslant 1$, we suppose that

$$\Gamma^{(m+1)}_{\rho_1\cdots\rho_{m+1}\mu_1\cdots\mu_n}=\Gamma^{(m+1)}_{(\rho_1\cdots\rho_{m+1})(\mu_1\cdots\mu_n)}, \tag{39}$$

$$\Gamma^{(m)}_{\rho_1\cdots\rho_m\mu_1\cdots\mu_n}=\Gamma^{(m)}_{(\rho_1\cdots\rho_m)(\mu_1\cdots\mu_n)}, \tag{40}$$

$$\Gamma^{(m)\lambda}{}_{\lambda\rho_3\cdots\rho_m\mu_1\cdots\mu_n}=0 \quad (m\geqslant 2), \tag{41}$$

$$\Gamma^{(m+1)}_{\rho_1\cdots\rho_m\rho_{m+1}\mu_1\cdots\mu_{n-1}\mu_n}=\Gamma^{(m)}_{\rho_1\cdots\rho_m\mu_1\cdots\mu_{n-1}\mu_n,\rho_{m+1}}-[n/(m+1)]\,\Gamma^{(m)}_{\rho_1\cdots\rho_m\rho_{m+1}(\mu_1\cdots\mu_{n-1},\mu_n)}, \tag{42}$$

and the freedom

$$\delta\Gamma^{(m)}_{\rho_1\cdots\rho_m\mu_1\cdots\mu_n}=a_m\,\xi_{\rho_1\cdots\rho_m(\mu_{m+2}\cdots\mu_n,\mu_1\cdots\mu_{m+1})}, \tag{43}$$

where $a_m=(-)^{n-m-1}\ (n-1)!/(n-m-1)!m!$. We then deduce the existence of $\Gamma^{(m-1)}$ satisfying the corresponding relations for $m-1$.

Equations (39), (40), and (42) imply

$$0=\Gamma^{(m+1)}_{\rho_1\cdots\rho_{m-1}[\rho_m\rho_{m+1}\mu_1]}{}^{\mu_2\cdots\mu_n}=\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1}{}^{\mu_2\cdots\mu_n}{}_{,\rho_{m+1}]}-[1/(m+1)]\,\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\rho_{m+1}}{}^{\mu_2\cdots\mu_n}{}_{,\mu_1]}, \tag{44}$$

and therefore

$$\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1}{}^{\mu_2\cdots\mu_n}{}_{,\rho_{m+1}]}=0. \tag{45}$$

Hence by the Poincaré Lemma,

$$\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1]\mu_2\cdots\mu_n} = G_{\rho_1\cdots\rho_{m-1}\mu_2\cdots\mu_n[\mu_1,\rho_m]}, \quad (46)$$

for some tensor G such that

$$G_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = G_{(\rho_1\cdots\rho_{m-1})(\mu_1\cdots\mu_{n-1})\mu_n}, \quad (47)$$

$$G^{\lambda}{}_{\lambda\rho_3\cdots\rho_{m-1}\mu_1\cdots\mu_n} = 0 \quad (m-1\geqslant 2). \quad (48)$$

Furthermore,

$$\delta G_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = H_{(\rho_1\cdots\rho_{m-1})(\mu_1\cdots\mu_{n-1}),\mu_n}, \quad (49)$$

with

$$H^{\lambda}{}_{\lambda\rho_3\cdots\rho_{m-1}\mu_1\cdots\mu_{n-1}} = 0 \quad (m-1\geqslant 2), \quad (50)$$

leaves $\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1]\mu_2\cdots\mu_n}$ invariant. Equations (40) and (46) imply

$$G_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_{n-2}[\mu_{n-1}\mu_n\lambda]} = 0. \quad (51)$$

This equation with (47), (48), and PLX1, imply the existence of a tensor H, as in Eqs. (49) and (50), so that $\hat{G}$ defined by

$$\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} \equiv G_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} - H_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_{n-1},\mu_n}, \quad (52)$$

satisfies

$$\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = \hat{G}_{(\rho_1\cdots\rho_{m-1})(\mu_1\cdots\mu_n)}, \quad (53)$$

$$\hat{G}^{\lambda}{}_{\lambda\rho_3\cdots\rho_{m-1}\mu_1\cdots\mu_n} = 0 \quad (m-1\geqslant 2), \quad (54)$$

$$\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1]\mu_2\cdots\mu_n} = \hat{G}_{\rho_1\cdots\rho_{m-1}\mu_2\cdots\mu_n[\mu_1,\rho_m]}. \quad (55)$$

It is not necessarily the case that

$$\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = G_{(\rho_1\cdots\rho_{m-1})(\mu_1\cdots\mu_n)}$$

or even that the two are proportional. The maximal residual freedom of $\hat{G}$ which preserves the symmetries of $\hat{G}$ and leaves $\Gamma^{(m)}$ invariant can be deduced exactly as for the sequence (30)–(38) and has the form

$$\delta\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = P_{\rho_1\cdots\rho_{m-1},\mu_1\cdots\mu_n}, \quad (56)$$

where

$$P_{\rho_1\cdots\rho_{m-1}} = P_{(\rho_1\cdots\rho_{m-1})}, \quad (57)$$

$$P^{\lambda}{}_{\lambda\rho_3\cdots\rho_{m-1}} = 0 \quad (m-1\geqslant 2). \quad (58)$$

On the other hand, (43) allows a freedom $\delta\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n}$ proportional to $\xi_{\rho_1\cdots\rho_{m-1}(\mu_{m+1}\cdots\mu_n,\mu_1\cdots\mu_m)}$. This causes $\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1]\mu_2\cdots\mu_n}$ to vary but only within the freedom (43) which leaves $\hat{F}$ invariant. Symmetry (39) and equations (42) and (55) imply

$$\begin{aligned} 0 &= \Gamma^{(m+1)\rho_1\cdots\rho_{m-1}}{}_{[\rho_m\rho_{m+1}]}{}^{\mu_1\cdots\mu_n} \\ &= \Gamma^{(m)\rho_1\cdots\rho_{m-1}}{}_{[\rho_m}{}^{\mu_1\cdots\mu_n}{}_{,\rho_{m+1}]} \\ &\quad - [n/(m+1)]\,\Gamma^{(m)\rho_1\cdots\rho_{m-1}}{}_{[\rho_m\rho_{m+1}]}{}^{(\mu_1\cdots\mu_{n-1},\mu_n)} \\ &= \Gamma^{(m)\rho_1\cdots\rho_{m-1}}{}_{[\rho_m}{}^{\mu_1\cdots\mu_n}{}_{,\rho_{m+1}]} \\ &\quad + [n/(m+1)]\,\hat{G}^{\rho_1\cdots\rho_{m-1}(\mu_1\cdots\mu_{n-1}}{}_{[\rho_m,\rho_{m+1}]}{}^{,\mu_n)}. \end{aligned} \quad (59)$$

Therefore the Poincaré Lemma implies that there exists a tensor J such that

$$\Gamma^{(m)}_{\rho_1\cdots\rho_m\mu_1\cdots\mu_n} + [n/(m+1)]\,\hat{G}_{\rho_1\cdots\rho_m(\mu_1\cdots\mu_{n-1},\mu_n)} = J_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n,\rho_m}, \quad (60)$$

$$J_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = J_{(\rho_1\cdots\rho_{m-1})(\mu_1\cdots\mu_n)}, \quad (61)$$

$$J^{\lambda}{}_{\lambda\rho_3\cdots\rho_{m-1}\mu_1\cdots\mu_n} = 0 \quad (m-1\geqslant 2). \quad (62)$$

Equations (59), (53), and (55) imply

$$\begin{aligned} J^{\rho_1\cdots\rho_{m-1}\mu_2\cdots\mu_n}{}_{[\mu_1,\rho_m]} &= \Gamma^{(m)\rho_1\cdots\rho_{m-1}}{}_{[\rho_m\mu_1]}{}^{\mu_2\cdots\mu_n} \\ &\quad + [1/(m+1)]\,\hat{G}^{\rho_1\cdots\rho_{m-1}}{}_{[\rho_m}{}^{\mu_2\cdots\mu_n,\mu_1]} \\ &= [m/(m+1)]\,\hat{G}^{\rho_1\cdots\rho_{m-1}\mu_2\cdots\mu_n}{}_{[\mu_1,\rho_m]}. \end{aligned} \quad (63)$$

Hence $J_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} - [m/(m+1)]\,\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n}$ obeys all the conditions of PLX2 which ensures the existence of a tensor P, as in (57) and (58), satisfying

$$J_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} - [m/(m+1)]\,\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = P_{\rho_1\cdots\rho_{m-1},\mu_1\cdots\mu_n}. \quad (64)$$

We use freedom (56) and rescale to define

$$\Gamma^{(m-1)}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = [m/(m+1)]\,\hat{G}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} - [m/(m+n)]\,P_{\rho_1\cdots\rho_{m-1},\mu_1\cdots\mu_m}, \quad (65)$$

or, equivalently,

$$\Gamma^{(m-1)}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = J_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} - [n/(m+n)]\,P_{\rho_1\cdots\rho_{m-1},\mu_1\cdots\mu_n}. \quad (66)$$

Clearly, $\Gamma^{(m-1)}$ satisfies (40) and (41) for $m-1$. Furthermore, by (52)

$$\Gamma^{(m)}_{\rho_1\cdots\rho_{m-1}[\rho_m\mu_1]\mu_2\cdots\mu_n} = [(m+1)/m]\,\Gamma^{(m-1)}_{\rho_1\cdots\rho_{m-1}\mu_2\cdots\mu_n[\mu_1,\rho_m]}, \quad (67)$$

and by (65), (66), and (60) one can readily verify (42) for $m-1$. Varying $\Gamma^{(m-1)}$ according to

$$\delta\Gamma^{(m-1)}_{\rho_1\cdots\rho_{m-1}\mu_1\cdots\mu_n} = a_{m-1}\,\xi_{\rho_1\cdots\rho_{m-1}(\mu_{m+1}\cdots\mu_n,\mu_1\cdots\mu_m)}, \quad (68)$$

varies $\Gamma^{(m)}$ in accordance with (43) and leaves $\hat{F}$ invariant.

The final result of carrying out all the above steps, for $m=n-1$ down to $m=1$, is to establish the existence of the completely symmetric potential $\Gamma^{(0)}{}_{\mu_1\cdots\mu_n} \equiv \varphi_{\mu_1\cdots\mu_n}$. The trace condition $W_{\mu_1\cdots\mu_n} \equiv \Gamma^{(2)\lambda}{}_{\lambda\mu_1\cdots\mu_n} = 0$, constitutes the field equation[12] for φ

$$\Box\varphi_{\mu_1\cdots\mu_n} - n\partial^{\lambda}\partial_{(\mu_1}\varphi_{\mu_2\cdots\mu_n)\lambda} + [n(n-1)/2]\,\partial_{(\mu_1}\partial_{\mu_2}\varphi_{\mu_3\cdots\mu_n)}{}^{\lambda}{}_{\lambda} = 0. \quad (69)$$

This equation is not itself derivable from a Lagrangian[12] but combined with a multiple of its own trace gives the equivalent standard bosonic spin n Lagrangian equation.[12–14]

The source constraint of Burgers[13] and Berends, Burgers and van Dam[14] is, for spin $\geqslant 3$, precisely the zero double trace of $\Gamma^{(3)}$, namely

$$\Gamma^{(3)\lambda}{}_{\lambda}{}^{\rho}{}_{\rho\mu_2\cdots\mu_n} = 0, \quad (70)$$

which implies

$$\partial^{\lambda} W_{\lambda\mu_2\cdots\mu_n} - [(n-1)/2]\, W^{\lambda}{}_{\lambda(\mu_2\cdots\mu_{n-1},\mu_n)} = 0, \tag{71}$$

where W is the left-hand side of the field equation. Expanding (71) in terms of the potential $\varphi_{\mu_1\cdots\mu_n}$ yields, for $n\geqslant 4$,

$$\varphi^{\lambda}{}_{\lambda}{}^{\rho}{}_{\rho(\mu_5\cdots\mu_n,\mu_2\mu_3\mu_4)} = 0, \tag{72}$$

and three successive applications of our third extension of the Poincaré Lemma (PLX3, Appendix E) yield the double-traceless condition[12] for the potential $\varphi_{\mu_1\cdots\mu_n}$, namely

$$\varphi^{\lambda}{}_{\lambda}{}^{\rho}{}_{\rho\mu_5\cdots\mu_n} = 0 \quad (j=n\geqslant 4). \tag{73}$$

We note that the requirement (E3) of Appendix E is supplied by the vanishing of our physical fields at infinity. With the potential obeying the symmetry (73) the source constraint is identically satisfied. The vanishing divergence of the trace of $\hat{F}$ for spin-2 and the antisymmetry of $\hat{F}\equiv F$ for spin-1 yield the lower-spin source constraints.

V. ARBITRARY HALF-ODD-INTEGER SPIN

The half-odd-integer spin analysis closely resembles the integer spin case. The hierarchy of generalized tensor–spinor Christoffel symbols of de Wit and Freedman[12] can be recreated exactly as for integer spin. However, in addition to the traceless condition (41) for $\Gamma^{(m)}$ $(m\geqslant 2)$ one can impose, by (14), a γ-traceless condition:

$$\gamma^{\lambda}\,\Gamma^{(m)}_{\lambda\rho_2\cdots\rho_m\mu_1\cdots\mu_n} = 0 \quad (m\geqslant 1). \tag{74}$$

The $\Gamma^{(m)}$ have a gauge freedom of the same form as (43) with the tensor ξ replaced by a tensor–spinor, which we denote by ϵ, obeying $\gamma^{\lambda}\epsilon_{\lambda\rho_2\cdots\rho_{n-1}} = 0$ $(j=n+\frac{1}{2}\geqslant\frac{5}{2})$. This condition, (74), supplies the appropriate first-order field equation for the potential ψ, namely

$$\partial\!\!\!/\,\psi_{\mu_1\cdots\mu_n} - n\gamma^{\lambda}\psi_{\lambda(\mu_2\cdots\mu_n,\mu_1)} = 0. \tag{75}$$

Again, this is equivalent, though not identical, to the standard Lagrangian field equation.[12–14]

For spin $j=n+\frac{1}{2}\geqslant\frac{7}{2}$ the combination

$$\tfrac{4}{3}\gamma^{\lambda}\,\Gamma^{(2)}{}_{\lambda}{}^{\rho}{}_{\rho\mu_2\cdots\mu_n} + \tfrac{2}{3}\gamma^{\kappa}\gamma^{\rho}\gamma^{\lambda}\,\Gamma^{(2)}_{\lambda\rho\kappa\mu_2\cdots\mu_n} = 0, \tag{76}$$

yields the appropriate source-constraint equation and also

$$\gamma^{\rho}\,\psi_{\rho}{}^{\lambda}{}_{\lambda(\mu_4\cdots\mu_n,\mu_2\mu_3)} = 0, \tag{77}$$

and hence, via PLX3, the zero double trace of $\psi_{\mu_1\cdots\mu_n}$

$$\gamma^{\rho}\,\psi_{\rho}{}^{\lambda}{}_{\lambda\mu_4\cdots\mu_n} = 0 \quad (j=n+\tfrac{1}{2}\geqslant\tfrac{7}{2}). \tag{78}$$

For spin-$\frac{3}{2}$ and -$\frac{5}{2}$ the source constraint arises from the equations satisfied by $\hat{F}$.

It is worth emphasizing that, for all spins, the sets of equations obeyed by the potentials φ or ψ are equivalent to the sets of equations obeyed by the field strengths. No equations have been especially selected or discarded in order to obtain the standard Lagrangian field equations and source constraints for arbitrary spin.

VI. ALTERNATIVE SYMMETRIES FOR THE LAGRANGIAN POTENTIALS

For spins-1 and -$\frac{3}{2}$ the integration from F to φ or ψ is completed in a single step and all the freedom arising from the integration corresponds to the usual gauge invariance of φ or ψ. For spins $\geqslant 2$, most of the freedom arising from each integration was used to fix the symmetries of the fields $\Gamma^{(m)}$ $(m=0,\ldots,n-1)$ in a conventional manner. Clearly, it is desirable that alternative choices of potential symmetries lead to equivalent dynamical equations. We will demonstrate here that this is the case for spins-2 and -$\frac{5}{2}$. Another desirable possibility would be for only one choice of symmetries to allow the series of integrations to be completed. This may be the case for spin $\geqslant 3$ though it seems likely that alternative integration schemes exist for higher spin as well.

For spin-2 we work directly with the original field strength $F_{\rho_1\mu_1\rho_2\mu_2}$. As before, there must exist a field A such that

$$F_{\rho\mu\lambda\nu} = A_{\rho\mu[\nu,\lambda]}. \tag{79}$$

However, this time we use the antisymmetry of F to directly impose

$$A_{\rho\mu\nu} = A_{[\rho\mu]\nu}. \tag{80}$$

These equations are preserved if A is varied by an arbitrary antisymmetric field B

$$\delta A_{\rho\mu\nu} = B_{\rho\mu,\nu}, \quad B_{\rho\mu} = B_{[\rho\mu]}. \tag{81}$$

Equation (4) implies $A_{\rho[\mu\nu,\lambda]} = 0$ and hence there exists a field C such that

$$A_{\rho[\mu\nu]} = C_{\rho[\mu,\nu]}. \tag{82}$$

By (80) we can write

$$A_{\rho\mu\nu} = C_{\rho[\mu,\nu]} - C_{\mu[\rho,\nu]} - C_{\nu[\rho,\mu]}. \tag{83}$$

Varying C according to $\delta C_{\mu\nu} = D_{\mu,\nu}$ with D arbitrary leaves A invariant. By splitting C into its antisymmetric and symmetric parts

$$C_{\mu\nu} = E_{\mu\nu} + G_{\mu\nu} = E_{[\mu\nu]} + G_{(\mu\nu)}, \tag{84}$$

we find that

$$A_{\rho\mu\nu} = 2G_{\nu[\mu,\rho]} + E_{\rho\mu,\nu}. \tag{85}$$

Hence we may use (81) to redefine

$$\hat{A}_{\rho\mu\nu} = A_{\rho\mu\nu} - E_{\rho\mu,\nu}. \tag{86}$$

All of the equations (79)–(83) hold for $\hat{A}$ replacing A and, in addition, one has

$$\hat{A}_{\rho\mu\nu} = 2G_{\nu[\mu,\rho]}, \tag{87}$$

with G symmetric.

The only variation of G which leaves $\hat{A}$ fixed is $\delta G_{\mu\nu} = H_{\mu,\nu}$ and to preserve the symmetry of G, PLX2 allows this to be expressed as $\delta G_{\mu\nu} = J_{,\mu\nu}$ where J is an arbitrary scalar. However, this variation can be incorporated into the more general variation

$$\delta G_{\mu\nu} = K_{(\mu,\nu)}, \tag{88}$$

which varies $\hat{A}$ according to $\delta\hat{A}_{\rho\mu\nu} = K_{[\mu,\rho],\nu}$, which is consistent with (81) and leaves F invariant.

The relation between F and $\frac{1}{2}G$ is precisely that between F and φ obtained by the formalism of Sec. IV. Therefore the field equations for φ and G are identical and (88) is just the appropriate spin-2 gauge invariance. Identifying $\frac{1}{2}G\equiv\varphi$ we find that A is just a (scaled) redefinition of $\Gamma^{(1)}$

$$\hat{A}_{\rho\mu\nu} = 2(\Gamma^{(1)}_{\rho\mu\nu} + \varphi_{\rho\mu,\nu}). \tag{89}$$

For spin-$\frac{5}{2}$ there is a similar alternative. However, the derivation of the field equation is slightly less straightforward. For spin-$\frac{5}{2}$, the tensors $A, B, \ldots, K$ become tensor–spinors and, in addition to Eqs. (79)–(88), we may impose

$$\gamma^\lambda A_{\lambda\mu\nu} = 0, \quad \gamma^\lambda B_{\lambda\mu} = 0, \quad \gamma^\lambda C_{\lambda\mu} = 0, \quad \gamma^\lambda D_\lambda = 0. \tag{90}$$

We cannot require that both G and E be γ traceless but we do have

$$\gamma^\lambda \hat{A}_{\lambda\mu\nu} = \gamma^\lambda A_{\lambda\mu\nu} - \gamma^\lambda E_{\lambda\mu,\nu} = \gamma^\lambda G_{\lambda\mu,\nu}, \tag{91}$$

which implies

$$2\gamma^\lambda G_{\nu[\mu,\lambda]} - \gamma^\lambda G_{\lambda\mu,\nu} = 0, \tag{92}$$

and hence the field equation for G is the same as that for the spin-$\frac{5}{2}$ field ψ. Again, $\frac{1}{2}G$ can be identified with ψ. The variation $\delta G_{\mu\nu} = \epsilon_{(\mu,\nu)}$ with $\gamma^\lambda \epsilon_\lambda = 0$ preserves the value of F and the symmetries of A.

It is probable that similar alternative methods can be applied to spin $\geqslant 3$. In each case we expect that alternative choices of symmetries will correspond to redefinitions of the intermediate fields so that the physically relevant quantities, the trivially invariant F and the nontrivially invariant φ or ψ, are unaltered.

VII. CONCLUSION

Berends *et al.*[14] show that interacting quantum field theories involving fields of helicity >2 encounter consistency problems. Berends *et al.* also note, however, that consistent higher-spin interaction may be possible with the inclusion of an infinite number of participating higher-spin fields.

The current importance of superstring theories[23] in unification studies of all the known interactions, including gravity, is closely related to the consistency with which they may be formulated. The close relationship between string states and those of an infinite sequence of particles of various spins points to the continued importance of higher-spin studies.

However, analyses of arbitrary spin fields can be exceedingly complex, if only because of the notational difficulties involved. This can be especially true when different higher-spin cases are considered individually and without full use of the systematic relationships that must tie together fields which are all irreps of the Poincaré group differing only in their irrep labels. Indeed, failure to capitalize on the systematic relationships can easily lead to misunderstandings of the origins of specific properties of higher-spin fields, especially in the highly nontrivial Lagrangian potential formulation which is routinely used for consideration of their mutual interactions.

In this paper, we have further developed the systematic study of arbitrary helicity fields by deriving the elegant relationships of de Wit and Freedman[12] from the exceedingly simple multispinor field strengths satisfying the equally simple Bargmann–Wigner equations. We have shown, by a direct method, which is, however, unlikely to be the most elegant available, that these simple equations may be systematically integrated, uniformly for arbitrary spin, to obtain all the properties of the Lagrangian potentials, their field equations, gauge freedoms, and source constraints (or Bianchi identities) without recourse to assumptions besides those of relativistic classical free fields.

This derivation of the Lagrangian formulation for arbitrary helicity explicitly confirms the equivalence (modulo gauge freedom and derivability from an action principle) of the massless unmixed spin irreps and the nontrivially gauge-invariant Lagrangian irreps of the Poincaré group.

ACKNOWLEDGMENTS

We are grateful for helpful suggestions from David Wiltshire, Peter Jarvis, Richard Farmer, and John Strathdee. The preparation of this manuscript was greatly facilitated by the use of the T^3 scientific word processing software of TCI Software, Inc.

APPENDIX A: CONVENTIONS

Dirac algebra:	$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}$,
Minknowski metric:	$\eta = (\eta_{\mu\nu}) = \text{diag}(+1, -1, -1, -1)$,
partial derivatives:	$A_{\mu,\nu} = \partial_\nu A_\mu, \quad \not{\partial} = \gamma^\mu \partial_\mu$,
complete symmetrization:	$T_{(\mu_1\cdots\mu_m)} = \frac{1}{m!}\sum_\pi T_{\mu_{\pi_1}\cdots\mu_{\pi_m}}$,
complete antisymmetrization:	$T_{[\mu_1\cdots\mu_m]} = \frac{1}{m!}\sum_\pi (-1)^\pi T_{\mu_{\pi_1}\cdots\mu_{\pi_m}}$,

where each sum is taken over all permutations, $\pi = (\pi_1\cdots\pi_m)$, of the numbers $1,\ldots,m$ and $(-1)^\pi$ is $+1$ or -1 for π even and odd, respectively.

APPENDIX B: POINCARÉ LEMMA

Completely antisymmetric tensors of rank n are essentially equivalent to n-forms and a very concise expression of the Poincaré Lemma makes use of the language of differential forms (Misner, Thorne, and Wheeler,[3] and Choquet-Bruhat, deWitt-Morette, and Dillard-Bleick[24]). For the purposes of this paper only the following two tensor expressions of the Poincaré Lemma[25] are needed. Both cases (and all their extensions) apply on a domain homeomorhic to $\mathbf{R}^d$ (with $d = 4$ used here).

Case 1: An arbitrary vector T_ν satisfying

$$\partial_{[\mu} T_{\nu]} = 0, \tag{B1}$$

can always be expressed in terms of a scalar S

$$T_\mu = \partial_\mu S, \tag{B2}$$

and the only variation of S allowed is by a constant R.

Case 2: The antisymmetric part of an arbitrary tensor $T_{\nu\lambda}$ obeying

$$\partial_{[\mu} T_{\nu\lambda]} = 0, \tag{B3}$$

can always be expressed in terms of a tensor S such that

$$T_{[\mu\nu]} = \partial_{[\mu} S_{\nu]}, \tag{B4}$$

and S may be varied by the gradient of an arbitrary tensor R

$$\delta S_\mu = \partial_\mu R. \tag{B5}$$

For each of these cases we may append any number of free indices simultaneously on T, S, and R and any symmetries possessed by T on these indices may also be imposed on the tensors S and R. It is also possible to apply the Poincaré Lemma to Majorana tensor–spinors (which have four real spinor components in an appropriately chosen representation: see, for example, Doughty and Collins[20]). As with symmetries on the free space-time indices, symmetries involving the spinor indices may also be transferred to the tensor–spinors corresponding to S and R above.

The following three appendices extend the Poincaré Lemma for specialized cases, required in Secs. IV–VI, involving highly symmetric tensors.

APPENDIX C: POINCARÉ LEMMA EXTENSION 1 (PLX1)

Suppose that $n \geqslant 2$ and the tensor A satisfies

$$A_{\mu_1\cdots\mu_{n-1}\mu_n} = A_{(\mu_1\cdots\mu_{n-1})\mu_n}, \tag{C1}$$

$$A_{\mu_1\cdots\mu_{n-2}[\mu_{n-1}\mu_n\lambda]} = 0. \tag{C2}$$

Then there exists a tensor B such that

$$B_{\mu_1\cdots\mu_{n-1}} = B_{(\mu_1\cdots\mu_{n-1})}, \tag{C3}$$

$$\Gamma_{\mu_1\cdots\mu_{n-1}\mu_n} \equiv A_{\mu_1\cdots\mu_{n-1}\mu_n} - B_{\mu_1\cdots\mu_{n-1},\mu_n}, \tag{C4}$$

$$\Gamma_{\mu_1\cdots\mu_n} = \Gamma_{(\mu_1\cdots\mu_n)}. \tag{C5}$$

Proof: (by induction) $n = 2$: The Poincaré Lemma and (C2) implies there exists a vector B satisfying $A_{[\mu_1\mu_2]} = B_{[\mu_1,\mu_2]}$ with $\Gamma_{\mu_1\mu_2} \equiv A_{\mu_1\mu_2} - B_{\mu_1,\mu_2}$ clearly satisfying $\Gamma_{\mu_1\mu_2} = \Gamma_{(\mu_1\mu_2)}$.

$n \geqslant 3$: Equation (C2) implies there exists a tensor C such that

$$A_{\mu_1\cdots\mu_{n-2}[\mu_{n-1}\mu_n]} = C_{\mu_1\cdots\mu_{n-2}[\mu_{n-1},\mu_n]}, \tag{C6}$$

and by (C1) we may take

$$C_{\mu_1\cdots\mu_{n-2}\mu_{n-1}} = C_{(\mu_1\cdots\mu_{n-2})\mu_{n-1}}. \tag{C7}$$

Using (C6) and (C1) yields

$$C_{\mu_1\cdots\mu_{n-3}[\mu_{n-2}\mu_{n-1},\mu_n]} = 0. \tag{C8}$$

These last two equations are just the assumptions (C1) and (C2) for $n-1$ so that by induction there exists a tensor D such that

$$D_{\mu_1\cdots\mu_{n-2}} = D_{(\mu_1\cdots\mu_{n-2})}, \tag{C9}$$

$$B_{\mu_1\cdots\mu_{n-2}\mu_{n-1}} \equiv C_{\mu_1\cdots\mu_{n-2}\mu_{n-1}} - D_{\mu_1\cdots\mu_{n-2},\mu_{n-1}}, \tag{C10}$$

$$B_{\mu_1\cdots\mu_{n-2}\mu_{n-1}} = B_{(\mu_1\cdots\mu_{n-2}\mu_{n-1})}. \tag{C11}$$

By (C6) and (C10)

$$B_{\mu_1\cdots\mu_{n-2}[\mu_{n-1},\mu_n]} = A_{\mu_1\cdots\mu_{n-2}[\mu_{n-1},\mu_n]}, \tag{C12}$$

and Γ defined in (C4), using the B of (C10), satisfies (C5). Q.E.D.

APPENDIX D: POINCARÉ LEMMA EXTENSION 2 (PLX2)

Suppose the tensor T satisfies

$$T_{\mu_1\cdots\mu_n} = T_{(\mu_1\cdots\mu_n)}, \tag{D1}$$

$$T_{\mu_1\cdots\mu_{n-1}[\mu_n,\lambda]} = 0. \tag{D2}$$

Then there exists a scalar P such that

$$T_{\mu_1\cdots\mu_n} = \partial_{\mu_1}\cdots\partial_{\mu_n} P. \tag{D3}$$

Proof: (By induction) $n = 1$: $T_{[\mu,\nu]} = 0$ implies $T_\mu = \partial_\mu P$ directly by the Poincaré Lemma.

$n \geqslant 2$: Equation (D2) implies

$$T_{\mu_1\cdots\mu_{n-1}\mu_n} = R_{\mu_1\cdots\mu_{n-1},\mu_n}. \tag{D4}$$

Equation (D1) implies we may take

$$R_{\mu_1\cdots\mu_{n-1}} = R_{(\mu_1\cdots\mu_{n-1})}, \tag{D5}$$

and, in addition, (D1) and (D4) imply

$$R_{\mu_1\cdots\mu_{n-2}[\mu_{n-1},\mu_n]} = 0. \tag{D6}$$

Hence, by the inductive assumption,

$$R_{\mu_1\cdots\mu_{n-1}} = \partial_{\mu_1}\cdots\partial_{\mu_{n-1}} P, \tag{D7}$$

or

$$T_{\mu_1\cdots\mu_n} = \partial_{\mu_1}\cdots\partial_{\mu_n} P. \tag{D8}$$

APPENDIX E: POINCARÉ LEMMA EXTENSION 3 (PLX3)

Suppose the tensor T satisfies

$$T_{\mu_1\cdots\mu_n} = T_{(\mu_1\cdots\mu_n)}, \tag{E1}$$

$$T_{(\mu_1\cdots\mu_n,\lambda)} = 0, \tag{E2}$$

and suppose that T and its first n gradients vanish at infinity, namely

$$T_{\mu_1\cdots\mu_n}\to 0,\ T_{\mu_1\cdots\mu_n,\lambda_1}\to 0,\ldots,\ T_{\mu_1\cdots\mu_n,\lambda_1\cdots\lambda_n}\to 0. \tag{E3}$$

Then

$$T_{\mu_1\cdots\mu_n} = 0. \tag{E4}$$

Proof: (by induction) $n = 1$: $T_{(\mu,\lambda)} = 0$ implies $T_{\mu,\lambda} = -T_{\lambda,\mu} = T_{[\mu,\lambda]}$. Therefore $0 = T_{\mu,[\lambda\rho]} = -T_{[\lambda,\rho]\mu}$ which implies $T_{[\lambda,\rho]}$ is constant. Thus by (E3) one has $0 = T_{[\lambda,\rho]} = T_{\lambda,\rho}$ and hence T_μ is also constant and similarly vanishes by (E3).

$n \geqslant 2$: (E1) and (E2) imply

$$T_{\mu_1\cdots\mu_n,\lambda} = -nT_{\lambda(\mu_2\cdots\mu_n,\mu_1)}, \tag{E5}$$

and

$$T_{\mu_1\cdots\mu_n,\lambda} = \frac{2}{n+1}\sum_{j=1}^{n} T_{\mu_1\cdots\hat{\mu}_j\cdots\mu_n[\mu_j\lambda]}. \tag{E6}$$

Where the caret indicates an omitted index, Eq. (E5) implies

$$0 = T^{\mu_1\cdots\mu_n}{}_{,[\lambda\rho]} = -nT_{[\lambda}{}^{(\mu_2\cdots\mu_n,\mu_1)}{}_{,\rho]}. \tag{E7}$$

This (and the induction) allows the theory for $n-1$ (with

two extra free indices, $\kappa\rho$) to be applied to the tensor $S_{\kappa\rho\mu_1\cdots\mu_{n-1}} = T_{\mu_1\cdots\mu_{n-1}[\kappa,\rho]}$ to deduce

$$T_{\mu_1\cdots\mu_{n-1}[\kappa,\rho]} = 0.$$

By (E6), $T_{\mu_1\cdots\mu_n\lambda} = 0$ so that $T_{\mu_1\cdots\mu_n}$ is constant and by (E3) must vanish everywhere.

