Multivariable biorthogonal continuous–discrete Wilson and Racah polynomials

M. V. Tratnik[a)]
Institute for Mathematics and its Applications, University of Minnesota, 514 Vincent Hall, 206 Church Street S. E., Minneapolis, Minnesota 55455

(Received 22 June 1989; accepted for publication 7 March 1990)

Several families of multivariable, biorthogonal, partly continuous and partly discrete, Wilson polynomials are presented. These yield limit cases that are purely continuous in some of the variables and purely discrete in the others, or purely discrete in all the variables. The latter are referred to as the multivariable biorthogonal Racah polynomials. Interesting further limit cases include the multivariable biorthogonal Hahn and dual Hahn polynomials.

I. INTRODUCTION

The Wilson polynomials[1,2] are a very general family that include as special or limiting cases all the classical orthogonal polynomials and many related families. They can be expressed as the following ${}_4F_3$ hypergeometric series:[3]

$$P_n(x) = (a+b)_n(a+c)_n(a+d)_n\ {}_4F_3\left(\begin{matrix} -n,n+a+b+c+d-1,a-ix,a+ix \\ a+b,a+c,a+d \end{matrix};1\right), \tag{1.1}$$

where a,b,c,d are complex parameters, $(\alpha)_n \equiv \Gamma(n+\alpha)/\Gamma(\alpha)$ denotes the usual Pochhammer symbol, and n is a non-negative integer. These are polynomials in x of degree $2n$ that one can show[1] are symmetric in all four of the parameters a,b,c,d. They are associated with the following weight function:

$$w(x) = \frac{\Gamma(a+ix)\Gamma(a-ix)\Gamma(b+ix)\Gamma(b-ix)\Gamma(c+ix)\Gamma(c-ix)\Gamma(d+ix)\Gamma(d-ix)}{\Gamma(2ix)\Gamma(-2ix)}, \tag{1.2}$$

and satisfy a complex orthogonality relation;

$$\int_C dx\, P_n(x)P_m(x)w(x) = \lambda_n\delta_{nm}, \tag{1.3}$$

where the normalization constant λ_n is given by

$$\lambda_n = 4\pi n!(n+a+b+c+d-1)_n \frac{\Gamma(n+a+b)\Gamma(n+a+c)\Gamma(n+a+d)\Gamma(n+b+c)\Gamma(n+b+d)\Gamma(n+c+d)}{\Gamma(2n+a+b+c+d)}, \tag{1.4}$$

and the contour C is deformed from the real axis so that it separates the increasing sequences of poles of the weight function from the decreasing sequences.

When the real parts of the parameters a,b,c,d are positive one can choose C to be the real axis. If in addition the parameters are real or occur in complex conjugate pairs then the polynomials and weight function are real and the latter is positive. In this case the Wilson polynomials satisfy a continuous orthogonality relation with respect to a positive measure on the real line. If the real part of one parameter is less than zero, let us say $\mathrm{Re}(a)<0$, $\mathrm{Re}(b,c,d)>0(2a,a+b,a+c,a+d \neq 0,-1,-2,\ldots)$, then C must be deformed from the real axis to pass over the decreasing sequence of poles given by $x=-ia-ij, j=0,1,2,\ldots$ and under the increasing sequence found at $x=ia+ij, j=0,1,2,\ldots$. This contour can then be deformed back to the real axis plus closed loops about a finite number of poles. If the closed loops are then evaluated by the method of residues the inner product in (1.3) can be written as an integral over the real axis plus a finite discrete sum. In this case the Wilson polynomials satisfy a partly continuous and partly discrete orthogonality relation

$$\int_{-\infty}^{\infty} dx\, w(x)P_n(x)P_m(x) + \sum_{j=0}^{\mathrm{Re}(a)+j<0} \tilde{w}(j)P_n(ia+ij)P_m(ia+ij) = \lambda_n\delta_{nm}, \tag{1.5}$$

where the discrete part of the weight function $\tilde{w}(j)$ is given by

$$\tilde{w}(j) = (4\pi)\Gamma(a+b)\Gamma(a+c)\Gamma(a+d)\frac{\Gamma(b-a)\Gamma(c-a)\Gamma(d-a)}{\Gamma(-2a)}\frac{(2a)_j(a+1)_j(a+b)_j(a+c)_j(a+d)_j}{(a)_j(a-b+1)_j(a-c+1)_j(a-d+1)_j j!}. \tag{1.6}$$

Formula (1.5) also yields a purely discrete orthogonality relation. Take $a+b=-\Delta+\epsilon$, where Δ is a non-negative integer, divide (1.5) by $\Gamma(a+b)=\Gamma(-\Delta+\epsilon)$, and then take the limit $\epsilon\to 0$. The continuous term vanishes because $1/\Gamma(-\Delta+\epsilon)\to 0$ but the discrete part survives leaving

[a)] Permanent address: CNLS and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

 0022-2488/90/071559-17$03.00

$$\sum_{j=0}^{\Delta} \rho(j) P_n(ia+ij) P_m(ia+ij) = \lambda'_n \delta_{nm}, \quad 0 \leqslant n,m \leqslant \Delta, \tag{1.7}$$

where the weight function and normalization constant are given by

$$\rho(j) = \frac{(2a)_j (a+1)_j (-\Delta)_j (a+c)_j (a+d)_j}{(a)_j (2a+\Delta+1)_j (a-c+1)_j (a-d+1)_j j!},$$

$$\lambda'_n = n!(n-\Delta+c+d-1)_n (2a+1)_\Delta (1-c-d)_\Delta \frac{(-\Delta)_n (a+c)_n (a+d)_n (c-a-\Delta)_n (d-a-\Delta)_n (c+d)_n}{(-\Delta+c+d)_{2n} (a-c+1)_\Delta (a-d+1)_\Delta}. \tag{1.8}$$

The orthogonality relation (1.7) is equivalent to Racah's orthogonality for what are called Racah coefficients or $6j$ symbols. Accordingly the polynomials in (1.7) are referred to in the literature as the Racah polynomials. It is customary to redefine the parameters and write these polynomials as

$$r_n(x) = (\alpha+1)_n (\gamma+1)_n (\beta+\delta+1)_n \, {}_4F_3\left(\begin{matrix} -n, n+\alpha+\beta+1, -x, x+\gamma+\delta+1 \\ \alpha+1, \beta+\delta+1, \gamma+1 \end{matrix} ;1 \right), \tag{1.9}$$

and then the orthogonality relation becomes

$$\sum_{x=0}^{\Delta} \frac{(\gamma+\delta+1)_x (\gamma/2+\delta/2+3/2)_x (\alpha+1)_x (\beta+\delta+1)_x (\gamma+1)_x}{(\gamma/2+\delta/2+1/2)_x (\gamma+\delta-\alpha+1)_x (\gamma-\beta+1)_x (\delta+1)_x x!} r_n(x) r_m(x) = \lambda_n \delta_{nm},$$

$$\lambda_n = n!(\alpha+1)_n (\beta+1)_n (\gamma+1)_n (\alpha-\delta+1)_n (\alpha+\beta-\gamma+1)_n (\beta+\delta+1)_n \tag{1.10}$$

$$\times \frac{(n+\alpha+\beta+1)_n}{(\alpha+\beta+2)_{2n}} \frac{\Gamma(\gamma+\delta-\alpha+1)\Gamma(-\beta-\alpha-1)\Gamma(\gamma-\beta+1)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+2)\Gamma(-\beta)\Gamma(\gamma-\beta-\alpha)\Gamma(\delta-\alpha)},$$

where $\alpha+1$, $\beta+\delta+1$, or $\gamma+1 = -\Delta$, and $0 \leqslant n,m \leqslant \Delta$.

Two interesting limit cases are the Hahn and dual Hahn polynomials. The limit $\delta \to \infty$ with $\gamma+1 = -\Delta$ gives the Hahn polynomial orthogonality

$$\sum_{x=0}^{\Delta} \frac{(\alpha+1)_x (-\Delta)_x}{(-\Delta-\beta)_x x!} h_n(x) h_m(x) = \lambda_n \delta_{nm}, \tag{1.11}$$

where

$$h_n(x) = {}_3F_2\left(\begin{matrix} -n, n+\alpha+\beta+1, -x \\ \alpha+1, -\Delta \end{matrix} ;1 \right), \quad 0 \leqslant n \leqslant \Delta,$$

$$\lambda_n = \left(\frac{(\Delta-n)!n!}{\Delta!} \right) \left(\frac{(n+\alpha+\beta+1)_\Delta (\beta+1)_n}{(\alpha+1)_n (\beta+1)_\Delta} \right) \left(\frac{(\Delta+n+\alpha+\beta+1)}{(2n+\alpha+\beta+1)} \right). \tag{1.12}$$

Letting $\beta \to \infty$ in (1.10) with $\alpha+1 = -\Delta$ gives the dual Hahn orthogonality

$$\sum_{x=0}^{\Delta} \binom{\Delta}{x} \frac{(\gamma+\delta+1)_x (\gamma/2+\delta/2+3/2)_x (\gamma+1)_x}{(\gamma/2+\delta/2+1/2)_x (\delta+1)_x (\Delta+\gamma+\delta+2)_x} d_n(x) d_m(x) = \lambda_n \delta_{nm}, \tag{1.13}$$

with

$$d_n(x) = {}_3F_2\left(\begin{matrix} -n, -x, x+\gamma+\delta+1 \\ -\Delta, \gamma+1 \end{matrix} ;1 \right), \quad 0 \leqslant n \leqslant \Delta,$$

$$\lambda_n = \left(\frac{(\Delta-n)!n!}{\Delta!} \right) \left(\frac{(\gamma+\delta+2)_\Delta}{(\gamma+1)_n (\delta+1)_{\Delta-n}} \right). \tag{1.14}$$

A generalization of the Wilson polynomials to p variables $x_1, x_2, \ldots, x_p$ is given by the following four families:[4]

$$P\left(\begin{matrix} x_1, x_2, \ldots, x_p \\ n_1, n_2, \ldots, n_p \end{matrix} \middle| \begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_p \end{matrix} \middle| c,d \right) = \left[\prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{\Gamma(a_k+b_k)} \right] (A+c)_N (A+d)_N$$

$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1} \left(\begin{matrix} N+A+B+c+d-1, A-iX: -n_1, a_1+ix_1; \ldots; -n_p, a_p+ix_p \\ A+c, A+d: a_1+b_1; \ldots; a_p+b_p \end{matrix} \right), \tag{1.15}$$

$$\overline{P}\left(\begin{matrix} x_1, x_2, \ldots, x_p \\ n_1, n_2, \ldots, n_p \end{matrix} \middle| \begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_p \end{matrix} \middle| c,d \right) = \left[\prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{\Gamma(a_k+b_k)} \right] (B+c)_N (B+d)_N$$

$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1} \left(\begin{matrix} N+A+B+c+d-1, B+iX: -n_1, b_1-ix_1; \ldots; -n_p, b_p-ix_p \\ B+c, B+d: a_1+b_1; \ldots; a_p+b_p \end{matrix} \right), \tag{1.16}$$

$$Q\begin{pmatrix} x_1,x_2,\ldots,x_p \\ n_1,n_2,\ldots,n_p \end{pmatrix}\begin{vmatrix} a_1,a_2,\ldots,a_p \\ b_1,b_2,\ldots,b_p \end{vmatrix} c,d\Bigg) = \left[\prod_{k=1}^{p}\frac{\Gamma(n_k+a_k+b_k)}{\Gamma(a_k+b_k)}\right](c-iX)_N(d-iX)_N$$

$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1}\begin{pmatrix} -N-c-d+1,B+iX: -n_1,a_1+ix_1;\ldots;-n_p,a_p+ix_p \\ -N-c+iX+1,-N-d+iX+1:a_1+b_1;\ldots;a_p+b_p \end{pmatrix}, \quad (1.17)$$

$$\overline{Q}\begin{pmatrix} x_1,x_2,\ldots,x_p \\ n_1,n_2,\ldots,n_p \end{pmatrix}\begin{vmatrix} a_1,a_2,\ldots,a_p \\ b_1,b_2,\ldots,b_p \end{vmatrix} c,d\Bigg) = \left[\prod_{k=1}^{p}\frac{\Gamma(n_k+a_k+b_k)}{\Gamma(a_k+b_k)}\right](c+iX)_N(d+iX)_N$$

$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1}\begin{pmatrix} -N-c-d+1,A-iX: -n_1,b_1-ix_1;\ldots;-n_p,b_p-ix_p \\ -N-c-iX+1,-N-d-iX+1:a_1+b_1;\ldots;a_p+b_p \end{pmatrix}, \quad (1.18)$$

where $a_1,a_2,\ldots,a_p,b_1,b_2,\ldots,b_p,c,d$ are complex parameters and $F^{q:\ell_1;\ldots;\ell_p}_{r:v_1;\ldots;v_p}$ is the generalized multivariable Kampé de Fériet hypergeometric series[5] defined as

$$F^{q:\ell_1;\ldots;\ell_p}_{r:v_1;\ldots;v_p}\begin{pmatrix} \alpha_1,\ldots,\alpha_q:\beta^{(1)}_1,\ldots,\beta^{(1)}_{\ell_1};\ldots;\beta^{(p)}_1,\ldots,\beta^{(p)}_{\ell_p}; \\ \gamma_1,\ldots,\gamma_r:\xi^{(1)}_1,\ldots,\xi^{(1)}_{v_1};\ldots;\xi^{(p)}_1,\ldots,\xi^{(p)}_{v_p}; \end{pmatrix} z_1,z_2,\ldots,z_p\Bigg)$$

$$= \sum_{\{j_k\}} \frac{\Pi^q_{i=1}(\alpha_i)_J \Pi^{\ell_1}_{i=1}(\beta^{(1)}_i)_{j_1}\cdots\Pi^{\ell_p}_{i=1}(\beta^{(p)}_i)_{j_p}}{\Pi^r_{i=1}(\gamma_i)_J \Pi^{v_1}_{i=1}(\xi^{(1)}_i)_{j_1}\cdots\Pi^{v_p}_{i=1}(\xi^{(p)}_i)_{j_p}} \frac{z_1^{j_1}}{j_1!}\frac{z_2^{j_2}}{j_2!}\cdots\frac{z_p^{j_p}}{j_p!}, \quad (1.19)$$

where $\{j_k\}$ denotes summation indices $j_1,j_2,\ldots,j_p$ that run over all non-negative integers and we have introduced the following shorthand notation:

$$X\equiv\sum_{k=1}^{p}x_k,\quad N\equiv\sum_{k=1}^{p}n_k,\quad J\equiv\sum_{k=1}^{p}j_k,\quad A\equiv\sum_{k=1}^{p}a_k,\quad B\equiv\sum_{k=1}^{p}b_k, \quad (1.20)$$

and in the absence of specifying the arguments $z_1,z_2,\ldots,z_p$, unity is to be understood. The overbars in (1.16) and (1.18) denote distinct families of polynomials and should not be confused with complex conjugation. The p tuple of non-negative integers $n_1,n_2,\ldots,n_p$ labels the different polynomials whose degrees are given by $2N$ where N is defined in (1.20). These polynomials are associated with the following multivariable weight function:

$$w\begin{pmatrix} x_1,x_2,\ldots,x_p \end{pmatrix}\begin{vmatrix} a_1,a_2,\ldots,a_p \\ b_1,b_2,\ldots,b_p \end{vmatrix} c,d\Bigg)$$

$$= \left[\prod_{k=1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\frac{\Gamma(A-iX)\Gamma(B+iX)\Gamma(c+iX)\Gamma(c-iX)\Gamma(d+iX)\Gamma(d-iX)}{\Gamma(2iX)\Gamma(-2iX)}, \quad (1.21)$$

which notice is symmetric under the interchange of c and d as are all four families of polynomials. When no ambiguity arises we simply write $P_n(x)$, $\overline{P}_n(x)$, $Q_n(x)$, $\overline{Q}_n(x)$, and $w(x)$ for the polynomials and weight function, respectively.

These satisfy the following biorthogonality relations[4]

$$P_n\cdot Q_m = \overline{P}_n\cdot\overline{Q}_m = \lambda_n\prod_{k=1}^{p}\delta_{n_k m_k},\quad P_n\cdot\overline{P}_m = Q_n\cdot\overline{Q}_m = 0,\quad \text{if } N\neq M, \quad (1.22)$$

where the normalization constant is given by

$$\lambda_n = 2(2\pi)^p\left[\prod_{k=1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right]\frac{\Gamma(N+A+c)\Gamma(N+A+d)\Gamma(N+B+c)\Gamma(N+B+d)\Gamma(N+c+d)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)}, \quad (1.23)$$

and for positive real parts of the parameters the inner product is defined as

$$P_n\cdot Q_m \equiv \int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_1\, w(x_1\cdots x_p)P_n(x_1\cdots x_p)Q_m(x_1\cdots x_p),\quad \mathrm{Re}(a_1,a_2,\ldots,a_p,b_1,b_2,\ldots,b_p,c,d)>0, \quad (1.24)$$

where the integration contours are simply the real axes.

In Sec. II we extend these purely continuous multivariable biorthogonal Wilson polynomials to several "mixed" cases where the inner product is partly continuous and partly discrete. In Sec. III we discuss the purely discrete family which are the multivariable biorthogonal Racah polynomials. Taking appropriate limits then yields multivariable biorthogonal Hahn and dual Hahn polynomials.

II. MULTIVARIABLE MIXED-TYPE INNER PRODUCTS

The biorthogonality relations (1.22) are still valid for negative real parts of the parameters provided each of the contours are suitably deformed to separate the increasing sequences of poles of the weight function from the decreasing sequences, assuming these two sets are disjoint. However, due to the multiple integrals involved it is not always clear where the poles lie in each of the variables and what the appropriate contours are.

We consider several specific cases, the first being the following parameter domain:

$$\begin{aligned}&\mathrm{Re}(a_1,a_2,\ldots,a_p,b_1,b_2,\ldots,b_p,d)>0,\quad \mathrm{Re}(c)<0,\\&\mathrm{Re}(a_1+c),\quad \mathrm{Re}(b_1+c)>0,\quad \mathrm{Re}(c),\quad 2c,\quad c+d\neq 0,-1,-2,\ldots,\end{aligned}\tag{2.1}$$

for which the x_1 contour C_1 is deformed from the real axis to pass above the decreasing sequence of poles given by $x_1 = -X_2^p - ic - ij, j = 0,1,2,\ldots$ and underneath the increasing sequence found at $x_1 = -X_2^p + ic + ij, j = 0,1,2,\ldots$, and also above and below the remaining decreasing and increasing sequences, respectively. We have introduced the following shorthand notation to denote partial sums:

$$X_j^l \equiv \sum_{k=j}^{l} x_k,\quad A_j^l \equiv \sum_{k=j}^{l} a_k,\quad B_j^l \equiv \sum_{k=j}^{l} b_k.\tag{2.2}$$

If we furthermore choose C_1 sufficiently close to (and under) $-X_2^p + ic$ and also sufficiently close to (and over) $-X_2^p - ic$, then the remaining contours $C_2,\ldots,C_p$ can be chosen on the real axes.

Let us first demonstrate that this choice of contours leads to the norm of the weight function as given by (1.23) with $N = 0$. Making a change of variables from $x_1,x_2,\ldots,x_p$ to $X,x_2,\ldots,x_p$ and reversing the order of the integrations gives

$$\begin{aligned}&\int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \int_{C_1} dx_1\, w(x_1,\ldots,x_p)\\&\quad=\int_C dX \frac{\Gamma(A-iX)\Gamma(B+iX)\Gamma(c+iX)\Gamma(c-iX)\Gamma(d+iX)\Gamma(d-iX)}{\Gamma(2iX)\Gamma(-2iX)}\\&\quad\quad\times\int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_p\, \Gamma(a_1+iX-iX_2^p)\Gamma(b_1-iX+iX_2^p)\left[\prod_{k=2}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right],\end{aligned}\tag{2.3}$$

where the contour C passes over the decreasing sequence of poles $X = -ic - ij, j = 0,1,2,\ldots$ and under the increasing sequence $X = ic + ij, j = 0,1,2,\ldots$, and also above and below the remaining decreasing and increasing sequences, respectively. To evaluate the $x_2,x_3,\ldots,x_p$ integrations we introduce the following single variable integral formula:[3]

$$\begin{aligned}&\int_{C'} dx\, \Gamma(\alpha+ix)\Gamma(\beta+ix)\Gamma(\gamma-ix)\Gamma(\delta-ix) = (2\pi)\frac{\Gamma(\alpha+\gamma)\Gamma(\alpha+\delta)\Gamma(\beta+\gamma)\Gamma(\beta+\delta)}{\Gamma(\alpha+\beta+\gamma+\delta)}\\&\qquad\qquad\alpha+\gamma,\quad \alpha+\delta,\quad \beta+\gamma,\quad \beta+\delta\neq 0,-1,-2,\ldots,\end{aligned}\tag{2.4}$$

where the contour C' separates the increasing sequences of poles of the integrand from the decreasing sequences; in the special case when $\mathrm{Re}(\alpha,\beta,\gamma,\delta)>0$, C' can be chosen on the real axis. Returning to (2.3) and recalling that $\mathrm{Re}(a_1+c)$, $\mathrm{Re}(b_1+c)>0$, the contour C is assumed to pass sufficiently close to (and over) $-ic$ so that $\mathrm{Re}(a_1+iX)>0$ and also sufficiently close to (and under) $+ic$ so that $\mathrm{Re}(b_1-iX)>0$. In this case formula (2.4) with C' on the real axis and induction can be used to evaluate the $x_2,x_3,\ldots,x_p$ integrations leading to

$$\begin{aligned}&\int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_p\, \Gamma(a_1+iX-iX_2^p)\Gamma(b_1-iX+iX_2^p)\left[\prod_{k=2}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\\&\quad=(2\pi)^{p-1}\left[\prod_{k=1}^{p}\Gamma(a_k+b_k)\right]\frac{\Gamma(A+iX)\Gamma(B-iX)}{\Gamma(A+B)},\end{aligned}\tag{2.5}$$

and if this is substituted into (2.3) the norm of the weight function becomes

$$\begin{aligned}&\int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \int_{C_1} dx_1\, w(x_1,\ldots,x_p) = (2\pi)^{p-1}\left[\prod_{k=1}^{p}\Gamma(a_k+b_k)\right][\Gamma(A+B)]^{-1}\int_C dX\, \Gamma(A+iX)\Gamma(A-iX)\\&\qquad\times\frac{\Gamma(B+iX)\Gamma(B-iX)\Gamma(c+iX)\Gamma(c-iX)\Gamma(d+iX)\Gamma(d-iX)}{\Gamma(2iX)\Gamma(-2iX)},\end{aligned}\tag{2.6}$$

which is simply proportional to the single variable integral given by (1.2)–(1.4) with $n = m = 0$. Using this result in (2.6) then yields the multivariable norm as defined in (1.23) with $N = 0$.

Next we express the multiple integral in the left of (2.6) as a finite discrete sum and integrations over the real axes. This is achieved by deforming C_1 to the real axis plus closed loops about a finite number of poles and then evaluating the latter by the method of residues. This leads to two additional terms involving discrete sums that by using (2.5) one can show are equal. The norm then becomes

$$\int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \int_{C_1} dx_1\, w(x_1,\ldots,x_p) = \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_1\, w(x_1,\ldots,x_p) + \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \sum_{j=0}^{\mathrm{Re}(c)+j<0} \widetilde{w}(j,x_2,\ldots,x_p), \tag{2.7}$$

where all the integrals on the right are over the real axes and the "mixed" weight function in the second term is given by

$$\widetilde{w}(j,x_2,\ldots,x_p) = (4\pi)\Gamma(a_1 - c - j - iX_2^p)\Gamma(b_1 + c + j + iX_2^p)\left[\prod_{k=2}^{p} \Gamma(a_k + ix_k)\Gamma(b_k - ix_k)\right] \times \frac{\Gamma(A+c+j)\Gamma(B-c-j)\Gamma(2c+j)\Gamma(d-c-j)\Gamma(d+c+j)}{\Gamma(2c+2j)\Gamma(-2c-2j)} \frac{(-1)^j}{j!}. \tag{2.8}$$

The first purely continuous term in the right of (2.7) is that which arises for positive real parts of the parameters. The second mixed term represents the contribution arising from the negative real part of the parameter c.

The inner product of the multivariable Wilson polynomials is defined as

$$P_n \cdot Q_m \equiv \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \int_{C_1} dx_1\, w(x_1,\ldots,x_p) P_n(x_1,\ldots,x_p) Q_m(x_1,\ldots,x_p), \tag{2.9}$$

which in analogy with (2.7) can be written as

$$P_n \cdot Q_m = \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_1\, w(x_1,\ldots,x_p) P_n(x_1,\ldots,x_p) Q_m(x_1,\ldots,x_p) + \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \times \sum_{j=0}^{\mathrm{Re}(c)+j<0} \widetilde{w}(j,x_2,\ldots,x_p) P_n(-X_2^p + ic + ij, x_2,\ldots,x_p) Q_m(-X_2^p + ic + ij, x_2,\ldots,x_p), \tag{2.10}$$

and similarly for $\overline{P}_n \cdot \overline{Q}_m$, $P_n \cdot \overline{P}_m$, and $Q_n \cdot \overline{Q}_m$. Having verified the norm of the weight function it then follows by the same proof as for the purely continuous family[4] that these inner products satisfy biorthogonality relations (1.22).

Formula (2.10) also yields a simpler and even more interesting mixed type inner product. Take $c + d = -\Delta_1 + \epsilon$, where Δ_1 is a non-negative integer, divide the biorthogonality relations (1.22) by $\Gamma(c+d) = \Gamma(-\Delta_1 + \epsilon)$, and then take the limit $\epsilon \to 0$. The first purely continuous term in (2.10) vanishes because $1/\Gamma(-\Delta_1 + \epsilon) \to 0$ but the second mixed term survives leaving (writing x_1 in place of j)

$$P_n^{(1)} \cdot Q_m^{(1)} \equiv \int_{-\infty}^{\infty} dx_p \cdots \int_{-\infty}^{\infty} dx_2 \sum_{x_1=0}^{\Delta_1} w^{(1)}(x_1,x_2,\ldots,x_p) P_n^{(1)}(x_1,x_2,\ldots,x_p) Q_m^{(1)}(x_1,x_2,\ldots,x_p),$$

$$P_n^{(1)} \cdot Q_m^{(1)} = \overline{P}_n^{(1)} \cdot \overline{Q}_m^{(1)} = \lambda_n^{(1)} \prod_{k=1}^{p} \delta_{n_k m_k}, \quad P_n^{(1)} \cdot \overline{P}_m^{(1)} = Q_n^{(1)} \cdot \overline{Q}_m^{(1)} = 0, \quad \text{if } N \neq M, \tag{2.11}$$

where the mixed weight function and normalization constant are given by

$$w^{(1)}(x_1,\ldots,x_p) = \Gamma(a_1 + d + \Delta_1 - x_1 - iX_2^p)\Gamma(b_1 + c + x_1 + iX_2^p)\left[\prod_{k=2}^{p} \Gamma(a_k + ix_k)\Gamma(b_k - ix_k)\right] \times \left(\frac{\Gamma(A+c+x_1)}{\Gamma(A+c)}\right)\left(\frac{\Gamma(B+d+\Delta_1-x_1)}{\Gamma(B+d)}\right)\left(\frac{\Gamma(1+2d+2\Delta_1-2x_1)}{\Gamma(1+2d+2\Delta_1-x_1)}\right) \times \left(\frac{\Gamma(2d+\Delta_1-x_1)}{\Gamma(2d+2\Delta_1-2x_1)}\right)\binom{\Delta_1}{x_1}(-1)^{x_1}, \tag{2.12}$$

$$\lambda_n^{(1)} = (2\pi)^{p-1}\left[\prod_{k=1}^{p} \Gamma(n_k + a_k + b_k) n_k!\right] \frac{\Delta_1!}{(\Delta_1 - N)!}(-1)^N (A+c)_N (B+d)_N \times \frac{\Gamma(N+A+d)\Gamma(N+B+c)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)}, \tag{2.13}$$

and the polynomials are defined as

$$\begin{aligned} P_n^{(1)}(x_1,\ldots,x_p) &\equiv P_n(-X_2^p + ic + ix_1, x_2,\ldots,x_p), \quad N = 0,1,2,\ldots,\infty, \\ \overline{P}_n^{(1)}(x_1,\ldots,x_p) &\equiv \overline{P}_n(-X_2^p + ic + ix_1, x_2,\ldots,x_p), \quad N = 0,1,2,\ldots,\infty, \\ Q_n^{(1)}(x_1,\ldots,x_p) &\equiv Q_n(-X_2^p + ic + ix_1, x_2,\ldots,x_p), \quad 0 < N < \Delta_1, \\ \overline{Q}_n^{(1)}(x_1,\ldots,x_p) &\equiv \overline{Q}_n(-X_2^p + ic + ix_1, x_2,\ldots,x_p), \quad 0 < N < \Delta_1, \end{aligned} \tag{2.14}$$

with $c + d = -\Delta_1$. These biorthogonality relations can be verified independently of the limiting procedure. To calculate the norm of the weight function one uses (2.4) and induction to perform the $x_2,x_3,...,x_p$ integrations and then the following summation theorem:[3]

$$
{}_5F_4\left(\begin{matrix}2\alpha,\alpha+1,\alpha+\beta,\alpha+\gamma,\alpha+\delta\\ \alpha,\alpha-\beta+1,\alpha-\gamma+1,\alpha-\delta+1\end{matrix};1\right)
=\frac{\Gamma(\alpha-\beta+1)\Gamma(\alpha-\gamma+1)\Gamma(\alpha-\delta+1)\Gamma(-\alpha-\beta-\gamma-\delta+1)}{\Gamma(2\alpha+1)\Gamma(-\beta-\gamma+1)\Gamma(-\beta-\delta+1)\Gamma(-\gamma-\delta+1)},\quad \mathrm{Re}(\alpha+\beta+\gamma+\delta)<1, \tag{2.15}
$$

to evaluate the x_1 sum. Having calculated the norm the biorthogonality relations can then be verified in the same manner as was proved for the purely continuous family.[4] Also, some of the original restrictions in (2.1) can be removed from (2.11) leaving only

$$
\mathrm{Re}(a_1,a_2,...,a_p,b_1,b_2,...,b_p,d)>0,\quad \mathrm{Re}(b_1+c)>0,\quad c+d=-\Delta_1, \tag{2.16}
$$

where recall Δ_1 is a non-negative integer.

We consider a further generalization of (2.11)–(2.14) by also allowing some or all of the a parameters to have negative real parts. The parameter domain in this case is defined by

$$
\begin{aligned}
&\mathrm{Re}(a_1,a_2,...,a_r)<0,\quad \mathrm{Re}(a_{r+1},...,a_p,b_1,...,b_p,d)>0,\\
&\mathrm{Re}(b_1+c)>0,\quad \mathrm{Re}(A_1^r+d)>0,\quad c+d=-\Delta_1,\\
&\mathrm{Re}(a_2,...,a_r),\quad a_2+b_2,...,a_r+b_r\neq 0,-1,-2,...,\quad r=1,2,...,p,
\end{aligned} \tag{2.17}
$$

for which the contours $C_2,...,C_r$, that in (2.11) were on the real axes, are here deformed to pass underneath the increasing sequences of poles $x_k=ia_k+ij_k, j_k=0,1,2,...,k=2,3,...,r$, while still passing above the decreasing sequences. We furthermore choose these contours sufficiently close to (and under) ia_k, $k=2,3,...,r$, so that $\mathrm{Re}(a_1+d-iX_2^r)>0$, which is always possible in light of the restriction $\mathrm{Re}(A_1^r+d)>0$. In this case the remaining contours $C_{r+1},...,C_p$ can be chosen to lie on the real axes. The norm of the weight function is calculated in the same manner as was done for (2.11). That is, one uses (2.4) and induction to perform the $x_2,...,x_p$ integrations and then the ${}_5F_4$ summation theorem (2.15) to evaluate the x_1 sum resulting in (2.13) with $N=0$. Having verified the norm one can then prove the biorthogonality relations (2.11) in the same manner as was done for the purely continuous family,[4] but with the more general inner product defined above or as re-expressed in (2.18).

The contour integrals can be transformed to discrete sums and real integrations by deforming $C_2,...,C_r$ to the real axes plus closed loops about a finite number of poles and then evaluating the latter by the method of residues. The inner product of the polynomials can then be schematically written as

$$
P_n^{(1)}\cdot Q_m^{(1)}\equiv\int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_{r+1}\left(\prod_{k=2}^{r}\left\{\int_{-\infty}^{\infty}dx_k+(2\pi i)\sum_{j_k=0}^{\mathrm{Re}(a_k)+j_k<0}\mathrm{res}(x_k=ia_k+ij_k)\right\}\right)
\times\sum_{x_1=0}^{\Delta_1}w^{(1)}(x_1,...,x_p)P_n^{(1)}(x_1,...,x_p)Q_m^{(1)}(x_1,...,x_p), \tag{2.18}
$$

and similarly for $\overline{P}_n^{(1)}\cdot\overline{Q}_m^{(1)}$, $P_n^{(1)}\cdot\overline{P}_m^{(1)}$, and $Q_n^{(1)}\cdot\overline{Q}_m^{(1)}$, where $\mathrm{res}(x_k)$ denotes the residue at x_k. The right side of (2.18) represents a multitude of mixed type terms involving integrations over the real axes and finite discrete sums.

This example, which is a generalization of a limit case of (2.10) has itself an interesting limit case. Set $a_k+b_k=-\Delta_k+\epsilon$, $k=2,3,...,r$, where Δ_k are non-negative integers, divide the biorthogonality relations by $\Pi_{k=2}^r\Gamma(a_k+b_k)=\Pi_{k=2}^r\Gamma(-\Delta_k+\epsilon)$, and then take the limit $\epsilon\to0$. Since $1/\Gamma(-\Delta_k+\epsilon)\to0$ the only term in (2.18) that survives is the one with r discrete sums leaving (writing $x_2,...,x_r$ in place of $j_2,...,j_r$ and transforming $x_1\to\Delta_1-x_1$)

$$
\begin{gathered}
P_n^{(2)}\cdot Q_m^{(2)}\equiv\int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_{r+1}\sum_{x_r=0}^{\Delta_r}\cdots\sum_{x_2=0}^{\Delta_2}\sum_{x_1=0}^{\Delta_1}w^{(2)}(x_1,...,x_p)P_n^{(2)}(x_1,...,x_p)Q_m^{(2)}(x_1,...,x_p),\\
P_n^{(2)}\cdot Q_m^{(2)}=\overline{P}_n^{(2)}\cdot\overline{Q}_m^{(2)}=\lambda_n^{(2)}\prod_{k=1}^{p}\delta_{n_km_k},\quad P_n^{(2)}\cdot\overline{P}_m^{(2)}=Q_n^{(2)}\cdot\overline{Q}_m^{(2)}=0,\quad \text{if } N\neq M,
\end{gathered} \tag{2.19}
$$

where the weight function, normalization constant, and polynomials are given by

$$
\begin{aligned}
w^{(2)}&(x_1,...,x_p)\\
&=\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\Gamma(A_1^r+d+X_1^r-iX_{r+1}^p)\Gamma(B_1^r+c+\Delta_1^r-X_1^r+iX_{r+1}^p)\\
&\quad\times\left(\frac{\Gamma(A+c+\Delta_1-x_1)}{\Gamma(A+c)}\right)\left(\frac{\Gamma(B+d+x_1)}{\Gamma(B+d)}\right)\left(\frac{\Gamma(1+2d+2x_1)}{\Gamma(1+2d+\Delta_1+x_1)}\right)\left(\frac{\Gamma(2d+x_1)}{\Gamma(2d+2x_1)}\right)(-1)^{\Delta_1-x_1},
\end{aligned} \tag{2.20}
$$

$$
\lambda_n^{(2)}=(2\pi)^{p-r}\Gamma(n_1+a_1+b_1)n_1!\left[\prod_{k=2}^{r}\frac{\Delta_k!}{(\Delta_k-n_k)!}n_k!(-1)^{n_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right]
$$

$$\times\frac{\Delta_1!}{(\Delta_1-N)!}(-1)^N(A+c)_N(B+d)_N\frac{\Gamma(N+A+d)\Gamma(N+B+c)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)}, \tag{2.21}$$

$$\begin{aligned}
&P_n^{(2)}(x_1,...,x_p)\equiv P_n(-iA_2^r-iX_1^r-X_{r+1}^p-id,ia_2+ix_2,...,ia_r+ix_r,x_{r+1},...,x_p),\\
&\overline{P}_n^{(2)}(x_1,...,x_p)\equiv \overline{P}_n(-iA_2^r-iX_1^r-X_{r+1}^p-id,ia_2+ix_2,...,ia_r+ix_r,x_{r+1},...,x_p),\\
&\qquad 0\leqslant n_k\leqslant\Delta_k,\quad k=2,3,...,r,\quad n_1,n_{r+1},...,n_p=0,1,2,...,\infty,\\
&Q_n^{(2)}(x_1,...,x_p)\equiv Q_n(-iA_2^r-iX_1^r-X_{r+1}^p-id,ia_2+ix_2,...,a_r+ix_r,x_{r+1},...,x_p),\\
&\overline{Q}_n^{(2)}(x_1,...,x_p)\equiv \overline{Q}_n(-iA_2^r-iX_1^r-X_{r+1}^p-id,ia_2+ix_2,...,a_r+ix_r,x_{r+1},...,x_p),\\
&\qquad 0\leqslant n_k\leqslant\Delta_k,\quad k=2,3,...,r,\quad 0\leqslant N\leqslant\Delta_1,
\end{aligned} \tag{2.22}$$

with $c+d=-\Delta_1$ and $a_k+b_k=-\Delta_k$, $k=2,3,...,r$. As an independent verification of the norm of the weight function one uses (2.4) and induction to evaluate the $x_{r+1},...,x_p$ integrations, the following summation theorem[6] to perform the $x_2,...,x_r$ sums,

$$F^{1:1;...;1}_{1:0;...;0}\left(\begin{matrix}\alpha:\beta^{(1)};...;\beta^{(p)};\\ \gamma:-;...;-;\end{matrix}\ 1,1,...,1\right)=\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta^{(1)}-\cdots-\beta^{(p)})}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta^{(1)}-\cdots-\beta^{(p)})}, \tag{2.23}$$

and then theorem (2.15) to evaluate the remaining x_1 sum, resulting in (2.21) with $N=0$. The biorthogonality relations (2.19) can then be independently verified in the manner described for the purely continuous family.[4] Also, the restriction $\mathrm{Re}(a_2,...,a_r)\neq 0,-1,-2,...$, is removable from (2.19)–(2.22).

Returning to the purely continuous family (1.22)–(1.24) we consider another mixed type generalization arising from the following parameter domain:

$$\begin{aligned}
&\mathrm{Re}(a_1,a_2,...,a_r)<0,\quad \mathrm{Re}(a_{r+1},a_{r+2},...,a_p,b_1,b_2,...,b_p,c,d)>0,\\
&\mathrm{Re}(A_1^r+A),\quad \mathrm{Re}(A_1^r+c),\ \mathrm{Re}(A_1^r+d)>0,\quad r=1,2,...,p-1,\\
&\mathrm{Re}(a_1,a_2,...,a_r),\quad a_1+b_1,a_2+b_2,...,a_r+b_r\neq 0,-1,-2,...,
\end{aligned} \tag{2.24}$$

for which the first r contours $C_1,C_2,...,C_r$ are deformed below the real axes to pass underneath the increasing sequences of poles $x_k=ia_k+ij_k$, $k=1,2,...,r$, $j_k=0,1,2,...$, while still passing above the decreasing sequences. If these are chosen sufficiently close to (and under) ia_k, $k=1,2,...,r$ then the remaining contours $C_{r+1},...,C_p$ can be chosen to lie on the real axes.

To show that these contours give (1.23) with $N=0$ for the norm of the weight function we begin with a change of variables from $x_1,x_2,...,x_p$ to $x_1,x_2,...,x_r$, $X_{r+1}^p,x_{r+2},...,x_p$ yielding

$$\begin{aligned}
&\int_{C_1}dx_1\cdots\int_{C_r}dx_r\int_{-\infty}^{\infty}dx_{r+1}\cdots\int_{-\infty}^{\infty}dx_p\, w(x_1,...,x_p)\\
&\quad=\int_{C_1}dx_1\cdots\int_{C_r}dx_r\left[\prod_{k=1}^{r}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\\
&\quad\times\int_{-\infty}^{\infty}dX_{r+1}^p\,\frac{\Gamma(A-iX)\Gamma(B+iX)\Gamma(c+iX)\Gamma(c-iX)\Gamma(d+iX)\Gamma(d-iX)}{\Gamma(2iX)\Gamma(-2iX)}\\
&\quad\times\int_{-\infty}^{\infty}dx_{r+2}\cdots\int_{-\infty}^{\infty}dx_p\,\Gamma(a_{r+1}+iX_{r+1}^p-iX_{r+2}^p)\Gamma(b_{r+1}-iX_{r+1}^p+iX_{r+2}^p)\\
&\quad\times\left[\prod_{k=r+2}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right],
\end{aligned} \tag{2.25}$$

and then the $x_{r+2},...,x_p$ integrations are performed by using (2.4) and induction leading to

$$\begin{aligned}
&\int_{C_1}dx_1\cdots\int_{C_r}dx_r\int_{-\infty}^{\infty}dx_{r+1}\cdots\int_{-\infty}^{\infty}dx_p\, w(x_1,...,x_p)\\
&\quad=(2\pi)^{p-r-1}\left[\prod_{k=r+1}^{p}\Gamma(a_k+b_k)\right][\Gamma(A_{r+1}^p+B_{r+1}^p)]^{-1}\int_{C_1}dx_1\cdots\int_{C_r}dx_r\left[\prod_{k=1}^{r}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\\
&\quad\times\int_{-\infty}^{\infty}dX_{r+1}^p\,\Gamma(A_{r+1}^p+iX_{r+1}^p)\Gamma(B_{r+1}^p-iX_{r+1}^p)\Gamma(A-iX)\Gamma(B+iX)\\
&\quad\times\frac{\Gamma(c+iX)\Gamma(c-iX)\Gamma(d+iX)\Gamma(d-iX)}{\Gamma(2iX)\Gamma(-2iX)}.
\end{aligned} \tag{2.26}$$

Another change of variables from $x_1,...,x_r$, X_{r+1}^p to $X,x_1,...,x_r$ then transforms this into

$$\int_{C_1} dx_1 \cdots \int_{C_r} dx_r \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w(x_1,\dots,x_p)$$
$$= \frac{(2\pi)^{p-r-1}}{\Gamma(A_{r+1}^p + B_{r+1}^p)} \left[\prod_{k=r+1}^{p} \Gamma(a_k + b_k) \right] \int_C dX \frac{\Gamma(A - iX)\Gamma(B + iX)\Gamma(c + iX)\Gamma(c - iX)\Gamma(d + iX)\Gamma(d - iX)}{\Gamma(2iX)\Gamma(-2iX)}$$
$$\times \int_{C_1} dx_1 \cdots \int_{C_r} dx_r \, \Gamma(A_{r+1}^p + iX - iX_1^r)\Gamma(B_{r+1}^p - iX + iX_1^r) \left[\prod_{k=1}^{r} \Gamma(a_k + ix_k)\Gamma(b_k - ix_k) \right], \quad (2.27)$$

where C passes underneath iA_1^r. The contours $C_1,\dots,C_r$, which recall pass underneath the increasing sequences $x_k = ia_k + ij_k$, $k = 1,2,\dots,r$, $j_k = 0,1,2,\dots$, are assumed to pass sufficiently close to (and under) ia_k, $k = 1,2,\dots,r$ so that $\mathrm{Re}(A_{r+1}^p - iX_1^r) > 0$, which is possible on account of $\mathrm{Re}(A_1^r + A) > 0$. Also $\mathrm{Re}(B_{r+1}^p + iX_1^r) > 0$ since $\mathrm{Re}(b_1,\dots,b_p) > 0$ and X_1^r has negative or zero imaginary part. In this case the sequences of poles in the variable X do not cross the real axis and so C can be deformed to this axis. With X real the $x_1,\dots,x_r$ integrations can then be performed by (2.4) and induction giving

$$\int_{C_1} dx_1 \cdots \int_{C_r} dx_r \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w(x_1,\dots,x_p)$$
$$= (2\pi)^{p-1} \left[\prod_{k=1}^{p} \Gamma(a_k + b_k) \right] [\Gamma(A + B)]^{-1} \int_{-\infty}^{\infty} dX \, \Gamma(A + iX)\Gamma(A - iX)\Gamma(B + iX)\Gamma(B - iX)$$
$$\times \frac{\Gamma(c + iX)\Gamma(c - iX)\Gamma(d + iX)\Gamma(d - iX)}{\Gamma(2iX)\Gamma(-2iX)}, \quad (2.28)$$

which is simply proportional to the single variable integral given by (1.2)–(1.4) with $n = m = 0$. Using this result in (2.28) then yields the multivariable norm (1.23) with $N = 0$. Having verified the norm the biorthogonality relations (1.22) then follow by the same proof as for the purely continuous family,[4] but with the more general inner product defined above or as re-expressed in (2.30).

As before the contour integrals in the left of (2.28) can be expressed as multiple finite sums and real integrations by deforming $C_1,C_2,\dots,C_r$ to the real axes plus closed loops about a finite number of poles and then evaluating the latter by the method of residues. The norm can then be schematically written as

$$\int_{C_1} dx_1 \cdots \int_{C_r} dx_r \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w(x_1,\dots,x_p)$$
$$= \left(\prod_{k=1}^{r} \left\{ \int_{-\infty}^{\infty} dx_k + (2\pi i) \sum_{j_k=0}^{\mathrm{Re}(a_k) + j_k < 0} \mathrm{res}(x_k = ia_k + ij_k) \right\} \right) \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w(x_1,\dots,x_p), \quad (2.29)$$

representing a multitude of mixed type terms involving real integrations and finite discrete sums. Accordingly the inner product of the polynomials becomes

$$P_n \cdot Q_m \equiv \left(\prod_{k=1}^{r} \left\{ \int_{-\infty}^{\infty} dx_k + (2\pi i) \sum_{j_k=0}^{\mathrm{Re}(a_k) + j_k < 0} \mathrm{res}(x_k = ia_k + ij_k) \right\} \right)$$
$$\times \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w(x_1,\dots,x_p) P_n(x_1,\dots,x_p) Q_m(x_1,\dots,x_p), \quad (2.30)$$

and similarly for $\overline{P}_n \cdot \overline{Q}_m$, $P_n \cdot \overline{P}_m$, and $Q_n \cdot \overline{Q}_m$.

Formula (2.30) also yields a much simpler mixed type inner product. Set $a_k + b_k = -\Delta_k + \epsilon$, $k = 1,2,\dots,r$, where Δ_k are non-negative integers, divide the biorthogonality relations (1.22) by $\Pi_{k=1}^r \Gamma(a_k + b_k) = \Pi_{k=1}^r \Gamma(-\Delta_k + \epsilon)$, and then take the limit $\epsilon \to 0$. The only term in (2.30) that survives is the one with r discrete sums leaving (writing $x_1,\dots,x_r$ in place of $j_1,\dots,j_r$)

$$P_n^{(3)} \cdot Q_m^{(3)} \equiv \sum_{x_1=0}^{\Delta_1} \sum_{x_2=0}^{\Delta_2} \cdots \sum_{x_r=0}^{\Delta_r} \int_{-\infty}^{\infty} dx_{r+1} \cdots \int_{-\infty}^{\infty} dx_p \, w^{(3)}(x_1,\dots,x_p) P_n^{(3)}(x_1,\dots,x_p) Q_m^{(3)}(x_1,\dots,x_p),$$
$$P_n^{(3)} \cdot Q_m^{(3)} = \overline{P}_n^{(3)} \cdot \overline{Q}_m^{(3)} = \lambda_n^{(3)} \prod_{k=1}^{p} \delta_{n_k m_k}, \quad P_n^{(3)} \cdot \overline{P}_m^{(3)} = Q_n^{(3)} \cdot \overline{Q}_m^{(3)} = 0, \quad \text{if } N \neq M, \quad (2.31)$$

where the mixed weight function and normalization constant are given by ($\Delta_1^r \equiv \Sigma_{k=1}^r \Delta_k$)

$$w^{(3)}(x_1,\dots,x_p) = \left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\Gamma(A+A_1^r+X_1^r-iX_{r+1}^p)\Gamma(B+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)$$
$$\times\Gamma(c+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)\Gamma(c+A_1^r+X_1^p-iX_{r+1}^p)$$
$$\times\frac{\Gamma(d+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)\Gamma(d+A_1^r+X_1^r-iX_{r+1}^p)}{\Gamma(2B_1^r+2\Delta_1^r-2X_1^r+2iX_{r+1}^p)\Gamma(2A_1^r+2X_1^r-2iX_{r+1}^p)}, \tag{2.32}$$

$$\lambda_n^{(3)} = 2(2\pi)^{p-r}\left[\prod_{k=1}^{r}\frac{\Delta_k!}{(\Delta_k-n_k)!}n_k!(-1)^{n_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right]$$
$$\times\frac{\Gamma(N+A+c)\Gamma(N+A+d)\Gamma(N+B+c)\Gamma(N+B+d)\Gamma(N+c+d)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)}, \tag{2.33}$$

and the polynomials are defined as

$$\begin{aligned}
&P_n^{(3)}(x_1,\dots,x_p)\equiv P_n(ia_1+ix_1,\dots,ia_r+ix_r,x_{r+1},\dots,x_p),\\
&\overline{P}_n^{(3)}(x_1,\dots,x_p)\equiv \overline{P}_n(ia_1+ix_1,\dots,ia_r+ix_r,x_{r+1},\dots,x_p),\quad 0\leqslant n_k\leqslant\Delta_k,\quad k=1,2,\dots,r,\\
&Q_n^{(3)}(x_1,\dots,x_p)\equiv Q_n(ia_1+ix_1,\dots,ia_r+ix_r,x_{r+1},\dots,x_p),\quad n_{r+1},\dots,n_p=0,1,2,\dots,\infty,\\
&\overline{Q}_n^{(3)}(x_1,\dots,x_p)\equiv \overline{Q}_n(ia_1+ix_1,\dots,ia_r+ix_r,x_{r+1},\dots,x_p),
\end{aligned} \tag{2.34}$$

where the indicated range of the indices applies to all four families.