Remark: As with the Poincaré Lemma, the three extensions above generalize to tensors A or T with extra free spacetime or spinorial indices and any symmetries involving only the free indices can be imposed on the tensors B, Γ, and T.

[1]M. Fierz and W. Pauli, Proc. R. Soc. London Ser. A **173**, 211 (1939).
[2]A. Proca, C. R. Acad. Sci. **202**, 1490 (1936).
[3]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[4]D. G. Boulware and A. Deser, Phys. Rev. D **6**, 3368 (1971).
[5]W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[6]A. Kawakami and S. Kamefuchi, Nuovo Cimento **48**, 239 (1967).
[7]S.-J. Chang, Phys. Rev. **161**, 1308 (1967).
[8]L. P. S. Singh and C. R. Hagen, Phys. Rev. D **9**, 898, 910 (1974).
[9]C. Fronsdal, Phys. Rev. D **18**, 3624 (1978).
[10]J. Fang and C. Fronsdal, Phys. Rev. D **18**, 3630 (1978).
[11]D. L. Wiltshire, M. Sc. thesis, University of Canterbury, 1983 (unpublished).
[12]B. de Wit and Daniel Z. Freedman, Phys. Rev. D **21**, 358 (1980).
[13]G. J. H. Burgers, Ph. D. thesis, Rijksuniversiteit te Leiden, 1985.
[14]F. A. Berends, G. J. H. Burgers, and H. van Dam, Nucl. Phys. B **260**, 295 (1985); "Explicit construction of conserved currents for massless fields of arbitrary spin," preprint, Instituut–Lorentz, Leiden, July 1985.
[15]N. A. Doughty and G. P. Collins, J. Phys. A: Gen. Phys. **19**, L887 (1986).
[16]S. Weinberg, Phys. Rev. B **138**, 988 (1965).
[17]M. A. Rodriguez and M. Lorente, Nuovo Cimento A **83**, 249 (1984).
[18]N. A. Doughty and D. L. Wiltshire, J. Phys. A Gen. Phys. 1986 (in press).
[19]G. P. Collins, M. Sc. thesis, University of Canterbury, 1985 (unpublished).
[20]N. A. Doughty and G. P. Collins, J. Math. Phys. **27**, 1639 (1986).
[21]V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. **34**, 211 (1946).
[22]F. A. E. Pirani, *Introduction to Gravitational Radiation Theory*, edited by S. Deser and K. W. Ford; *Lectures on General Relativity: 1964 Brandeis Summer Institute in Theoretical Physics*, Vol. I (Prentice–Hall, Englewood Cliffs, NJ, 1965).
[23]M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984); D. J. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985); G. B. West, "The construction of gauge-invariant actions for arbitrary spin and bosonic field theories," preprint LA-UR-86-1136, Los Alamos National Laboratory, New Mexico, 1986.
[24]Y. Choquet-Bruhat, C. deWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[25]W. E. Britten, W. R. Smythe, and W. Wyss, Am. J. Phys. **50**, 693 (1982).

Dirac spinor orbits

Richard L. Hall
Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8

(Received 18 March 1985; accepted for publication 24 September 1986)

It is supposed that a single fermion with Hamiltonian $H = \alpha \cdot p + \beta\mu(r) + \phi(r)$, where $\mu(r)$ and $\phi(r)$ are central potentials, obeys the Dirac equation. If $\psi_1(r)$ and $\psi_2(r)$ are the radial factors in the Dirac spinor, then the graph $\{\psi_1(r), \psi_2(r)\}$ for $r \in (0, \infty)$ is called a spinor orbit. In cases where discrete eigenvalues exist, the corresponding spinor orbit eventually returns to the origin. However, if there is a constant $a \geqslant 0$ such that, for $r > a$, the three functions $\phi(r)$, $\phi(r)/\mu(r)$, and $r\mu(r)$ increase monotonically without bound, then it is proved that the spinor orbit must eventually be confined to an annular region excluding the origin. Consequently, the spinor orbit approaches a "spinor circle," the spinor is not L^2, and there are no eigenvalues. This happens, for example, if μ is constant and $\phi(r)$ is any monotone increasing and unbounded potential. In such cases the radius of the spinor circle is sensitive to the energy, and instead of eigenvalues one finds a sequence of resonant energies for which the radii of the spinor circles are local minima.

I. INTRODUCTION

We consider a single spin-$\frac{1}{2}$ fermion moving in a static central field $\phi(r)$. This potential is actually the time component of a four-vector, just like the Coulomb potential of the hydrogen atom. We also allow for a central scalar potential by writing the mass as $m = \mu(r)$. Thus the Dirac Hamiltonian has the form:

$$H = \alpha \cdot p + \beta\mu(r) + \phi(r), \tag{1.1}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and β are the usual Dirac operators defined, for example, in the book by Messiah.[1] The main result of this paper (which we prove in Sec. IV) is to establish sufficient conditions on unbounded potentials which imply that the Hamiltonian H has no eigenvalues. We suppose that there is a distance $a \geqslant 0$, such that for $r > a$ the following three conditions hold:

$$\phi(r),\ \phi(r)/\mu(r), \text{ and } r\mu(r) \text{ are positive monotone increasing without bound.} \tag{1.2}$$

These conditions describe a competition between the scalar and vector potentials. The conditions are satisfied if the vector component $\phi(r)$ dominates the scalar potential $\mu(r)$ sufficiently strongly. For example, if $\mu(r)$ is constant and $\phi(r)$ increases without bound, then the three conditions are met and we know that there are no eigenvalues.

There is extensive literature on the Dirac equation going back to 1928. Interest in the nature of the spectrum when the potentials are unbounded has emerged in recent years mainly because of the importance of the problem for applications to quark physics. Specific results, which we have to date,[2–4] are consistent with the present work, in the general sense that bound states are found not to exist whenever an unbounded vector potential dominates the scalar potential. The purpose of this article is to treat this question more generally, and by a new method that may also prove to be useful as a tool for treating other aspects of the Dirac problem.

Our technique is to study what we call Dirac spinor orbits. These were first introduced in Ref. 5 as a method for finding Dirac eigenvalues with the aid of a microcomputer. For central potentials, we can choose the total angular momentum and the parity and then solve the Dirac equation for a given value E of the energy. The Dirac spinors so generated can be constructed (more details will be given in the next section) with the aid of just two radial functions $\psi_1(r)$ and $\psi_2(r)$, which are sometimes called the large and small radial functions. With the initial conditions

$$\psi_1(0) = \psi_2(0) = 0, \tag{1.3}$$

the energy E is an eigenvalue of H if the boundary condition

$$\int_0^\infty \{\psi_1^2(r) + \psi_2^2(r)\} dr < \infty, \tag{1.4}$$

that is to say, if the spinor is L^2. A Dirac spinor orbit is the graph $\{\psi_1(r), \psi_2(r)\}$ for $r \in (0, \infty)$. In Fig. 1 we show the spinor orbit for the hydrogenic atom with $\mu(r) = 1$, $\phi(r) = -0.2/r$, $J = \frac{5}{2}$, parity $= +1$, and the energy $E = 0.999\,591\,308$, which is (approximately) the energy of the fifth excited radial state. It is, of course, not clear from such a figure whether or not the orbit is L^2 because we have no idea how fast the orbit returns to the origin. However, if the orbit never returned to the origin, we would certainly know that the spinor was not L^2 and that E was not an eigenvalue.

This is how we prove our result. We prove that for the class of potentials satisfying (1.2) the Dirac spinor orbit never returns to the origin; in fact, in those cases we shall prove that, for any choice of E, the orbit is eventually confined to an annular region excluding the origin. Such orbits approach a circular orbit we call a spinor circle. Our proof depends on the convergence properties of some delicate conditionally convergent integrals. These integrals are discussed in Sec. III.

 0022-2488/87/020457-06$02.50 © 1987 American Institute of Physics

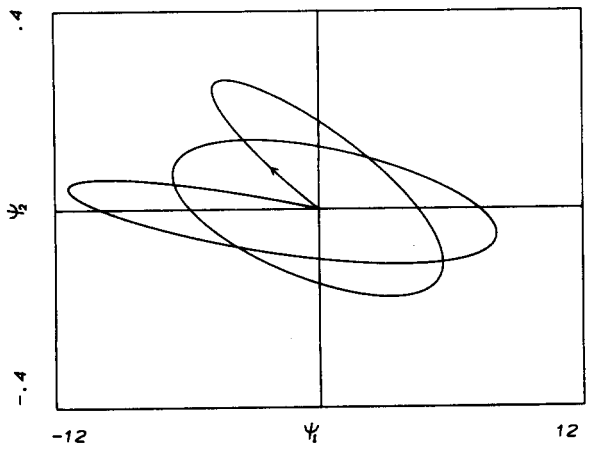


FIG. 1. The spinor orbit for the hydrogenlike problem $\mu = 1$, $\phi(r) = -0.2/r$, and the eigenvalue $E^P_{nj} = E^+_{5/2} = 0.999\,513\,08$.

Some illustrations of spinor orbits with spinor circles and a discussion of the resonancelike properties of the orbits when no eigenvalues exist are presented in Sec. V.

II. THE ORBIT EQUATIONS

We are concerned with central potentials and it is therefore possible to work entirely in a subspace labeled by the total angular momentum quantum numbers JM and the parity P. It is very convenient to follow the exposition of Messiah[1] and introduce two new quantum numbers τ and k defined by

$$P = (-1)^{J+(1/2)\tau}, \quad k = J + \tfrac{1}{2}, \tag{2.1}$$

where τ (called ω by Messiah) takes the two values $\tau = \pm 1$. The spinor can be constructed with the aid of just two radial functions, which we write in the form $t^{-1}\psi_1(t)$ and $t^{-1}\psi_2(t)$, along with spherical harmonics and spin functions. We now use t instead of r for the radial variable in order to suggest the idea that the orbit is generated by an abstract dynamical system in the plane. The coupled radial equations for the central field problem may then be written as the dynamical system

$$\left\{\frac{-d}{dt} + \frac{\tau k}{t}\right\}\psi_2(t) = \{E - \mu(t) - \phi(t)\}\psi_1(t),$$
$$\left\{\frac{d}{dt} + \frac{\tau k}{t}\right\}\psi_1(t) = \{E + \mu(t) - \phi(t)\}\psi_2(t), \tag{2.2}$$

with the initial conditions

$$\psi_1(0) = \psi_2(0) = 0. \tag{2.3}$$

The dynamical system (2.2) can also be written in the vector form

$$\Psi'(t) = S(t)\Psi(t), \quad \text{where } \Psi(t) = [\psi_1(t)\psi_2(t)]^T, \tag{2.4}$$

and $S(t)$ is given by

$$S(t) = \begin{bmatrix} -k\tau/t & (\mu + E - \phi) \\ (\mu - E + \phi) & k\tau/t \end{bmatrix}. \tag{2.5}$$

The singularity in the differential operator $S(t)$ at $t = 0$ allows this system to move away from the origin.

We now introduce an important argument. We suppose that the potentials $\mu(t)$ and $\phi(t)$ are such that there are no singularities in $S(t)$ for $t > 0$. In this case, if $\Psi(t)$ is zero (that is to say, the orbit is at the origin) for $t = t_1 > 0$, then we can run the system both forwards and backwards from t_1, and, since (2.4) is linear and homogenous, we conclude that $\Psi(t)$ is zero for all t. Hence, nontrivial solutions *cannot* return to the origin for finite positive t.

Now we look at an argument concerning the asymptotic behavior of a particular class of potentials. Much of the rest of this paper is directed towards making this kind of argument secure. Suppose $\phi(t)$ is non-negative, increases without bound, and dominates $\mu(t)$ in the sense that $\phi(t)/\mu(t)$ also increases without bound. Then, for large t, $S(t)$ has the asymptotic form

$$S(t) = \begin{bmatrix} 0 & -\phi \\ \phi & 0 \end{bmatrix}. \tag{2.6}$$

We can therefore immediately write down an asymptotic solution to the system equations in the form

$$\psi_1(t) = R\cos\left(\int \phi(t)dt\right),$$
$$\psi_2(t) = R\sin\left(\int \phi(t)dt\right), \tag{2.7}$$

where R is a constant which, for a nontrivial solution, cannot be zero. The graph of this asymptotic solution is a spinor circle and the spinor is therefore not L^2 and there can be no eigenvalues. The weak point of this argument is that, although in the above asymptotic solution R cannot be zero, in the (unknown) exact solution a corresponding time-dependent radial factor may approach zero sufficiently fast for the orbit to be L^2. This possibility must be ruled out.

We shall continue the discussion of the orbits by using the polar variables:

$$r(t) = \{\Psi_1^2(t) + \Psi_2^2(t)\}^{1/2},$$
$$\theta(t) = \arctan\{\psi_2(t)/\psi_1(t)\}. \tag{2.8}$$

In terms of these polar variables the system equations (2.2) become

$$r'(t)/r(t) = \mu(t)\sin(2\theta(t)) - (k\tau/t)\cos(2\theta(t)), \tag{2.9}$$
$$\theta'(t) = (\phi(t) - E) + \mu(t)\cos(2\theta(t)) + (k\tau/t)\sin(2\theta(t)). \tag{2.10}$$

These equations are valid for $t > 0$. The initial condition (2.3) becomes $r(0) = 0$. The initial angle $\theta(0)$ depends on the form of $\phi(t)$ and $\mu(t)$ and can usually be calculated by using a series approximation about $t = 0$.

We now start the work towards the proof of our main result. We observe that (2.10) does not involve $r(t)$. By using our hypotheses (1.2) in the θ equation (2.10) we see that a positive number a exists such that for $t > a$, $\theta'(t) > 0$; also,

$\theta'(t)$ increases without bound as $t\to\infty$. Moreover, if we choose a sufficiently large, we also have $\theta(t)>0$ and therefore, on the patch (a,∞), $\theta(t)$ is positive and monotone increasing without bound. We may therefore change variables and work with θ instead of t. If we let $\theta_1=\theta(a)$, we can now formally integrate Eq. (2.9) and write r in the form

$$r=r(a)\exp[J(\theta)], \tag{2.11}$$

where (with $s=2\theta$) we have

$$J(\theta)=\frac{1}{2}\int_{2\theta_1}^{2\theta}\frac{\sin(s)-\sigma(s)\cos(s)}{g(s)+\cos(s)+\sigma(s)\sin(s)}\,ds, \tag{2.12}$$

and we define $g(s)$ and $\sigma(s)$ by the expressions

$$g(2\theta)=\{\phi(t)-E\}/\mu(t),\quad \sigma(2\theta)=k\tau[t\mu(t)]^{-1}. \tag{2.13}$$

The direction of our argument now is as follows. If we can show that $J(\theta)$ is bounded, say $|J(\theta)|<B$, for $\theta\in(\theta_1,\infty)$, then (2.12) implies that the spinor orbit lies inside an annular region with boundary radii $r(a)\exp(-B)$ and $r(a)\exp(B)$. Meanwhile, as we showed above, $r(a)$ cannot be zero. Hence, such an orbit will not be L^2 and there can be no eigenvalues.

Since, by hypothesis (for t sufficiently large), both $\phi(t)$ and the ratio $\phi(t)/\mu(t)$ increase monotonically without bound, and since we know that $\theta'(t)>0$, and $\theta'(t)$ increases without bound as $t\to\infty$, we can easily prove from (2.13) that $g(s)$ is positive and monotone increasing without bound for s sufficiently large. Similarly, our hypotheses (1.2) guarantee that $|\sigma(s)|$ decreases monotonically to zero as $s\to\infty$ (i.e., $t\to\infty$). We shall assume that properties of $g(s)$ and $\sigma(s)$ for the remainder of the paper.

If $g(s)$ were to increase, for example, like s^α, for $\alpha>1$, then $J(\infty)$ would be absolutely convergent. However, this only happens for potentials which increase faster than exponential. The power-law potentials, for example, cannot be accommodated this way. We can see this by the following argument: suppose $\mu=$ const, and $\phi(t)=t^q$, then, for large t, $\Theta(t)\sim t^{q+1}$, and $\phi\sim\theta^{q/(q+1)}$; consequently, $g(s)\sim s^{q/(q+1)}$ increases too slowly. Therefore in order to handle such potentials as the linear potential, we must treat the integral $J(\theta)$ as a conditionally convergent integral as $\theta\to\infty$.

Since $g(s)$ increases monotonically without bound, comparison of $J(\theta)$ with the well-known convergent improper integral

$$K=\int_{2\theta_1}^{\infty}\frac{\sin(s)}{g(s)}\,ds \tag{2.14}$$

would appear to settle the issue immediately in favor of all increasing potentials. However, it turns out that the convergence of K is easily disturbed by small perturbations such as we have in $J(\theta)$. These convergence questions are treated in Sec. III and we return to complete the proof of our main result in Sec. IV.

III. SOME INTEGRALS

Throughout this section we shall assume that the constant $\theta_1>0$ and that the function $g(\theta)$ is defined and positive and differentiable on (θ_1,∞). In addition we shall assume that $g(\theta)$ is monotone increasing without bound as $\theta\to\infty$. The monotone behavior is essential whereas the differentiability is merely an analytical convenience.

The first result is the well-known convergent integral.

Theorem 3.1:

$$J_1=\int_{\theta_1}^{\infty}\frac{\sin(\theta)}{g(\theta)}\,d\theta. \tag{3.1}$$

Proof of Theorem 3.1: See, for example, the text by Widder,[6] p. 331.

We now demonstrate that the convergence of (3.1) can be upset by a periodic perturbation of the denominator.

Theorem 3.2:

$$J_2=\int_{\theta_1}^{\infty}\frac{\sin(\theta)}{\{g(\theta)-\sin(\theta)\}}\,d\theta \tag{3.2}$$

diverges if the integral $\int_{\theta_1}^{\infty}\{g(\theta)\}^{-2}\,d\theta$ diverges. In particular, the case $g(\theta)=\theta^{1/2}$ diverges.

Proof of Theorem 3.2: We construct a lower bound. Let K_n be the integral in (3.2) from $\theta=t_n=2n\pi$ to $\theta=2(n+1)\pi$, where n is a positive integer. Then, by change of variables, we may write this integral in the form:

$$K_n=\int_0^{\pi}\frac{\sin(t)\{g(t_n+\pi+t)-g(t_n+t)+2\sin(t)\}}{\{g(t_n+t)-\sin(t)\}\{g(t_n+\pi+t)+\sin(t)\}}\,dt. \tag{3.3}$$

Since g is monotone increasing, we therefore have

$$K_n\geq\pi\{g(t_n+2\pi)+1\}^{-2}.$$

This inequality establishes Theorem (3.2). We note that the changes of variables $\theta\to\theta+\pi$ and $\theta\to\theta+\frac{1}{2}\pi$ leave the proof unchanged so that the sin functions can both be changed to cos, and the sign in the denominator can be changed from $-$ to $+$ without altering the divergence properties of the integral.

Before considering the next integral, we first prove a lemma that is really a mean-value theorem for the type of integral we are discussing.

Lemma 3.1: Suppose that the functions $p(\theta)$ and $q(\theta)$ are continuous and suppose that $g(\theta)+p(\theta)\neq0$ for $\theta\in[a,b]$, where $b>a>0$, then there exists a constant $\theta_1\in[a,b]$ such that the integral

$$P=\int_a^b\frac{p(\theta)}{\{g(\theta)+q(\theta)\}}\,d\theta$$

has the representation $P=G(\theta_1)$ where $G(x)$ is defined by

$$G(x)=\int_a^b\frac{p(\theta)}{\{g(x)+q(\theta)\}}\,d\theta.$$

Proof of Lemma 3.1: The function $G(x)$ is continuous. Because $g(\theta)$ is monotone, we see that, as x varies between a and b $G(x)$ varies monotonically, assuming all values between $G(a)$ and $G(b)$. Meanwhile, the integral P lies between $G(a)$ and $G(b)$. This establishes Lemma 3.1.

Theorem 3.3: The following integral is convergent:

$$J_3=\int_{\theta_1}^{\infty}\frac{\sin(\theta)}{\{g(\theta)+\cos(\theta)\}}\,d\theta. \tag{3.4}$$

Proof of Theorem 3.3: We let K_n represent the integral in (3.4) between the limits $t_n=n\pi$ and $(n+1)\pi$, where n is a positive integer. We may now employ the mean-value

Lemma (3.1) along with the appropriate changes of variables to write K_n in the form

$$K_n = (-1)^{(n+1)} G(t_n + \bar{t}_n), \tag{3.5}$$

where the function $G(x)$ is given by

$$G(x) = \ln|(g(x)+1)/(g(x)-1)|, \tag{3.6}$$

and $\bar{t}_n \in [0,\pi]$. Since g is monotone increasing to infinity, it follows that the function $G(x)$ in (3.6) is monotone decreasing to zero. Meanwhile, the sequence $\{t_n + \bar{t}_n\}$ is monotone increasing to infinity. Consequently a sum of terms of the form K_n is a convergent alternating series and we have proved that the integral (3.4) converges. We note that by making the changes of variable, $\theta \to \theta + \frac{1}{2}\pi$ and $\theta \to \theta + \pi$, the proof again goes through and we can conclude that the convergence is unaltered if the sin and cos are interchanged and also if the sign in the denominator is reversed.

IV. THE SPINOR CIRCLE THEOREM

We now continue the proof of our main result, which we began in Sec. II. We need to show that for the class of potentials (1.2) the integral $J(\theta)$ given in (2.12) is bounded for $\theta \in (\theta_1, \infty)$. We shall prove this by showing that, by a suitable change of variables, we can write $J = J(\infty)$ as the sum $J = J_3 + J_4$ of two integrals, where J_3 is the conditionally convergent integral of Theorem 3.3 and J_4 is absolutely convergent. We begin by defining the angle $\gamma(s)$, $-\frac{1}{2}\pi \leqslant \gamma(s) \leqslant \frac{1}{2}\pi$, and the quantity $\rho(s)$ as follows:

$$\gamma(s) = \arctan(\sigma(s)), \quad \rho(s) = \{1 + \sigma^2(s)\}^{1/2}. \tag{4.1}$$

We recall that $g(s)$ is monotone increasing without bound and $\sigma^2(s)$ is monotone decreasing to zero as $s \to \infty$ [the sign of $\sigma(s)$ itself is the same as τ]. The integral J now takes the form

$$J = \frac{1}{2}\int_{2\theta_1}^{\infty} \frac{\sin(s-\gamma(s))}{\{g(s)/\rho(s) + \cos(s-\gamma(s))\}}\, ds. \tag{4.2}$$

It is natural at this point to make the change of variables

$$u = s - \gamma(s), \quad f(u) = g(s)/\rho(s). \tag{4.3}$$

It follows from (4.1) and our hypotheses (1.2) and the result from Sec. II that $\theta'(t) > 0$, that

$$\gamma'(s) = \tfrac{1}{2}\sigma(s)\sigma'(s)\rho^{-2}(s) < 0. \tag{4.4}$$

Hence we have the differential relation

$$ds = du + \gamma'(s)ds. \tag{4.5}$$

Meanwhile, because $\gamma(s)$ is bounded and s increases without bound, we conclude that the variable u increases monotonically without bound as $s \to \infty$. The integral now has the form $J = J_3 + J_4$, where J_3 is given by

$$J_3 = \frac{1}{2}\int_{u_1}^{\infty} \frac{\sin(u)}{\{f(u) + \cos(u)\}}\, du, \tag{4.6}$$

where $u_1 = 2\theta_1 - \gamma(2\theta_1)$, and J_4 is given by

$$J_4 = \frac{1}{2}\int_{2\theta_1}^{\infty} \frac{\sin(u)\gamma'(s)}{\{f(u) + \cos(u)\}}\, ds. \tag{4.7}$$

The function $f(u) = g(s)/\rho(s)$ is an increasing function of u for we have

$$\frac{df}{ds} = \frac{d\{g/\rho\}}{ds} = g'(s)\rho^{-1}(s) - g(s)\sigma(s)\sigma'(s)\rho^{-3}(s) > 0, \tag{4.8}$$

and therefore, since $du/ds > 0$, it follows that $f'(u) > 0$. Also, since $u \to \infty \Rightarrow s \to \infty$, we know also that $f(u)$ increases without bound as $u \to \infty$. Consequently, the integral in (4.6) has exactly the form of the conditionally convergent integral J_3 in Theorem 3.3.

The integral J_4 is absolutely convergent because the factor $\gamma'(s)$ in the integrand decreases to zero sufficiently fast and the other factor is bounded. In fact we have

$$|2J_4| \leqslant |\gamma(2\theta_1)|[g(2\theta_1)/\rho(2\theta_1) - 1]^{-1}. \tag{4.9}$$

Thus the integral $J(\infty)$ is bounded and so therefore is $J(\theta)$ for $\theta \in (\theta_1, \infty)$. Hence there exists a positive number B such that $|J(\theta)| \leqslant B$ for $\theta \in (\theta_1, \infty)$. Consequently, from (2.11) we have that the spinor orbit is confined for $t > a$ to the annulus bounded by $r = r(a)\exp(\pm B)$. Since the orbit is confined to an annular region excluding the origin, we can now rely on the argument we introduce in Sec. II that shows that the orbit approaches a circle as $t \to \infty$.

This completes our proof that no eigenvalues exist and the orbit is asymptotically circular if the potentials satisfy the conditions (1.2).

V. AN EXAMPLE

We now consider an illustration of the circle phenomenon which we have thus far characterized in general mathematically. We consider the case of a harmonic oscillator with constant mass $\mu(t) = 1$ and vector potential $\phi(t) = (0.02)\, t^2$. We first suppose that in all innocence we were to look for an "eigenvalue" with $J = \frac{1}{2}$, parity $P = -1$, and the fourth radial state (counting the bottom as the first), that is to say $n = 4$, where n is a radial quantum number. First, from (2.1) we see that $\tau = k = 1$. Examination of the orbit equations near $r = 0$ leads us to the conclusion that the orbit should start from the origin initially along the Ψ_2 axis. The absolute size of the orbit is not significant in this discussion, but all the orbits we show are drawn to the same scale. As we shall see, the relative sizes are important.

In order to find the energy $E^P_{nj} = E^{-}_{4\,1/2}$, we therefore integrate the Dirac equations (2.2) and try to find a value for E which causes the orbit to have three nodes in $\psi_1(t)$ and after this the orbit should (for the $n = 4$ eigenvalue) approach the origin. If we arrange for a microcomputer to perform this search automatically we get the Dirac spinor orbit shown in Fig. 2(a) which was obtained with $E = 2.385\,28$.

However, although all seems well at first glance, difficulties are encountered with the algorithm that searches for that energy which would allow the orbit to approach the origin after the required number of nodes. In fact, as we know, no such approach is possible. Figure 2(b) shows the central part of the orbit and it is the best we could do in choosing E to minimize the size of the spinor circle. The radius of this circle is about 1% of the overall size of the complete orbit. The T inscribed next to the orbit indicates the point at which the equation

$$E = \phi(t) - \mu(t) \tag{5.1}$$

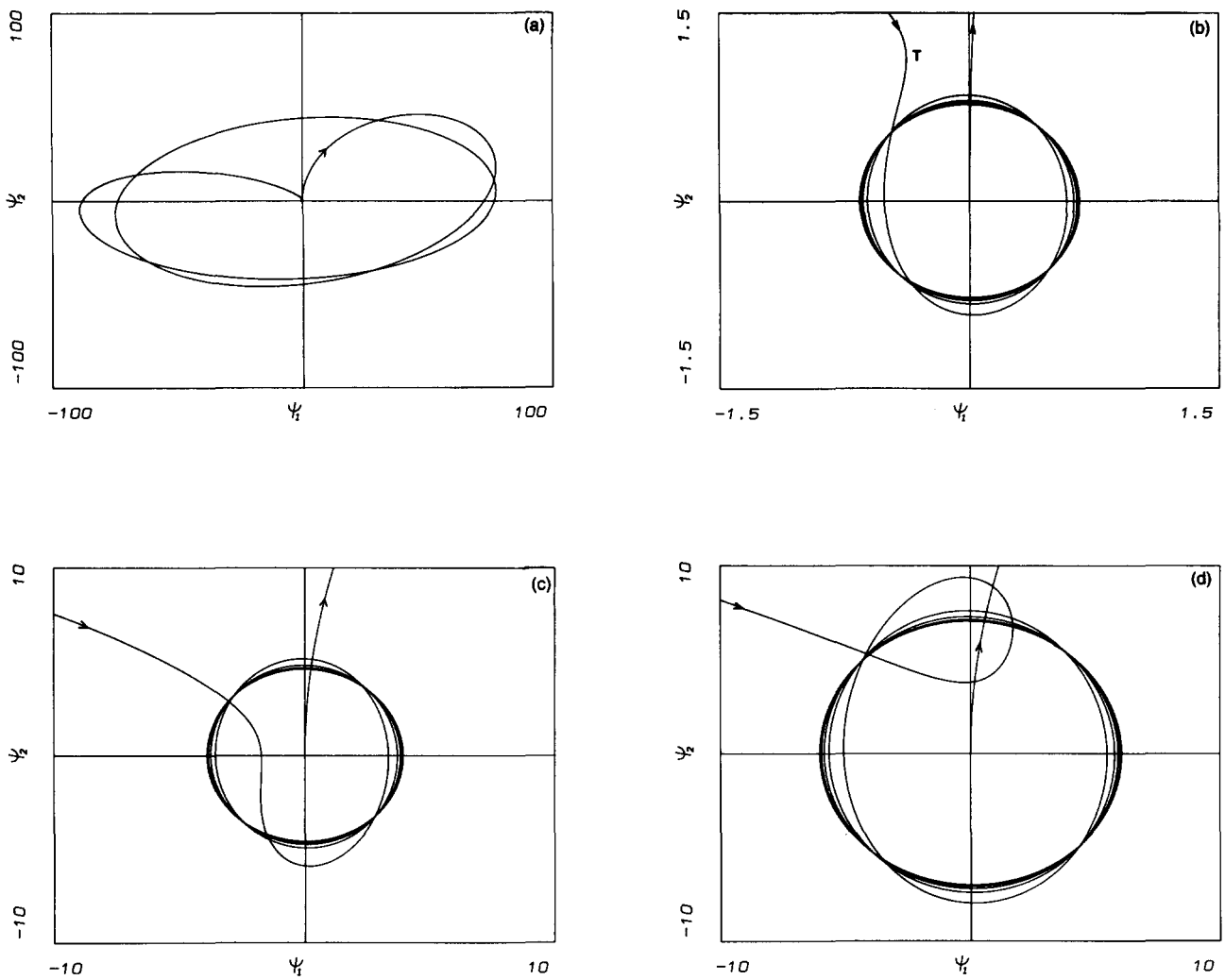


FIG. 2. (a) The spinor orbit for the harmonic-oscillator potential $\mu = 1$, $\phi(r) = 0.02\, r^2$, and the resonance $E_{\bar{4}1/2} = 2.382\,58$. (b) The central part of the spinor orbit in Fig. 2(a) for the harmonic oscillator, showing the approach to the minimal spinor circle. (c) The central part of the spinor orbit for the harmonic-oscillator potential $\mu = 1$, $\phi(r) = 0.02\, r^2$ of Fig. 2(a), but with the energy value $E = E_{\bar{4}1/2} - 10^{-4}$. (d) The central part of the spinor orbit for the harmonic-oscillator potential $\mu = 1$, $\phi(r) = 0.02\, r^2$ of Fig. 2(a), but with the energy value $E = E_{\bar{4}1/2} + 10^{-4}$.

is satisfied. After this point, the initial clockwise direction of the orbit is reversed and the orbit now assumes the anticlockwise direction of the spinor circle according to the matrix (2.5).

In Figs. 2(c) and 2(d) we show what happens to the spinor circle if the energy E is, respectively, decreased and increased by only 10^{-4}, in both cases the circle radius increases by a factor of about 10. What we have here is a phenomenon exactly analogous to resonance in scattering. One can define resonance energies in scattering to be precisely those values for which, for a given ingoing probability current, the outgoing probability current is minimized. In the present problem, the Dirac particle therefore "sees" (in the Schrödinger sense) not an unbounded confining oscillator potential r^2 but something (qualitatively) more like $r^2 - r^4$: the spinor circle is simply a representation for the outgoing wave after tunneling through the barrier.

We note that the same phenomena as we have seen with the oscillator are found with other examples that meet the hypotheses of the circle theorem. For example, we have found similar graphs for the very slowly increasing unbounded potential $\phi(t) = \ln(\ln(t+3))$ with $\mu(t) = 1$. In all cases one can find a sequence of resonance energies that locally minimize the radius of the spinor circle.

VI. CONCLUSION

We have proved that there are no eigenvalues to the one-particle Dirac problem if the vector potential $\phi(r)$ increases without bound and dominates the scalar potential $\mu(r)$ in the sense that $\phi(r)/\mu(r)$ increases without bound as $r \to \infty$.

We also require that the scalar potential does not diminish with distance too fast, that is to say, that $r\mu(r)$ increases without bound. For potentials satisfying these hypotheses, we have proved that the spinor orbit eventually lies inside an annular region excluding the origin and it approaches a spinor circle asymptotically as $r \to \infty$.

In certain special cases it has been argued by Su and Ma[7] that, if the vector and scalar potentials are equal, then there are no eigenvalues. It does not seem possible to prove this result in general by our method.

There may be another class of potentials for which the orbit can eventually approach the origin but not sufficiently fast for the spinor to be square integrable. We do not know what such potentials might look like except that they would have to be bounded.

ACKNOWLEDGMENTS

The author would like to thank the Natural Sciences and Engineering Research Council Canada for partial financial support of this work by Grant No. A3438.

[1]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1961), Vol. II. (The Dirac Hamiltonian with central potentials is discussed on p. 924.)
[2]C. Long and D. Robson, Phys. Rev. D **27**, 644 (1983).
[3]P. M. Fishbane, S. G. Gasiorowicz, D. C. Johansen, and P. Kaus, Phys. Rev. D **27**, 2433 (1983).
[4]R. K. Su and Y. Zhang, J. Phys. A **17**, 851 (1984).
[5]R. L. Hall, J. Phys. A **19**, 2079 (1986).
[6]D. V. Widder, *Advanced Calculus* (Prentice–Hall, Englewood Cliffs, NJ, 1961). Theorem 3.1 of the present paper is proved on p. 333 of this book.
[7]R. K. Su and Z. Q. Ma, J. Phys. A **19**, 1739 (1986).

Spontaneous symmetry breaking and the Higgs mechanism for quaternion fields

C. Nash and G. C. Joshi
School of Physics, University of Melbourne, Parkville, Victoria, 3052, Australia

(Received 27 February 1985; accepted for publication 10 September 1986)

In the context of quaternion valued fields spontaneous symmetry breaking and the Higgs mechanism are investigated. In particular, for the potential $-(\mu^2/2)\bar{\phi}\phi - (\lambda/4)(\bar{\phi}\phi)^2$, Goldstone's theorem is studied and, for the gauge theory of automorphisms of the quaternions, a Higgs mechanism investigated.

I. INTRODUCTION

The fact that the Birkhoff and von Neumann[1] formulation of quantum mechanics admits quaternionic scalars has invited study of so-called "quaternionic quantum mechanics" (QQM) by a number of authors.[2–9] Independently, Kaneno[2] and Finkelstein *et al.*[3] first looked at QQM in the early 1960s. Finkelstein *et al.* and Rembielinski[7] have investigated the quaternionic quantum Hilbert space.

This interest in QQM began concurrently with interest in gauge theories; not surprisingly, gauge theory was incorporated in QQM at the beginning. As a result, several gauge symmetries have been proposed.[4,5,8,9] Since we do not find these symmetries exactly reflected in nature, we must find a mechanism by which to break them. Spontaneous symmetry breaking (SSB) and the Higgs mechanism are successful in breaking gauge symmetries in the complex theory.[10]

In a quaternionic theory the classical fields are quaternion valued functions of space time. The Lagrangian will be a function of these quaternionic fields and so the form of it will be restricted by quaternionic multiplication. There will, however, be no restriction on the nature of the fields like those on Chkareuli's[8] "quasireal" fields. Equally, the derivative will not be restricted as Morita's[9] is; this can be done without losing the association between infinitisimal changes and derivatives.

We will produce SSB and the Higgs mechanism for quaternionic valued fields, though it is not the case that we give a physical interpretation of the full quaternionic field apart from its components. That is, we wish to consider the physical fields as the components of the quaternionic fields.

The plan of this paper is as follows: in Sec. III, a quaternionic SSB will be investigated and a quaternionic Goldstone's theorem produced for the potential $-(\mu^2/2)\bar{\phi}\phi - (\lambda/4)(\bar{\phi}\phi)^2$. In Sec. IV, quaternionic gauge theory is discussed and the preferred example of an automorphic gauge theory examined. In Sec. V a quaternionic Higgs mechanism is produced in the context of automorphic gauge theory.[4] In Sec. VI a simple Lagrangian is expanded in terms of quaternionic components for a general vacuum vector and possible decompositions of the gauge boson *à la* Finkelstein *et al.*[4] are investigated.

II. NOTATION

We will use Einstein summation notation. Greek indices will take the values 0, 1, 2, and 3 and Latin indices will take the values 1, 2, and 3 unless otherwise specified.

Let the quaternions Q be the noncommutative algebra $\{q_\alpha e_\alpha : q_\alpha \in \mathbf{R}\}$ where addition is defined by $q_\alpha e_\alpha + p_\alpha e_\alpha = (q_\alpha + p_\alpha)e_\alpha$ and multiplication is defined by

$$q_\alpha e_\alpha \cdot r_\beta e_\beta = q_\alpha r_\beta (e_\alpha \cdot e_\beta),$$

where

$$e_0 \cdot e_0 = e_0, \quad e_i \cdot e_0 = e_0 \cdot e_i = e_i,$$

$$e_i e_j = -\delta_{ij} e_0 + \epsilon_{ijk} e_k,$$

with ϵ_{ijk}, the alternating symbol with $\epsilon_{123} = 1$.

Let $\phi(x)$ be a quaternion valued field and define

$$\phi(x) = \phi_\alpha(x) e_\alpha = \phi(x) e_0 + \phi_i(x) e_i$$
$$= \phi_0(x) e_0 + \underline{\phi}(x) \cdot \underline{e},$$

where an underbar indicates the vector component. Note that $\phi_\alpha(x)$ is a real function $\mathbf{R}^4 \to \mathbf{R}$. Let

$$\bar{\phi}(x) = \phi_\alpha(x) \bar{e}_\alpha,$$

where

$$\bar{e}_0 = e_0 \quad \text{and} \quad \bar{e}_i = -e_i.$$

Let

$$\partial_\mu \phi(x) = \partial_\mu \phi_\alpha(x) e_\alpha.$$

Let

$$S(\phi) = \tfrac{1}{2}(\phi + \bar{\phi}) = \phi_0 e_0.$$

Let

$$|q| = \bar{q}q \text{ for } q \in Q.$$

This is a norm on the quaternions and $\bar{q}q = q_\alpha q_\alpha e_\alpha$. It is noteworthy that $|q_1 q_2| = |q_1|\,|q_2|$.

III. QUATERNIONIC SPONTANEOUS SYMMETRY BREAKING

We begin by looking at the classical nongauge invariant case of SSB for quaternion valued fields. Variations δV in the potential $V(\phi)$ due to variations in ϕ are approximated by $(\partial V/\partial \phi_\alpha)\delta\phi_\alpha$. So for symmetries which rotate the ϕ_α amongst themselves SSB and Goldstone's theorem follow that of four real fields.

The following shows what happens in quaternion language. Remembering that multiplication is quaternionic we consider the Lagrangian

$$\mathcal{L} = \partial_\mu \bar{\phi}\, \partial_\mu \phi - V(\phi), \tag{3.1a}$$

 0022-2488/87/020463-05$02.50 © 1987 American Institute of Physics

$$V(\phi) = (\mu^2/2)\bar{\phi}\phi + (\lambda/4)(\bar{\phi}\phi)^2,$$
$$\mu^2 < 0, \quad \lambda > 0. \tag{3.1b}$$

For future use we note the action of the derivative on the function $V: Q \to \mathbf{R}$ is

$$dV(\phi, d\phi) = (\mu^2 + \lambda\bar{\phi}\phi)S(\phi\, \overline{d\phi}), \tag{3.2}$$

$$d(dV)(\phi, d\phi) = 2\lambda S(\phi\ \overline{d\phi})^2 + (\mu^2 + \lambda\bar{\phi}\phi)S(d\phi\ \overline{d\phi}). \tag{3.3}$$

We now consider an infinitesimal variation $\delta\phi$ of ϕ with

$$\delta\phi = \omega^i R^{\ i}_{\beta\gamma}\phi_\gamma e_\beta, \tag{3.4}$$

where ω^i is infinitesimal and real, $R^{\ i}_{\beta\gamma}$ is real, and i takes the values 1,2,...,n. The variation in V to first order (see Refs. 11, 12, and 13) is

$$\delta V = (\mu^2 + \lambda\bar{\phi}\phi)S(\phi\ \overline{\delta\phi}). \tag{3.5}$$

Differentiating (3.5) we find

$$d(\delta V) = (\mu^2 + \lambda\bar{\phi}\phi)d\left[S(\phi\ \overline{\delta\phi})\right] + 2\lambda S(\phi\ \overline{d\phi})S(\phi\ \overline{\delta\phi}). \tag{3.6}$$

The vacuum state is the state of minimum energy and can be found by minimizing the potential. So the vacuum state satisfies

$$dV(v, d\phi) = 0, \tag{3.7}$$

for all $d\phi \in Q$, where v is the value of the field in the vacuum state. From (3.2) and (3.7) we have either $\bar{v}v = -\mu/\lambda$ or $S(v, d\phi) = 0$ for all $d\phi \in Q$, and $S(v, d\phi) = 0$ for all $d\phi \in Q$ implies $v = 0$. So as $V(v)$ is a minimum of V, and $\mu^2 < 0$ and $\lambda > 0$ implies $V(v) < 0$, we have

$$\bar{v}v = -\mu^2/\lambda. \tag{3.8}$$

Such a vacuum state is in general not invariant under variations (3.4).

We now assume that the Lagrangian and the potential are invariant under variations (3.4). That is that $\delta V = 0$ for all w^i, $i = 1,...,n$. Therefore

$$d(\delta V)|_{\phi = v} = 0, \tag{3.9}$$

for all $d\phi \in Q$ and for all ω^i, $i = 1,...,n$. Now as $v \neq 0$, $S(v\ \overline{d\phi})$ can be made to take any real value by suitable choice of $d\phi$. So using (3.8) and (3.6) we have

$$S(vR^{\ i}_{\beta\gamma}v_\gamma\bar{e}_\beta) = 0, \tag{3.10}$$

for all $i = 1,...,n$.

A real valued function V of a quaternionic variable ϕ has the following "Taylor" series expansion[12]:

$$V(\phi) = V(v) + dV|_{\substack{\phi = v \\ d\phi = \phi'}} + \tfrac{1}{2}ddV|_{\substack{\phi = v \\ d\phi_2 = d\phi_1 = \phi'}} + \cdots, \tag{3.11}$$

where we have expanded $V(\phi)$ about v with $\phi' = \phi - v$.

For the potential (3.1b) using (3.2), (3.3), and (3.8) we find that its series expansion to second order is

$$V(\phi) = V(v) + \lambda S(v\phi')^2. \tag{3.12}$$

Alternatively by straight substitution of $\phi' + v$ for ϕ in (3.1b) we find using (3.8) that

$$V(\phi) = V(v) + \lambda S(v\ \overline{\phi'})^2 + (\lambda/4)(\ \overline{\phi'}\phi')^2 + (\lambda/2)(\ \overline{\phi'}\phi')(\phi' v + \bar{v}\phi'). \tag{3.13}$$

Thus if ϕ' is a linear combination of the $R^{\ i}_{\alpha\beta}v_\beta e_\alpha$, then by (3.10) there are no second-order terms in ϕ' in the potential. Consider the subspace G (the quaternions may be considered as a four-dimensional real vector space with the unit vectors e_0, e_1, e_2, e_3) of the quaternions spanned by $\{R^{\ i}_{\alpha\beta}v_\alpha e_\beta\colon\ i = 1,...,n\}$. As we physically interpret ϕ' at the component level, and as any component of ϕ' in G will have no second-order term in the Lagrangian, then any component of ϕ' in G will correspond to a massless particle. Now the dimension of G is the same as the number of independent components of ϕ', which can be massless, and this is equal to the number of independent quaternions $R^{\ i}_{\alpha\beta}v_\beta e_\alpha$, with i taking values 1, 2,...,n. This is Goldstone's theorem.

IV. QUATERNIONIC GAUGE THEORY

We now examine the rudiments of quaternionic gauge theory for classical scalar fields. Take a field $\phi(x) \in Q$ and assume that the Lagrangian governing the motion is invariant under transformations $\phi \to \phi + \delta\phi$, where

$$\delta\phi = \omega^i R^{\ i}_{\alpha\beta}\phi_\beta e_\alpha \tag{4.1}$$

for infinitesimal $\omega^i \in \mathbf{R}$, $i = 1,2,...,n$. Note that we are considering rotations within a quaternion value field, that is, rotations between the components of a field not rotations between quaternion valued fields. However, with the interpretation that the quaternion components are the physical fields then we are still rotating the particles amongst themselves.

If we allow ω^i to depend upon the space-time coordinate x then to ensure the invariance of the Lagrangian under these "local" transformations we need to introduce a covariant derivative. We require any gauge field introduced in the covariant derivative to be quaternion valued and that it appears in the Lagrangian as a quaternion and not in components. As a result, although we introduce several gauge bosons at a component level, we can only produce the interaction terms by a combination of quaternion multiplication and quaternion addition of the gauge boson field and the other particle fields, so the quaternionic form of the covariant derivative is dependent on the R's.

In a quantum theory, the field and the canonical momentum will satisfy some (anti)commutation relations. For complex theories U(1) symmetry is guaranteed to leave the (anti)commutation relations alone as $\exp(i\theta)$ commutes with all the operators.[14] However, not all of the R's above generate transformations that leave the form of the (anti)commutation relations of the quaternionic field alone. One transformation that does not have this problem is ϕ going to $q\phi q^{-1}$, where q is a quaternion. Since π_μ the conjugate momentum is a function of ϕ and $\partial_\mu\phi$ then π_μ goes to $q\pi_\mu q^{-1}$ and so the form of the (anti)commutation relations is unchanged.

So we will follow Finkelstein *et al.*[4] and consider the automorphisms of the quaternions as the symmetry of the Lagrangian. This means that the physics described by the Lagrangian does not vary with changes in the quaternions that leave the algebra invariant.

These automorphisms form the group O(3) acting on the vector part of the quaternions. That is that every auto-

morphic gauge transformation leaves ϕ_0 alone and transforms $\underline{\phi}$ by an O(3) gauge transformation.

As a global symmetry of the Lagrangian, we have that the Lagrangian is invariant under the transformation

$$\phi(x) \to q\phi(x)q^{-1}, \tag{4.2}$$

for $q \in Q$. In infinitesimal form this transformation is

$$\phi(x) \to (1_{\alpha\beta} + 2\epsilon_i Z^i_{\alpha\beta})\phi_\beta e_\alpha, \tag{4.3}$$

where $i = 1,2,3$, and can be written

$$\phi(x) = \phi(x) + [\phi(x),E], \tag{4.4}$$

with

$$Z^i_{\alpha\beta} = 0,$$

for $\alpha = 0$ or $\beta = 0$, and

$$Z^i_{\alpha\beta} = \epsilon_{i\alpha\beta},$$

for $a \neq 0$ and $\beta \neq 0$. Here, ϵ_{ijk} is the totally antisymmetric symbol with $\epsilon_{123} = 1$, E is defined as $\epsilon_i e_i$, and ϵ_i is an infinitesimal. As a local symmetry, we have that the Lagrangian is invariant under transformations

$$\phi(x) \to q(x)\phi(x)q(x)^{-1}, \tag{4.5}$$

for any function q: $\mathbb{R}^4 \to Q$ such that $dq(x,dx)$ exists for all x and dx. In infinitesimal form this is

$$\phi(x) \to (1_{\alpha\beta} + 2\epsilon_i(x)Z^i_{\alpha\beta})\phi_\beta e_\alpha. \tag{4.6}$$

This can be written

$$\phi(x) \to \phi x + [\phi(x),E(x)] \tag{4.7}$$

with Z^i as above and with $\epsilon_i(x)$ infinitesimal and differentiable for all x, while $E(x)$ is still $\epsilon_i(x)e_i$ and $dE(x,dx)$ exists.

The requirement of local symmetry of the Lagrangian is, as such, not satisfied by all Lagrangians that satisfy the global symmetry. The problem lies with the derivatives of the fields. Assuming that the Lagrangian has no derivative of greater than first order then local symmetry is satisfied by a Lagrangian that satisfies global symmetry if we replace the derivative, ∂_μ, by a covariant derivative, D_μ, such that

$$q(x)(D_\mu\phi)q(x)^{-1} = D'_\mu(q(x)\phi(x)q(x)^{-1}). \tag{4.8}$$

We let D_μ have the form

$$D_\mu\phi = \partial_\mu\phi + q[A_\mu,\phi], \tag{4.9}$$

where A_μ is a quaternionic vector field with $A_\mu = A_{\mu\alpha}e_\alpha$. The requirement that $D_\mu\phi$ transforms as in (4.8) is a necessary and sufficient condition for A_μ, defined in (4.9), to transform

$$A_\mu(x) \to q(x)A_\mu(x)q(x)^{-1} + (1/g)(\partial_\mu q(x))q(x)^{-1}, \tag{4.10}$$

which has the infinitesimal form

$$A_\mu(x) \to A_\mu(x) + [A_\mu(x),E(x)] + (1/g)\partial_\mu E(x), \tag{4.11}$$

where E is defined above.

A kinetic energy term for the gauge field A_μ must be gauge invariant and involve first derivatives of A_μ. Such a term is $\overline{K_{\mu\nu}}K_{\mu\nu}$, where

$$K_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - g[A_\mu,A_\nu]. \tag{4.12}$$

Explicitly, we have

$$\begin{aligned}\overline{K_{\mu\nu}}K_{\mu\nu} = {} & 2(\overline{\partial_\mu A_\nu}\partial_\mu A_\nu - \overline{\partial_\mu A_\nu}\partial_\nu A_\mu) \\ & - 2g(\overline{\partial_\mu A_\nu}[A_\mu,A_\nu] + [\overline{A_\mu,A_\nu}]\partial_\mu A_\nu) \\ & + g^2[\overline{A_\mu,A_\nu}][A_\mu,A_\nu].\end{aligned} \tag{4.13}$$

V. THE HIGGS MECHANISM FOR A QUATERNIONIC GAUGE THEORY

If there are to exist gauge symmetries apart from electromagnetism and color, then the zero mass Goldstone bosons that are predicted are phenomenologically unacceptable. The Higgs mechanism transforms the fields to a gauge—the unitary gauge—in which the offending fields no longer lie in the space of massless particles that Goldstone's theorem provided, thus avoiding unknown massless particles. In complex theories it is fortuitous that in the unitary gauge not only is Goldstone's theorem a null result but that the gauge bosons decouple from the scalar field to first order and that some gauge bosons acquire mass.

To look at the Higgs mechanism quaternionically we combine the two examples already used. When local automorphic gauge invariance, of Sec. IV, is required of the Lagrangian (3.1a), it becomes

$$\mathcal{L} = \overline{D_\mu\phi}D_\mu\phi - V(\phi) + \overline{K_{\mu\nu}}K_{\mu\nu}, \tag{5.1}$$

where $V(\phi)$ is defined in (3.1b), $D_\mu\phi$ is defined in (4.9), and the kinetic term of (4.13) has been added.

Goldstone's theorem is—in this language—for a vacuum quaternion v satisfying $\bar{v}v = -\mu^2/\lambda$ that if a component of $\phi' = \phi - v$ is in the space G generated by $\{Z^i_{\alpha\beta}v_\beta e_\alpha: i = 1,2,3\}$, then that component has no second-order terms in V. The Higgs mechanism becomes finding a gauge—the unitary gauge—such that in this gauge the shifted field has no components in G. That is, if the unitary gauge is the physical gauge then none of the physical Higgs bosons may be forced to be massless through Goldstone's theorem.

Now writing gauge transformations as $\phi \to O(\phi)$, the transformation to the unitary gauge is required by the last paragraph in Sec. III to satisfy "$O(\phi) - v$ has no component in G." This means that

$$(O(\phi) - v)_\alpha Z^i_{\alpha\beta}v_\beta = 0, \tag{5.2}$$

for $i = 1,2,3$. By the definition of $Z^i_{\alpha\beta}$ we see that $v_\alpha z^i_{\alpha\beta}v_\beta = 0$ for $i = 1,2,3$. Therefore the unitary gauge condition becomes

$$O(\phi)_j\epsilon_{ijk}v_k = 0. \tag{5.3}$$

Remember that the gauge transformations act as O(3) on $\underline{\phi}$ and leave ϕ_0 alone. It is the case that for every gauge transformation there is an O(3) transformation and vice versa such that

$$\underline{O_{\text{gauge}}(\phi)\cdot e} = O_{\text{gauge}}(\underline{\phi}\cdot\underline{e}) = O_{O(3)}(\underline{\phi})\cdot\underline{e}. \tag{5.4}$$

This along with (5.3) and the fact that ϵ_{ijk}, $i = 1,2,3$, are the generators of O(3) allows us, following Weinberg,[15] to conclude that a transformation O satisfying (5.3) exists. The expression (5.3) may also be written $[O(\phi),v] = 0$ and is equivalent to "$O(\phi)$ is parallel to $\underline{v}$."

We now expand $\overline{D_\mu\phi}D_\mu\phi$ in the unitary gauge in terms of the shifted fields. Remember that $D_\mu\phi$ is designed that

under gauge transformations $D_\mu\phi \to q(x)D_\mu\phi q(x)^{-1}$. Thus $\overline{D_\mu\phi}D_\mu\phi$ is invariant under gauge transformations. So we have

$$\begin{aligned}\overline{D_\mu\phi}\, D_\mu\phi &= \overline{D^0_\mu\phi^0}\, D^0_\mu\phi^0 \\ &= \overline{D^0_\mu(\phi^{0'}+v)}\, D^0_\mu(\phi^{0'}+v),\end{aligned} \tag{5.5}$$

where the ^{0} denotes the transformation to the unitary gauge, so that $\underline{\phi}^0$ is parallel to $\underline{v}$, $\phi^{0'}$ is defined as $\phi^0 - v$. Now we drop the ^{0}'s and leave the unitary gauge as implicit in what follows. We have

$$\begin{aligned}&\overline{D_\mu(\phi'+v)}\, D_\mu(\phi'+v) \\ &\quad= \overline{\partial_\mu\phi'}\,\partial_\mu\phi' + g\,\overline{\partial_\mu\phi'}\,[A_\mu,\phi'+v] \\ &\qquad + g[\,\overline{A_\mu,\phi'+v}\,]\partial_\mu\phi' \\ &\qquad + g^2[\,\overline{A_\mu,\phi'+v}\,]\,[A_\mu,\phi'+v].\end{aligned} \tag{5.6}$$

As $\underline{\phi} = \underline{\phi'} + \underline{v}$ is parallel to $\underline{v}$ then so is $\underline{\phi'}$. As $\underline{v}$ is a constant, and as $\underline{\phi'}$ is parallel to $\underline{v}$, $\partial_\mu\underline{\phi'}$ is parallel to $\underline{v}$ and therefore to $\underline{\phi'+v}$. That $\partial_\mu\underline{\phi'}$ is parallel to $\underline{\phi'+v}$ means, as the reader can verify, that the second and third terms in (5.6) cancel each other. In the analogous complex theory these terms both vanish independently. This difference results from using quaternions explicitly rather than using the components. The terms vanish independently if we write

$$\overline{D_\mu(\phi'+v)}\, D_\mu(\phi'+v)$$

as

$$(D_\mu(\phi'+v))_\alpha\,(D_\mu(\phi'+v))_\alpha e_0.$$

In either case we have that

$$\begin{aligned}&\overline{D_\mu(\phi'+v)}\, D_\mu(\phi'+v) \\ &\quad= \overline{\partial_\mu\phi'}\partial_\mu\phi' + g^2[\,\overline{A_\mu,v}\,]\,[A_\mu,v] \\ &\qquad + 2g^2S([\,\overline{A_\mu,\phi'}\,]\,[A_\mu,v]) + g^2[\,\overline{A_\mu,\phi'}\,]\,[A_\mu,\phi'],\end{aligned} \tag{5.7}$$

where first order A_μ, ϕ cross terms have disappeared as well as this a second-order "mass" term for A_μ has appeared. Then the Lagrangian in the unitary gauge using (3.13), (4.13), and (5.7) can be written

$$\begin{aligned}\mathscr{L} &= \overline{D_\mu(\phi'+v)}D_\mu(\phi'+v) - V(\phi'+v) + \overline{K_{\mu\nu}}K_{\mu\nu} \\ &= \overline{\partial_\mu\phi'}\partial_\mu\phi' + g^2[\,\overline{A_\mu,v}\,]\,[A_\mu,v] - \lambda S(\bar\phi' v)^2 \\ &\quad + 2(\,\overline{\partial_\mu A_\nu}\,\partial_\mu A_\nu - \overline{\partial_\mu A_\nu}\,\partial_\nu A_\mu) \\ &\quad + 2g^2S([\,\overline{A_\mu,\phi'}\,]\,[A_\mu,v]) + g^2[\,\overline{A_\mu,\phi'}\,]\,[A_\mu,\phi'] \\ &\quad - V(v) - \lambda/4(\bar\phi'\phi')^2 - \lambda(\bar\phi'\phi')S(\bar\phi' v) \\ &\quad - 4gS(\,\overline{\partial_\mu A_\nu}\,[A_\mu,A_\nu]) + g^2[\,\overline{A_\mu,A_\nu}\,]\,[A_\mu,A_\nu].\end{aligned} \tag{5.8}$$

The term $\lambda S(\bar\phi' v)^2$ is not forced by Goldstone's theorem to be zero for any component of ϕ' as ϕ', in the unitary gauge, has no component in G. As has already been noted the first-order cross terms have disappeared and a second-order mass term for A_μ has appeared.

VI. POSSIBLE VACUUM VECTORS

As physical interpretation of the quaternion fields is on a component level we will express the Higg's mechanism on a component level. We first look at the vacuum quaternion v in components. The most general possible vacuum quaternion is

$$v = ae_0 + be_1 + ce_2 + de_3, \tag{6.1}$$

with

$$a^2 + b^2 + c^2 + d^2 = -\mu^2/\lambda.$$

The Higgs mechanism finds a gauge transformation O such that $O(\phi)$ is parallel to $\underline{v}$. We then define ϕ' such that $O(\phi) = \phi' + v$. Now global gauge transformations of $O(\phi)$, which have the form $O(\phi) \to qO(\phi)q^{-1}$, where q is not dependent on x, will leave the Lagrangian invariant and will send $\phi' + v$ to $q\phi'q^{-1} + qvq^{-1}$. So we can transform v to $v' = qvq^{-1} = a'e_0 + b'e_1$, where $a'^2 + b'^2 = -\mu^2/\lambda$, by such a global transformation. The unitary gauge condition is invariant under such a transformation and so we might well have chosen v' as v in the first place. We now choose

$$v = ae_0 + be_1, \tag{6.2}$$

where

$$a^2 + b^2 = -\mu^2/\lambda, \tag{6.3}$$

as the general vacuum quaternion.

Now the Lagrangian (5.8) in component form in the unitary gauge for the general vacuum vector and shifted fields to second order is

$$\begin{aligned}\mathscr{L} = \{&\partial_\mu\phi_0'\,\partial_\mu\phi_0' + \partial_\mu\phi_1'\,\partial_\mu\phi_1' + 4g^2A_{\mu 2}A_{\mu 2}b^2 \\ &+ 4g^2A_{\mu 3}A_{\mu 3}b^2 - \lambda(\phi_0'a + \phi_1'b)^2 \\ &+ 2(\partial_\mu A_{\alpha\nu}\,\partial_\mu A_{\alpha\nu} - \partial_\mu A_{\alpha\nu}\,\partial_\nu A_{\alpha\mu})\}e_0.\end{aligned} \tag{6.4}$$

Notice that two components of the gauge boson A_μ have gained mass and two components of the scalar field ϕ' have disappeared. Note also that one component of ϕ', namely $\phi_0'a + \phi_1'b$, has mass, and that, if a is not zero, $\phi_0'a - \phi_1'b$ is massless and if a is zero, ϕ_0' is massless. This last point is not in contradiction to the Higg's mechanism as the massless state is not forced by Goldstone's theorem but occurs independently of it.

Now it is possible to split A_μ into

$$A_\mu = B_\mu + C_\mu, \tag{6.5}$$

where

$$B_\mu \to q_{(x)}B_\mu q_{(x)}^{-1}, \tag{6.6a}$$

and

$$C_\mu \to q_{(x)}C_\mu q_{(x)}^{-1} + (1/g)\partial_\mu(q_{(x)})q_{(x)}^{-1}, \tag{6.6b}$$

under gauge transformations. The transformations (6.6a) of B_μ allow B_μ to have a gauge invariant mass term $m\overline{B}_\mu B_\mu$ added to the Lagrangian. We can make the decomposition (6.5) in two ways: first we may decompose A_μ via existing fields as Finkelstein *et al.*[4] do. This does not change the number of degrees of freedom. Second we may introduce new degrees of freedom and have B_μ and C_μ as totally independent quaternion fields. In either case the Lagrangian is now

$$\mathcal{L} = \overline{D_\mu(\phi' - v)} D_\mu(\phi' + v) - V(\phi' + v)\,\overline{K_{\mu\nu}} K_{\mu\nu} + m\,\overline{B_\mu} B_\mu. \tag{6.7}$$

In the first case take an example and consider, as did Finkelstein *et al.*,[4] the following simple situation. The fields B_μ and C_μ are restricted to satisfy

$$\{B_\mu, \eta\} = 0, \tag{6.8a}$$

and

$$[C_\mu, \eta] = \partial_\mu \eta, \tag{6.8b}$$

where η is an existing field or combination of fields satisfying

$$\eta^2 = -1, \tag{6.9a}$$

and

$$\eta \to q(x)\eta q(x)^{-1}, \tag{6.9b}$$

under gauge transformations. To demonstrate what happens we further particularize the situation to the case where we identify η as $\underline{\phi}'/|\underline{\phi}'|$. Now in the unitary gauge and for the general vacuum vector

$$\phi' = \phi_0' e_0 + \phi_1' e_1. \tag{6.10}$$

From which follows that

$$\eta = e_1, \tag{6.11a}$$

$$B_\mu = B_{2\mu} e_2 + B_{3\mu} e_3, \tag{6.11b}$$

$$C_\mu = C_{1\mu} e_1. \tag{6.11c}$$

Here the Lagrangian (3.7) becomes in component form

$$\begin{aligned}\mathcal{L} = \{&\partial_\mu \phi_0' \,\partial_\mu \phi_0' + \partial_\mu \phi_1' \,\partial_\mu \phi_1' + 4g^2 b^2 B_{2\mu} B_{2\mu} \\ &+ 4g^2 b^2 B_{3\mu} B_{3\mu} + m B_{2\mu} B_{2\mu} + m B_{3\mu} B_{3\mu} \\ &- \lambda(\phi_0 a + \phi_1 b)^2 + 2(\partial_\mu A_{\alpha\nu} \partial_\mu A_{\alpha\nu} \\ &- \partial_\mu A_{\alpha\nu} \partial_\nu A_{\alpha\mu})\} e_0,\end{aligned} \tag{6.12}$$

to second order. This is open to the same interpretation as (6.4). However, this method may produce new interpretations if we can identify η as something that is not parallel to $\underline{v}$.