These results can also be verified independently of the limit. To calculate the norm of the weight function begin with a change of variables from $x_{r+1},\dots,x_p$ to $X_{r+1}^p,x_{r+2},\dots,x_p$ and use (2.4) and induction to evaluate the $x_{r+2},\dots,x_p$ integrations. This gives

$$\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_r=0}^{\Delta_r}\int_{-\infty}^{\infty}dx_{r+1}\cdots\int_{-\infty}^{\infty}dx_p\, w^{(3)}(x_1,\dots,x_p)$$
$$=\frac{(2\pi)^{p-r-1}}{\Gamma(A_{r+1}^p+B_{r+1}^p)}\left[\prod_{k=r+1}^{p}\Gamma(a_k+b_k)\right]\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_r=0}^{\Delta_r}\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]$$
$$\times\int_{-\infty}^{\infty}dX_{r+1}^p\,\Gamma(A_{r+1}^p+iX_{r+1}^p)\Gamma(B_{r+1}^p-iX_{r+1}^p)$$
$$\times\Gamma(A+A_1^r+X_1^r-iX_{r+1}^p)\Gamma(B+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)\Gamma(c+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)$$
$$\times\Gamma(c+A_1^r+X_1^r-iX_{r+1}^p)\frac{\Gamma(d+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)}{\Gamma(2B_1^r+2\Delta_1^r-2X_1^r+2iX_{r+1}^p)}\frac{\Gamma(d+A_1^r+X_1^r-iX_{r+1}^p)}{\Gamma(2A_1^r+2X_1^r-2iX_{r+1}^p)}, \tag{2.35}$$

which by a further change of variable from X_{r+1}^p to $Z\equiv X_{r+1}^p+i(A_1^r+X_1^r)$ becomes

$$\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_r=0}^{\Delta_r}\int_{-\infty}^{\infty}dx_{r+1}\cdots\int_{-\infty}^{\infty}dx_p\, w^{(3)}(x_1,\dots,x_p)$$
$$=\frac{(2\pi)^{p-r-1}}{\Gamma(A_{r+1}^p+B_{r+1}^p)}\left[\prod_{k=r+1}^{p}\Gamma(a_k+b_k)\right]\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_r=0}^{\Delta_r}\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]$$
$$\times\int_{C(X_1^r)}dZ\,\Gamma(A+X_1^r+iZ)\Gamma(B+\Delta_1^r-X_1^r-iZ)$$
$$\times\frac{\Gamma(A-iZ)\Gamma(B+iZ)\Gamma(c+iZ)\Gamma(c-iZ)\Gamma(d+iZ)\Gamma(d-iZ)}{\Gamma(2iZ)\Gamma(-2iZ)}, \tag{2.36}$$

where the contours $C(X_1^r)$, $X_1^r=0,1,2,\dots,\Delta_1^r$ run parallel to the real axis with imaginary part $i(A_1^r+X_1^r)$. Taking into account (2.24) one finds that no poles of the integrand lie in the region bounded by and including $C(X_1^r)$ and the real axis. In this case each of the contours $C(X_1^r)$ can be deformed to the real axis and the integral can be brought outside of the multiple sums. The latter are then evaluated by theorem (2.23) resulting in

$$\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_r=0}^{\Delta_r}\int_{-\infty}^{\infty}dx_{r+1}\cdots\int_{-\infty}^{\infty}dx_p\, w^{(3)}(x_1,\dots,x_p)$$
$$=\frac{(2\pi)^{p-r-1}}{\Gamma(A+B)}\left[\prod_{k=r+1}^{p}\Gamma(a_k+b_k)\right]$$
$$\times\int_{-\infty}^{\infty}dZ\,\frac{\Gamma(A+iZ)\Gamma(A-iZ)\Gamma(B+iZ)\Gamma(B-iZ)\Gamma(c+iZ)\Gamma(c-iZ)\Gamma(d+iZ)\Gamma(d-iZ)}{\Gamma(2iZ)\Gamma(-2iZ)}, \tag{2.37}$$

which is proportional to the single variable integral given by (1.2)–(1.4) with $n = m = 0$. Using this result in (2.37) then confirms the norm as given by (2.33) with $N = 0$. The biorthogonality relations (2.31) are then verified in the same manner as for the purely continuous family.[4] Also, the restriction $\mathrm{Re}(a_1,a_2,...,a_r) \neq 0, -1, -2,...$ is removable from (2.31)–(2.34).

III. MULTIVARIABLE BIORTHOGONAL RACAH POLYNOMIALS

The mixed type family (2.31), which is a limit case of (2.30), requires that at least one of the a parameters have a positive real part. We consider a further generalization where all of the a parameters have negative real parts. Choosing $r = p - 1$ in (2.31) we then define the parameter domain as

$$\begin{aligned}&\mathrm{Re}(a_1,a_2,...,a_p) < 0, \quad \mathrm{Re}(b_1,b_2,...,b_p,c,d), \quad \mathrm{Re}(A_1^{p-1} + c), \quad \mathrm{Re}(A_1^{p-1} + d) > 0,\\ &2A, \quad A + B, \quad A + c, \quad A + d, \quad \mathrm{Re}(A + A_1^{p-1}), \quad \mathrm{Re}(a_p), \quad a_p + b_p \neq 0, -1, -2,...,\\ &a_k + b_k = -\Delta_k, \quad k = 1,2,...,p-1,\end{aligned} \tag{3.1}$$

for which the x_p contour C_p, which in (2.31) was on the real axis, is here deformed to pass underneath the increasing sequence $x_p = ia_p + ij_p, j_p = 0,1,2,...$ and above the decreasing sequence $x_p = -iA - iA_1^{p-1} - ij, j = 0,1,2,...$ while still passing above and below the remaining decreasing and increasing sequences, respectively.

To evaluate the norm of the weight function one begins with a change of variable from x_p to $Z \equiv x_p + i(A_1^{p-1} + X_1^{p-1})$ giving

$$\begin{aligned}&\sum_{x_1=0}^{\Delta_1} \cdots \sum_{x_{p-1}=0}^{\Delta_{p-1}} \int_{C_p} dx_p\, w^{(3)}(x_1,...,x_p)\\ &\quad = \sum_{x_1=0}^{\Delta_1} \cdots \sum_{x_{p-1}=0}^{\Delta_{p-1}} \left[\prod_{k=1}^{p-1} \binom{\Delta_k}{x_k}\right] \int_{C(X_1^{p-1})} dZ\, \Gamma(A + X_1^{p-1} + iZ)\Gamma(B + \Delta_1^{p-1} - X_1^{p-1} - iZ)\\ &\quad\quad \times \frac{\Gamma(A - iZ)\Gamma(B + iZ)\Gamma(c + iZ)\Gamma(c - iZ)\Gamma(d + iZ)\Gamma(d - iZ)}{\Gamma(2iZ)\Gamma(-2iZ)},\end{aligned} \tag{3.2}$$

and then the contours $C(X_1^{p-1}), X_1^{p-1} = 0,1,2,...,\Delta_1^{p-1}$ are deformed back to the real axes but with indentations to pass underneath the increasing sequence $Z = iA + ij, j = 0,1,2,...$ and above the decreasing sequence $Z = -iA - ij, j = 0,1,2,...$ while still passing above and below the remaining decreasing and increasing sequences, respectively. The multiple sums can then be brought inside the integral and evaluated by theorem (2.23) resulting in

$$\begin{aligned}&\sum_{x_1=0}^{\Delta_1} \cdots \sum_{x_{p-1}=0}^{\Delta_{p-1}} \int_{C_p} dx_p\, w^{(3)}(x_1,...,x_p)\\ &\quad = \frac{\Gamma(a_p + b_p)}{\Gamma(A + B)} \int_C dZ\, \Gamma(A + iZ)\Gamma(A - iZ)\\ &\quad\quad \times \frac{\Gamma(B + iZ)\Gamma(B - iZ)\Gamma(c + iZ)\Gamma(c - iZ)\Gamma(d + iZ)\Gamma(d - iZ)}{\Gamma(2iZ)\Gamma(-2iZ)},\end{aligned} \tag{3.3}$$

which is proportional to the single variable integral given by (1.2)–(1.4) with $n = m = 0$. Substituting this result into (3.3) then yields the norm defined by (2.33) with $r = p - 1$ and $N = 0$. The biorthogonality relations (2.31) then once again follow in the same manner as was proved for the purely continuous family[4] but with the inner product defined above or as expressed in (3.5).

Deforming C_p to the real axis plus closed loops about a finite number of poles and then evaluating the latter by the method of residues allows us to write schematically

$$\int_{C_p} dx_p = \int_{-\infty}^{\infty} dx_p + (2\pi i) \sum_{j_p=0}^{\mathrm{Re}(a_p) + j_p < 0} \mathrm{res}(x_p = ia_p + ij_p) - (2\pi i) \sum_{j=0}^{\mathrm{Re}(A + A_1^{p-1}) + j < 0} \mathrm{res}(x_p = -iA - iA_1^{p-1} - ij), \tag{3.4}$$

and then the inner product of the polynomials can be expressed as

$$\begin{aligned}P_n^{(3)} \cdot Q_m^{(3)} \equiv \sum_{x_1=0}^{\Delta_1} \cdots \sum_{x_{p-1}=0}^{\Delta_{p-1}} &\left\{\int_{-\infty}^{\infty} dx_p + (2\pi i) \sum_{j_p=0}^{\mathrm{Re}(a_p) + j_p < 0} \mathrm{res}(x_p = ia_p + ij_p)\right.\\ &\left. - (2\pi i) \sum_{j=0}^{\mathrm{Re}(A + A_1^{p-1}) + j < 0} \mathrm{res}(x_p = -iA - iA_1^{p-1} - ij)\right\} w^{(3)}(x_1,...,x_p) P_n^{(3)}(x_1,...,x_p) Q_m^{(3)}(x_1,...,x_p),\end{aligned} \tag{3.5}$$

and similarly for $P_n^{(3)} \cdot \overline{P}_m^{(3)}$, $Q_n^{(3)} \cdot \overline{Q}_m^{(3)}$, and $\overline{P}_n^{(3)} \cdot \overline{Q}_m^{(3)}$.

Formula (3.5) has a limit to a purely discrete inner product. Take $a_p + b_p = -\Delta_p + \epsilon$, where Δ_p is a non-negative integer, divide by $\Gamma(a_p + b_p) = \Gamma(-\Delta_p + \epsilon)$ and then take the limit $\epsilon \to 0$. The two purely discrete terms survive giving (writing x_p in place of j_p)

$$\lim_{\epsilon\to 0}\frac{P_n^{(3)}\cdot Q_m^{(3)}}{\Gamma(-\Delta_p+\epsilon)}$$

$$= (2\pi)\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_p=0}^{\Delta_p}\left[\prod_{k=1}^{p}\binom{\Delta_k}{x_k}\right]$$

$$\times\Gamma(2A+X)\Gamma(B-A-X)\frac{\Gamma(c+A+X)\Gamma(c-A-X)\Gamma(d+A+X)\Gamma(d-A-X)}{\Gamma(2A+2X)\Gamma(-2A-2X)}$$

$$\times P_n^{(3)}(x_1,\ldots,x_{p-1},ia_p+ix_p)Q_m^{(3)}(x_1,\ldots,x_{p-1},ia_p+ix_p)$$

$$+(2\pi)\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_{p-1}=0}^{\Delta_{p+1}}\sum_{j=0}^{\Delta+X_1^{p-1}}\left[\prod_{k=1}^{p-1}\binom{\Delta_k}{x_k}\right]\frac{\Delta_p!}{(\Delta+X_1^{p-1}-j)!(j-X_1^{p-1})!}(-1)^{\Delta+\Delta_p}$$

$$\times\Gamma(2A+j)\Gamma(B-A+\Delta_1^{p-1}-j)\Gamma(c+A-X_1^{p-1}+j)\Gamma(c-A+X_1^{p-1}-j)$$

$$\times\frac{\Gamma(d+A-X_1^{p-1}+j)\Gamma(d-A+X_1^{p-1}-j)}{\Gamma(2A-2X_1^{p-1}+2j)\Gamma(-2A+2X_1^{p-1}-2j)}$$

$$\times P_n^{(3)}(x_1,\ldots,x_{p-1},-iA-iA_1^{p-1}-ij)Q_m^{(3)}(x_1,\ldots,x_{p-1},-iA-iA_1^{p-1}-ij), \tag{3.6}$$

and in turn one can show that these two terms are equal. To demonstrate this make a change of summation index in the second term, $j\to j+X_1^{p-1}$, substitute representations (1.15) and (1.17) for $P_n(x)$ and $Q_m(x)$, and then use theorem (2.23) to evaluate the $x_1,x_2,\ldots,x_{p-1}$ sums. This leaves only the j sum but if (2.23) is used again with a different choice of parameters this can be re-expressed as a multiple sum that is equal to the first term in (3.6). The inner product can then be taken as twice the first term

$$\lim_{\epsilon\to 0}\frac{P_n^{(3)}\cdot Q_m^{(3)}}{\Gamma(-\Delta_p+\epsilon)}$$

$$= (4\pi)\sum_{x_1=0}^{\Delta_1}\cdots\sum_{x_p=0}^{\Delta_p}\left[\prod_{k=1}^{p}\binom{\Delta_k}{x_k}\right]\Gamma(2A+X)\Gamma(B-A-X)$$

$$\times\frac{\Gamma(c+A+X)\Gamma(c-A-X)\Gamma(d+A+X)\Gamma(d-A-X)}{\Gamma(2A+2X)\Gamma(-2A-2X)}$$

$$\times P_n^{(3)}(x_1,\ldots,x_{p-1},ia_p+ix_p)Q_m^{(3)}(x_1,\ldots,x_{p-1},ia_p+ix_p), \tag{3.7}$$

and similarly for $\overline{P}_n^{(3)}\cdot\overline{Q}_m^{(3)}$, $P_n^{(3)}\cdot\overline{P}_m^{(3)}$, and $Q_n^{(3)}\cdot\overline{Q}_m^{(3)}$. The biorthogonality relations (2.31) then yield in this limit, with a change in notation,

$$\begin{gathered}R_n\cdot W_m\equiv\sum_{\{x_k\}}\rho(x_1,\ldots,x_p)R_n(x_1,\ldots,x_p)W_m(x_1,\ldots,x_p),\\ R_n\cdot W_m=\overline{R}_n\cdot\overline{W}_m=\lambda_n\prod_{k=1}^{p}\delta_{n_km_k},\quad R_n\cdot\overline{R}_m=W_n\cdot\overline{W}_m=0,\quad\text{if } N\neq M,\end{gathered}\tag{3.8}$$

where $R_n(x)$, $\overline{R}_n(x)$, $W_n(x)$, and $\overline{W}_n(x)$ are the multivariable biorthogonal Racah polynomials and $\{x_k\}$ denotes the p discrete variables $x_1,x_2,\ldots,x_p$. After a customary redefinition of the parameters to $\alpha_1,\alpha_2,\ldots,\alpha_p$, β, δ, γ ($\alpha\equiv\Sigma_{k=1}^{p}\alpha_k$) this family is given by

$$\rho(x_1,\ldots,x_p)=\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(\alpha_k+1)x_k!}\right]\frac{(\gamma+\delta+1)_X(\gamma/2+\delta/2+3/2)_X}{(\gamma/2+\delta/2+1/2)_X(\gamma+\delta-\alpha-p+2)_X}\left(\frac{(\beta+\delta+1)_X(\gamma+1)_X}{(\gamma-\beta+1)_X(\delta+1)_X}\right), \tag{3.9}$$

$$\begin{aligned}\lambda_n=&\left[\prod_{k=1}^{p}n_k!(\alpha_k+1)_{n_k}\right](\beta+1)_N(\gamma+1)_N(\alpha-\delta+p)_N(\alpha+\beta-\gamma+p)_N(\beta+\delta+1)_N\\ &\times\frac{(N+\alpha+\beta+p)_N}{(\alpha+\beta+p+1)_{2N}}\frac{\Gamma(\gamma+\delta-\alpha-p+2)\Gamma(-\beta-\alpha-p)\Gamma(\gamma-\beta+1)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+2)\Gamma(-\beta)\Gamma(\gamma-\beta-\alpha-p+1)\Gamma(\delta-\alpha-p+1)},\end{aligned}\tag{3.10}$$

$$\begin{aligned}R_n(x_1,\ldots,x_p)=&\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\beta+\delta+1)_N(\gamma+1)_N\\ &\times F_{2:1;\ldots;1}^{2:2;\ldots;2}\left(\begin{matrix}N+\alpha+\beta+p,X+\gamma+\delta+1:-n_1,-x_1;\ldots;-n_p,-x_p\\ \beta+\delta+1,\gamma+1:\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right),\end{aligned}\tag{3.11}$$

$$\overline{R}_n(x_1,\ldots,x_p)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\alpha+\beta-\gamma+p)_N(\alpha-\delta+p)_N$$

$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1}\left(\begin{matrix}N+\alpha+\beta+p,-X+\alpha-\gamma-\delta+p-1:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ \alpha+\beta-\gamma+p,\alpha-\delta+p:\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right), \tag{3.12}$$

$$W_n(x_1,\ldots,x_p)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](X+\beta+\delta+1)_N(X+\gamma+1)_N$$
$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1}\left(\begin{matrix}-N-\beta,-X+\alpha-\gamma-\delta+p-1:-n_1,-x_1;\ldots;-n_p,-x_p\\ -N-X-\beta-\delta,-N-X-\gamma:\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right), \tag{3.13}$$

$$\overline{W}_n(x_1,\ldots,x_p)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](-X+\beta-\gamma)_N(-X-\delta)_N$$
$$\times F^{2:2;\ldots;2}_{2:1;\ldots;1}\left(\begin{matrix}-N-\beta,X+\gamma+\delta+1:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ -N+X+\gamma-\beta+1,-N+X+\delta+1:\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right), \tag{3.14}$$

which according to the present derivation satisfy relations (3.8) when $\alpha_k+1=-\Delta_k$, $k=1,2,\ldots,p$, and the $\{x_k\}$ sum is over the region $0\leqslant x_k\leqslant\Delta_k$, $k=1,2,\ldots,p$. As an independent evaluation of the norm of the weight function one uses theorem (2.23) to perform the $x_1,\ldots,x_{p-1}$ summations at constant X and then (2.15) to evaluate the X sum resulting in (3.10) with $N=0$. The biorthogonality relations (3.8) can then be independently verified in the same manner as for the purely continuous family[4] apart from a redefinition of the parameters. In this way one finds that (3.8) are also valid for $\beta+\delta+1$ or $\gamma+1=-\Delta_{p+1}$, where Δ_{p+1} is another non-negative integer, and in this case the $\{x_k\}$ sum is over the region $0\leqslant X\leqslant\Delta_{p+1}$. Another possibility is to have a combination of $\beta+\delta+1$ or $\gamma+1=-\Delta_{p+1}$ and only a subset of the α parameters satisfying $\alpha_k+1=-\Delta_k$. These different possibilities and the corresponding regions of the $\{x_k\}$ sums are summarized below:

$$\begin{aligned}&\alpha_k+1=-\Delta_k,\quad k=1,2,\ldots,p,\quad 0\leqslant x_k\leqslant\Delta_k,\\ &\beta+\delta+1\ \text{or}\ \gamma+1=-\Delta_{p+1},\quad 0\leqslant X\leqslant\Delta_{p+1},\\ &\beta+\delta+1\ \text{or}\ \gamma+1=-\Delta_{p+1},\end{aligned} \tag{3.15}$$

and

$$\alpha_k+1=-\Delta_k,\quad k\in S\subseteq(1,2,\ldots,p),\quad (0\leqslant x_k\leqslant\Delta_k)\cap(0\leqslant X\leqslant\Delta_{p+1}).$$

In the special case of a single variable all four families of polynomials (3.11)–(3.14) reduce, through a transformation formula satisfied by the ${}_4F_3(1)$ hypergeometric function, to (1.9) and the biorthogonality relations (3.8) reduce to the single orthogonality relation (1.10).

Alternatively one could have chosen the inner product to be twice the second term in (3.6), which leads to a different but within a change of variables equivalent family.

Another inequivalent multivariable generalizaiton of the Racah polynomials has been studied by Gustafson.[7] These are closely related to the so-called $U(n)$ multivariable hypergeometric series introduced by Holman *et al.*[8] and Holman,[9] and which have been q extended by Milne.[10–15] Gustafson's polynomials are associated with a different weight function than (3.9) and are orthogonal as opposed to biorthogonal. The difference in these two families is a reflection of the distinct hypergeometric series to which they are related, the Kampé de Fériet series (1.19) in the present case and the $U(n)$ series in Gustafson's case.

IV. MULTIVARIABLE BIORTHOGONAL HAHN AND DUAL HAHN POLYNOMIALS

In analogy with the single variable case there is an interesting limit to a family of multivariable Hahn polynomials. These are obtained upon dividing the Racah polynomials by δ^N and taking the limit $\delta\to\infty$. The first two families $R_n(x)$ and $\overline{R}_n(x)$ limit to the same Hahn polynomials $H_n(x)$ while $W_n(x)$ and $\overline{W}_n(x)$ limit to the same biorthogonal counterparts $\overline{H}_n(x)$,

$$\begin{aligned}&\lim_{\delta\to\infty}\delta^{-N}R_n(x)=\lim_{\delta\to\infty}\delta^{-N}\overline{R}_n(x)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\gamma+1)_NH_n(x),\\ &\lim_{\delta\to\infty}\delta^{-N}W_n(x)=\lim_{\delta\to\infty}\delta^{-N}\overline{W}_n(x)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\gamma+1)_N\overline{H}_n(x),\\ &\lim_{\delta\to\infty}\rho(x)=\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(\alpha_k+1)x_k!}\right]\frac{(\gamma+1)_X}{(\gamma-\beta+1)_X},\end{aligned} \tag{4.1}$$

where $H_n(x)$ and $\overline{H}_n(x)$ are given by

$$H_n(x)=F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}N+\alpha+\beta+p:-n_1,-x_1;\ldots;-n_p,-x_p\\ \gamma+1:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix}$$
$$=(-1)^N\frac{(\alpha+\beta-\gamma+p)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}N+\alpha+\beta+p:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ \alpha+\beta-\gamma+p:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix},$$
$$\overline{H}_n(x)=\frac{(X+\gamma+1)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}-N-\beta:-n_1,-x_1;\ldots;-n_p,-x_p\\ -N-X-\gamma:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix} \tag{4.2}$$
$$=(-1)^N\frac{(-X+\beta-\gamma)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}-N-\beta:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ -N+X+\gamma-\beta+1:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix}.$$

These satisfy the biorthogonality relations

$$H_n\cdot\overline{H}_m\equiv\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(\alpha_k+1)x_k!}\right]\frac{(\gamma+1)_X}{(\gamma-\beta+1)_X}H_n(x)\overline{H}_m(x),$$
$$H_n\cdot\overline{H}_m=\lambda_n\prod_{k=1}^{p}\delta_{n_km_k},\quad H_n\cdot H_m=\overline{H}_n\cdot\overline{H}_m=0,\quad \text{if } N\neq M, \tag{4.3}$$
$$\lambda_n=(-1)^N\left[\prod_{k=1}^{p}\frac{n_k!}{(\alpha_k+1)_{n_k}}\right](\alpha+\beta-\gamma+p)_N\frac{(\beta+1)_N}{(\gamma+1)_N}\frac{(N+\alpha+\beta+p)_N}{(\alpha+\beta+p+1)_{2N}}\frac{\Gamma(-\beta-\alpha-p)\Gamma(\gamma-\beta+1)}{\Gamma(-\beta)\Gamma(\gamma-\beta-\alpha-p+1)},$$

where the $\{x_k\}$ sum is over one of the regions

$$\begin{aligned}&\alpha_k+1=-\Delta_k,\quad k=1,2,\ldots,p,\quad 0\leqslant x_k\leqslant\Delta_k,\\ &\gamma+1=-\Delta_{p+1},\quad 0\leqslant X\leqslant\Delta_{p+1},\\ &\gamma+1=-\Delta_{p+1}\ \text{and}\ \alpha_k+1=-\Delta_k,\quad k\in S\subseteq(1,2,\ldots,p),\ (0\leqslant x_k\leqslant\Delta_k)\cap(0\leqslant X\leqslant\Delta_{p+1}).\end{aligned} \tag{4.4}$$

These polynomials have already been discussed in detail[16] for the specific case $\gamma+1=-\Delta_{p+1}$ and the equivalence of each pair of representations in (4.2) has been demonstrated. Among other interesting properties these polynomials possess discrete Rodrigues formulas.

Another important limit not yet studied are the multivariable biorthogonal dual Hahn polynomials. These result upon dividing the Racah polynomials by β^N and taking the limit $\beta\to\infty$. In this case $R_n(x)$ and $\overline{W}_n(x)$ limit to the same dual Hahn family $D_n(x)$ while $\overline{R}_n(x)$ and $W_n(x)$ limit to the same biorthogonal counterparts $\overline{D}_n(x)$,

$$\lim_{\beta\to\infty}\beta^{-N}R_n(x)=\lim_{\beta\to\infty}\beta^{-N}\overline{W}_n(x)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\gamma+1)_ND_n(x),$$
$$\lim_{\beta\to\infty}\beta^{-N}\overline{R}_n(x)=\lim_{\beta\to\infty}\beta^{-N}W_n(x)=\left[\prod_{k=1}^{p}(\alpha_k+1)_{n_k}\right](\gamma+1)_N\overline{D}_n(x), \tag{4.5}$$
$$\lim_{\beta\to\infty}\rho(x)=\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(\alpha_k+1)x_k!}\right]\frac{(\gamma+\delta+1)_X(\gamma/2+\delta/2+3/2)_X(\gamma+1)_X}{(\gamma/2+\delta/2+1/2)_X(\gamma+\delta-\alpha-p+2)_X(\delta+1)_X}(-1)^X,$$

where $D_n(x)$ and $\overline{D}_n(x)$ are given by

$$D_n(x)=F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}X+\gamma+\delta+1:-n_1,-x_1;\ldots;-n_p,-x_p\\ \gamma+1:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix}$$
$$=\frac{(-X-\delta)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}X+\gamma+\delta+1:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ -N+X+\delta+1:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix},$$
$$\overline{D}_n(x)=\frac{(\alpha-\delta+p)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}-X+\alpha-\gamma-\delta+p-1:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ \alpha-\delta+p:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix} \tag{4.6}$$
$$=\frac{(X+\gamma+1)_N}{(\gamma+1)_N}F^{1:2;\ldots;2}_{1:1;\ldots;1}\begin{pmatrix}-X+\alpha-\gamma-\delta+p-1:-n_1,-x_1;\ldots;-n_p,-x_p\\ -N-X-\gamma:\alpha_1+1;\ldots;\alpha_p+1\end{pmatrix}.$$

The four biorthogonality relations satisfied by the Racah polynomials, in this limit, imply the single relation

$$\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(\alpha_k+1)x_k!}\right]\frac{(\gamma+\delta+1)_X(\gamma/2+\delta/2+3/2)_X(\gamma+1)_X(-1)^X}{(\gamma/2+\delta/2+1/2)_X(\gamma+\delta-\alpha-p+2)_X(\delta+1)_X}D_n(x)\overline{D}_m(x)$$
$$=\left[\prod_{k=1}^{p}\frac{n_k!}{(\alpha_k+1)_{n_k}}\right]\frac{(\alpha-\delta+p)_N}{(\gamma+1)_N}\frac{\Gamma(\gamma+\delta-\alpha-p+2)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+2)\Gamma(\delta-\alpha-p+1)}\prod_{k=1}^{p}\delta_{n_km_k}, \tag{4.7}$$

where the $\{x_k\}$ sum is again over one of the regions given in (4.4).

To demonstrate the equivalence of the two representations of $D_n(x)$ we first use (2.23) to deduce the identity

$$\frac{\Gamma(X+\delta+1)}{\Gamma(X+\delta+1-N+J)}=\sum_{\{\ell_k\}}\left[\prod_{k=1}^{p}\binom{n_k-j_k}{\ell_k}\right]\frac{\Gamma(N-L+X+\delta+\gamma+1)}{\Gamma(J+X+\delta+\gamma+1)}\frac{\Gamma(L-N-\gamma)}{\Gamma(-N-\gamma)}, \tag{4.8}$$

where j_k denotes the summation indices in the Kampé de Fériet hypergeometric series (1.19). This is substituted into the second expression for $D_n(x)$, the $\{j_k\}$ and $\{\ell_k\}$ sums are interchanged, and (2.23) is used again to evaluate the former. Reversing the remaining $\{\ell_k\}$ sums by transforming $\ell_k\to n_k-\ell_k$ then yields the first expression for $D_n(x)$. For the $\overline{D}_n(x)$ representations one begins instead with the following identity also deduced from (2.23),

$$\frac{\Gamma(N+\alpha-\delta+p)}{\Gamma(J+\alpha-\delta+p)}=\sum_{\{\ell_k\}}\left[\prod_{k=1}^{p}\binom{n_k-j_k}{\ell_k}\right]\frac{\Gamma(N-L+\alpha+p-X-\delta-\gamma-1)}{\Gamma(J+\alpha+p-X-\delta-\gamma-1)}\left(\frac{\Gamma(L+X+\gamma+1)}{\Gamma(X+\gamma+1)}\right). \tag{4.9}$$

Substituting this into the first expression for $\overline{D}_n(x)$ and proceeding as before then yields the second representation.

There are also mixed type counterparts to these Hahn and dual Hahn polynomials which are obtained as limit cases of the mixed type Wilson polynomials. In (2.19)–(2.22) put

$$\begin{aligned}&a_k=a_k'+\tfrac12 iw_k,\quad b_k=b_k'-\tfrac12 iw_k,\quad k=1,2,\ldots,p,\\&x_k=x_k',\quad k=1,2,\ldots,r,\quad x_k=x_k'-\tfrac12 w_k,\quad k=r+1,\ldots,p,\\&c=c'+\tfrac12 iW,\quad d=d'-\tfrac12 iW,\quad W\equiv\sum_{k=1}^{p}w_k,\end{aligned} \tag{4.10}$$

divide the polynomials by $(iW)^N$ and then take the limit $W\to\infty$ (and drop the primes). The mixed Wilson families $P_n^{(2)}(x)$ and $\overline{P}_n^{(2)}(x)$ both limit to the same mixed Hahn polynomials while $Q_n^{(2)}(x)$ and $\overline{Q}_n^{(2)}(x)$ limit to the same biorthogonal counterparts

$$\begin{aligned}&\lim_{W\to\infty}(iW)^{-N}P_n^{(2)}(x)=\lim_{W\to\infty}(iW)^{-N}\overline{P}_n^{(2)}(x)=H_n^{(2)}(x),\\&\lim_{W\to\infty}(iW)^{-N}Q_n^{(2)}(x)=\lim_{W\to\infty}(iW)^{-N}\overline{Q}_n^{(2)}(x)=\overline{H}_n^{(2)}(x),\\&\lim_{W\to\infty}w^{(2)}(x)=\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\\&\qquad\times\Gamma(A_1^r+d+X_1^r-iX_{r+1}^p)\Gamma(B_1^r+c+\Delta_1^r-X_1^r+iX_{r+1}^p),\end{aligned} \tag{4.11}$$

where $H_n^{(2)}(x)$ and $\overline{H}_n^{(2)}(x)$ are given by

$$\begin{aligned}H_n^{(2)}(x)&=\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](A+d)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}N+A+B+c+d-1:-n_1,A_1^r+d\\A+d:a_1+b_1;\end{matrix}\right.\\&\qquad\left.\begin{matrix}+X_1^r-iX_{r+1}^p;-n_2,-x_2;\ldots;-n_r,-x_r;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p\\-\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right)\\&=(-1)^N\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](B+c)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}N+A+B+c+d-1:-n_1,B_1^r+c+\Delta_1^r\\B+c:a_1+b_1;\end{matrix}\right.\\&\qquad\left.\begin{matrix}-X_1^r+iX_{r+1}^p;-n_2,-\Delta_2+x_2;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p\\-\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right),\\\overline{H}_n^{(2)}(x)&=(-1)^N\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](-\Delta_1+x_1)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}\Delta_1-N+1:-n_1,B_1^r+c+\Delta_1^r\\\Delta_1-N-x_1+1:a_1+b_1;\end{matrix}\right.\\&\qquad\left.\begin{matrix}-X_1^r+iX_{r+1}^p;-n_2,-\Delta_2+x_2;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p\\-\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right)\\&=\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](-x_1)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}\Delta_1-N+1:-n_1,A_1^r+d+X_1^r-iX_{r+1}^p;\\-N+x_1+1:a_1+b_1;\end{matrix}\right.\end{aligned} \tag{4.12}$$

$$-n_2,-x_2;\ldots;-n_r,-x_r;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p \atop -\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\Bigg),$$

with $c+d=-\Delta_1$ and $a_k+b_k=-\Delta_k$, $k=2,3,\ldots,r$. These satisfy

$$H_n^{(2)}\cdot\overline{H}_m^{(2)}\equiv\int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_{r+1}\sum_{x_r=0}^{\Delta_r}\cdots\sum_{x_1=0}^{\Delta_1}\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]$$
$$\times\Gamma(A_1^r+d+X_1^r-iX_{r+1}^p)\Gamma(B_1^r+c+\Delta_1^r-X_1^r+iX_{r+1}^p)H_n^{(2)}(x)\overline{H}_m^{(2)}(x),$$

$$H_n^{(2)}\cdot\overline{H}_m^{(2)}=\lambda_n^{(2)}\prod_{k=1}^{p}\delta_{n_km_k},\quad H_n^{(2)}\cdot H_m^{(2)}=\overline{H}_n^{(2)}\cdot\overline{H}_m^{(2)}=0,\quad\text{if } N\neq M, \tag{4.13}$$
$$\lambda_n^{(2)}=(2\pi)^{p-r}\Gamma(n_1+a_1+b_1)n_1!\left[\prod_{k=2}^{r}(-\Delta_k)_{n_k}n_k!\right]\left[\prod_{k=r+1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right]$$
$$\times\frac{\Delta_1!}{(\Delta_1-N)!}\frac{\Gamma(N+A+d)\Gamma(N+B+c)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)},$$

and the equivalence of each pair of representations in (4.12) follows from (4.2) upon a redefinition of the parameters and a change of variables.

Another mixed type Hahn family is obtained from (2.31)–(2.34) using the same limiting procedure. That is, transform the parameters and variables as in (4.10) and then take the limit $W\to\infty$. This yields,

$$\lim_{W\to\infty}(iW)^{-N}P_n^{(3)}(x)=\lim_{W\to\infty}(iW)^{-N}\overline{P}_n^{(3)}(x)=H_n^{(3)}(x),$$
$$\lim_{W\to\infty}(iW)^{-N}Q_n^{(3)}(x)=\lim_{W\to\infty}(iW)^{-N}\overline{Q}_n^{(3)}(x)=\overline{H}_n^{(3)}(x), \tag{4.14}$$
$$\lim_{W\to\infty}\frac{w^{(3)}(x)}{\Gamma(A'+c'+iW)\Gamma(B'+d'-iW)}=\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]$$
$$\times\Gamma(A_1^r+d+X_1^r-iX_{r+1}^p)\Gamma(B_1^r+c+\Delta_1^r-X_1^r+iX_{r+1}^p),$$

where the polynomials are given by

$$H_n^{(3)}(x)=\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](A+d)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\Bigg(\begin{matrix}N+A+B+c+d-1:-n_1,-x_1;\\ A+d:-\Delta_1;\end{matrix}$$
$$\begin{matrix}-n_2,-x_2;\ldots;-n_r,-x_r;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p\\ -\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\Bigg)$$
$$=(-1)^N\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](B+c)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\Bigg(\begin{matrix}N+A+B+c+d-1:-n_1,-\Delta_1+x_1;\\ B+c:-\Delta_1;\end{matrix} \tag{4.15}$$
$$\begin{matrix}-n_2,-\Delta_2+x_2;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p\\ -\Delta_2;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\Bigg),$$
$$\overline{H}_n^{(3)}(x)=\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](d+A_1^r+X_1^r-iX_{r+1}^p)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\Bigg(\begin{matrix}-N-c-d+1:\\ -N-d-A_1^r-X_1^r\end{matrix}$$
$$\begin{matrix}-n_1,-x_1;\ldots;-n_r,-x_r;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p\\ +iX_{r+1}^p+1:-\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\Bigg)$$
$$=(-1)^N\left[\prod_{k=1}^{p}(a_k+b_k)_{n_k}\right](c-A_1^r-X_1^r+iX_{r+1}^p)_NF^{1:2;\ldots;2}_{1:1;\ldots;1}\Bigg(\begin{matrix}-N-c-d+1:\\ -N-c+A_1^r+X_1^r\end{matrix}$$
$$\begin{matrix}-n_1,-\Delta_1+x_1;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p\\ -iX_{r+1}^p+1:-\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\Bigg),$$

with $a_k+b_k=-\Delta_k$, $k=1,2,\ldots,r$, and these satisfy

$$H_n^{(3)}\cdot\overline{H}_m^{(3)}\equiv\int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_{r+1}\sum_{x_r=0}^{\Delta_r}\cdots\sum_{x_1=0}^{\Delta_1}\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]$$
$$\times\Gamma(A_1^r+d+X_1^r-iX_{r+1}^p)\Gamma(B_1^r+c+\Delta_1^r-X_1^r+iX_{r+1}^p)H_n^{(3)}(x)\overline{H}_m^{(3)}(x),$$

$$H_n^{(3)}\cdot\overline{H}_m^{(3)}=\lambda_n^{(3)}\prod_{k=1}^{p}\delta_{n_k m_k},\quad H_n^{(3)}\cdot H_m^{(3)}=\overline{H}_n^{(3)}\cdot\overline{H}_m^{(3)}=0,\quad \text{if } N\neq M, \tag{4.16}$$

$$\lambda_n^{(3)}=(2\pi)^{p-r}\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}n_k!\right]\left[\prod_{k=r+1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right](-1)^N$$
$$\times\frac{\Gamma(N+A+d)\Gamma(N+B+c)\Gamma(N+c+d)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)}.$$

Notice that the weight function in (4.16) is equivalent to that of the first mixed type Hahn family (4.13); the two families of polynomials are however distinct. The equivalence of each pair of representations in (4.15) again follows from (4.2) upon a redefinition of the parameters and a change of variables.

The analogous mixed type dual Hahn family is obtained from (2.31)–(2.34) in the limit $d\to\infty$. In this case $P_n^{(3)}(x)$ and $\overline{Q}_n^{(3)}(x)$ limit to the same dual Hahn polynomials while $\overline{P}_n^{(3)}(x)$ and $Q_n^{(3)}(x)$ limit to the same biorthogonal counterparts

$$\lim_{d\to\infty} d^{-N}P_n^{(3)}(x)=\lim_{d\to\infty} d^{-N}\overline{Q}_n^{(3)}(x)=D_n^{(3)}(x),$$
$$\lim_{d\to\infty} d^{-N}\overline{P}_N^{(3)}(x)=\lim_{d\to\infty} d^{-N}Q_n^{(3)}(x)=\overline{D}_n^{(3)}(x), \tag{4.17}$$
$$\lim_{d\to\infty}\frac{w^{(3)}(x)}{\Gamma^2(d)}=\rho^{(3)}(x),$$

where

$$\rho^{(3)}(x)=\left[\prod_{k=1}^{r}\binom{\Delta_k}{x_k}\right]\left[\prod_{k=r+1}^{p}\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]\Gamma(A+A_1^r+X_1^r-iX_{r+1}^p)$$
$$\times\Gamma(B+B_1^r+\Delta_1^r-X_1^r+iX_{r+1}^p)\frac{\Gamma(c+B_1^r+\Delta_1^r+X_1^r+iX_{r+1}^p)\Gamma(c+A_1^r-X_1^r-iX_{r+1}^p)}{\Gamma(2B_1^r+2\Delta_1^r-2X_1^r+2iX_{r+1}^p)\Gamma(2A_1^r+2X_1^r-2iX_{r+1}^p)},$$

$$\begin{aligned}
D_n^{(3)}(x)&=\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}\right]\left[\prod_{k=r+1}^{p}(a_k+b_k)_{n_k}\right](A+c)_N F^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}A+A_1^r+X_1^r-iX_{r+1}^p: \\ A+c:\end{matrix}\right.\\
&\quad\left.\begin{matrix}-n_1,-x_1;\ldots;-n_r,-x_r;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p \\ -\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right)\\
&=\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}\right]\left[\prod_{k=r+1}^{p}(a_k+b_k)_{n_k}\right](c-A_1^r-X_1^r+iX_{r+1}^p)_N F^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}A+A_1^r+X_1^r \\ -N-c+A_1^r\end{matrix}\right.\\
&\quad\left.\begin{matrix}-iX_{r+1}^p:-n_1,-\Delta_1+x_1;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p \\ +X_1^r-iX_{r+1}^p+1:-\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right),\\
\overline{D}_n^{(3)}(x)&=\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}\right]\left[\prod_{k=r+1}^{p}(a_k+b_k)_{n_k}\right](B+c)_N F^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}B-A_1^r-X_1^r+iX_{r+1}^p: \\ B+c:\end{matrix}\right.\\
&\quad\left.\begin{matrix}-n_1,-\Delta_1+x_1;\ldots;-n_r,-\Delta_r+x_r;-n_{r+1},b_{r+1}-ix_{r+1};\ldots;-n_p,b_p-ix_p \\ -\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right)\\
&=\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}\right]\left[\prod_{k=r+1}^{p}(a_k+b_k)_{n_k}\right](c+A_1^r+X_1^r-iX_{r+1}^p)_N F^{1:2;\ldots;2}_{1:1;\ldots;1}\left(\begin{matrix}B-A_1^r-X_1^r \\ -N-c-A_1^r\end{matrix}\right.\\
&\quad\left.\begin{matrix}+iX_{r+1}^p:-n_1,-x_1;\ldots;-n_r,-x_r;\ldots;-n_{r+1},a_{r+1}+ix_{r+1};\ldots;-n_p,a_p+ix_p \\ -X_1^r+iX_{r+1}^p+1:-\Delta_1;\ldots;-\Delta_r;a_{r+1}+b_{r+1};\ldots;a_p+b_p\end{matrix}\right),
\end{aligned} \tag{4.18}$$

with $a_k+b_k=-\Delta_k$, $k=1,2,\ldots,r$. These satisfy the biorthogonality relation

$$\int_{-\infty}^{\infty}dx_p\cdots\int_{-\infty}^{\infty}dx_{r+1}\sum_{x_r=0}^{\Delta_r}\cdots\sum_{x_1=0}^{\Delta_1}\rho^{(3)}(x)D_n^{(3)}(x)\overline{D}_m^{(3)}(x)$$
$$=2(2\pi)^{p-r}\left[\prod_{k=1}^{r}(-\Delta_k)_{n_k}n_k!\right]\left[\prod_{k=r+1}^{p}\Gamma(n_k+a_k+b_k)n_k!\right]\Gamma(N+A+c)\Gamma(N+B+c)\prod_{k=1}^{p}\delta_{n_k m_k}, \tag{4.19}$$

and the equivalence of each pair of representations in (4.18) follows from (4.6) upon a redefinition of the parameters and a change of variables.

The purely continuous multivariable biorthogonal Hahn[17] and dual Hahn[4] polynomials are also known.

ACKNOWLEDGMENTS

I would like to thank E. G. Kalnins, W. Miller, Jr., and D. Stanton for stimulating conversations. This research was supported in part by the Institute of Mathematics and its Applications (IMA) with funds provided by the National Science Foundation. I also thank the staff of the IMA for their warm hospitality during my long term visit when part of this work was done. This research was also supported by the Natural Sciences and Engineering Research Council of Canada and by the United States Department of Energy.

[1] J. A. Wilson, SIAM J. Math. Anal. **11**, 690 (1980).
[2] R. Askey and J. Wilson, SIAM J. Math. Anal. **13**, 651 (1982).
[3] W. N. Bailey, *Generalized Hypergeometric Series* (Cambridge U. P., London, 1935).
[4] M. V. Tratnik, J. Math. Phys. **30**, 2001 (1989).
[5] H. M. Srivastava and P. W. Karlsson, *Multiple Gaussian Hypergeometric Series* (Ellis Horwood, New York, 1985), p. 38.
[6] P. Appell and J. Kampé de Fériet, *Fonction Hypergéométriques et Hypersphériques* (Gauthier-Villars, Paris, 1926).
[7] R. A. Gustafson, SIAM J. Math. Anal. **18**, 495 (1987).
[8] W. J. Holman, III, L. C. Biedenharn, and J. D. Louck, SIAM J. Math. Anal. 7, 529 (1976).
[9] W. J. Holman, III, SIAM J. Math. Anal. **11**, 523 (1980).
[10] S. C. Milne, Adv. Math. **58**, 1 (1985).
[11] S. C. Milne, J. Math. Anal. Appl. **122**, 223 (1987).
[12] S. C. Milne, Adv. Math. **57**, 34 (1985).
[13] S. C. Milne, Adv. Math. **57**, 14 (1985).
[14] S. C. Milne, Adv. Math. **57**, 71 (1985).
[15] S. C. Milne, J. Math. Anal. Appl. **118**, 263 (1986).
[16] M. V. Tratnik, J. Math. Phys. **30**, 627 (1989).
[17] M. V. Tratnik, J. Math. Phys. **29**, 1529 (1988).

The unitary irreducible representations of SU(2,1)

Itzhak Bars and Zhong-jian Teng
Physics Department, University of Southern California, Los Angeles, California 90089-0484

(Received 12 September 1989; accepted for publication 7 March 1990)

This paper analyzes the irreducible unitary representations of SU(2,1) in a basis labeled as $|p,q;jmy\rangle$, where p,q correspond to quantum numbers associated with the quadratic and cubic Casimir operators, j,m label states of the SU(2) subgroup, and y labels the quantum number with respect to the U(1) subgroup. All the irreducible representations are found and the allowed range of these quantum numbers for each representation are given. The results are expressed in the form of diagrams that show the allowed values in a (j,y) plot for fixed values of p,q. A (p,q) plot is also provided that indicates the allowed values of these quantum numbers.

I. INTRODUCTION

Recently it has been shown that noncompact affine current algebras can be used as building blocks in the construction of unitary conformal field theories that describe vacuum configurations of string theory.[1,2] In this construction it is necessary to use unitary representations of noncompact groups in which the states are labeled by the eigenvalues of the Casimir operators and by the quantum numbers of the maximal compact subgroup. Thus, for SU(1,1), the labeling is $|jm\rangle$, where j labels the Casimir and m labels the U(1) subgroup. For SU(2,1) the labeling must be done as indicated in the abstract. In Ref. 2 a construction using harmonic oscillators was used in order to analyze some discrete representations of SU(N,M), including SU(2,1) and applying them in conformal field theory. However, one needs to know all the irreducible representations with the complete allowed range of the quantum numbers in order to carry out the full analysis of the conformal field theory. The most important noncompact groups in this application are SU(1,1) and SU(2,1).[2] The SU(1,1) case is fully understood[1,2] while for SU(2,1) we have results only for certain discrete representations.[2]

The mathematical literature[3] on noncompact groups is mainly developed in the Iwasawa decomposition (that uses triangular subgroups) while some mathematical physics literature[4,5] uses still a different basis than the one useful in our application. Furthermore, most of the available discussion is limited to the discrete series representations that can be much more easily handled in terms of oscillators as in Refs. 2 and 6.

For our application we have thus found it necessary to develop the full set of representations in the SU(2)×U(1) basis, labeled with the quantum numbers $|p,q;jmy\rangle$. This is the same basis used for SU(3) in particle physics applications, and was developed fully by Biedenharn and other authors[7] in the case of SU(3). Actually, SU(2,1) was studied in the basis of interest to us sometime ago by Biedenharn,[8] but this work seems to have been forgotten and we became aware of it when it was pointed out to us by the editor after submitting our paper to the journal. Our method is similar to Refs. 7 and 8, which amounts to methods of quantum mechanics applied to the analysis of a system of operators that form a Lie algebra. This aproach yields all the unitary representations and the allowed range of all quantum numbers as demonstrated in this paper. The set of representations that we found are considerably larger than those of Ref. 8 due to the fact that the authors put a global restriction that we do not impose. In our application in conformal field theory based on SU(2,1) we need to discuss the covering group and hence the global condition of Ref. 8 should not *a priori* be imposed. Hence we discover a much larger set of representations that were not previously discussed in our basis even for ordinary SU(2,1).

It is useful to point out that, as is well known, all unitary representations of SU(3) are finite dimensional and can be obtained by taking direct products of the fundamental 3 and $\bar{3}$ representations. The most convenient form corresponds to traceless tensors $T^{j_1 j_2 \cdots j_q}_{i_1 i_2 \cdots i_p}$ with completely symmetrized p lower indices and q upper indices. A single lower index $i = 1,2,3$ correponds to three-dimensional fundamental representation while a single upper index $j = 1,2,3$ is the three-dimensional complex conjugate representation. The quadratic and cubic Casimir operators have eigenvalues on this tensor, and they can be shown to be given in terms of the nonnegative integers (p,q). The (p,q) labels that we use for SU(2,1) in $|p,q;jmy\rangle$ correspond to the same ones as the SU(3), except that, as we shall see they take on not only integer values, but also other values on the real line as well as in the complex plane. When (p,q) are integers, the remaining quantum numbers (j, m, y) take values in different regions for SU(3) or SU(2,1) cases, as seen in the plots that we obtain. It is instructive to keep in mind the SU(3) case as we develop SU(2,1) representations and note the differences, as we shall do below.

II. GENERATORS AND COMMUTATION RULES

We prefer the matrix form of SU(2,1) generators, decomposed according to the maximal compact SU(2)×U(1) subgroup

$$Q^A_B = \begin{pmatrix} J^a_b + Y/2 & iK^a \\ iK^\dagger_a & -Y \end{pmatrix}, \tag{2.1}$$

where the hypercharge operator Y is the generator of the

 0022-2488/90/071576-11$03.00

U(1) subgroup and the matrix form of the subgroup SU(2) is

$$J^a_b = \begin{pmatrix} J_0 & J_+ \\ J_- & -J_0 \end{pmatrix}. \tag{2.2}$$

Furthermore,

$$K^a = \begin{pmatrix} K_{1/2} \\ K_{-1/2} \end{pmatrix} \tag{2.3}$$

form a doublet $j = 1/2$ operator under the SU(2) subgroup and carries U(1) charge $+1$. Thus, the low case letters a,b take the values $\pm\frac{1}{2}$. The Hermitian conjugate of K^a is denoted as $K^\dagger_a = (K^\dagger_{-1/2}, K^\dagger_{1/2})$ and it forms a row matrix that transforms as a doublet under SU(2) from the right and has hypercharge -1. Thus, when an upper index a has the value $+\frac{1}{2}$ the corresponding lower index has the value $-\frac{1}{2}$, etc. Note the factor of i that appears in the definition of $K, K^\dagger$ in (2.1). This factor of i would be missing if we were considering SU(3) instead of SU(2,1).

The commutation rules for both SU(2,1) and SU(3) may be given in the form

$$[Q^A_B, Q^C_D] = \delta^C_B Q^A_D - \delta^A_D Q^C_B, \tag{2.4}$$

from which we may extract

$$\begin{aligned}
&[J_0, J_\pm] = \pm J_\pm, \quad [J_+, J_-] = 2J_0, \quad [Y, J^a_b] = 0,\\
&[J_0, K_{\pm 1/2}] = \pm\tfrac{1}{2} K_{\pm 1/2}, \quad [Y, K_{\pm 1/2}] = K_{\pm 1/2},\\
&[J_+, K_{-1/2}] = K_{1/2}, \quad [J_-, K_{1/2}] = K_{-1/2},\\
&[K_{\pm 1/2}, K^\dagger_{\mp 1/2}] = \mp J_0 - \tfrac{3}{2} Y,\\
&[K_{\pm 1/2}, K^\dagger_{\pm 1/2}] = -J_\pm,\\
&[J_+, K_{1/2}] = [J_-, K_{-1/2}] = 0,\\
&[K_{1/2}, K_{-1/2}] = [K^\dagger_{-1/2}, K^\dagger_{1/2}] = 0,
\end{aligned} \tag{2.5}$$

and the commutation rules among $K^\dagger_a$ with $J\pm$, J_0, Y are obtained by taking Hermitian conjugates of those involving K^a. The commutation rules for SU(3) differ from the above only by having the opposite signs on the right-hand side of the commutators $[K^a, K^\dagger_b]$. The origin of this difference is the analytic continuation provided by the insertion of the factors of i in (2.1).

A harmonic oscillator representation may be given for these generators.[2,6] To do so we introduce annihilation operators $A_a(p)$ and $B(p)$ and creation operators $A^{\dagger a}$, $B^\dagger$, with their lower and upper indices in one-to-one correspondence with the indices above. The label $p = 1,2,...,P$ is an additional index corresponding to P copies of these oscillators, where P is an arbitrary positive integer. By taking many copies of oscillators we are able to construct more general representations. For SU(3) all representations may be constructed with this method, but for SU(2,1) only certain discrete integer representations can be described with harmonic oscillators, as we shall see below. Thus we can now write for SU(2,1)

$$\begin{aligned}
&J^a_b = \sum_p \big(A^{\dagger a}(p) A_b(p) - \tfrac{1}{2}\delta^a_b A^{\dagger c}(p) A_c(p)\big),\\
&Y = \tfrac{1}{3}(A^\dagger \cdot A + 2B^\dagger B + 2P),\\
&K^a = \sum_p A^{\dagger a}(p) B^\dagger(p), \quad K^\dagger_a = \sum_p A_a(p) B(p).
\end{aligned} \tag{2.6}$$

It can be checked that these correctly satisfy the SU(2,1) algebra. For SU(3), B and $B^\dagger$ are exchanged in the construction of K, $K^\dagger$ and the factor involving $2P$ is omitted in Y. In this paper we will first develop the representations generally without using harmonic oscillators. We will then compare the more limited harmonic oscillator results to the general situation.

According to quantum mechanics, given an algebra of operators as above, we need to choose a maximal number of commuting operators that can simultaneously be diagonalized, and then use their eigenvalues in order to provide a complete set of labels in the Hilbert space. The obvious candidates include Y, J_0, which form the Cartan subalgebra, the quadratic Casimir operator for SU(2) given by $\mathbf{J}\cdot\mathbf{J} = \frac{1}{2}\operatorname{Tr}(J)^2$, and the quadratic and cubic Casimir operators of SU(2,1) given by

$$\begin{aligned}
C_2 &= \tfrac{1}{2}\operatorname{Tr} Q^2 \equiv \tfrac{1}{2}\operatorname{Tr} Q^{T^2}\\
&= \mathbf{J}\cdot\mathbf{J} + \tfrac{3}{4} Y(Y-2) - K^a K^\dagger_a\\
&= \mathbf{J}\cdot\mathbf{J} + \tfrac{3}{4} Y(Y+2) - K^\dagger_a K^a,
\end{aligned} \tag{2.7}$$

$$\begin{aligned}
C_3 &= (1/3!)\operatorname{Tr}\{Q^3 + Q^{T^3}\}\\
&= \mathbf{J}\cdot\mathbf{J}(Y+1) - \tfrac{1}{4} Y(Y+1)(Y+2)\\
&\quad - K^\dagger_a (J^a_b - \delta^a_b) K^b + \tfrac{1}{2} K^\dagger_a Y K^a\\
&= \mathbf{J}\cdot\mathbf{J}(Y-1) - \tfrac{1}{4} Y(Y-1)(Y-2)\\
&\quad - (J^a_b - \delta^a_b) K^b K^\dagger_a + \tfrac{1}{2} Y K^a K^\dagger_a,
\end{aligned} \tag{2.8}$$

where Q^t is the transpose of the matrix $\mathbf{Q}$. In the two forms of the Casimir operators we have used the commutation rules in order to move either K or $K^\dagger$ to the right. For SU(3), the Casimir operators differ from the above because they have minus signs in front of all the terms involving $K, K^\dagger$.

The eigenvalue of the SU(2) Casimir operator will be parametrized as usual by $\mathbf{J}\cdot\mathbf{J} = j(j+1)$, while the eigenvalues of the SU(2,1) Casimir operators will be parametrized by

$$\begin{aligned}
C_2 &= p + q + \tfrac{1}{3}(p^2 + pq + q^2)\\
&= -1 - (x_1 x_2 + x_2 x_3 + x_3 x_1),\\
C_3 &= \tfrac{1}{27}(p-q)(p+2q+3)(q+2p+3) = x_1 x_2 x_3,
\end{aligned} \tag{2.9}$$

where

$$\begin{aligned}
x_1 &= -(p+2q+3)/3,\\
x_2 &= (q-p)/3,\\
x_3 &= (2p+q+3)/3.
\end{aligned} \tag{2.10}$$

Note that $x_1 + x_2 + x_3 = 0$, which reduces the independent parameters to two among the (x_1, x_2, x_3). As we shall see, the x_1, x_2, x_3 parametrization will be handy in our analysis although this parametrization is equivalent to that of (p,q). As already mentioned earlier, for SU(3) the integers (p,q) are related to the rank of the tensor, but for SU(2,1) their values will differ significantly from SU(3) and they have no tensor interpretation.

III. UNITARITY CONDITIONS

The five mutually commuting operators (C_3, C_2; $\mathbf{J}\cdot\mathbf{J}$, J_0, Y) are simultaneouly diagonal on the states labeled by $|p,q;jmy\rangle$, where (m,y) are the eigenvalues of (J_0,Y) that form the Cartan subalgebra, while the eigenvalues of the Casimirs have been parametrized above. Our task is to determine the action of the remaining generators on these states and obtain the allowed values of (p, q, j, m, y) so that this basis provides a unitary representation of SU(2,1). Unitarity demands that the generators (or their matrix representations in this basis) are Hermitian and that the norms of all states are positive. From the Hermiticity of the $SU(2)\times U(1)$ generators we already know that the eigenvalues (j, m, y) must be real. As we shall see later the half-integer values of j and the values of y will be required to be within certain ranges depending on the values of (p, q).

Hermiticity of the Casimir operators C_2, C_3 also requires that their eigenvalues be real, that is, $x_1x_2 + x_2x_3 + x_3x_1 = x_1^*x_2^* + x_2^*x_3^* + x_3^*x_1^*$ and $_1x_2x_3 = x_1^*x_2^*x_3^*$. In general this is satisfied if either all (x_1, x_2, x_3) are real or if one of them is real and the other two are complex conjugates of each other. Thus, (p,q) will generally be allowed to be complex.

There are three possibilities if (p,q) are complex, depending on which x_i is real

$$\begin{aligned} p &= -1-r+is, \quad q = -1+r+is \quad (x_2=\text{real}),\\ p &= -1+r-is, \quad q = -1-i2s \quad (x_3=\text{real}), \qquad (3.1)\\ p &= -1+i2s, \quad q = -1-r-is \quad (x_1=\text{real}), \end{aligned}$$

where $n = 0,1,2,3,\ldots$ and as in (4.3), the points $y = y_0$ at $j = j_0 + n/2$ are included only if n is an even integer.

$$C_2 = r_2/3 - s^2 - 1, \quad C_3 = (2r/27)(r^2 + 9s^2). \quad (3.2)$$

When (p,q) are real we can divide the real (p,q) plane into six regions that correspond to six possible ways of ordering (x_1, x_2, x_2) on the real line. It is evident from (2.9) that any one of these regions will produce the same real eigenvalues for (C_2, C_3). It is therefore sufficient to take them in the region $(p\geqslant -1, q\geqslant -1)$, which corresponds to the order $x_1\leqslant x_2\leqslant x_3$. The mapping $(p,q)\to(C_2, C_3)$ is evidently six to one with the equivalent points $(A_1\sim A_{-1}\sim A_2\sim A_{-2}\sim A_3\sim A_{-3}$ as shown in Fig. 1) given by

$$\begin{aligned} (p,q) &\sim (-q-2, -p-2),\\ &\sim(-p-2, p+q+1)\sim(-(p+q+3),p), \qquad (3.3)\\ &\sim(p+q+1, -q-2)\sim(q, -(p+q+3)), \end{aligned}$$

with the first factor $A_1 = (p,q)$ taken in the region $(p>1, q>-1)$.

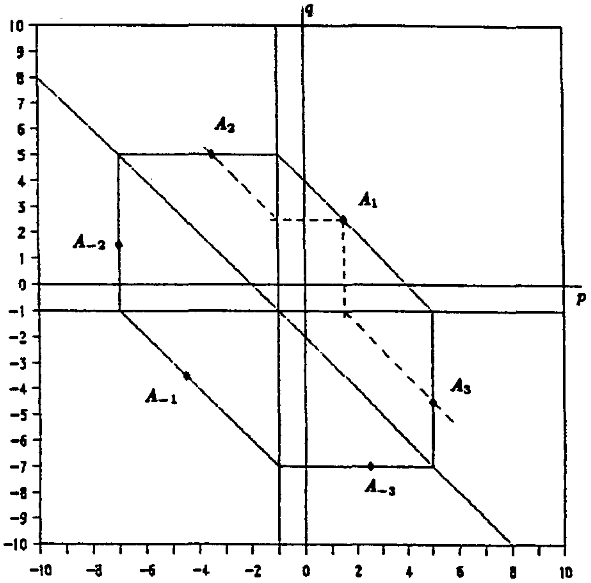


FIG. 1. The six equivalent points in p-q diagram. The following six points are equivalent: $A_1\sim A_{-1}\sim A_2\sim A_{-2}\sim A_3\sim A_{-3}$, $A_1 = (p,q)$, $A_{-1} = (-q-2, -p-2)$; $A_2 = (-p-2, p+q+1)$, $A_{-2} = (-(p+q+3),p)$; $A_3 = (p+q+1, -q-2)$, $A_{-3} = (q, -(p+q+3))$, *graphic parameters $A_1 = (p,q) = (1.5, 2.5)$. In this diagram, one can reach A_2 and A_3 from A_1 following the dotted lines shown. Finally, the other points A_{-1}, A_{-2}, A_{-3} are obtained by a reflection from the solid line that makes a $3\pi/4$ angle with the p axis and passes the point $(p,q) = (-1, -1)$.

Although the Lie algebra is represented equally in any of these regions of the complex or real (p,q) plane, the global properties of the SU(2,1) group representations may very well be different in each of these regions. [For an example of such global properties consider the case of the supplementary series for SU(1,1) whose Casimir eigenvalue depends only on the absolute value of a parameter σ which lies in the range $-1\leqslant\sigma\leqslant 1$, but which has global representations that differ depending on the sign of σ.[9]] In this paper we will concentrate only on the representations of the Lie algebra and will not discuss global properties of the group representations. Therefore, our discussion will concentrate on the region $(p\geqslant -1, q\geqslant -1)$ when (p,q) are real, and the case of $x_2 = \text{real}$ when (p,q) are complex. We find that we do not miss any representations by concentrating in this region provided we include in our list of representations the complex conjugate representations of those discussed below, if they are not already included automatically.

When the unitarity requirements are applied to the SU(2) subgroup they yield the familiar results, namely $m = -j, -j+1,\ldots, j-1, j$ and $j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2,\ldots$. Furthermore, the action of $J_\pm$ on the states is

$$J_\pm |pq;jmy\rangle = \sqrt{j(j+1) - m(m\pm 1)}\,|pq;j(m\pm 1)y\rangle. \quad (3.4)$$

There remains to figure out the action of K^a and $K_a^\dagger$ on the states. Since K^a, $K_a^\dagger$ carry "hypercharge" $y = 1, -1$, respectively, their action on states shifts the eigenvalue y by the corresponding amount (± 1). Furthermore since, from the point of view of SU(2), these are tensor operators with "spin" $j=\frac{1}{2}$, we can use the Wigner–Eckhart theorem[10] to write their action on the states up to "reduced" matrix elements in the form,

$$K_{\pm 1/2}|p,q;jmy\rangle$$
$$=c_+(jy)\sqrt{\frac{j\pm m+1}{2j+1}}\left|p,q;\left(j+\frac{1}{2}\right)\left(m\pm\frac{1}{2}\right)(y+1)\right\rangle$$
$$\pm c_-(jy)\sqrt{\frac{j\mp m}{2j}}\left|p,q;\left(j-\frac{1}{2}\right)\left(m\pm\frac{1}{2}\right)(y+1)\right\rangle, \tag{3.5}$$

$$K^\dagger_{\pm 1/2}|p,q;jmy\rangle$$
$$=\mp d_+(jy)\sqrt{\frac{j\pm m+1}{2j+1}}$$
$$\left|p,q;\left(j+\frac{1}{2}\right)\left(m\pm\frac{1}{2}\right)(y-1)\right\rangle$$
$$+d_-(jy)\sqrt{\frac{j\mp m}{2j}}\left|p,q;\left(j-\frac{1}{2}\right)\left(m\pm\frac{1}{2}\right)(y-1)\right\rangle.$$

The reduced matrix elements $(c_\pm, d_\pm)$ can depend only on the quantum numbers (p,q,j,y) since all the m dependence is already displayed in the SU(2) Clebsch–Gordan coefficients (the square roots) dictated by the Wigner–Eckhart theorem. In order to obtain these four coefficients we apply the $[K,K^\dagger]$ commutators on the states, use the above formulas for acting on an arbitrary state, and further demand that the quadratic and cubic Casimir operators be diagonal with the eigenvalue parametrized as in Eqs. (2.9) and (2.10). These conditions completely fix $(c_\pm, d_\pm)$. We then find that we can parametrize these coefficients in terms of the cubic function

$$G(x)=(x-x_1)(x-x_2)(x-x_3)$$
$$=x^3-(C_2+1)x-C_3, \tag{3.6}$$

where (x_1, x_2, x_3) are given in terms of (p,q) as in (2.10), and x is expressed in terms of (j,y) depending on the coefficient as follows:

$$c_+(jy)=\sqrt{\frac{G(j+y/2+1)}{2j+2}},$$
$$c_-(jy)=\sqrt{\frac{-G(-j+y/2)}{2j+1}}, \tag{3.7}$$
$$d_+(jy)=\sqrt{\frac{-G(-j+y/2-1)}{2j+2}},$$
$$d_-(jy)=\sqrt{\frac{G(j+y/2)}{2j+1}}.$$

It can be verified that this form provides the general representation of the SU(2,1) commutation rules on the states.

Given the reality conditions on (j,m,y) (or x) and the reality (or complex) conditions placed on (x_1, x_2, x_3) (or C_2, C_3) arrived at in the first paragraph of this section, we see that in a unitary representation $G(x)$ is real. For unitarity we should have positive norms and Hermiticity of matrix representations for all generators. Requiring a positive norm

$$\langle p,q;jmy|p',q';j'm'y'\rangle=\delta_{pp'}\delta_{qq}\delta_{jj'}\delta_{mm'}\delta_{yy'}, \tag{3.8}$$

and demanding that the matrix elements of K^a be the Hermitian conjugates of those of $K^\dagger_a$, that is,

$$\langle p,q;jmy|K^\dagger_{\mp 1/2}|p,q;j'm'y'\rangle$$
$$=(\langle p,q;j'm'y'|K_{\pm 1/2}|p,q;jmy\rangle), \tag{3.9}$$

corresponds to insuring that $(c_\pm, d_\pm)$ be real. This is possible provided the arguments of each of the four square roots, in (3.7) are positive:

$$G(j+\tfrac{1}{2}y+1)\geqslant 0,\quad -G(-j+\tfrac{1}{2}y)\geqslant 0,$$
$$-G(-j+\tfrac{1}{2}y-1)\geqslant 0,$$
$$G(j+\tfrac{1}{2}y)\geqslant 0. \tag{3.10}$$

This restricts severely the allowed ranges of (p,q,j,y). [For SU(3) we need to require the opposite sign for each $G(x)$ since we need to multiply it by an extra minus sign. This corresponds to multiplying $(c_\pm, d_\pm)$ by a factor of $-i$ that is equivalent to removing the factor of i in (2.1).]

IV. THE SERIES OF IRREDUCIBLE UNITARY REPRESENTATIONS

It is easy to check from (3.5) that each time $K,K^\dagger$ are applied on the states the variable x appearing in $c_\pm$ is increased by 1 or 0, while the argument x appearing in $d_\pm$ is decreased by 1 or 0. This implies that in a (j,y) plot the action of $K,K^\dagger$ can be seen to correspond to stepping motions along straight lines on which the states are designated as discrete points, as seen in the plots that we present below. Taking this stepping behavior of x into account we examine the inequalities (3.10). It is evident that in order to avoid getting out of the allowed region by the aplication of power of K or $K^\dagger$ we must require that the stepping operation terminates by satisfying

$$(x-x_1)(x-x_2)(x-x_3)=0. \tag{4.1}$$

Applying this condition with the four types of arguments $x=(j+y/2+1)$ or $(-j+y/2)$ or $(-j+y/2-1)$ or $(j+y/2)$ maps out the boundaries of the allowed regions. In all the (j,y) plots below (Fig. 2 and Figs. 3–11) these boundaries are indicated as dotted lines. Then, taking into account the inequalities, we can easily identify graphically the regions allowed by the unitarity conditions (3.10). In general there will be several disconnected regions allowed by the inequalities. However, this is not all. For any given region to be unitary, we must also insure that the spin eigenvalues j included in that region are *half-integer*. In all our plots the lowest spin state of an allowed region is indicated with a heavy dot. [As we shall note later, the irreducible representations of SU(3) as well as SU(2,1) can be both allowed in some integer series plots, e.g. in Fig. 3, in which case we label the lowest spin state of SU(3) by a circle instead of a heavy dot.] In some of the figures there are regions that are excluded when the spin condition is not satisifed while the same figure contains regions that do satisfy the spin condition. The half-integer spin condition may put restrictions on the values of (p,q) beyond those imposed by Hermiticity

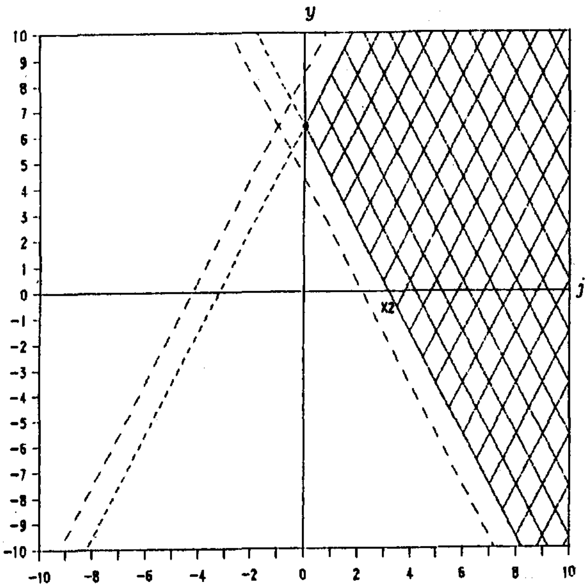


FIG. 2. Principal series. $p+1=-r+si$, $q+1=r+si$, $x_1=-r/3-si$, $x_2=2r/3=$ real, $x_3=-r/3+si$, where $r,s=$ any real number, $C_2+1=(r^2-s^2)/3$, $C_3=2r(r^2/9+s^2)/3$, *Graphic parameter $r=4.8$.

which was discussed in the beginning of Sec. III. These will become evident as we discuss the various possible values of (p,q) case by case below.

A. Principal series, Fig. 2

When (p,q) are complex as in Eq. (3.1), as already explained, we will consider only the parametrization of which x_2 is real since all other cases are equivalent at the algebra level. We will refer to this case as the *principal series*. Since $x_1+x_2+x_3=0$ we can write $x_1=(-x_2+i\rho)/2=x_3$, so that $G(x)=[(x+x_2/2)^2+\rho^2/4](x-x_2)$ showing that the boundaries for $G(x)=0$ arise only from $x=x_2$. [As we shall note later, the irreducible representations of SU(3) as well as SU(2,1) can be both allowed in some integer series plots, e.g. in Fig. 3, in which case we label the lowest spin state of SU(3) by a circle instead of a heavy dot.] For the four cases of x indicated in (3.10) we draw the four dotted lines in Fig. 2. The region allowed by the inequalities (3.10) is shaded by lines that correspond to steps taken by x when $K,K^\dagger$ are applied. Recall that K increases the eigenvalue y while $K^\dagger$ decreases it. The states are the points where these lines cross. The state of lowest spin is unique. In this case it has the spin-hypercharge (j_0, y_0) quantum numbers

$$j_0=0, \quad y_0=2x_2=4r/3\,, \tag{4.2}$$

and is indicated by a heavy dot. The hypercharge y_0 is not *a priori* quantized without imposing additional global constraints that may be required in physical applications. Note that this is *not* the lowest (or highest state) in the traditional sense of weights that are given as the eigenvalues of the Cartan generators (m,y). However, this state plays a similar role in that all states are obtained from this one by applying $K,K^\dagger$ repeatedly. We will call this state the *generating state* to distinguish it from the lowest or highest weight states that will appear in some of the plots below. For the principal series the generating state does not put new conditions on (p,q) beyond those already indicated in (3.1). The rest of the allowed states have quantum numbers (j,y)

$$j=n/2, \quad y=y_0-n, y_0-n+2,\ldots,y_0-2,[y_0]y_0+2,\ldots,y_0+n-2,y_0+n, \tag{4.3}$$

for every non-negative integer $n=0,1,2,3,\ldots$. Note that the point with $y=y_0$ is included among the $j=n/2$ states only if n is even.

B. Discrete series of $p,q,(p+q)$ types, Figs. 8, 9(a), 9(b)

When (p,q) are real as in (3.3), then they are allowed to be only as indicated in Fig. 12. Namely, they can take values on any vertical line for which $p=$ integer or any horizontal line for which $q=$ integer or any slanted line on which $p+q=$ integer or they can take values in the shaded regions. It will become apparent how these restrictions arise as we discuss each case separately. As already indicated, we will restrict our analysis to the quadrant defined by $(p>-1, q>-1)$. When p is integer and q is not we call the representation the *p-discrete series*. Similarly if q is integer and p is not we call it the *q-discrete series*. Finally, if neither p nor q are integers but $p+q$ is an integer, we call it the $(p+q)$*-discrete series*. These lie in Fig. 12 on any vertical, horizontal, slanted line, respectively, but not on any intersection of these lines. In addition we need to consider the points of intersection and the shaded regions of Fig. 12 which we will discuss separately.

First consider the p-discrete series, as in Fig. 9(a). Here, $G(x)$ can vanish when x equals any real x_i. Taking into account the four possible expressions of x as in (3.10), we draw the 12 dotted lines (six slanted forward and six slanted backward) that provide the boundaries. Then we look for the regions that satisfy the inequalities (3.10). There are three such regions but only in one of them the spin j takes non-negative half-integer values thanks to the fact that p is

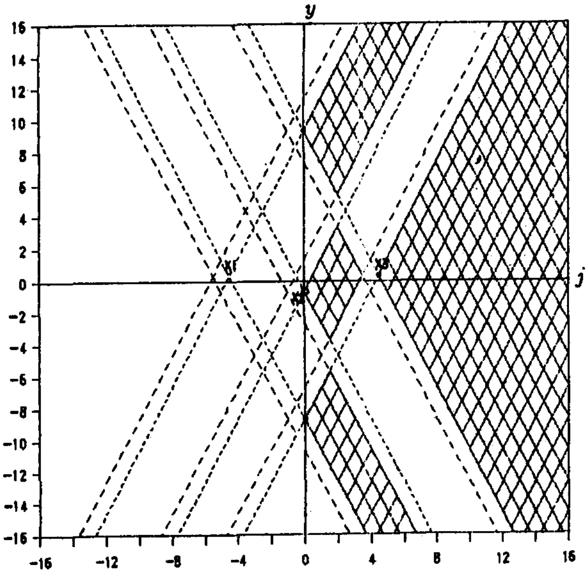


FIG. 3. Integer series. $p+1=$ int, $q+1=$ int, *graphic parameters $p=4$, $q=3$.

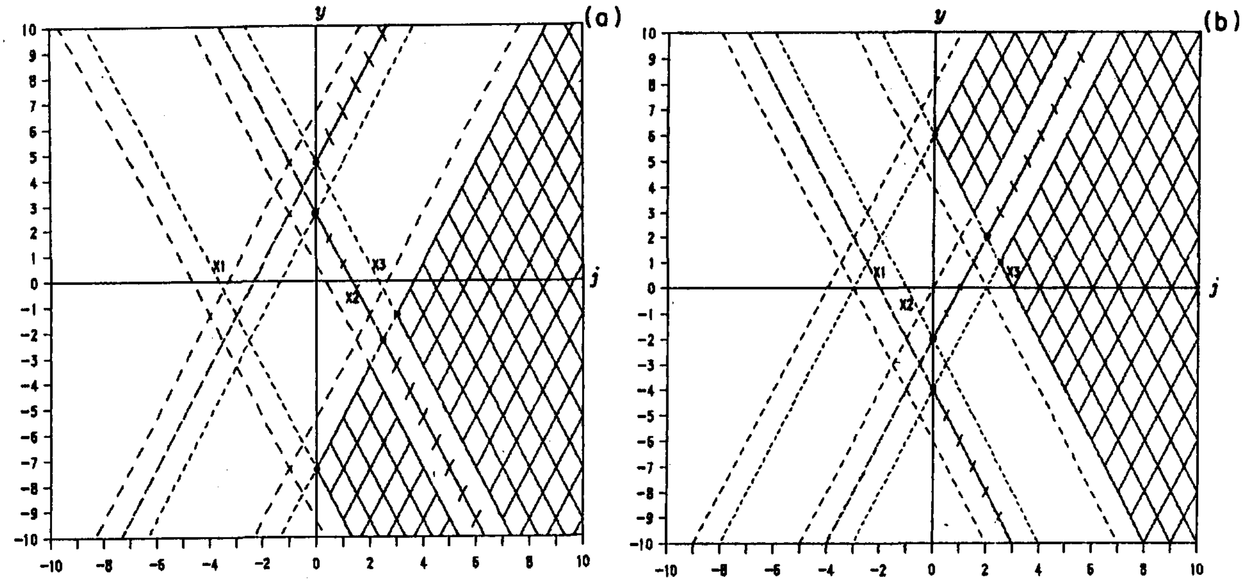


FIG. 4. Integer series. (a) $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=0$, $q=4$. (b). Integer series. $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=3$, $q=0$.

an integer (in considering the 12 boundary lines the half-integer requirement for the spin was ignored at first). Had p not been an integer we would have to discard that region as well. This is why (p,q) needs to be on one of the vertical lines in Fig. 12 (similarly for the horizontal lines by exchanging the roles of p and q). The allowed shaded region is indicated in Fig. 9(a) by lines that correspond to stepping via the generators $K,K^\dagger$. Note that in this figure $p=3$ corresponds to the maximum 3 steps that can be taken by applying purely powers of $K^\dagger$ on any state. This is one way of seeing that p needs to be an integer. The state with the lowest spin, the generating state, is indicated by a heavy dot and has

$$j_0=0,\quad y_0=2x_3=\tfrac{2}{3}(2p+q+3). \tag{4.4a}$$

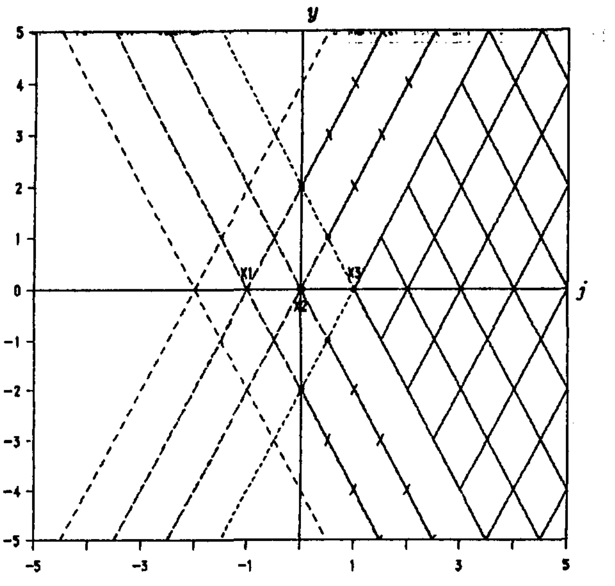


FIG. 5. Integer series. $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=0$, $q=0$.

There is a state with lowest hypercharge at the bottom of the shaded region. This is a spin multiplet that contains the lowest weight state in the traditional sense. Its coordinates are

$$\tilde{j}_0=p/2,\quad \tilde{y}_0=1-x_1=2+(p+2q)/3. \tag{4.4b}$$

The lowest state has $m=-\tilde{j}=-p/2$. All states may be generated from this one (or from the generating state) by applying powers of $K,K^\dagger$. Thus we can characterize all states in the p-discrete series by

$$|\tilde{j}_0-n/2|\leqslant j\leqslant \tilde{j}_0+n/2,\quad y=\tilde{y}_0+n, \tag{4.5}$$

where $n=0,1,2,3,\ldots$. The expression for the limits on j is consistent with the figure, but they also can be regarded to arise through addition of angular momentum.

The q-discrete series in Fig. 9(b) is derived with very similar arguments. In this case the generating state is identified as

$$j_0=0,\quad y_0=2x_1=-\tfrac{2}{3}(p+2q+3). \tag{4.6a}$$

The highest state has $\tilde{y}_0=1+x_1, \tilde{j}_0=q/2, m=q/2$. All the states may be generated either from the generating state or from the highest state and they be characterized by

$$|\tilde{j}_0-n/2|\leqslant j\leqslant \tilde{j}_0+n/2,\quad y=\tilde{y}_0-n, \tag{4.6b}$$

Note that in Fig. 9(b), $q=4$ corresponds to the maximum four steps that are possible by applying powers of only $K_{\pm 1/2}$.

The $(p+q)$-discrete series of Fig. 8 is derived through similar procedures as the above two cases but now the allowed region is quite different. The generating state has an allowed half-integer spin thanks to the fact that $(p+q)$ is an integer. The other regions are excluded because p,q are not individually integers. This explains the allowed slanted lines in Fig. 12. The generating state is identified as

$$j_0=(x_3-x_1)/2=1+\tfrac{1}{2}(p+q),\quad y_0=-x_2=\tfrac{1}{3}(p+q), \tag{4.7}$$

while the set of allowed states is given by

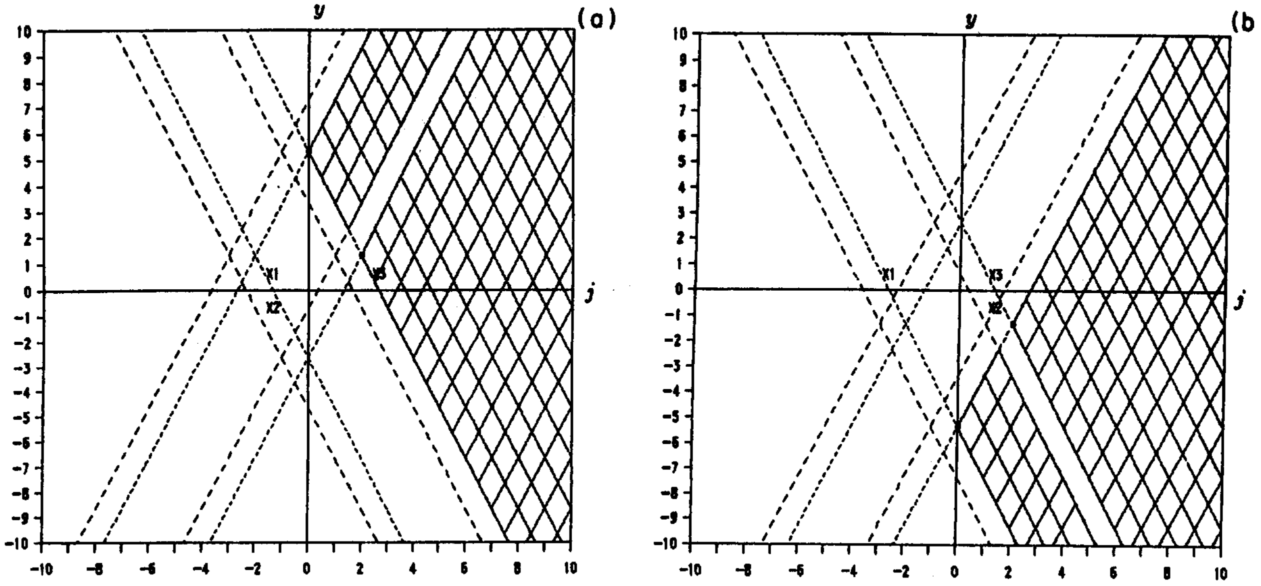


FIG. 6. (a) Integer series. $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=3$, $q=-1$. (b) Integer series. $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=-1$, $q=3$.

$$j=j_0+n/2,\quad y=y_0-n, y_0-n+2,\ldots,y_0-2,[y_0],$$
$$y_0+2,\ldots,y_0+n-2,y_0+n, \quad (4.8)$$

where $(r,s)=\text{real}$. All three parametrizations produce the same eigenvalues for (C_2, C_3),

C. Integer series

When both p and q are integers then all three regions corresponding to the p-discrete, q-discrete, and $(p+q)$-discrete series are allowed simultaneously as seen in Fig. 3. Thus, in this case the eigenvalues of the Casimir operators do not uniquely label the irreducible representations. One must in addition supply the region (that is not overlapping with any of the others). The generating states and the highest or lowest states as well as the sets of all states have formally the same expressions as (4.4)–(4.8) except for the fact that both p and q are integers. This makes the values of y one-third integer in general.

The center of Fig. 3 contains a finite-dimensional SU(3) representation that corresponds to the tensor with p-lower and q-upper indices as mentioned in the Introduction. This is not part of the allowed region for SU(2,1) given in (3.10), but it is the only allowed region if the orientation of the inequalities in (3.10) is reversed, corresponding to SU(3). Note that in Fig. 3, $p=4$, $q=3$ corresponds to the maximum number of steps allowed by pure K motions or pure $K^\dagger$

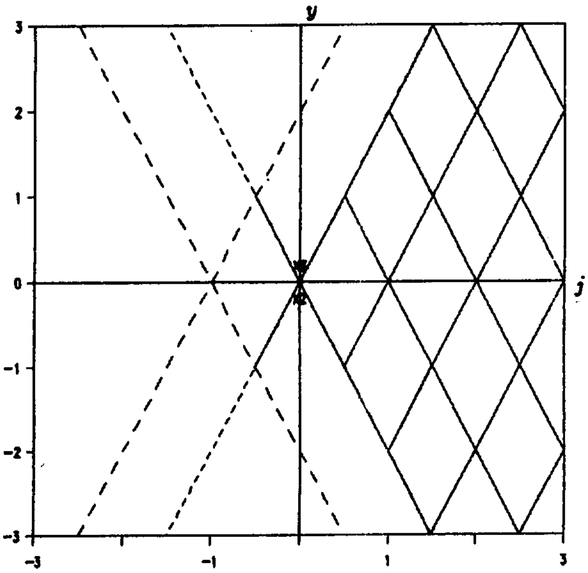


FIG. 7. Integer series. $p+1=\text{int}$, $q+1=\text{int}$, *graphic parameters $p=-1$ $q=-1$.

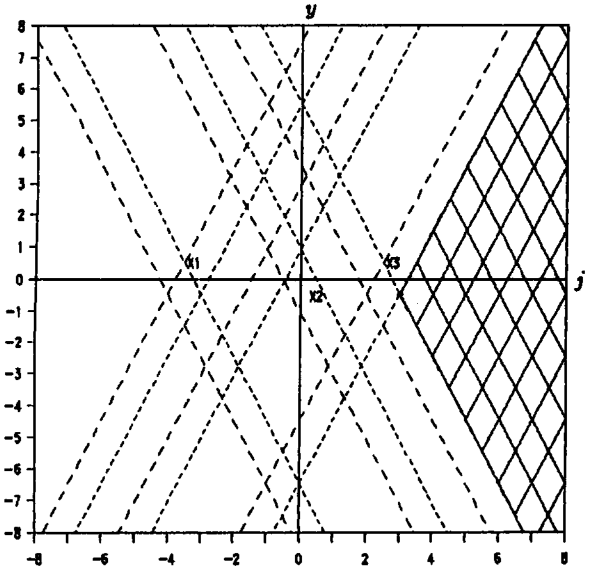


FIG. 8. Discrete series [$(p+q)$-type]. $p+q+2=\text{int}$, but $p>0$, $q>0$, *graphic parameters $p=1.3$, $q=2.7$.

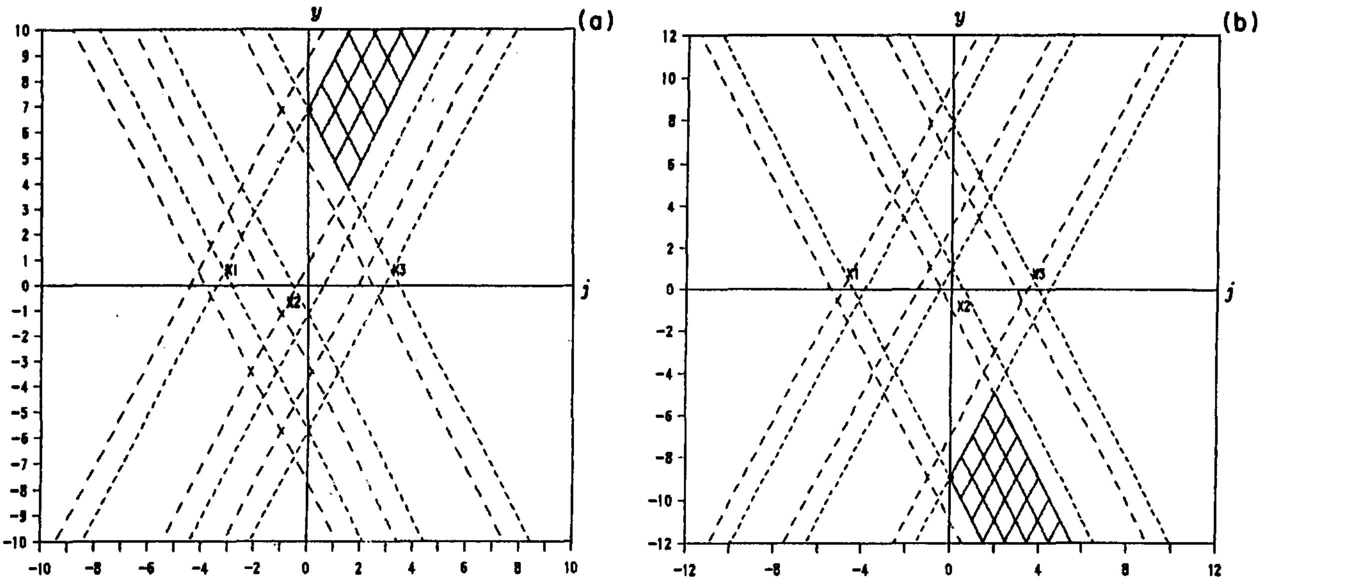


FIG. 9. (a) Discrete series [p-type]. $p+1=\text{int}$, and $q>0$, *graphic parameters $p=3$, $q=1.3$. (b) Discrete series [q-type]. $q+1=\text{int}$, and $p>0$, *graphic parameters $p=2.4$, $q=4$.

motions, respectively, for either the SU(3) representation or the SU(2,1) p-discrete/q-discrete representations.

In Fig. 4(a) we specialize to the case of $p=0$ and $q=$ integer. The p-discrete region has shrunk to a line [so did the SU(3) region] while the q-discrete and $(p+q)$-discrete regions remain as before. However, now another representation has made its appearance in the form of a line stuck in between the q-discrete and $(p+q)$-discrete regions. This line as well as the single line p-discrete region correspond to the merging of two boundary dotted lines. Thus, for these values of (p,q) there are four rather than three distinct SU(2,1) representations and the region must be specified in addition to the Casimirs in order to uniquely identify the irreducible representation. The states for the three regions that we have seen before are correctly described by the previous formulas (except that we need to take the $p=0$ limit). The states of the fourth representation are generated from the highest state and are given by

$$j=\tfrac{1}{2}(q+1+n), \quad y=-n-1-q/3. \tag{4.9}$$

In Fig. 4(b) we have $q=0$, $p=$ integer. Therefore, this case corresponds to interchanging the roles of p and q in the previous paragraph. The information conveyed by the figure is self-evident. The states of the fourth representation are given by

$$j=\tfrac{1}{2}(p+1+n), \quad y=n+1+p/3. \tag{4.10}$$

In Fig. 5 we consider the case of $p=q=0$, or

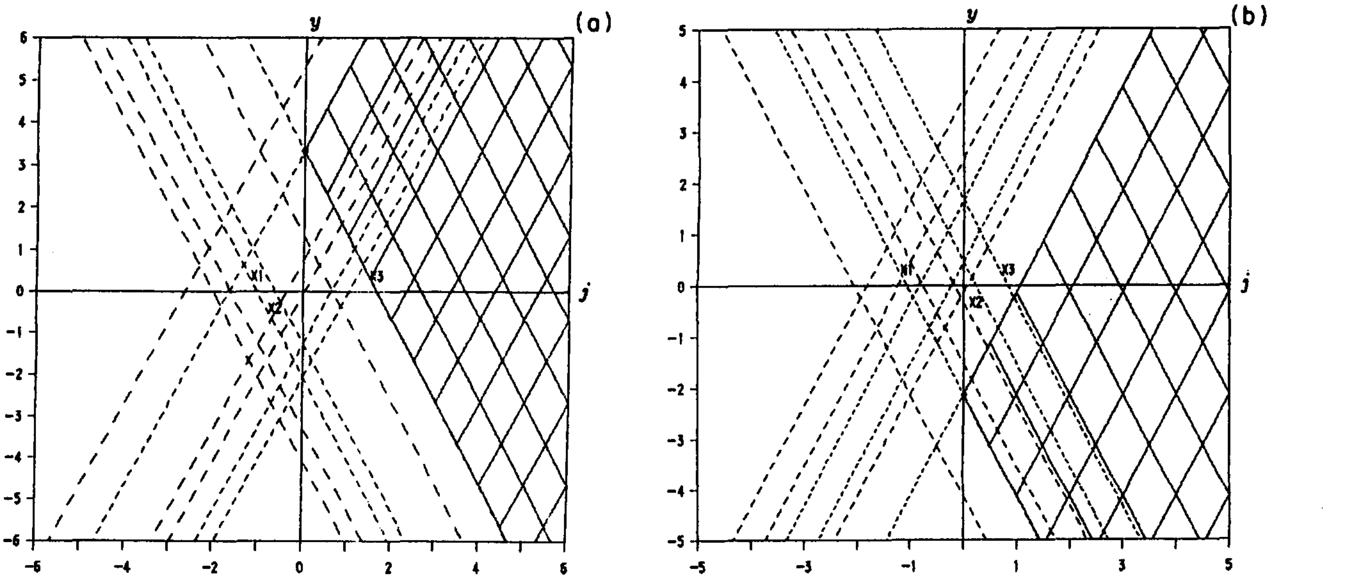


FIG. 10. (a) Supplementary series. $p>0$, $-1<q<0$ and $\tilde{p}'+\tilde{q}'<1$, *graphic parameters $p=1.3$, $q=-0.6$. (b) Supplementary series. $q>0$, $-1<p<0$ and $\tilde{p}'+\tilde{q}'<1$, *graphic parameters $p=-0.4$, $q=0.3$.

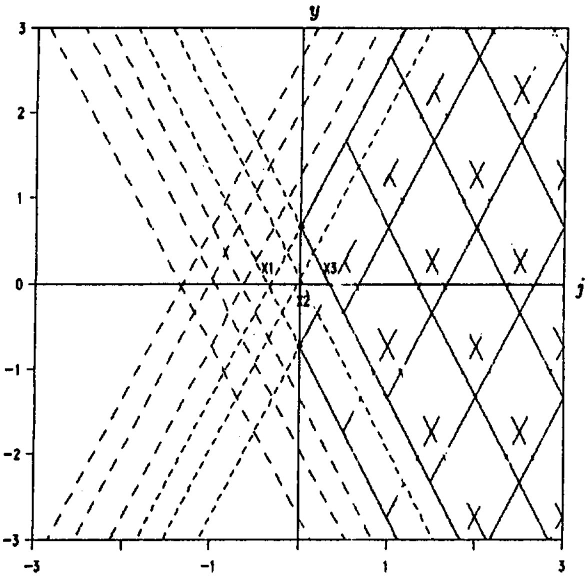


FIG. 11. Supplementary series. $-1<p<0$, $-1<q<0$ and $\tilde{p}'+\tilde{q}'<1$, *graphic parameters $p=-0.7$, $q=-0.6$.

$x_2=0, x_3=1=-x_1$. There are now five representations for the same eigenvalues of the Casimirs. The mechanism by which they arise are evident from the previous discussion. Only the $(p+q)$-discrete region fills an area as before, while all other four regions correspond to single lines. Even though these are neighboring lines one cannot jump from one to the other because they are collapsed boundary lines that satisfy (3.10) and (4.1). The quantum numbers of the states in these five irreducible representations are obtained from the above formulas by specializing to $p=q=0$.

In Fig. 6(a) we consider $q=-1, p=$ integer. There are now only two representations for the same Casimirs, the p-discrete and the $(p+q)$-discrete regions. The q-integer

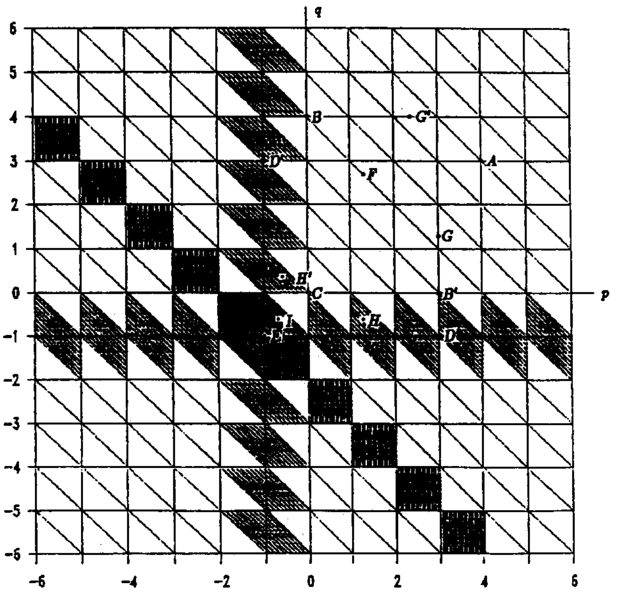


FIG. 12. State-labelling diagram—Allowed values of (p,q). The letters A, B, C, etc. correspond to the (p,q) values used in Figs. 3–11.

region has shrunk away completely. The generating state and the highest lowest weight states of the p-discrete region are given by Eqs. (4.4a) and (4.4b) with $q=-1$, whle the set of all states in this representation is given by (4.5). The states of the $(p+q)$-discrete region are given by (4.7) and (4.8).

Similarly, Fig. 6(b) corresponds to the interchange of p,q with $p=-1$, $q=$ integer. There are again two representations for the same Casimirs: the q-discrete and the $(p+q)$-discrete regions, with the p-discrete region vanished.

Finally, in Fig. 7 we consider the $p=q=-1$ case. This is the only integer point for which there is a unique irreducible representation for given Casimirs $C_2=-1$, $C_3=0$ $(x_1=x_2=x_3=0)$. The surviving region is the $(p+q)$-discrete case, with the other ones vanishing. The generating state is located at $(j_0=0, y_0=0)$ and the set of all states is given in (4.8) as specialized to the current generating state.

D. Supplementary series

The shaded regions in Fig. 12 (for $p>-1, q>-1$) is characterized by either p or q or both being in the region between -1 and 0 and by the condition $\tilde{p}+\tilde{q}=1$, where $(\tilde{p},\tilde{q})$ are the decimal parts of (p,q). The boundaries of the shaded regions are excluded from the supplementary series and they have already been discussed as part of the p-discrete or q-discrete or $(p+q)$-discrete or integer representations. Thus, we are concerned only with the inside of the shaded regions.

Unlike the continuous (p,q) values in the white regions of Fig. 12 for which it is impossible to satisfy the spin condition, the (p,q) values inside the shaded region can satisfy it because the q-discrete or p-discrete regions cease to exist on their own and combine with the previously $(p+q)$-discrete region. Thus, for any (p,q) values inside a shaded region there is an allowed area whose generating state is located at $j=0$, and the figures are *identical.* Thus when q is positive we get a generating state [Fig. 10(a)]

$$j_0=0, \quad y_0=2x_3=\tfrac{2}{3}(2p+q+3), \tag{4.11}$$

When p is positive we get a generating state [Fig. 10(b)]

$$j_0=0, \quad y_0=2x_1=-\tfrac{2}{3}(p+2q+3). \tag{4.12}$$

When both p and q are between -1 and 0, we get two generating weights for the same Casimirs (Fig. 11) whose expressions are the same as (4.11) and (4.12). In order to clearly distinguish the states generated by each one of these generating states, they are indicated by crosses of solid and continuous lines in Fig. 11, respectively. The full set of states are given by (4.8) as applied to the current generating states.

V. HARMONIC OSCILLATOR REPRESENTATIONS

We consider the commutation rules and construction of generators from harmonic oscillators as in Eqs. (2.2)–(2.6). We will work in the Fock space based on the usual vacuum annihilated by all annihilation operators $A^a(p), B(p)$. The harmonic oscillator representation is obviously unitary in Fock space and it allows the construction of a highest weight representation as follows. We choose states $|\Omega\rangle$ that are annihilated by $K_a^\dagger = A_a \cdot B$. In addition, we demand that $|\Omega\rangle$ form a collection of states that transform irreducibly under $SU(2)\times U(1)$. This is the multiplet that contains the lowest weight. Applying on these states all possible powers of $K^a = A^{\dagger a}\cdot B^\dagger$ generates all the states in the SU(2,1) representation. All of these states can be written in terms of irreducible representations with respect to the subgroup $SU(2)\times U(1)$ by working out the Clebsch–Gordan series in the direct products of $|\Omega\rangle$ and the symmetrized powers of K^a's. This is a straightforward exercise for the $SU(2)\times U(1)$ group. Thus, if the lowest weight is identified to have spin and hypercharge $(\tilde{j}_0, \tilde{y}_0)$ then the rest of the states in the irreducible representation have spin and hypercharge given by Eq. (4.5). We wish to find the values of $(\tilde{j}_0, \tilde{y}_0)$ that may be constructed with the harmonic oscillator representation and identify the (p,q) quantum numbers for these representations.

The possible candidates for $|\Omega\rangle$ can be listed;

$$\begin{aligned} &|0\rangle, A^{\dagger a}(p)|0\rangle, \quad B^\dagger|0\rangle,\\ &A^{\dagger a_1}(p_1)A^{\dagger a_2}(p_2)|0\rangle, \quad B^\dagger(p_1)B^\dagger(p_2)|0\rangle,\\ &A^{\dagger a}(p_1)B^\dagger(p_2)|0\rangle, p_1\neq p_2,\\ &A^{\dagger a_1}(p_1)A^{\dagger a_2}(p_2)B^\dagger(p_3)|0\rangle, \quad p_1,p_2\neq p_3, \end{aligned} \tag{5.1}$$

and so on, with any number of $A^\dagger$'s or $B^\dagger$'s, provided the p indices on the A's are different than those on the B's. This structure insures that any of these states is a candidate for the highest state. However, before identifying a specific $|\Omega\rangle$ we must first symmetrize–antisymmetrize the indices of the A's among themselves according to the rules of SU(2) Young tableaux to make an irreducible SU(2) representations. Note that if $P=1$, then the only possibility is completely symmetric tensors of SU(2) corresponding to single row Young tableaux with a boxes, giving $j=a/2$. $P\neq 1$ allows Young tableaux up to P rows but for SU(2) we can only go up to two rows. For a two-row Young tableau with a_1,a_2 boxes, the spin is $j=(a_1-a_2)$. The U(1) quantum number is calculated by adding the U(1) charges of the vacuum and of the A and B oscillators applied on the vacuum. The vacuum has hypercharge $y_{\rm vac}=2P/3$, the $A^{\dagger a}$ has hypercharge $\frac{1}{3}$ and the $B^\dagger$ oscillator has hypercharge $\frac{2}{3}$ all of which is seen from the hypercharge operator in (2.6). Using these rules we can identify the $SU(2)\times U(1)$ content of each of the above candidates for the highest weight in the form $|\Omega\rangle \sim (\tilde{j}_0, \tilde{y}_0)$.

Thus, we give a few examples:

$$\begin{aligned} &|0\rangle \sim (0,2P/3), \quad A^{\dagger a}(p)|0\rangle \sim (\tfrac{1}{2},(2P+1)/3),\\ &B^\dagger(p)|0\rangle \sim (0,2(P+1)/3),\\ &A^{\dagger a}(p)B^\dagger(q)|0\rangle \sim [(\tfrac{1}{2},(3+2P)/3)],\\ &A^{\dagger(a_1}(p_1)A^{\dagger a_2)}(p_2)|0\rangle \sim [(1,2(P+1)/3)],\\ &A^{\dagger[a_1}(p_1)A^{\dagger a_2]}(p_2)|0\rangle \sim [(0,2(1+P)/3)], \end{aligned} \tag{5.2}$$

where the latin indices are symmetrized in line 2 and antisymmetrized in line 3. In this way, it is clear that all values of $\tilde{j}_0 =$ half-integer will be possible through the A oscillators, while the U(1) charge takes the form

$$\tilde{y}_0 = (a + 2(b+P))/3, \tag{5.3}$$

where (a,b) are, respectively, the number of A and B oscillators applied on the vacuum to construct the lowest state $|\Omega\rangle$. Note that P appears only in the combination $b+P=\tilde{b}$. Thus y is quantized and positive since $a,\tilde{b}$ are integers.

We see then that the harmonic oscillator representation reproduces the p-discrete branch of the *integer series* of Fig. 3 and Eq. (4.5), with $\tilde{j}_0 = a/2 = p/2$ and $\tilde{y}_0 = (a+2\tilde{b})/3 = 2 + (p+2q)/3$. This allows us to identify

$$a=p, \quad \tilde{b} = q+3. \tag{5.4}$$

It is possible to also reproduce the q-discrete branch of the integer series in Fig. 3 with harmonic oscillators by taking all the above oscillators to have negative norm. Equivalently, we may keep the definition of the generators as before but change the definition of the vacuum so that it is annihilated by $A_a^\dagger$, $B^\dagger$. The effect of this is to reverse the sign of the hypercharge of the states that we have constructed above, thus corresponding to the q-discrete branch.

Note that when $P=1$ the only representations that are possible with harmonic oscillators are the single line p-discrete branch of Figs. 4(a) and 5 or the single line q-discrete branch of Figs. 4(b) and 5. The extra single line branches of Figs. 4(a) and (b), and 5 cannot be reproduced with harmonic oscillators alone. The integer series $(p+q)$-discrete branches of Figs. 3, 4(a), 4(b), 5, 6(a), 6(b), 7, or the other nonfully integer representations cannot be reproduced with the harmonic oscillator construction of (2.6).