In the case that we introduce B_μ and C_μ as independent fields, the Lagrangian (6.3) becomes

$$\begin{aligned}\mathcal{L} = \{&\partial_\mu \phi_0' \,\partial_\mu \phi_0' + \partial_\mu \phi_1' \,\partial_\mu \phi_1' - \lambda(\phi_0' a + \phi_1' b)^2 \\ &+ m B_{1\mu} B_{1\mu} + (4b^2 g^2 + m) B_{2\mu} B_{2\mu} \\ &+ (4b^2 g^2 + m) B_{3\mu} B_{3\mu} + 8b^2 g^2 C_{2\mu} B_{2\mu} \\ &+ 8b^2 g^2 C_{3\mu} B_{3\mu} + 4b^2 g^2 C_{2\mu} C_{2\mu} + 4b^2 g^2 C_{3\mu} C_{3\mu} \\ &+ 2(\partial_\mu A_{\alpha\nu} \,\partial_\mu A_{\alpha\nu} - \partial_\mu A_{\alpha\nu} \,\partial_\nu A_{\alpha\mu})\} e_0,\end{aligned} \tag{6.13}$$

to second order. This can be written in a more explicit form

$$\begin{aligned}\mathcal{L} = \{&\partial_\mu \phi_0' \,\partial_\mu \phi_0' + \partial_\mu \phi_1' \,\partial_\mu \phi_1' - \lambda(\phi_0' s + \phi_1' b)^2 \\ &+ (B_{1\lambda}, C_{1\lambda}) \left[2\begin{pmatrix} \overleftarrow{\partial}_\mu \overrightarrow{\partial}_\mu & 0 \\ 0 & 0 \end{pmatrix} \delta_{\lambda k} - \begin{pmatrix} 0 & 0 \\ 0 & \overleftarrow{\partial}_k \overrightarrow{\partial}_\lambda \end{pmatrix} \right. \\ &\left. + \delta_{\lambda k} \begin{pmatrix} m & 0 \\ 0 & 0 \end{pmatrix} \right] \begin{pmatrix} B_{1k} \\ C_{1k} \end{pmatrix} \\ &+ \sum_{\alpha=2,3} (B_{\alpha\lambda}, C_{\alpha\lambda}) \left[2\left(\overleftarrow{\partial}_\mu \overrightarrow{\partial}_\mu \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \delta_{\lambda k} - \overleftarrow{\partial}_k \overrightarrow{\partial}_\lambda \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right) \right. \\ &\left. + \delta_{\lambda k} \begin{pmatrix} 4b^2 g^2 + m & 4b^2 g^2 \\ 4b^2 g^2 & 4b^2 g^2 \end{pmatrix} \right] \begin{pmatrix} B_{\alpha k} \\ C_{\alpha k} \end{pmatrix} \Big\} e_0,\end{aligned} \tag{6.14}$$

to second order. This is reminiscent of the current mixing of Hung and Sakurai.[16] However, we will not examine (6.14) in that light here.

VII. CONCLUSION

SSB and the Higg's mechanism has been successfully produced for classical quaternionic fields. Further, it seems the extension from complex theories is nontrivial. More work is required to produce a Hilbert space formulation of SSB and the Higgs mechanism with the quaternions as the scalar field of the Hilbert space. It is not obvious whether or not the extension to quaternions will be significant there. The gauge boson splitting in Sec. VI especially with reference to current mixing[16] warrants more investigation. As yet we have not examined fermions and it is clear that this will need to be done.[5,6,8,9]

ACKNOWLEDGMENTS

One of us C.G.N. wishes to thank R. Volkas for many useful discussions and J. Nash for assistance with portions of the manuscript.

C.G.N. would also like to thank the University of Melbourne for its financial support.

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[2] T. Kaneno, Prog. Theor. Phys. **23**, 17 (1960).
[3] D. Finkelstein, J. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962).
[4] D. Finkelstein, J. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **4**, 788 (1963).
[5] B. Ward, Nuovo Cimento A **51**, 209 (1979).
[6] U. Wolff, Phys. Lett. A **84**, 89 (1981).
[7] J. Rembielinski, J. Phys. **14** 2609 (1981).
[8] Dz. Chkareuli, JETP Lett. **27**, 557 (1978).
[9] K. Morita, Prog. Theor. Phys. **67**, 1860 (1982).
[10] R. Gatto, in *Proceedings of the International School of Physics "Enrico Fermi,"* LXXI Varenna, edited by M. Ceolin (Academic, New York, 1979).
[11] W. Hamilton, *Lectures on Quaternions* (Hodgers and Smith, Dublin, 1853).
[12] C. Joly, *Manual of Quaternions* (McMillan, London, 1905).
[13] H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1950).
[14] J. Jauch and F. Rohrolich, *The Theory of Photons and Electrons* (Springer, New York, 1976), 2nd ed.
[15] S. Weinberg, Rev. Mod. Phys. **46**, 255 (1974).
[16] P. Hung and J. Sakurai, Nucl. Phys. B **143**, 81 (1978).

Geometric approach to classical improvement of lattice actions

P. Mitra
Saha Institute of Nuclear Physics, 92, Acharya Prafulla Chandra Road, Calcutta 700 009, India

(Received 20 May 1986; accepted for publication 10 September 1986)

The classical improvement program in lattice gauge theory is examined from the geometric point of view. Improved actions are derived for pure gauge theories as well as for fermions.

I. INTRODUCTION

The lattice treatment of quantum chromodynamics has been a very fruitful one. Numerical calculations have given us a surprisingly large amount of information on nonperturbative aspects of the theory. Much of this information, however, has to be regarded as preliminary, because the desired accuracy necessitates values of the lattice spacing that are too small to be handled by existing computers in a reasonable amount of time. To overcome this difficulty, one may modify the lattice action in such a way that the continuum limit remains unchanged but is—loosely speaking—approached faster. If the *dynamical* continuum limit is sought to be approached faster, renormalization group transformations are involved and the modified action has to be calculated on the computer itself. Alternatively, the approach to the *perturbative* or even classical continuum limit may be desired to be hastened; in this case, the new action can be determined analytically. It is this kind of modification or improvement—pioneered by Symanzik[1]—with which we shall be concerned in this paper.

It is fairly simple to adjust the classical Lagrangian in such a way that the leading corrections to the classical continuum limit get canceled. Nearest-neighbor terms have to be supplemented by terms involving next-nearest or even more remote neighbors. In principle, one can continue in this fashion and remove nonleading corrections order by order. All this involves the classical Lagrangian, which may be regarded as the tree-level approximation to the quantum effective Lagrangian. An order by order improvement in the coupling constant also may be carried out. This involves loop calculations and is obviously more complicated than mere classical improvement. In any case, high-order improvement in either the lattice spacing or the coupling constant makes the Lagrangian too unwieldy for practical work, so that only the leading tree-level and one-loop modifications are usually attempted. We shall be even more restrictive and ignore all quantum (loop) corrections.

Classical improvement of pure gauge theory has been discussed very clearly by Weisz and Wohlert[2] on the basis of Lüscher's ideas; see also Grensing and Grensing.[3] A more complicated action has been advocated by Eguchi and Kawamoto,[4] who were really interested in improving the fermion action proposed by Wilson; Hamber and Wu[5] also have considered the Wilson action, as have Sheikholeslami and Wohlert.[6] The other popular action for fermions, viz., the staggered one, has been improved by Mitra.[7]

Our aim in this paper is to discuss the improvement of the classical action from the point of view of differential geometry on the lattice. The standard action for pure gauge theory on the lattice can very easily be written in a geometric language closely paralleling the continuum situation—see, e.g., Mack.[8] Similarly, staggered lattice fermions can be formulated in an elegant geometric way.[9,10] It is these languages that we wish to use in our study of improvement.

In Sec. II, we first rewrite some of the possible improved actions for pure gauge theory, including a new one, in an extension of the available geometric formalism. Then we show that the formalism, without some such extension, cannot naturally incorporate improvement. A modification of the formalism is found to pick out the improved action suggested by Eguchi and Kawamoto.[4]

In Sec. III, we consider fermions. The geometric formalism is used to derive an improved action. It is equivalent to the one found earlier[7] in a different framework. But the earlier derivation left it unclear whether this is the *simplest* improved action for staggered fermions. This defect is remedied in the present derivation.

A concluding discussion is given in Sec. IV.

II. IMPROVEMENT OF THE ACTION FOR THE PURE GAUGE THEORY

The dynamical variables of a pure gauge theory on the lattice are gauge group elements associated with links. We write $U(x,\mu)$ to denote the group element associated with the one-chain $(x,\{\mu\})$, i.e., the link going from the site x to the next site in the μ direction. Introducing the basic one-cochains $d^{x\{\mu\}}$ dual to the basic one-chains $(x,\{\mu\})$, one can construct the one-cochain

$$U=\sum_{x,\mu}U(x,\mu)_R\,d^{x\{\mu\}}, \tag{2.1}$$

which is the analog of the one-form in the continuum describing the gauge field, taking values, however, in the representation R of the gauge group rather than in the Lie algebra. To obtain the field strength, one considers the commutator between parallel transports along two different paths[8]; taking these to be along the sides of an elementary cell, one is led to

$$F_{\mu\nu}(x)=U(x,\mu)_R\,U(x+e_\mu,\nu)_R - U(x,\nu)_R\,U(x+e_\nu,\mu)_R, \tag{2.2}$$

where e_μ stands for translation to the next site in the μ direction. A two-cochain can now be constructed:

$$F=\frac{1}{2}\sum_{x,\mu,\nu}\epsilon_{\mu\nu}F_{\mu\nu}(x)d^{x\{\mu,\nu\}}. \tag{2.3}$$

Here $d^{x\{\mu,\nu\}}$ is the basic two-cochain dual to the basic two-chain $(x,\{\mu,\nu\})$ and may be identified with $dx^\mu\wedge dx^\nu$ up to

 0022-2488/87/020468-04$02.50 © 1987 American Institute of Physics

a sign in the continuum limit, the sign being $\epsilon_{\mu\nu} = \pm 1$, depending on whether μ is less or greater than ν. Using the wedge product of cochains,[10] one can actually write

$$F = U \wedge U . \tag{2.4}$$

The standard lattice action for pure gauge theory is related to the norm of F:

$$\begin{aligned}(F,F) &= \frac{1}{2}\sum_{x,\mu,\nu} \mathrm{tr}\{F^{+}_{\mu\nu}(x)F_{\mu\nu}(x)\} \\ &= 2\,\mathrm{Re}\,\mathrm{tr}\sum_{p}(1_R - U_p) .\end{aligned} \tag{2.5}$$

In the second line, the sum is over all (unoriented) plaquettes p and U_p denotes the product of the link variables around p. The usual lattice action is thus

$$S_{\text{Wilson}} = (1/2g^2)(F,F) . \tag{2.6}$$

If one wishes to describe the standard improved lattice action[2] in the same language, one has to introduce alternatives to the above F and take a suitable combination of the norms of the different cochains. These alternatives are

$$\begin{aligned}F_{+} &= \sum_{\substack{x,\mu,\nu \\ \mu<\nu}} F_{(+)\mu\nu}(x) d^{x\{\mu^2,\nu\}} , \\ F_{-} &= \sum_{\substack{x,\mu,\nu \\ \mu<\nu}} F_{(-)\mu\nu}(x) d^{x\{\mu,\nu^2\}} ,\end{aligned} \tag{2.7}$$

where

$$\begin{aligned}&F_{(+)\mu\nu}(x) \\ &\quad = \tfrac{1}{2}[\,U(x,\mu)_R\,U(x+e_\mu,\mu)_R\,U(x+2e_\mu,\nu)_R \\ &\qquad - U(x,\nu)_R\,U(x+e_\nu,\mu)_R\,U(x+e_\mu+e_\nu,\mu)_R\,] , \\ &F_{(-)\mu\nu}(x) \\ &\quad = \tfrac{1}{2}[\,U(x,\mu)_R\,U(x+e_\mu,\nu)_R\,U(x+e_\mu+e_\nu,\nu)_R \\ &\qquad - U(x,\nu)_R\,U(x+e_\nu,\nu)_R\,U(x+2e_\nu,\mu)_R\,] ,\end{aligned} \tag{2.8}$$

and $(x,\{\mu^2,\nu\})$ indicates a doubled distance in the μ direction (a rectangular cell instead of a square one). The continuum limit of the norm is unchanged if F is replaced by $F_{\pm}$. But the corrections—terms of higher order in the lattice spacing—do change. By taking an appropriate combination of these terms, the first nonleading term (or equivalently, the leading correction) can be canceled. The normalization of the combination is fixed by demanding that the continuum limit be the same as for (F,F). The resulting improved action is

$$S_{\text{improved}} = (1/6g^2)[5(F,F) - (F_+,F_+) - (F_-,F_-)] . \tag{2.9}$$

However, this is not the only way in which the first nonleading piece in (F,F) can be canceled. One can use cochains different from both F and $F_{\pm}$—F corresponds to squares and $F_{\pm}$ to 2×1 rectangles. The next thing to try is obviously F_* corresponding to 2×2 squares:

$$F_* = \frac{1}{2}\sum_{x,\mu,\nu} F_{(*)\mu\nu}(x) d^{x\{\mu^2,\nu^2\}} , \tag{2.10}$$

$$\begin{aligned}F_{(*)\mu\nu}(x) \equiv \tfrac{1}{4}[\,&U(x,\mu)_R\,U(x+e_\mu,\mu)_R\,U(x+2e_\mu,\nu)_R \\ &\times U(x+2e_\mu+e_\nu,\nu)_R - (\mu\leftrightarrow\nu)\,] .\end{aligned} \tag{2.11}$$

If one is willing to use both $F_{\pm}$ and F_*, then conditions of improvement to leading order and normalization are not sufficient to determine all the coefficients. One may think of canceling the leading as well as the first nonleading corrections, but there are two kinds of terms making up the first nonleading correction, and both cannot be canceled by the single degree of freedom available. Hence some other criterion is needed to fix the coefficients. Equation (2.9) follows from the desire to avoid 2×2 squares altogether. One might instead decide to omit the rectangles. This would lead to the improved action

$$S'_{\text{improved}} = (1/6g^2)[4(F,F) - (F_*,F_*)] . \tag{2.12}$$

This action takes less time on the computer than (2.9), mainly because it involves only two terms instead of three. It is true that larger cells are involved here than in the standard case, but in both cases the restriction on the lattice size is that it should be much bigger than 2 in each direction. Thus, from the computational point of view, S'_{improved} is the improved action to be preferred. An improved action involving both 2×1 rectangles and 2×2 squares has been used[4] in the literature:

$$\begin{aligned}S''_{\text{improved}} = (1/18g^2)[&16(F,F) - 4(F_+,F_+) \\ &- 4(F_-,F_-) + (F_*,F_*)] .\end{aligned} \tag{2.13}$$

It has the advantage that the propagator factorizes and thereby makes weak-coupling calculations somewhat easier. We shall show below that it has some aesthetic advantage as well: it can be motivated by the geometric approach!

The three improved actions discussed above have been *expressed* in the cochain language, but the formalism is not used in their *derivation*. The elementary cell is supplemented by more complicated rectangles in a more or less arbitrary way, a two-cochain being then introduced separately for each kind of rectangle. We would like to take a different approach, making the departure from the unimproved action (2.6) less arbitrary and more fundamental. Thus we try to retain (2.6) and even the form (2.4) for F, replacing the cochain U by

$$\begin{aligned}\tilde{U} = \sum_{x,\mu}\{&\lambda_1 U(x,\mu)_R\, d^{x\{\mu\}} \\ &+ (\lambda_2/2)\,U(x,\mu)_R\,U(x+e_\mu,\mu)_R\,d^{x\{\mu^2\}}\} ,\end{aligned} \tag{2.14}$$

in an attempt at improvement. The coefficients λ_1 and λ_2 are to be determined by the conditions of improvement and normalization. First of all we need

$$\begin{aligned}\tilde{F} &\equiv \tilde{U} \wedge \tilde{U} \\ &= \lambda_1^2 F + \lambda_1\lambda_2(F_+ + F_-) + \lambda_2^2 F_* .\end{aligned} \tag{2.15}$$

The norm is

$$\begin{aligned}(\tilde{F},\tilde{F}) = |\lambda_1|^4(F,F) + |\lambda_1|^2|\lambda_2|^2\{&(F_+,F_+) \\ &+ (F_-,F_-)\} + |\lambda_2|^4(F_*,F_*) .\end{aligned} \tag{2.16}$$

Note that all coefficients are non-negative here. Since the leading corrections to the continuum limits of the different terms are all of the same sign, it immediately follows that improvement cannot be achieved in this way.

We therefore have to develop a different formalism for the action. We go back to the unimproved case. In addition

to the cochain U of (2.1), we introduce

$$U^{(-1)} \equiv \sum_{x,\mu} U^{-1}(x,\mu)_R d^{x\{\mu\}} . \tag{2.17}$$

Correspondingly, we define

$$F^{(-1)} \equiv U^{(-1)} \overleftarrow{\wedge} U^{(-1)} , \tag{2.18}$$

where the left-directed arrow on the wedge indicates that the matrices in the cochains are to be taken in the reversed order. Explicitly,

$$F^{(-1)} = \frac{1}{2} \sum_{x,\mu,\nu} \epsilon_{\mu\nu} \{ U^{-1}(x+e_\mu,\nu)_R U^{-1}(x,\mu)_R - U^{-1}(x+e_\nu,\mu)_R U^{-1}(x,\nu)_R \} d^{x\{\mu,\nu\}} . \tag{2.19}$$

Finally, we write

$$S_{\text{Wilson}} = (1/2g^2) \langle F^{(-1)}, F \rangle , \tag{2.20}$$

where the inner product $\langle\ ,\ \rangle$ is bilinear, i.e., dispenses with complex conjugation in the first factor, in contrast to the Hermitian inner product $(\ ,\ )$ used earlier. This construction may look unnecessarily involved in the context of the unimproved action, but something like this is necessary if the sign problem of (2.16) is to be avoided.

Going back to the cochain $\tilde{U}$ of (2.14), we introduce, in analogy with (2.17),

$$\tilde{U}^{(-1)} = \sum_{x,\mu} \{ \lambda_1 U^{-1}(x,\mu)_R d^{x\{\mu\}} + (\lambda_2/2) U^{-1}(x+e_\nu,\mu)_R U^{-1}(x,\mu)_R d^{x\{\mu^2\}} \} . \tag{2.21}$$

Similarly, in analogy with (2.18), we set

$$\tilde{F}^{(-1)} = \tilde{U}^{(-1)} \overleftarrow{\wedge} \tilde{U}^{(-1)} . \tag{2.22}$$

The inner product entering the action is

$$\langle \tilde{F}^{(-1)}, \tilde{F} \rangle = \lambda_1^4 (F,F) + \lambda_1^2 \lambda_2^2 \{ (F_+,F_+) + (F_-,F_-) \} + \lambda_2^4 (F_*,F_*) , \tag{2.23}$$

which is the same as (2.16) except in the crucial matter of the phases of the coefficients. The condition for improvement can be shown to be

$$\lambda_1^4 + 5\lambda_1^2 \lambda_2^2 + 4\lambda_2^4 = 0 , \tag{2.24}$$

while the normalization demands that

$$\lambda_1^4 + 2\lambda_1^2 \lambda_2^2 + \lambda_2^4 = 1 . \tag{2.25}$$

These imply that

$$\lambda_1^4 = \tfrac{16}{9}, \quad \lambda_2^2/\lambda_1^2 = -\tfrac{1}{4} , \tag{2.26}$$

yielding the action S''_{improved} of (2.13). Thus, out of the one-parameter class of classically improved actions obtainable by supplementing the elementary cell with 2×1 rectangles and 2×2 squares, our geometrical approach chooses a specific one, and it is the same one that has been motivated earlier[4] by a factorization property. Unfortunately, it is hard on the computer!

III. IMPROVEMENT OF THE STAGGERED ACTION FOR FERMIONS

A geometrical formalism[9,10] for lattice fermions can be derived from the continuum formalism due to Kähler,[11] which involves four species of fermions. It turns out to be a new representation for staggered fermions on a lattice of doubled spacing.[10] The fields can be written as a combination of p-forms with p running all the way from 0 to 4. On the lattice, this translates to a combination of p-cochains. It may be expanded in basic cochains as

$$\phi = \sum_{x,H} \phi(x,H) d^{xH} , \tag{3.1}$$

where H goes over all subsets of $\{0,1,2,3\}$. In this representation, the Dirac operator takes the form $d-\delta$ or $i(d+\delta)$, where d and δ are the lattice versions of the usual coboundary (exterior derivative) and boundary (adjoint of exterior derivative) operators. Ignoring the mass term, we can write the action for free fermions as

$$S_{\text{staggered}} = (\bar{\phi}, (d-\delta)\phi) , \tag{3.2}$$

where $\bar{\phi}$ is a cochain as in (3.1), but independent of ϕ. The introduction of gauge interactions involves the replacement of d and δ by appropriate covariant versions[12] and will not bother us here. Let us rewrite (3.2) in terms of the components of ϕ and $\bar{\phi}$:

$$S_{\text{staggered}} = \sum_{x,H} \bar{\phi}(x,H) \Big[\sum_{\mu\in H} \epsilon_{\mu,H\setminus\{\mu\}} \{ \phi(x+e_\mu, H\setminus\{\mu\}) - \phi(x,H\setminus\{\mu\}) \} + \sum_{\mu\notin H} \epsilon_{\mu,H} \{ \phi(x,H\cup\{\mu\}) - \phi(x-e_\mu,H\cup\{\mu\}) \} \Big] . \tag{3.3}$$

Here $\epsilon_{\mu,K}$ is equal to -1 raised to a power given by the number of elements in K which are less than μ.

A special symmetry of (3.3) will be very important in our development. If we make the transformations

$$e_\mu \to -e_\mu , \qquad \phi(x,H) \to \phi(x-e_H,H) \tag{3.4}$$

(and similarly for $\bar{\phi}$), both terms in (3.3) change sign. Thus

$$\begin{aligned} &\bar{\phi}(x,H) \{ \phi(x+e_\mu,H\setminus\{\mu\}) - \phi(x,H\setminus\{\mu\}) \} \\ &\quad \to \bar{\phi}(x-e_H,H) \{ \phi(x-e_\mu-(e_H-e_\mu),H\setminus\{\mu\}) - \phi(x-(e_H-e_\mu),H\setminus\{\mu\}) \} \\ &\quad = -\bar{\phi}(x-e_H,H) \{ \phi(x-e_H+e_\mu,H\setminus\{\mu\}) - \phi(x-e_H,H\setminus\{\mu\}) \} . \end{aligned} \tag{3.5}$$

The first and last lines give equal and opposite results when all values of x are summed over. This symmetry is recognized to be the invariance of the action under a formal change of sign of the lattice spacing when one recalls[10] that $\phi(x,H)$ is to be associated with the site $x+\frac{1}{2}e_H$ of a lattice with halved spacing. Under this formal change of sign, the chain (x,H) gets rewritten as $(x+e_H,H)$, i.e., the origin is shifted from one extremity to the opposite one. This transformation of basic chains is equivalent to the replacement of $\phi(x,H)$ by $\phi(x-e_H,H)$ as is clear from (3.1).

To improve the action (3.2), we add similar terms where $(d-\delta)$ is replaced by $(d-\delta)_r$, which is the same as $(d-\delta)$ except that the hopping distance is multiplied by a factor r; thus, in the analog of (3.3), the rth forward and

backward differences will occur instead of the 1st. If this were the only difference between the different terms in the improved action, the linearity of the action (3.2) in $(d-\delta)$ would permit us to rewrite the improved action in the form (3.2) with $(d-\delta)$ replaced by an improved $(d-\delta)$, i.e., a sum of several $(d-\delta)_r$. However, with such a structure it is not possible to preserve the symmetry discussed above. It is necessary to change ϕ and $\bar{\phi}$ when we change r in $(d-\delta)_r$. We take

$$\phi^{(s)} \equiv \sum_{x,H} \phi(x + se_H, H) d^{xH}, \tag{3.6}$$

and consider

$$\begin{aligned}(\bar{\phi}^{(s)}, (d-\delta)_r \phi^{(s)}) &= \frac{1}{r} \sum_{x,H} \bar{\phi}(x + se_H, H) \Bigg[\sum_{\mu \in H} \epsilon_{\mu, H\setminus\{\mu\}} \\ &\times \{\phi(x + re_\mu + s(e_H - e_\mu), H\setminus\{\mu\}) \\ &- \phi(x + s(e_H - e_\mu), H\setminus\{\mu\})\} \\ &+ \sum_{\mu \notin H} \epsilon_{\mu,H} \{\phi(x + s(e_H + e_\mu), H \cup \{\mu\}) \\ &- \phi(x + s(e_H + e_\mu) - re_\mu, H \cup \{\mu\})\}\Bigg]. \end{aligned} \tag{3.7}$$

Under the transformation (3.4), a term in the first piece (with se_H absorbed in x) changes as follows:

$$\begin{aligned}&\bar{\phi}(x,H)\{\phi(x + (r-s)e_\mu, H\setminus\{\mu\}) - \phi(x - se_\mu, H\setminus\{\mu\})\} \\ &\quad \to \bar{\phi}(x - e_H, H)\{\phi(x - (e_H - e_\mu) - (r-s)e_\mu, H \setminus \\ &\qquad \{\mu\}) - \phi(x - (e_H - e_\mu) + se_\mu, H\setminus\{\mu\})\}\end{aligned} \tag{3.8}$$

It is clear that this expression (after the absorption of e_H in x) will be the negative of what it was if

$$r = 2s + 1. \tag{3.9}$$

The same condition ensures the desired transformation of the second piece of (3.7) as well. Thus we see that the symmetry can be preserved if in the different terms s and r are related in the above manner. One immediate observation is that r must be odd. Note also that changing the *sign* of r does not change (3.7).

The usual action (3.2) has $r = 1$ and $s = 0$. To improve it, the simplest thing that can be done is the introduction of a term with $r = 3$ and $s = 1$:

$$S_{\text{improved}} = \lambda_1(\bar{\phi}, (d-\delta)\phi) + \lambda_2(\bar{\phi}^{(1)}, (d-\delta)_3 \phi^{(1)}). \tag{3.10}$$

The parameters λ_1 and λ_2 are to be determined by the conditions of improvement and normalization. Because of the symmetry considered above, second derivatives do not arise in the expansion of (3.10) about its continuum limit. The cancellation of the third derivatives imposes the condition

$$\lambda_1 + \lambda_2 \cdot 3^2 = 0. \tag{3.11}$$

Since normalization demands that

$$\lambda_1 + \lambda_2 = 1, \tag{3.12}$$

it follows that

$$S_{\text{improved}} = \tfrac{9}{8}(\bar{\phi}, (d-\delta)\phi) - \tfrac{1}{8}(\bar{\phi}^{(1)}, (d-\delta)_3 \phi^{(1)}). \tag{3.13}$$

This action is equivalent to the one arrived at earlier[7] in a different representation for staggered fermions. There it was shown in greater detail that this action and its gauge-invariant version effect improvement, but in that approach it was not clear that this is the *simplest* improved action. The first term in (3.13) corresponds to a nearest-neighbor term (also on the lattice with halved spacing) while the second term involves third nearest neighbors, so that the question arises whether improvement can be effected with second nearest neighbors. The present derivation, through the result that r must be odd, demonstrates that second nearest neighbors cannot be used.

IV. CONCLUSION

In this paper we have studied both pure gauge and fermion theories on the lattice, the aim being to look at what are called improved actions from a geometric point of view. In both cases we have come up with specific improved actions. In the pure gauge case, some alternative (nongeometric) actions exist, so that our work may be invoked to motivate one particular choice. In the case of (staggered) fermions, our action coincides with the only improved action known. Now this action, involving third nearest neighbors instead of second nearest ones, is not simple from a computational point of view. Our work shows that a simpler improved action cannot be found. In both cases we have restricted ourselves to improvement at the leading order. Extension to higher orders is straightforward.

[1] K. Symanzik, in *Mathematical Problems in Theoretical Physics*, edited by R. Schrader, R. Seiler, and D. A. Uhlenbrock (Springer, Berlin, 1982).
[2] P. Weisz and R. Wohlert, Nucl. Phys. B **236**, 397 (1984).
[3] D. Grensing and G. Grensing, Phys. Rev. D **30**, 2669 (1984).
[4] T. Eguchi and N. Kawamoto, Nucl. Phys. B **237**, 609 (1984).
[5] H. W. Hamber and C. M. Wu, Phys. Lett. B **133**, 351 (1983).
[6] B. Sheikholeslami and R. Wohlert, Nucl. Phys. B **259**, 572 (1985).
[7] P. Mitra, "Staggered fermions: Continuum limit and improvement," Saha Institute Report SINP/TNP/85-36 (unpublished).
[8] G. Mack, Nucl. Phys. B **235**, 197 (1984).
[9] J. M. Rabin, Nucl. Phys. B **201**, 315 (1982).
[10] P. Becher and H. Joos, Z. Phys. C **15**, 343 (1982).
[11] E. Kähler, Rend. Mat. **21**, 425 (1962).
[12] P. Mitra, Nucl. Phys. B **227**, 349 (1983).

Some remarks on scattering theory in supersymmetric quantum systems

Asao Arai[a)]
Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan

(Received 20 May 1986; accepted for publication 24 September 1986)

First it is shown in an abstract mathematical framework of supersymmetric quantum theory that, under some conditions for the perturbation, the wave and the scattering operators on bosonic states have some connections with those on fermionic ones. In particular, a unitary equivalence between the two scattering operators is proved. Then, a class of supersymmetric quantum mechanical models on $\mathbb{R}^n$ is discussed and the abstract general results are applied to obtain some more explicit results on the supersymmetric potential scattering.

I. INTRODUCTION

In this paper we study some mathematical aspects of scattering in supersymmetric quantum theory (SSQT).[1-3] The main interest is to understand on a mathematically rigorous basis what the supersymmetry implies for scattering quantities such as wave and scattering operators. On a level of model studies, such an attempt has recently been made,[4] where supersymmetric quantum mechanical models in one and three dimensions are discussed, although their discussions are not completely mathematically rigorous. In the present paper, however, we first consider SSQT in an abstract framework and derive some general results on the scattering theory, which include the results in Ref. 4 as special cases. Then, we discuss a class of models of supersymmetric quantum mechanics (SSQM) on $\mathbb{R}^n$. We hope that our approach would clarify in some mathematical generality the role of supersymmetry in scattering phenomena.

The outline of this paper is as follows. In Sec. II, we start with a mathematically precise definition of SSQT with a functional analytical flavor.[2,5] Then, after preparing some terminologies from the standard abstract scattering theory (e.g., Refs. 6 and 7), we state the main result (Theorem 2.2). In particular, we note that, under some conditions for the perturbation, a unitary equivalence between the scattering operator on bosonic states and that on fermionic ones holds. This theorem is proved by a series of lemmas. An essential point lies in the fact that the bosonic and the fermionic part of the total Hamiltonian H restricted to the orthogonal subspace of the kernel of H are unitarily equivalent (Lemma 2.3), which is due to the supersymmetry. In Sec. III, we consider a class of models of SSQM on $\mathbb{R}^n$, which contains the one-dimensional Witten model.[1,3-5] By applying the general results, we can get an explicit relation between the T-matrix elements on bosonic states and those on fermionic ones. Furthermore, the spectrum of the Hamiltonian is exactly identified.

II. WAVE AND SCATTERING OPERATORS IN AN ABSTRACT SSQT

We first give a mathematically precise definition of SSQT in an abstract framework.[2,5]

Definition 2.1: Let N be a positive integer. Then, an SSQT with N-supersymmetry is a quadruple $\{\mathbf{H}, \{Q_i\}_{i=1}^N, H, N_F\}$ consisting of a complex Hilbert space $\mathbf{H}$, a set of self-adjoint operators $\{Q_i\}_{i=1}^N$ (the "supercharges"), self-adjoint operators H (the "supersymmetric Hamiltonian"), and N_F the "Fermion number operator"), which has the following properties.

(a) $\mathbf{H}$ is decomposed into two mutually orthogonal closed subspaces $\mathbf{H}_B$ and $\mathbf{H}_F$:

$$\mathbf{H} = \mathbf{H}_B \oplus \mathbf{H}_F . \tag{2.1}$$

(b) For all $i,j = 1,...,N$,

$$Q_i^2 = Q_j^2 , \tag{2.2}$$

and H is written as

$$H = Q_i^2 , \quad i = 1,...,N . \tag{2.3}$$

(c) For all Ψ in $\mathbf{H}_B$ (resp. $\mathbf{H}_F$),

$$N_F\Psi = +\Psi \ \text{(resp. } -\Psi) . \tag{2.4}$$

(d) N_F: $D(Q_i) \to D(Q_i)$, $N_F Q_i + Q_i N_F = 0$,

$$i = 1,...,N , \tag{2.5}$$

on $D(Q_i)$, where $D(Q_i)$ denotes the domain of Q_i.

(e) $Q_iQ_j + Q_jQ_i = 0$, $i \neq j, i,j = 1,...,N$,

on $D(Q_iQ_j) \cap D(Q_jQ_i)$.

In physics literature (e.g., Refs. 1 and 3), elements in the Hilbert space $\mathbf{H}_B$ (resp. $\mathbf{H}_F$) are called *bosonic* (resp. *fermionic*) *states.*

From the above definition, one can easily see that H is non-negative and reduced by $\mathbf{H}_B$ and $\mathbf{H}_F$ (Ref. 5). We shall denote by H_B (resp. H_F) the reduced part of H to the subspace $\mathbf{H}_B$ (resp. $\mathbf{H}_F$). The operator H_B (resp. H_F) is called the *bosonic* (resp. *fermionic*) *part* of the total supersymmetric Hamiltonian H. It is also obvious from (c) and (d) in Definition 2.1 that

$$Q_i: D(Q_i) \cap \mathbf{H}_B \ \text{(resp. } \mathbf{H}_F) \to \mathbf{H}_F \ \text{(resp. } \mathbf{H}_B) , \tag{2.6}$$

and

$$\text{Ker } Q_i = \text{Ker } H , \quad i = 1,...,N . \tag{2.7}$$

In what follows, we concentrate our attention on properties of H and hence we write Q_i simply as Q:

$$H = Q^2 . \tag{2.8}$$

[a)] Address after 1 June 1986: Department of Mathematics, Hokkaido University, 060 Sapporo, Japan.

 0022-2488/87/020472-05$02.50 © 1987 American Institute of Physics

Remark: In the following discussions, we do not use property (e) in Definition 2.1.

In order to consider scattering theory, we first have to prepare some terminologies from the standard abstract scattering theory (e.g., Refs. 6 and 7). Given a self-adjoint operator A, we denote by $\mathbf{H}_{ac}(A)$ [resp. $P_{ac}(A)$] the subspace of absolute continuity with respect to A [resp. the orthogonal projection onto $\mathbf{H}_{ac}(A)$]. Let A, B be two self-adjoint operators acting in Hilbert spaces $\mathbf{H}_2$ and $\mathbf{H}_1$, respectively, and J be a bounded operator from $\mathbf{H}_1$ to $\mathbf{H}_2$. Then, wave operators are defined by

$$W_{\pm}(A,B;J) = \text{s-lim}_{t\to\pm\infty} e^{itA} J e^{-itB} P_{ac}(B) , \tag{2.9}$$

provided that the rhs exists. The scattering operator $S(A,B;J)$ is given by

$$S(A,B;J) = W_{+}(A,B;J)^{*} W_{-}(A,B;J) . \tag{2.10}$$

If $\mathbf{H}_1 = \mathbf{H}_2$ and $J = I$ (identity), then we write

$$W_{\pm}(A,B) = W_{\pm}(A,B;I) , \tag{2.11}$$

$$S(A,B) = S(A,B;I) . \tag{2.12}$$

If

$$\text{Ran}\, W_{\pm}(A,B) = \mathbf{H}_{ac}(A) ,$$

then $W_{\pm}(A,B)$ are said to be complete. It is well known that, if $W_{\pm}(A,B)$ are complete, then $W_{\pm}(B,A)$ exist and are complete.[6,7]

As a "free" part in the supersymmetric quantum scattering system, we take another SSQT, $\{\mathbf{H}, \{Q_{0i}\}_{i=1}^{N}, H_0, N_F\}$, where $\mathbf{H}$ and N_F are identical with the previous ones. We shall write Q_{0i} simply as Q_0 and denote by H_{0B} (resp. H_{0F}) the bosonic (resp. fermionic) part of H_0. Let $\{E(\lambda)\}$ [resp. $\{E_0(\lambda)\}$] be the spectral family of Q (resp. Q_0) and define

$$U = I - E(0) - E(-0) , \tag{2.13}$$

$$U_0 = I - E_0(0) - E_0(-0) . \tag{2.14}$$

Then, our main result is the following.

Theorem 2.2: Suppose that (a) $W_{\pm}(H,H_0)$ exist and are complete; (b) $D(Q_0) \subset D(Q)$ and, for all $\Psi \in \mathbf{H}_{ac}(H_0)$ and for all $\epsilon > 0$, there exist functions $F_{\pm}$ on $[\epsilon,\infty)$ such that, for all $s \in [\epsilon,\infty)$ and $t \in \mathbf{R}$,

$$\|(Q - Q_0)(Q_0 \pm is)^{-1} e^{-itH_0}\Psi\| \leqslant F_{\pm}(s) , \tag{2.15}$$

and

$$\int_{\epsilon}^{\infty} \frac{1}{s} F_{\pm}(s)\, ds < \infty ;$$

and (c) for all $\Psi \in \mathbf{H}_{ac}(H_0) \cap D(H_0)$,

$$\lim_{t\to\pm\infty} \|(Q - Q_0) e^{-itH_0}\Psi\| = 0 .$$

Then, we have

$$W_{\pm}(H_B,H_{0B}) = U W_{\pm}(H_F,H_{0F}) U_0 P_{ac}(H_{0B}) , \tag{2.16}$$

$$S(H_B,H_{0B}) = U_0 S(H_F,H_{0F}) U_0 P_{ac}(H_{0B}) . \tag{2.17}$$

Remark: (1) By symmetry, formulas (2.16) and (2.17) with "B(F)" in place of "F(B)" also hold. The same is true for formulas appearing in what follows.

(2) Formula (2.17) shows that the scattering operator on the bosonic states is unitarily equivalent to that on the fermionic ones. In other words, (2.17) is equivalent to that the total scattering operator

$$S = \begin{pmatrix} S(H_B,H_{0B}) & 0 \\ 0 & S(H_F,H_{0F}) \end{pmatrix} ,$$

is commutative with U_0 on the space of absolute continuity of H_0. Since U_0 may be regarded as a "normalized" generator of the supersymmetry in the "free" system, this may be a rigorous version of the statement that the supersymmetry is a symmetry consistent with scattering (cf. Ref. 8).

(3) Condition (2.15) can be weakened: It is sufficient for (2.15) to hold for all Ψ in a dense domain D in $\mathbf{H}_{ac}(H_0)$ (see Proof of Lemma 2.6).

To prove Theorem 2.2, we need some lemmas. A fundamental fact in SSQT is given by the following.

Lemma 2.3: Let U be given by (2.13). Then, U is a partial isometry from $\mathbf{H}_B$ (resp. $\mathbf{H}_F$) to $\mathbf{H}_F$ (resp. $\mathbf{H}_B$) with $\text{Ker}\, U = \text{Ker}\, Q$ and $\text{Ran}\, U = \overline{\text{Ran}\, Q}$ and we have

$$H_B = U H_F U , \tag{2.18}$$

on $\mathbf{H}_B \cap D(H_B)$.

Proof: Since Q is self-adjoint, it has the polar decomposition as

$$Q = U|Q| ,$$

(see, e.g., Ref. 6, p. 358). Noting that $|Q| = H^{1/2}$, one can easily see that $|Q|$ is also reduced by $\mathbf{H}_B$ and $\mathbf{H}_F$. Therefore (2.6) implies that

$$U: \text{Ran}|Q| \cap \mathbf{H}_B \ (\text{resp. } \mathbf{H}_F) \to \mathbf{H}_F \ (\text{resp. } \mathbf{H}_B) ,$$

and is isometric. Since $\overline{(\text{Ran}|Q|)}^{\perp} = \text{Ker}|Q| = \text{Ker}\, Q$ and $\text{Ker}\, Q = \text{Ker}\, U$, U gives a partial isometry from $\mathbf{H}_B$ (resp. $\mathbf{H}_F$) to $\mathbf{H}_F$ (resp. $\mathbf{H}_B$). Using the self-adjointness of Q and U, we have

$$H = QQ^{*} = U|Q|\,|Q|U = UHU$$

on $D(H)$. Then, the restriction of both sides to $\mathbf{H}_B \cap D(H_B)$ gives (2.18). □

Remarks: (1) Equation (2.18) implies remarkable properties on the spectrum of H_B and H_F. Namely we have

$$\sigma(H)\setminus\{0\} = \sigma(H_B)\setminus\{0\} = \sigma(H_F)\setminus\{0\} , \tag{2.19}$$

and

$$\sigma_p(H)\setminus\{0\} = \sigma_p(H_B)\setminus\{0\} = \sigma_p(H_F)\setminus\{0\} , \tag{2.20}$$

with the equal multiplicity of each corresponding strictly positive eigenvalue between H_B and H_F, where $\sigma(A)$ [resp. $\sigma_p(A)$] denotes the (resp. point) spectrum of the operator A. [Equation (2.20) is actually well known on a formal level in the physics literature and it has been expected or assumed that (2.19) is also true. It seems, however, that, in physics literature, no rigorous proofs have been given for (2.19) so far.]

(2) If $\text{Ker}\, Q = \{0\}$, then U is unitary and (2.18) represents a unitary equivalence between H_B and H_F.

Lemma 2.4: Suppose that $W_{\pm}(H,H_0)$ and $W_{\pm}(H_0,H)$ exist. Then, $W_{\pm}(H_0,H_0;U)$ exist and we have

$$W_{\pm}(H_0,H_0;U) = W_{\pm}(H_0,H) U W_{\pm}(H,H_0) . \tag{2.21}$$

Proof: Let Ψ be in $\mathbf{H}_{ac}(H_0)$ and put

$$\Psi(t) \equiv e^{itH_0} U e^{-itH_0}\Psi$$
$$= e^{itH_0} e^{-itH} U e^{itH} e^{-itH_0}\Psi ,$$

where we have used the commutativity of U with H. Therefore we have

$$\|\Psi(t) - W_{\pm}(H_0,H)UW_{\pm}(H,H_0)\Psi\|$$
$$\leqslant \|[e^{itH} e^{-itH_0} - W_{\pm}(H,H_0)]\Psi\|$$
$$+ \|[e^{itH_0} e^{-itH} - W_{\pm}(H,H_0)]UW_{\pm}(H,H_0)\Psi\| . \tag{2.22}$$

On the other hand, it is well known[6,7] that $W_{\pm}(H,H_0)\Psi \in \mathbf{H}_{ac}(H)$. Since U maps $\mathbf{H}_{ac}(H)$ onto $\mathbf{H}_{ac}(H)$, $UW_{\pm}(H,H_0)\Psi$ is in $\mathbf{H}_{ac}(H)$. Hence the rhs of (2.22) converges to zero as $t \to \pm\infty$. □

Lemma 2.5: Under the same assumption as in Lemma 2.4, we have

$$W_{\pm}(H_B,H_{0B}) = UW_{\pm}(H_F,H_{0F})W_{\pm}(H_{0F},H_{0B};U) , \tag{2.23}$$

$$S(H_B,H_{0B})$$
$$= W_{+}(H_{0B},H_{0F};U)S(H_F,H_{0F})W_{-}(H_{0F},H_{0B};U) . \tag{2.24}$$

Proof: Using (2.18) and Lemma 2.4, we have

$$UW_{\pm}(H_B,H_{0B})$$
$$= \text{s-lim}_{t\to\pm\infty} e^{itH_F} e^{-itH_{0F}} e^{itH_{0F}} U e^{-itH_{0B}} P_{ac}(H_{0B})$$
$$= W_{\pm}(H_F,H_{0F})W_{\pm}(H_{0F},H_{0B};U) ,$$

which imply (2.23). Eq. (2.24) is a direct consequence of (2.23). □

Lemma 2.6: Under the same assumption as in Theorem 2.2, we have

$$W_{\pm}(H_0,H_0;U) = U_0 P_{ac}(H_0) . \tag{2.25}$$

In particular,

$$W_{\pm}(H_{0B},H_{0F};U) = U_0 P_{ac}(H_{0F}) . \tag{2.26}$$

Proof: It is easy to see that the linear subspace

$$D_{ac} = \{Q_0\Phi \mid \Phi \in D(H_0) \cap \mathbf{H}_{ac}(H_0)\}$$

is dense in $\mathbf{H}_{ac}(H_0)$ and hence we need only to prove (2.25) on D_{ac}. Let $\Psi = Q_0\Phi \in D_{ac}$. Noting that

$$U = \text{s} - \lim_{\substack{\rho\to\infty \\ \delta\to 0}} \frac{2}{\pi}\int_{\delta}^{\rho} Q(Q^2+s^2)^{-1}\,ds$$

(see Ref. 6, p. 359), we can write as

$$\Psi(t) \equiv (e^{itH_0} U e^{-itH_0} - U_0)\Psi$$
$$= (2/\pi)\{I^{\epsilon}(t;Q) + I^{\epsilon}(0;Q_0)\} - (1/\pi)I^{\epsilon}(t) ,$$

where $\epsilon > 0$ is arbitrary and

$$I^{\epsilon}(t;A) = s - \int_0^{\epsilon} ds\, e^{itH_0} A(A^2+s^2)^{-1} e^{-itH_0}\Psi ,$$

$$I^{\epsilon}(t) = s - \int_{\epsilon}^{\infty} ds\, e^{itH_0}$$
$$\times\{(Q+is)^{-1}(Q-Q_0)(Q_0+is)^{-1}$$
$$+ (Q-is)^{-1}(Q-Q_0)(Q_0-is)^{-1}\} e^{-itH_0}\Psi .$$

We have

$$\|I^{\epsilon}(t)\| \leqslant \int_{\epsilon}^{\infty} ds \frac{1}{s}\{\|(Q-Q_0)(Q_0+is)^{-1}e^{-itH_0}\Psi\|$$
$$+ \|(Q-Q_0)(Q_0-is)^{-1}e^{-itH_0}\Psi\|\} .$$

The integrand in the rhs is dominated by $s^{-1}[F_+(s) + F_-(s)]$ for $s \in [\epsilon,\infty)$, which, by the assumption, is integrable on $[\epsilon,\infty)$. Therefore, by the dominated convergence theorem and condition (c) in Theorem 2.2, we get

$$\lim_{t\to\pm\infty} \|I^{\epsilon}(t)\| = 0 .$$

Thus we have

$$\lim_{t\to\pm\infty} \|\Psi(t)\|$$
$$\leqslant (2/\pi)\left\{\overline{\lim_{t\to\pm\infty}} \|I^{\epsilon}(t;Q)\| + \|I^{\epsilon}(0;Q_0)\|\right\} . \tag{2.27}$$

One can easily see that

$$\|I^{\epsilon}(t;Q)\|^2 = \int_{\mathbf{R}} \left[\arctan\frac{\epsilon}{|\lambda|}\right]^2 d(E(\lambda)\Phi(t),\Phi(t)) , \tag{2.28}$$

where

$$\Phi(t) = e^{-itH_0}\Psi .$$

By condition (c) in Theorem 2.2, for any $\delta > 0$, there exists a constant $t_0 > 0$ such that, for all $|t| \geqslant t_0$,

$$\|(Q-Q_0)e^{-itH_0}\Phi\| < \delta .$$

Hence, for any $\alpha > 0$, we have

$$\|E(-\alpha,\alpha)\Phi(t)\| < \delta + \|E(-\alpha,\alpha)Qe^{-itH_0}\Phi\|$$
$$\leqslant \delta + \alpha\|\Phi\| .$$

Then, by dividing the integral interval in (2.28) into two parts $|\lambda| \geqslant \alpha$ and $|\lambda| < \alpha$, we have for $|t| \geqslant t_0$

[rhs of (2.28)]
$$\leqslant 2(\pi/2)^2(\delta^2 + \alpha^2\|\Phi\|^2) + (\arctan(\epsilon/\alpha))^2\|\Psi\|^2 .$$

Taking the limit $t \to \pm\infty$ first, $\delta \to 0$ second, $\epsilon \to 0$ third, and finally $\alpha \to 0$, we get

$$\lim_{t\to\pm\infty} \|I^{\epsilon}(t;Q)\| = 0 . \tag{2.29}$$

It is obvious that

$$\lim_{\epsilon\to 0} \|I^{\epsilon}(0;Q_0)\| = 0 . \tag{2.30}$$

From (2.27), (2.29), and (2.30), we obtain

$$\lim_{t\to\pm\infty} \|\Psi(t)\| = 0 ,$$

which is the desired result. □

Theorem 2.2 now follows easily from Lemmas 2.5 and 2.6.

III. EXAMPLES

Let $n,p \geqslant 1$ be positive integers and let us consider the operator of Dirac type

$$L = i\gamma^{\mu}\partial_{\mu} + i\Phi(x) , \tag{3.1}$$

acting in the Hilbert space

$$\mathbf{K} = \overset{p}{\oplus} L^2(\mathbb{R}^n, dx) . \tag{3.2}$$

Here $\{\gamma^\mu\}_{\mu=1}^n$ is a family of $p \times p$ Hermitian matrices ("gamma matrices") satisfying the Euclidean Dirac algebra

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\delta^{\mu\nu}, \quad \mu,\nu = 1,\ldots,n ,$$

and $\Phi(x)$ is a real-valued function on $\mathbb{R}^n$. The operator L is considered by Callias[9] from a different context. Throughout this section, we assume for simplicity the following statement.

(A) Φ is bounded and differentiable on $\mathbb{R}^n$ with its partial derivatives being also bounded on $\mathbb{R}^n$.

By the boundedness of Φ, L is closed with domain $D(L) = \mathbf{H}_1(\mathbb{R}^n;\mathbb{C}^p)$ (the first Sobolev space of C^p-valued tempered distributions) and the adjoint is given by

$$L^* = i\gamma^\mu \partial_\mu - i\Phi(x) , \tag{3.3}$$

with $D(L^*) = D(L)$.

Let $\mathbf{H}_B$ and $\mathbf{H}_F$ be copies of $\mathbf{K}$ and put

$$\mathbf{H} = \mathbf{H}_B \oplus \mathbf{H}_F . \tag{3.4}$$

Further, let

$$Q_1 = \begin{pmatrix} 0 & L^* \\ L & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & iL^* \\ -iL & 0 \end{pmatrix}. \tag{3.5}$$

Then, Q_1 and Q_2 are self-adjoint with $D(Q_1) = D(Q_2) = D(L) \oplus D(L^*)$ and satisfy

$$Q_1 Q_2 + Q_2 Q_1 = 0 , \tag{3.6}$$

on $D(L^*L) \oplus D(LL^*)(= D(Q_1Q_2) = D(Q_2Q_1))$. We also have

$$H \underset{\text{def}}{\equiv} Q_1^2 = Q_2^2 = \begin{pmatrix} L^*L & 0 \\ 0 & LL^* \end{pmatrix}. \tag{3.7}$$

Finally, we define

$$N_F = \begin{pmatrix} 1_p & 0 \\ 0 & -1_p \end{pmatrix}, \tag{3.8}$$

with 1_p being the $p \times p$ identity matrix. Then, we conclude that the quadruple $\{\mathbf{H},\{Q_1,Q_2\},H,N_F\}$ with $\mathbf{H},\{Q_1,Q_2\},H$, and N_F given by (3.4), (3.5), (3.7), and (3.8), respectively, is an SSQT with $N = 2$ supersymmetry. [Conditions (c) and (d) in Definition 2.1 are easily checked.]

From (3.7), we have

$$H_B = L^*L = H_{0B} + V_B , \tag{3.9}$$

$$H_F = LL^* = H_{0F} + V_F , \tag{3.10}$$

with

$$H_{0B} = H_{0F} = -\Delta , \tag{3.11}$$

$$V_{\substack{F \\ B}} = \pm \gamma^\mu \partial_\mu \Phi + \Phi^2 , \tag{3.12}$$

where Δ is the n-dimensional Laplacian. The total Hamiltonian is written as

$$H = H_0 + V , \tag{3.13}$$

with

$$H_0 = \begin{pmatrix} -\Delta & 0 \\ 0 & -\Delta \end{pmatrix}, \quad V = \begin{pmatrix} V_B & 0 \\ 0 & V_F \end{pmatrix}. \tag{3.14}$$

Let

$$Q_0 = \begin{pmatrix} 0 & i\gamma^\mu \partial_\mu \\ i\gamma^\mu \partial_\mu & 0 \end{pmatrix}. \tag{3.15}$$

We must check the conditions in Theorem 2.2 for the quantum scattering system $\{H,H_0\}$ described above. For the existence and the completeness of the wave operators, we employ a standard result in potential scattering (see, e.g., Ref. 10).

Lemma 3.1: Let $\chi_R(x)$ be the characteristic function of the set $\{x \in \mathbb{R}^n \mid |x| \geqslant R\}$ and suppose that

$$h(R) = \| V(H_0 + 1)^{-1} \chi_R \|$$

is in $L^1((0,\infty),dx)$ as a function of $R > 0$. Then, $W_\pm(H,H_0)$ exist and are complete.

Remark: If

$$\|V(x)\| \leqslant c(1 + |x|)^{-\delta} ,$$

with some constants $c > 0$ and $\delta > 1$, then V satisfies the above assumption.

Lemma 3.2: Suppose, in addition to (A), that Φ is in $L^2(\mathbb{R}^n,dx)$. Then (a) for all $\epsilon > 0$, we have

$$\int_\epsilon^\infty ds \frac{1}{s} \| (Q - Q_0)(Q_0 \pm is)^{-1} \| < \infty ;$$

and (b) for all $\Psi \in \mathbf{H}$, we have

$$\lim_{t \to \pm\infty} \| (Q - Q_0) e^{-itH_0} \Psi \| = 0 .$$

Proof: (a) Since $(Q - Q_0)$ is bounded and Q_0 is self-adjoint, we have

$$\| (Q - Q_0)(Q_0 \pm is)^{-1} \| < c/s ,$$

with some constant $c > 0$. Hence we get the desired result.

(b) It is sufficient to prove that

$$\lim_{t \to \pm\infty} \| \Phi e^{it\Delta} f \| = 0 , \quad f \in \mathbf{K} . \tag{3.16}$$

By using the explicit representation of the integral kernel of $\exp(it\Delta)$, one can easily see that

$$\| \Phi e^{it\Delta} f \| \leqslant [1/(4\pi t)^{n/2}] \| \Phi \|_2 \| f \|_1 .$$

Hence, for all f in $L^1 \cap L^2(\mathbb{R}^n,dx;\mathbb{C}^p)$, we have (3.16). By a limiting argument, we can extend the result to all f in $\mathbf{K}$. □

Thus, under conditions in Lemmas 3.1 and 3.2 for Φ, assumptions in Theorem 2.2 are satisfied and we have, in particular,

$$S(H_B, -\Delta) = u_0 S(H_F, -\Delta) u_0 P_{ac}(-\Delta) , \tag{3.17}$$

where the operator u_0 on $\mathbf{K}$ is defined by

$$(u_0 f)\hat{}(k) = (-\gamma^\mu k_\mu / |k|) \hat{f}(k) , \quad k \in \mathbb{R}^n , \quad f \in \mathbf{K} , \tag{3.18}$$

with $\hat{}$ denoting the Fourier transform. Eq. (3.17) gives an explicit relation between the corresponding T-matrices:

$$T_B(k,k') = (\gamma^\mu k_\mu / |k|) T_F(k,k') \gamma^\mu k'_\mu / |k'| ,$$
$$k,k' \in \mathbb{R}^n \quad (k^2 = k'^2) , \tag{3.19}$$

where T_B (resp. T_F) is the T-matrix defined from $S(H_B, -\Delta)$ [(resp. $S(H_F, -\Delta)$]. (For the definition of T-matrix, see, e.g., Ref. 7.) In particular, in the case $n = 1$ where $\gamma^1 = 1$, one has

$$T_B(k,k') = (\operatorname{sgn} k)(\operatorname{sgn} k') T_F(k,k') ,$$
$$k,k' \in \mathbb{R} , \quad k^2 = k'^2 , \tag{3.20}$$

which gives the result in Ref. 4 with respect to the one-dimensional model.

As a by-product, under the same assumptions for Φ, we have proved that

$$\sigma(-\Delta \pm \gamma^{\mu}\partial_{\mu}\Phi + \Phi^2) = \sigma_{ac}(-\Delta \pm \gamma^{\mu}\partial_{\mu}\Phi + \Phi^2) = [0,\infty), \tag{3.21}$$

where $\sigma_{ac}(A)$ denotes the absolute continuous spectrum of the operator A. This follows from (i) that, by the completeness of the wave operators, $H_B \upharpoonright \mathbf{H}_{ac}(H_B)$ (resp. $H_F \upharpoonright \mathbf{H}_{ac}(H_F)$ is unitarily equivalent to $-\Delta \upharpoonright \mathbf{H}_{ac}(-\Delta)$, (ii) that $\sigma(-\Delta) = \sigma_{ac}(-\Delta) = [0,\infty)$, and (iii) that, by the spectral property of the supersymmetric Hamiltonian, $\sigma(H_B) = \sigma(H_F) \subset [0,\infty)$.

Remark: One can extend the discussions in this section to the case of a more general case of Φ and also apply the general result (Theorem 2.2) to other models in SSQM.

IV. CONCLUDING REMARKS

We have derived in a general framework of SSQT some relations between the wave and the scattering operators on bosonic states and those on fermionic ones and seen their concrete implications in a class of models of SSQM on $\mathbb{R}^n$. In some sense, our result on the scattering operators gives a rigorous proof for the consistency of the supersymmetry with scattering. We remark that our analysis in the present paper is restricted essentially to nonrelativistic cases, because, in relativistic quantum field theories, the concept of wave operators does not make sense in general, which is due to Haag's theorem (e.g., Refs. 11 and 12). We hope, however, that, if one employs the Haag–Ruelle scattering theory in the axiomatic quantum field theory (e.g., Ref. 11), one may obtain results similar to those in the present paper.

ACKNOWLEDGMENT

The author would like to thank Dr. S. Kusuoka for pointing out an error in the first draft.

[1] E. Witten, Nucl. Phys. B **188**, 513 (1981).
[2] E. Witten, J. Differential Geom. **17**, 661 (1982).
[3] E. Witten, Nucl. Phys. B **202**, 253 (1982).
[4] J. C. D'olivo, L. F. Urrutia, and F. Zertuche, Phys. Lett. B **159**, 307 (1985).
[5] A. Arai, J. Func. Anal. **60**, 378 (1985).
[6] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.
[7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (Academic, New York, 1979).
[8] R. Haag, J. T. Lopuszauski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975).
[9] C. Callias, Commun. Math. Phys. **62**, 213 (1978).
[10] V. Enss, Commun. Math. Phys. **61**, 285 (1978).
[11] N. N. Bogoliubov, A. A. Logonov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, New York, 1975).
[12] R. F. Streater and A. S. Wightman, PCT, *Spin and Statistics, and All That* (Benjamin, New York, 1964).

Superspace formulation of $N=2$ pseudomechanics and superpotentials

L. Bouquiaux,[a] P. Dauby,[b] and V. Hussin[c]
Université de Liège, Physique théorique et mathématique, Institut de Physique au Sart Tilman, Bâtiment B.5, B-4000 Liege 1, Belgium

(Received 3 July 1986; accepted for publication 1 October 1986)

The general formalism of $N = 2$ pseudomechanics in the superspace is presented for one spatial dimension. Taking into account the (super) symmetries of the Lagrangian, an exhaustive classification of superpotentials into three categories is obtained. The first class contains the harmonic oscillator potential, the free particle constant potential, and the superposition of both with a linear potential. The second one contains the λ^2/q^2 potential and its superposition with the harmonic oscillator potential and a constant one. The third class contains all other potentials. Through Noether's theorem, conserved quantities are associated with (super)symmetry properties, and, for each class, we, respectively, get the following superalgebras: osp(2,2) $\square$ sh(1), osp(2,2), and spl(1,1) $\square$ so(2).

I. INTRODUCTION

In the 1970's supersymmetry[1,2] emerged as one of the most elegant creations in theoretical physics. The associated supersymmetry transformations turn bosons into fermions and conversely, a very interesting property for the description of fundamental interactions in particle physics.

Supersymmetry[2] deals with fusion between space-time and internal symmetries, gauge invariance, spontaneous breaking of symmetry, string picture, local quantum field theory, asymptotic freedom in non-Abelian gauge theories, *but also* with classical and quantum pseudomechanics.[3–7] This last context is very useful for getting acquainted with supersymmetry in advanced realistic field theories and asks for relatively simple but meaningful applications.

In recent years many quantum-mechanical systems have been treated and solved completely within supersymmetry developments. In particular, the supersymmetry of the Pauli equation in the presence of a magnetic monopole has already been pointed out[8] as well as for the cases[9,10] of the $1/r^2$ and Coulomb potentials. The so important harmonic oscillator has also been extensively studied in the context of supersymmetry[6,11–16] by dealing in particular with superconformal[7] transformations.

In fact, within the superfield formalism introduced by Salam, Strathdee,[17] and others,[2] $N = 1$ and $N = 2$ supersymmetric quantum mechanics have been partially developed in a very elegant way.[11,18] In particular, D'Hoker and Vinet[18] have presented the $(N = 1)$ superspace formulation for the dynamical supersymmetry of the Pauli system in the presence of a Dirac magnetic monopole: they discovered[8,18] that Osp(1, 1) is the largest dynamical invariance group of this system. More recently, one of us[14] discussed the particular $(N = 2)$ example of the harmonic oscillator and obtained osp(2, 2) $\square$ sh(1), the semidirect sum of osp(2, 2) and of the Heisenberg superalgebra sh(1), as the largest invariance superalgebra of this system.

This article deals with a general discussion of the one-dimensional $N = 2$ pseudomechanics[3–7] in the superspace formulation. Our purpose is to obtain a classification of *superpotentials* in connection with the associated (super)symmetry properties and to get the corresponding invariance superalgebras. We restrict ourselves to superpotentials depending only on the superposition Z. Let us insist on the fact that we here obtain an *exhaustive* classification of superpotentials leading to only three kinds of invariance superalgebras. This classification corresponds essentially to the results already obtained by Durand,[19] but by another method. Indeed his work, which does *not* refer to the superspace formulation, deals with the study of the invariance of the quantized supersymmetric Schrödinger equation containing explicit potentials following the procedure issued from the works of Niederer[20] and Boyer.[21]

Let us now describe the contents of this paper. Section II is devoted to the construction of all the necessary superspace elements we need for the study of symmetries of the superspace Lagrangian and for the determination of associated conserved quantities. We then give a classification of the superpotentials entering into the theory. In Sec. III, we discuss the symmetries leading to the superalgebras associated with the corresponding systems described by these potentials. In this context we show that potentials such as the $1/q^2$ potential, the harmonic oscillator, the linear, the constant ones, and some superposition of these correspond to precise sets of symmetries leading to the existence of invariance superalgebras such as spl(1, 1) $\square$ so(2), osp(2, 2), and osp(2, 2) $\square$ sh(1).

II. $N=2$-SUPERSPACE FORMULATION AND SUPERPOTENTIALS

The $N = 2$-supersymmetric pseudomechanics[3–7] in one (spatial) dimension can be formulated in superspace.[18] Section A contains the definition of the superspace Lagrangian $\mathscr{L}$ leading to the expected equations of motion and to an easy quantization. In Sec. B, from the study of the symmetries of $\mathscr{L}$, we classify all the potentials of the theory. Finally, the superspace Noether theorem is given in Sec. C.

[a] Chercheur Institut Interuniversitaire des Sciences Nucleaires.
[b] Boursier Institute pour la Recherche Scientifique dans l'Industrie et l'Agriculture (I.R.S.I.A.).
[c] Chargé de recherches Fonds National de la Recherche Scientifique (F.N.R.S.).

 0022-2488/87/020477-08$02.50 © 1987 American Institute of Physics

A. Superspace Lagrangian and equations of motion

Following Ravndal,[11] for example, it is sufficient in order to construct the $N=2$-superspace formulation of pseudomechanics, to extend the ordinary time variable (t) to a new supertime involving two Grassmann time variables (Θ and $\bar{\Theta}$). Then in the $(t, \Theta, \bar{\Theta})$ superspace we call *"superfield"* any arbitrary function of t, Θ, and $\bar{\Theta}$. These superfields are characterized by their *"components,"* which are the coefficients of their Taylor expansion in powers of Θ and $\bar{\Theta}$. In particular, we define the *"superposition"* $Z(t, \Theta, \bar{\Theta})$ as

$$Z(t, \Theta, \bar{\Theta}) = q(t) + i\bar{\Theta}\psi(t) + i\Theta\bar{\psi}(t) + \Theta\bar{\Theta}A(t), \tag{2.1}$$

where the usual position variable $q(t)$ and the function $A(t)$ are bosonic variables while $\psi(t)$ and $\bar{\psi}(t)$ are fermionic ones describing spin degrees of freedom. Let us also introduce two superderivatives

$$D = \partial_{\bar{\Theta}} - i\Theta\,\partial_t, \quad \bar{D} = \partial_{\Theta} - i\bar{\Theta}\,\partial_t, \tag{2.2}$$

satisfying

$$\{D, \bar{D}\} = -2i\,\partial_t. \tag{2.3}$$

Our physical system is governed by the action

$$I = \int dt\, L, \tag{2.4}$$

where L is the usual supersymmetric Lagrangian. In this superspace formulation L is given by

$$L = \int d\Theta\, d\bar{\Theta}\, \mathcal{L}(Z, DZ, \bar{D}Z), \tag{2.5}$$

where $\mathcal{L}(Z, DZ, \bar{D}Z)$ is called the *superspace Lagrangian*. Here we shall suppose that $\mathcal{L}$ reads

$$\mathcal{L}(Z, DZ, \bar{D}Z) = \tfrac{1}{2}\bar{D}Z\,DZ - W(Z), \tag{2.6}$$

where the so-called *superpotential* $W(Z)$ is an arbitrary function of the superposition Z. The Taylor expansion of the superpotential in powers of Θ and $\bar{\Theta}$ is

$$W(Z) = W(q) + i\Theta(W'(q)\,\bar{\psi}) + i\bar{\Theta}(W'(q)\psi) + \Theta\bar{\Theta}(W'(q)A + W''(q)\,\bar{\psi}\psi), \tag{2.7}$$

where the primes denote derivatives with respect to q.