VI. CONCLUSION

The above analysis exhausts all irreducible representations since it covers all possible allowed eigenvalues in the complete labeling $|p,q;jmy\rangle$. As emphasized in Sec. III and Eq. (3.3), when (p,q) are real it is sufficient to discuss the region $(p>-1, q>-1)$ for representing the algebra, but global properties of the group may require more discussion in the entire (p,q) plane. A similar comment applies to the three possible parametrizations when (p,q) are complex, as in Eq. (3.1).

The main motivation for the present paper was to prepare the group theory background for applying SU(2,1) affine current algebras to the construction of conformal and superconformal field theories in $1+1$ dimensions. This will be presented in a separate publication.[11] We hope that in this process we have produced a clear and complete description of all unitary irreducible representations of SU(2,1).

ACKNOWLEDGMENTS

We would like to thank H. Zhang for her advice in computer programming that was helpful for producing this paper. We also thank L. Biedenharn for pointing out to us his work of Ref. 8.

Z.J.T. is supported in part by the U.S. Department of Energy under Contract DE-FG03-84ER40168.

[1] L. Dixon, J. Lykken, and M. E. Peskin, SLAC-PUB-4884 (1989).

[2]I. Bars, USC Preprint USC-89/HEP02, May 1989.
[3]David A. Vogan Jr., *Unitary Representations of Reductive Lie Groups* (Princeton U.P., Princeton, NJ, 1987); N. Wallach, *Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups* (Princeton U.P., Princeton, NJ, 1980); A. W. Knapp, J. Funct. An. **45**, 41 (1982).
[4]J. Patera, J. Math. Phys. **14**, 279 (1973).
[5]B. Gruber, J. Math. Phys. **19**, 2009 (1978).
[6]M. Günaydin and C. Saclioglu, Commun. Math. Phys. **87**, 159 (1982).
[7]L. C. Biedenharn, Phys. Lett. **3**, 69, 254 (1962); J. Math. Phys. **4**, 436 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).
[8]L. C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. Henri Poincaré A **3**, 13 (1965); L. C. Biedenharn, *Some Remarks on Representations of the Noncompact Unitary Groups SU(p,l)*, Lecture, 1967 (Gordon and Breach, New York, 1967).
[9]F. Gürsey and S. Orfanidis, Phys. Rev. D **7**, 2414 (1973), and references therein.
[10]See any quantum mechanics book.
[11]I. Bars and Z. J. Teng (to be published).

Symplectic orbits in quantum state space

Dariusz Chruściński
Institute of Physics, Nicholas Copernicus University, Toruń, Poland

(Received 6 September 1989; accepted for publication 24 January 1990)

All symplectic orbits of the action of an arbitrary compact connected Lie group on the space of density operators are found. It is shown that there is only one orbit that is Kähler (orbit of coherent states).

I. INTRODUCTION

Let $\mathscr{H}$ be a complex finite dimensional Hilbert space and let G be an arbitrary compact connected Lie group. Suppose that we are given an irreducible unitary representation of G on $\mathscr{H}$. Then there is a natural action of G on the Lie algebra su (N) (where N is a dimension of $\mathscr{H}$):

$$\Phi: G \times \mathrm{su}(N) \to \mathrm{su}(N),$$
$$\forall g \in G \quad \forall \alpha \in \mathrm{su}(N) \quad \Phi(g,\alpha) = g\alpha g^+ , \tag{1}$$

where we identify G with its unitary representation and g^+ denotes Hermitian conjugation of g.

Let $O(\alpha)$ and $\hat{O}(\alpha)$ denote orbits of the adjoint representation of SU(N) and of the action (1), respectively, passing through $\alpha \in \mathrm{su}(N)$. It is a well-known fact that $O(\alpha)$ is a symplectic manifold with canonical Kirillov form[1–4] [Kirillov form is a canonical form on orbits of the coadjoint representation but, for su(N), we have a natural isomorphism between adjoint and coadjoint orbits via Killing form[5]].

Let $i_\alpha : \hat{O}(\alpha) \to O(\alpha)$ be an inclusion. There is a question: Which $\alpha \in \mathrm{su}(N)$ gives rise to a symplectic manifold $(\hat{O}(\alpha), i_\alpha^* \omega_\alpha)$? This problem was solved by Kostant and Sternberg[6] only for $\alpha \in P(\mathscr{H})$ (corresponding projective space). In this paper, we give a general solution.

Why is this important for physics? There are at least two reasons. First motivation is connected with a generalization of Hartree–Fock theory[7–9] (HF theory is generalized so as to apply to nondeterminantal densities) and other approximations[10] and the second comes from control theory (a control theoretic aspects of this problem will be published elsewhere). In control theory, $\mathscr{H}$ is a Hilbert space of N-level quantum system S and the Lie algebra $\mathfrak{g}$ of G is generated by a time-dependent Hamiltonian of S. In both cases a state of a physical system is represented by a density operator [Hermitian, semidefinite operator on $\mathscr{H}$ with unit trace (in HF theory trace may differ from unity but it is important that trace is fixed)]. Let $\hat{\mathscr{P}}$ denote the state space. We have the unique decomposition

$$i\hat{\rho} = (i/N)I + \rho , \tag{2}$$

where $\hat{\rho} \in \hat{\mathscr{P}}$ and ρ lives in su(N). Let $\mathscr{P} = \{\rho \in \mathrm{su}(N) : \hat{\rho} \in \hat{\mathscr{P}}\}$. Elements of $\mathscr{P}$ will also be called states. So we are looking for states in $\mathscr{P}$ that lie on symplectic orbits of G.

II. SYMPLECTIC ORBITS

Let us remind the reader of the definition of the Kirillov form ω_ρ on $O(\rho)$ for $\rho \in \mathscr{P}$ (we identify adjoint and coadjoint orbits by Killing form). Any tangent vector $A_\rho \in T_\rho O(\rho)$ is of the form $A_\rho = [A,\rho]$, where $A \in \mathrm{su}(N)$ and

$$\omega_\rho (A_\rho, B_\rho) = (\rho,[A,B]) , \tag{3}$$

where (,) denotes a Killing form on su(N).

But $(A,B) = \mathrm{Tr}(AB)$,[11] thus

$$\omega_\rho (A_\rho, B_\rho) = \mathrm{Tr}(\rho[A,B]) = 2 \,\mathrm{Re}\, \mathrm{Tr}(B\rho A) . \tag{4}$$

If $\rho \in P(\mathscr{H})$ (orbit of pure states) then $\hat{\rho} = |v\rangle\langle v|$, where $v \in \mathscr{H}$ and

$$\omega_{[v]}(A_{[v]}, B_{[v]}) = 2\,\mathrm{Im}\langle Av|Bv\rangle = i < v|[A,B]v> , \tag{5}$$

where $\langle | \rangle$ denotes scalar product in $\mathscr{H}$.

Proposition 1: $\Phi: G \times O(\rho) \to O(\rho)$ defined by (1) is a Hamiltonian action.

Proof: (1) Φ is symplectic:

$$\begin{aligned}(\Phi_g^*\omega)_\rho (A_\rho, B_\rho) &= \omega_{\Phi(g,\rho)} (\Phi_{g*}A_\rho, \Phi_{g*}B_\rho)\\ &= (g\rho g^+, g[A,B]g^+)\\ &= \mathrm{Tr}(g\rho g^+ g[A,B]g^+)\\ &= \mathrm{Tr}(\rho[A,B]) = \omega_\rho(A_\rho, B_\rho)\end{aligned}$$

(2) Φ is Hamiltonian since G is semisimple and cohomology groups $H^1(\mathfrak{g}) = H^2(\mathfrak{g}) = 0$.

Let $J_\rho O(\rho) \to \mathfrak{g}^*$ denote the moment map for this action defined by

$$\forall \rho' \in O(\rho) \quad \forall A \in \mathfrak{g} \quad J_\rho(\rho')(A) = (\rho', A) . \tag{6}$$

We identify $\mathfrak{g}^*$ with $\mathfrak{g}$ so $J_\rho(\rho')$ lives in $\mathfrak{g}$.

Theorem 1: Let $G \times P \to P$ be a Hamiltonian action of a semisimple Lie group G on a symplectic manifold P. Then the G orbit passing through $p \in P$ is symplectic if and only if the stabilizer group of p is equal to the stabilizer group of $J(p)$ ($J: P \to \mathfrak{g}^*$ is a moment map for this action).

For the proof see Ref. 4.

The stabilizer group of $J_\rho(\rho')$ contains some maximal torus T and hence $J_\rho(\rho')$ belongs to the Cartan subalgebra t of $\mathfrak{g}$ (t is a Lie algebra of T). Let h_t denote the Cartan subalgebra of su(N) such that $t \subset h$. From Theorem 1, it follows that if $\hat{O}(\rho')$ is symplectic then $\rho' \in h_t \cap \mathscr{P}$. We must now determine which states $\rho' \in h_t \cap \mathscr{P}$ give rise to symplectic orbits of G. Let $\mathfrak{g}^c$ denote the complexification of $\mathfrak{g}$. We have a decomposition

$$\mathfrak{g}^c = t^c + \oplus CE_\alpha , \tag{7}$$

where E_α denotes a root vector corresponding to the root α and α ranges over all the roots. Similarly, we have decomposition of the real form of $\mathfrak{g}^c$

$$\mathfrak{g} = t + \oplus R(E_\alpha - E_{-\alpha}) + \oplus Ri(E_\alpha + E_{-\alpha}) , \tag{8}$$

where α ranges over all positive roots. Subspaces $\mathscr{C}_\alpha$ spanned by $[E_\alpha - E_{-\alpha}, \rho']$ and $i[E_\alpha + E_{-\alpha}, \rho']$ are mutu-

 0022-2488/90/071587-02$03.00

ally orthogonal with respect to $\omega_{\rho'}$ for different positive roots. So to check that $i^*_{\rho'}\omega_{\rho'}$ is a symplectic form we need to know that if $\omega_{\rho'}$ vanishes on $\mathscr{C}_\alpha$ then tangent vectors are zero. From (4) it follows that if $(\rho',[E_\alpha,E_{-\alpha}])=0$ then $E_\alpha\rho'=\rho'E_\alpha=0$ (since $E_\alpha^+=E_{-\alpha}$). But $[E_\alpha,E_{-\alpha}]=t_\alpha\in t$ and $(\rho',t_\alpha)=\alpha(\rho')$ (roots are elements of t^*). We have thus proven Theorem 2.

Theorem 2: An orbit $\widehat{O}(\rho)$ is symplectic if and only if ρ is an element of the Cartan subalgebra of su(N) and satisfy the following condition: if $\alpha(\rho)=0$ then $E_\alpha\rho=\rho E_\alpha=0$ for every root α.

III. KÄHLER MANIFOLD

Now we find which symplectic orbits $\widehat{O}(\rho)$ described by Theorem 2 are Kähler manifolds. It is well known that if G is a connected complex Lie group then the only Kähler homogeneous space for G is of the form G/P, where P is a parabolic subgroup of G. Let us reiterate that a Borel subgroup of G is a maximal complex solvable Lie subgroup and a parabolic subgroup is a complex Lie subgroup containing a Borel subgroup. A Borel subalgebra of $\mathfrak{g}^c$ (standard relative to t) is of the form

$$b=t+\bigoplus_{\alpha>0}CE_\alpha\,. \tag{9}$$

We are looking for states that are stabilized by a Borel subgroup $B=\exp b$. Let $\hat\rho\in\widehat{\mathscr{P}}$ be such a state. Thus for every $\alpha>0$ and $s\in R$

$$e^{sE_\alpha}\hat\rho e^{sE_{-\alpha}}=\hat\rho \tag{10}$$

[we must define this action on $\widehat{\mathscr{P}}$ because the action of G^c does not conserve I in gl(N,C)]. Differentiating (10) with respect to s and putting $s=0$ we obtain

$$\forall_{\alpha>0}\quad E_\alpha\hat\rho+\hat\rho E_{-\alpha}=0\,. \tag{11}$$

The Lie algebra su$(N)^c$ is of the type A_{N-1} (Ref. 11) and for any $i,j=1,2,\ldots,N$ vector E_{ij} such that $(E_{ij})_{kl}=\delta_{ik}\,\delta_{jl}$ is a root vector for su$(N)^c$. A root α_{ij} is positive if $i<j$. Root vectors for $\mathfrak{g}^c$ have the following form: $E_\alpha=\Sigma_{(ij)}\,\lambda_{ij}^{(\alpha)}E_{ij}$ where $\lambda_{ij}^{(\alpha)}\in$R. A root α is positive if in this sum all pairs (ij) have the property $i<j$. Let $\hat\rho=\Sigma_{k=1}^N\lambda_kE_{kk}$ and α be a positive root of $\mathfrak{g}^c$. From (11) we obtain

$$\sum_{(ij)}\sum_{k=1}^{N}\lambda_{ij}^{(\alpha)}\lambda_k(E_{ij}E_{kk}+E_{kk}E_{ji})=0\,. \tag{12}$$

Since $E_{ij}E_{kl}=\delta_{jk}E_{il}$ thus

$$\sum_{(ij)}\lambda_{ij}^{(\alpha)}\lambda_j(E_{ij}+E_{ji})=0 \tag{13}$$

and hence $\lambda_{ij}^{(\alpha)}\,\lambda_j=0$. Since representation of G is irreducible there exists a positive root α and a number $i<j$ such that $\lambda_{ij}^{(\alpha)}\neq0$. Thus $\lambda_j=0$ for any $j>1$ and because $\mathrm{Tr}\,\hat\rho=1$ there must be $\lambda_1=1$. So $\hat\rho=E_{11}=|v_{\max}\rangle\,\langle v_{\max}|$, where $|v_{\max}\rangle$ is the maximal weight vector.

Theorem 3: There is only one orbit that is Kähler and that is the orbit passing through $\hat\rho=|v_{\max}\rangle\,\langle v_{\max}|$.

The Kähler orbit has a clear physical interpretation. This is an orbit of so-called coherent states.[12]

IV. EXAMPLE

Let us consider $G=\mathrm{SU}(2)$, which plays a fundamental role in control theory. For any N there exists irreducible unitary representation of SU(2) in a Hilbert space C^N [so-called spin-$\frac12(N-1)$ representation]. Let us take for simplicity $N=3$. Since SU(2) has only one positive root α, we have

$$E_\alpha=\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}\;;\;E_{-\alpha}=\begin{pmatrix}0&0&0\\1&0&0\\0&1&0\end{pmatrix}\,. \tag{14}$$

The Cartan subalgebra t is spanned by

$$[E_\alpha,E_{-\alpha}]=t_\alpha=\begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix}\,. \tag{15}$$

Let $\rho\in h_t$ so $\rho=\Sigma_{k=1}^3\lambda_kE_{kk}$, where $\Sigma_{k=1}^3\lambda_k=0$:

$$\alpha(\rho)=(\rho,t_\alpha)=\mathrm{Tr}(\rho t_\alpha)=\lambda_1-\lambda_3\,. \tag{16}$$

If $\alpha(\rho)=0$ then

$$\rho=\begin{pmatrix}\lambda&0&0\\0&-2\lambda&0\\0&0&\lambda\end{pmatrix}. \tag{17}$$

But $E_\alpha\rho\neq0$ for $\lambda\neq0$ and hence $\widehat{O}(\rho)$ is not symplectic. In particular, for $\lambda=-\frac13$, $\hat\rho=|v\rangle\,\langle v|$ where $\langle v|=(0,1,0)$.

ACKNOWLEDGMENT

The author is grateful to Professor A. Jamiolkowski for helpful discussions.

[1]B. Kostant, *Quantization and Unitary Representations*, Lecture Notes in Math. Vol. 170 (Springer, New York, 1970).

[2]A. A. Kirillov, *Elements of the Theory of Representations* (Springer, New York, 1976).

[3]J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).

[4]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U. P., Cambridge, 1984).

[5]D. Chruściński, "Symplectic Structure of the von-Neumann Equation," **29** (1990), Rep. Math. Phys. **29** (1990).

[6]B. Kostant and S. Sternberg, ' — Symplectic Projective Orbits," in *New Directions in Applied Mathematics* (Springer, New York, 1982).

[7]G. Rosensteel, Phys. Rev. A **23** 2794, (1981).

[8]G. Rosensteel and D. J. Rowe, Phys. Rev. A **24** 673, (1981).

[9]D. J. Rowe, *Many-Body Quantum Mechanics as Symplectic Dynamics*, Lecture Notes in Phys. Vol. 135 (Springer, New York, 1980).

[10]P. Kramer and M. Saraceno, *Geometry of the Time-dependent Variational Principle in Quantum Mechanics*, Lecture Notes in Phys. Vol. 140 (Springer, New York, 1981).

[11]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Warszawa, 1980).

[12]A. M. Perelemov, Comm. Math. Phys. **26**, 222 (1972).

Relationship between S_N and U(n) isoscalar factors and higher-order U(n) invariants

Xiangzhu Li[a] and Josef Paldus[b),c)]
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

(Received 28 July 1989; accepted for publication 14 February 1990)

General relationships expressing U(n) and S_N coupling and transformation isoscalar factors in terms of U(n) Racah and 9λ coefficients are derived. The absolute values of U(n) Racah coefficients involving at most k-column irreducible representations are shown to be identical with SU(k) Racah coefficients. In particular, an explicit relationship is established between the U(n) and SU(2) approaches to the many-electron correlation problem.

I. INTRODUCTION

The unitary group Racah–Wigner algebras[1-5] have been widely exploited in many different fields ranging from particle and nuclear physics to molecular and solid state physics and chemistry. In quantum chemical applications, the so-called unitary group approach (UGA)[6-15] has been primarily exploited in large scale electronic structure calculations based on the molecular orbital model (cf. Refs. 11–15 and references therein). This approach relies heavily on the representation theory of unitary groups, or rather, of their Lie algebras, since the electronic spin-independent Hamiltonian for an n-orbital model of atomic and molecular systems can be expressed in terms of U(n) generators and the corresponding exact wave function can be expanded in terms of canonical basis vectors spanning the carrier spaces for U(n) irreducible representations (irreps), (e.g., Refs. 6 and 16). Consequently, the basic parameters arising in the representation theory, such as matrix representatives of generators, Clebsch–Gordan (CG) coefficients, isoscalar factors, or reduced Wigner coefficients as well as higher-order invariants, such as the U(n) analogs of Racah coefficients and $9j$ symbols (cf., e.g., Refs. 17–19), play the central role in such applications. Although all such invariants appear implicitly in various exploitations of the UGA formalism, they have not always been recognized as such, particularly when various *ad hoc* procedures were developed for specific algorithms used in atomic and molecular electronic structure computations.

At this point, we must recall that there exists a very close relationship between the representation theory of the unitary and symmetric groups. In applications to many-electron systems, where at most two-column irreps are involved, there is an additional intimate relationship of both the orbital group U(n) and the particle symmetry group S_N irreps with those of the angular momentum or spin group SU(2). Indeed, these interrelationships enabled a useful cross fertilization in the development of various aspects of UGA formalism and stem from a radically different yet closely related viewpoints offered by these three groups. For example, Shavitt's graphical representation of UGA bases,[8,11,12,15] which naturally derives from the ABC (or Paldus) tableau formalism, suggested an analogous representation in the S_N-based approaches.[20,21] Similarly, the necessary segment values for an efficient evaluation of matrix elements of U(n) generator products were first derived[9] by exploiting SU(2) graphical methods of spin algebras,[22] following an earlier exploitation of this technique within the context of the S_N-based formalism.[23]. The approach based on Green–Gould[24] representation theory for unitary and orthogonal groups also relies heavily on various concepts of tensor representations and proceeds via evaluation of reduced Wigner coefficients.[25,26]

In our recent papers,[27,28] we have elaborated explicitly the implications of the S_N–U(n) duality, and derived rank n-independent expressions for U(n) isoscalar factors and Racah coefficients by exploiting the symmetric group based formalism.[29] Since the matrix elements of any tensor operator are given by a product of a reduced matrix element and a CG coefficient, the latter being in turn expressible as a product of isoscalar factors, the matrix element segmentation that is widely exploited in UGA (e.g., Refs. 8–15) represents, in fact, a particular case of Racah factorization with relevant segment values being given by properly scaled U(n) isoscalar factors.[30] We also note in this context that Kent and Schlesinger[31] have recently exploited these ideas for general multicolumn U(n) irreps, expressing the generator matrix elements through higher-order Racah coefficients and U(n) $3n-j$ symbols, although the explicit expressions for these quantities remain to be worked out. Very recently a new progress in the development of the U(n) Racah–Wigner algebra[5] was achieved with the help of the vector coherent state theory,[32-35] which enabled to establish the relations between certain classes of reduced Wigner coefficients (or projective operators) and $6j$ and $9j$ symbols.

It is thus clear that there exists a close relationship between various approaches that are employed in the derivation of required group theoretical invariants. It is the purpose of this paper to elucidate the relationship between the isoscalar factors and higher-order U(n) invariants as well as between the U(n) quantities and SU(2) Racah coefficients in case of many-electron systems. In Sec. II, we present a simple and unified formulation of the relationship between the U(n) analogues of Racah and $9j$ symbols and U(n) and

[a] Permanent address: Department of Chemistry, Xiamen University, Xiamen, Fujian, People's Republic of China.
[b] Also at: Department of Chemistry and Guelph-Waterloo Center for Graduate Work in Chemistry, Waterloo Campus, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
[c] Killam Research Fellow 1987–9.

 0022-2488/90/071589-11$03.00

S_N isoscalar factors. In Sec. III, we consider many-electron systems, in which case a correspondence with SU(2) invariants can be established. It is believed that a clarification of these relationships will provide us with a better understanding of U(n) tensor operator calculus and will enable its more versatile exploitation in more general situations when spin-dependent interactions are present or particle number is not conserved.

II. ISOSCALAR FACTORS AND HIGHER-ORDER U(n) INVARIANTS

A. General relationships

The irreps of U(n) can be labeled by Young diagrams $\langle\lambda\rangle$, or simply λ, having an arbitrary (finite) number of boxes N that we can interpret as a particle number N and that are arranged in at most n rows, being otherwise n independent. The same diagrams will also label the irreps of the symmetric group S_N, in which case we designate them by $[\lambda]$ in order to distinguish them from the U(n) irreps $\langle\lambda\rangle$ or λ.

Let us first recall that we can define two basic types of isoscalar factors for U(n).[27]

(i) The *standard* isoscalar (or coupling) factors, also referred to as *reduced Wigner coefficients*, that are associated with the canonical subgroup chain $U(m)\supset U(m-1)$, $(m<n)$. These factors are associated with the coupling that leads to Gel'fand–Tsetlin states[36] and we designate them as the I_u factors,[27,28]

$$I_u\begin{pmatrix}\lambda & \mu & \alpha\nu\\ \lambda' & \mu' & \alpha'\nu'\end{pmatrix}\equiv\left(\begin{matrix}\langle\lambda\rangle & \langle\mu\rangle\\ \langle\lambda'\rangle & \langle\mu'\rangle\end{matrix}\middle|\begin{matrix}\alpha\langle\nu\rangle\\ \alpha'\langle\nu'\rangle\end{matrix}\right)$$
$$\equiv\left(\begin{matrix}\lambda & \mu\\ \lambda' & \mu'\end{matrix}\middle|\begin{matrix}\alpha\nu\\ \alpha'\nu'\end{matrix}\right), \tag{1}$$

α, α' being the multiplicity labels, if necessary.

(ii) *Transformation* factors, designated as I_t factors,[27,28] that correspond to a basis transformation from Gel'fand–Tsetlin states to arbitrary partitioned states that are adapted to the subgroup chain $U(n_1+n_2)\supset U(n_1)\otimes U(n_2)$,

$$I_t\left(\lambda\ \begin{matrix}\mu\\ \mu'\end{matrix}\middle|\begin{matrix}\alpha\nu\\ \alpha'\nu'\end{matrix}\right)\equiv\left(\langle\lambda\rangle\ \begin{matrix}\langle\mu\rangle\\ \langle\mu'\rangle\end{matrix}\middle|\begin{matrix}\alpha\langle\nu\rangle\\ \alpha'\langle\nu'\rangle\end{matrix}\right)$$
$$\equiv\left(\lambda\ \begin{matrix}\mu\\ \mu'\end{matrix}\middle|\begin{matrix}\alpha\nu\\ \alpha'\nu'\end{matrix}\right). \tag{2}$$

A product of the I_u factors yields U(n) Clebsch–Gordan (CG) coefficients, while a product of the I_t factors gives subduction coefficients for the above given chain. The interrelationship between the I_u and I_t factors has been established in our recent study[28] and is given by the formula

$$I_t\left(\lambda\ \begin{matrix}\mu\\ \mu_1\end{matrix}\middle|\begin{matrix}\alpha\nu\\ \beta\nu_1\end{matrix}\right)=\left(\frac{H_\nu H_{\mu_1}}{H_{\nu_1}H_\mu}\right)^{1/2} I_u\left(\begin{matrix}\lambda & \mu\\ \lambda & \mu_1\end{matrix}\middle|\begin{matrix}\alpha\nu\\ \beta\nu_1\end{matrix}\right), \tag{3}$$

where α and β are multiplicity labels and H_λ designates the product of hook lengths for the Young diagram λ

$$H_\lambda=\prod_i h_i. \tag{4}$$

The latter determines the dimension f_λ of the irrep $[\lambda]$ of S_{N_λ},

$$f_\lambda=N_\lambda!/H_\lambda, \tag{5}$$

N_λ being the number of boxes in the Young diagram λ, so that we can also write that

$$H_\lambda=N_\lambda!/f_\lambda. \tag{5'}$$

Analogously to the above given $U(n)$ factors I_u and I_t, we can define the isoscalar factors I_o and I_s for the symmetric group S_N, namely:[27–30]

(iii) The *outer product* isoscalar factors I_o, which are associated with the coupling leading from the Young–Yamanouchi bases for S_{N_1} and S_{N_2} to that for $S_{N_1+N_2}$, and

(iv) the *subduction* isoscalar factors I_s, which enable a transformation from the standard S_N basis to the nonstandard basis adapted to the chain $S_N\supset S_{N_1}\otimes S_{N_2}$. The relationship between the I_o and I_s factors is similarly given by[27]

$$I_s\left([\lambda]\ \begin{matrix}[\mu]\\ [\mu_1]\end{matrix}\middle|\begin{matrix}\alpha[\nu]\\ \beta[\nu_1]\end{matrix}\right)\equiv\left([\lambda]\ \begin{matrix}[\mu]\\ [\mu_1]\end{matrix}\middle|\begin{matrix}\alpha[\nu]\\ \beta[\nu_1]\end{matrix}\right)$$
$$=\left[\frac{N_\nu f_{\nu_1} f_\mu}{N_\mu f_\nu f_{\mu_1}}\right]^{1/2}$$
$$\times I_o\left(\begin{matrix}[\lambda] & [\mu]\\ [\lambda] & [\mu_1]\end{matrix}\middle|\begin{matrix}\alpha[\nu]\\ \beta[\nu_1]\end{matrix}\right), \tag{6}$$

where $[\mu_1]$ and $[\nu_1]$ are obtained from $[\mu]$ and $[\nu]$ by a removal of a single box. In fact, the relationship given by Eq. (6) represents a special case of Eq. (3) that relates the U(n) isoscalar factors. The relationships between the I_o and I_u factors, and between the I_s and I_t factors, have also been established and can be found in Refs. 28 and 29. For the sake of simplicity, we drop in the following symbols I_u, I_t, I_o, and I_s, since we can easily recognize these factors by their structure and the irrep symbols used (λ or $\langle\lambda\rangle$ for U(n) and $[\lambda]$ for S_N).

A wide variety of notations is employed in the literature for the recoupling coefficients involving three, four, or more irreps. In the case of the SU(2) group, the most often exploited are the $6j$ and $9j$ (or, generally, $3n-j$) coefficients, which possess a very high symmetry in their arguments. However, for nonsimple reducible groups, as is the case of general U(n) groups, this advantage is lost since the mixed symmetry may result when high multiplicities are involved as shown by Derome[37] (cf. also Ref. 29). In such mixed symmetry cases, it is impossible to define a $3j$ symbol that changes only its phase under an arbitrary permutation of its columns. Consequently, the symmetry properties of higher-order invariants, such as the U(n) analogs of the $6j$ and $9j$ symbols, may be rather involved. In view of this fact we prefer to investigate primitive recoupling matrices of lower symmetry, thus avoiding an unnecessarily complex notation, even though the mixed symmetry is nowadays fairly well understood for CG coupling coefficients, and in some cases[38] even for Racah coefficients. We thus adopt the following notation for the U(n) Racah coefficients and $9j$ symbols:

$$U(\lambda_1\lambda_2\lambda\lambda_3;\lambda_{12}\lambda_{23})_{\alpha_{12}\alpha_{23}\alpha\alpha'}$$
$$:=\langle(\lambda_1\lambda_2)\alpha_{12}\lambda_{12},\lambda_3;\alpha\lambda\,|\lambda_1,(\lambda_2\lambda_3)\alpha_{23}\lambda_{23};\alpha'\lambda\rangle \quad (7)$$
$$X(\lambda_1\lambda_2\lambda_3\lambda_4;\lambda_{12}\lambda_{34};\lambda_{13}\lambda_{24};\lambda)_{\alpha_{12}\alpha_{34}\alpha_{13}\alpha_{24}\alpha\alpha'}$$
$$:=\langle(\lambda_1\lambda_2)\alpha_{12}\lambda_{12},(\lambda_3\lambda_4)\alpha_{34}\lambda_{34};\ \alpha\lambda\,|(\lambda_1\lambda_3)\alpha_{13}\lambda_{13},(\lambda_2\lambda_4)\alpha_{24}\lambda_{24};\alpha'\lambda\rangle. \quad (8)$$

In the SU(2) case, this notation was employed by Jahn.[39]

It is well known that Racah coefficients can be expressed as a sum of products of four CG coefficients. Using the orthogonality properties of the latter, we easily find that [cf., e.g., Eq. (139) of Ref. 28; for a more general form, see Ref. 4]

$$\sum_{\alpha'} U(\lambda_1\lambda_2\lambda\lambda_3;\lambda_{12}\lambda_{23})_{\alpha_{12}\alpha_{23}\alpha\alpha'}\left\langle\begin{matrix}\alpha'\lambda\\ W\end{matrix}\middle|\begin{matrix}\lambda_1 & \lambda_{23}\\ W_1 & W_{23}\end{matrix}\right\rangle$$
$$=\sum_{W_2,W_3,W_{12}}\left\langle\begin{matrix}\lambda_1 & \lambda_2\\ W_1 & W_2\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ W_{12}\end{matrix}\right\rangle\left\langle\begin{matrix}\lambda_{12} & \lambda_3\\ W_{12} & W_3\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ W\end{matrix}\right\rangle\times\left\langle\begin{matrix}\lambda_2 & \lambda_3\\ W_2 & W_3\end{matrix}\middle|\begin{matrix}\alpha_{23}\lambda_{23}\\ W_{23}\end{matrix}\right\rangle, \quad (9)$$

an expression that involves fewer terms in the summation on the rhs than the defining relationship having all four CG coefficients on the rhs. Expressing next each CG coefficient as a product of an isoscalar factor for $U(n)\supset U(n-1)$ and a CG coefficient for $U(n-1)$, i.e.,

$$\left\langle\begin{matrix}\lambda_{12} & \lambda_3\\ W_{12} & W_3\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ W\end{matrix}\right\rangle=\sum_{\beta}\left(\begin{matrix}\lambda_{12} & \lambda_3\\ \mu_{12} & \mu_3\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ \beta\mu\end{matrix}\right)\left\langle\begin{matrix}\mu_{12} & \mu_3\\ \tilde{W}_{12} & \tilde{W}_3\end{matrix}\middle|\begin{matrix}\beta\mu\\ \tilde{W}\end{matrix}\right\rangle, \quad (10)$$

where λ_i's and μ_i's are irreps of $U(n)$ and $U(n-1)$, respectively, while W_i's and $\tilde{W}_i$'s are relevant Weyl tableaux labeling the corresponding $U(n)$ and $U(n-1)$ states, we obtain the following recursion formula for the Racah coefficients involving the $U(n)\supset U(n-1)$ isoscalar factors:

$$\sum_{\alpha'} U(\lambda_1\lambda_2\lambda\lambda_3;\lambda_{12}\lambda_{23})_{\alpha_{12}\alpha_{23}\alpha\alpha'}\left(\begin{matrix}\alpha'\lambda\\ \beta'\mu\end{matrix}\middle|\begin{matrix}\lambda_1 & \lambda_{23}\\ \mu_1 & \mu_{23}\end{matrix}\right)$$
$$=\sum_{\mu_2,\mu_3,\mu_{12}}\ \sum_{\beta_{12},\beta_{23},\beta} U(\mu_1\mu_2\mu\mu_3;\mu_{12}\mu_{23})_{\beta_{12}\beta_{23}\beta\beta'}\times\left(\begin{matrix}\lambda_1 & \lambda_2\\ \mu_1 & \mu_2\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ \beta_{12}\mu_{12}\end{matrix}\right)\left(\begin{matrix}\lambda_{12} & \lambda_3\\ \mu_{12} & \mu_3\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ \beta\mu\end{matrix}\right)\times\left(\begin{matrix}\lambda_2 & \lambda_3\\ \mu_2 & \mu_3\end{matrix}\middle|\begin{matrix}\alpha_{23}\lambda_{23}\\ \beta_{23}\mu_{23}\end{matrix}\right). \quad (11)$$

This formula will be used to derive a simple relationship between Racah coefficients and isoscalar factors.

The same procedure can be extended to higher-order invariants. For 9λ symbols [or X coefficients of Eq. (8)], which can be similarly expressed in terms of six CG coefficients, we find

$$\sum_{\alpha'} X(\lambda_1\lambda_2\lambda_3\lambda_4;\lambda_{12}\lambda_{34};\lambda_{13}\lambda_{24};\lambda)_{\alpha_{12}\alpha_{34}\alpha_{13}\alpha_{24}\alpha\alpha'}\left\langle\begin{matrix}\alpha'\lambda\\ W\end{matrix}\middle|\begin{matrix}\lambda_{13} & \lambda_{24}\\ W_{13} & W_{24}\end{matrix}\right\rangle$$
$$=\sum_{W_1,W_2,W_3,W_4}\ \sum_{W_{12},W_{34}}\left\langle\begin{matrix}\lambda_1 & \lambda_2\\ W_1 & W_2\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ W_{12}\end{matrix}\right\rangle\left\langle\begin{matrix}\lambda_3 & \lambda_4\\ W_3 & W_4\end{matrix}\middle|\begin{matrix}\alpha_{34}\lambda_{34}\\ W_{34}\end{matrix}\right\rangle\left\langle\begin{matrix}\lambda_{12} & \lambda_{34}\\ W_{12} & W_{34}\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ W\end{matrix}\right\rangle\left\langle\begin{matrix}\lambda_1 & \lambda_3\\ W_1 & W_3\end{matrix}\middle|\begin{matrix}\alpha_{13}\lambda_{13}\\ W_{13}\end{matrix}\right\rangle\left\langle\begin{matrix}\lambda_2 & \lambda_4\\ W_2 & W_4\end{matrix}\middle|\begin{matrix}\alpha_{24}\lambda_{24}\\ W_{24}\end{matrix}\right\rangle, \quad (12)$$

as well as a corresponding recursion formula

$$\sum_{\alpha'} X(\lambda_1\lambda_2\lambda_3\lambda_4;\lambda_{12}\lambda_{34};\lambda_{13}\lambda_{24};\lambda)_{\alpha_{12}\alpha_{34}\alpha_{13}\alpha_{24}\alpha\alpha'}\left(\begin{matrix}\alpha'\lambda\\ \beta'\mu\end{matrix}\middle|\begin{matrix}\lambda_{13} & \lambda_{24}\\ \mu_{13} & \mu_{24}\end{matrix}\right)$$
$$=\sum_{\mu_1,\mu_2,\mu_3,\mu_4,\mu_{12},\mu_{34}}\ \sum_{\beta_{12},\beta_{34},\beta_{13},\beta_{24},\beta} X(\mu_1\mu_2\mu_3\mu_4;\mu_{12}\mu_{34};\mu_{13}\mu_{24};\mu)_{\beta_{12}\beta_{34}\beta_{13}\beta_{24}\beta\beta'}$$
$$\times\left(\begin{matrix}\lambda_1 & \lambda_2\\ \mu_1 & \mu_2\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ \beta_{12}\mu_{12}\end{matrix}\right)\left(\begin{matrix}\lambda_3 & \lambda_4\\ \mu_3 & \mu_4\end{matrix}\middle|\begin{matrix}\alpha_{34}\lambda_{34}\\ \beta_{34}\mu_{34}\end{matrix}\right)\left(\begin{matrix}\lambda_{12} & \lambda_{34}\\ \mu_{12} & \mu_{34}\end{matrix}\middle|\begin{matrix}\alpha\lambda\\ \beta\mu\end{matrix}\right)\left(\begin{matrix}\lambda_1 & \lambda_3\\ \mu_1 & \mu_3\end{matrix}\middle|\begin{matrix}\alpha_{13}\lambda_{13}\\ \beta_{13}\mu_{13}\end{matrix}\right)\left(\begin{matrix}\lambda_2 & \lambda_4\\ \mu_2 & \mu_4\end{matrix}\middle|\begin{matrix}\alpha_{24}\lambda_{24}\\ \beta_{24}\mu_{24}\end{matrix}\right). \quad (13)$$

Racah coefficients and 9λ symbols possess many useful properties just as in the SU(2) case. We note, in particular, their orthogonality and symmetry properties as well as their reduction to simpler quantities, when one or more of the irreps involved is a trivial scalar irrep. Some of these properties that will be useful to us later are summarized below.

(i) *Symmetry properties.* Using the symmetry of CG coefficients relative to the interchange of coupled irreps [cf. Eq. (136) of Ref. 28],

$$\left\langle\begin{matrix}\lambda_1 & \lambda_2\\ W_1 & W_2\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ W_{12}\end{matrix}\right\rangle=(-1)^{\lambda_1+\lambda_2+\lambda_{12}}\Theta(\lambda_1,\lambda_2,\lambda_{12})_{\alpha_{12}}\times\left\langle\begin{matrix}\lambda_2 & \lambda_1\\ W_2 & W_1\end{matrix}\middle|\begin{matrix}\alpha_{12}\lambda_{12}\\ W_{12}\end{matrix}\right\rangle, \quad (14)$$

with the phase factor $\Theta(\lambda_1,\lambda_2,\lambda_{12})_{\alpha_{12}}$ equal to 1 in case of unique multiplicity, we get for the X-coefficients that

$$X(\lambda_1\lambda_2\lambda_3\lambda_4;\lambda_{12}\lambda_{34};\lambda_{13}\lambda_{24};\lambda)_{\alpha_{12}\alpha_{34}\alpha_{13}\alpha_{24}\alpha\alpha'}$$
$$=\phi\cdot\Theta(\lambda_1\lambda_2\lambda_{12})_{\alpha_{12}}\cdot\Theta(\lambda_3\lambda_4\lambda_{34})_{\alpha_{34}}\cdot\Theta(\lambda_{13}\lambda_{24}\lambda)_{\alpha'}\cdot X(\lambda_2\lambda_1\lambda_4\lambda_3;\lambda_{12}\lambda_{34};\lambda_{24}\lambda_{13};\lambda)_{\alpha_{12}\alpha_{34}\alpha_{24}\alpha_{13}\alpha\alpha'}$$
$$=\phi\cdot\Theta(\lambda_1\lambda_3\lambda_{13})_{\alpha_{13}}\cdot\Theta(\lambda_2\lambda_4\lambda_{24})_{\alpha_{24}}\cdot\Theta(\lambda_{12}\lambda_{34}\lambda)_{\alpha}\cdot X(\lambda_3\lambda_4\lambda_1\lambda_2;\lambda_{34}\lambda_{12};\lambda_{13}\lambda_{24};\lambda)_{\alpha_{34}\alpha_{12}\alpha_{13}\alpha_{24}\alpha\alpha'}, \quad (15)$$

where

$$\phi = (-1)^{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_{12}+\lambda_{34}+\lambda_{13}+\lambda_{24}+\lambda}. \quad (16)$$

We do not consider permutational symmetry involving the third column in CG coefficients, Eq. (14), which can involve a mixed symmetry. We stress, however, that Eq. (14) holds even when the mixed symmetry does occur. We also recall similar properties for the U coefficients, given by Eqs. (134-7) of Ref. 28.

(ii) *Reduction* of the X coefficients when one or more irreps are trivial scalar irrep $\langle 0\rangle$. We list only independent cases, the remaining ones follow by applying the above given symmetry rules,

$$X(\lambda_1 0\lambda_3\lambda_4;\lambda_1\lambda_{34};\lambda_{13}\lambda_4;\lambda)_{-\alpha_{34}\alpha_{13}-\alpha\alpha'} = U(\lambda_1\lambda_3\lambda\lambda_4;\lambda_{13}\lambda_{34})_{\alpha_{13}\alpha_{34}\alpha'\alpha}, \quad (17)$$

$$\begin{aligned} X(00\lambda_3\lambda_4;0\lambda_{34};\lambda_3\lambda_4;\lambda_{34})_{-\alpha_{34}---\alpha_{34}} &= X(0\lambda_2 0\lambda_4;\lambda_2\lambda_4;0\lambda_{24};\lambda_{24})_{---\alpha_{24}\alpha_{24}-} \\ &= X(\lambda_1 00\lambda_4;\lambda_1\lambda_4;\lambda_1\lambda_4;\lambda_{14})_{----\alpha_{14}\alpha_{14}} = 1. \end{aligned} \quad (18)$$

Note that when $\lambda_1 = \lambda_4 = \langle 0\rangle$, the X coefficient is given by the phase factor appearing in Eq. (15). Finally, when any three of the irreps λ_i, $i = 1,\dots 4$ are trivial irreps $\langle 0\rangle$, then $X = 1$. We also note that we can use the reduction of the X coefficients to Racah coefficients, Eq. (17), when one of the irreps is $\langle 0\rangle$, to find corresponding symmetry and reduction properties for the latter, by exploiting Eqs. (14)–(18).

B. Relationship between S_N isoscalar factors and U(n) Racah coefficients

We shall first derive a useful relationship between Racah coefficients and symmetric group isoscalar factors. We thus consider U(n) isoscalar factors that are associated with the so-called special Gel'fand states,[40] in which all single particle states are singly occupied. Such isoscalar factors are then identical with those for the symmetric group S_N when using standard Young–Yamanouchi basis.

We employ Eq. (11) for the special case when $\langle\lambda_1\rangle = \langle 1\rangle$ and $\langle\mu_1\rangle = \langle 0\rangle$, so that $\mu = \mu_{23}$ and we set $\lambda_{23} = \mu_{23} = \nu'$. Relabeling other irreps to simplify the notation, we thus get

$$\begin{aligned} &U(\langle 1\rangle\mu'\nu\lambda;\mu\nu')_{-\alpha_{23}\alpha-}\begin{pmatrix}\nu & \langle 1\rangle & \nu' \\ \nu' & \langle 0\rangle & \nu'\end{pmatrix} \\ &= \sum_{\mu_2,\mu_3,\mu_{12}}\sum_{\beta_{23},\beta} U(\langle 0\rangle\mu_2\nu'\mu_3;\mu_{12}\nu')_{-\beta_{23}\beta-} \\ &\times\begin{pmatrix}\langle 1\rangle & \mu' & \mu \\ \langle 0\rangle & \mu_2 & \mu_{12}\end{pmatrix}\begin{pmatrix}\mu & \lambda & \alpha\nu \\ \mu_{12} & \mu_3 & \beta\nu'\end{pmatrix}\begin{pmatrix}\mu' & \lambda & \alpha_{23}\nu' \\ \mu_2 & \mu_3 & \beta_{23}\nu'\end{pmatrix}. \end{aligned} \quad (19)$$

The isoscalar factors on the rhs imply that we must have $\mu_2 = \mu_{12}$ and thus $\beta = \beta_{23}$, so that also $\mu_2 = \mu'$, $\mu_3 = \lambda$ and $\alpha_{23} = \beta_{23}$ in view of the last factor in Eq. (19). We thus get

$$\begin{aligned} &U(\langle 1\rangle\mu'\nu\lambda;\mu\nu')_{-\beta\alpha-}\begin{pmatrix}\nu & \langle 1\rangle & \nu' \\ \nu' & \langle 0\rangle & \nu'\end{pmatrix} \\ &= \begin{pmatrix}\langle 1\rangle & \mu' & \mu \\ \langle 0\rangle & \mu' & \mu'\end{pmatrix}\begin{pmatrix}\mu & \lambda & \alpha\nu \\ \mu' & \lambda & \beta\nu'\end{pmatrix}, \end{aligned} \quad (20)$$

since the Racah coefficient and the last isoscalar factor on the rhs in Eq. (19) reduce to 1 [cf. Eq. (138) of Ref. 28]. The remaining isoscalar factors that involve irreps $\langle 1\rangle$ and $\langle 0\rangle$ are easily evaluated [cf. Eq. (145) of Ref. 28]

$$\begin{pmatrix}\lambda & \langle 1\rangle & \nu \\ \lambda & \langle 0\rangle & \lambda\end{pmatrix} = \left(\frac{f_\nu}{N_\nu f_\lambda}\right)^{1/2}, \quad (21)$$

so that

$$\begin{aligned} &U(\langle 1\rangle\langle\mu'\rangle\langle\nu\rangle\langle\lambda\rangle;\langle\mu\rangle\langle\nu'\rangle)_{-\beta\alpha-} \\ &= \left[\frac{N_\nu f_\nu f_\mu}{N_\mu f_\nu f_{\mu'}}\right]^{1/2}\begin{pmatrix}\langle\lambda\rangle & \langle\mu\rangle & \alpha\langle\nu\rangle \\ \langle\lambda\rangle & \langle\mu'\rangle & \beta\langle\nu'\rangle\end{pmatrix} \\ &= U(\langle\lambda\rangle\langle\mu'\rangle\langle\nu\rangle\langle 1\rangle;\langle\nu'\rangle\langle\mu\rangle)_{\beta--\alpha}, \end{aligned} \quad (22)$$

where we used the symmetry relation, Eq. (137) of Ref. 28, in the last equation. The resulting relationship generalizes that of Eq. (148) of Ref. 28. It is important to note the following.

(1) Only one box is involved in subductions from $\langle\nu\rangle$ to $\langle\nu'\rangle$ and from $\langle\mu\rangle$ to $\langle\mu'\rangle$, so that

$$N_\mu = N_{\mu'} + 1, \quad N_\nu = N_{\nu'} + 1. \quad (23)$$

Consequently, the isoscalar factor appearing in Eq. (22) corresponds to a fundamental shift

$$\langle\nu'\rangle = \langle\nu\rangle - \langle\Delta(\tau)\rangle, \quad (24)$$

where

$$\langle\Delta(\tau)\rangle = \langle\dot{0}1\dot{0}\rangle \quad (25)$$

with 1 occurring in the τ's place.

(2) The I_u factor appearing in Eq. (22) is equivalent with the corresponding I_o factor for S_N, namely,

$$\begin{pmatrix}\langle\lambda\rangle & \langle\mu\rangle & \alpha\langle\nu\rangle \\ \langle\lambda\rangle & \langle\mu'\rangle & \beta\langle\nu'\rangle\end{pmatrix} = \begin{pmatrix}[\lambda] & [\mu] & \alpha[\nu] \\ [\lambda] & [\mu'] & \beta[\nu']\end{pmatrix}. \quad (26)$$

We thus find that this I_o factor is equal to a U(n) Racah coefficient times a simple factor depending on the dimension and particle number of the S_N irreps involved.

(3) The bottom row labels of the isoscalar factor in Eq. (22), i.e., λ, μ', and ν', are of course the irreps of U($n-1$) or S_{N-1}. However, they must be regarded as the U(n) or U($n-1$) irreps in the Racah coefficient in Eq. (22). [Note that a U($n-1$) irrep can always be regarded as a U(n) irrep but not conversely.] This reflects a basic fact that the Young diagram labeling of U(n) irreps is n independent as long as n equals or exceeds the number of rows in the Young diagram.

Exploiting, finally, the relationship between the I_o and I_s isoscalar factors for S_N, Eq. (6), we can rewrite Eq. (22) in the form

$$\begin{aligned} &\begin{pmatrix}[\lambda] & [\mu] & \alpha[\nu] \\ & [\mu'] & \beta[\nu']\end{pmatrix} \\ &= U(\langle\lambda\rangle\langle\mu'\rangle\langle\nu\rangle\langle 1\rangle;\langle\nu'\rangle\langle\mu\rangle)_{\beta--\alpha}, \end{aligned} \quad (27)$$

with $N_\mu = N_{\mu'} + 1$ and $N_\nu = N_{\nu'} + 1$. Thus a general S_N subduction factor [i.e., a special transformation factor for U(n)] is identical with a special U(n) Racah coefficient, in which one of the six irreps equals to the fundamental irrep $\langle 1\rangle$.

C. General relationship between isoscalar factors and higher-order U(n) invariants

Consider next a more general case when more than one box is removed from a given irrep of U(n). We recall that in a general I_u factor

$$\begin{pmatrix} \lambda & \mu & \nu \\ \lambda_1 & \mu_1 & \nu_1 \end{pmatrix},$$

we remove f_1, f_2, and $f(=f_1+f_2)$ boxes from the Young diagrams representing the irreps λ, μ, and ν, respectively, obtaining the irreps λ_1, μ_1, and ν_1. Likewise, a general I_t factor has the form

$$\left(\lambda \quad \begin{matrix} \mu \\ \mu_1 \end{matrix} \middle| \begin{matrix} \alpha\nu \\ \beta\nu_1 \end{matrix}\right),$$

with f boxes removed from both μ and ν irrep labels. We shall consider the case in which ν is obtained by coupling $\langle\nu_1\rangle$ with the symmetric representation $\langle f\rangle$, i.e., $\langle\nu\rangle$ results by adding of f boxes from $\langle f\rangle$ to $\langle\nu_1\rangle$. The Littlewood–Richardson rule then requires that no more than one box is added to any column. Invoking, thus, Eq. (13) and setting (we drop again the angular brackets)

$$\lambda_2=f_1,\quad \lambda_4=f_2,\quad \lambda_{24}=f=f_1+f_2, \tag{28}$$

while

$$\mu_{13}=\lambda_{13}=\mu \text{ and } \mu_{24}=0, \tag{29}$$

we get

$$X(\lambda_1 f_1 \lambda_3 f_2;\lambda_{12}\lambda_{34};\lambda_{13} f;\lambda)_{--\alpha_{13}-\alpha-}\begin{pmatrix} \lambda & \lambda_{13} & f \\ \lambda_{13} & \lambda_{13} & 0 \end{pmatrix}$$
$$=\begin{pmatrix} \lambda_1 & f_1 & \lambda_{12} \\ \lambda_1 & 0 & \lambda_1 \end{pmatrix}\begin{pmatrix} \lambda_3 & f_2 & \lambda_{34} \\ \lambda_3 & 0 & \lambda_3 \end{pmatrix}\begin{pmatrix} \lambda_{12} & \lambda_{34} & \alpha\lambda \\ \lambda_1 & \lambda_3 & \alpha_{13}\lambda_{13} \end{pmatrix}, \tag{30}$$

since $X(\lambda_1 0\lambda_3 0;\lambda_1\lambda_3;\lambda_{13}0;\lambda_{13})=1$, and the last two isoscalar factors on the rhs of Eq. (13) also equal 1 implying that $\mu_1=\lambda_1$ and $\mu_3=\lambda_3$, $\mu_2=\mu_4=0$, and $\beta_{13}=\alpha_{13}=\beta$. The first three isoscalar factors appearing in the resulting Eq. (30) and involving the trivial scalar irrep $\langle 0\rangle$ are given by a simple expression [cf. Eq. (154) of Ref. 28],

$$\begin{pmatrix} \lambda & f & \nu \\ \lambda & 0 & \lambda \end{pmatrix}=\left(f!\frac{H_\lambda}{H_\nu}\right)^{1/2}, \tag{31}$$

with H_λ given by Eq. (5′), since $N_\lambda+f=N_\nu$. Exploiting this formula and relabeling the irreps to obtain a more symmetric form, we get

$$\begin{pmatrix} \lambda & \mu & \alpha\nu \\ \lambda_1 & \mu_1 & \beta\nu_1 \end{pmatrix}=\left[\frac{H_\lambda H_\mu H_{\nu_1}}{H_\nu H_{\lambda_1} H_{\mu_1}}\binom{f}{f_1}\right]^{1/2}$$
$$\times X(\lambda_1 f_1 \mu_1 f_2;\lambda\mu;\nu_1 f;\nu)_{--\beta-\alpha-}, \tag{32}$$

with

$$N_\nu=N_{\nu_1}+f,\quad N_\lambda=N_{\lambda_1}+f_1,\quad N_\mu=N_{\mu_1}+f_2,$$

and

$$f=f_1+f_2.$$

We thus find that a general isoscalar factor is equal to a special $9j$-type (i.e., 9λ) symbol times a factor involving dimensions and corresponding particle numbers [cf. Eq. (5′)] for the corresponding S_N irreps. Again, the resulting relationship, Eq. (32), is rank independent.