The corresponding Euler–Lagrange equation evidently reads

$$\frac{\partial\mathcal{L}}{\partial Z} - D\frac{\partial\mathcal{L}}{\partial(DZ)} - \bar{D}\frac{\partial\mathcal{L}}{\partial(\bar{D}Z)} = 0, \tag{2.8}$$

leading with $\mathcal{L} \equiv (2.6)$ to the equation

$$\tfrac{1}{2}[D, \bar{D}]Z = \frac{\partial W}{\partial Z}. \tag{2.9}$$

In components, Eq. (2.9) explicitly gives

$$A = W'(q), \tag{2.10a}$$

$$\dot{\bar{\psi}} = iW''(q)\bar{\psi}, \tag{2.10b}$$

$$\dot{\psi} = -iW''(q)\psi, \tag{2.10c}$$

$$\ddot{q} = -W''(q)A - W'''(q)\bar{\psi}\psi. \tag{2.10d}$$

From Eqs. (2.10a) and (2.10d) we have

$$\ddot{q} = -(\tfrac{1}{2}W'(q)^2)' - W'''(q)\bar{\psi}\psi. \tag{2.11}$$

Thus, we can identify $\frac{1}{2}W'(q)^2$ with the classical potential $V(q)$.

This formulation is of course equivalent to the usual one. Indeed, using the components of the superfields, the Lagrangian (2.5) can be written

$$L = \tfrac{1}{2}\dot{q}^2 + \tfrac{1}{2}A^2 - W'A + (\dot{\tfrac{1}{2}})(\bar{\psi}\dot{\psi} - \dot{\bar{\psi}}\psi) - W''\bar{\psi}\psi. \tag{2.12}$$

This Lagrangian is a constrained[22,23,5] one. It gives rise to three primary constraints

$$\varphi_1 = \pi_\psi + (\dot{\tfrac{1}{2}})\bar{\psi}, \tag{2.13a}$$

$$\varphi_2 = \pi_{\bar{\psi}} + (\dot{\tfrac{1}{2}})\psi, \tag{2.13b}$$

$$\varphi_3 = \pi_A, \tag{2.13c}$$

and a secondary one

$$\varphi_4 = A - W', \tag{2.13d}$$

where π_ψ, $\pi_{\bar{\psi}}$, π_A are the momenta conjugated to ψ, $\bar{\psi}$, and A, respectively. All these constraints are second class ones. Thus from the canonical Poisson brackets,[4,5] we can define Dirac brackets. Strongly realizing the second class constraints, the only two brackets different from zero are

$$\{\psi, \bar{\psi}\}_D = -i, \quad \{q, p\}_D = 1, \tag{2.14}$$

where p is the momentum conjugated to q. Once more, realizing the constraints, the total Hamiltonian is

$$H_T = \tfrac{1}{2}p^2 + \tfrac{1}{2}(W')^2 + W''\bar{\psi}\psi, \tag{2.15}$$

and the equations of motion are

$$\dot{p} = \{p, H_T\}_D = -(\tfrac{1}{2}(W')^2 + W''\bar{\psi}\psi)', \tag{2.16a}$$

$$\dot{q} = \{q, H_T\}_D = p, \tag{2.16b}$$

$$\dot{\bar{\psi}} = \{\bar{\psi}, H_T\}_D = iW''\bar{\psi}, \tag{2.16c}$$

$$\dot{\psi} = \{\psi, H_T\}_D = -iW''\psi. \tag{2.16d}$$

These equations are clearly equivalent to (2.10) [and (2.11)]. Let us also notice that the Lagrangian

$$L = \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}(W')^2 + (\dot{\tfrac{1}{2}})(\bar{\psi}\dot{\psi} - \dot{\bar{\psi}}\psi) - W''\bar{\psi}\psi \tag{2.17}$$

would give the same results. It means that it is not necessary to consider a Lagrangian depending explicitly on A. This variable appears in (2.12) only because of the superspace formulation.

Finally, the quantization[4,5] of this supersymmetric pseudomechanics is realized by imposing

$$\{\psi, \bar{\psi}\} = i\hbar\{\psi, \bar{\psi}\}_D = 1, \quad [q, p] = i\hbar\{q, p\}_D = i, \tag{2.18}$$

where we have taken $\hbar = 1$ and where $[\,,\,]$ and $\{\,,\,\}$ denote as usual commutators and anticommutators, respectively.

B. Symmetries of the superspace Lagrangian

Let us now study infinitesimal transformations

$$Z \to Z + \delta Z, \tag{2.19}$$

leaving the action (2.4) unchanged. This happens if the variation of $\mathcal{L}(Z, DZ, \bar{D}Z)$ is given by

$$\delta\mathcal{L} = \partial_t a + \partial_\Theta\alpha + \partial_{\bar{\Theta}}\bar{\alpha}, \tag{2.20}$$

or

$$\delta\mathcal{L} = D\bar{\Lambda} + \bar{D}\Lambda, \tag{2.21}$$

where

$$\bar{\Lambda} = (i/2)\bar{D}(a + i\,\bar{\Theta}\alpha + i\Theta\,\bar{\alpha}) + \bar{\alpha}, \tag{2.22a}$$

and

$$\Lambda = (i/2)D(a + i\,\bar{\Theta}\alpha + i\Theta\,\bar{\alpha}) + \alpha \tag{2.22b}$$

[in fact, it is easy to see that the condition (2.21) is equivalent to the usual condition $\delta L = df/dt$]. In the following, we will consider transformations (2.19) characterized by

$$\delta Z = \delta t\,\partial_t Z + \delta\Theta\,\partial_\Theta Z + \delta\,\bar{\Theta}\,\partial_{\bar{\Theta}} Z + K(t, \Theta, \bar{\Theta})Z + Z_0(t, \Theta, \bar{\Theta}), \tag{2.23}$$

with

$$\delta t = F(t, \Theta, \bar{\Theta}), \quad \delta\Theta = \Xi(t, \Theta, \bar{\Theta}), \quad \delta\,\bar{\Theta} = \bar{\Xi}(t, \Theta, \bar{\Theta}), \tag{2.24}$$

where F, K, and Z_0 are arbitrary even[24] functions of t, Θ, and $\bar{\Theta}$, while Ξ and $\bar{\Xi}$ are odd[24] ones. Using the superderivatives (2.2), Eq. (2.23) becomes

$$\delta Z = \tfrac{1}{2} G\{D, \bar{D}\}Z + \Xi\,\bar{D}Z + \bar{\Xi}DZ + KZ + Z_0, \tag{2.25}$$

with

$$G = iF + \Theta\,\bar{\Xi} + \bar{\Theta}\Xi\,. \tag{2.26}$$

With $\mathcal{L} \equiv$ (2.6), we can then compute

$$\delta\mathcal{L} = \tfrac{1}{2}\,\bar{D}\delta Z\,DZ + \tfrac{1}{2}\,\bar{D}Z\,D\delta Z + \delta Z\,\frac{\partial W(Z)}{\partial Z}. \tag{2.27}$$

Using (2.25), after rather simple calculations, we get

$$\begin{aligned}\delta\mathcal{L} = {} & -\tfrac{1}{4}D(G\,\bar{D}Z\,\bar{D}DZ + \bar{D}K\,Z^2 + 2\bar{D}Z_0\,Z + 2GW'\bar{D}Z) + \tfrac{1}{4}\bar{D}(G\,D\bar{D}Z\,DZ + DK\,Z^2 + 2DZ_0\,Z \\ & + 2GW'DZ) + \tfrac{1}{2}(2K + \bar{D}\Xi + D\,\bar{\Xi})\bar{D}Z\,DZ + \tfrac{1}{4}(2\Xi - DG)(\bar{D}Z\,D\bar{D}Z - 2W'\bar{D}Z) \\ & - \tfrac{1}{4}(2\bar{\Xi} - \bar{D}G)(\bar{D}DZ\,DZ + 2W'DZ) + \tfrac{1}{4}([D, \bar{D}\,]K)Z^2 + (\tfrac{1}{2}\,[D, \bar{D}\,]Z_0)Z - KW'Z - Z_0W'.\end{aligned}$$

We notice that condition (2.21) will be satisfied if and only if

$$\Xi = \tfrac{1}{2}\,DG, \quad \bar{\Xi} = \tfrac{1}{2}\,\bar{D}G, \quad K = -\tfrac{1}{4}\{D, \bar{D}\}G, \tag{2.28}$$

and

$$\tfrac{1}{4}([D, \bar{D}\,]K)Z^2 + \tfrac{1}{2}([D, \bar{D}\,]Z_0)Z - KW'Z - Z_0W' = D\,\bar{\Omega} + \bar{D}\Omega, \tag{2.29}$$

where Ω and $\bar{\Omega}$ are functions of t, Θ, and $\bar{\Theta}$.

These expressions lead to the quantities

$$\begin{aligned}\Lambda = {} & \tfrac{1}{4}\,G\,D\bar{D}Z\,DZ - \tfrac{1}{16}D\{D, \bar{D}\}G\,Z^2 \\ & - \tfrac{1}{2}\,GW'DZ + \tfrac{1}{2}\,DZ_0\,Z + \Omega\end{aligned} \tag{2.30}$$

and

$$\begin{aligned}\bar{\Lambda} = {} & -\tfrac{1}{4}\,G\,\bar{D}Z\,\bar{D}DZ + \tfrac{1}{16}\bar{D}\{D, \bar{D}\}G\,Z^2 \\ & - \tfrac{1}{2}\,GW'\bar{D}Z + \tfrac{1}{2}\,\bar{D}Z_0\,Z + \bar{\Omega}\,.\end{aligned}$$

Let us discuss the system (2.28) and (2.29). We notice that Eqs. (2.28) give the expressions of Ξ, $\bar{\Xi}$, and K in terms of G but are independent on the specific form of the superpotential W while Eq. (2.29) does depend on it and now becomes

$$(MZ + N)W' = D\,\bar{\Omega} + \bar{D}\Omega - M'Z^2 - N'Z, \tag{2.31}$$

where

$$M = \tfrac{1}{4}\{D, \bar{D}\}G, \quad M' = -\tfrac{1}{16}[D, \bar{D}\,]\{D, \bar{D}\}G,$$
$$N = -Z_0, \quad N' = \tfrac{1}{2}\,[D, \bar{D}\,]Z_0\,.$$

Let us now solve Eq. (2.31) according to different possible superpotentials.

If W' is not of the form

$$W' = \omega Z + \mu + (\lambda/Z) + \alpha Z^2, \quad \omega, \mu, \lambda, \alpha \text{ constants}, \tag{2.32}$$

we immediately see that Eq. (2.31) implies the vanishing of all the functions M, M', N, and N'. For such a potential, $\delta\mathcal{L}$ reduces to a superbidivergence if and only if

$$\{D, \bar{D}\}G = 0 \tag{2.33a}$$

and

$$Z_0 = 0. \tag{2.33b}$$

In this case, we have

$$\Omega = \bar{\Omega} = 0. \tag{2.34}$$

If W' has the form (2.32), Eq. (2.31) gives, by identification of the coefficients of the powers of Z,

$$\alpha M = 0, \tag{2.35a}$$
$$\omega M + \alpha N = -M', \tag{2.35b}$$
$$\mu M + \omega N = -N', \tag{2.35c}$$
$$\mu N + \lambda M = D\bar{\Omega} + \bar{D}\Omega, \tag{2.35d}$$
$$\lambda N = 0. \tag{2.35e}$$

From Eqs. (2.35) we immediately deduce that, if $\alpha \neq 0$, we have to impose restrictions (2.33) in order to get Eq. (2.31). So, let us suppose that $\alpha = 0$. Then Eq. (2.35e) suggests considering separately the cases $\lambda = 0$ and $\lambda \neq 0$. Let us begin with $\lambda \neq 0$. Equation (2.35e) gives $N = 0$, which also implies $N' = 0$. Equation (2.35c) tells us that, if $\mu \neq 0$, M has to be zero, and this leads us once more to restrictions (2.33). So we consider $\mu = 0$, i.e., we study the potential

$$W(Z) = (\omega/2)Z^2 + \lambda \ln|Z|, \quad \lambda \neq 0, \omega \text{ arbitrary constants}. \tag{2.36}$$

Equations (2.35) now become

$$\omega M = -M', \quad N = 0, \quad \lambda M = D\bar{\Omega} + \bar{D}\Omega,$$

and tell us that the potential (2.36) satisfies Eq. (2.31) if and only if

$$([D, \bar{D}\,] - 4\omega)\{D, \bar{D}\}G = 0 \tag{2.37a}$$

and

$$Z_0 = 0, \tag{2.37b}$$

with

$$\bar{\Omega} = (\lambda/4)\bar{D}G \quad \text{and} \quad \Omega = (\lambda/4)DG. \tag{2.38}$$

Let us now consider $\lambda = 0$. We study the potential

$$W(Z) = (\omega/2)Z^2 + \mu Z, \quad \omega, \mu \text{ arbitrary constants.} \tag{2.39}$$

Equations (2.35) become

$$\omega M = -M', \quad \mu M + \omega N = -N', \quad \mu N = D\bar{\Omega} + \bar{D}\Omega.$$

So the potential (2.39) satisfies Eq. (2.31) if and only if

$$([D, \bar{D}] - 4\omega)\{D, \bar{D}\}G = 0 \tag{2.40a}$$

and

$$([D, \bar{D}] - 2\omega)Z_0 = (-\mu/2)\{D, \bar{D}\}G, \tag{2.40b}$$

with Ω and $\bar{\Omega}$ determined by

$$D\bar{\Omega} + \bar{D}\Omega = -\mu Z_0. \tag{2.41}$$

In summary, we can classify all the potentials in three categories.

(1) If $W(Z)$ is any potential different from (2.36) and (2.39), we ensure Eq. (2.21) if and only if relations (2.33) are satisfied.

(2) If $W(Z) \equiv$ (2.36), we have (2.21) if and only if relations (2.37) are satisfied. If $\omega = 0$, this case corresponds to the potential $V(q) = \lambda^2/q^2$, while if $\omega \neq 0$, we get the superposition of the λ^2/q^2 potential with the harmonic oscillator potential and a constant one.

(3) If $W(Z) \equiv$ (2.39), we have (2.21) if and only if relations (2.40) and (2.41) are satisfied. If $\mu = 0$, we get the harmonic oscillator potential. If $\omega = 0$, we get the free particle potential. If $\omega \neq 0$ and $\mu \neq 0$, we get the superposition of the harmonic oscillator potential with a linear one and a constant one.

Let us notice that the solutions of Eq. (2.33) are, of course, also solutions of (2.37) and (2.40). This means that, for any potential $W(Z)$, the Lagrangian (2.5) will have the symmetries characterized by Eq. (2.33).

These results are particularly meaningful in connection with Niederer's[20] and Boyer's[21] discussions on classes of potentials within the nonrelativistic "conformal invariance" of the Schrödinger equation. In the following section, we will discuss the symmetries and supersymmetries of such superpotentials.

C. Superspace Noether's theorem

Let us now end this section by giving Noether's theorem in the $N = 2$ superspace formulation, which enables us to get the conserved quantities associated with the symmetries we have found above. It reads as follows: If the transformation (2.19) is a symmetry of the Lagrangian $\mathcal{L}(Z, DZ, \bar{D}Z)$ [i.e., it gives rise to (2.21)], there exists a conserved quantity called a "superbicurrent" defined by

$$\Sigma = \delta Z \frac{\partial L}{\partial \bar{D} Z} - \Lambda, \quad \bar{\Sigma} = \delta Z \frac{\partial L}{\partial DZ} - \bar{\Lambda}, \tag{2.42}$$

such that

$$D\bar{\Sigma} + \bar{D}\Sigma = 0. \tag{2.43}$$

Assuming that the Euler–Lagrange equation (2.9) is satisfied, the proof is straightforward. Let us notice that this superspace theorem is strictly equivalent to its fourth component which is the usual Noether theorem.

With

$$\Sigma = \sigma + \Theta\bar{s} + \bar{\Theta}s + i\Theta\bar{\Theta}\sigma_0 \tag{2.44a}$$

and

$$\bar{\Sigma} = \bar{\sigma} + \Theta\bar{u} + \bar{\Theta}u + i\Theta\bar{\Theta}\bar{\sigma}_0, \tag{2.44b}$$

Eq. (2.43) reads

$$u + \bar{s} - i\Theta(\bar{\sigma}_0 + \dot{\bar{\sigma}}) + i\bar{\Theta}(\sigma_0 - \dot{\sigma}) + i\Theta\bar{\Theta}(\dot{\bar{s}} - \dot{u}) = 0. \tag{2.45}$$

The first three components do not teach us anything. So, in the following, we will only take into account the conservation law given by the fourth component

$$\frac{d}{dt}(\bar{s} - u) = 0. \tag{2.46}$$

III. SUPERPOTENTIALS AND INVARIANCE SUPERALGEBRAS

Through Noether's theorem, we associate conserved quantities with symmetries described by Eqs. (2.33), (2.37), and (2.40). These quantities obey the structure equations of a superalgebra and we discuss separately the superalgebras associated with the various potentials, our choice of presentation going from the smallest to the largest superstructures.

So we will get in Sec. III A the superalgebra spl(1, 1) $\square$ so(2) corresponding to the (super)symmetries characterized by Eqs. (2.33), i.e., the (super)symmetries of the Lagrangian (2.6) including an arbitrary superpotential. In the particular cases of the superpotentials (2.36) (Sec. B) and (2.39) (Sec. C), we will obtain the superalgebras osp(2, 2) and osp(2, 2) $\square$ sh(1), respectively. As expected, we see that the arbitrary superpotential case leads to a supersubalgebra of the superalgebras associated with both the other cases.

A. Arbitrary superpotentials

The symmetries of the Lagrangian (2.6) are associated with the transformations on Z, t, Θ, and $\bar{\Theta}$ given by Eqs. (2.23) and (2.24) such that we have Eqs. (2.28) and (2.29). So we get

$$\begin{aligned} &\delta t = \tfrac{i}{2}[-2G + \Theta\bar{D}G + \bar{\Theta}DG], \\ &\delta\Theta = \tfrac{1}{2}DG, \quad \delta\bar{\Theta} = \tfrac{1}{2}\bar{D}G \end{aligned} \tag{3.1}$$

and

$$\delta Z = \tfrac{1}{2}G\{D, \bar{D}\}Z + \tfrac{1}{2}DG\,\bar{D}Z + \tfrac{1}{2}\bar{D}G\,DZ, \tag{3.2}$$

where G satisfies Eq. (2.33a). The general solution G can then be written

$$G(\Theta, \bar{\Theta}) = ic - 2\alpha\bar{\Theta} - 2\bar{\alpha}\Theta + 2d\Theta\bar{\Theta}, \tag{3.3}$$

where c and id are ordinary real parameters while α and $\bar{\alpha}$ are Grassmann ones. From Eqs. (3.1) with (3.3) we explicitly get

$$\delta t = c + i\alpha\bar{\Theta} + i\bar{\alpha}\Theta, \quad \delta\Theta = \alpha - d\Theta, \quad \delta\bar{\Theta} = \bar{\alpha} + d\bar{\Theta}, \tag{3.4}$$

and from Eq. (3.2), definition (2.1), and Eq. (3.3), we determine

$$\delta q = c\dot{q} + i\alpha\,\overline{\psi} + i\,\overline{\alpha}\psi, \quad \delta\,\overline{\psi} = c\,\dot{\overline{\psi}} - \overline{\alpha}(\dot{q} + iA) - d\overline{\psi},$$
$$\delta\psi = c\,\dot{\psi} - \alpha(\dot{q} - iA) + d\,\psi, \quad \delta A = c\,\dot{A} + \alpha\,\dot{\overline{\psi}} - \overline{\alpha}\,\dot{\psi}. \tag{3.5}$$

The conserved superbicurrent is then given from (2.42), with (2.30) and (2.34), by

$$\Sigma = \tfrac{1}{4}G\,\bar{D}DZ\,DZ + \tfrac{1}{4}DG\,\bar{D}Z\,DZ + \tfrac{1}{2}GW'DZ$$

and (3.6)

$$\overline{\Sigma} = -\tfrac{1}{4}G\,D\bar{D}Z\,\bar{D}Z - \tfrac{1}{4}\bar{D}G\,DZ\,\bar{D}Z + \tfrac{1}{2}GW'\,\bar{D}Z.$$

In components, with the definitions (2.44) the general charge is

$$C = \bar{s} - u = \Sigma|_{\Theta} - \overline{\Sigma}|_{\bar{\Theta}}$$
$$= -i(cH_T + dY + i\,\overline{\alpha}Q + i\alpha\,\bar{Q}), \tag{3.7}$$

where

$$H_T = \tfrac{1}{2}p^2 + \tfrac{1}{2}(W'(q))^2 + \tfrac{1}{2}\,W''(q)\,[\overline{\psi}, \psi], \tag{3.8a}$$
$$Y = \tfrac{i}{2}\,[\overline{\psi}, \psi], \tag{3.8b}$$
$$Q = (p + iW'(q))\psi, \tag{3.8c}$$
$$\bar{Q} = (p - iW'(q))\overline{\psi}. \tag{3.8d}$$

In these charges (3.8), we have written $p = (\partial L/\partial\dot{q}) = \dot{q}$ and strongly realized the second class constraint (2.13d): $A = W'$. Moreover we have made the skew symmetrization of the terms proportional to $\overline{\psi}\,\psi$. Let us also notice that the total Hamiltonian (3.8a) admits the decomposition into a bosonic part H_B, i.e., containing no fermionic degree of freedom, and a fermionic part H_F such that

$$H_T = H_B + H_F, \tag{3.9a}$$

with

$$H_B = \tfrac{1}{2}p^2 + \tfrac{1}{2}(W'(q))^2 = \tfrac{1}{2}p^2 + V(q), \tag{3.9b}$$
$$H_F = \tfrac{1}{2}\,W''(q)\,[\overline{\psi}, \psi]. \tag{3.9c}$$

Since we have

$$H_F = -iW''(q)\,Y, \tag{3.10}$$

the bosonic and fermionic parts are separately conserved if and only if W'' is independent of q.

Charges (3.8) form a closed structure for the commutator and the anticommutator (2.18) which turns out to be the semidirect sum of spl(1, 1) (generated by H_T, Q and $\bar{Q}$) and so(2) (generated by Y). The brackets different from zero are

$$[Y, Q] = -iQ, \quad [Y, \bar{Q}] = i\bar{Q}, \quad \{Q, \bar{Q}\} = 2H_T. \tag{3.11}$$

B. Superpotentials $W(Z) \equiv (2.36)$

The symmetries of Lagrangian (2.6) are now associated with the transformations on t, Θ, and $\overline{\Theta}$ such that we have Eqs. (3.1) and on Z such that

$$\delta Z = \tfrac{1}{2}\,G\{D, \bar{D}\}Z + \tfrac{1}{2}\,DG\,\bar{D}Z + \tfrac{1}{2}\,\bar{D}G\,DZ - \tfrac{1}{4}\,\{D, \bar{D}\}G\,Z, \tag{3.12}$$

where G has to satisfy Eq. (2.37a). Let us then find the symmetries by considering separately the cases $\omega \neq 0$ and $\omega = 0$.

For $\omega \neq 0$, we get the general solution of Eq. (2.37a)

$$G(t,\Theta,\overline{\Theta}) = ic - 2\alpha\overline{\Theta} - 2\overline{\alpha}\Theta + 2d\Theta\,\overline{\Theta}$$
$$+ i(a\cos 2\omega t + b\sin 2\omega t)(1 + 2\omega\Theta\overline{\Theta})$$
$$- 2\beta e^{-2i\omega t}\,\overline{\Theta} - 2\overline{\beta}\,e^{2i\omega t}\,\Theta, \tag{3.13}$$

where the first four terms are evidently those given in Eq. (3.3). From Eqs. (3.1) with Eq. (3.13), we explicitly get

$$\delta t = c + a\sin 2\omega t + b\cos 2\omega t + i\alpha\,\overline{\Theta} + i\,\overline{\alpha}\,\Theta$$
$$+ i\beta\,\overline{\Theta}e^{-2i\omega t} + i\,\overline{\beta}\Theta e^{2i\omega t},$$
$$\delta\Theta = \alpha + \beta(1 - 2\omega\Theta\,\overline{\Theta})e^{-2i\omega t}$$
$$+ a\omega e^{-2i\omega t}\Theta - ib\omega e^{-2i\omega t}\Theta - d\Theta, \tag{3.14}$$
$$\delta\,\overline{\Theta} = \overline{\alpha} + \overline{\beta}(1 - 2\omega\Theta\,\overline{\Theta})e^{2i\omega t} + a\omega e^{2i\omega t}\overline{\Theta}$$
$$+ ib\omega e^{2i\omega t}\overline{\Theta} + d\,\overline{\Theta},$$

and from Eq. (3.2) with the solution (3.13), we explicitly have

$$\delta q = c\dot{q} + a(\dot{q}\sin 2\omega t - \omega q\cos 2\omega t) + b(\dot{q}\cos 2\omega t$$
$$+ \omega q\sin 2\omega t) + i\alpha\,\overline{\psi} + i\,\overline{\alpha}\psi + i\beta e^{-2i\omega t}\overline{\psi} + i\,\overline{\beta}e^{2i\omega t}\psi,$$
$$\delta\,\dot{\overline{\psi}} = c\,\dot{\overline{\psi}} - \overline{\alpha}(\dot{q} + iA) + i\,\overline{\beta}e^{2i\omega t}(-A + i\dot{q} + 2\omega q) - d\,\dot{\overline{\psi}}$$
$$+ a(\dot{\overline{\psi}} - i\omega\,\overline{\psi})\sin 2\omega t + b(\dot{\overline{\psi}} - i\omega\,\overline{\psi})\cos 2\omega t, \tag{3.15}$$
$$\delta\psi = c\,\dot{\psi} - \alpha(\dot{q} - iA) + i\beta e^{-2i\omega t}(-A + i\dot{q} - 2\omega q)$$
$$+ d\psi + a(\dot{\psi} + i\omega\psi)\sin 2\omega t + b(\dot{\psi} + i\omega\psi)\cos 2\omega t,$$
$$\delta A = c\,\dot{A} + a(\dot{A}\sin 2\omega t + (A\omega - 2\omega^2 q)\cos 2\omega t)$$
$$+ b(\dot{A}\cos 2\omega t - (A\omega - 2\omega^2 q)\sin 2\omega t)$$
$$+ \alpha\,\dot{\overline{\psi}} - \overline{\alpha}\dot{\psi} + \beta e^{-2i\omega t}\dot{\overline{\psi}} - \overline{\beta}e^{2i\omega t}\dot{\psi}.$$

The superbicurrent (2.42) becomes, with Eqs. (2.30) and (2.38),

$$\Sigma = \tfrac{1}{4}G\,\bar{D}DZ\,DZ + \tfrac{1}{4}DG\,\bar{D}Z\,DZ$$
$$+ \tfrac{1}{2}GW'DZ - \tfrac{1}{8}\{D, \bar{D}\}G\,Z\,DZ$$
$$+ \tfrac{1}{16}\,D\{D, \bar{D}\}G\,Z^2 - (\lambda/4)DG \tag{3.16a}$$

and

$$\overline{\Sigma} = -\tfrac{1}{4}G\,D\bar{D}Z\,\bar{D}Z - \tfrac{1}{4}\bar{D}G\,DZ\,\bar{D}Z$$
$$+ \tfrac{1}{2}\,GW'\bar{D}Z + \tfrac{1}{8}\{D, \bar{D}\}G\,Z\,\bar{D}Z$$
$$- \tfrac{1}{16}\,\bar{D}\{D, \bar{D}\}G\,Z^2 - (\lambda/4)\bar{D}G. \tag{3.16b}$$

By computation of the Θ component of Σ and the $\overline{\Theta}$ component of $\overline{\Sigma}$, the general charge is easily found to be

$$C = -i(cH_T + dY + aC_1 + bC_2$$
$$+ i\,\overline{\alpha}Q + i\alpha\,\bar{Q} + i\,\overline{\beta}S + i\beta\,\bar{S}), \tag{3.17}$$

where H_T, Y, Q, and $\bar{Q}$ are given in (3.8) with $W' = \omega q + (\lambda/q)$. Using $p = \dot{q}$ and $A = W'$ and defining $C_{\pm} = C_1 \pm iC_2$, we get the following:

$$C_+ = \tfrac{i}{2}\,e^{-2i\omega t}((p + i\omega q)^2 + (\lambda/q^2)(\lambda - [\overline{\psi}, \psi])), \tag{3.18a}$$

$$C_- = -\tfrac{i}{2}\,e^{2i\omega t}((p - i\omega q)^2 + (\lambda/q^2)(\lambda - [\overline{\psi}, \psi])), \tag{3.18b}$$

$$S = e^{2i\omega t}(p - i(\omega q - (\lambda/q)))\psi, \tag{3.18c}$$

$$\bar{S} = e^{-2i\omega t}(p + i(\omega q - (\lambda/q)))\bar{\psi}, \tag{3.18d}$$

which are the additional charges [with respect to Eqs. (3.8)], appearing for the particular potential we are considering here. Let us notice that in this case the bosonic and fermionic parts given by Eqs. (3.9b) and (3.9c) are not separately conserved.

We can verify that these charges form a closed structure which is the superalgebra osp(2, 2). The brackets different from zero are, in addition to those given in (3.11),

$$\begin{aligned}
&[H_T, C_\pm] = \pm 2\omega C_\pm, \\
&[C_+, C_-] = -4\omega(H_T + i\omega Y), \\
&[H_T, S] = -2\omega S, \quad [H_T, \bar{S}] = 2\omega\bar{S}, \\
&[Y, S] = -iS, \quad [Y, \bar{S}] = i\bar{S}, \\
&[C_+, \bar{Q}] = -2i\omega\bar{S}, \quad [C_+, S] = -2i\omega Q, \\
&[C_-, Q] = -2i\omega S, \quad [C_-, \bar{S}] = -2i\omega\bar{Q}, \\
&\{Q, \bar{S}\} = -2iC_+, \quad \{\bar{Q}, S\} = 2iC_-, \\
&\{S, \bar{S}\} = 2(H_T + 2i\omega Y).
\end{aligned} \tag{3.19}$$

For $\omega = 0$, Eq. (2.37a) admits the general solution

$$G(t, \Theta, \bar{\Theta}) = ic - 2\alpha\bar{\Theta} - 2\bar{\alpha}\Theta + 2d\Theta\bar{\Theta} + iat^2 + ibt + 2\beta t\bar{\Theta} + 2\bar{\beta}t\Theta, \tag{3.20}$$

so that we get from Eqs. (3.1)

$$\begin{aligned}
&\delta t = c + at^2 + bt + i\alpha\bar{\Theta} + i\bar{\alpha}\Theta - i\beta t\bar{\Theta} - i\bar{\beta}t\Theta, \\
&\delta\Theta = \alpha - \beta(t - i\Theta\bar{\Theta}) + at\Theta + (b/2)\Theta - d\Theta, \\
&\delta\bar{\Theta} = \bar{\alpha} - \bar{\beta}(t + i\Theta\bar{\Theta}) + at\bar{\Theta} + (b/2)\bar{\Theta} + d\bar{\Theta},
\end{aligned} \tag{3.21}$$

and from Eq. (3.2) with (3.20)

$$\begin{aligned}
\delta q &= c\dot{q} + a(t^2\dot{q} - tq) + b(t\dot{q} - \tfrac{1}{2}q) \\
&\quad + i\alpha\bar{\psi} + i\bar{\alpha}\psi - i\beta t\bar{\psi} - i\bar{\beta}t\psi, \\
\delta\bar{\psi} &= c\dot{\bar{\psi}} - \bar{\alpha}(\dot{q} + iA) + \bar{\beta}(\dot{q}t + itA - q) \\
&\quad - d\bar{\psi} + at^2\dot{\bar{\psi}} + bt\dot{\bar{\psi}}, \\
\delta\psi &= c\dot{\psi} - \alpha(\dot{q} - iA) + \beta(\dot{q}t - itA - q) \\
&\quad + d\psi + at^2\dot{\psi} + bt\dot{\psi}, \\
\delta A &= c\dot{A} + a(t^2\dot{A} + tA) + b(t\dot{A} + \tfrac{1}{2}A) \\
&\quad + \alpha\dot{\bar{\psi}} - \bar{\alpha}\dot{\psi} - \beta t\dot{\bar{\psi}} + \bar{\beta}t\dot{\psi}.
\end{aligned} \tag{3.22}$$

The conserved superbicurrent is given by Eqs. (3.16) so that in this case the general charge reads

$$C = -i(cH_T + dY + bD + aK + i\bar{\alpha}Q + i\alpha\bar{Q} + i\beta\bar{S} + i\bar{\beta}S), \tag{3.23}$$

where H_T, Y, Q, and $\bar{Q}$ are once more given in (3.8) (with $W' = \lambda/q$) and ($p = \dot{q}$, $A = W'$),

$$D = tH - \tfrac{1}{4}\{q, p\}, \tag{3.24a}$$

$$K = -t^2H + 2tD + \tfrac{1}{2}q^2, \tag{3.24b}$$

$$S = -tQ + q\psi, \tag{3.24c}$$

$$\bar{S} = -t\bar{Q} + q\bar{\psi}. \tag{3.24d}$$

These charges are the ones typically conserved for the superpotential $W(Z) = \lambda \ln|Z|$. In this case, the bosonic and fermionic parts of the Hamiltonian are not separately conserved.