At this point, we must note that Eq. (32) represents a generalization of an earlier result by Committee Ališauskas, *et al.* [see Eq. (B1) of Ref. 41], which is easily obtained from our general result when one of the irreps is totally symmetric, e.g., $\langle\mu\rangle=\langle p\rangle$ and $\langle\mu_1\rangle=\langle q\rangle$ with $p-q=f_2$, so that

$$\begin{pmatrix} \lambda & p & \nu \\ \lambda_1 & q & \nu_1 \end{pmatrix}=\left[\frac{f_\nu f_{\lambda_1} N_\lambda! N_{\nu_1}! p! f!}{f_\lambda f_{\nu_1} N_\nu! N_{\lambda_1}! q! f_1! (p-q)!}\right]^{1/2}$$
$$\times X(\lambda_1 f_1 q(p-q);\lambda p;\nu_1 f;\nu)$$
$$=\left[\frac{H_\lambda H_{\nu_1} f!}{H_\nu H_{\lambda_1} f_1!}\binom{p}{q}\right]^{1/2}$$
$$\times X(\lambda_1 f_1 q(p-q);\lambda p;\nu_1 f;\nu). \tag{33}$$

Thus our general result shows that the relationship between an arbitrary U(n) isoscalar factor I_u and a special 9λ symbol, Eq. (32), has a universal validity and is not restricted to the multiplicity-free cases examined by Ališauskas *et al.*[41]

We can likewise generalize the relationship given by Eq. (22), setting $f_1=0$ in Eq. (32), so that $f_2=f$ and $\lambda_{12}=\lambda$, obtaining

$$\begin{pmatrix} \lambda & \mu & \alpha\nu \\ \lambda & \mu_1 & \beta\nu_1 \end{pmatrix}=\left[\frac{H_\mu H_{\nu_1}}{H_\nu H_{\mu_1}}\right]^{1/2} U(\lambda\mu_1\nu f;\nu_1\mu)_{\beta--\alpha}, \tag{34}$$

where we used Eq. (17), and where $N_\mu=N_{\mu_1}+f$ and $N_\nu=N_{\nu_1}+f$. The isoscalar factor on the lhs has now the same form as the I_t factor, so that we can exploit Eq. (3), obtaining

$$\left(\lambda \quad \begin{matrix} \mu \\ \mu_1 \end{matrix} \middle| \begin{matrix} \alpha\nu \\ \beta\nu_1 \end{matrix}\right)=U(\lambda\mu_1\nu f;\nu_1\mu)_{\beta--\alpha}. \tag{35}$$

This relationship generalizes that of Eq. (27) for the S_N isoscalar factor I_s to an arbitrary I_t factor for U(n). Thus a general U(n) I_t factor, that describes a subduction from the Gel'fand–Tsetlin basis to a partitioned $U(n_1+n_2)\supset U(n_1)\otimes U(n_2)$ basis, is identical with a special Racah coefficient involving the coupling of an f box totally symmetric irrep $\langle f\rangle$.

D. Racah and 9λ coefficients for many-electron systems

In our recent paper,[28] we have explored the U(n) tensor algebra that is relevant for many-electron systems. We derived explicit algebraic expressions for U(n) isoscalar factors that involve at most two-column irreps.[28,30] Exploiting the above given relationships, Eqs. (32) and (35), we can now obtain corresponding Racah and 9λ coefficients that arise in applications to many-electron systems. Thus, for example, Eq. (35) gives immediately that

$$U((a,b)(d-1,e)(s,t)(1,0);(s-1,t)(d,e))$$
$$=\left((a,b) \quad \begin{matrix} (d,e) \\ (d-1,e) \end{matrix} \middle| \begin{matrix} (s,t) \\ (s-1,t) \end{matrix}\right)=1, \tag{36}$$

where we employed the shorthand notation of Ref. 28 for two-column irreps, i.e., $(a,b) \equiv \langle 2^a 1^b \dot{0} \rangle$. In Eq. (36), the total particle number N equals $2a + b + 2d + e = 2s + t$. The other nonvanishing Racah coefficients and 9λ symbols are given in Tables I and II, respectively. Note that any column (row) of Table I is normalized, while the entire Table II is normalized, as required. Note also that all the absolute values of these coefficients depend only on the lengths b, e, and y of the single-column part of the Young diagrams involved, while the two-column length appears at most in the phase factors.

III. UNITARY GROUP FORMALISM FOR MANY-ELECTRON SYSTEMS AND ITS SU(2) COUNTERPART

It is well known that the S_N, U(n), and SU(2) descriptions of the many-electron correlation problem for spin-$\frac{1}{2}$ fermion n-orbital models are very closely related. Each formulation has its own merits and their interrelationship brings often a deeper insight into the problem, enabling a formulation of efficient algorithms for large scale numerical computations of accurate energies and wave functions of various models employed in studies of many-electron systems.[10–15]

In this section, we thus turn our attention to the two-column irreps of U(n) that characterize the so-called unitary group approach (UGA) to many-electron correlation problem.[6–15] These applications often involve unitary groups U(n) of a fairly large rank n, since a large number of atomic or molecular orbitals may be required for a reliable and accurate model description of such systems, particularly when the number of electrons N that are involved is large as well.

From the formal viewpoint, this special case represents a multiplicity-free variant of the general formalism, and as such facilitates the discussion while providing us with some new insights into the structure of the general U(n) Racah–Wigner algebra. It is thus useful to examine directly the relationship between the U(n) tensor algebra and the SU(2) based formalism.

An n-orbital spin-independent model of any many-electron system is characterized by the U(n) irrep $\langle 2^a 1^b \dot{0} \rangle$ with $a = \frac{1}{2}N - S$ and $b = 2S$, where N is the total electron number and S is the total spin of the state considered. An important feature of higher-order U(n) invariants for N-electron systems is their orbital and particle number independence. Let us first illustrate this fact on the case of Racah coefficients, showing that their absolute values only depend on the spin of relevant irreps.

A. Spin-dependence of U(n) invariants

Consider a general Racah coefficient [cf. Eq. (129) of Ref. 28]

$$U(\langle 2^a 1^b \rangle \langle 2^d 1^e \rangle \langle 2^x 1^y \rangle \langle 2^g 1^h \rangle ; \langle 2^s 1^t \rangle \langle 2^u 1^v \rangle) = U((a,b)(d,e)(x,y)(g,h);(s,t)(u,v)). \quad (37)$$

The particle numbers involved satisfy the following relations:

$$2a + b + 2d + e = 2s + t \quad (s \geqslant a + d), \quad (38a)$$

$$2s + t + 2g + h = 2x + y \quad (x \geqslant g + s), \quad (38b)$$

$$2d + e + 2g + h = 2u + v \quad (u \geqslant d + g), \quad (38c)$$

$$2u + v + 2a + b = 2x + y \quad (x \geqslant a + u). \quad (38d)$$

Exploiting the recursion formulas for Racah coefficients, Eq. (11), for the following choice of irreps:

$$\mu_1 = (a-1,b), \quad \mu_{23} = (u,v), \quad \mu = (x-1,y),$$

we get that

$$
\begin{aligned}
&U((a,b)(d,e)(x,y)(g,h);(s,t)(u,v))\begin{pmatrix} (x,y) & \Big| & (a,b) & (u,v) \\ (x-1,y) & \Big| & (a-1,b) & (u,v) \end{pmatrix} \\
&= \sum_{\mu_2,\mu_3,\mu_{12}} U((a-1,b)\mu_2(x-1,y)\mu_3;\mu_{12}(u,v))\begin{pmatrix} (a,b) & (d,e) & \Big| & (s,t) \\ (a-1,b) & \mu_2 & \Big| & \mu_{12} \end{pmatrix}\begin{pmatrix} (s,t) & (g,h) & \Big| & (x,y) \\ \mu_{12} & \mu_3 & \Big| & (x-1,y) \end{pmatrix} \\
&\quad\times\begin{pmatrix} (d,e) & (g,h) & \Big| & (u,v) \\ \mu_2 & \mu_3 & \Big| & (u,v) \end{pmatrix} \\
&= U((a-1,b)(d,e)(x-1,y)(g,h);(s-1,t)(u,v))\begin{pmatrix} (a,b) & (d,e) & \Big| & (s,t) \\ (a-1,b) & (d,e) & \Big| & (s-1,t) \end{pmatrix}\begin{pmatrix} (s,t) & (g,h) & \Big| & (x,y) \\ (s-1,t) & (g,h) & \Big| & (x-1,y) \end{pmatrix}.
\end{aligned}
\quad (39)
$$

TABLE I. Racah coefficients $U((a,b)(d,e)(x,y)(0,1);(s,t)(u,v))$ for the two-column U(n) irreps $(a,b) \equiv \langle 2^a 1^b \dot{0} \rangle$, etc.

(s,t)	$(u,v) = (d+1,e-1)$	$(u,v) = (d,e+1)$
$(x-1,y+1)$	$\frac{1}{2}\left[\frac{(b+e+y+3)(e-b+y+1)}{(e+1)(y+1)}\right]^{1/2}$	$\frac{-(-1)^{a+d+e+x}}{2}\left[\frac{(b-e+y+1)(b+e-y+1)}{(e+1)(y+1)}\right]^{1/2}$
$(x,y-1)$	$\frac{(-1)^{a+d+e+x}}{2}\left[\frac{(b+e-y+1)(b-e+y+1)}{(e+1)(y+1)}\right]^{1/2}$	$\frac{1}{2}\left[\frac{(b+e+y+3)(e-b+y+1)}{(e+1)(y+1)}\right]^{1/2}$

TABLE II. 9λ coefficients $X((a,b)(0,1)(d,e)(0,1);(s,t)(u,v);(x-1,y)(1,0);(x,y))$ for the two-column U(n)irreps$(a,b)\equiv\langle 2^a 1^b \dot{0}\rangle$, etc.

(s,t)	$(u,v)=(d+1,e-1)$	$(u,v)=(d,e+1)$
$(a+1,b-1)$	$\frac{(-1)^{d+e}}{2}\left[\frac{(b+e-y)(b+e+y+2)}{2(b+1)(e+1)}\right]^{1/2}$	$-\frac{(-1)^{a+x}}{2}\left[\frac{(b-e+y)(e-b+y+2)}{2(b+1)(e+1)}\right]^{1/2}$
$(a,b+1)$	$-\frac{(-1)^{a+b+x+y}}{2}\left[\frac{(b-e+y+2)(e-b+y)}{2(b+1)(e+1)}\right]^{1/2}$	$\frac{(-1)^{d}}{2}\left[\frac{(b+e+y+4)(b+e-y+2)}{2(b+1)(e+1)}\right]^{1/2}$

The explicit expressions for the isoscalar factors appearing in this equation were given in Ref. 28. In particular, we have[28]

$$\begin{pmatrix} (u,v) & (a,b) & \vert & (x,y) \\ (u,v) & (a-1,b) & \vert & (x-1,y)\end{pmatrix} = \left[\frac{a(a+b+1)}{x(x+y+1)}\right]^{1/2}, \tag{40}$$

which implies that Eq. (39) represents the following simple recursion formula:

$$\begin{aligned} &U((a,b)(d,e)(x,y)(g,h);(s,t)(u,v)) \\ &\quad = U((a-1,b)(d,e)(x-1,y)(g,h);(s-1,t)(u,v)). \end{aligned} \tag{41}$$

Using successively this formula and the symmetry properties of Racah coefficients, we finally arrive at

$$\begin{aligned} &U((a,b)(d,e)(x,y)(g,h);(s,t)(u,v)) \\ &\quad = U((0,b)(d,e)(x-a-g,y)(0,h); \\ &\qquad (s-a,t)(u-g,v)). \end{aligned} \tag{42}$$

As will be shown later, we can also remove the "two-column part" from the second irrep $\langle\lambda_2\rangle\equiv(d,e)$, obtaining finally

$$\begin{aligned} &U((a,b)(d,e)(x,y)(g,h);(s,t)(u,v)) \\ &\quad = U((0,b)(0,e)(x-a-d-g,y)(0,h); \\ &\qquad (s-a-d,t)(u-d-g,v)). \end{aligned} \tag{43}$$

Thus, the two-column parts of the three primitive irreps, namely $\langle\lambda_1\rangle\equiv(a,b)$, $\langle\lambda_2\rangle\equiv(d,e)$ and $\langle\lambda_3\rangle\equiv(g,h)$, can be excluded from our considerations so that the values of the corresponding Racah coefficients will only depend on the spin of these primitive irreps.

At this point, it should be recalled that in our recent derivations of the algebraic expressions for the U(n) isoscalar factors (cf. Appendix of Refs. 27 and 28) we have exploited the fact that the transformation coefficients relating the canonical and partitioned bases are independent of the two-column parts, which can be regarded as consisting of doubly occupied orbitals. This fact, that greatly facilitated our derivations, immediately follows from Eq. (43) since both the I_t and I_s factors can be represented as Racah coefficients, as shown in the preceding section. In view of this result, our earlier derivations could be further simplified.

Another useful corollary of Eq. (43) results when one of the primitive irreps (a,b), (d,e), or (g,h) contains only the two-column part, in which case the relevant Racah coefficient is equal to 1. Thus, for example,

$$U((a,0)(d,e)(x,y)(g,h);(s,t)(u,v)) = 1. \tag{44}$$

This corollary also immediately yields Eq. (36).

So far we have only discussed the properties of Racah coefficients. However, the same results can be derived for higher-order invariants as well. Indeed, the 9λ symbols (or X coefficients) are expressible in terms of Racah coefficients, so that their values will only depend on the spins of primitive irreps while being independent of the orbital and particle numbers.

B. Recursion formulas for Racah coefficients

Let us now derive some useful recursion formulas for Racah coefficients involving two-column irreps. We can restrict our attention to the coefficients appearing on the rhs of Eq. (42). Exploiting again Eq. (11), as well as the symmetry properties of Racah coefficients, we can write

$$\begin{aligned} &U((0,b)(d,e)(x,y)(0,h);(s,t)(u,v))\begin{pmatrix} (x,y) & \vert & (s,t) & (0,h) \\ (x,y-1) & \vert & (s,t) & (0,h-1)\end{pmatrix} \\ &\quad = U((0,b)(d,e)(x,y-1)(0,h-1);(s,t)(u,v-1))\begin{pmatrix} (d,e) & (0,h) & \vert & (u,v) \\ (d,e) & (0,h-1) & \vert & (u,v-1)\end{pmatrix}\begin{pmatrix} (0,b) & (u,v) & \vert & (x,y) \\ (0,b) & (u,v-1) & \vert & (x,y-1)\end{pmatrix} \\ &\qquad + U((0,b)(d,e)(x,y-1)(0,h-1);(s,t)(u-1,v+1)) \\ &\qquad \times\begin{pmatrix} (d,e) & (0,h) & \vert & (u,v) \\ (d,e) & (0,h-1) & \vert & (u-1,v+1)\end{pmatrix}\begin{pmatrix} (0,b) & (u,v) & \vert & (x,y) \\ (0,b) & (u-1,v+1) & \vert & (x,y-1)\end{pmatrix}. \end{aligned} \tag{45}$$

Substituting the known expressions for the isoscalar factors [cf. Table II of Ref. 28 or Table I of this paper for Racah coefficients together with the relationship (34)], we get

$$\begin{aligned} &U((0,b)(d,e)(x,y)(0,h);(s,t)(u,v))\cdot[4v(v+1)(v+2)(h-t+y)(h+t+y+2)]^{1/2} \\ &\quad = U((0,b)(d,e)(x,y-1)(0,h-1);(s,t)(u,v-1)) \\ &\qquad \times[(v+2)(h-e+v)(h+e+v+2)(b+v+y+2)(v-b+y)]^{1/2} \end{aligned}$$

$$+(-1)^{d+e+x}U\big((0,b)(d,e)(x,y-1)(0,b-1);(s,t)(u+1,v-1)\big)$$
$$\times[v(h+e-v)(e-h+v+2)(b+v-y+2)(b-v+y)]^{1/2}. \tag{46}$$

Note that we have used Eq. (38c) which implies that $(-1)^{h+v}=(-1)^{e}$, thus simplifying the phase factor in the last term. In a completely analogous way, we obtain

$$U\big((0,b)(d,e)(x,y)(0,h);(s,t)(u,v)\big)[4v(v+1)(v+2)(h+t-y)(t-h+y+2)]^{1/2}$$
$$=U\big((0,b)(d,e)(x-1,y+1)(0,h-1);(s,t)(u,v-1)\big)(-1)^{e+s+u}[(v+2)(h-e+v)$$
$$\times(b+v-y)(h+e+v+2)(b-v+y+2)]^{1/2}+U\big((0,b)(d,e)(x-1,y+1)(0,h-1);(s,t)(u-1,v+1)\big)$$
$$\times(-1)^{d+s+u+x}[v(h+e-v)(v-b+y+2)(e-h+v+2)(b+v+y+4)]^{1/2}. \tag{47}$$

Using Eqs. (46) and (47), we can now recursively evaluate any Racah coefficient involving two-column $U(n)$ irreps, starting with the simplest Racah coefficients given in Table I. As an example, we present Racah coefficients with $h=2$ in Table III.

We wish to stress again that in Eqs. (46) and (47), similarly as in Table I, the variables d, s, u, and x, that characterize the two-column part of the coupled irreps, can only appear in the phase factors. Moreover, the phase factors appearing in these formulas, namely $(-1)^{d+e+x}$, $(-1)^{e+s+u}$, and $(-1)^{d+s+u+x}$, are invariant with respect to an arbitrary shift of the two-column parts, i.e., to the transformation

$$d\to d-\tau,\quad s\to s-\tau,\quad u\to u-\tau,\quad x\to x-\tau. \tag{48}$$

Consequently, both

$$U\big((0,b)(d,e)(x,y)(0,h);(s,t)(u,v)\big)$$

and

$$U\big((0,b)(d-\tau,e)(x-\tau,y)(0,h);(s-\tau,t)(u-\tau,v)\big)$$

satisfy the same recursion formulas, Eqs. (46) and (47). Since they also involve the same starting point, Table I, we can conclude that Eq. (43) holds.

As already mentioned in the Introduction, there is a close relationship between the S_N, $U(n)$, and $SU(2)$ based approaches to the many-electron correlation problem. The above presented developments reveal an explicit form of this connection. In particular, we note that:

(i) The absolute values of simple $U(n)$ Racah coefficients, Tables I and III, are identical with the $SU(2)$ Racah coefficients. In particular,

TABLE III. $U(n)$ Racah coefficients of the type $U\big((0,b)(d,e)(x,y)(0,2);(s,t)(u,v)\big)$ involving two-column irreps $(s,t)\equiv\langle 2^s1^t\dot{0}\rangle$, etc.

(u,v)	$(s,t)\equiv(x,y-2)$
$(d,e+2)$	$\dfrac{1}{4}\left[\dfrac{(e-b+y)(b+e+y+2)(e-b+y+2)(b+e+y+4)}{(e+1)(e+2)y(y+1)}\right]^{1/2}$
$(d+1,e)$	$\dfrac{(-1)^{d+e+x}}{2}\left[\dfrac{(b-e+y)(e-b+y)(b+e+y+2)(b+e-y+2)}{2e(e+2)y(y+1)}\right]^{1/2}$
$(d+2,e-2)$	$\dfrac{1}{4}\left[\dfrac{(b-e+y)(b+e-y)(b+e-y+2)(b-e+y+2)}{e(e+1)y(y+1)}\right]^{1/2}$
(u,v)	$(s,t)=(x-1,y)$
$(d,e+2)$	$\dfrac{-(-1)^{d+e+x}}{2}\left[\dfrac{(b-e+y)(b+e-y+2)(e-b+y+2)(b+e+y+4)}{2(e+1)(e+2)y(y+2)}\right]^{1/2}$
$(d+1,e)$	$\dfrac{e(e+2)+y(y+2)-b(b+2)}{2[e(e+2)y(y+2)]^{1/2}}$
$(d+2,e-2)$	$\dfrac{(-1)^{d+e+x}}{2}\left[\dfrac{(b+e-y)(b-e+y+2)(b+e+y+2)(e-b+y)}{2e(e+1)y(y+2)}\right]^{1/2}$
(u,v)	$(s,t)=(x-2,y+2)$
$(d,e+2)$	$\dfrac{1}{4}\left[\dfrac{(b+e-y)(b-e+y)(b+e-y+2)(b-e+y+2)}{(e+1)(e+2)(y+1)(y+2)}\right]^{1/2}$
$(d+1,e)$	$\dfrac{-(-1)^{d+e+x}}{2}\left[\dfrac{(b+e-y)(b-e+y+2)(e-b+y+2)(b+e+y+4)}{2e(e+2)(y+1)(y+2)}\right]^{1/2}$
$(d+2,e-2)$	$\dfrac{1}{4}\left[\dfrac{(e-b+y)(b+e+y+2)(e-b+y+2)(b+e+y+4)}{e(e+1)(y+1)(y+2)}\right]^{1/2}$

$$|U((a,b)(d,e)(x,y)(g,1);(s,t)(u,v))|$$

$$= \left|\mathscr{U}\left(\frac{b}{2},\frac{e}{2},\frac{y}{2},\frac{1}{2};\frac{t}{2},\frac{v}{2}\right)\right|$$

$$= \sqrt{(t+1)(v+1)}\left|\begin{Bmatrix} \frac{b}{2} & \frac{e}{2} & \frac{t}{2} \\ \frac{1}{2} & \frac{y}{2} & \frac{v}{2} \end{Bmatrix}\right| \tag{49}$$

and

$$|U((a,b)(d,e)(x,y)(g,2);(s,t)(u,v))|$$

$$= \left|\mathscr{U}\left(\frac{b}{2},\frac{e}{2},\frac{y}{2},1;\frac{t}{2},\frac{v}{2}\right)\right|$$

$$= \sqrt{(t+1)(v+1)}\left|\begin{Bmatrix} \frac{b}{2} & \frac{e}{2} & \frac{t}{2} \\ 1 & \frac{y}{2} & \frac{v}{2} \end{Bmatrix}\right|, \tag{50}$$

where U and $\mathscr{U}$ designate the U(n) and SU(2) Racah coefficients, respectively. The above relations also show the connection with the usual $6j$ symbols.

(ii) The SU(2) Racah coefficients satisfy a number of recursion formulas, which enable their efficient evaluation. The above presented recursion formulas, Eqs. (46) and (47), for the U(n) Racah coefficients involving two-column irreps are again identical with the corresponding SU(2) formulas, except for the phases involved. For example, Biedenharn's formula [cf., e.g., Eq. (6.3.5) of Ref. 42]

$$\mathscr{U}(j_1j_2jj_3;j_{12}j_{23})[2j_{23}(j_{23}+1)(2j_{23}+1)(j_3-j_{12}+j)(j_{12}+j_3+j+1)]^{1/2}$$

$$= (-1)^{2j_1+2j_3}\left\{\mathscr{U}\left(j_1,j_2,j-\frac{1}{2},j_3-\frac{1}{2};j_{12},j_{23}-\frac{1}{2}\right)[(j_{23}+1)(j_3-j_2+j_{23})(j_2+j_3+j_{23}+1)\right.$$

$$\times(j_1+j_{23}+j+1)(j_{23}-j_1+j)]^{1/2} - \mathscr{U}\left(j_1,j_2,j-\frac{1}{2},j_3-\frac{1}{2};j_{12},j_{23}+\frac{1}{2}\right)$$

$$\left.\times[j_{23}(j_2+j_3-j_{23})(j_2-j_3+j_{23}+1)(j_1+j_{23}-j+1)(j_1-j_{23}+j)]^{1/2}\right\}, \tag{51}$$

is equivalent with Eq. (46), except for the phases, which can be either opposite for the two terms on the rhs as in Eq. (51), or the same in view of the extra factor $(-1)^{d+e+x}$ in the second term on the rhs of Eq. (46).

We can thus conclude that up to the phase factors, the U(n) Racah coefficients, as well as the higher-order invariants (which are expressible in terms of Racah coefficients), that involve at most two-column irreps which are relevant in many-electron problems, are identical with the corresponding SU(2) invariants. In fact, in view of the Clifford algebra UGA representation[43] of U(n) states, we can expect that the Racah coefficients involving k-column irreps will behave analogously as the SU(k) invariants.

IV. DISCUSSION

This paper contains two distinct types of results, which we discuss in turn. The first one involves general relationships between various U(n) and S_N isoscalar factors and higher-order U(n) invariants, while the second one investigates an explicit form of the relationship between the U(n) and SU(2) approaches to the many-electron correlation problem.

Let us first stress the striking simplicity of the relationships which relate various U(n) and S_N isoscalar factors with U(n) Racah coefficients or 9λ symbols, as well as an equally striking simplicity of the derivation of these results. Although it is expected that higher-order invariants do reduce to lower-order ones when one or more irreps involved is a particularly simple one, it is indeed surprising that any U(n) isoscalar factor is just a special Racah or 9λ symbol. Moreover, an analogous relationship must exist for any group (considering finite-dimensional irreps), although it may not be as simple as for the U(n) case considered, since the isoscalar factors and higher-order invariants used are defined analogously for any group chain. It is worth remarking at this occasion that very recently LeBlanc and Biedenharn[35] found that the two classes of isoscalar factors (projective operators) can be related with the 9λ symbols. This result was obtained using the vector coherent state technique[32–35] for the U(3) ⊃ U(2) chain and its validity was conjectured for the general U(n) ⊃ U($n-1$) case. We also note that the approach employed in this paper is very similar to the so-called "build-up procedure" for the isoscalar factors and recoupling coefficients.[44] Finally, we must caution the reader that throughout this paper we assume that the rank n of U(n) is sufficiently large so that U(n) possesses all the irreps with a given N (equal to the number of boxes in a Young diagram labeling a given irrep), determining in turn S_N, which may be rather small. Consequently, minor adjustments may be required when applying these results to U(n) groups with small rank, in which case certain irreps (or corresponding terms containing them) may have to be omitted should their row number exceed n.

It thus remains to discuss the relationship between the U(n) and SU(2) approaches, which is relevant to special two-column irreps characterizing the many-electron correlation problem. In view of the UGA formalism,[6–16] we can call this relationship a *conjugation property* and assume its validity in general. Moreover, we expect that, for example, the U(n) Racah coefficients involving three-column irreps and the SU(3) Racah coefficients will be identical (up to the phase) in case when no multiplicity arises or can be so chosen even when multiplicity does occur. In fact, this conjugation property is implied by the Littlewood–Richardson rules. It is well known that the outer-product reduction $[\lambda] \otimes [\mu] \downarrow [\nu]$ is identical with that for the conjugate reps, i.e., $[\tilde{\lambda}] \otimes [\tilde{\mu}] \downarrow [\tilde{\nu}]$. Although the corresponding U(n) irreps $\langle \lambda \rangle$ and $\langle \tilde{\lambda} \rangle$ have different dimensions, the result still holds. In fact, since Racah coefficients represent transformations between different coupling schemes and are independent of the chosen basis, we can choose the Weyl tableau W in Eq. (9) [or, similarly, in Eq. (12)], representing the final coupled state, to be a special Gel'fand basis vector[40] with all orbitals singly occupied. Then, all other Weyl tableaux labeling various coupled states must also correspond to special Gel'fand states, and the coupling reduces to the S_N case, in which the outer product coupling coefficients for mutually conjugate irreps are identical except for the phase, and so are the final Racah coefficients. In the particular case of two-column irreps, when the relationship between U(n) and SU(2) is involved, we can understand this conjugation property in view of the following facts: (i) primitive irreps $\lambda_1, \lambda_2, \lambda_3$ can be reduced to single-column irreps, (ii) the coupling of three single-column irreps is conjugate to that for three single-row irreps [they are, in fact, SU(2) irreps] and, finally, (iii) U(n) isoscalar factors for coupling of two single-row irreps to a two-row irrep are just the SU(2) CG coefficients and similarly for higher-order invariants. Note, however, that in the general case, coupling of two-column irreps can yield a three-column one, requiring the knowledge of SU(3) quantities, or a four-column one, when SU(4) comes into play.

We note, finally, two implications of the present results for future work: (i) It enables a novel approach to the evaluation of U(n) isoscalar factors. For example, knowing the SU(3) Racah coefficients, we can obtain (similarly as above) the U(n) Racah coefficients involving three-column irreps, which in turn will yield U(n) isoscalar factors involving three-column irreps using general relations established in Sec. II. (ii) It enables a canonical solution to the multiplicity problem, since we can exploit the conjugation property to define the multiplicity labels. Thus, for example, because the SU(3) tensor operators have been completely evaluated and classified (cf. e.g., Ref. 45), the corresponding multiplicity problem for the three-column U(n) irreps may be resolved in a conjugate scheme. This fact is of considerable practical significance, since in actual applications the number of columns in each irrep is determined by the physics of the problem (e.g., two column irreps for $\frac{1}{2}$-spin systems,[6] four-column irreps for $\frac{1}{2}$–spin–isospin systems,[16] etc.). Clearly, a major existing problem is an easy determination of relevant phase factors.

ACKNOWLEDGMENTS

One of us (X.L.) would like to express his thanks to Professor J. Paldus for his hospitality during his stay in Waterloo.

Continued support by NSERC (J.P.) is gratefully acknowledged.

[1] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[2] L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[3] J. D. Louck, J. Am. Phys. **38**, 3 (1970).
[4] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970); **12**, 173 (1971).
[5] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics, Theory and Application* (Addison–Wesley, Reading, MA, 1981); L. C. Biedenharn and J. D. Louck, *The Racah-Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, MA, 1981).
[6] J. Paldus, J. Chem. Phys. **61**, 5321 (1974).
[7] J. Paldus, in *Theoretical Chemistry: Advances and Perspectives*, edited by H. Eyring and D. Henderson (Academic, New York, 1976), Vol. 2, pp. 131–290.
[8] I. Shavitt, Int. J. Quantum Chem. **S11**, 131 (1977); **S12**, 5 (1978).
[9] J. Paldus and M. J. Boyle, Phys. Scripta **21**, 295 (1980).
[10] *The Unitary Group for the Evaluation of Electronic Energy Matrix Elements*, Lecture Notes in Chemistry, Vol. 22, edited by J. Hinze (Springer, Berlin, 1981).
[11] M. S. Robb and U. Niazi, Comp. Phys. Rep. **1**, 127 (1984).
[12] F. A. Matsen and R. Pauncz, *The Unitary Group in Quantum Chemistry* (Elsevier, Amsterdam, 1986).
[13] J. Paldus, in *Symmetries in Science II*, edited by B. Gruber and R. Lenczewski (Plenum, New York, 1986), pp. 429–446.
[14] J. Paldus, in *Mathematical Frontiers in Computational Chemical Physics*, IMA Series Vol. 15, edited by D. G. Truhlar (Springer, Berlin, 1988), pp. 262–299.
[15] I. Shavitt, in Ref. 14, pp. 300–340.
[16] M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1968).
[17] K. T. Hecht, Commun. Math. Phys. **41**, 135 (1975).
[18] K. T. Hecht and J. P. Draayer, Nucl. Phys. A **223**, 285 (1974).
[19] P. Kramer and M. Moshinsky, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968), Vol. 2, pp. 339–468.
[20] W. Duch and J. Karwowski, Comp. Phys. Rep. **2**, 93 (1985).
[21] W. Duch, *GRMS or Graphical Representation of Model Spaces*, Lecture Notes in Chemistry, Vol. 42 (Springer, Berlin, 1986).
[22] A. P. Jucys, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Institute of Physics and Mathematics of the Academy of Sciences of the Lithuanian S.S.R., Mintis, Vilnius, 1960) (in Russian), English translations: (Israel Program for Scientific Translations, Jerusalem, 1962), and (Gordon and Breach, New York, 1964); A. P. Jucys and A. A. Bandzaitis, *The Theory of Angular Momentum in Quantum Mechanics* (Mokslas, Vilnius, 1977), 2nd ed. (in Russian); D. M. Brink and G. R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1968), 2nd ed.; E. El Baz and B. Castel, *Graphical Methods of Spin Algebras in Atomic, Nuclear and Particle Physics* (Dekker, New York, 1972); I. Lindgren and J. Morrison, *Atomic Many-Body Theory* (Springer, Berlin, 1972).
[23] G. W. Drake and M. Schlesinger, Phys. Rev. A **15**, 1990 (1977).
[24] H. S. Green, J. Math. Phys. **12**, 2106 (1971); M. D. Gould, J. Math. Phys. **21**, 444 (1980); **22**, 15 (1981); **27**, 1944 (1986).
[25] M. D. Gould and G. S. Chandler, Int. J. Quantum Chem. **25**, 1089 (1984) and references therein; **26**, 44 (1984).
[26] M. D. Gould and J. Paldus, Int. J. Quantum Chem. **30**, 327 (1986); Phys. Rev. A **34**, 804 (1986); M. D. Gould, Int. J. Quantum Chem. **30**, 364 (1986).
[27] X. Li and J. Paldus, Int. J. Quantum Chem. **36**, 177 (1989).
[28] X. Li and J. Paldus, J. Math. Chem. (in press).
[29] Q. Zhang and X. Li, J. Phys. A: Math. Gen. **20**, 6185 (1987).
[30] X. Li and J. Paldus, unpublished results.
[31] R. D. Kent and M. Schlesinger, Phys. Rev. A **40**, 536 (1989).
[32] K. T. Hecht, R. LeBlanc, and R. J. Rowe, J. Phys. A: Math. Gen. **20**, 2241 (1987).
[33] R. LeBlanc and K. T. Hecht, J. Phys. A: Math. Gen. **20**, 4613 (1987).
[34] R. LeBlanc, J. Phys. A: Math. Gen. **20**, 5015 (1987).
[35] R. LeBlanc and L. C. Biedenharn, J. Phys. A: Math. Gen. **22**, 31 (1989).
[36] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 825, 1070 (1950).

[37] J. R. Derome, J. Math. Phys. **7**, 612 (1966).
[38] B. G. Searle and P. H. Butler, J. Phys. A: Math. Gen. **21**, 3313 (1988).
[39] H. A. Jahn, Proc. R. Soc. London, Ser. A **205**, 192 (1951).
[40] M. Moshinsky, J. Math. Phys. **7**, 691 (1966).
[41] S. J. Ališauskas, A.-A. A. Jucys and A. P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[42] A. E. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[43] J. Paldus and C. R. Sarma, J. Chem. Phys. **83**, 5135 (1985); C. R. Sarma and J. Paldus, J. Math. Phys. **26**, 1140 (1985); M. D. Gould and J. Paldus, J. Math. Phys. **28**, 2304 (1987).
[44] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[45] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).

On some Racah coefficients of U(n)

C. Quesne[a)]

Physique Nucléaire Théorique et Physique Mathématique CP229, Université Libre de Bruxelles, Campus de la Plaine, Bd du Triomphe, B-1050 Brussels, Belgium

(Received 20 July 1989; accepted for publication 7 March 1990)

Some U(n) Racah coefficients appearing in various applications of vector coherent state and K-matrix theories are calculated: For such a purpose, use is made of their definition in terms of U(n):U($n-1$) reduced Wigner coefficients and U($n-1$) Racah coefficients. By starting from known U(2) Racah coefficients, the recursion relations obtained are solved by induction over n.

I. INTRODUCTION

During the last few years, the vector coherent state[1,2] and K-matrix[2–6] theories have played an ever-increasing role in the representation theory of Lie algebras (see Refs. 5 and 6 and references therein). Indeed, their combination provides a simple systematic way of determining the explicit matrices for the ladder irreducible representations (irreps) of all classical Lie algebras in bases that reduce their u(n) or gl(n) subalgebra.

In all cases, the analytic formulas or numerical algorithms obtained contain some U(n) Racah coefficients. Although the values of most of the coefficients are given in the literature, some are unknown. The purpose of the present paper is to calculate some Racah coefficients appearing in various applications of the vector coherent state and K-matrix theories.[6,7]

In Secs. II and III, we respectively prove the following results:

$$\begin{aligned}&U([1\dot{0}]_n[1\dot{0}-1]_n[f+\Delta^{(1)}(i)]_n[f]_n;\\&\quad[1\dot{0}]_n[f]_n(\rho=1))\\&\quad=[(n-1)n(n+1)g([f]_n)]^{-1/2}\\&\quad\quad\times\left[\sum_{j=1}^{n}f_j-n(f_i-i+1)\right]\end{aligned}\tag{1.1}$$

and

$$\begin{aligned}&U([1^2\dot{0}]_n[1\dot{0}-1]_n[f+\Delta^{(2)}(i,j)]_n[f]_n;\\&\quad[2\dot{0}]_n[f]_n(\rho=1))\\&\quad=[2(n-1)g([f]_n)]^{-1/2}[(f_i-f_j+j-i-1)\\&\quad\quad\times(f_i-f_j+j-i+1)]^{1/2},\end{aligned}\tag{1.2}$$

where the U coefficients are Racah coefficients in unitary form.[8,9] By using the symmetry properties of (1.2) (Refs. 8 and 9), we obtain the relation

$$\begin{aligned}&U([f+\Delta^{(2)}(i,j)]_n[\dot{0}-2]_n[f]_n[1\dot{0}-1]_n;\\&\quad[f_n](\rho=1)[\dot{0}(-1)^2]_n)\\&\quad=-[2(n+1)g([f]_n)]^{-1/2}[(f_i-f_j+j-i-1)\\&\quad\quad\times(f_i-f_j+j-i+1)]^{1/2},\end{aligned}\tag{1.3}$$

which was proved for $n=3$ and surmized for higher values of n by Rowe *et al.*[6]

The notations in Eqs. (1.1)–(1.3) are defined as follows. The U(n) irreps are characterized by partitions into n integers $[f]_n\equiv[f_1f_2\cdots f_n]$, $f_1\geqslant f_2\geqslant\cdots\geqslant f_n$. An overdot over a numeral implies that this numeral is repeated as often as necessary. The symbols $\Delta^{(1)}(i)$ and $\Delta^{(2)}(i,j)$, respectively, denote row vectors of dimension n with vanishing entries everywhere except for the components i or i and j, which have value unity. All couplings in (1.1)–(1.3) are multiplicity free except for that of $[1\dot{0}-1]_n$, with $[f]_n$ giving $[f]_n$, which has a multiplicity of $n-1$ for the generic irreps $[f]_n$: The multiplicity index $\rho=1$ (corresponding to $q=1$ in Louck and Biedenharn's notations[10]) refers to the case where this coupling arises from the matrix elements of the SU(n) generators $E_{ij}-n^{-1}\delta_{ij}\sum_k E_{kk}$, $i,j=1,\ldots,n$. Finally, the rhs of (1.1)–(1.3) contain the function

$$g([f]_n)=n^{-1}\sum_{i<j}(f_i-f_j)(f_i-f_j+2j-2i),\tag{1.4}$$

which is the eigenvalue of the SU(n) Casimir operator

$$G=\sum_{i,j=1}^{n}E_{ij}E_{ji}-n^{-1}\left(\sum_{i=1}^{n}E_{ii}\right)^2\tag{1.5}$$

corresponding to the irrep $[f]_n$.

II. PROOF OF EQ. (1.1)

The proof of Eq. (1.1) is based on the definition of U(n) Racah coefficients in terms of U(n):U($n-1$) reduced Wigner coefficients and U($n-1$) Racah coefficients, which in the present case reads as

$$\begin{aligned}&U([1\dot{0}]_n[1\dot{0}-1]_n[f+\Delta^{(1)}(i)]_n[f]_n;[1\dot{0}]_n[f]_n(\rho=1))\\&\quad=\sum\langle[1\dot{0}]_n[m^{(1)}]_{n-1};[1\dot{0}-1]_n[m^{(2)}]_{n-1}\|[1\dot{0}]_n[m^{(12)}]_{n-1}\rangle\\&\quad\quad\times\langle[1\dot{0}]_n[m^{(12)})]_{n-1};[f]_n[m^{(3)}]_{n-1}\|[f+\Delta^{(1)}(i)]_n[m]_{n-1}\rangle\end{aligned}$$

a) Director of Research, F.N.R.S.

 0022-2488/90/071600-05$03.00

$$\times\langle[1\dot{0}-1]_n[m^{(2)}]_{n-1};[f]_n[m^{(3)}]_{n-1}\|[f]_n[m^{(23)}]_{n-1}\rangle_{\rho=1}$$
$$\times\langle[1\dot{0}]_n[m^{(1)}]_{n-1};[f]_n[m^{(23)}]_{n-1}\|[f+\Delta^{(1)}(i)]_n[m]_{n-1}\rangle$$
$$\times \mathrm{U}([m^{(1)}]_{n-1}[m^{(2)}]_{n-1}[m]_{n-1}[m^{(3)}]_{n-1};[m^{(12)}]_{n-1}[m^{(23)}]_{n-1}(\rho=1)). \tag{2.1}$$

On the rhs of Eq. (2.1), the $\mathrm{U}(n-1)$ irrep $[m]_{n-1}\equiv[m_1m_2\cdots m_{n-1}]$ is kept fixed, while the summation runs over $[m^{(1)}]_{n-1}$, $[m^{(2)}]_{n-1}$, $[m^{(3)}]_{n-1}$, $[m^{(12)}]_{n-1}$, $[m^{(23)}]_{n-1}$; in the $\mathrm{U}(n-1)$ Racah coefficient, the multiplicity index $\rho=1$ is required only when $[m^{(2)}]_{n-1}=[1\dot{0}-1]_{n-1}$, since the remaining recoupling coefficients are multiplicity free.

To reduce the number of contributing terms in Eq. (2.1), it is convenient to choose for $[m]_{n-1}$ the highest weight $\mathrm{U}(n-1)$ irrep contained in $[f+\Delta^{(1)}(i)]_n$:

$$\begin{aligned}[m]_{n-1}&=[f+\Delta^{(1)}(i)]_{n-1}, \quad \text{if } i<n,\\ &=[f]_{n-1}, \quad \text{if } i=n,\end{aligned} \tag{2.2}$$

where $[f]_{n-1}\equiv[f_1f_2\cdots f_{n-1}]$ and $\Delta^{(1)}(i)$ is now a row vector of dimension $n-1$. The $\mathrm{U}(n-1)$ irreps corresponding to such nonzero terms are listed in Table I. For $i<n$, there are only two contributing terms, while for $i=n$ there are five, the first four of which contain a summation over $j=1,\ldots,n-1$.

Among the four types of reduced Wigner coefficients appearing on the rhs of Eq. (2.1), three belong to the class of fundamental reduced Wigner coefficients and can be easily evaluated by using the Biedenharn and Louck pattern calculus rules.[11] The remaining reduced Wigner coefficient, corresponding to the coupling of $[1\dot{0}-1]_n$ with $[f]_n$, can be determined from the results of Louck and Biedenharn[10] or directly calculated from the known matrix elements of the $\mathrm{U}(n)$ generators in the Gel'fand basis.[12] On the other hand, the $\mathrm{U}(n-1)$ Racah coefficients are either of the same type as the $\mathrm{U}(n)$ Racah coefficient of the lhs of Eq. (2.1) or have a trivial value, because one of the coupled irreps is $[\dot{0}]_{n-1}$.

Hence, Eq. (2.1) leads to the following recursion relations for the Racah coefficients:

$$(n-1)R_i([f]_n)=nR_i([f]_{n-1})-\left(\sum_{j=1}^{n-1}f_j-(n-1)f_n\right), \quad i<n, \tag{2.3}$$

$$\begin{aligned}(n-1)R_n([f]_n)=nA([f]_n)\Bigg(&\sum_{j=1}^{n-1}(f_j-f_n+n-j-1)^{-1}B_j([f]_{n-1})R_j([f-\Delta^{(1)}(j)]_{n-1})\\ &+\sum_{j=1}^{n-1}f_j-(n-1)f_n+2(n-1)^2-1\Bigg)-\left(\sum_{j=1}^{n-1}f_j-(n-1)f_n-n\right),\end{aligned} \tag{2.4}$$

where $R_i([f]_n)$, $A([f]_n)$, and $B_j([f]_{n-1})$ are, respectively, defined by

$$R_i([f]_n)=[(n-1)n(n+1)g([f]_n)]^{1/2}\mathrm{U}([1\dot{0}]_n[1\dot{0}-1]_n[f+\Delta^{(1)}(i)]_n[f]_n;[1\dot{0}]_n[f]_n(\rho=1)), \tag{2.5}$$

$$A([f]_n)=\prod_{k=1}^{n-1}[(f_k-f_n+n-k-1)(f_k-f_n+n-k)^{-1}], \tag{2.6}$$

and

$$B_j([f]_{n-1})=\prod_{\substack{k=1\\k\neq j}}^{n-1}[(f_j-f_k+k-j-1)(f_j-f_k+k-j)^{-1}]. \tag{2.7}$$

By starting from the well-known $\mathrm{U}(2)$ Racah coefficients and proceeding by induction over n, it can be shown that the solution of Eqs. (2.3) and (2.4) is

$$R_i([f]_n)=\sum_{j=1}^{n}f_j-n(f_i-i+1), \tag{2.8}$$

TABLE I. The $\mathrm{U}(n-1)$ irreps which give nonzero contributions to the rhs of Eq. (2.1) when $[m]_{n-1}$ is chosen according to Eq. (2.2).

i	$[m^{(1)}]_{n-1}$	$[m^{(2)}]_{n-1}$	$[m^{(12)}]_{n-1}$	$[m^{(3)}]_{n-1}$	$[m^{(23)}]_{n-1}$
$i<n$	$[1\dot{0}]_{n-1}$	$[1\dot{0}-1]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$
$i=n$	$[1\dot{0}]_{n-1}$	$[1\dot{0}-1]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[\dot{0}-1]_{n-1}$	$[\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$
	$[\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f-\Delta^{(1)}(j)]_{n-1}$	$[f]_{n-1}$
	$[\dot{0}]_{n-1}$	$[\dot{0}]_{n-1}$	$[\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$

therefore leading to Eq. (1.1). The proof of Eq. (2.3) is straightforward, while that of Eq. (2.4) proceeds as follows. After introducing the value of $R_j([f-\Delta^{(1)}(j)]_{n-1})$ coming from Eq. (2.8) into the rhs of Eq. (2.4) and after recombining the various terms, one obtains the relation

$$\begin{aligned}(n-1)R_n([f]_n) &= nA([f]_n)\Big[-(n-1)\sum_{j=1}^{n-1}B_j([f]_{n-1})+\Big(\sum_{k=1}^{n-1}f_k-(n-1)f_n+(n-1)^2-1\Big)\\ &\quad\times\sum_{j=1}^{n-1}(f_j-f_n+n-j-1)^{-1}B_j([f]_{n-1})+\sum_{j=1}^{n-1}f_j-(n-1)f_n+2(n-1)^2-1\Big]\\ &\quad-\Big(\sum_{j=1}^{n-1}f_j-(n-1)f_n-n\Big).\end{aligned} \tag{2.9}$$

The two sums of the $B_j([f]_{n-1})$ functions can be easily evaluated using the complex function residue theory[13] and are given by

$$\sum_{j=1}^{n-1}B_j([f]_{n-1})=n-1 \tag{2.10}$$

and

$$\sum_{j=1}^{n-1}(f_j-f_n+n-j-1)^{-1}B_j([f]_{n-1})=[A([f]_n)]^{-1}-1, \tag{2.11}$$

respectively. By taking Eqs. (2.10) and (2.11) into account, Eq. (2.9) can be transformed into the relation

$$(n-1)R_n([f]_n)=(n-1)\Big(\sum_{j=1}^{n-1}f_j-(n-1)f_n+n(n-1)\Big), \tag{2.12}$$

thus completing the proof of Eq. (2.8).

III. PROOF OF EQ. (1.2)

The proof of Eq. (1.2) proceeds in the same way as that of Eq. (1.1). We therefore start from the definition

$$\begin{aligned}&\mathbf{U}([1^2\dot{0}]_n[1\dot{0}-1]_n[f+\Delta^{(2)}(i,j)]_n[f]_n;[2\dot{0}]_n[f]_n(\rho=1))\\ &\quad=\sum\langle[1^2\dot{0}]_n[m^{(1)}]_{n-1};[1\dot{0}-1]_n[m^{(2)}]_{n-1}\|[2\dot{0}]_n[m^{(12)}]_{n-1}\rangle\\ &\qquad\times\langle[2\dot{0}]_n[m^{(12)}]_{n-1};[f]_n[m^{(3)}]_{n-1}\|[f+\Delta^{(2)}(i,j)]_n[m]_{n-1}\rangle\\ &\qquad\times\langle[1\dot{0}-1]_n[m^{(2)}]_{n-1};[f]_n[m^{(3)}]_{n-1}\|[f]_n[m^{(23)}]_{n-1}\rangle_{\rho=1}\\ &\qquad\times\langle[1^2\dot{0}]_n[m^{(1)}]_{n-1};[f]_n[m^{(23)}]_{n-1}\|[f+\Delta^{(2)}(i,j)]_n[m]_{n-1}\rangle\\ &\qquad\times\mathbf{U}([m^{(1)}]_{n-1}[m^{(2)}]_{n-1}[m]_{n-1}[m^{(3)}]_{n-1};[m^{(12)}]_{n-1}[m^{(23)}]_{n-1}(\rho=1)),\end{aligned} \tag{3.1}$$

where we take, for $[m]_{n-1}$, the highest weight $\mathrm{U}(n-1)$ irrep contained in $[f+\Delta^{(2)}(i,j)]_n$:

$$\begin{aligned}[m]_{n-1}&=[f+\Delta^{(2)}(i,j)]_{n-1}, \quad \text{if } i<j<n,\\ &=[f+\Delta^{(1)}(i)]_{n-1}, \quad \text{if } i<j=n.\end{aligned} \tag{3.2}$$

The $\mathrm{U}(n-1)$ irreps corresponding to the contributing terms in Eq. (3.1) are listed in Table II. For $i<j<n$, the rhs of Eq. (3.1) comprises a single term, while for $i<j=n$ it includes five terms, the first two of which contain a summation over $k=1,\ldots,n-1$, $k\neq i$ and the third a summation over $k=1,\ldots,n-1$.

Among the four types of reduced Wigner coefficients appearing on the rhs of Eq. (3.1), two belong to the class of elementary reduced Wigner coefficients and are given in Ref. 11, another involves the totally symmetric irrep $[2\dot{0}]_n$ and is calculated in the Appendix, while the remaining one corresponds to the coupling of $[1\dot{0}-1]_n$ with $[f]_n$ and is evaluated as explained in Sec. II. On the other hand, the $\mathrm{U}(n-1)$ Racah coefficients fall into four categories: (i) coefficients of the same type as the $\mathrm{U}(n)$ Racah coefficient of the lhs of Eq. (3.1), (ii) coefficients given in Eq. (1.1), (iii) coefficients that can be calculated from those given in the Appendix of Ref. 14 by using some symmetry properties[8,9] of Racah coefficients, and (iv) coefficients that have a trivial value because one of the coupled irreps is $[\dot{0}]_{n-1}$.

By taking the above results and Eq. (2.7) into account, Eq. (3.1) can be put into the following form:

$$R_{ij}([f]_n)=R_{ij}([f]_{n-1}), \quad i<j<n, \tag{3.3}$$

TABLE II. The $\mathrm{U}(n-1)$ irreps which give nonzero contributions to the rhs of Eq. (3.1) when $[m]_{n-1}$ is chosen according to Eq. (3.2).

ij	$[m^{(1)}]_{n-1}$	$[m^{(2)}]_{n-1}$	$[m^{(12)}]_{n-1}$	$[m^{(3)}]_{n-1}$	$[m^{(23)}]_{n-1}$
$i<j<n$	$[1^2\dot{0}]_{n-1}$	$[1\dot{0}-1]_{n-1}$	$[2\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$
$i<j=n$	$[1^2\dot{0}]_{n-1}$	$[1\dot{0}-1]_{n-1}$	$[2\dot{0}]_{n-1}$	$[f-\Delta^{(1)}(k)]_{n-1}$	$[f-\Delta^{(1)}(k)]_{n-1}$
	$[1^2\dot{0}]_{n-1}$	$[\dot{0}-1]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f-\Delta^{(1)}(k)]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[2\dot{0}]_{n-1}$	$[f-\Delta^{(1)}(k)]_{n-1}$	$[f]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[1\dot{0}-1]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$
	$[1\dot{0}]_{n-1}$	$[\dot{0}]_{n-1}$	$[1\dot{0}]_{n-1}$	$[f]_{n-1}$	$[f]_{n-1}$

$$R_{in}([f]_n) = C_i([f]_n)\Bigg(-\sum_{\substack{k=1\\k\neq i}}^{n-1} [(f_i-f_k+k-i+1)(f_k-f_n+n-k-1)]^{-1}[(f_i-f_k+k-i)$$
$$\times (f_i-f_k+k-i+2)]^{1/2} B_k([f]_{n-1}) R_{ik}([f-\Delta^{(1)}(k)]_{n-1})$$
$$-2\sum_{k=1}^{n-1} B_k([f]_{n-1}) - (f_i-f_n-i+1)\Bigg), \quad i<j=n, \tag{3.4}$$

where $R_{ij}([f]_n)$ and $C_i([f]_n)$ are, respectively, defined by

$$R_{ij}([f]_n) = R_{ji}([f]_n) = [2(n-1)g([f]_n)]^{1/2}$$
$$\times \mathrm{U}([1^2\dot{0}]_n [1\dot{0}-1]_n [f+\Delta^{(2)}(i,j)]_n [f]_n;[2\dot{0}]_n [f]_n(\rho=1)), \quad i<j, \tag{3.5}$$

and

$$C_i([f]_n) = -[(f_i-f_n+n-i-1)(f_i-f_n+n-i+1)^{-1}]^{1/2}$$
$$\times \prod_{\substack{k=1\\k\neq i}}^{n-1} [(f_k-f_n+n-k-1)(f_k-f_n+n-k)^{-1}]. \tag{3.6}$$

By induction over n, it can be shown that the solution of Eqs. (3.3) and (3.4) is

$$R_{ij}([f]_n) = [(f_i-f_j+j-i-1)(f_i-f_j+j-i+1)]^{1/2}, \tag{3.7}$$

hence leading to Eq. (1.2). In the case of Eq. (3.4), after introducing the value of $R_{ik}([f-\Delta^{(1)}(k)]_{n-1})$ coming from Eq. (3.7) into its rhs and after recombining the various terms, one obtains the relation

$$R_{in}([f]_n) = C_i([f]_n)\Bigg(\sum_{\substack{k=1\\k\neq i}}^{n-1} (f_i-f_k+k-i)(f_i-f_k+k-i+1)^{-1} B_k([f]_{n-1})$$
$$-(f_i-f_n+n-i+1)\sum_{\substack{k=1\\k\neq i}}^{n-1} (f_i-f_k+k-i)[(f_i-f_k+k-i+1)(f_k-f_n+n-k-1)]^{-1}$$
$$\times B_k([f]_{n-1}) - 2\sum_{k=1}^{n-1} B_k([f]_{n-1}) - (f_i-f_n-i+1)\Bigg), \tag{3.8}$$

where the three sums of the $B_k([f]_{n-1})$ functions are given by

$$\sum_{\substack{k=1\\k\neq i}}^{n-1} (f_i-f_k+k-i)(f_i-f_k+k-i+1)^{-1} B_k([f]_{n-1}) = n-2, \tag{3.9}$$

$$\sum_{\substack{k=1\\k\neq i}}^{n-1} (f_i-f_k+k-i)[(f_i-f_k+k-i+1)(f_k-f_n+n-k-1)]^{-1} B_k([f]_{n-1})$$
$$= -[(f_i-f_n+n-i-1)(f_i-f_n+n-i+1)^{-1}]^{1/2}[C_i([f]_n)]^{-1} - 1, \tag{3.10}$$

and Eq. (2.10), respectively. Equations (3.9) and (3.10) can be directly derived from Eqs. (2.10) and (2.11) by changing $n-1$ into $n-2$ and $f_1 \cdots f_{n-1}$ into $f_1 \cdots f_{i-1} f_{i+1} \cdots f_{n-1}$. Their substitution into Eq. (3.8) finally completes the proof of Eq. (3.7).

APPENDIX: U(n):U($n-1$) REDUCED WIGNER COEFFICIENTS INVOLVING THE IRREP $[2\dot{0}]_n$

According to Table II, three different types of $\mathrm{U}(n)$:$\mathrm{U}(n-1)$ reduced Wigner coefficients involving $[2\dot{0}]_n$ appear on the rhs of Eq. (3.1), namely

$$\langle [2\dot{0}]_n [2\dot{0}]_{n-1};[f]_n [f]_{n-1} \| [f+\Delta^{(2)}(i,j)]_n [f+\Delta^{(2)}(i,j)]_{n-1}\rangle \quad \text{for } i<j<n,$$

$$\langle [2\dot{0}]_n [2\dot{0}]_{n-1};[f]_n [f-\Delta^{(1)}(k)]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \quad \text{for } i,k<n,$$

and

$$\langle [2\dot{0}]_n [1\dot{0}]_{n-1};[f]_n [f]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \quad \text{for } i<n.$$

Since the first of the above coefficients is trivially equal to 1, we are only left with the calculation of the remaining two, which is the purpose of this Appendix.

This calculation can be easily performed by using the relation

$$\begin{aligned}
&\langle [2\dot{0}]_n [m^{(12)}]_{n-1};[f]_n [m^{(3)}]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \\
&\quad = [\mathrm{U}([1\dot{0}]_n [1\dot{0}]_n [f+\Delta^{(2)}(i,n)]_n [f]_n;[2\dot{0}]_n [f+\Delta^{(1)}(n)]_n)]^{-1} \\
&\qquad \times \sum \langle [1\dot{0}]_n [m^{(1)}]_{n-1};[1\dot{0}]_n [m^{(2)}]_{n-1} \| [2\dot{0}]_n [m^{(12)}]_{n-1}\rangle \\
&\qquad \times \langle [1\dot{0}]_n [m^{(2)}]_{n-1};[f]_n [m^{(3)}]_{n-1} \| [f+\Delta^{(1)}(n)]_n [m^{(23)}]_{n-1}\rangle \\
&\qquad \times \langle [1\dot{0}]_n [m^{(1)}]_{n-1};[f+\Delta^{(1)}(n)]_n [m^{(23)}]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \\
&\qquad \times \mathrm{U}([m^{(1)}]_{n-1} [m^{(2)}]_{n-1} [f+\Delta^{(1)}(i)]_{n-1} [m^{(3)}]_{n-1};[m^{(12)}]_{n-1} [m^{(23)}]_{n-1}),
\end{aligned} \tag{A1}$$

resulting from the definition of the $\mathrm{U}(n)$ Racah coefficients and the orthogonality properties of $\mathrm{U}(n)$:$\mathrm{U}(n-1)$ reduced Wigner coefficients. In Eq. (A1), all couplings are multiplicity free and on the rhs the summation runs over $[m^{(1)}]_{n-1}$, $[m^{(2)}]_{n-1}$, and $[m^{(23)}]_{n-1}$.

For $[m^{(12)}]_{n-1}=[2\dot{0}]_{n-1}$, $[m^{(3)}]_{n-1}=[f-\Delta^{(1)}(k)]_{n-1}$, as well as for $[m^{(12)}]_{n-1}=[1\dot{0}]_{n-1}$, $[m^{(3)}]_{n-1}=[f]_{n-1}$, the sum contains a single term corresponding to $[m^{(1)}]_{n-1}=[m^{(2)}]_{n-1}=[1\dot{0}]_{n-1}$, $[m^{(23)}]_{n-1}=[f]_{n-1}$, and $[m^{(1)}]_{n-1}=[1\dot{0}]_{n-1}$, $[m^{(2)}]_{n-1}=[\dot{0}]_{n-1}$, $[m^{(23)}]_{n-1}=[f]_{n-1}$, respectively. Since all the reduced Wigner coefficients are fundamental ones and are given in Ref. 11 and since the $\mathrm{U}(n)$ and $\mathrm{U}(n-1)$ Racah coefficients can be determined from those calculated in the Appendix of Ref. 14 by using some symmetry properties[8,9] of Racah coefficients, it is straightforward to obtain the following results:

$$\begin{aligned}
&\langle [2\dot{0}]_n [2\dot{0}]_{n-1};[f]_n [f-\Delta^{(1)}(k)]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \\
&\quad = (-1)^{n-k-1}\Bigg[(f_i-f_k+k-i+2)(f_i-f_n+n-i-1) \\
&\qquad \times [(f_i-f_k+k-i+1)(f_i-f_n+n-i+1)(f_k-f_n+n-k-1)]^{-1} \\
&\qquad \times \left(\prod_{\substack{j=1\\ j\neq k}}^{n-1}(f_j-f_k+k-j+1)(f_j-f_k+k-j)^{-1}\right)\left(\prod_{\substack{j=1\\ j\neq i}}^{n-1}(f_j-f_n+n-j-1)(f_j-f_n+n-j)^{-1}\right)\Bigg]^{1/2}
\end{aligned} \tag{A2}$$

and

$$\begin{aligned}
&\langle [2\dot{0}]_n [1\dot{0}]_{n-1};[f]_n [f]_{n-1} \| [f+\Delta^{(2)}(i,n)]_n [f+\Delta^{(1)}(i)]_{n-1}\rangle \\
&\quad = (-1)^{n-1}\left[(f_i-f_n+n-i-1)(f_i-f_n+n-i+1)^{-1}\left(\prod_{\substack{j=1\\ j\neq i}}^{n-1}(f_j-f_n+n-j-1)(f_j-f_n+n-j)^{-1}\right)\right]^{1/2}.
\end{aligned} \tag{A3}$$

[1] J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984); J. Math. Phys. **26**, 2705 (1985); C. Quesne, J. Math. Phys. **27**, 428, 869 (1986).

[2] D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984); D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).

[3] J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982); J. Phys. A: Math. Gen. **17**, L405 (1984).

[4] O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985); O. Castaños, P. Kramer, and M. Moshinsky, *ibid.* **27**, 924 (1986).

[5] K. T. Hecht, *The Vector Coherent State Method and Its Application to Problems of Higher Symmetries* (Springer, Berlin, 1987).

[6] D. J. Rowe, R. Le Blanc, and K. T. Hecht, J. Math. Phys. **29**, 287 (1988).

[7] C. Quesne, J. Phys. A: Math. Gen. **23**, L43 (1990); "Vector coherent state theory of the non-compact orthosymplectic superalgebras: I. General theory" (submitted).

[8] K. T. Hecht, E. J. Reske, T. H. Seligman, and W. Zahn, Nucl. Phys. A **356**, 146 (1981).

[9] K. T. Hecht, R. Le Blanc, and D. J. Rowe, J. Phys. A: Math. Gen. **20**, 2241 (1987).

[10] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).

[11] L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).

[12] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).

[13] R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **20**, L681 (1987).

[14] R. Le Blanc and K. T. Hecht, J. Phys. A: Math. Gen. **20**, 4613 (1987).

Remarks on tensor operators

Dan Flath
Department of Mathematics, University of South Alabama, Mobile, Alabama 36688

(Received 11 October 1989; accepted for publication 7 March 1990)

The notion of an algebra of tensor operators for a simple Lie algebra is discussed. A model for the finite-dimensional irreducible representations of sl(4) is constructed. Explicit Wigner operators acting on the model are defined. Striking commutation properties for these operators are conjectured that resolve a sequence of nontrivial multiplicity problems.

I. INTRODUCTION

Evidence is presented that a theory of tensor operators of sl_4 can be built that is analogous to that found recently for sl_3 in Refs. 1 and 2.

Progress for sl_3 was based on a close study of commutators of sl_3 Wigner operators that revealed a simpler pattern than might have been expected. This paper initiates a similar study for sl_4. We present the first computational evidence that simple patterns exist for sl_4 as well. We remain far from a definitive result.

Section II is a mathematician's view of the first goal of a theory of tensor operators, namely the construction of a suitable "algebra of tensor operators" for each simple Lie algebra. The algebra will be an algebra of linear endomorphisms of a "model" for the irreducible finite-dimensional representations of the Lie algebra. (A model contains each isomorphism type with multiplicity one.) Certain of the tensor operators are called "Wigner operators." The problem is to find enough natural Wigner operators. It should be emphasized that this view is only a small part of a vast program initiated by Biedenharn that has regrettably attracted less attention from mathematicians than from physicists.

In Sec. III a model for the finite-dimensional irreducible representations of sl_4 is constructed. The model is realized as the null space of a family of polynomial differential operators. Two commutative algebras, each isomorphic to the shape algebra of sl_4, act on the model. Together they generate an algebra D, analogous to a Weyl algebra. I would guess that the desired algebra of tensor operators would be found within D. This suggests an avenue of investigation for sl_n.

Section IV presents the computational discoveries for sl_4. They are that the commutators of certain Wigner operators are themselves (unexpectedly) Wigner operators. Iteration of commutators thus produces a sequence of Wigner operators that, since they belong to spaces of multiplicity two, resolve a sequence of multiplicity problems. Since the Wigner operators are presented in Sec. IV as differential operators in 14 variables, the computations are too lengthy to be done by hand. The results of Sec. IV were both discovered and verified by massive computation on an IBM 3081 using the software REDUCE. It is in the nature of things that the computer suggests more than it proves; so I have had to phrase the results as conjectures with given computer evidence.

II. THE PROBLEM

Let g be a simple complex Lie algebra.

Choose a Cartan subalgebra h of g, and let P be the group of weights of (g,h). Let P_{++} be the set of dominant weights relative to some ordering on P that we fix once for all. For each $\lambda \in P_{++}$, let W_λ be a simple g module of highest weight λ. For each $\lambda \in P - P_{++}$, let $W_\lambda = (0)$. Let $W = \oplus_{\lambda \in P} W_\lambda$, so that W is a model for the finite-dimensional irreducible representations of g. We will write $g \subset \mathrm{End}_{\mathbf{C}}(W)$ (a slight abuse of notation) and view $\mathrm{End}_{\mathbf{C}}(W)$ as a g module via the adjoint action of g, the usual way.

We next define some terms that have become more or less standard in the physics literature but are unfamiliar to most mathematicians.

The elements of $\mathrm{End}_{\mathbf{C}}(W)$ are called *tensor operators.* Since every linear endomorphism of the model W is admitted as a tensor operator, the notion is not very restrictive. Part of the problem is to restrict it.

Let T be a tensor operator.

We say that T is a Δ-*shift operator* for $\Delta \in P$ iff $T(W_\lambda) \subset W_{\lambda+\Delta}$ for all $\lambda \in P$. It is a *shift operator* iff it is a Δ-shift operator for some $\Delta \in P$.

We say that T is a Λ *operator* for $\Lambda \in P_{++}$ iff T generates a simple g submodule of $\mathrm{End}_{\mathbf{C}}(W)$ with highest weight Λ or $T=0$. It is *irreducible* iff it is a nonzero Λ operator for some $\Lambda \in P_{++}$; thus, T is an irreducible operator if and only if T generates a finite-dimensional simple g submodule of $\mathrm{End}_{\mathbf{C}}(W)$.

We say that T is a *Wigner operator* iff T is an irreducible shift operator or $T=0$. We will say that T is of *type* $\binom{\Delta}{\Lambda}$ iff T is both a Δ-shift operator and a Λ operator. A *complete Wigner operator* is a simple g submodule of $\mathrm{End}_{\mathbf{C}}(W)$ that is generated by a nonzero Wigner operator. It carries the *type* of its nonzero elements.

Note, for example, that g itself is a complete Wigner operator, effecting shift $\Delta = 0$.

It is known that there exist nonzero Wigner operators of type $\binom{\Delta}{\Lambda}$ if and only if Δ is a weight of W_Λ.

The first problem in the subject is that of constructing a suitable subalgebra A of $\mathrm{End}_{\mathbf{C}}(W)$. (The literature suggests that once A is got right, physicists would restrict the word "tensor operator" to mean an element of A.) The vague word "suitable" is not defined, but can be taken to include the

 0022-2488/90/071605-05$03.00

following points.

(i) A should contain g (and hence contain a copy of the universal enveloping algebra of g.)

(ii) Every endomorphism of every finite-dimensional subspace of W should lift to an element of A.

(iii) A should have a basis consisting of Wigner operators. Equivalently, A should be a direct sum of complete Wigner operators.

(iv) The commutant of g in A, which is the (commutative) algebra of elements of A that act as scalar multiplications on each W_λ, should be manageable. Probably it or a natural quotient of it should be isomorphic to a polynomial ring in rank(g) variables. (It can be extended as needed.) Note that this commutative algebra can be described as the algebra of all Wigner operators in A that are of type $\binom{0}{0}$.

(v) Let $B\binom{\Delta}{\Lambda}$ be the space of Wigner operators in A that are of type $\binom{\Delta}{\Lambda}$ and of g weight Λ. Then $B\binom{\Delta}{\Lambda}$ should be a free $B\binom{0}{0}$ module of rank equal to the multiplicity of the weight Δ in the representation W_Λ. This property reflects the fact that the multiplicity of W_Λ as a g submodule of $\mathrm{Hom}_{\mathbf{C}}(W_\mu, W_{\mu+\Delta})$, viewed as a function of μ for fixed Λ and Δ, attains a maximum equal to the multiplicity of the weight Δ in W_Λ.

(vi) The multiplication law in A should be surprisingly simple. Though this is vague, it is important. I do not know how to be precise here. I will just give an example of the kind of phenomenon I am looking for. Let U, V, be complete Wigner operators of types $\binom{\delta}{\lambda}$, $\binom{\Delta}{\Lambda}$. Then the g module $[U,V]$ that is spanned by the commutators of elements of U and V need not be simple. It may contain any or all irreducible constituents of $W_\lambda \otimes W_\Lambda$ for which $\delta+\Delta$ is a weight. Nevertheless, for many U and V in A, the module $[U,V]$ should be smaller than the previous sentence might suggest. This is a "property" possessed also by the universal enveloping algebra of g.

The only cases I know of for which such an algebra A has been constructed are those of $g=\mathrm{sl}_2$ and $g=\mathrm{sl}_3$. For $g=\mathrm{sl}_2$ there are several candidates for A, including a Weyl algebra and some quotients of the universal enveloping algebra of sl_3. The main result of Refs. 1 and 2 is the construction for $g=\mathrm{sl}_3$ of an A that is a quotient of the universal enveloping algebra of so_8.

The common feature in these constructions is that the generators for A are represented by nice formulas, essentially polynomial differential operators, acting on the model W, which is a subspace of the symmetric algebra of the sum of the fundamental representations of g. This motivates the construction of the model for the representations of sl_4 that is presented in Sec. III of the present paper.

It is possible that A can be generated by a well-known Lie algebra that has a basis of Wigner operators, but I would expect such a Lie algebra to be infinite dimensional for $g=\mathrm{sl}_n$, $n\geqslant 4$.

In all my work I have started with the vague property (vi) above. In Sec. IV of this paper I will present computational evidence that something can be done with property (vi) in the case $g=\mathrm{sl}_4$. More precisely, I will find some pairs of complete Wigner operators U, V for which $[U,V]$ is (unexpectedly) a simple g module (and is hence a complete Wigner operator.)

III. A MODEL FOR THE REPRESENTATIONS OF sl_4

Henceforth we will be concerned only with the Lie algebra $g=\mathrm{sl}_4$ of 4×4 complex matrices with trace equal to 0. Denote by h the Cartan subalgebra of diagonal matrices in g and by n the subalgebra of strictly upper triangular matrices in g. We write E_{ij} for the matrix all of whose entries equal 0 except for the ijth which equals 1.

The group P of weights of (g,h) will be identified with $\mathbf{Z}^4/\langle(1,1,1,1)\rangle$ as follows: For $\lambda=(m_1m_2m_3m_4)\in P$ and $H=\Sigma h_iE_{ii}\in h$, define $\lambda(H)=\Sigma m_ih_i$. Take the ordering on P determined by the Borel subalgebra $b=h\oplus n$. Then $(m_1m_2m_3m_4)\in P_{++}$ if and only if $m_i\geqslant m_{i+1}$ for $1\leqslant i\leqslant 3$.

The principal result of this paper is the production of complete Wigner operators V_n for sl_4 of type

$$\begin{pmatrix} n & 0 & n & n\\ n+1 & n & n-1 & 0\end{pmatrix}$$

for $n\geqslant 1$ such that $[V_m,V_n]=V_{m+n}$.

All g modules that concern us will be spaces of polynomials and of polynomial differential operators. For convenience, we adopt the notation $\bar{x}$ to indicate the formal differential operator $\partial/\partial x$ with respect to a variable x. Thus, for example, $[\bar{x},x]=1$ and $[\bar{x},y]=0$.

We first produce a model for the finite-dimensional irreducible representations of g. Begin with the defining g module $W^1=\mathbf{C}^4$, which has highest weight $\varpi_1=(1000)$. Let $W_2=\wedge^2W_1$ and $W_3=\wedge^3W_1$, simple g modules whose highest weights are the other two fundamental dominant weights $\varpi_2=(1100)$ and $\varpi_3=(1110)$. Let $C=\mathrm{Sym}(W_1\oplus W_2\oplus W_3)$.

Let $\{a_i\}_{1\leqslant i\leqslant 4}$ be the standard basis of W_1. The g action on W_1 can be described by the formula $E_{ij}=a_i\bar{a}_j$ for $i\neq j$. Write a_{ij} for $a_i\wedge a_j$, $1\leqslant i<j\leqslant 4$, and write a_{ijk} for $a_i\wedge a_j\wedge a_k$, $1\leqslant i<j<k\leqslant 4$. The space C is nothing but the polynomial ring in the 14 independent commuting variables a_i, a_{ij}, a_{ijk}. It is easy to work out explicitly the action of g on this ring. It is through polynomial differential operators as follows:

$$E_{ij}=a_i\bar{a}_j+a_{ik}\bar{a}_{jk}+a_{il}\bar{a}_{jl}+a_{ikl}\bar{a}_{jkl},$$

for $ijkl$ a permutation of 1234,

where for I' a permutation of I whose sign is ϵ, we let $a_{I'}=\epsilon a_I$.

Let $\lambda=(mnp0)\in P_{++}$, and let $x_\lambda=a_1^{m-n}a_{12}^{n-p}a_{123}^p\in C$. Let W_λ be the g submodule of C that is generated by x_λ. Clearly x_λ is primitive (i.e., annihilated by n) of weight λ, so that W_λ is a simple g module of highest weight λ.

Let $W=\oplus_{\lambda\in P_{++}}W_\lambda$, so that W is a model for the irreducible representations of g.

There are two naturally occurring commutative algebras of linear endomorphisms of W that I want to describe next.

Define the ten *shuffle polynomials* in C as follows.

$$\begin{aligned}
A &= -a_{12}a_{34} + a_{13}a_{24} - a_{23}a_{14},\\
B_1 &= a_1a_{23} - a_2a_{13} + a_3a_{12},\\
B_2 &= a_1a_{24} - a_2a_{14} + a_4a_{12},\\
B_3 &= a_1a_{34} - a_3a_{14} + a_4a_{13},\\
B_4 &= a_2a_{34} - a_3a_{24} + a_4a_{23},\\
C_1 &= a_{12}a_{134} - a_{13}a_{124} + a_{14}a_{123},\\
C_2 &= a_{12}a_{234} - a_{23}a_{124} + a_{24}a_{123},\\
C_3 &= a_{13}a_{234} - a_{23}a_{134} + a_{34}a_{123},\\
C_4 &= a_{14}a_{234} - a_{24}a_{134} + a_{34}a_{124},\\
D &= a_1a_{234} - a_2a_{134} + a_3a_{124} - a_4a_{123}.
\end{aligned}$$

Both A and D are annihilated by g. The four B_i and the four C_i each span four-dimensional g submodules of C. The shuffle terminology derives from Ref. 3, where a combinatorial description of the shuffle polynomials can be found.

Let I be the ideal in C generated by the ten shuffle polynomials. It is a g submodule of C.

Let $\overline{C}$ be the polynomial ring in the 14 commuting variables $\bar{a}_i$, $\bar{a}_{ij}$, $\bar{a}_{ijk}$. Define the *shuffle operators* $\overline{A}$, $\overline{B}_i$, $\overline{C}_i$, $\overline{D}$ to be the elements of $\overline{C}$ derived from the shuffle polynomials by replacing the as by $\bar{a}$s. Let $\overline{I}$ denote the ideal in $\overline{C}$ that is generated by the ten shuffle operators.

Proposition 1: The ideal I is a vector space complement to the model W in C. In other words, $C = W \oplus I$.

Proposition 2: $W = \{ q \in C \mid \bar{f}q = 0 \text{ for all } \bar{f} \in \overline{I}\}$.

Proposition 2 is very useful because the definition of W was so indirect.

Proposition 1 goes back at least to Hodge.[4] The quotient C/I, which is the coordinate ring of an affine flag variety, has been called the shape algebra in Refs. 5 and 6. Proposition 2 may be known, too, but I have been unable to trace it in the literature.

From Propositions 1 and 2 we can derive actions of the (isomorphic) quotient rings C/I and $\overline{C}/\overline{I}$ on W. Let D be the subalgebra of $\mathrm{End}_{\mathbf{C}}(W)$ that is generated by C/I and $\overline{C}/\overline{I}$. I would suggest that D merits close study. The algebra D, or something very close to it, may provide the suitable algebra A of Sec. II. Since shuffle operators and the natural analog of Proposition 1 are known for sl_n, one can formulate Proposition 2 and try to construct D for sl_n. For sl_2, D will equal the A of Ref. 7.

There is a subalgebra of D that also deserves attention, namely F, the algebra of endomorphisms of W that are restrictions to W of polynomial differential operators d in the variables a_i, a_{ij}, a_{ijk} such that $d(W) \subset W$. For sl_2 there is equality $F = D$, but this is not so for sl_3. The precise relationship of D to F is worth determining. Nearly all my calculations have taken place within F. F contains the A for sl_3 constructed in Ref. 2.

We turn to the proofs of the propositions. They will be based upon an explicit determination of the algebra of primitive vectors in C.

Lemma: Let $X_1 = a_1$, $X_2 = a_{12}$, $X_3 = a_{123}$, $X_4 = A$, $X_5 = B_1$, $X_6 = C_1$, $X_7 = D$, $X_8 = a_2C_1 - a_1C_2$.

(i) $\bar{f}(q) = 0$ for all $\bar{f} \in n$ and $q \in \mathbf{C}[X]$.

(ii) The X_i are algebraically independent elements of C.

(iii) C is generated as a g module by $\mathbf{C}[X]$; $\mathbf{C}[X] = \{ q \in C \mid nq = 0\}$.

Proof: Part (i) follows from the computation $E_{jk}(X_i) = 0$ for $j<k$ and $1 \leqslant i \leqslant 8$.

Each of the formulas for the polynomials in the sequence $X_1, X_2, X_3, X_5, X_6, X_8, X_7, X_4$ involves at least one variable not present in any of the earlier formulas; one may take the sequence of new variables to be a_1, a_{12}, a_{123}, a_{23}, a_{14}, a_{234}, a_4, a_{34}. Part (ii) follows.

In view of (i), the two statements of part (iii) are equivalent. In view of (ii) they can be proved by showing that the dimension of the space of homogeneous polynomials belonging to C of fixed degree d equals the sum of the dimensions of the simple g modules generated by the monomials in the X_i that are of degree d in the a variables. We omit the details.

Proof of Proposition 1: $\mathbf{C}[X] \subset I + W$, and $I + W$ is a g submodule of C (because I and W are g submodules.) Therefore, by part (iii) of the lemma, $I + W = C$. We will show that $I \cap W = (0)$ by constructing a homomorphism ϕ with domain C such that $I \subset \ker(\phi)$ and $W \cap \ker(\phi) = (0)$.

Let $\mathbf{C}[y]$ be the polynomial ring in the twelve variables y_{ik}, $1 \leqslant i \leqslant 4$, $1 \leqslant k \leqslant 3$. We give $\mathbf{C}[y]$ the structure of g module via the formulas $E_{ij} = \Sigma_{k=1}^{3} y_{ik}\bar{y}_{jk}$ for $i \neq j$.

Define an algebra homomorphism $\phi: C \to \mathbf{C}[y]$ as follows:

$$\phi(a_i) = y_{i1},$$

$$\phi(a_{ij}) = \det \begin{pmatrix} y_{i1} & y_{i2} \\ y_{j1} & y_{j2} \end{pmatrix},$$

$$\phi(a_{ijk}) = \det \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} \\ y_{j1} & y_{j2} & y_{j3} \\ y_{k1} & y_{k2} & y_{k3} \end{pmatrix}.$$

Since ϕ is a g map, its kernel must be a g submodule of C. By Sylvester's identities, $\phi(X_i) = 0$ for $i = 4, 5, 6, 7$ and so $I \subset \ker(\phi)$. Because ϕ is nonzero on each monomial in the first three X_i alone, $W_\lambda \cap \ker(\phi) = (0)$ for all $\lambda \in P_{++}$. Therefore, $W \cap \ker(\phi) = (0)$.

Proof of Proposition 2: To show that $W \subset \{ q \in C \mid \overline{I}q = (0)\}$, it suffices to show that $\{ q \in C \mid \overline{I}q = (0)\}$ is a g submodule of C that contains all the monomials $a_1^m a_{12}^n a_{123}^p$. This is so because the shuffle operators, which generate $\overline{I}$, span a g subspace of $\overline{C}$ and are easily seen to annihilate all the $a_1^m a_{12}^n a_{123}^p$.

Now consider the reverse inclusion, $\{ q \in C \mid \overline{I}q = (0)\} \subset W$. Since both sides are g modules, it will suffice to prove that $\{ q \in C \mid \overline{I}q = (0)$ and $nq = (0)\} \subset \{ q \in W \mid nq = (0)\}$. By the lemma and the definition of W, the inclusion to be proved is $\{ q \in \mathbf{C}[X] \mid \overline{I}q = (0)\} \subset \mathbf{C}[X_1, X_2, X_3]$.

Let $q \in \mathbf{C}[X]$ and suppose that q is annihilated by all ten shuffle operators. We must prove that q is a polynomial in the variables X_1, X_2, X_3 alone.

The elements of $\overline{I}$ act on $\mathbf{C}[X]$ as differential operators in the X_i with coefficients in C. They can be worked out explicitly. We present four of the formulas that will get the proof started. In each case I describe an element of $\overline{I}$ by giving its effect on the monomial $\Pi_{i=1}^{8} X_i^{m_i}$. We let $\overline{E}_6 = \bar{a}_4\overline{C}_3 - \bar{a}_3\overline{C}_4 \in \overline{I}$.

$\overline{\mathrm{A}}:(m_2+m_4+m_5+m_6+m_8+2)\overline{X}_4-X_1X_3\overline{X}_5\overline{X}_6,$
$\overline{\mathrm{C}}_3:B_2\overline{X}_4\overline{X}_7-a_{124}X_1\overline{X}_6\overline{X}_7-X_2(X_1\overline{X}_5\overline{X}_6+\overline{X}_3\overline{X}_4),$
$\overline{\mathrm{C}}_4:X_1X_3\overline{X}_6\overline{X}_7-X_5\overline{X}_4\overline{X}_7,$
$\overline{\mathrm{E}}_6:((m_3+m_5+m_7+1)X_2+X_8\overline{X}_7)\overline{X}_4\overline{X}_7.$

The equation $\overline{\mathrm{E}}_6(q)=0$ can be shown to imply that $\overline{X}_4\overline{X}_7(q)=0$. It follows immediately, then, from $\overline{\mathrm{C}}_4(q)=0$, that $\overline{X}_6\overline{X}_7(q)=0$. Next, the equation $\overline{\mathrm{C}}_3(q)=0$ implies that $-X_1\overline{X}_5\overline{X}_6(q)=\overline{X}_3\overline{X}_4(q)$, which combines with $\overline{\mathrm{A}}(q)=0$ to prove that $\overline{X}_4(q)=\overline{X}_5\overline{X}_6(q)=0$. Thus the variable X_4 does not occur in q.

The rest of the analysis is similar. The equations $\overline{\mathrm{C}}_2(q)=0$ and $\overline{\mathrm{D}}(q)=0$ now imply that X_7 and X_8 are absent from q as well. At this point we know that q is a polynomial in X_1, X_2, X_3, X_5, X_6 and that $\overline{X}_5\overline{X}_6(q)=0$. It is an easy matter to use $\overline{\mathrm{C}}_1$ to prove that $\overline{X}_6(q)=0$ and to use $\overline{\mathrm{B}}_1$ to prove that $\overline{X}_5(q)=0$.

IV. THE DISCOVERY

We can now present the main construction of this paper. Let

$$v_1=(1+G_1)a_1\bar a_{34}+M\bar a_{34}-N\bar a_{234},$$

where

$$\begin{aligned}G_1&=a_1\bar a_1+a_2\bar a_2+a_3\bar a_3+a_4\bar a_4,\\ M&=a_{12}(a_1\bar a_{12}-a_3\bar a_{23}-a_4\bar a_{24})\\ &\quad+a_{13}(a_1\bar a_{13}+a_2\bar a_{23}-a_4\bar a_{34})\\ &\quad+a_{14}(a_1\bar a_{14}+a_2\bar a_{24}+a_3\bar a_{34}),\\ N&=a_{123}(a_1\bar a_{13}+a_2\bar a_{23}-a_4\bar a_{34})\\ &\quad+a_{124}(a_1\bar a_{14}+a_2\bar a_{24}+a_3\bar a_{34}).\end{aligned}$$

After verifying that $[\bar I, v_1]$ lies in the left ideal of $\mathrm{End}_{\mathbf{C}}(C)$ that is generated by $\bar I$, we can assert (by Proposition 2) that $v_1(W)\subset W$ and therefore write $v_1\in F\subset \mathrm{End}_{\mathbf{C}}(W)$. A computation shows that v_1 commutes with n and is a Wigner operator of type

$$\begin{pmatrix}1&0&1&1\\2&1&0&0\end{pmatrix}.$$

Let V_1 be the g submodule of $\mathrm{End}_{\mathbf{C}}(W)$ that is generated by v_1. Define $V_n=[V_1, V_{n-1}]$ and $v_n=[E_{21}v_1, E_{32}v_{n-1}]$ for all $n>1$. Note that $v_n\in V_n$.

Conjecture:

(1) V_n is a complete Wigner operator of type

$$\begin{pmatrix}n&0&n&n\\n+1&n&n-1&0\end{pmatrix},$$

and v_n is a vector of highest weight, for all $n\geqslant 1$,

(2) $[V_m, V_n]=V_{m+n}$, all $m, n\geqslant 1$.

(3) $[E_{21}v_m, E_{32}v_n]=[E_{21}v_n, E_{32}v_m]$, for all $m, n\geqslant 1$.

I have verified on the computer using REDUCE conjecture 1 for $n\leqslant 5$, conjectures 2 and 3 for $m+n\leqslant 4$ and $(m, n)=(1,4)$.

It follows from the conjectures that $[E_{21}v_m, E_{32}v_n]=c_{mn}v_{m+n}$ for some $c_{mn}\neq 0\in\mathbf{Q}$. By definition $c_{1n}=1$; and the conjectures imply that $c_{mn}=c_{nm}$. I have verified that $c_{22}=1$.

The conjectures resolve a sequence of multiplicity problems in the following way. Fix n, and let $\lambda_n=(n+1, n, n-1, 0)\in P_{++}$. Let $\Delta=(1,0,1,1)\in P$. Since the multiplicity of the weight $n\Delta$ in the representation W_{λ_n} is 2, the multiplicity of W_{λ_n} in $\mathrm{Hom}_{\mathbf{C}}(W_\mu, W_{\mu+n\Delta})$ is equal to 2 for generic μ. Thus it is not at all clear how to find an *irreducible* subspace of $\mathrm{Hom}_{\mathbf{C}}(W_\mu, W_{\mu+n\Delta})$ that has highest weight λ_n. I propose taking V_n. A choice of V_1 for $n=1$ had to be made, but the other V_n are then generated automatically.

We proceed to describe the computer verifications of the conjectures.

For $n\geqslant 1$, let U_n be a simple g module of highest weight $(n+1, n, n-1, 0)$. We ask for those simple g modules of $U_m\otimes U_n$ that have $(m+n, 0, m+n, m+n)$ as a weight, since they and only they could potentially arise in the commutator $[V_m, V_n]$. The list is not long and is independent of (m, n). Let u, v be highest weight vectors in U_m, U_n. We list the five relevant dominant weights and the six corresponding primitive elements w_i of $U_m\otimes U_n$:

(i) $(m+n+2, m+n, m+n-2, 0)$, $w_1=u\otimes v$.
(ii) $(m+n+1, m+n+1, m+n-2, 0)$, $w_2=E_{21}u\otimes v-u\otimes E_{21}v$.
(iii) $(m+n+2, m+n-1, m+n-1, 0)$, $w_3=E_{32}u\otimes v-u\otimes E_{32}v$.
(iv) $(m+n+1, m+n, m+n-1, 0)$

$$w_s=-3(E_{21}u\otimes E_{32}v+E_{32}u\otimes E_{21}v)+(E_{32}E_{21}u\otimes v+u\otimes E_{32}E_{21}v)+(E_{21}E_{32}u\otimes v+u\otimes E_{21}E_{32}v).$$

$$w_a=(E_{21}u\otimes E_{32}v-E_{32}u\otimes E_{21}v)-(E_{32}E_{21}u\otimes v-u\otimes E_{32}E_{21}v)+(E_{21}E_{32}u\otimes v-u\otimes E_{21}E_{32}v).$$

(v) $(m+n, m+n, m+n, 0)$

$$\begin{aligned}w_5=&-3(u\otimes E_{32}E_{21}E_{32}E_{21}v+E_{32}E_{21}E_{32}E_{21}u\otimes v)+3(E_{21}u\otimes E_{32}E_{21}E_{32}v+E_{32}E_{21}E_{32}u\otimes E_{21}v)\\ &+3(E_{21}E_{32}E_{21}u\otimes E_{32}v+E_{32}u\otimes E_{21}E_{32}E_{21}v)+(E_{32}E_{21}u\otimes E_{21}E_{32}v+E_{21}E_{32}u\otimes E_{32}E_{21}v)\\ &-2E_{32}E_{21}u\otimes E_{32}E_{21}v-2E_{21}E_{32}u\otimes E_{21}E_{32}v.\end{aligned}$$

Now make the substitutions $u \mapsto v_m$ and $v \mapsto v_n$. Then conjectures 1 and 2 amount to the assertions that under the maps $V_m \otimes V_n \to \mathrm{End}_{\mathbf{C}}(W)$, $\alpha \otimes \beta \mapsto [\alpha, \beta]$, we have $w_1 = w_2 = w_3 = w_s = w_5 = 0$ and $w_a = \lambda v_{m+n}$ for some $\lambda \neq 0 \in \mathbf{C}$. Note that in the case $m = n$, then w_1, w_s, and w_5 are zero automatically because they are symmetric tensors.

Further, if $w_1 = 0$, then

$$w_2 = 2E_{21}u \otimes v,$$

$$w_3 = 2E_{32}u \otimes v,$$

and

$$w_s = -5(E_{21}u \otimes E_{32}v + E_{32}u \otimes E_{21}v).$$

If

$$w_1 = w_2 = w_3 = w_s = 0, \text{ then}$$

$$w_a = 6E_{21}u \otimes E_{32}v = -6E_{32}u \otimes E_{21}v$$

and

$$w_5 = -12(u \otimes (E_{32}E_{21})^2 v + (E_{32}E_{21})^2 u \otimes v).$$

I simply checked on the computer that

$$[u, v] = 0, \quad [E_{21}u, v] = 0, \quad [E_{32}u, v] = 0,$$

$$[E_{21}u, E_{32}v] + [E_{32}u, E_{21}v] = 0,$$

$$[u, (E_{32}E_{21})^2 v] + [(E_{32}E_{21})^2 u, v] = 0.$$

It would have been sufficient to verify only the second, third, and fifth of these last five equations.

[1] L. C. Biedenharn and D. Flath, Commun. Math. Phys. **93**, 143 (1984).
[2] D. Flath, and L. C. Biedenharn, Can. J. Math. **37**, 710 (1985).
[3] K. Baclawski, SIAM J. Alg. Disc. Meth. **3**, 592 (1982).
[4] W. V. D. Hodge, Proc. Cambridge Philos. Soc. **39**, 22 (1943).
[5] J. Towber, J. Algebra **47**, 80 (1977).
[6] J. Towber, J. Algebra **61**, 414 (1979).
[7] D. Flath, L'Enseignement Math. **29**, 339 (1983).

Moment invariants for the Vlasov equation

D. D. Holm
Theoretical Division, MS-B284, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
W. P. Lysenko
Accelerator Technology Division, MS-H808, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
J. C. Scovel
Computer Division, MS-B265, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 2 October 1989; accepted for publication 31 January 1990)

Moment invariants [functions of the moments of a Vlasov distribution that are invariant under Sp(6)] are classified using Young diagrams. The connection between the moment invariants and the Poincaré invariants is established. An application using the moment invariants as phase space coordinates is considered for a matching section in a particle-beam accelerator, and a Lie–Poisson numerical integration algorithm for the moment dynamics is proposed.

I. INTRODUCTION

For particle beams accurately described by the Vlasov equation, the moment description is useful since the moments correspond to laboratory quantities. For example, in 1D, the moments consisting of the beam centroid $\langle q\rangle$ and the rms width Sqrt($\langle q^2\rangle$) are measurable quantities, where the brackets denote integration against the particle distribution. The rms emittance Sqrt ($\langle q^2\rangle\langle p^2\rangle - \langle pq\rangle^2$) is invariant under linear symplectic motions. In Lysenko and Overley,[1] generalizations of the rms emittance in 1, 2, and 3 D are presented, which are also moment invariants (i.e., also functions of the moments that are invariant under linear motions). In addition, they find an infinite number other moment invariants and suggest, in the context of a matching section, that these would provide useful variables for describing beam dynamics.

In this paper we use the Lie–Poisson structure of the Vlasov equation as discussed in Marsden *et al.*,[2] and show the moments to be a projection of the particle distribution determining Lie–Poisson dynamics dual to a subalgebra or quotient algebra of the algebra corresponding to the group of symplectomorphisms. The new phase space for the Vlasov equation is the symmetric tensor algebra over single particle phase space. In this setting a complete list of tensor invariants is provided, which explains and extends the moment invariants found in Ref. 1. Hilbert's theorem on polynomial ideals shows[3] that the moment invariants depending on moments of order k or less are finitely generated, suggesting a numerical integration algorithm that uses the generators as beam coordinates.

II. THE MOMENTS

Let $Z=R^3\times R^3$ with usual symplectic two-form $\omega=\Sigma_{i,j}J_{ij}\,dz^i\wedge dz^j$, where

$$J=\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix}$$

is the symplectic matrix. Let μ be a distribution that rapidly decays at infinity, and f a real-valued function on Z. Let φ be a symplectic map ($\varphi^*\omega=\omega$) and define $\mathrm{Ad}_\varphi f=f\cdot\varphi$ and $\varphi^*\mu(z)=\mu(\varphi(z))$ for $z\in Z$.

Define

$$\langle f\rangle_\mu=\int f\mu.$$

The change of variables formula is $\langle\mathrm{Ad}_\varphi f\rangle_{\varphi^*\mu}=\langle f\rangle_\mu$ so that

$$\langle\mathrm{Ad}_\varphi f\rangle_\mu=\langle\mathrm{Ad}_\varphi f\rangle_{\varphi^*\varphi^{*-1}\mu}=\langle f\rangle_{\varphi^{*-1}\mu}=\langle f\rangle_{(\varphi^{-1})^*\mu},$$

and $\langle f\rangle_\mu$ can be thought of as a tensor in f with μ fixed or a tensor in μ with f fixed that transforms contragradiently. For a collection of functions f^i, their product $\Pi_i f^i$ transforms according to $\mathrm{Ad}_\varphi(\Pi_i f^i)=\Pi_i(\mathrm{Ad}_\varphi f^i)$, so that $\langle\Pi_i f^i\rangle_\mu$ transforms as a tensor product under the Kronecker representation of Ad_φ.[3]

Consider those f's that are *linear* functions ($\in Z^*$) and φ in the linear subgroup Sp(6). Then μ defines a linear functional X^k_μ: $S^k(Z^*)\to R$ given by

$$X^k_\mu(f^1\vee f^2\vee\cdots\vee f^k)=\int f^1\cdots f^k\mu$$

on decomposable elements, and hence an element $X^k\in S^k(Z^*)^*=S^k(Z)$. Here, $S^k(Z)$ denotes the symmetric k tensors and $f^1\vee f^2$ the symmetric tensor product, cf. Ref. 4.

Here, $X^k\in S^k(Z)$ transforms contragradiently to those in $S^k(Z^*)$ and so transforms under the Kronecker representation $(\otimes S)^k$ for $S\in\mathrm{Sp}(6)$. Then, X^k comprises the k th-order moments of the distribution μ. For the classical moments, choose a basis z^i, $i=1,\ldots 6$ of Z^*. Then, the basis elements of $S^k(Z^*)$ are of the form $z^{i_1}\vee\cdots\vee z^{i_k}$ and the value of X^k_μ on this element is its component

$$X^{i_1,\cdots i_k}=\int z^{i_1}\cdots z^{i_k}\mu$$

a k th-order moment. In particular, $X^1\in Z$ is the center of mass.

The space of moments of signed distributions is the symmetric tensor algebra $S\equiv S(Z)=\oplus_{k>1}S^k(Z)$, and we refer to a point $X=\oplus_{k>1}X^k\in S$ as a moment or its moments. In all that follows, when a tensor has one index, that index indicates the tensor's rank. Otherwise, the notation is standard. The brackets $\langle\rangle_\mu$ indicate integration against a distribution while $\langle,\rangle$ denotes the pairing between a space and its dual such as between contravariant and covariant tensors. Distributions are assumed to be non-negative in that integration over any subdomain should be non-negative; otherwise the distribution functions are referred to as signed distributions. Einstein's summation convention is used throughout.

Haviland[5] demonstrates necessary and sufficient conditions for a given set of moments to have arisen from a (non-negative) distribution. Let $P \in S(Z^*)$ correspond to the formal power series $\Sigma \langle P_k, \otimes_k z\rangle$. The summation is over the rank, k, going from 1 to infinity while each term is $\langle P_k, \otimes_k z\rangle = P_{i_1 \cdots i_k} z^{i_1} \cdots z^{i_k}$. Let $P_n^X = \Sigma_{k<n} \langle P_k, X^k\rangle$ be the truncated sum. Here, X is said to be non-negative if $P_n^X \geqslant 0$ whenever $\Sigma_{k<n} \langle P_k, \otimes_k z\rangle \geqslant 0$ for all z in Z and all n. Haviland shows that given a set of moments X, for there to be a corresponding distribution with these moments, it is necessasry and sufficient that X be non-negative. In addition, he establishes sufficient conditions for the distribution to be uniquely determined from a set of moments.

The moments come about via a general construction (Marsden–Weinstein reduction, e.g., Ref. 2) that justifies their consideration as phase space. Consider the imbedding $i: \mathscr{L} \to \mathscr{G}$ of a Lie subalgebra $\mathscr{L}$ of the algebra $\mathscr{G}$. The adjoint of this map $i^*: \mathscr{G}^* \to \mathscr{L}^*$ is a projection and a momentum map. For the Vlasov equation $\mathscr{G} = C^\infty(R^3 \times R^3)$ and $\mathscr{G}^*$ is the signed distributions. If we choose $\mathscr{L}$ to be the polynomial algebra and $\mu \in \mathscr{G}^*$, then $i^*(\mu)$ are the moments. In particular, $i^*(\mu)$ will enjoy contragradiently any symmetries that $\mathscr{L}$ has. Under the isomorphism

$$\mathscr{G}^* \approx \mathscr{L}^* \oplus \mathscr{G}^*/\mathscr{L}^*,$$

vector fields determined by polynomials in $\mathscr{L}$ have a trivial projection onto the second factor. Therefore the subalgebra $\mathscr{L}$ is appropriate when the vector field can be well approximated by a polynomial one. On the other hand, if $\mathscr{E}$ is an ideal of $\mathscr{G}$, then the quotient algebra $\mathscr{G}/\mathscr{E} = \mathscr{L}$ is a Lie algebra. This remark has found applications in Ref. 6 for Lie–Poisson integration of particle-beam models. Consider the canonical projection $P: \mathscr{G} \to \mathscr{G}/\mathscr{E} = \mathscr{L}$ with adjoint imbedding $P^*: \mathscr{L}^* \to \mathscr{G}^*$. If $\mu \in \mathscr{G}^*$ is near the image of $\mathscr{L}^*$ under P^*, then $\mathscr{L}$ can initially be thought of as a good approximation. However, the full dynamics might take μ away from $P^*\mathscr{L}^*$ in which case we refer to the approximation $\mathscr{L}$ as unstable. For example, let $\mathscr{E}$ be the exponentially flat functions at the origin in Z, so that $\mathscr{G}/\mathscr{E}$ is the power series algebra. Suppose that $\mu \in \mathscr{G}^*$ has support in $|z| \leqslant \epsilon$. Then, for $e \in \mathscr{E}$, $|\langle \mu, e\rangle|$ is exponentially small so that initially $P^*\mu \approx \mu$. However this will not remain so if the vector field moves points near the origin too far. We refer to the first case as a global approximation and the latter as a local one.

The general problem is to construct a momentum map (for example, the map from distributions to moments) so that the Hamiltonian is approximately collective (in the sense of Guillemin and Sternberg[7]). In the case of the moments, this implies that the Hamiltonian can be well approximated by functions of the moments.

III. MOMENT DYNAMICS

We begin by describing the Lie–Poisson structure of the Vlasov equation discussed in Ref. 2. Consider the group of symplectomorphisms of $Z, \mathscr{G} = C^\infty(Z)$, its Lie algebra, with the dual space $\mathscr{G}^*$ being the signed distributions. The space $C^\infty(\mathscr{G}^*)$ is a Poisson algebra in that it is a Lie algebra such that the bracket is a derivation. Let $F,H \in C^\infty(\mathscr{G}^*)$, $\mu \in \mathscr{G}^*$. Then $\{F,H\}(\mu) \equiv \langle \mu, [\partial F/\partial \mu, \partial H/\partial \mu]\rangle$ (where $[,]$ is the bracket in $\mathscr{G}$) determines the Lie–Poisson bracket $\{,\}$ on $C^\infty(\mathscr{G}^*)$. The Lie–Poisson system $\dot\mu = -\mathrm{ad}^*_{\partial H/\partial\mu}\mu$ becomes $\dot\mu = -\{\mu, \mathscr{H}\}$ under the identification $\mu \to \mu d\mathrm{Vol.}$, where

$$H(\mu) = \int \mathscr{H}(\mu)\mu$$

is the Vlasov Hamiltonian. The use of the bracket in $\mathscr{G}$ requires the identification $\mathscr{G}^{**} \approx \mathscr{G}$. In this case $\mathscr{G}^{**}$ is surely much larger than $\mathscr{G}$ and it is not clear what $C^\infty(\mathscr{G}^*)$ should be. We try to resolve this problem by going to the formal power series algebra. Questions of topology will be ignored here but the topology of linear compactification could be introduced as in Guillemin[8] and would still maintain many of the important results of classical Lie algebra theory. In this spirit, consider a graded algebra $\mathscr{L} = \oplus \mathscr{L}_i$, where each summand is finite dimensional. Such a direct sum is a graded Lie algebra if $[\mathscr{L}_i, \mathscr{L}_j] \subset \mathscr{L}_{i+j-e}$, where e is some constant. The subalgebra $\mathscr{L}_e$ will be referred to as the core subalgebra. Let $S(\mathscr{L}) = \oplus S^n(\mathscr{L})$, where the symmetric n tensors of $S^n(\mathscr{L})$ are to be identified with homogeneous functions of homogeneity n on $\mathscr{L}^*$. Let μ be in $\mathscr{L}^*$ and $s \in S^n(\mathscr{L})$. Then the homogeneous function on $\mathscr{L}^*$ corresponding to s is defined by $\bar{s}(\mu) = \langle s, \otimes_n \mu\rangle$. To write with indices, let e_i be a basis of $\mathscr{L}$ determining a dual basis. In terms of these basis elements s has coefficients $s^{i_1 \cdots i_n}$, μ has coefficients μ_i, and $\bar{s}(\mu) = s^{i_1 \cdots i_n}\mu_{1_1} \cdots \mu_{i_n}$. The bracket determined by reduction satisfies

$$\{\bar{e}_i, \bar{e}_j\}(\mu) = \left\langle \mu, \left[\frac{\partial \bar{e}_i}{\partial \mu}, \frac{\partial \bar{e}_j}{\partial \mu}\right]\right\rangle,$$

but since $\partial \bar{e}_i/\partial\mu = e_i$ this becomes $\{\bar{e}_i, \bar{e}_j\} = \overline{[e_i, e_j]}$. The derivation property of the bracket determines that

$$\{\bar{s}_1, \bar{s}_2\}(\mu) = \left\langle \mu, \frac{\partial s_1}{\partial e_i} \{e_i, e_j\} \frac{\partial s_2}{\partial e_j}\right\rangle$$

$$= \left\langle \mu, \frac{\partial s_1}{\partial e_i} C_{ij}^k e_k \frac{\partial s_2}{\partial e_j}\right\rangle,$$

where C_{ij}^k are the structure constants of $\mathscr{L}$. It is evident that $\{S^m, S^n\} \subset S^{m+n-1}$ so that $S(\mathscr{L}) = \oplus S^n(\mathscr{L})$ is a graded Lie algebra with core subalgebra $S^1 = \mathscr{L}$, the linear functions on $\mathscr{L}^*$.