We find that the superalgebra associated with this case is once more osp(2, 2). The brackets different from zero are, apart from those given in (3.11),

$$\begin{aligned}
&[H_T, D] = iH, \\
&[H_T, K] = 2iD, \\
&[D, K] = iK, \\
&[H_T, S] = -iQ, \quad [H_T, \bar{S}] = -i\bar{Q}, \\
&[Y, S] = -iS, \quad [Y, \bar{S}] = i\bar{S}, \\
&[D, Q] = -\tfrac{i}{2}Q, \quad [D, \bar{Q}] = -\tfrac{i}{2}\bar{Q}, \\
&[D, S] = \tfrac{i}{2}S, \quad [D, \bar{S}] = \tfrac{i}{2}\bar{S}, \\
&[K, Q] = iS, \quad [K, \bar{Q}] = i\bar{S}, \\
&\{Q, \bar{S}\} = -2D + Y, \quad \{\bar{Q}, S\} = -2D - Y, \\
&\{S, \bar{S}\} = 2K.
\end{aligned} \tag{3.25}$$

As a last comment let us notice that here we recover Durand's results[19] obtained for the above two cases, but not from the study of the supersymmetric Schrödinger equation following the method of Niederer[20] and Boyer.[21] In fact the supersymmetries of the $1/q^2$ potential have already been determined in the work of Fubini–Rabinovici[7] and have been combined[9] with those of the field of the magnetic monopole. The superposition of the harmonic oscillator and the $1/q^2$ potentials (plus a constant one) has been studied by Boyer[21] and Durand[19] as far as symmetries and supersymmetries, respectively, are concerned. Our approach differs from the preceding ones by the fact that we construct the superspace formulation as D'Hoker–Vinet[18] in the $N = 1$ context.

C. Superpotentials $W(Z) = (2.39)$

The symmetries of the Lagrangian (2.6) are associated with the transformations (3.1) on t, Θ, and $\bar{\Theta}$ and the transformation δZ on Z such that

$$\delta Z = \tfrac{1}{2}G\{D, \bar{D}\}Z + \tfrac{1}{2}DG\bar{D}Z + \tfrac{1}{2}\bar{D}GDZ - \tfrac{1}{4}\{D, \bar{D}\}GZ + Z_0, \tag{3.26}$$

where G and Z_0 have to satisfy Eqs. (2.40). Since Eq. (2.40a) is identical to Eq. (2.37a), we get solutions (3.13) for $\omega \neq 0$ and (3.20) for $\omega = 0$. As a consequence, we again consider separately the cases $\omega \neq 0$ and $\omega = 0$.

For $\omega \neq 0$, inserting solution (3.13) into Eq. (2.40b) we get

$$\begin{aligned}
Z_0(t, \Theta, \bar{\Theta}) = &-\mu(a\cos 2\omega t + b\sin 2\omega t)(1 + 4\omega\Theta\bar{\Theta}) \\
&+ 2\mu\bar{\beta}e^{2i\omega t}\Theta - 2\mu\beta e^{-2i\omega t}\bar{\Theta} \\
&+ (e\cos\omega t + f\sin\omega t)(1 + \omega\Theta\bar{\Theta}) \\
&+ i\bar{\gamma}e^{i\omega t}\Theta + i\gamma e^{-i\omega t}\bar{\Theta}.
\end{aligned} \tag{3.27}$$

This solution contains two ordinary real parameters e, f, and two Grassmann ones $\gamma, \bar{\gamma}$. Let us notice that the transformations associated with these parameters affect only the superposition Z and not the variables $(t, \Theta, \bar{\Theta})$. Indeed, we have

Eq. (3.14) while from Eq. (3.26) with Eqs. (3.13) and (3.27), we get

$$\begin{aligned}
\delta q &= c\dot{q} + a(\dot{q}\sin 2\omega t - (\omega q+\mu)\cos 2\omega t)\\
&\quad + b(\dot{q}\cos 2\omega t + (\omega q+\mu)\sin 2\omega t) + e\cos\omega t\\
&\quad + f\sin\omega t + i\alpha\bar{\psi} + i\bar{\alpha}\psi + i\beta e^{-2i\omega t}\bar{\psi} + i\bar{\beta}e^{2i\omega t}\psi,\\
\delta\bar{\psi} &= c\dot{\bar{\psi}} + a(\dot{\bar{\psi}} - i\omega\bar{\psi})\sin 2\omega t\\
&\quad + b(\dot{\bar{\psi}} - i\omega\bar{\psi})\cos 2\omega t - d\bar{\psi} - \bar{\alpha}(\dot{q}+iA)\\
&\quad + i\bar{\beta}e^{2i\omega t}(-A + i\dot{q} + 2(\omega q+\mu)) - \bar{\gamma}e^{i\omega t},\\
\delta\psi &= c\dot{\psi} + a(\dot{\psi} + i\omega\psi)\sin 2\omega t\\
&\quad + b(\dot{\psi} - i\omega\psi)\cos 2\omega t + d\psi - \alpha(\dot{q} - iA) \qquad (3.28)\\
&\quad + i\beta e^{-2i\omega t}(-A + i\dot{q} - 2(\omega q+\mu)) - \gamma e^{-i\omega t},\\
\delta A &= c\dot{A} + a(\dot{A}\sin 2\omega t + (A\omega - 2\omega^2 q - 2\omega\mu)\cos 2\omega t)\\
&\quad + b(\dot{A}\cos 2\omega t + (A\omega - 2\omega^2 q - 2\omega\mu)\sin 2\omega t)\\
&\quad + e\omega\cos\omega t + f\omega\sin\omega t + \alpha\dot{\bar{\psi}} - \bar{\alpha}\dot{\psi}\\
&\quad + \beta e^{-2i\omega t}\dot{\bar{\psi}} - \bar{\beta}e^{2i\omega t}\dot{\psi}.
\end{aligned}$$

The conserved superbircurrent (2.42) then explicitly reads

$$\begin{aligned}
\Sigma &= \tfrac{1}{4}G\bar{D}DZ\,DZ + \tfrac{1}{4}DG\bar{D}Z\,DZ + \tfrac{1}{2}W'G\,DZ\\
&\quad - \tfrac{1}{8}\{D,\bar{D}\}G\,Z\,DZ + \tfrac{1}{16}D\{D,\bar{D}\}G\,Z^2\\
&\quad + \tfrac{1}{2}Z_0\,DZ - \tfrac{1}{2}ZDZ_0 - \Omega, \qquad (3.29a)
\end{aligned}$$

and

$$\begin{aligned}
\bar{\Sigma} &= -\tfrac{1}{4}G\,D\bar{D}Z\,\bar{D}Z - \tfrac{1}{4}\bar{D}G\,DZ\,\bar{D}Z + \tfrac{1}{2}W'G\,\bar{D}Z\\
&\quad + \tfrac{1}{8}\{D,\bar{D}\}G\,Z\,\bar{D}Z - \tfrac{1}{16}\bar{D}\{D,\bar{D}\}G\,Z^2\\
&\quad - \tfrac{1}{2}Z_0\,\bar{D}Z + \tfrac{1}{2}Z\,\bar{D}Z_0 - \bar{\Omega}, \qquad (3.29b)
\end{aligned}$$

where Ω and $\bar{\Omega}$ can be obtained from Eq. (2.41) with $Z_0 \equiv (3.27)$. Going to the components, the general charge reads

$$\begin{aligned}
C &= -i(cH_T + dY + aC_1 + bC_2 + eC_3 + fC_4 + i\bar{\alpha}Q\\
&\quad + i\alpha\bar{Q} + i\bar{\beta}S + i\beta\bar{S} + i\gamma\bar{T} + i\bar{\gamma}T), \qquad (3.30)
\end{aligned}$$

where H_T, Y, Q, and $\bar{Q}$ are given in (3.8) with $W' = \omega q + \mu$. Using $p = \dot{q}$ and $A = W'$, and defining $C_\pm = C_1 \pm iC_2$, $P_\pm = C_4 \pm iC_3$ we have

$$\begin{aligned}
C_+ &= \tfrac{1}{2}e^{-2i\omega t}((p + i(\omega q+\mu))^2 - \mu^2),\\
C_- &= -\tfrac{1}{2}e^{2i\omega t}((p - i(\omega q+\mu))^2 - \mu^2), \qquad (3.31a)
\end{aligned}$$

$$\begin{aligned}
S &= e^{2i\omega t}(p - i(\omega q+\mu))\psi,\\
\bar{S} &= e^{-2i\omega t}(p + i(\omega q+\mu))\bar{\psi}, \qquad (3.31b)
\end{aligned}$$

$$\begin{aligned}
P_+ &= ie^{-i\omega t}(p + i(\omega q+\mu)),\\
P_- &= -ie^{i\omega t}(p - i(\omega q+\mu)), \qquad (3.31c)
\end{aligned}$$

$$\begin{aligned}
T &= e^{i\omega t}\psi,\\
\bar{T} &= e^{-i\omega t}\bar{\psi}. \qquad (3.31d)
\end{aligned}$$

In this case, let us notice that the bosonic and fermionic parts $\equiv (3.9)$ of the total Hamiltonian are separately conserved, the fermionic part H_F essentially being the charge Y. It is easy to show that the charges (3.8a), (3.31a), and (3.31c) generating the so-called Schrödinger group[25] are exactly those which would be obtained by considering the problem where the fermionic degrees of freedom have been suppressed. This was not the case in the preceding section.

The charges (3.31a) together with (3.8) again lead to the superalgebra osp(2, 2). Moreover we also have four additional charges [(3.31c) and (3.31d)] which come from the nonannulation of Z_0. Finally, if we add the identity I to them, we obtain the superalgebra sh(1) characterized by the nontrivial relations

$$[P_+, P_-] = 2\omega I, \quad \{T, \bar{T}\} = I. \qquad (3.32)$$

We can then check that, taken altogether, the charges (3.8) and (3.31) form the semidirect sum osp(2, 2) $\square$ sh(1). Indeed we have

$$\begin{aligned}
&[H_T, P_+] = \omega P_+, \quad [H_T, P_-] = -\omega P_-,\\
&[H_T, T] = -\omega T, \quad [H_T, \bar{T}] = \omega\bar{T},\\
&[Y, T] = -i\omega T, \quad [Y, \bar{T}] = i\omega\bar{T},\\
&[C_+, P_-] = 2i\omega P_+, \quad [C_-, P_+] = 2i\omega P_-,\\
&[P_+, \bar{Q}] = -2i\omega\bar{T}, \quad [P_-, Q] = -2i\omega T, \qquad (3.33)\\
&[P_+, S] = -2i\omega T, \quad [P_-, \bar{S}] = -2i\omega\bar{T},\\
&\{T, \bar{Q}\} = iP_-, \quad \{\bar{T}, Q\} = -iP_+,\\
&\{T, \bar{S}\} = -iP_+, \quad \{\bar{T}, S\} = iP_-.
\end{aligned}$$

For $\omega = 0$, inserting the solution (3.20) into Eq. (2.40b) we get

$$\begin{aligned}
Z_0(t, \Theta, \bar{\Theta}) &= et + f - \mu\bar{\beta}t\Theta + \mu\beta t\bar{\Theta}\\
&\quad + i\bar{\gamma}\Theta + i\gamma\bar{\Theta} + \tfrac{1}{2}\mu b\Theta\bar{\Theta}, \qquad (3.34)
\end{aligned}$$

so that in components $\delta Z \equiv (3.26)$ reads

$$\begin{aligned}
\delta q &= c\dot{q} + a(t^2\dot{q} - tq) + b(t\dot{q} - \tfrac{1}{2}q)\\
&\quad + i\alpha\bar{\psi} + i\bar{\alpha}\psi - i\beta t\bar{\psi} - i\bar{\beta}t\psi + et + f,\\
\delta\bar{\psi} &= c\dot{\bar{\psi}} + at^2\dot{\bar{\psi}} + bt\dot{\bar{\psi}} - d\bar{\psi}\\
&\quad - \bar{\alpha}(\dot{q} + iA) + \bar{\beta}(\dot{q}t + itA - q - i\mu t) - \bar{\gamma},\\
\delta\psi &= c\dot{\psi} + at^2\dot{\psi} + bt\dot{\psi} + d\psi - \alpha(\dot{q} - iA) \qquad (3.35)\\
&\quad + \beta(\dot{q}t - itA - q + i\mu t) - \gamma,\\
\delta A &= c\dot{A} + a(t^2\dot{A} + tA) + b(t\dot{A} + \tfrac{1}{2}(A+\mu))\\
&\quad + \alpha\dot{\bar{\psi}} - \bar{\alpha}\dot{\psi} - \beta t\dot{\bar{\psi}} + \bar{\beta}t\dot{\psi}.
\end{aligned}$$

The conserved superbicurrent is given by (3.29), Ω and $\bar{\Omega}$ satisfying Eq. (2.41) with $Z_0 = (3.34)$. Going to the components, the general charge is now

$$\begin{aligned}
C &= -i(cH_T + dY + aK + bD + eE + fF + i\bar{\alpha}Q\\
&\quad + i\alpha\bar{Q} + i\bar{\beta}S + i\beta\bar{S} + i\gamma\bar{T} + i\bar{\gamma}T), \qquad (3.36)
\end{aligned}$$

where H_T, Y, Q, and $\bar{Q}$ are given by Eqs. (3.8) with $W' = \mu$, D, K, S, and $\bar{S}$ are given by Eqs. (3.24) and, with $p = \dot{q}$,

$$E = tp - q, \quad F = p, \quad T = \psi, \quad \bar{T} = \bar{\psi}. \qquad (3.37)$$

Just as for $\omega \neq 0$, we have to add the identity I in order to get a closed structure. If we want to see directly that this structure is the semidirect sum osp(2, 2) $\square$ sh(1), we have to consider the linear combination

$$\begin{aligned}
&Q' = Q - i\mu T = p\psi, \quad \bar{Q}' = \bar{Q} + i\mu\bar{T} = p\bar{\psi},\\
&H'_T = H_T - \tfrac{1}{2}\mu^2 I = \tfrac{1}{2}p^2, \qquad (3.38)
\end{aligned}$$

instead of Q, $\bar{Q}$, and H_T. It can then be shown that these new charges verify the same commutation rules (3.11) as the old ones. One can also check that, together with D, K, S, and $\bar{S}$, the charges (3.38) verify relation (3.25), which characterizes the osp(2, 2) algebra. Moreover E, F, T, and $\bar{T}$ turn out to obey the (anti)commutation relations of the sh(1) superalgebra. The other nonzero brackets are then given by

$$\begin{aligned}
&[H'_T, E] = iF,\\
&[D, E] = \tfrac{i}{2} E,\\
&[D, F] = -\tfrac{i}{2} F,\\
&[Y, T] = -iT, \quad [Y, \bar{T}] = i\bar{T},\\
&[Q', E] = iT, \quad [\bar{Q}', E] = i\bar{T},\\
&[S, F] = iT, \quad [\bar{S}, F] = i\bar{T},\\
&\{Q', \bar{T}\} = \{\bar{Q}', T\} = F,\\
&\{S, \bar{T}\} = \{\bar{S}, T\} = -E.
\end{aligned} \tag{3.39}$$

Finally, let us notice that from the superspace formulation we have shown that we recover some results already known[14–16,19] about the supersymmetries of the physical systems such as the free particle and the harmonic oscillator. Let us recall that since Niederer[25] we know that a change of variables realizes the one-to-one correspondence between the free particle and the harmonic oscillator showing that the superalgebras for both cases are isomorphic ones.

In our approach we have (additionally with respect to Durand's[19] work) considered (when $\omega \neq 0$ in the preceding discussion) the potential

$$V(q) = \tfrac{1}{2}\omega^2 q^2 + \tfrac{1}{2}\mu^2 + \omega\mu q, \tag{3.40}$$

which corresponds to the superpotential $W(Z) = (2.39)$ including the harmonic oscillator and constant potentials as well as the linear one. We have shown that it admits a symmetry also associated with the superalgebra osp(2, 2) $\square$ sh(1). The correspondence is evident since the substitution $q' = q + (\mu/\omega)$ leads to the harmonic oscillator potential. Let us notice that with $V(q) \equiv (3.40)$ we cannot recover the linear potential alone. It was expected since as already noticed by Durand,[19] this case does not admit a large set of supersymmetries. More precisely, we can say here that such a case admits only the superalgebra (3.8) of symmetries when we have considered $W' = \nu q^{1/2}$. This supersymmetric system breaks down some symmetries of the nonsupersymmetric one. Indeed the symmetries[25] of the Schrödinger equation with the linear potential are associated with the Schrödinger algebra Schr(1) and there exists[20] a one-to-one correspondence between the linear potential case and the free one.

ACKNOWLEDGMENT

Very stimulating discussions with Professor J. Beckers are cordially acknowledged.

[1] Yu. A. Gol'fand and E. P. Likhtman, JETP Lett. **13**, 323 (1971); P. Ramond, Phys. Rev. D **3**, 2415 (1971); J.-L. Gervais and B. Sakita, Nucl. Phys. B **34**, 633 (1971); A. Neveu and J. Schwarz, *ibid.* **31**, 86 (1971); J. Wess and B. Zumino, *ibid.* **70**, 39 (1974), Phys. Lett. B **49**, 52 (1974).
[2] P. Fayet and S. Ferrara, Phys. Rep. C **32**, 249 (1977), and references therein.
[3] F. A. Berezin and M. S. Marinov, JETP Lett. **21**, 30 (1975), Ann. Phys. (NY) **104**, 336 (1977); A. Barducci, R. Casalbuoni, and L. Lusanna, Nuovo Cimento A **35**, 377 (1976); F. Ravndal, Phys. Rev. D **21**, 2823 (1980); C. A. P. Galvao and C. Tetelboim, J. Math. Phys. **21**, 1863 (1980); P. D. Mannheim, Phys. Rev. D **32**, 898 (1985).
[4] R. Casalbuoni, Nuovo Cimento A **33**, 115 (1976).
[5] R. Casalbuoni, Nuovo Cimento A **33**, 389 (1976).
[6] E. Witten, Nucl. Phys. B **188**, 513 (1981); **202**, 253 (1982); P. Salomonson and J. W. Van Holten, Nucl. Phys B **196**, 509 (1982); M. De Crombrugghe and V. Rittenberg, Ann. Phys. (NY) **151**, 99 (1983).
[7] S. Fubini and E. Rabinovici, Nucl. Phys. B **245**, 17 (1984).
[8] E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984).
[9] E. D'Hoker and L. Vinet, Comm. Math. Phys. **97**, 391 (1985).
[10] E. D'Hoker and L. Vinet, Nucl. Phys. B **260**, 79 (1985).
[11] F. Ravndal, *Proceedings of the Cern School of Physics* (CERN, Geneva, 1984), p. 300.
[12] A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985).
[13] J. Gamboa and J. Zanelli, Phys. Lett. B **165**, 91 (1985).
[14] V. Hussin, Bull. Soc. R. Sc. Liège **55**, 363 (1986).
[15] J. Beckers and V. Hussin, Phys. Lett. A **118**, 319 (1986).
[16] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A (to be published).
[17] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[18] E. D'Hoker and L. Vinet, Lett. Math. Phys. **8**, 439 (1984).
[19] S. Durand, Maîtrise ès Sciences, Université de Montréal, 1985.
[20] U. Niederer, Helv. Phys. Acta **47**, 167 (1974).
[21] C. P. Boyer, Helv. Phys. Acta **47**, 590 (1974).
[22] P. A. M. Dirac, *Lectures on Quantum Mechanics, Belfer Graduate School of Science* (Yeshiva U., New York, 1964).
[23] A. Hanson, T. Regge, and C. Tetelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Liucei, Rome, 1976).
[24] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[25] U. Niederer, Helv. Phys. Acta **45**, 802 (1972); **46**, 191 (1973).

On the color factor of n-gluon decay from quarkonium

Minoru Biyajima
Department of Physics, Matsumoto, Faculty of Liberal Arts, Shinshu University, Matsumoto 390, Japan
Tetsuji Kawabe
Department of Physics, Kyushu Institute of Design, Fukuoka 815, Japan
Naomichi Suzuki
Matsushou Gakuen Junior College, Matsumoto 390-12, Japan

(Received 30 May 1986; accepted for publication 15 October 1986)

A method for evaluating the weight of the color factor necessary for gluon decay in the context of quantum chromodynamics (QCD) is presented. As concrete examples, the method is applied to the three- and four-gluon decays from heavy quarkonium.

I. INTRODUCTION

To study the decay of any quarkonium in quantum chromodynamics (QCD) (Ref. 1), it is necessary to evaluate the color factor involved in an amplitude for the n-gluon annihilation of a color singlet state as shown in Fig. 1.

For the case of charmonium decay, two- and three-gluon decays with definite charge conjugation (c.c.) states have been studied from realistic viewpoints.[2] The amplitude for the two-gluon decay of c.c. even states is given by

$$4\pi\alpha_s M_{\mu\nu}\,\mathrm{Tr}(\lambda^a\lambda^b/2^2)e_\mu^a(1)e_\nu^b(2)/\sqrt{3}, \tag{1}$$

where $a,b = 1,2,...,8$, the λ^a are the standard SU(3) matrices, $e_\mu{}^a(1)$ and $e_\nu{}^a(2)$ are the gluon wave functions, and α_s is the coupling constant of the quark–gluon interaction.

Similarly, the amplitude for the three-gluon decay of c.c. odd states is given by

$$4\pi\alpha_s{}^{3/2}M_{\mu\nu\rho}w_3{}^{1/2}e_\mu{}^a(1)e_\nu{}^b(2)e_\rho{}^c(3)/\sqrt{3}, \tag{2}$$

with

$$w_3{}^{1/2} = \mathrm{Tr}(\lambda^a\lambda^b\lambda^c/2^3)_{\mathrm{sym}},$$

where the suffix "sym" means that the structure constants for the totally symmetric states, i.e., d coupling, are only taken into account. In the above expressions (1) and (2), the trace parts describing the color factors are easily calculated by using the general properties of λ matrices.[2] However, recent argument[3] for heavier quarkonium, i.e., upsilon ($b\bar{b}$), seems to need the contribution of four-gluon decay in addition to the three-gluon decay. If this is the case, we have to evaluate the amplitude involving the four λ matrices under the constraint of the charge conjugation as follows:

$$4\pi\alpha_s{}^2M_{\mu\nu\rho\sigma}w_4{}^{1/2}e_\mu{}^a(1)e_\nu{}^b(2)e_\rho{}^c(3)e_\sigma{}^d(4)/3, \tag{3}$$

with

$$w_4{}^{1/2} = \mathrm{Tr}(\lambda^a\lambda^b\lambda^c\lambda^d/2^4)_{\text{c.c. odd}}.$$

The actual calculation of the trace w_4 is, however, not so easy as that of w_3 in (2) because the way of combining the λ matrices to form the c.c. odd states is very complicated. Unfortunately, the explicit value seems not to be given in any literature, to the best of our knowledge.[4] Furthermore, from the physical viewpoint the color factor w_n with $n\geqslant 4$ will become important for the study of heavier quarkonium such as topponium ($t\bar{t}$).

II. METHOD OF CALCULATIONS

In this paper we give the explicit method to evaluate the color factor w_n for n-gluon decay by using the Young tableaux and the concept of the nth symmetric group S_n. And we apply our method to three- and four-gluon decays with c.c. odd states as concrete examples.

The systematic evaluation for the trace of the product of λ matrices is obtained from the following rule derived by the method of the Young tableaux[5,6]:

$$\mathrm{Tr}(\underbrace{\lambda^a\lambda^b\cdots\lambda^k}_{n}/2^n) = (1/\sqrt{n!})\mathrm{Tr}(\underbrace{\square\times\square\times\cdots\times\square}_{n}). \tag{4}$$

The parentheses containing n boxes on the rhs of (4) stand for all possible normalized basis functions of the nth symmetric group S_n. The basis functions are constructed by the usual procedure[6,7] as follows:

$$|l,m\rangle = \sum_R D_{lm}(R)R\Psi, \tag{5}$$

where $l,m = 1,2,...,d$ (= dimension of the representation), and $D_{lm}(R)$ denotes the (l,m) component of orthogonal matrices of irreducible representations of S_n belonging to the group element R. Here, Ψ represents a state vector of n identical particles.

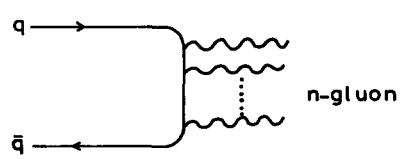


FIG. 1. The decay of the quarkonium state into an n-gluon. The solid and wavy lines stand for quark and gluon, respectively.

 0022-2488/87/020485-04$02.50 © 1987 American Institute of Physics

III. CALCULATION OF S_3

For the case of S_3 there are the following three possible Young tableaux:

$$\phi_s = \square\square\square \quad , \quad \phi_a = \begin{array}{|c|}\hline \\ \hline \\ \hline \end{array} \, , \quad \phi_m = \begin{array}{|c|c|}\hline \\ \hline \end{array}\!\!\begin{array}{|c|}\hline \\ \hline \end{array} \, , \tag{6}$$

where ϕ_s (ϕ_a) represents a state which is totally (anti)symmetric under the interchange of any two of the particles and stands for a basis vector of a one-dimensional irreducible unitary representation of S_3. On the other hand, ϕ_m denotes the tableau of mixed symmetry for the two functions which are the basis vectors of a two-dimensional irreducible representation (doublets) of S_3. In S_3, there are six elements such as E, (1 2), (1 3), (2 3), (1 2 3), and (1 3 2). Thus we obtain from (5) with $\Psi = ABC$ for ϕ_s and ϕ_a,

$$\phi_s = (1/\sqrt{6})(ABC + BAC + CBA + ACB + CAB + BCA), \tag{7}$$

$$\phi_a = (1/\sqrt{6})(ABC - BAC - CBA - ACB + CAB + BCA), \tag{8}$$

where A, B, and C denote $\lambda^a/2$, $\lambda^b/2$, and $\lambda^c/2$, respectively, and also stand for gluons. For four functions of ϕ_m, it is sufficient to consider only one of them, e.g., $l = m = 1$, to show the present procedure

$$\begin{aligned}\phi_m^{(1,1)} &= (1/\sqrt{3})|1,1\rangle \\ &= (1/\sqrt{3})[D_{11}(E)E + D_{11}(1\,2) + D_{11}(1\,3) \\ &\quad + D_{11}(2\,3)(2\,3) \\ &\quad + D_{11}(1\,2\,3)(1\,2\,3) + D_{11}(1\,3\,2)(1\,3\,2)]ABC \\ &= (1/2\sqrt{3})(2ABC + 2BAC - CBA \\ &\quad - ACB - BCA - CAB),\end{aligned} \tag{9}$$

where we use the same orthogonal matrices for D as those given by Hamermesh.[6] Here it is worth giving the correspondence of notations between the present expressions for the basis functions of S_3 and those given by Lichtenberg[5] $\psi_s = \sqrt{6}\phi_s$, $\psi_a = \sqrt{6}\phi_a$, $\psi = -|1,1\rangle + \sqrt{3}|2,1\rangle$, $\psi_1 = |1,1\rangle + |2,1\rangle/\sqrt{3}$, $\phi = 2|2,2\rangle$, and $\chi = -2|1,2\rangle/\sqrt{3}$. Since the trace of the mixed symmetric states vanishes as easily as proved from (9), the trace of the Young tableaux in the rhs of (4) consists of the totally symmetric state and the antisymmetric one

$$\mathrm{Tr}(\lambda^a\lambda^b\lambda^c/2^3) = (1/\sqrt{6})\mathrm{Tr}(\phi_s + \phi_a). \tag{10}$$

For the charge conjugation we notice that a gluon field $G^\alpha{}_\beta$ goes into $-G^\beta{}_\alpha$ ($\alpha,\beta = 1,2,3$), under the charge conjugation so that $\mathrm{Tr}(ABC)$ and $\mathrm{Tr}(BAC)$ go into $-\mathrm{Tr}(BAC)$ and $-\mathrm{Tr}(ABC)$, respectively. This means that the c.c. odd state corresponds to the symmetric state ϕ_s as known from (7) and (8). From the general properties of λ matrices as

$$\{\lambda^a/2,\ \lambda^b/2\} = \delta_{ab}/3 + d_{abc}\lambda^c/2, \tag{11}$$

$$\mathrm{Tr}\,\lambda = 0, \tag{12}$$

we get the following relation:

$$\mathrm{Tr}\,\phi_s = (\sqrt{3}/\sqrt{2})\mathrm{Tr}(\{A,B\}C) = (\sqrt{3}/2\sqrt{2})d_{abc}. \tag{13}$$

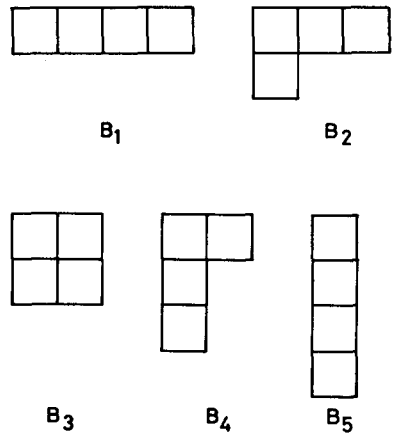


FIG. 2. The Young tableaux for S_4.

Thus we obtain the well-known result[2] for the weight w_3 as

$$w_3 = \sum_{a,b,c}\left[\left(\frac{1}{\sqrt{6}}\right)\mathrm{Tr}(\square\square\square)\right]^2 = \sum_{a,b,c}\left(\frac{d_{abc}}{4}\right)^2 = \frac{5}{6}. \tag{14}$$

IV. CALCULATION OF S_4

For the case of S_4 there are five Young tableaux as shown in Fig. 2. In S_4, there are 24 elements such as E, (1 2), (1 3), (1 4), (2 3), (2 4), (3 4), (1 2)(3 4), (1 3)(2 4), (1 4)(2 3), (1 2 3), (1 3 2), (1 2 4), (1 4 2), (1 3 4), (1 4 3), (2 3 4), (2 4 3), (1 2 3 4), (1 2 4 3), (1 3 2 4), (1 3 4 2), (1 4 2 3), and (1 4 3 2). In the same way as the case of S_3, we first write down the basis functions of their irreducible representation for all Young tableaux contained in S_4 by means of (5) with $\Psi = ABCD$. Next we select the nonvanishing Young tableaux by taking the trace of the basis functions. Then we find that there remain the totally symmetric tableaux (B_1) and one of the mixed symmetric ones (B_4). Furthermore, it is easily shown that the mixed symmetric tableau B_4 corresponds to the c.c. odd state of four-gluon decay.