We now make the choice $\mathscr{L}_n = S^n(Z^*)$, the homogeneous functions on Z of homogeneity n. This choice induces another grading that has the Poisson algebra $S(\mathrm{sp}(6))$ as its core subalgebra. To see this, note that the bracket relations satisfy $[\mathscr{L}_i, \mathscr{L}_j] \subset \mathscr{L}_{i+j-2}$ so that $\mathscr{L}$ is a graded Lie algebra with core subalgebra $\mathscr{L}_2$, the quadratic functions. Denote by N a function on the integers with non-negative integer values. Then we can write

$$S^n(\oplus \mathscr{L}_m) = \oplus \vee_i \mathscr{L}_i^{N(i)},$$

$$\sum_i N(i) = n$$

where $\vee$ is the symmetric tensor product, $\mathscr{L}_i^{N(i)} = \mathscr{L}_i \vee \cdots \vee \mathscr{L}_i$, [$N(i)$ times], and the sum is taken over all such functions, N, such that $\sum N(i) = n$. Here, N is considered as an index. Note that this is the symmetric tensor product of vector spaces not of symmetric tensors. Hence, the result will not be, in general, a space of symmetric tensors. As functions of $\mathscr{L}_i^*$, $\mathscr{L}_i$ are the ith-order moments and $\mathscr{L}_i^{N(i)}$ are $N(i)$th-order functions of the ith-order moments. For example, $S^2(\oplus \mathscr{L}_m) = \mathscr{L}_1^2 \oplus \mathscr{L}_1 \mathscr{L}_2 \oplus \mathscr{L}_2^2 \oplus \cdots$. Namely it is the quadratic functions on $\mathscr{L}_i^*$ that are the quadratic functions of the first-order moments plus bilinear functions of the first- and second-order moments, plus the quadratic functions of the second-order moments, etc. Define $Q^N = \vee_i \mathscr{L}_i^{N(i)}$, the homogeneous functions of homogeneity $N(i)$ of the ith-order moments. Then

$$S(\mathscr{L}) = \underset{N,n}{\oplus} Q^N,$$

$$\sum_i N(i) = n.$$

The Lie–Poisson bracket satisfies $\{Q^N, Q^M\} \subset \oplus_{s,t} Q^{K(s,t)}$, where

$$K(s,t)(i) = N(i) + M(i), \quad i \neq s,t,s+t-2,$$

$$K(s,t)(s) = N(s) + M(s) - 1, \quad s \neq t$$

$$= N(s) + M(s) - 2, \quad s = t,$$

$$K(s,t)(s+t-2) = N(s+t-2) + M(s+t-2) + 1.$$

Let I be the functional $I(N) = \Sigma_i N(i)(i-2)$. Then the above relations show that $I(K(s,t)) = I(N) + I(M)$ so that I determines the grading $S(\mathscr{L}) = \oplus_i B_i$, where

$$B_i = \underset{N,n}{\oplus} Q^N,$$

$$\sum_j N(j) = n,$$

$$I(N) = i.$$

Because the Lie–Poisson bracket is linear in μ, $S^0(\mathscr{L})$ can be neglected in the algebra. So can $\mathscr{L}_0$ since for symplectic motion $\int \mu = \text{const}$. However, $\mathscr{L}_1$ cannot so easily be removed. The choice of origin in Z is still available and if it can be chosen as a fixed point for the vector fields under consideration, then we can consider the subalgebra

$$S^+ = \underset{N,n>1}{\oplus} Q^N,$$

$$\sum_{j>1} N(j) = n,$$

where $Q^N = \vee_{i>2} \mathscr{L}_i^{N(i)}$ so that we can consider functions of second-order moments and higher. Consequently the function I is non-negative determining the core subalgebra of $\oplus_{i>0} B_i$ to be $B_0 = \oplus_{s>1} \mathscr{L}_2^s$ where $\mathscr{L}_2 = \mathrm{sp}(6)$, so that B_0 is the formal power series Poisson algebra on $\mathrm{sp}(6)^*$.

IV. THE MOMENT INVARIANTS

We now describe the moment invariants (invariant functions of the moments) for Sp(6).

A. Quadratic moment invariants

The calculation of quadratic moment invariants can be accomplished by the use of Casimirs, cf., Micu,[9] who also proves the functional dependence of a particular list of moment invariants. We proceed here with a different, simpler, and more direct technique even if it is not applicable to higher-order moments.

The symplectic form $\omega \in Z^* \otimes Z^*$ determines an invertible map $j: Z \to Z^*$ by $jx = \omega(x,\cdot)$. Consider the tensor operator $j \otimes i: S^2(Z) \to Z^* \otimes Z$ where i is the identity. (No confusion between indices and maps should arise from this notation.) For X in $S^2(Z)$, the tensor $X' = j \otimes i \cdot X$ transforms according to

$$(j \otimes i)\cdot(s \otimes s)\cdot X = (s^{*-1} \otimes s)\cdot(j \otimes i)\cdot X,$$

for $s \in \mathrm{Sp}(6)$ since $s^{*-1} j = js$. Now $(s^{*-1} \otimes s)\cdot X'$ goes to $s\widetilde{X}s^{-1}$ under the isomorphism $Z^* \otimes Z \approx L(Z,Z)$, the linear maps on Z sending X' to $\widetilde{X}$. Hence, the invariant functions of the tensor X are $\mathrm{tr}(\widetilde{X}^n)$, $n = 1,2,\ldots$. The Cayley–Hamilton theorem says that $\widetilde{X}$ is a root of its characteristic polynomial. Multiplying this equation by powers of $\widetilde{X}$ and taking traces determines the $\mathrm{tr}(\widetilde{X}^n)$, $n > 6$ as algebraic functions of the $\mathrm{tr}(\widetilde{X}^n)$, $n \leqslant 6$. Let X^{ij} be the components of $X \in S^2(Z)$ in the symplectic basis e_i where $\omega(e_i, e_j) = J_{ij}$. Then $\widetilde{X} = J_{ij} X^{jk}$ and

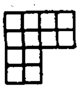
$$\mathrm{tr}(\widetilde{X}^n) = J_{i_1 j_1} X^{i_1 j_2} J_{i_2 j_2} X^{i_2 j_3} \cdots J_{i_n j_n} X^{i_n j_1},$$

are the functions of the quadratic moments which are invariant under Sp(6).

B. Higher-order moment invariants

Invariant functions of the moments are graded according to their homogeneity $N(k)$ with respect to kth-order moments. Therefore, any such function is determined by an invariant tensor in $\otimes (S^i(Z))^{N(i)*} = \otimes (S^i(Z^*))^{N(i)} \subset T^k(Z^*)$, where $I(N) = k$ and T^k denotes the k tensors. This can be seen as follows: it suffices to consider a form $T(X,\ldots,X)$ of degree k in $S^i(Z^*)$. The polarization[3]

$$D_{X_{a_1}} T = \partial_1 T \cdot a_1 = k \cdot T(a_1, X, X, \ldots, X)$$

is an invariant and the inverse is $T = (1/k) D_{XX} T$. The complete polarization $D_{X_{a_1}} D_{X_{a_1}} \cdots D_{X_{a_1}} T$ is an invariant tensor determining T to be an invariant tensor in $(S^i(Z))^k$. The description of such invariant tensors is most conveniently described by Young symmetrizers and Young diagrams, cf. Ref. 3,10. Any real analytic invariant for $\mathrm{Sp}(6,R)$ becomes complex analytic upon the field extension to $\mathrm{Sp}(6,C)$, and since the latter is semisimple all such invariants may be obtained in this way by Weyl's unitarian trick.[3,11] So from here on we use the complex field. Then the Young symmetrizers can be used to project the tensor space $T^k(Z^*)$ onto irreducible representation spaces of the symmetric group and also of $GL(6,C)$. The restriction to the subgroup $Sp(6)$ does not preserve irreducibility and a more complicated reduction of the Kronecker product as found in Ref. 12 is required to deduce the irreducible representations. However, presently we are only interested in determining the invariant tensors and the Young symmetrizers will suffice for this purpose. The symmetry class is represented by a Young diagram $(f_1,\dots,f_n)$ where $f_1 \geqslant f_2 \geqslant \cdots f_n$, with f_1 boxes in the first row, f_2 boxes in the second row, etc. For example, (4,4,2,2) appears as  .

The Young symmetrizer corresponding to the diagram is $h = PQ$, where $P = \Sigma_p p$ and $Q = \Sigma_q \operatorname{sign}(q) q$, where P is symmetrization with respect to the rows and Q is alternation with respect to the columns. Here, h, P, and Q are all projections after rescaling and since both P and Q are equivariant with respect to $Sp(6)$ these projections commute with the projection onto the invariant tensors. Since Q acting on the invariant tensors is a projection onto tensor products of skew-symmetric $Sp(6)$ invariant tensors its image is the linear span of tensor products of the Poincaré invariants $\omega^i = \omega \wedge \cdots \wedge \omega$ (i times). More precisely, if ϕ is an invariant tensor with diagram $(f_1,\dots,f_n)$ and λ_i is the number of columns of length i, then $Q\phi$ is a constant multiple of a tensor product of skew symmetric invariant tensors such that each column of length i corresponds the Poincaré invariant $\omega^{i/2}$. In particular, λ_i vanishes for odd i so that $f_1 = f_2 \geqslant f_3 = f_4 \geqslant \cdots$. For example, $Q\phi$ for (4,4,2,2) schematically appears as  so that $Q\phi$ must be a multiple of $\omega^2 \otimes \omega^2 \otimes \omega^1 \otimes \omega^1$, and the corresponding invariant tensor is $P(\omega^2 \otimes \omega^2 \otimes \omega^1 \otimes \omega^1)$. To explain the notation first note that in a symplectic basis for Z, ω^1 has the matrix $\omega^1_{ij} = J_{ij}$ and $\omega^2 = \omega \wedge \omega$ has the matrix skew $(J_{ij} J_{kl})$, where skew means to skew symmetrize with respect to all indices. This can be written $J \wedge J$, where $\wedge$ is the skew symmetric tensor product. The Poincaré invariants $\omega^i = \omega \wedge \cdots \wedge \omega$ (i times), written with indices, are $J \wedge \cdots \wedge J$. The columns in the diagram above must be of even length and into each column is inserted a Poincaré invariant of the correct rank, and an index corresponding to each box. Then, a tensor product of the moment tensors is formed so that each index corresponds to a box in the diagram, and the moment tensor index is summed against a Poincaré invariant index in each box. There are many ways to do this and each may provide a different moment invariant. However, some of these produce products and sums of simpler moment invariants. For example, for the diagram (4,4,2,2) write with indices $Y = X^{i_1 i_2 i_3 i_4} X^{j_1 j_2 j_3 j_4} X^{k_1 k_2} X^{l_1 l_2}$, where the X's are symmetric and the indices of the first X correspond to the first row and the indices of the second X correspond to the second row, etc. Then $PY = Y$ since the symmetries of Y correspond to the symmetries of the diagram and the value of this basic invariant is

$$\langle P(\omega^2 \otimes \omega^2 \otimes \omega^1 \otimes \omega^1), Y \rangle = \omega^2_{i_1 j_1 k_1 l_1} \omega^2_{i_2 j_2 k_2 l_2} \omega^1_{i_3 j_3} \omega^1_{i_4 j_4} \times X^{i_1 i_2 i_3 i_4} X^{j_1 j_2 j_3 j_4} X^{k_1 k_2} X^{l_1 l_2}.$$

When $PY = Y$, we refer to the invariant as a basic moment invariant, for which there is the special form $\langle \otimes_i (\otimes \omega^{i/2})^{\lambda_i}, Y \rangle$. On the other hand, if we consider $Y = X^{i_1 i_2} X^{i_3 i_4} X^{j_1 j_2} X^{j_3 j_4} X^{k_1 k_2} X^{l_1 l_2}$, then the first term in the moment invariant is

$$\omega^2_{i_1 j_1 k_1 l_1} \omega^2_{i_2 j_2 k_2 l_2} \omega^1_{i_3 j_3} \omega^1_{i_4 j_4} X^{i_1 i_2} X^{i_3 i_4} X^{j_1 j_2} X^{j_3 j_4} X^{k_1 k_2} X^{l_1 l_2},$$

which is

$$\omega^2_{i_1 j_1 k_1 l_1} \omega^2_{i_2 j_2 k_2 l_2} \times X^{i_1 i_2} X^{j_1 j_2} X^{k_1 k_2} X^{l_1 l_2} \times \omega^1_{i_3 j_3} \omega^1_{i_4 j_4} \times X^{i_3 i_4} X^{j_3 j_4}.$$

This is a product of the basic moment invariants corresponding to  and  .

Also, if we choose

$$Y = X^{i_1} X^{i_2} X^{i_3} X^{i_4} X^{j_1} X^{j_2} X^{j_3} X^{j_4} X^{k_1 k_2} X^{l_1 l_2},$$

then the corresponding moment invariant vanishes due to antisymmetry of the Poincaré invariants.

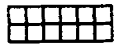
As a general rule, for a diagram with $f_1 = f_2 \geqslant f_3 = f_4 \geqslant \cdots$ the tensor invariant is $P(\otimes_i (\otimes \omega^{i/2})^{\lambda_i})$, and the basic moment invariant is

$$\langle \otimes_i (\otimes \omega^{i/2})^{\lambda_i}, X^{f_1} \otimes \cdots X^{f_n} \rangle,$$

where the indices of the Poincaré invariants run from X^{f_1} to X^{f_2} to X^{f_3}, etc. so that each row of length f_i corresponds to the tensor X^{f_i} by placing its indices in the boxes of that row and each column corresponds to a Poincaré invariant by putting its indices in the boxes of the column.

Thus, the Poincaré invariants in the single-particle phase space generate infinite families of invariants in the moment phase space. (However, not all of the moment invariants in a given family are functionally independent.)

Example 1: Consider a moment invariant corresponding to the diagram

$(n,n,0,0,\dots)$  , where we place the indices of only the second-order moment tensors in the boxes. Then, this invariant must be a function of the quadratic moment invariants from Sec. IV A

$$\phi_k = J_{i_1 j_1} X^{i_1 j_2} J_{i_2 j_2} X^{i_2 j_3} \cdots J_{i_k j_k} X^{i_k j_1}, \quad k \leqslant 6.$$

This can be seen explicitly as follows. The tensor J in ϕ_k connects the second index of the first X with the first index of the second X. Likewise, another J connects the second index of the second X with the first index of the third X, and so on until finally the second index of the last X is connected with the first index of the first X, forming a cycle of length k. Now, place the indices of $X = X^2$ into the boxes and, without permutations, sum against the indices of $(\otimes J)^n$. In this sum, the first index of one X are connected by J to an index of another X. The latter's remaining index is then connected to another index until the cycle (of length m suppose) closes back to the first X, giving plus or minus ϕ_m. Thus the whole sum is plus or minus the product of some of the ϕ_m, $m = 1,\dots,n$. The same is true under permutation of the indices, so that upon application of the Young symmetrizer P, the moment invariant must be a linear combination of products of the ϕ_m. Since these are functionally dependent upon ϕ_m, $m \leqslant 6$, we are done.

Example 2: Lysenko and Overley[1] show that $I_n = \langle (\mathcal{Q} - \mathcal{P})^n \rangle_{\text{sym}}$ is an invariant for all n, where $\mathcal{Q} = q^1 + q^2 + q^3$ and $\mathcal{P} = p^1 + p^2 + p^3$ and

$$\left\langle \prod_{k=1}^{3} (q^k)^{i_k} (p^k)^{j_k} \right\rangle_{\text{sym}} = \left\langle \prod_{k=1}^{3} (q^k)^{i_k} (p^k)^{j_k} \right\rangle \left\langle \prod_{k=1}^{3} (q^k)^{j_k} (p^k)^{i_k} \right\rangle.$$

This invariant is expressible as $\langle (\otimes \omega)^n, X^n \otimes X^n \rangle$, also corresponding to the diagram (n,n,0,0...)  .

Proof: $(\mathcal{Q} - \mathcal{P})^n = (-\Sigma_i J_{i\bar{j}} z^j)^n = (-1)^n (J_{\bar{j}} z^j)^n$ since $J_{ij} \neq 0$ only when $i = \bar{j}$, where $\bar{1} = 4$, $\bar{4} = 1$, $\bar{2} = 5$, etc. Barring indices corresponds to exchanging q^j and p^j, i.e., if $z^j = q^k$ then $z^{\bar{j}} = p^k$ and vice versa. Consequently,

$$\begin{aligned} (J_{\bar{j}j} z^j)^n &= J_{\bar{i}_1 i_1} z^{i_1} \cdots J_{\bar{i}_n i_n} z^{i_n} \\ &= (-1)^n J_{i_1 \bar{i}_1} \cdots J_{i_n \bar{i}_n} z^{i_1} \cdots z^{i_n}. \end{aligned}$$

Letting $\bar{i}_1 \cdots \bar{i}_n$ be free variables and integration over the particle distribution shows that, as claimed,

$$\begin{aligned} I_n &= J_{i_1 j_1} \cdots J_{i_n j_n} X^{i_1 \cdots i_n} X^{j_1 \cdots j_n} \\ &= \langle (\otimes \omega)^n, X^n \otimes X^n \rangle. \end{aligned}$$

These two examples emphasize that the description of moment invariants via Young diagrams refers to *classes* of moment invariants rather than individual moment invariants. Note that the rms emittance is the basic moment invariant corresponding to the diagram (2,2),  , and so is a special case of Examples 1 and 2.

V. MOMENT INVARIANT INEQUALITIES

Dragt *et al.*[13] show the invariants of Example 1 are non-negative (in 1D, $\langle q^2 \rangle \langle p^2 \rangle - \langle pq \rangle^2 \geqslant 0$ follows from Schwarz's inequality). The second order moment tensor X^2 is non-negative and generically it is positive definite. In this case it can be put into real normal form,[14] so that under a symplectic coordinate change $X^2 = \mathrm{diag}(\alpha_1, \alpha_2, \alpha_3, \alpha_1, \alpha_2, \alpha_3)$ thereby determining the moment invariants of Example 1 to be

$$\mathrm{tr}(\widetilde{X}^n) = 2(\alpha_1^n + \alpha_2^n + \alpha_3^n),$$

for n even and zero for n odd, which provides a proof of the Dragt *et al.* result since the positive definite tensors are dense in the non-negative tensors.

The basic moment invariants depending on even order moments can also be shown to be non-negative. In this case, each Poincaré invariant appears an even number of times. Consider the basic moment invariant in Sec. IV B with diagram (4,4,2,2). The general case follows similarly. Consider variable tensors x^i, y^i, z^i, w^i and form the function

$$\begin{aligned} &\omega^2_{i_1 j_1 k_1 l_1} \omega^2_{i_2 j_2 k_2 l_2} \omega^1_{i_3 j_3} \omega^1_{i_4 j_4} x^{i_1} x^{i_2} x^{i_3} x^{i_4} y^{j_1} y^{j_2} y^{j_3} y^{j_4} z^{k_1} z^{k_2} w^{l_1} w^{l_2} \\ &\quad = (\omega^2_{ijkl} x^i y^j z^k w^l)^2 \times (\omega^1_{ij} x^i y^j)^2. \end{aligned}$$

This is a non-negative function and integration with respect to the variable x over any particle distribution also produces a non-negative function. Iterated integration over each variable against different particle distributions shows that the *mixed* moment invariant,

$$\omega^2_{i_1 j_1 k_1 l_1} \omega^2_{i_2 j_2 k_2 l_2} \omega^1_{i_3 j_3} \omega^1_{i_4 j_4} \times X^{i_1 i_2 i_3 i_4} Y^{j_1 j_2 j_3 j_4} Z^{k_1 k_2} W^{l_1 l_2},$$

depending on distinct moment tensors, X, Y, Z, W is non-negative, and the specification $X = Y = Z = W$ gives the result.

The actual determination of the distribution corresponding to a numerically computed moment tensor may not be necessary since the moments are the important laboratory quantities. However, it is necessary that each tensor be the moment of a distribution. With this in mind, consider the cone D in $\mathcal{G}^*$ consisting of the distributions. It is closed under multiplication by non-negative functions and is a convex set. Since the coadjoint action is $\mathrm{Ad}^*_g \mu = \mu \cdot g$, where g is a symplectic map, the coadjoint orbit through any element in D is contained in D. Consider also the cone N in L^* consisting of non-negative tensors. The map $i^*: \mathcal{G}^* \to \mathcal{L}^*$ maps coadjoint orbits to coadjoint orbits and maps D onto N. The conclusion is that the coadjoint orbit through a non-negative tensor lies completely in the non-negative tensors. At present all Lie–Poisson integrators use the coadjoint action to generate the Poisson map.[6,15,16] For such schemes, this implies that Lie–Poisson integration with initial value a moment of a distribution stays in the space of tensors which are moments of distributions. Moreover, the moment-invariant inequalities are also preserved by Lie–Poisson integration.

VI. DISCUSSION

The evolution of a Lie–Poisson system is a Poisson map in that it preserves bracket relations and coadjoint orbits. Ge and Marsden[15] have demonstrated a Lie–Poisson version of Hamilton–Jacobi theory and Ge[16] has demonstrated a generating function for the Poisson map that can be used to construct Lie–Poisson integrators. Preliminary calculations by Channell and Scovel[6] show that Lie–Poisson integrators possess the same stability characteristics as the symplectic integrators demonstrated in Channell and Scovel,[17] and therefore we propose to integrate the Vlasov equation as a Lie–Poisson system. A related approach can be found in Dragt *et al.*[13]

Dobrushin[18] demonstrates the existence and uniqueness of solutions to the Vlasov equation for both signed measures and Baire measures. If the moment of a distribution, $X \in S(Z)$, is integrated numerically according to $\dot{\mu} = -\mathrm{ad}^*_{\partial H/\partial \mu} \mu$, the tensor may not remain the moment of a distribution. However, we have shown that it does so under Lie–Poisson integration.

As an application of the phase space of moment invariants we consider a matching section in a particle-beam accelerator. Suppose the matching section can provide any desired linear symplectic transformation; so that phase space (defined in Sec. III) is quotiented by Sp(6). The moment invariants will determine a good set of coordinates if the invariants form a complete set, in that any two sets of moments with the same value for the invariants are conjugate under Sp(6). X is said to be conjugate to Y if $X^k = (\otimes S)^k Y^k$ for all k and some fixed $S \in \mathrm{Sp}(6)$.

For two moments X and Y that come from distributions, the two tensors X^2 and Y^2 must be non-negative. When they are also definite they have real normal forms[14] so they are conjugate.

Should the completeness of the set of invariants be established, we may then consider the moment invariant dynamics

$$\dot{I} = \{H,I\},$$

where I and H are in $S(\mathscr{L})$. Since the quadratic moment invariants describe the coadjoint orbits in $\mathrm{sp}(6)^* = \mathscr{L}_2^*$ they are invariant under the flow generated by the full symmetric tensor Poisson algebra over $\mathrm{sp}(6)^*$, $S(\mathrm{sp}(6)) = S(\mathscr{L}_2)$.

The moment invariants form a small subset of the full algebra and they are functionally dependent. For example, the quadrant moment invariants $\mathrm{tr}((JX^2)^n)$ for $n \geqslant 6$ are dependent upon $\mathrm{tr}((jX^2)^n)$ for $n < 6$ so that we can limit ourselves to the latter. Likewise, Weyl[3] uses Hilbert's theorem on polynomial ideals to show that the moment invariants which depend on k th-order moments or less are finitely generated and Schwarz[19] has estimated the number of generators and relations, so one could try to find and use the generators. However, at present, these generators are not known.

Note added in proof: The requirement imposed at the end of Sec. III—that the first moments be fixed—can be removed by writing all moments with respect to the center of mass, and keeping the first moments variable. Channell and Scovel[20] apply this technique to the finite-dimensional truncation problem.

ACKNOWLEDGMENTS

This work was completed while the third author was visiting the ETH in Zürich. He would like to thank Jürgen Moser for this fine opportunity, and to thank Rahel Boler for an excellent typing of the original manuscript. We would also like to express thanks to Paul Channell, Jim Louck, Gian-Carlo Rota, Gerald Schwarz, Alexander Veselov, and Alan Weinstein for helpful conversations.

Support from MSRI at Berkeley and the Center for Nonlinear Studies at Los Alamos is gratefully acknowledged.

[1] W. P. Lysenko and M. Overley, in *Linear Accelerators and Beam Optics Codes*, AIP Conference Proceedings 177, edited by C. R. Eminhizer (AIP, New York, 1988), p. 323.

[2] J. Marsden, T. Ratiu, R. Schmid, R. Spencer, and A. Weinstein, in *Proceedings of the IUTAM-ISIMM Symposium on Modern Developments in Analytical Mechanics*, Torino, 7–11 June, 1982 (Atti della Scienze di Torino, Torino, Italy, 1982), pp. 117, 289–340.

[3] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, NJ, 1946).

[4] R. Shaw, *Linear Algebra and Group Representations* (Academic, London, 1983), Vol. 2.

[5] E. K. Haviland, Am. J. Math. **57**, 562–568 (1935).

[6] P. J. Channell and J. C. Scovel, "Integrators for Lie–Poisson Dynamical Systems," in preparation.

[7] V. Guillemin and S. Sternberg, Ann. Phys. **127**, 220–253 (1980).

[8] V. Guillemin, J. Diff. Geom. **4**, 257–282 (1970).

[9] M. Micu, Nucl. Phys. **60**, 353–362 (1964).

[10] C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95–120 (1966).

[11] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations* (Springer, New York, 1984).

[12] M. Fischler, J. Math. Phys. **22**, 637–648 (1981).

[13] A. J. Dragt, R. L. Gluckstern, F. Neri, and G. Rangarajan, in *Proceedings of the 1989 International Conference on Particle Accelerators, Chicago*, to appear.

[14] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Addison-Wesley, Redwood City, 1978).

[15] Z. Ge and J. Marsden, Phys. Lett A **133**, 134–139 (1988).

[16] Z. Ge, "The Generating Function for the Poisson Map," preprint.

[17] P. J. Channell and J. C. Scovel, "Symplectic Integration of Hamiltonian Systems," to appear in Nonlinearity.

[18] R. L. Dobrushin, Funct. Anal. Appl. **13**, 115–123 (1979).

[19] G. W. Schwarz, Ann. Inst. Fourier **37**, 191–216 (1987).

[20] P. J. Channell and J. C. Scovel, "Finite Dimensional Truncation of Infinite Dimensional Lie–Poisson Dynamical Systems," in preparation.

Nonlinear evolution equations related to the first-order system: $\phi_x = \lambda J\phi + P\phi$

J. A. Cavalcante and J. A. Silva
Departamento de Matemática, Universidade de Brasília, 70.910 Brasília D. F., Brazil

(Received 21 June 1989; accepted for publication 28 February 1990)

Let $U_t = F(U,\partial U/\partial x,..., \partial^r U/\partial x^r,t)$ or $U_{xt} = F(U,\partial U/\partial x,...,\partial^r U/\partial x^r,t)$ be the nonlinear evolution equations that are the compatibility conditions between $\phi_x = \lambda J\phi + P\phi$ and $\phi_t = A\phi$ for $P = U(x,t)$ or $P = U_x(x,t)$, respectively. In this paper, it is proved that if $A(Z_0,...,Z_{r-1},t,\lambda)$ is a continuous function such that $(\partial A/\partial Z_k)$ $(Z_0,...,Z_{r-1},t,\lambda)$ $k = 0,1,...,$ exists $(Z_k = \partial^k U/\partial x^k)$, then for $P = U(x,t)$, A is a polynomial in λ of degree r and for the case $P = U_x$, $A = A_{-1}/\lambda + A_0 + \cdots + A_{r-1}\lambda^{r-1}$. The case where $P = Z_m$, $m \geqslant 2$ is also analyzed.

I. INTRODUCTION

In a previous study,[1] we have derived a recursive formula to obtain nonlinear evolution equations that are the compatibility condition between

$$\phi_x = \lambda J\phi + P\phi \tag{1.1a}$$

and

$$\phi_t = A\phi, \tag{1.1b}$$

where ϕ, J, P, and A are $n\times n$ matrices, J is a fixed diagonal matrix $(J_{jj} \neq J_{kk}; j\neq k)$, and P is an off-diagonal matrix. We denote that $M_n(C)$ is the space of the $n\times n$ matrices, $\mathbf{D}_n(C) \subset M_n(C)$ is the space of diagonal matrices, and $\tau_n(C) \subset M_n(C)$ is the space of the off-diagonal matrices. Let $\{E_{jk}\in M_n(C);\ (E_{jk})_{lm} = \delta_{jl}\delta_{km}\}$ be a basis of $M_n(C)$. The compatibility condition between (1.1a) and (1.1b) is given by

$$A_x = \lambda\,[J,A] + [P,A] + P_t. \tag{1.2}$$

This equation is equivalent to the system

$$D_x = \mathbf{P}_d[P,T], \tag{1.3a}$$

$$T_x = \lambda\,\mathbf{J}T + [P, D] + \mathbf{P}_0[P,T] + P_t, \tag{1.3b}$$

where $\mathbf{P}_d$ and $\mathbf{P}_0$ are the orthogonal projections on $\mathbf{D}_n(C)$ and $\tau_n(C)$, respectively, $A = D + T$ with $D = \mathbf{P}_d A$, $T = \mathbf{P}_0 A$, $\operatorname{tr} D = 0$, and $\mathbf{J}$: $M_n(C) \to M_n(C)$ $\mathbf{J}(A) = [J,A]$ is a linear mapping that is a bijection on $\tau_n(C)$. If we take $P = U(x,t)$ or $P = U_x(x,t)$ we get the equation

$$P_t = F\left(U,\frac{\partial U}{\partial x},...,\frac{\partial^r U}{\partial x^r},t\right). \tag{1.4}$$

In Ref. 1 we suppose that D, T, and F are smooth functions on the variables Z_0, Z_1,... $(Z_j = \partial^j U/\partial x^j, j = 0,1,...)$ and that $\lambda = 0$ is at most an isolated singularity of D and T. Therefore, we can expand D and T around $\lambda = 0$,

$$D = \sum_{k=-\infty}^{\infty} \lambda^k D_k, \quad T = \sum_{k=-\infty}^{\infty} \lambda^k T_k,$$

to obtain the following relations.

(i) For $P = U(x,t)$, D is a polynomial in λ of degree r and T is also a polynomial in λ of degree $r - 1$. Therefore, A is a polynomial in λ of degree r.

(ii) For $P = U_x(x,t)$, D and T have, at most, a simple pole at $\lambda = 0$ and their regular part are polynomials in λ of degree $r - 1$ and $r - 2$, respectively. Hence, $A = A_{-1}/\lambda + A_0 + \cdots + \lambda^{r-1}A_{r-1}$.

We assume now that if A is a continuous function on the variables Z_k, $k = 0$, 1,..., and $(\partial A/\partial Z_k)$ $k = 0,1,...,$ exists then A and F will be C^∞ functions on the variables Z_k, $k = 0,1,...,$ and we prove in Sec. II the following.

(iii) For $P = U(x,t)$, A is a C^∞ function on the variables Z_k, λ $k = 0,1,...,$ $\lambda\in C$ (A is analytic on λ). Hence, we can expand A around $\lambda = 0$ and we get from (i) that A is a polynomial in λ of degree r.

(iv) For $P = U_x(x,t)$, A is a C^∞ function on the variables Z_k, λ $k = 0,1,...,$ $\lambda\in C\setminus\{0\}$ and A has at most a simple pole at $\lambda = 0$. It follows then (ii) that $A = A_{-1}/\lambda + A_0 + ... + A_{r-1}\lambda^{r-1}$.

Therefore, if A is a smooth function on Z_k, $k = 0,1,...,r - 1$, then the most general nonlinear evolution equation that is solvable by inverse scattering associated to (1.1a) and (1.1b) for $P = U$ or $P = U_x$ comes from A in the forms (i) or (ii), respectively, that are given in Ref. 1 (see, also, Ablowitz[2]).

In Sec. III, we study the case when $P = Z_m$ and Eq. (1.4) becomes

$$P_t = F\left(\underbrace{\int\cdots\int}_{m\text{-times}} P,...,\int P,P,...,t\right)$$

or

$$\left(\frac{\partial^m U}{\partial x^m}\right)_t = F\left(U,\frac{\partial U}{\partial x},...,t\right).$$

We show that if the entries of Z_{m-1} are linearly independent functions, then D, T are polynomial in λ and D, T, and F are functions independent of the variables $Z_0, Z_1,..., Z_{m-1}$. Hence, the nonlinear evolution equations related to this case are the same as of the case $P = U$. As an example, when the entries of U are linearly dependent we study the case

$$P = z_m\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \partial^m u/\partial x^m \\ \partial^m u/\partial x^m & 0 \end{pmatrix},$$

and we show that D, T have, at most, a simple pole at $\lambda = 0$, their regular part are polynomials in λ, the residues are functions depending only on z_{m-1}, and the functions D, T, and F are independent of $z_0,z_1,...,z_{m-2}$. Therefore, the substitution $v = z_{m-1}$ reduces this case to the case $v_{xt} = F(v,\partial v/\partial x,...,t)$ so no new equations are obtained.

 0022-2488/90/071616-05$03.00

II. THE SMOOTHNESS OF A ON THE VARIABLES Z_k, λ $k=0,1,...$

We recall the definition of the operator $\mathbf{L}$ and the definition of a type of a function that is given in Ref. 1.

Definition 2.1: Let G be a function depending on a finite number of variable Z_k $k=0,1,...$, such that $\partial G/\partial Z_k$ exists. Define

$$(\mathbf{L}G)(Z_0,Z_1,...)=\sum_{j\geqslant 0}\frac{\partial G(Z_0,Z_1,...)}{\partial Z_j}(Z_{j+1}).$$

Equation (1.2) is now written as

$$\mathbf{L}A=\lambda\,\mathbf{J}A+[Z_m,A]+F. \tag{2.1}$$

Definition 2.2: We say that the function f depending on $Z_0,Z_1,...$, is of type m and write $\tau(f)=m$ if

$$\frac{\partial f}{\partial Z_j}=0,\quad j\geqslant m+1,$$

and f depends on Z_m (if $f\in C^1$, $\partial f/\partial Z_m\neq 0$). For f a constant function we define $\tau(f)<0$.

Remark 1: Let $\tau(f)=m$, $\tau(g)=m_1$, and suppose that $(\partial f/\partial Z_k)$ $k=0,1,...$ exists, then the following is true.

(a) $\tau(\mathbf{L}f)=m+1$ and

$$\mathbf{L}f=\sum_{j=0}^{m}\frac{\partial f}{\partial Z_j}(Z_{j+1}),$$

If $\tau(f)<0$, then $\tau(\mathbf{L}f)<0$.

(b) Let $B{:}\,M_n(C)\times M_n(C)$ be a bilinear mapping, then $\tau(B(f,g))\leqslant\max\{m,m_1\}$,

(c) If in (2.1) we have $\tau(F)\geqslant m+1$ then $\tau(A)=\tau(F)-1$, if $\tau(F)<m$ then $\tau(A)=m-1$, and if $\tau(F)=m$ then $\tau(A)\leqslant\tau(F)-1$.

We also need the following lemma proved in Ref. 1.

Lemma 2.3: Let $A\in M_n(C)$ and $T\in\tau_n(C)$.

(i) If $[A,T]=0\ \forall T\in\tau_n(C)$, then $A=\alpha I$, in particular, $\operatorname{tr}A=0$ implies $A=0$.

(ii) If $[A,T]\in\tau_n(C)$ $\forall A\in M_n(C)$, then $T=0$.

A. $P=U(x,t),\tau(F)=r$

Let $P=U=Z_0\in\tau_n(C)$, then (2.1) gives us

$$\mathbf{L}(A)=\lambda\,\mathbf{J}A+[Z_0,A]+F \tag{2.2}$$

and we have the following theorem.

Theorem 2.4: Suppose A is a C^∞ function on the variables Z_k, $k=0,1,...,r-1$, then A is a C^∞ function on the variables Z_k,λ $k=0,1,...,r-1$, $\lambda\in C$.

Proof: Initially we prove by induction that $(\partial A/\partial Z_j)$ $j=0,1,...,r-1$ is C^∞ on Z_k,λ $k=0,1,...,r-1$, $\lambda\in C$ $(r\geqslant 1)$.

(i) Taking the derivative of (2.2) with respect to Z_r yields

$$\frac{\partial A}{\partial Z_{r-1}}=\frac{\partial F}{\partial Z_r},$$

hence $\partial A/\partial Z_{r-1}$ is C^∞ on Z_k,λ $k=0,1,...,r-1$, $\lambda\in C$.

(ii) Suppose that $\partial A/\partial Z_{j+1}$ is C^∞ on Z_k,λ $k=0,1,...,r-1$, $\lambda\in C$ for $0\leqslant j\leqslant r-2$, then taking derivative of (2.2) with respect to Z_{j+1} we find

$$\frac{\partial A}{\partial Z_j}+\mathbf{L}\frac{\partial A}{\partial Z_{j+1}}=\lambda\,\mathbf{J}\frac{\partial A}{\partial Z_{j+1}}+\left[Z_0,\frac{\partial A}{\partial Z_{j+1}}\right]+\frac{\partial F}{\partial Z_{j+1}}.$$

So $\partial A/\partial Z_j$ is C^∞ on Z_k,λ $k=0,1,...$, $\lambda\in C$. Since

$$\begin{aligned}A(Z_0,...,Z_{r-1},t,\lambda)&=A(0,...,0,t,\lambda)+\int_0^1\frac{\partial A}{\partial s}(sZ_0,...,sZ_{r-1},t,\lambda)\,ds\\&=A(0,...,0,t,\lambda)+\int_0^1\sum_{j=0}^{r-1}\frac{\partial A}{\partial Z_j}(Z_{j+1})\,ds,\end{aligned}$$

we need to show only that $A(0,...,0,t,\lambda)$ is C^∞ on λ. Let $A_0(Z_0,t,\lambda)=A(Z_0,0,...,0,t,\lambda)$ and $F_0(Z_0,t)=F(Z_0,0,...,0,t)$. Choosing $Z_1=Z_2=\cdots=Z_r=0$ in (2.2) we get

$$\lambda\,\mathbf{J}A_0+[Z_0,A_0]+F_0=0,$$

or

$$\mathbf{P}_d[Z_0,T_0]=0, \tag{2.3a}$$

$$\lambda\,\mathbf{J}T_0+[Z_0,D_0]+\mathbf{P}_0[Z_0,T_0]+F_0=0, \tag{2.3b}$$

where $D_0=\mathbf{P}_dA_0$ and $T_0=\mathbf{P}_0A_0$. Taking the derivatives of (2.3a) and (2.3b) with respect to Z_0 and letting $Z_0=0$ we obtain

$$\mathbf{P}_d[\cdot,T_0(0,t,\lambda)]=0, \tag{2.4a}$$

$$\lambda\,\mathbf{J}\frac{\partial T_0(0,t,\lambda)}{\partial Z_0}(\cdot)+[\cdot,D_0(0,t,\lambda)]+\mathbf{P}_0[\cdot,T_0(0,\lambda,t)]+\frac{\partial F(0,t)}{\partial Z_0}=0. \tag{2.4b}$$

From (2.4a) and from (ii) of Lemma 2.3 we get that $T_0(0,\lambda,t)=0$. From (2.4b) it follows that $[\cdot,D_0(0,t,\lambda)]$ is C^∞ on λ. Let

$$D_0(0,t,\lambda)=\sum_{J=1}^{n}d_j(t,\lambda)E_{jj},$$

with

$$\sum_{j=1}^{n}d_j=0(\operatorname{tr}A=\operatorname{tr}D=0),$$

then

$$[E_{jk},D_0(0,t,\lambda)]=(d_k-d_j)E_{jk}.$$

Therefore, (d_j-d_k) is C^∞ on λ. Since

$$d_j=\frac{1}{n}\sum_{k=1}^{n}(d_j-d_k),$$

we get that d_j is analytic on λ and $A(0,...,0,t,\lambda)=D_0(0,t,\lambda)$, is analytic on λ. If $r=0$, then A is constant on Z_k, $k=0,1,...$, and A is equal to $A(0,...,0,t,\lambda)$. □

B. $P=U_x(x,t),\tau(F)=r$

Let $P=U_x(x,t)=Z_1$, then (2.1) gives us

$$\mathbf{L}A=\lambda\,\mathbf{J}A+[Z_1,A]+F,$$

or

$$\mathbf{L}D=\mathbf{P}_d[Z_1,T], \tag{2.5a}$$

$$\mathbf{L}T=\lambda\,\mathbf{J}T+[Z_1,D]+\mathbf{P}_0[Z_1,T]+F, \tag{2.5b}$$

and then we have the following theorem.

Theorem 2.5: Suppose that A is C^∞ on Z_k, $k=0,1,\ldots$, then A is a C^∞ function on the variables Z_k, λ $k=0,1,\ldots$, $\lambda \in C\setminus\{0\}$ and A has, at most, a simple pole at $\lambda=0$.

Proof ($r\geqslant 2$): Similarly, as Theorem 2.4 shows, we can prove by induction that $(\partial A/\partial Z_j)$ $1\leqslant j\leqslant r-1$ is C^∞ on Z_k, λ $k=0,1,\ldots,r-1$ $\lambda\in C$. Therefore,

$$A=A(Z_0,0,\ldots,0,t,\lambda)$$
$$+\int_0^1 \frac{\partial A}{\partial s}(Z_0,sZ_1,\ldots,sZ_{r-1},t,\lambda)\,ds.$$

Let

$$A_0(Z_0,t,\lambda)=A(Z_0,0,\ldots,0,t,\tau),$$
$$D_0(Z_0,t,\tau)=\mathbf{P}_d A_0(Z_0,t,\lambda),$$
$$T_0(Z_0,t,\lambda)=\mathbf{P}_0 A_0(Z_0,t,\lambda),$$

and

$$F_0(Z_0,t)=F(Z_0,0,\ldots,0,t).$$

Then, choosing $Z_1=Z_2=\cdots=Z_r=0$ in (2.5b) we get

$$0=\lambda\,\mathbf{J}T_0(Z_0,t,\lambda)+F_0(Z_0,t).$$

So $T_0(Z_0,t,\lambda)$ has a simple pole at $\lambda=0$ and is analytic on $C\setminus\{0\}$. Taking $Z_2=Z_3=\cdots=Z_r=0$ in (2.5b), differentiating with respect to Z_1, and equaling Z_1 to zero we obtain

$$\frac{\partial T_0(Z_0,t,\lambda)}{\partial Z_0}(\cdot)$$
$$=\lambda\,\mathbf{J}\,\frac{\partial T(Z_0,0,\ldots,0,t,\lambda)}{\partial Z_1}(\cdot)+[\cdot,D_0(Z_0,t,\lambda)]$$
$$+\mathbf{P}_0[\cdot,T_0(Z_0,t,\lambda)]+\frac{\partial F(Z_0,0,\ldots,0,t)}{\partial Z_1}.$$

Evaluating in E_{jk} it follows that d_j-d_k has a simple pole at $\lambda=0$, where

$$D_0(Z_0,t,\lambda)=\sum_{j=1}^n d_j(Z_0,t,\lambda)E_{jj}$$

with

$$\sum_{j=1}^n d_j=0.$$

Since

$$d_j=\frac{1}{n}\sum_{k=1}^n(d_j-d_k)$$

we get that $D_0(Z_0,t,\lambda)$ has a simple pole at $\lambda=0$. For $r=0,1$ $\tau(A)<0$. Therefore, A depends mostly on Z_0 and is equal to A_0. □

Remark 2: In Ref. 1 we proved that if the entries of U are functions linearly independent then the residues of D and T are zero and the functions D, T, and F are independent of $Z_0=U$. So the nonlinear evolution equations related to the case $P=U_x$ are identical to the nonlinear evolution equations related to $P=U$.

III. $P=Z_m$

We start with Lemma 3.1 that is a generalization of Lemma 3.4 of Ref. 1.

Lemma 3.1: Let $Z=(z_{jk})\in\tau_n(C)$ with the entries z_{jk} being linearly independent functions. Let $D\in\mathbf{D}_n(C)$, $T\in\tau_n(C)$, $L_d(Z)\in\mathbf{L}(\tau_n(C);\mathbf{D}_n(C))$ be differentiable functions with respect to Z and let $L_0(Z)\in\mathbf{L}(\tau_n(C),\tau_n(C))$ be a twice differentiable function with respect to Z such that

$$\frac{dD(Z)}{dZ}(\cdot)=\mathbf{P}_d[\cdot,T(Z)]+L_d(Z)(\cdot)\quad(\mathrm{tr}D(Z)=0),\tag{3.1a}$$

$$\frac{dT(Z)}{dZ}(\cdot)=[\cdot,D(Z)]+\mathbf{P}_0[\cdot,T(Z))]+L_0(Z)(\cdot).\tag{3.1b}$$

Then if

$$D=\sum_{j=1}^n d_jE_{jj}\ \text{ and }\ T=\sum_{j,k=1}^n t_{jk}E_{jk}$$

we get

$$t_{jk}=\frac{1}{2}\left\{\left\langle\frac{\partial L_0(Z)}{\partial z_{jk}}(E_{jk})-\frac{\partial L_0(Z)}{\partial z_{kj}}(E_{jk}),E_{jk}\right\rangle\right.$$
$$\left.-\langle L_d(Z)(E_{jk}),E_{kk}-E_{jj}\rangle\right\},\tag{3.2a}$$

$$d_j-d_k=-\frac{\partial t_{jk}}{\partial z_{jk}}+\langle L_0(Z)(E_{jk}),E_{jk}\rangle,\tag{3.2b}$$

$$d_j=\frac{1}{n}\sum_{k=1}^n(d_j-d_k),\tag{3.2c}$$

where $\langle A,B\rangle=\mathrm{tr}(AB^*)$.

Proof: The following relations hold:

$$\frac{\partial d_j}{\partial z_{kj}}=-t_{jk}+\langle L_d(E_{jk}),E_{jj}\rangle,\tag{3.3a}$$

$$\frac{\partial d_k}{\partial z_{kj}}=t_{jk}+\langle L_d(E_{kj}),E_{kk}\rangle,\tag{3.3b}$$

$$\frac{\partial t_{jk}}{\partial z_{jk}}=d_k-d_j+\langle L_0(E_{jk}),E_{jk}\rangle,\tag{3.3c}$$

$$\frac{\partial t_{jk}}{\partial z_{kj}}=\langle \mathrm{L}_0(E_{kj}),E_{jk}\rangle.\tag{3.3d}$$

Let us verify (3.3a), the others follow similarly

$$\frac{\partial d_j}{\partial z_{kj}}=\frac{\partial}{\partial z_{kj}}\langle D,E_{jj}\rangle=\langle\frac{\partial D}{\partial z_{kj}},E_{jj}\rangle.$$

With the help of (3.1a) we get

$$\frac{\partial d_j}{\partial z_{kj}}=\langle\mathbf{P}_d[E_{kj},T]+L_d(E_{kj}),E_{jj}\rangle.$$

Now, using $\langle[A,B],C\rangle=\langle B,[A^*,C]\rangle$, it follows that

$$\frac{\partial d_j}{\partial z_{kj}}=\langle[E_{kj},T],E_{jj}\rangle+\langle L_d(E_{kj}),E_{jj}\rangle$$
$$=\langle T,[E_{jk},E_{jj}]\rangle+\langle L_d(E_{kj}),E_{jj}\rangle$$
$$=-\langle T,E_{jk}\rangle+\langle L_d(E_{kj}),E_{jj}\rangle$$
$$=-t_{jk}+\langle L_d(E_{kj}),E_{jj}\rangle.$$

Subtracting (3.3a) from (3.3b) we get

$$t_{jk} = \frac{1}{2}\left\{\frac{\partial(d_k - d_j)}{\partial z_{kj}} - \langle L_d(E_{kj}),E_{kk} - E_{jj}\rangle\right\}. \tag{3.4}$$

From (3.3c) we obtain

$$\frac{\partial(d_k - d_j)}{\partial z_{kj}} = \frac{\partial^2 t_{jk}}{\partial z_{kj}\partial z_{jk}} - \frac{\partial}{\partial z_{kj}}\langle L_0(E_{jk}),E_{jk}\rangle,$$

and changing the order of differentiation it follows that

$$\begin{aligned}\frac{\partial(d_k - d_j)}{\partial z_{kj}} &= \frac{\partial}{\partial z_{jk}}\langle L_0(E_{kj}),E_{jk}\rangle \\ &\quad - \frac{\partial}{\partial z_{kj}}\langle L_0(E_{jk}),E_{jk}\rangle \\ &= \left\langle\frac{\partial L_0(E_{kj})}{\partial z_{jk}} - \frac{\partial L_0(E_{jk})}{\partial z_{kj}},E_{jk}\right\rangle.\end{aligned} \tag{3.5}$$

From (3.4) and (3.5) we get (3.2a) and (3.2b). □

Theorem 3.2: Let $\tau(F) = r$ and suppose that the entries of Z_{m-1} are linearly independent functions, then

(a) if $r \geqslant m$ then A is C^∞ on Z_k, λ $k = 0,1,\ldots,r-1, \lambda \in C$.

(b) if $r < m$ then $A = 0 = F$.

Proof: Let us prove (b) first.

(i) Suppose that $m = 0$ then A and F are constant on $Z_0,Z_1,\ldots$. Then from (2.1) we get

$$\lambda\, \mathbf{J}A + [Z_0,A] + F = 0.$$

Therefore $[Z_0,A] = 0 \forall Z_0 \in \tau_n(C)$. By (i) of Lemma 2.3 we get $A = 0$ so $F = 0$.

(ii) suppose $m > 0$. Then (2.1) gives us

$$\sum_{j=0}^{k-1} \frac{\partial A}{\partial Z_j}(Z_{j+1}) = \lambda\, \mathbf{J}A + [Z_m,A] + F.$$

Differentiating the equation with respect to Z_m we obtain

$$\frac{\partial A}{\partial Z_{m-1}}(\cdot) = [\cdot,A],$$

or

$$\frac{\partial D}{\partial Z_{m-1}}(\cdot) = \mathbf{P}_d[\cdot,T],$$

$$\frac{\partial T}{\partial Z_{m-1}}(\cdot) = [\cdot,D] + \mathbf{P}_0[\cdot,T],$$

and by Lemma 3.1 we get $D = T = 0$. So $F = 0$. Let us now prove (a).