Then we obtain for (4) as follows:

$$\mathrm{Tr}(\lambda^a\lambda^b\lambda^c\lambda^d/2^4)_{\text{c.c. odd}} = (1/\sqrt{4!})\mathrm{Tr}\begin{array}{|c|c|}\hline \\ \hline \\ \hline \end{array}\!\!\begin{array}{|c|}\hline \\ \hline \end{array}\,. \tag{15}$$

This tableau denotes the basis vector of three three-dimensional representations (triplets) for the mixed symmetry. For the function $|1,1\rangle$ we obtain

$$\begin{aligned}|1,1\rangle &= [D_{1\,1}(E)E + D_{1\,1}(1\,2)(1\,2) + \cdots \\ &\quad + D_{1\,1}(1\,4\,3\,2)(1\,4\,3\,2)]ABCD \\ &= ABCD + BACD + \cdots + DABC/2.\end{aligned} \tag{16}$$

The normalized function ϕ_1 is then given by $|1,1\rangle/2\sqrt{2}$. Similarly, we have all the orthogonal normalized basis functions $\phi_i(\propto|i,1\rangle)$, $\chi_i(\propto|i,2\rangle)$, and $\psi_i(\propto|i,3\rangle)$ ($i = 1,2,3$) belonging to each triplet (3 2 1 1), (3 1 2 1), and (1 3 2 1) in the Yamanouchi symbols.[6] The explicit forms of them are as follows:

$$\phi_1 = (1/4\sqrt{2})(2ABCD - CABD - DACB - BCAD - BDCA - 2BADC + ACDB + ADBC + CBDA + DBAC + 2BACD - ACBD - ADCB - CBAD - DBCA - 2ABDC + CADB + DABC + BCDA + BDAC), \tag{17}$$

$$\phi_2 = (1/4\sqrt{6})(-3CABD - DACB + 3BCAD + BDCA - ACDB - 3ADBC + 2CDAB + CBDA - 2DCBA + 3DBAC + 3ACBD + ADCB - 3CBAD - DBCA + CADB + 3DABC - 2DCAB - BCDA + 2CDBA - 3BDAC), \tag{18}$$

$$\phi_3 = (1/2\sqrt{3})(DACB - BDCA + ACDB + CDAB - CBDA - DCBA - ADCB + DBCA - CADB - DCAB + BCDA + CDBA), \tag{19}$$

$$\chi_1 = (1/4\sqrt{6})(3CABD + DACB - 3BCAD - BDCA - 3ACDB - ADBC + 2CDAB + 3CBDA - 2DCBA + DBAC + 3ACBD + ADCB - 3CBAD - DBCA - 3CADB - DABC + 2DCAB + 3BCDA - 2CDBA + BDAC), \tag{20}$$

$$\chi_2 = (1/12\sqrt{2})(6ABCD - 3CABD + 5DACB - 3BCAD + 5BDCA + 2BADC - ACDB - ADBC - 4CDAB - CBDA - 4DCBA - DBAC - 6BACD + 3ACBD - 5ADCB + 3CBAD - 5DBCA - 2ABDC + CADB + DABC + 4DCAB + BCDA + 4CDBA + BDAC), \tag{21}$$

$$\chi_3 = \tfrac{1}{6}(DACB + BDCA - 2BADC + ACDB - 2ADBC + CDAB + CBDA + DCBA - 2DBAC - ADCB - DBCA + 2ABDC - CADB + 2DABC - DCAB - BCDA - CDBA + 2BDAC), \tag{22}$$

$$\psi_1 = (1/2\sqrt{3})(-DACB + BDCA + ADBC + CDAB - DCBA - DBAC - ADCB + DBCA + DABC + DCAB - CDBA - BDAC), \tag{23}$$

$$\psi_2 = \tfrac{1}{6}(DACB + BDCA - 2BADC - 2ACDB + ADBC + CDAB - 2CBDA + DCBA + DBAC - ADCB - DBCA + 2ABDC + 2CADB - DABC - DCAB + 2BCDA - CDBA - BDAC), \tag{24}$$

$$\psi_3 = (1/6\sqrt{2})(3ABCD + 3CABD - DACB + 3BCAD - BDCA - BADC - ACDB - ADBC - CDAB - CBDA - DCBA - DBAC - 3BACD - 3ACBD + ADCB - 3CBAD + DBCA + ABDC + CADB + DABC + DCAB + BCDA + CDBA + BDAC). \tag{25}$$

The trace of this mixed symmetry becomes

$$\mathrm{Tr}\ \square\!\!\square = k\,[\mathrm{Tr}\phi_1 + (1/\sqrt{3})\mathrm{Tr}\chi_1 + (2\sqrt{2}/\sqrt{3})\mathrm{Tr}\psi_1], \tag{26}$$

with

$$\mathrm{Tr}\,\phi_1 = 4\,\mathrm{Tr}(\{A,B\}[C,D]) = 2id_{abm}f_{cdm}, \tag{27}$$

$$\mathrm{Tr}\,\chi_1 = 4\,\mathrm{Tr}(\{B,C\}[D,A]) + 4\mathrm{Tr}(\{A,C\}[B,D]) = 2i(d_{bcm}f_{dam} + d_{acm}f_{bdm}), \tag{28}$$

$$\mathrm{Tr}\,\psi_1 = 2\,\mathrm{Tr}([B,C]\{A,D\}) + 2\,\mathrm{Tr}([C,A]BD) + 2\,\mathrm{Tr}(AC[B,D]) = \tfrac{1}{2}i(2d_{adm}f_{bcm} + f_{cam}d_{mbd} + f_{bdm}d_{acm}), \tag{29}$$

and

$$k = (1/\sqrt{2} + 1/\sqrt{3} + 1/\sqrt{6})/4.$$

Here we use (1 1), (1 2), and the property as

$$[\lambda^a/2, \lambda^b/2] = if_{abc}\lambda^c/2. \tag{30}$$

Thus we obtain the value of w_4 from (15),

$$w_4 = \sum_{\substack{a,b \\ c,d}} \left[\left(\frac{1}{\sqrt{24}}\right)\mathrm{Tr}\left(\square\!\!\square\right)\right]^2 = 3.581. \tag{31}$$

V. CONCLUDING REMARKS

Our method can be directly applied to arbitrary n-gluon states with the definite charge conjugation. As a summary, we briefly describe the procedures in order. First, we write down all Young diagrams belonging to S_n and find all expressions of the basis functions for their irreducible representation by (5). Second, we take the trace of them and select the nonvanishing Young tableaux. Third, we check the charge conjugation for their basis functions. Thus we can construct, straightforwardly, appropriate basis functions for n-gluon decay with definite c.c. states as shown above for the case of n = three- and four-gluon decays.

ACKNOWLEDGMENTS

The authors would like to thank D. B. Lichtenberg and Y. Tanabe for their kind correspondences. One of the authors (T.K.) would like to thank S. Kawabe for many helpful discussions.

[1]W. Marciano and H. Pagels, Phys. Rep. C **36**, 137 (1978); Yu. L. Dokshitzer, D. I. Dyakonov, and S. I. Troyan, *ibid.* **58**, 269 (1980).
[2]V. A. Novikov, L. B. Okun, M. A. Shifman, A. I.Vainshtein, M. B. Voloshin, and V. I. Zakharov, Phys. Rep. **41**, 1 (1978).
[3]A. C. Melissinos, *Proceedings of the 1st Hellenic School on Elementary Particle Physics* (World Scientific, Corfu, Greece, 1982), p. 303; P. M. Mackenzie and G. P. Lepage, Phys. Rev. Lett. **47**, 1244 (1981).
[4]Y. Tanabe (private communication); D. B. Lichtenberg (private communication).
[5]D. B. Lichtenberg, *Unitary Symmetry and Elementary Particles* (Academic, New York, 1978), 2nd ed.
[6]M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison–Wesley, Reading, MA, 1962); D. E. Rutherford, *Substitutional Analysis* (Edinburgh Univ., Edinburgh, 1948).
[7]T. Inui, Y. Tanabe, and Y. Onodera, *Group Theory and Its Applications in Physics* (Syōkabō, Tokyo, 1980) (in Japanese).

On dissipativity of quantum optical systems

J. Andres[a)]
Joint Laboratory of Optics of the Czechoslovakian Academy of Science and Palacký University, 771 46 Olomouc, Czechoslovakia

(Received 9 May 1986; accepted for publication 8 October 1986)

Applying a suitably modified Liapunov–Yoshizawa direct method [T. Yoshizawa, *Stability Theory by Liapunov's Second Method* (Math. Soc. Japan, Tokyo, 1966)], a rigorous mathematical proof of dissipativity in the sense of Levinson [N. Rouche, P. Habets, and M. Laloy, *Stability Theory by Liapunov's Direct Method* (Springer, Berlin, 1977)] to the majority of effective optical processes has been carried out. The ability of an upper final estimation of the average number of photons is demonstrated (as an example) on the well-known second-harmonic generation process with classical pumping.

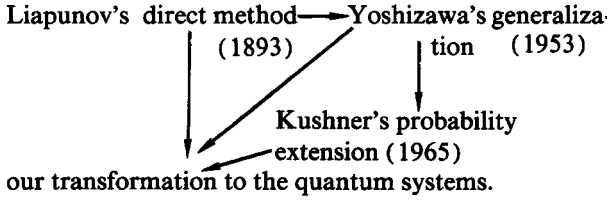
I. INTRODUCTION

It is very interesting that the ingenious idea of Liapunov, concerning the construction of comparing functions to the stability analysis, a straightforward expression of which has become his second (direct) method, grew rich essentially only after more than half a century. This famous generalization due to the Japanese mathematician Yoshizawa[1] consists mainly of an extension of the analyzed object into the (semi-) invariant sets.[1] Recently, Habets and Peiffer[2] performed a classification of the Liapunov–Yoshizawa functions (more precisely, of the single types of attractors studied by them) in detail. Moreover, Kushner[3] transformed the actual deterministic stability theory by an implementation of the probability element on the stochastic models in the 1960s. Therefore it seems to be only a little step to employ Liapunov's idea for the quantum dynamical systems.

Although some attempts have been done in this field [namely the potential[4] defined by virtue of the stationary (quasi-) distribution or the generalized entropy[5] defined by virtue of the density operator have been used for the construction of the Liapunov functions], the corresponding general theory does not yet exist. Maybe an operator nature of the variables included (in the Heisenberg picture) makes the largest difficulties in this respect.

In spite of the intuitively clear (from the physical point of view) qualitative behavior of the lossy systems, here, we would like to prove rigorously the dissipativity in the sense of Levinson[2] to some nonlinear optical phenomena, when applying the quantum theory of damping.[6,7] As we will see, sometimes their final state can be deduced quantitatively in the same manner (at the simpler effects).

The successive generalization of Liapunov's idea can be schematically sketched as follows:



There are de facto three conceptions of stability in our diagram, regarding the character of the level respected. The mutual correspondence is reached by the projection into the phase space of amplitudes through the basis of the coherent states.[6] Here, it should be noted that such an averaging is not done *a priori* at the studied systems, but at the conservation laws of their "conservative parts" (which are taken just as the Yoshizawa's functions), because an influence of the quantum fluctuations is not eliminated in this way.

II. DERIVATION OF RATE EQUATIONS (QUANTUM THEORY OF DAMPING)

Let us consider the Hamiltonian $\hat{H} = \hat{H}_L + \hat{H}_N$, where

$$\hat{H}_L = \sum_{k=1}^{m} (f_k(t)\hat{a}_k^+ \hat{a}_k + p_k(t)\hat{a}_k^+ + p_k^*(t)\hat{a}_k),$$

$$\hat{H}_N = F(t,\hat{a}_1^+,...,\hat{a}_m^+,\hat{a}_1,...,\hat{a}_m).$$

The functions $f_k(t) = \mathrm{Re} f_k(t)$ are continuous and the $p_k(t)$ are integrable (in Lebesgue's sense) everywhere, $F = F^+$ is a continuous multilinear form without the "unmixed terms" (i.e., without those like $[g_{j...k}(t)\hat{a}_j...\hat{a}_k + \text{H.c.}]$) and $\hat{a}_k^+$ ($\hat{a}_k$) are the creation (annihilation) photon operators[6] which can be represented, e.g., as infinitely dimensional matrixes of some Hilbert space.

Thus the corresponding Heisenberg equations take the form ($\hbar = 1$)

$$i\dot{\hat{a}}_j = f_j(t)\hat{a}_j + p_j(t) + \frac{\partial F}{\partial \hat{a}_j^+}, \quad \text{for } j = 1,...,m. \tag{2.1}$$

Assuming the actual (lossy) mechanism as to be connected with an infinite reservoir boson system in each of the modes (the quantum theory of damping), (2.1) can be replaced[6,7] by the following Langevin equations:

$$i\dot{\hat{a}}_j = \left(f_j(t) - i\frac{\gamma_j}{2}\right)\hat{a}_j + \frac{\partial F}{\partial \hat{a}_j^+} + p_j(t) + i\hat{L}_j, \quad \text{for } j = 1,...,m, \tag{2.2}$$

with Langevin forces $\hat{L}_j$ representing the quantum noise contributions of the reservoir and γ_j being the positive damping constants.

In order for the Bose–Einstein statistical rules to be satisfied for the ensemble of the radiation field photons [whose dynamics is determined by (2.2)], we will traditionally require that

a) Present address: Department of Mathematical Analysis and Numerical Mathematics, Faculty of Science, Palacký University, 771 46 Olomouc, Czechoslovakia.

 0022-2488/87/020489-03$02.50 © 1987 American Institute of Physics

$$[\hat{a}_k,\hat{a}_j^+] = \delta_{kj} \quad \text{(the Kronecker delta)},$$
$$[\hat{a}_k,\hat{a}_j] = [\hat{a}_k^+,\hat{a}_j^+] = 0, \qquad \text{for } j,k = 1,...,m.$$

Since $p_j(t)$ represents the pumping of the jth mode, (2.1) turns out under $p_j(t) \equiv 0$ $(j = 1,...,m)$ to be a conservative system with the conservation law

$$\sum_{k=1}^{m} c_k \hat{a}_k^+ \hat{a}_k = \text{const}, \tag{2.3}$$

where the c_k are suitable positive reals. Indeed, if we multiply (2.3) by the density operator $\hat{\rho}$, take the trace and derivative, we obtain a zero identity, i.e. ($\dot{\hat{\rho}} = 0$ according to the Liouville theorem[6]),

$$\frac{d}{dt}\sum_{k=1}^{m} \mathrm{Tr}\,\hat{\rho} c_k \hat{a}_k^+ \hat{a}_k = \sum_{k=1}^{m} \mathrm{Tr}\,\hat{\rho} c_k (\dot{\hat{a}}_k^+ \hat{a}_k + \hat{a}_k^+ \dot{\hat{a}}_k) = 0, \tag{2.4}$$

where we have substituted for $\dot{\hat{a}}_k$ and $\dot{\hat{a}}_k^+$ from the right-hand sides of (2.1) and have used the ability to find constants c_k such that

$$\sum_{k=1}^{m} c_k \hat{a}_k^+ \frac{\partial F}{\partial \hat{a}_k^+} = \sum_{k=1}^{m} c_k \frac{\partial F}{\partial \hat{a}_k} \hat{a}_k. \tag{2.5}$$

III. DEFINITION OF YOSHIZAWA'S FUNCTION

At first, let us recall the simplified version of Yoshizawa's theorem (see Ref. 4, p. 38).

Yoshizawa's Theorem: Let there exist a function $V(\langle n_1\rangle,...,\langle n_m\rangle)$ with continuous first-order partial derivatives with respect to $\langle n_1\rangle,...,\langle n_m\rangle$ for all $(\langle n_1\rangle,...,\langle n_m\rangle)$ such that

$$\sum_{k=1}^{m} |\langle n_k\rangle| \geqslant \delta > 0,$$

where δ is a suitable constant. If the following relations

(i) $\lim V(\langle n_1\rangle,...,\langle n_m\rangle) = \infty$, for $\sum_{k=1}^{m} |\langle n_k\rangle| \to \infty$,

(ii) $\frac{d}{dt} V(\langle n_1\rangle,...,\langle n_m\rangle) \leqslant -\epsilon < 0$, for all $\sum_{k=1}^{m} |\langle n_k\rangle| \geqslant \delta$, $t \geqslant 0$,

where ϵ is a suitable number (time derivatives are respected to the continuous system of equations considered), are satisfied, then such a system is dissipative in the sense of Levinson[2]; i.e., such a constant D (common for all the solutions of the given system) can be found that

$$\limsup_{t\to\infty} \sum_{k=1}^{m} |\langle n_k(t)\rangle| \leqslant D. \tag{3.1}$$

Hence defining [cf. (2.4)]

$$V(\langle n_1\rangle,...,\langle n_m\rangle) \overset{\text{def.}}{=} \sum_{k=1}^{m} \mathrm{Tr}\,\hat{\rho} c_k \hat{a}_k^+ \hat{a}_k = \sum_{k=1}^{m} c_k \langle \hat{a}_k^+ \hat{a}_k\rangle = \sum_{k=1}^{m} c_k \langle n_k\rangle, \tag{3.2}$$

where

$$\hat{\rho} = \hat{\rho}_s \hat{\rho}_r, \quad \hat{\rho}_s = \int \phi_{\mathcal{N}}(0,\alpha_1,...,\alpha_m) \prod_{j=1}^{m} |\alpha_j\rangle\langle\alpha_j| d^2\alpha_j$$

is the Glauber–Sudarshan representation[6] of the probability density operator for the field in the mixture of pure coherent states[6] $|\alpha_j\rangle$ with a quasiprobability[6] $\phi_{\mathcal{N}}$ as a weight function, it is obvious that ($\hat{\rho}_r$ is related to the reservoir modes)

$$\sum_{k=1}^{m} \langle n_k\rangle = \sum_{k=1}^{m} |\langle n_k\rangle| = \int \phi_{\mathcal{N}} \sum_{k=1}^{m} |\alpha_k|^2 d^2\{\alpha\}, \tag{3.3}$$

where

$$\langle n_k\rangle = \int \phi_{\mathcal{N}} \langle\{\alpha\}|\hat{a}_k^+ \hat{a}_k|\{\alpha\}\rangle d^2\{\alpha\} = \int \phi_{\mathcal{N}} |\alpha_k|^2 d^2\{\alpha\}, \quad \{\alpha\} = (\alpha_1,...,\alpha_m).$$

Expression (3.3) [also included in (3.1) and (3.2)] represents the sum of the mean number of photons in the single modes.

Furthermore, using the Schwarz inequality, we have

$$|\langle \hat{a}_k^+\rangle| = |\langle \hat{a}_k\rangle| = |\mathrm{Tr}\,\hat{\rho}\hat{a}_k^+| = |\mathrm{Tr}\,\hat{\rho}\hat{a}_k|$$
$$= \left|\int \phi_{\mathcal{N}} \langle\{\alpha\}|\hat{a}_k|\{\alpha\}\rangle d^2\{\alpha\}\right|$$
$$= \left|\int \phi_{\mathcal{N}} \alpha_k \, d^2\{\alpha\}\right| \leqslant \int \phi_{\mathcal{N}} |\alpha_k| d^2\{\alpha\}$$
$$\leqslant \left(\int \phi_{\mathcal{N}} |\alpha_k|^2 d^2\{\alpha\}\right)^{1/2} = \sqrt{\langle n_k\rangle}, \tag{3.4}$$

and mainly (see, e.g., Ref. 7, p. 169)

$$\langle \hat{L}_k^+ \hat{a}_k + \hat{a}_k^+ \hat{L}_k\rangle = \langle \hat{L}_k^+ \hat{a}_k + \hat{a}_k^+ \hat{L}_k\rangle_r = \gamma_k \langle n_k^{(d)}\rangle \ (= \text{const}), \tag{3.5}$$

where $\langle n_k^{(d)}\rangle$ is the mean number of the noise photons (more precisely, of the reservoir photons for the k th mode).

Deriving the comparing function of (3.2) with respect to (2.2), we get by means of (2.4), (2.5), and (3.3)–(3.5) the following important inequality:

$$\frac{d}{dt} V(\langle n_1\rangle,...,\langle n_m\rangle)_{(2.2)}$$
$$= \sum_{k=1}^{m} \mathrm{Tr}\,\hat{\rho} c_k \left\{ -\gamma_k \hat{a}_k^+ \hat{a}_k + \hat{L}_k^+ \hat{a}_k + \hat{a}_k^+ \hat{L}_k - i\left[p_k(t)\hat{a}_k^+ - p_k^*(t)\hat{a}_k + \hat{a}_k^+ \frac{\partial F}{\partial \hat{a}_k^+} - \left(\hat{a}_k^+ \frac{\partial F}{\partial \hat{a}_k^+}\right)^+\right]\right\}$$
$$\leqslant -\sum_{k=1}^{m} c_k \gamma_k (\langle n_k\rangle - \langle n_k^{(d)}\rangle) + 2\sum_{k=1}^{m} |p_k(t)| \sqrt{\langle n_k\rangle}. \tag{3.6}$$

IV. FORMAL PROBABILITY EXTENSION FOR STOCHASTIC PUMPING

If the functions $p_k(t)$ are of a stochastic nature and if they are bounded by suitable constants P_k almost everywhere ("a.e.," i.e., except the zero measure set)

$$|p_k(t)| \leqslant P_k \text{ a.e. for } t \geqslant 0, \tag{4.1}$$

then, regarding (3.6), it is clear[3] that condition (ii) of Yoshizawa's theorem is satisfied a.e. as well; namely such positive constants ϵ,δ can be found that

$$\begin{aligned}\frac{d}{dt}V(\langle n_1\rangle,\ldots,\langle n_m\rangle)_{(2.2)} &\leqslant -\sum_{k=1}^{m} c_k\gamma_k\langle n_k\rangle + 2\sum_{k=1}^{m} c_k P_k\sqrt{\langle n_k\rangle} + C\langle n_d\rangle \\ &\leqslant -\epsilon \end{aligned} \tag{4.2}$$

holds a.e. for $\Sigma_{k=1}^{m}\langle n_k\rangle \geqslant \delta$, where

$$C\langle n_d\rangle = \sum_{k=1}^{m} c_k\gamma_k\langle n_k^{(d)}\rangle.$$

Since condition (i) of the same theorem is trivially satisfied, its assertion can be formally extended,[3] namely that the system (2.2) is under (4.1) dissipative with probability 1:

$$\lim_{T\to\infty} P_{\{n\}}\left(\sup_{\infty > t \geqslant T}\sum_{k=1}^{m}\langle n_k(t)\rangle > D\right) = 0, \tag{4.3}$$

where $P_{\{n\}}$ is the Lebesgue–Stieltjes-like probability measure.

Theorem: If $F = F^+$ is a multilinear form of the variables $\hat{a}_1^+,\ldots,\hat{a}_m^+,\hat{a}_1,\ldots,\hat{a}_m^+$, without the unmixed terms (see Sec. II), then the system (2.2) is under (4.1) dissipative in the sense of Levinson with probability 1.

Remark 1: The assumptions of our theorem comprise[6] the majority of the effective quantum optical processes, but those of multilinear absorption and emission.

Remark 2: The theorem asserts that the sum of the average number of photons is ultimately (finally) less than some D. However, using (4.2), we can, at least, approximately estimate this D.

V. APPLICATION TO THE SECOND-HARMONIC GENERATION PROCESS WITH CLASSICAL PUMPING

This process is described (see Ref. 7, p. 226) by the Hamiltonian ($\hbar = 1$)

$$\begin{aligned}\hat{H} = {} & \omega_1\hat{a}_1^+\hat{a}_1 + \omega_2\hat{a}_2^+\hat{a}_2 - g\hat{a}_1^2\hat{a}_2^+ - g^*\hat{a}_1^{+2}\hat{a}_2 \\ & + i\hat{a}_1^+ p(t)\exp(-i\omega_1 t) - i\hat{a}_1 p(t)\exp(i\omega_1 t),\end{aligned}$$

where

$$\hat{H}(\hat{a}_1^+,\hat{a}_2^+,\hat{a}_3^+,\hat{a}_1,\hat{a}_2,\hat{a}_3) = \hat{H}(\hat{a}_1^+,\hat{a}_2^+,\hat{a}_1^+,\hat{a}_1,\hat{a}_2,\hat{a}_1) = \hat{H},$$

$$\operatorname{Im}\omega_1 = 0 = \operatorname{Im}\omega_2, \quad \sup_{t\in\langle 0,\infty\rangle}|p(t)| \leqslant P,$$

and g is a coupling constant. Furthermore, we assume that the sum-frequency mode is pumped only for simplicity.

Thus system (2.2) reads

$$\begin{aligned}\dot{\hat{a}}_1 &= -(i\omega_1 + (\gamma_1/2))\hat{a}_1 + 2ig^*\hat{a}_1^+\hat{a}_2 + p(t)\exp(-i\omega_1 t), \\ \dot{\hat{a}}_2 &= -(i\omega_2 + \gamma_2/2)\hat{a}_2 + ig\hat{a}_1^2. \end{aligned} \tag{5.1}$$

Defining Yoshizawa's function as $V(\langle n_1\rangle,\langle n_2\rangle) = \langle n_1\rangle + 2\langle n_2\rangle$ [i.e., substituting $c_1 = 1, c_2 = 2$ in (3.2) in order for (2.5) to be satisfied], we have, according to (4.2), that

$$\begin{aligned}\frac{d}{dt}V(\langle n_1\rangle,\langle n_2\rangle)_{(5.1)} &\leqslant -\gamma_1\langle n_1\rangle - 2\gamma_2\langle n_2\rangle \\ &\quad + P\sqrt{\langle n_1\rangle} + C\langle n_d\rangle.\end{aligned}$$

From here we can estimate (see Remark 2) the constant D of (3.2) or (4.3) as follows:

$$\begin{aligned}&\limsup_{t\to\infty}(\langle n_1(t)\rangle + \langle n_2(t)\rangle) \\ &\quad \leqslant 2\max(C\langle n_d\rangle, 4P^2/\gamma_1\min(\gamma_1,2\gamma_2)).\end{aligned}$$

Although the last relation determines (rather roughly) the corresponding attracting set, an invariant probability measure of the attractor itself may be much less.

VI. CONCLUSIONS

In the results above we have illustrated a dissipativity (and consequently also a Lagrange-like stability[2]) in the sense of Levinson to the majority of the lossy optical processes. Nevertheless, those need not be stable in the sense of Liapunov. For example, considering just the second-harmonic generation process as above, when $C = 0$ and $p(t) \equiv P$, one of the stationary solutions of (5.1) is (Liapunov-like) nonstable already for $P > \gamma_1\gamma_2/|g|$, while the second one is stable. The same is true even for the more general parametric generation process[8] (when $\hat{a}_1 \neq \hat{a}_3$).

ACKNOWLEDGMENT

I would like to thank Dr. J. Peřina for a stimulating discussion and his comments about the manuscript.

[1] T. Yoshizawa, *Stability Theory by Liapunov's Second Method* (Math. Soc. Japan, Tokyo, 1966).

[2] N. Rouche, P. Habets, and M. Laloy, *Stability Theory by Liapunov's Direct Method* (Springer, Berlin, 1977).

[3] H. J. Kushner, *Stochastic Stability and Control* (Academic, New York, 1967).

[4] R. Graham, "Statistical theory of instabilities in stationary nonequilibrium systems with applications to lasers and nonlinear optics," in *Quantum Statistics in Optics and Solid-State Physics, Springer Tracts in Modern Physics*, Vol. 66 (Springer, Berlin, 1973).

[5] A. R. Bulsara and W. C. Schieve, "A description of laser stability near threshold by Liapunov's second method," in *Coherence and Quantum Optics IV*, edited by L. Mandel and E. Wolf ((Plenum, New York, 1978), pp. 649–666 (see also the references included).

[6] W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).

[7] J. Peřina, *Quantum Statistics of Linear and Nonlinear Optical Phenomena* (Reidel, Dordrecht, 1984).

[8] J. Andres, "On dissipativity of parametric generation process," preprint, 1986.

ERRATUM

Erratum: The Lorentz group and the Thomas precession. II. Exact results for the product of two boosts [J. Math. Phys. 27, 157 (1986)]

N. A. Salingaros
Division of Mathematics, Computer Science and Systems Design, The University of Texas at San Antonio, San Antonio, Texas 78285

(Received 30 September 1986; accepted for publication 8 October 1986)

Reference 1 attempted to derive the exact combination of two finite nonparallel Lorentz boosts and to calculate the rotation correction that is otherwise identified with the Wigner angle. The derivation in Ref. 1 sheds some light on this neglected aspect of special relativity. Some additional references that also derive the Wigner angle from the product of two Lorentz boosts include Refs. 2–6 (see also Ref. 7). The result of two consecutive Lorentz boosts, first by **a**, then by **b**, is given by

$$\mathbf{L}(\mathbf{b}) \vee \mathbf{L}(\mathbf{a}) = \mathbb{R}(\theta) \vee \mathbb{L}(\mathbf{d}), \tag{1}$$

where θ is the rotation correction (Wigner angle). What was a most surprising result in Ref. 1 is that the net boost vector **d** is not equal to the standard combination of the boosts **a** and **b**, which is denoted by **s**. Instead, there is a further rotation correction by another angle ϕ as follows:

$$\begin{aligned}\mathbf{L}(\mathbf{b}) \vee \mathbf{L}(\mathbf{a}) &= \mathbb{R}(\theta) \vee \mathbb{L}(\mathbf{d}) \\ &= \mathbb{R}(\theta) \vee \mathbb{R}(\phi) \vee \mathbb{L}(\mathbf{s}) \vee \mathbb{R}(-\phi).\end{aligned} \tag{2}$$

The extra correction angle ϕ is entirely the result of using the Clifford algebra in Minkowski space-time $\mathbb{N}_4$ (see Ref. 8) to realize the Lorentz Lie algebra.[1,9] It should be emphasized that this realization is isomorphic to the Dirac gamma-matrix realization of the Lorentz Lie algebra, which provides the standard representation of the Lorentz group in field theory.[10,11] The rotation corrections to the product of boosts are not, however, easily derivable using explicit matrix representations except in the infinitesimal case. The additional correction angle ϕ in (2) is entirely absent from the corresponding derivation of the Wigner angle in the Clifford algebra in three-dimensional Euclidean space $\$_1$ (see Ref. 12). That algebra is isomorphic to the Pauli algebra, which has the well-known sigma-matrix representation, and is employed in Refs. 2–6 to calculate the product of Lorentz boosts. In the Pauli algebra $\$_1$, decomposition (1) implies that $\mathbf{d} = \mathbf{s}$, i.e., that $\phi = 0$ in (2).

The puzzle is the following: apparently, the Lorentz group depends upon which algebra is used. Both algebras $\mathbb{N}_4$ and $\$_1$ provide a faithful representation of the Lorentz Lie algebra $\mathbf{so}(1,3)$ via the commutator bracket. Yet the two Clifford algebras are entirely distinct, and $\$_1$ is a subalgebra of $\mathbb{N}_4$ (see Refs. 13 and 14). Included in the Lie algebra $\mathbf{so}(1,5)$ of the Clifford algebra $\mathbb{N}_4$ is a duality rotation that defines a $\mathbf{U}(1)$ group outside the Lorentz group $\mathbf{SO}(1,3)$. This is not true in $\$_1$, which strictly contains the Lie algebra $\mathbf{so}(1,3)$.[13,14] Exponentials of elements of the Lorentz algebra in $\mathbb{N}_4$ therefore generate an intrinsic duality rotation which may be responsible for the additional rotation correction angle ϕ in (2).

The additional angle ϕ identified in Ref. 1 was unfortunately calculated incorrectly in Ref. 1. There is an algebraic error following Eq. (37) of Ref. 1 so that ϕ does not equal $-\frac{1}{2}\theta$. Consequently, formulas (39), (40), (42), and (44) are not correct in substituting $-\frac{1}{2}\theta$. This error was pointed out in Ref. 6. The angle ϕ can be calculated from the discussion in Ref. 1, and one obtains the expression

$$\begin{aligned}\tan\phi &= \frac{A\sin\xi - (B + C\cos\xi)\tan\frac{1}{2}\theta}{B + C\cos\xi + A\sin\xi\tan\frac{1}{2}\theta}, \\ A &= y\alpha_s - x\beta_s, \quad B = x\alpha_s + y\beta_s, \\ C &= y\alpha_s + x\beta_s.\end{aligned} \tag{3}$$

In (3) θ is the Wigner angle, and ξ is the angle between the two boosts. The quantities x,y and α_s, β_s are defined in terms of the two boost parameters by Eqs. (27) and (34) of Ref. 1. Note, in particular, that the rotation ϕ is along the axis of the Wigner rotation, i.e., $\hat{\phi} = \hat{\theta}$. As should be expected, $\theta = 0$, $\phi = 0$ for parallel boosts. Nevertheless, the interesting special case of equal orthogonal boosts gives $\theta =$ maximum, $\phi = 0$. Otherwise, $\phi \neq 0$.

As the Lorentz group description in $\mathbb{N}_4$, which naturally includes the additional rotation correction ϕ, is isomorphic to the standard gamma-matrix representation of the Lorentz group,[10,11] I do not agree with the authors of Refs. 6 that Eq. (2) is mistaken. Rather, one has to seriously address the possible physical validity of the additional angle ϕ, and the important question of which Clifford algebra correctly describes the physical Lorentz group.

ACKNOWLEDGMENT

I would like to thank Professor W. E. Baylis for communicating his results to me prior to publication.

[1] N. Salingaros, J. Math Phys. **27**, 157 (1986).
[2] A. J. MacFarlane, J. Math. Phys. **3**, 1116 (1962).
[3] C. B. van Wyk, Am. J. Phys. **52**, 853 (1984).
[4] M. Rivas, M. A. Valle, and J. M. Aguirregabiria, European J. Phys. **7**, 1 (1986).
[5] A. C. Hirschfeld and F. Metzger, Am. J. Phys. **54**, 550 (1986).
[6] W. E. Baylis and G. Jones, University of Windsor, preprint.

 0022-2488/87/020492-02$02.50 © 1987 American Institute of Physics

[7]C. B. van Wyk, J. Math. Phys. **27**, 1306, 1311 (1986).
[8]N. Salingaros and M. Dresden, Adv. Appl. Math. **4**, 1 (1983).
[9]N. Salingaros, J. Math. Phys. **25**, 706 (1984).
[10]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).
[11]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[12]N. Salingaros and Y. Ilamed, Found. Phys. **14**, 777 (1984).
[13]N. Salingaros, J. Math. Phys. **22**, 226 (1981).
[14]N. Salingaros, J. Math. Phys. **23**, 1, 1231 (1982).