(iii) Assume $r > m$, then we can prove by induction that $\partial A/\partial Z_k$ is C^∞ on $Z_0,\ldots,Z_{r-1}, \lambda$ $\lambda \in C$ for $m \leqslant k \leqslant r-1$:

$$A = A(Z_0,\ldots,Z_{m-1},0,\ldots,0,t,\lambda) + \int_0^1 \frac{\partial}{\partial s} A(Z_0,\ldots,Z_{m-1},sZ_m,\ldots,sZ_{r-1},t,\lambda)\, ds,$$

and we need to show that

$$A_{m-1}(Z_0,\ldots,Z_{m-1},t,\lambda) = A(Z_0,\ldots,Z_{m-1},0,\ldots,0,t,\lambda)$$

is C^∞ on $Z_0,\ldots,Z_{r-1},\lambda$. Taking the derivative of (2.1) with respect to Z_m and making $Z_m = \cdots = Z_r = 0$, we find

$$\begin{aligned}\frac{\partial A_{m-1}}{\partial Z_{m-1}}(\cdot) &= [\cdot,A_{m-1}] + \lambda\, \mathbf{J}\frac{\partial A(Z_0,\ldots,Z_{m-1},0,\ldots,0,t,\lambda)}{\partial Z_m}(\cdot) \\ &\quad + \frac{\partial F(Z_0,\ldots,Z_{m-1},0,\ldots,0,t_1)}{\partial Z_m}(\cdot) \\ &\quad - \mathbf{L}\frac{\partial A(Z_0,\ldots,Z_{m-1},0,\ldots,0,t,\lambda)}{\partial Z_m}(\cdot).\end{aligned}$$

Taking the projections $\mathbf{P}_d$ and $\mathbf{P}_0$ we get from Lemma 3.1 that A_{m-1} is C^∞ on $Z_0,\ldots,Z_{m-1},\lambda$.

(iv) Suppose $r = m$. Differentiating (2.1) with respect to Z_m we obtain

$$\frac{\partial A}{\partial Z_{m-1}}(\cdot) = [\cdot,A] + \frac{\partial F}{\partial Z_m}(\cdot).$$

Therefore, using Lemma 3.1 it follows that A is C^∞ on $Z_0,\ldots,Z_{m-1}, \lambda$ $\lambda \in C$. □

Since A is C^∞ on $Z_0,\ldots,Z_{r-1}, \lambda$ $\lambda \in C$ $(r \geqslant m)$ then we have

$$A = \sum_{j=0}^{\infty} \lambda^j A_j, \quad \operatorname{tr} A_j = 0, \quad j = 0,1,\ldots .$$

Substituting in (2.1) we get

$$\mathbf{L}A_j = \mathbf{J}A_{j-1} + [Z_m,A_j], \quad j = 1,2,\ldots,$$
$$\mathbf{L}A_0 = [Z_m,A_0] + F,$$

or

$$\mathbf{L}D_j = \mathbf{P}_d[Z_m,T_j], \quad j = 0,1,\ldots \tag{3.6a}$$
$$\mathbf{L}T_j = \mathbf{J}T_{j-1} + [Z_m,D_j] + \mathbf{P}_0[Z_m,T_j], \quad j = 1,2,\ldots, \tag{3.6b}$$
$$\mathbf{L}T_0 = [Z_m,D_0] + \mathbf{P}_0[Z_m,T_0] + F. \tag{3.6c}$$

It follows from (3.6a) and (3.6b) that if $j \geqslant r - m + 1$ then $\tau(T_j) \leqslant m - 1$ and $\tau(D_j) \leqslant m - 1$. Therefore, differentiating (3.6a) and (3.6b) with respect to Z_m, and again using Lemma 3.1 we conclude that for $j \geqslant r - m + 1$, we have $D_j = 0 = T_j$. Equations (3.6a) and (3.6b) also show that $T_{r-m} = 0$ and $D_{r-m} = \text{const}$. Hence, using induction we obtain that D_j and $T_j, j = 0,\ldots,r-m$ $(m \geqslant 1)$ are functions independent of $Z_0,Z_1,\ldots,Z_{m-1}$. Using (3.6c) then F is also independent of $Z_0,Z_1,\ldots,Z_{m-1}$. So

$$D(Z_m,\ldots,Z_{r-1},t,\lambda) = \sum_{k=0}^{r-m} \lambda^k D_k(Z_m,\ldots,Z_{r-1},t,\lambda),$$

$$T(Z_m,\ldots,Z_{r-1},t,\lambda) = \sum_{k=0}^{r-m-1} \lambda^k T_k(Z_m,\ldots,Z_{r-1},t,\lambda),$$

$$F = F(Z_m,\ldots,Z_r,t),$$

and the substitution $V = Z_m$ reduces this case to the previous one $P = U$.

A. $P = z_m\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Let

$$P=\frac{\partial^m u}{\partial x^m}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}=z_m\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad C_0=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

$$C_+=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad C_-=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},$$

and

$$J=\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\text{ with } \alpha_1-\alpha_2=1.$$

Then $D=dC_0$, $T=t^+C_+ + t^-C_-$ and $F=fC_+$.

We can show that D and T are analytic for $\lambda\neq 0$. Hence,

$$D=\Big(\sum_{k=-\infty}^{\infty}\lambda^k d_k\Big)C_0,$$

$$T=\Big(\sum_{k=-\infty}^{\infty}\lambda^k t_k^+\Big)C_+ + \Big(\sum_{k=-\infty}^{\infty}\lambda^k t_k^-\Big)C_-.$$

Since $[C_+,C_-]=-2C_0$, $[C_+,C_0]=-2C_-$, $[C_-,C_0]=-2C_+$, $\mathbf{J}C_+=C_-$, and $\mathbf{J}C_-=C_+$ we get the following equations:

$$\mathbf{L}d_k=-2z_m t_k^-, \tag{3.7a}$$

$$\mathbf{L}t_k^-=t_{k-1}^+-2z_m d_k, \tag{3.7b}$$

$$\mathbf{L}t_k^+=t_{k-1}^-,\quad k\neq 0, \tag{3.7c}$$

$$\mathbf{L}t_0^+=t_{-1}^-+f, \tag{3.7d}$$

with

$$\mathbf{L}g=\sum_{j>0} z_{j+1}\frac{\partial g}{\partial z_j}.$$

(i) For $k<0$ we have $\tau(t_k^-)\leqslant m-1$, $\tau(d_k)\leqslant m-1$ and for $k<0$ $\tau(t_k^+)\leqslant m-1$. Deriving (3.7a) and (3.7b) with respect to z_m we find

$$\left.\begin{aligned}\frac{\partial d_k}{\partial z_{m-1}}&=-2t_k^-,\\ \frac{\partial t_k^-}{\partial z_{m-1}}&=-2d_k,\end{aligned}\right.\qquad k\leqslant 0.$$

The solution of this system is given by

$$d_k=C_k(z_0,\dots,z_{m-2})\cosh 2z_{m-1}, \tag{3.8a}$$

$$t_k^-=C_k(z_0,\dots,z_{m-2})\sinh 2z_{m-1}, \tag{3.8b}$$

where T obeys the normalization condition as $|x|\to\infty$ then $z_k\to 0$ and $T\to 0$. Taking $z_m=0$ in (3.7a) and (3.7b) and using (3.8a) and (3.8b) we find

$$\sum_{j=0}^{m-2} z_{j+1}\frac{\partial C_k}{\partial z_j}=0 \tag{3.9a}$$

and

$$t_{k-1}^+=0\ \text{ for }\ k\leqslant 0. \tag{3.9b}$$

Equation (3.9a) shows that C_k is constant. Substituting (3.9b) in (3.7c) yields

$$t_{k-1}^-=0\ \text{ for }\ k\leqslant -1,$$

so

$$d_k=0\ \text{ for }\ k\leqslant -2.$$

Therefore,

$$d_{-1}=C\cosh 2z_{m-1},$$

$$t_{-1}^-=C\sinh 2z_{m-1}.$$

(ii) Similarly, we show that the regular part of D and T are polynomials in λ, independent of $z_0,z_1,\dots,z_{m-1}$.

ACKNOWLEDGMENT

This work was partially supported by the CNPq-Brazil.

[1] J. A. Cavalcante and J. A. Silva (unpublished).
[2] M. J. Ablowitz, SIAM (1981).

Relations between hyperspherical harmonic transformations and generalized Talmi–Moshinsky transformations

Xian-hui Liu
Center of Theoretical Physics, CCAST (World Laboratory) Beijing, People's Republic of China and Institute of High Energy Physics, Academic Sinica, Beijing, People's Republic of China

(Received 12 July 1989; accepted for publication 14 February 1990)

The correlations between the hyperspherical harmonic transformations and the generalized Talmi–Moshinsky transformations are studied for the three-body and four-body systems. An optical approach for solving few-body problems through diagonalizing the Hamiltonian of a system in an optimal subset of the basis functions of harmonic oscillators in hyperspherical coordinates is proposed. The evaluations of the interaction matrix elements are achieved with the aid of the transformation properties of hyperspherical harmonics.

I. INTRODUCTION

The method of hyperspherical harmonics or k harmonics has been applied extensively in recent years to study many new types of atomic, molecular, and nuclear few-body problems.[1,2] In the hyperspherical coordinates, the $(N-1)$ radial variables for the internal degrees of freedom of an N-body system can be reduced to only one hyperradius continuous variable together with a set of hyperangular quantum numbers. The eigenstates of the total kinetic energy operator can be expressed by using hyperspherical harmonics where the most significant character is the decoupling between the hyperradius associated with the size of the system and the hyperangles associated with the shape of the system. The advantage of hyperspherical coordinates is that the hyperradius is an invariant quantity under coordinate transformation and the rearrangement collision can conveniently be treated with the aid of the transformation properties of hyperspherical harmonics. Progress has recently been made by Lin and Liu.[3] They proposed a method of solving three-body problems in hyperspherical coordinates in adiabatic approximation. The adiabatic channel function is expanded in terms of analytical functions expressed in different sets of Jacobi coordinates to describe each disassociation limit naturally. The evaluation of matrix elements between functions in different Jacobi coordinates is achieved through the known transformation properties of hyperspherical harmonics in these coordinates. At the same time, the generalized Talmi–Moshinsky transformations have also been extensively used in the calculations of few-body problems. The approach for obtaining the bound-state solutions of the systems with short-range interactions by using harmonic oscillator product states as basis functions is well developed by many authors.[4] It is a very convenient method, although it is not a method with the best accuracy. The evaluations of matrix elements can be carried out with the aid of the generalized Talmi–Moshinsky transformations.

In this paper, we attempt to further study certain correlations between the above-mentioned two kinds of basis functions from the point of view of few-body problems. In Sec. II, the basis functions of the harmonic oscillator in hyperspherical coordinates are presented. These basis functions are related with the harmonic oscillator product states in Jacobi coordinates through the orthogonal transformations. The correlations between the hyperspherical harmonic transformations and the generalized Talmi–Moshinsky transformations are found with the aid of the above-mentioned orthogonal transformations. An optional approach for solving the few-body problems through diagonalizing the Hamiltonian of the system in an optimal subset of basis functions of the hyperspherical harmonic oscillators is proposed. A brief conclusion is given in Sec. III.

II. CORRELATIONS BETWEEN THE HYPERSPHERICAL HARMONIC TRANSFORMATIONS AND THE GENERALIZED TALMI–MOSHINSKY TRANSFORMATIONS

Let us consider an N-body system. Let $\vec{\gamma}_i$ denote the coordinates of particle i with mass m_i in laboratory frame. A possible set of Jacobi coordinates is denoted by $\vec{\rho}_i$ with reduced mass μ_i. Let $M = \Sigma_i m_i$ be the total mass of the system and $\vec{u}$ be the coordinates of the center of mass. There exist two important identities for the different sets of Jacobi coordinates for the N-body system:

$$\sum_{i=1}^{N} m_i \vec{\gamma}_i^2 = \sum_{j=1}^{N-1} \mu_j \vec{\rho}_j^2 + M\vec{u}^2, \tag{1}$$

$$\sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial^2}{\partial^2 \vec{\gamma}_i} = \sum_{j=1}^{N-1} \frac{1}{\mu_j} \frac{\partial^2}{\partial^2 \vec{\rho}_j} + \frac{1}{M} \frac{\partial^2}{\partial^2 \vec{u}}. \tag{2}$$

If we further introduce a set of mass-weighted vectors,

$$\vec{\xi}_j = (\mu_j/\mu)^{1/2} \vec{\rho}_j, \tag{3}$$

where μ is arbitrary, Eqs. (1) and (2) will lead up to the following identities:

$$\sum_{j=1}^{N-1} \mu_j^{(\alpha)} \vec{\rho}_j^{(\alpha)^2} = \sum_{j=1}^{N-1} \mu_j^{(\beta)} \vec{\rho}_j^{(\beta)^2}, \tag{4a}$$

$$\sum_{j=1}^{N-1} \frac{1}{\mu_j^{(\alpha)}} \nabla^2_{\vec{\rho}_j^{(\alpha)}} = \sum_{j=1}^{N-1} \frac{1}{\mu_j^{(\beta)}} \nabla^2_{\vec{\rho}_j^{(\beta)}}, \tag{4b}$$

and

$$\sum_{j=1}^{N-1} \vec{\xi}_j^{(\alpha)^2} = \sum_{j=1}^{N-1} \vec{\xi}_j^{(\beta)^2}, \tag{5a}$$

$$\sum_{j=1}^{N-1} \nabla^2_{\vec{\xi}_j^{(\alpha)}} = \sum_{j=1}^{N-1} \nabla^2_{\vec{\xi}_j^{(\beta)}}, \tag{5b}$$

where α and β denote two arbitrary sets of different Jacobi coordinates. For a three-body system, there exists three sets

 0022-2488/90/071621-06$03.00

of different Jacobi coordinates as shown in Fig. 1. For a four-body system, there are 15 sets of different Jacobi coordinates, shown in Fig. 2 as examples. Equations (4a) and (4b) are the foundation of the generalized Talmi–Moshinsky transformations and Eqs. (5a) and (5b) are the foundation of the hyperspherical harmonic transformations. In the following we will discuss, in turn the three-body and four-body systems.

A. Three-body system

The kinetic energy operator for the three-body system in the center-of-mass coordinate system is expressed as

$$T = -\frac{\hbar^2}{2}\sum_{j=1}^{2}\frac{1}{\mu_j}\nabla^2_{\vec{\rho}_j}, \tag{6a}$$

or

$$T = -\frac{\hbar^2}{2\mu}\sum_{j=1}^{2}\nabla^2_{\vec{\xi}_j}. \tag{6b}$$

By defining a hyperspherical radius ξ and a hyperangle ϕ,

$$\xi^2 = \xi_1^2 + \xi_2^2, \quad \tan\phi = \xi_2/\xi_1, \tag{7}$$

Eq. (6b) is rewritten as

$$T = -\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial^2\xi} + \frac{5}{\xi}\frac{\partial}{\partial\xi} - \frac{\Lambda^2(\Omega)}{\xi^2}\right), \tag{8}$$

where Ω denotes the assembly of five angles, $\Omega = (\phi,\hat{\xi}_1,\hat{\xi}_2)$ and Λ^2 is the grand angular momentum operator

$$\Lambda^2(\Omega) = -\frac{1}{\sin^2\phi\cos^2\phi}\frac{d}{d\phi}\left(\sin^2\phi\cos^2\phi\frac{d}{d\phi}\right) + \frac{\hat{l}_1^2(\hat{\xi}_1)}{\sin^2\phi} + \frac{\hat{l}_2^2(\hat{\xi}_2)}{\cos^2\phi}. \tag{9}$$

The normalized eigenfunctions $Y_{[K]}$ of the operator $\Lambda^2(\Omega)$ are well known:

$$Y_{[K]}(\Omega) = Q_m^{l_2 l_1}(\phi)\,Y_{l_1 l_2 LM}(\hat{\xi}_1,\hat{\xi}_2), \tag{10}$$

satisfying the eigenequation

$$\Lambda^2(\Omega)Y_{[K]}(\Omega) = \lambda_K(\lambda_K+4)Y_{[K]}(\Omega) \tag{11}$$

$$\lambda_K = 2m + l_1 + l_2, \tag{12}$$

and $[k]$ denotes an aggregate of quantum numbers m,l_1,l_2,L,M, and

$$Q_m^{l_2 l_1}(\phi) = \theta_m^{l_2 l_1}\sin^{l_2}\phi\cos^{l_1}\phi \times P_m^{(l_2+1/2,\,l_1+1/2)}(\cos 2\phi), \tag{13}$$

where $\theta_m^{l_2 l_1}$ is the normalization constant and $P_m^{(\alpha,\beta)}(\cos 2\phi)$ is the Jacobi polynomial. In Eq. (12), $Y_{l_1 l_2 LM}$ is the coupled total orbital angular momentum function. The eigenfunctions $Y_{[K]}(\Omega)$ form a complete and orthonormal set satisfying

$$\int d\Omega\, Y^*_{[K']}(\Omega)Y_{[K]}(\Omega) = \delta_{[K'][K]}, \tag{14}$$

where the volume element $d\Omega = \cos^2\phi\sin^2\phi\, d\phi\, d\hat{\xi}_1\, d\hat{\xi}_2$.

The above equations, (6)–(14), can be written for each of the three sets of Jacobi coordinates. From Eqs. (5), (6b), and (8), it is clear that the grand angular momentum operator is independent of the chosen Jacobi coordinates

$$\Lambda^2(\Omega^\alpha) = \Lambda^2(\Omega^\beta) = \Lambda^2(\Omega^\gamma). \tag{15}$$

Therefore, it is possible to expand the eigenfunction of one set in terms of the eigenfunctions of another set within the given subset with $\lambda_{K'} = \lambda_K$:

$$Y_{[K]}(\Omega^\alpha) = \sum_{[K']} A^{[K]\alpha-\alpha'}_{[K']}(\eta_{\alpha\alpha'})Y_{[K']}(\Omega^{\alpha'}). \tag{16}$$

Here, α' denotes either a β or γ set and $\eta_{\alpha\alpha'}$ is related to the mass ratio of the particles. The expansion coefficients

$$A^{[K]\alpha-\alpha'}_{[K']}(\eta_{\alpha\alpha'}) = \int d\Omega^{\alpha'}\, Y^*_{[K']}(\Omega^{\alpha'})Y_{[K]}(\Omega^\alpha) \tag{17}$$

are called the transformation brackets of the hyperspherical harmonics (TB).

Now let us introduce a Hamiltonian:

$$H_0 = -\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial^2\xi} + \frac{5}{\xi}\frac{\partial}{\partial\xi} - \frac{\Lambda^2(\Omega)}{\xi^2}\right) + \frac{1}{2}\mu\omega^2\xi^2. \tag{18}$$

The eigenfunctions of H_0 are the wave functions of the harmonic oscillator in the hyperspherical coordinates with the formula

$$\Psi^{\text{hs}}_{[I]} = R_{N\nu}(\xi)Y_{[K]}(\Omega)/\xi^{5/2} = |Nml_1l_2;LM\rangle, \tag{19}$$

where $[I]$ denotes an aggregate of quantum numbers N,m,l_1,l_2,L,M. The corresponding eigenvalue is

$$E^0_{[I]} = (2N + \nu + \tfrac{3}{2})\hbar w. \tag{20}$$

Here, $\nu = \lambda_K + \frac{3}{2} = 2m + l_1 + l_2 + \frac{3}{2}$ and N is the radial quantum number of the harmonic oscillator. Also, $R_{N\nu}$ satisfies the following radial equation of the harmonic oscillator:

$$\left\{\frac{d^2}{d^2\xi} + \frac{2\mu}{\hbar^2}E^0_{[I]} - \frac{\mu^2\omega^2}{\hbar^2}\xi^2 - \frac{\nu(\nu+1)}{\xi^2}\right\}R_{N\nu} = 0. \tag{21}$$

The eigenfunctions $\Psi^{\text{hs}}_{[I]}$ form a complete and orthonormal set. The wave function of the three-body system can be expanded in terms of the basis functions $\Psi^{\text{hs}}_{[I]}$ and an approximate solution can be obtained through diagonalizing the Hamiltonian of the system in an optimal subset. This approach has been used to solve the three cluster structure of

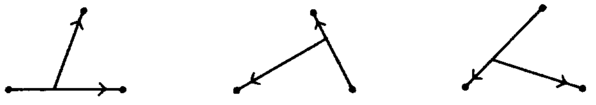


FIG. 1. Three different sets of Jacobi coordinates for the three-body systems.

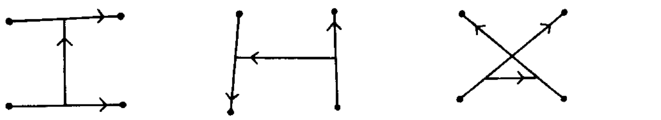


FIG. 2. The different sets of Jacobi coordinates for the four-body systems.

light hypernuclei and proved to be successful.[5] The evaluation of matrix elements is carried out with the aid of the transformation brackets of the hyperspherical harmonics. A variational parameter $\hbar w$ is introduced.

An optional method can be obtained through rewriting Eq. (18) as

$$H_0 = \sum_{j=1}^{2} \left(-\frac{\hbar^2}{2\mu} \nabla^2_{\xi_j} + \frac{1}{2} \mu w^2 \vec{\xi}_j^{\,2} \right)$$
$$= \sum_{j=1}^{2} \left(-\frac{\hbar^2}{2\mu_j} \nabla^2_{\bar{\rho}_j} + \frac{1}{2} \mu_j w^2 \vec{\rho}_j^{\,2} \right). \tag{22}$$

The eigenfunctions of Eq. (22) are the two harmonic oscillator product states:

$$\Psi^{\rm ho}_{[J]} = (\varphi_{n_1 l_1}(\vec{\xi}_1)\varphi_{n_2 l_2}(\vec{\xi}_2))_{LM}$$
$$= |n_1 l_1 n_2 l_2; LM\rangle, \tag{23}$$

with the eigenvalues $E^0_{[J]} = (2n_1 + l_1 + 2n_2 + l_2 + 3)\hbar w$. Here, $[J]$ denotes an aggregate of quantum numbers n_1, l_1, n_2, l_2, L, M. The functions $\Psi^{\rm ho}_{[J]}$ also form an orthonormal and complete set. The wave function of the three-body system can also be expanded in terms of the basis functions $\Psi^{\rm ho}_{[J]}$ and an approximate solution can be obtained through diagonalizing the Hamiltonian of the system in an optimal basis set. This approach has also successfully been used in the calculations of three-body problems with the short-range interactions.[6] The evaluation of matrix elements is achieved with the aid of the generalized Talmi–Moshinsky transformations.

The wave functions considered here include only the spatial part. For the systems including the identical particles, the spatial wave functions have to be combined properly with the spin functions to obtain correct overall symmetries for the total wave functions. This combination depends on whether the identical particles are Fermi particles or Bose particles.

Both of above-mentioned methods are intrinsically equivalent to each other. The $\Psi^{\rm hs}_{[I]}$ and $\Psi^{\rm ho}_{[J]}$ are the solutions of the same Hamiltonian with the same boundary conditions. Both of them are correlated to each other through an orthogonal transformation. Then $\Psi^{\rm hs}_{[I]}$ can be expanded in terms of $\Psi^{\rm ho}_{[J]}$ or converse:

$$|Nml_1 l_2; LM\rangle = \sum_{n_1, n_2} C^{Nm}_{n_1 n_2}(l_1 l_2)|n_1 l_1 n_2 l_2; LM\rangle,$$

i.e.,

$$\frac{R_{N\nu}(\xi)\,Y_{[K]}(\Omega)}{\xi^{5/2}} = \sum_{n_1, n_2 (n_1 + n_2 = N + m)} C^{Nm}_{n_1 n_2}(l_1 l_2)$$
$$\times (\varphi_{n_1 l_1}(\vec{\xi}_1)\varphi_{n_2 l_2}(\vec{\xi}_2))_{LM}. \tag{24}$$

Here, $C^{Nm}_{n_1 n_2}(l_1 l_2)$ is the expansion coefficient whose explicit formula is given in the Appendix and there exists a symmetry:

$$C^{Nm}_{n_1 n_2}(l_1 l_2) = (-1)^m C^{Nm}_{n_2 n_1}(l_2 l_1).$$

These formulas can be written for each of the three sets of Jacobi coordinates. Since the hyperradius ξ and so $R_{N\nu}(\xi)$ are invariant under the coordinate transformation, and by using the orthonormal property of $R_{N\nu}(\xi)/\xi^{5/2}$, we obtain

$$\langle Y^{(\alpha')}_{[K']}|Y^{(\alpha)}_{[K]}\rangle = \sum_{\substack{n_1, n_2 \\ (n_1 + n_2 = m)}} \sum_{\substack{n'_1, n'_2 \\ (n'_1 + n'_2 = m')}} C^{0\ m'}_{n'_1 n'_2}(l'_1 l'_2)\delta_{\lambda_K \lambda'_K}$$
$$\times {}_{(\alpha')}\langle n'_1 l'_1 n'_2 l'_2; L|n_1 l_1 n_2 l_2; L\rangle_{(\alpha)}$$
$$\times C^{0\ m}_{n_1 n_2}(l_1 l_2). \tag{25}$$

The left-hand side of Eq. (25) is just the transformation bracket of the hyperspherical harmonics, Eq. (17), and

$${}_{(\alpha')}\langle n'_1 l'_1 n'_2 l'_2; L|n_1 l_1 n_2 l_2; L\rangle_{(\alpha)}$$
$$= \langle (\varphi_{n'_1 l'_1}(\vec{\rho}_1^{(\alpha')})\varphi_{n'_2 l'_2}(\vec{\rho}_2^{(\alpha')}))_{LM}|$$
$$\times (\varphi_{n_1 l_1}(\vec{\rho}_1^{(\alpha)})\varphi_{n_2 l_2}(\vec{\rho}_2^{(\alpha)}))_{LM}\rangle$$

is just the generalized Talmi–Moshinsky transformation coefficient (GTM). Equations (24) and (25) are just the relations we want to find.

B. Four-body system

A procedure similar to the one done in Sec. II A can be shown for the four-body system. There exist 15 different sets of Jacobi coordinates for a four-body system (cf. Ref. 7). The kinetic-energy operator of the four-body system in an arbitrary set of Jacobi coordinates is given by

$$T = -\frac{\hbar^2}{2} \sum_{j=1}^{3} \frac{1}{\mu_j} \nabla^2_{\bar{\rho}_j}, \tag{26a}$$

or

$$T = -\frac{\hbar^2}{2\mu} \sum_{j=1}^{3} \nabla^2_{\xi_j}. \tag{26b}$$

By defining a hyperradius and two hyperangles,

$$\xi_1 = \xi \cos\phi_1,$$
$$\xi_2 = \xi \sin\phi_1 \cos\phi_2, \tag{27}$$
$$\xi_3 = \xi \sin\phi_1 \sin\phi_2,$$

the kinetic-energy operator is rewritten as

$$T = -\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial^2 \xi} + \frac{8}{\xi}\frac{\partial}{\partial \xi} - \frac{\Lambda^2(\omega)}{\xi^2} \right), \tag{28}$$

where ω denotes the assembly of eight angles, $\omega = (\phi_1, \phi_2, \hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3)$ and $\Lambda^2(\omega)$ is well known as the grand angular momentum operator:

$$\Lambda^2(\omega) = -\frac{\partial^2}{\partial^2 \phi_1} - (5\cot\phi_1 - 2\tan\phi_1)\frac{\partial}{\partial\phi_1}$$
$$+ \frac{\hat{l}_1^2(\hat{\xi}_1)}{\cos^2\phi_1} + \frac{1}{\sin^2\phi_2}\left(-\frac{\partial^2}{\partial^2\phi_2} - 2(\cot\phi_2 \right.$$
$$\left. - \tan\phi_2)\frac{\partial}{\partial\phi_2} + \frac{\hat{l}_2^2(\hat{\xi}_2)}{\cos^2\phi_2} + \frac{\hat{l}_3^2(\hat{\xi}_3)}{\sin^2\phi_2} \right), \tag{29}$$

where $\hat{l}_i(\hat{\xi}_i)$ is the orbital angular momentum operator associated with $\vec{\xi}_i$. The eigenfunctions of $\Lambda^2(\omega)$ are

$$Y_{[K]}(\omega) = Q^{l_1 l_2 l_3}_{m_1 m_2}(\phi_1,\phi_2)(y_{l_1}(\hat{\xi}_1)(y_{l_2}(\hat{\xi}_2)y_{l_3}(\hat{\xi}_3))_{l_{23}})_{LM} \tag{30}$$

fulfilling the eigenequation

$$\Lambda^2(\omega)Y_{[K]}(\omega) = \lambda_{[K]}(\lambda_{[K]}+7)Y_{[K]}(\omega). \tag{31}$$

Here, $[K]$ denotes an aggregate of eight quantum numbers $m_1,m_2,l_1,l_2,l_3,l_{23},L,M$ and $\lambda_{[K]} = 2m_1 + 2m_2 + l_1 + l_2 + l_3$,

$$Q^{l_1 l_2 l_3}_{m_1 m_2}(\phi_1,\phi_2) = \sin^{-3/2}\phi_1 Q^{\lambda_K + 3/2, l_1}_{m_1}(\phi_1) Q^{l_3,l_2}_{m_2}(\phi_2), \tag{32}$$

and $Q^{\alpha,\beta}_m$ is defined by Eq. (13), $\lambda_K = 2m_2 + l_2 + l_3$. The function $Y_{[K]}(\omega)$ form a complete and orthonormal set

$$\int dw\, Y^*_{[K']}(\omega)Y_{[K]}(\omega) = \delta_{[K'][K]}, \tag{33}$$

where the volume element

$$dw = \sin^5\phi_1 \cos^2\phi_1 \sin^2\phi_2 \cos^2\phi_2\, d\phi_1\, d\phi_2\, d\hat{\xi}_1\, d\hat{\xi}_2\, d\hat{\xi}_3.$$

Equations (26)–(33) can be written for each of the 15 sets of Jacobi coordinates. From Eqs. (5) and (28), it is quite obvious that the hyperradius ξ, T operator and so $\Lambda^2(\omega)$ are invariant under the coordinate transformations. Then, being similar to the three-body case, the eigenfunction in one set of Jacobi coordinates can be expanded in terms of the eigenfunctions in another set within the given subset with $\lambda_{[K']} = \lambda_{[K]}$, i.e.,

$$Y_{[K]}(\omega^\alpha) = \sum_{[K']} A^{[K]\alpha\to\beta}_{[K']}(\eta_{\alpha\beta})Y_{[K']}(\omega^\beta), \tag{34}$$

and the transformation bracket for the four-body system (TB) is

$$\begin{aligned} A^{[K]\alpha\to\beta}_{[K]}(\eta_{\alpha\beta}) &= \int dw^\beta\, Y^*_{[K']}(\omega^\beta)Y_{[K]}(\omega^\alpha) \\ &= \langle Y_{[K']}(\omega^\beta)|Y_{[K]}(\omega^\alpha)\rangle. \end{aligned} \tag{35}$$

Here, α and β denote two arbitrary sets of 15 different Jacobi coordinates.

We further introduce a four-body Hamiltonian now

$$H_0 = -\frac{\hbar^2}{2\mu}\left(\frac{\partial^2}{\partial^2\xi} + \frac{8}{\xi}\frac{\partial}{\partial\xi} - \frac{\Lambda^2(\omega)}{\xi^2}\right) + \frac{1}{2}\mu w^2\xi^2. \tag{36}$$

The eigensolutions of Eq. (36) can be written in the following formula:

$$\begin{aligned} \Psi^{\rm hs}_{[I]} &= R_{N\nu}(\xi)Y_{[K]}(\omega)/\xi^4 \\ &= |Nm_1m_2l_1l_2l_3l_{23};LM\rangle; \end{aligned} \tag{37}$$

$R_{N\nu}$ satisfies the following equation:

$$\left\{\frac{d^2}{d^2\xi} + \frac{2\mu}{\hbar^2}E^0_{[I]} - \frac{\mu^2\omega^2}{\hbar^2}\xi^2 - \frac{\nu(\nu+1)}{\xi^2}\right\}R_{N\nu} = 0, \tag{38}$$

where $\nu = \lambda_{[K]} + 3 = 2m_1 + 2m_2 + l_1 + l_2 + l_3 + 3$. The solutions of Eq. (38) are the typical radial wave functions of the harmonic oscillator with the quantum numbers N,ν, and the corresponding eigenvalues are

$$E^0_{[I]} = (2N + \nu + 3/2)\hbar w. \tag{39}$$

Here, $[I]$ denotes an aggregate of nine quantum numbers $(N,m_1,m_2,l_1,l_2,l_3,l_{23},L,M)$. The eigenfunctions $\Psi^{\rm hs}_{[I]}$ form a complete and orthonormal set. The wave functions of the four-body system can be expanded in terms of the basis functions $\Psi^{\rm hs}_{[I]}$ and an approximate solution of the system can be obtained through diagonalizing the Hamiltonian of the system in an optimal basis set. The evaluation of the interactions matrix elements can be carried out with the aid of the transformation brackets of the hyperspherical harmonics of the four-body system. A variational parameter $\hbar w$ is introduced.

The other optional method is to rewrite Eq. (30) as

$$H_0 = \sum_{j=1}^{3}\left(-\frac{\hbar^2}{2\mu}\nabla^2_{\vec{\xi}_j} + \frac{1}{2}\mu w^2\xi_j^2\right), \tag{40a}$$

or

$$H_0 = \sum_{j=1}^{3}\left(-\frac{\hbar^2}{2\mu_j}\nabla^2_{\vec{\rho}_j} + \frac{1}{2}\mu_j w^2\vec{\rho}_j^{\,2}\right). \tag{40b}$$

The eigenfunctions of Eq. (40) are the product states of three harmonic oscillators

$$\begin{aligned} \Psi^{\rm ho}_{[J]} &= (\varphi_{n_1l_1}(\vec{\xi}_1)(\varphi_{n_2l_2}(\vec{\xi}_2)\varphi_{n_3l_3}(\vec{\xi}_3))_{l_{23}})_{LM} \\ &= |n_1l_1n_2l_2n_3l_3l_{23};LM\rangle, \end{aligned} \tag{41}$$

with the eigenvalue

$$E^0_{[J]} = (2(n_1+n_2+n_3) + l_1 + l_2 + l_3 + 9/2)\hbar\omega.$$

Here, $[J]$ denotes the aggregate of nine quantum numbers $(n_1,l_1,n_2,l_2,n_3,l_3,l_{23},L,M)$. The eigenfunctions $\Psi^{\rm ho}_{[J]}$ also form a complete and orthonormal set. The wave function of the four-body system can be expanded in terms of the basis functions $\Psi^{\rm ho}_{[J]}$ and an approximate solution can be obtained through diagonalizing the Hamiltonian of the system in the optimal basis set. This approach has also been successfully used in the calculations of the four-body problems with the short-range interactions. The evaluation of matrix elements is achieved with the aid of the generalized Talmi–Moshinsky transformation of four body.

The wave functions $\Psi^{\rm ho}_{[J]}$ and $\Psi^{\rm hs}_{[I]}$ are the solutions of the same Hamiltonian with the same boundary conditions. Both of them are correlated to each other through an orthogonal transformation. Then we can expand $\Psi^{\rm hs}_{[I]}$ in terms of $\Psi^{\rm ho}_{[J]}$ or converse

$$\begin{aligned} &|Nm_1m_2l_1l_2l_3l_{23};LM\rangle \\ &\quad = \sum_{n_1,n_2,n_3} C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3)|n_1l_1n_2l_2n_3l_3l_{23};LM\rangle, \end{aligned}$$

i.e.,

$$R_{N\nu}(\xi)Y_{[K]}(\omega)/\xi^4 = \sum_{\substack{n_1,n_2,n_3 \\ (n_1+n_2+n_3=N+m_1+m_2)}} C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3) \times(\varphi_{n_1l_1}(\vec{\xi}_1)(\varphi_{n_2l_2}(\vec{\xi}_2)\varphi_{n_3l_3}(\vec{\xi}_3))_{l_{23}})_{LM}. \tag{42}$$

All the above formulas can be written for each of the 15 sets of Jacobi coordinates. Since the hyperradius ξ and so $R_{N\nu}$ are invariant under the coordinate transformation and by using the orthonormal property of $R_{N\nu}/\xi^4$, we have

$$\langle Y_{[K']}(\omega^\beta)|Y_{[K]}(\omega^\alpha)\rangle = \sum_{\substack{n_1,n_2,n_3 \\ (n_1+n_2+n_3=m_1+m_2)}} \sum_{\substack{n'_1,n'_2,n'_3 \\ (n'_1+n'_2+n'_3=m'_1+m'_2)}} \times C^{0m_1m_2}_{n_1n_2n_3}(l_1l_2l_3)C^{0m'_1m'_2}_{n'_1n'_2n'_3}(l'_1l'_2l'_3) \times{}_{(\beta)}\langle n'_1l'_1n'_2l'_2n'_3l'_3l'_{23};LM \times|n_1l_1n_2l_2n_3l_3l_{23};LM\rangle_{(\alpha)}. \tag{43}$$

The left-hand side of Eq. (43) is just the transformation brackets of hyperspherical harmonics shown in Eq. (35) and the symbol in the right-hand side

$$_{(\beta)}\langle n'_1l'_1n'_2l'_2n'_3l'_3l'_{23};LM|n_1l_1n_2l_2n_3l_3l_{23};LM\rangle_{(\alpha)} = \langle(\varphi_{n'_1l'_1}(\vec{\xi}_1^{(\beta)})(\varphi_{n'_2l'_2}(\vec{\xi}_2^{(\beta)})\varphi_{n'_3l'_3}(\vec{\xi}_3^{(\beta)}))_{l'_{23}})_{LM} \times|(\varphi_{n_1l_1}(\vec{\xi}_1^{(\alpha)})(\varphi_{n_2l_2}(\vec{\xi}_2^{(\alpha)})\varphi_{n_3l_3}(\vec{\xi}_3^{(\alpha)}))_{l_{23}})_{LM}\rangle$$

is the generalized Talmi–Moshinsky transformation coefficient (GTM) from the α set to the β set. The closed formula of the GTM coefficients was rederived by Bao and a computation program was also given.[7] There the GTM coefficients were represented by the symbol TM4 $(n_1,l_1,n_2,l_2,n_3,l_3,n'_1,l'_1, n'_2,l'_2,n'_3,l'_3,l_{23},l'_{23},L,X_1,X_2,X_3,X_4,K)$, and x_1,x_2,x_3,x_4 denote the masses of the four particles, respectively. Index K is used to classify the transformations from a certain set to another set (cf. Ref. 7). The coefficients $C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3)$ are given in the Appendix. There exist a symmetry relation

$$C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3) = (-1)^{m_2}C^{Nm_1m_2}_{n_1n_3n_2}(l_1l_3l_2).$$

Equations (42) and (43) are just the relations we attempt to find.

III. CONCLUSION

The correlations between the transformation brackets of hyperspherical harmonics and the generalized Talmi–Moshinsky transformation coefficients are obtained. This is significant for understanding the intrinsic connection between two kinds of basis functions mentioned above. It may be utilized for calculating one kind of both TB and GTM coefficients in terms of another kind and checking the formula of the transformation coefficients used in the calculations. An optional approach of solving few-body problems is proposed. In this method, the wave function of the system is expanded in terms of the basis functions of the harmonic oscillator in hyperspherical coordinates and the approximate solutions are achieved through diagonalizing the Hamiltonian of the system in an optimal basis set. The advantage of this approach is that the hyperradius ξ is an invariant quantity under the coordinate transformation. The evaluation of matrix elements can be carried out in terms of the transformation brackets of hyperspherical harmonics. The number of the transformation bracket needed in the calculations is less than that of the generalized Talmi–Moshinsky coefficients needed in calculating matrix elements since the quantum numbers taking part in the transformation in hyperspherical coordinates are less one than that in the harmonic oscillator product states. The similar procedure can be generalized to the cases of the number of particles $N\geqslant 5$.

ACKNOWLEDGMENTS

This work is supported by the National Nature Science Foundation of China.

APPENDIX

In this appendix we give the explicit expressions of $C^{Nm}_{n_1n_2}(l_1l_2)$ and $C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3)$ for three-body and four-body systems, respectively. In order to be concise, we rewrite the formulas mentioned in Sec. II as follows:

$$R_{nl}(\rho) = \sum_{p=0}^{n} B(nlp)\rho^{l+2p}\frac{\exp(-\rho^2/2)}{\gamma_\mu^{3/2}}.$$

Here,

$$B(nlp) = \left[\frac{2n!}{\Gamma(n+l+3/2)}\right]^{1/2}(-1)^p\binom{n+l+1/2}{n-p}\frac{1}{p!},$$

$$Q^{\tau_1\tau_2}_m(\phi) = \sum_{k=0}^{m} D(m\tau_1\tau_2k)\sin^{\tau_1+2(m-k)}\phi\cos^{\tau_2+2k}\phi.$$

Here,

$$D(m\tau_1\tau_2k) = \theta^{\tau_1\tau_2}_m(-1)^{m-k} \times\binom{m+\tau_1+1/2}{k}\binom{m+\tau_2+1/2}{m-k},$$

$$\theta^{\tau_1\tau_2}_m = \left(\frac{2(2n+\tau_1+\tau_2+2)m!\Gamma(m+\tau_1+\tau_2+2)}{\Gamma(m+\tau_1+3/2)\Gamma(m+\tau_2+3/2)}\right)^{1/2}$$

and define

$$E(x) = \int_0^\infty \exp(-\rho^2)\rho^x\,d\rho = \begin{cases} \frac{1}{2}\left(\frac{x-1}{2}\right)!, & \text{if } x=\text{odd}, \\ \frac{(x-1)!!}{2^{x/2+1}}\sqrt{\pi}, & \text{if } x=\text{even}, \end{cases}$$

$$SC(I,J) = \int_0^{\pi/2} d\phi\sin^I\phi\cos^J\phi = \epsilon\frac{(I-1)!!(J-1)!!}{(I+J)!!},$$

$$\epsilon = \begin{cases} \pi/2, & \text{if } I \text{ and } J \text{ are the even integers}, \\ 1 & \text{otherwise}. \end{cases}$$

Then we have

$$C^{Nm}_{n_1n_2}(l_1l_2) = \sum_{p_1}^{n_1}\sum_{p_2}^{n_2}\sum_{p_3}^{N} B(N\nu p_3)B(n_1l_1p_1)B(n_2l_2p_2)\,\frac{(2+m+l_1+l_2+p_1+p_2+p_3)!}{2}\sum_{k}^{m} D(ml_2l_1k)$$
$$\times \mathrm{Sc}(2(l_2+m-k+p_1+1),2(l_1+k+p_1+1))\delta_{N+m,n_1+n_2},$$

where,

$$\nu = 2m + l_1 + l_2 + 3/2,$$

$$C^{Nm_1m_2}_{n_1n_2n_3}(l_1l_2l_3) = \sum_{p_1}^{n_1} B(n_1l_1p_1)\sum_{p_2}^{n_2} B(n_2l_2p_2)\sum_{p_3}^{n_3} B(n_3l_3p_3)\sum_{p_4}^{N} B(N\nu p_4)E(\nu + 5 + l_1 + l_2 + l_3 + 2$$
$$\times (p_1+p_2+p_3+p_4))\sum_{k_1}^{m_1} D\left(m_1\left(\lambda_k + \frac{3}{2}\right)l_1k_1\right)\sum_{k_2}^{m_2} D(m_2l_3l_2k_2)SC(\lambda_k + 5 + l_2 + l_3$$
$$+ 2(m_1+p_2+p_3-k_1),2(l_1+k_1+p_1+1))SC(2(l_3+m_2-k_2+p_3+1),2(l_2+k_2+p_2+1))$$
$$\times \delta_{N+m_1+m_2,n_1+n_2+n_3},$$

where $\lambda_K = 2m_2 + l_2 + l_3$ and $\nu = 2(m_1+m_2) + l_1 + l_2 + l_3$.

[1]C. D. Lin, Adv. Mol. Phys. **22**, 77 (1986); Phys. Rev. A **29**, 1019 (1984).

[2]S. Y. Larsen, in *Few-Body Methods: Principles and Applications*, edited by T. K. Lim, G. Bao, D. P. Hou, and S. Huber (World Scientific, Singapore, 1986), p. 467; M. I. Haftel, *ibid.*, p. 793.

[3]C. D. Lin and X. H. Liu, Phys. Rev. A **37**, 2749 (1988).

[4]F. Miglietta, Nuovo. Cimento A **49**, 187 (1979). V. C. Aguilera-Mavarro and O. Portilho, Ann. Phys. **107**, 126 (1979). O. Portilho and S. A. Coon, Z. Phys. A **290**,93· (1979).

[5]X. H. Liu, preprint.

[6]C. G. Bao, in *Few-Body Methods: Principles and Applications*, edited by T. K. Lim, C. G. Bao, D. P. Hou, and S. Huber (World Scientific, Singapore, 1986) and references therein; C. Y. Zhang, F. X. Kong, and X. H. Liu, Nucl. Phys. A **500**, 627 (1989).

[7]Y. P. Gan, M. Z. Gong, C. E. Wu, and C. G. Bao, Comp. Phys. Comm. **34**, 387 (1985).

[8]J. Raynal, Nucl. Phys. A **259**, 272 (1976).

Lie group symmetries and invariants of the Hénon–Heiles equations

Barbara Abraham-Shrauner
Department of Electrical Engineering, Washington University, St. Louis, Missouri 63130

(Received 3 February 1989; accepted for publication 14 February 1990)

Lie group symmetries and invariants of the generalized Hénon–Heiles equations are found. The coupled second-order equations are invariant under translations in time, in general, and the stretching group (dilation) if the linear terms in the "force" are absent. The equivalent set of four coupled first-order equations is found to be invariant under a one-parameter group for three cases and the group generators are given. Three different approaches are reported: the "classical method" for determining Lie group symmetries, a modified method for finding Lie group symmetries with vector fields and the direct method for calculating the invariants. For the Hénon–Heiles equations the direct method is the most efficient.

I. INTRODUCTION

The integrability of the two-dimensional Hénon–Heiles equations, which have been regarded as a prototype of coupled nonlinear first-order differential equations, has been analyzed by several approaches.[1–11] The equations were originally found for gravitating stars in a cylindrical galaxy.[1] Presented here is an attempt to determine the Lie point group symmetries of these equations and to relate the symmetries to the determination of a second invariant. The Hénon–Heiles equations can be viewed as the equations of motion of a dynamical system whose Hamiltonian H is the energy E, where E is the first invariant. We call a system of equations integrable if the corresponding dynamical system in a $2n$-dimensional phase space has n constants of motion or invariants in involution (independent) since the Liouville theorem[11] states that the other n invariants can be found once n invariants are known. For our system, the phase space is four dimensional and we need find only two independent invariants.

The generalized Hénon–Heiles equations are

$$\ddot{x} + Ax + 2Dxy = 0, \tag{1}$$

$$\ddot{y} + By + Dx^2 - Cy^2 = 0, \tag{2}$$

where A, B, C, and D are constants and the overdots denote differentiation with respect to time t. These equations are autonomous since there is no explicit dependence on time t in the equations. Multiplication of Eq. (1) by $\dot{x}$ and Eq. (2) by $\dot{y}$ and integration with respect to t gives the first invariant, the energy E,

$$E = \frac{\dot{x}^2}{2} + \frac{\dot{y}^2}{2} + \frac{Ax^2}{2} + \frac{By^2}{2} + Dx^2y - \frac{Cy^3}{3}, \tag{3}$$

where $f = Ax^2/2 + By^2/2 + Dx^2y - Cy^3/3$ is used later on. An additional invariant quadratic in the velocities has been found previously for two special cases. The first for $A = B$, $C = -D$ is (Ref. 2)

$$G_1 = \dot{x}\dot{y} + Axy + Dx^3/3 + Dxy^2, \tag{4}$$

where we do not rescale the coefficients here as is frequently done.[2,6] The second invariant for $C = -6D$ is (Ref. 4)

$$G_2 = \dot{x}(\dot{y}x - \dot{x}y) + \frac{(4A-B)x^2}{4D} + \frac{Dx^4}{4} + Dx^2y^2 + Ax^2y + \frac{A(4A-B)x^2}{4D}. \tag{5}$$

Finding the first invariant by a canonical transformation was fairly straightforward, but finding the second invariant required ingenuity. A more general procedure for finding the invariants is the Painlevé method[4,5,7,8] that searches for solutions of the differential equations for which the only movable singularities are simple poles. This method also indicated an invariant for $C = -16D$ and $B/A = 16$. The actual invariant was found by a variation of the direct method[6] and is

$$G_3 = \frac{\dot{x}^4}{4} + \left(\frac{Ax^2}{2} + Dx^2y\right)\dot{x}^2 - \frac{Dx^3\dot{x}\dot{y}}{3} + \frac{A^2x^4}{4} - \frac{ADx^4y}{3} - \frac{D^2x^6}{18} - \frac{D^2x^4y^2}{3}. \tag{6}$$

The direct method begins with an assumed form of the velocity dependence of the invariant, sets the time derivative of the invariant to zero and then equates the coefficients of each distinct functional form of $\dot{x}$ and $\dot{y}$ to zero. These equations, which are the counterpart of the determining equations in the classic Lie point group method, are then solved for the coefficients which depend on the spatial variables.

II. LIE GROUP OF COUPLED HÉNON–HEILES EQUATIONS BY THE CLASSICAL METHOD

First the invariance of the coupled second-order Hénon–Heiles equations is checked by the "classical method" as described in a number of references.[12–25] We exclude conformal invariance here but it should give the same result. The twice-extended group generator needed here is

$$U'' = \xi(x,y,t)\frac{\partial}{\partial t} + \eta^x(x,y,t)\frac{\partial}{\partial x} + \eta^y(x,y,t)\frac{\partial}{\partial y} + \eta,_{xx}^x\frac{\partial}{\partial \ddot{x}} + \eta,_{yy}^y\frac{\partial}{\partial \ddot{y}}, \tag{7}$$

where formulas for $\eta,_{xx}^x$ and $\eta,_{yy}^y$ are available on page 159 of

 0022-2488/90/071627-05$03.00

Ref. 12. Then, the differential operator acting on Eqs. (1) and (2) together with the requirement that Eqs. (1) and (2) hold simultaneously is the invariance condition. This condition is an identity, consequently, the coefficients of the various combinations of $\dot{x}$ and $\dot{y}$ are set equal to zero. The resulting *linear* partial differential equations for the coordinate functions ξ, η^x, and η^y are the determining equations. For the general case, the group generator was only $U_1 = \partial/\partial t$ and the equations invariant under time t had a single invariant, the energy E. With the linear terms absent, $A = B = 0$, the equations are invariant under the stretching group and the group generator U_1

$$U_2 = t\frac{\partial}{\partial t} - 2x\frac{\partial}{\partial x} - 2y\frac{\partial}{\partial y}, \quad U_1 = \frac{\partial}{\partial t}. \tag{8}$$

The invariance under the stretching group does not lead to a second invariant by Noether's theorem.[15] The theorem offers a prescription for the calculation of an invariant for a dynamical system invariant under a Lie point group but restrictions occur that are not satisfied here. The invariants found from Noether's theorem for Lie point symmetries may not include all the invariants for the differential equations. The question then is how to find any invariants from the stretching group. The classical Lie group method calculates the generalized similarity variables from the characteristic equations found from the first extension of U_2. These are $xt^2, yt^2, \dot{x}t^3$, and $\dot{y}t^3$. The invariants with these variables, in general, will have dependence on time but the known invariants, E, G_1, G_2, and G_3 do not depend on time. These four similarity variables can be combined to form an independent set of three variables, x/y, $\dot{x}/x^{3/2}$, $\dot{y}/y^{3/2}$, which do not depend on time but the known invariants cannot be constructed from these variables alone. Hence, the Vlasov equation for the invariant in these three similarity variables cannot be expected to be solved for the known invariants. The Vlasov equation in the four similarity variables or the reduced set of similarity variables may have other invariants but integration of the characteristic equations of the relevant Vlasov equations has not been successful so far. The characteristic equations found from the first extension of U_2 may be augmented by an auxiliary variable, here t, such that the characteristic equations may be combined with appropriate functions of the similarity variables to give the invariants E and G_1, for example (Leach, private communication). This nonclassical method is a generalization of the one discussed by Ince.[26] However, finding the correct combination of equations in the similarity variables is similar to finding the invariants by taking combinations of the original set of Eqs. (1) and (2). Another tack is tried here.

The Hénon–Heiles equations are rewritten as a set of four first-order nonlinear, ordinary differential equations. The Lie point group symmetries of this set includes the higher symmetries;[15] therefore, we may find all the symmetries for this set whereas the contact and higher symmetries are not found for Eqs. (1) and (2) by finding the Lie point symmetries. In addition, we can and shall write the group generator in a form in which the coordinate function multiplying the t derivative is zero but the other coordinate functions depend on time t (Ref. 15):

$$\dot{x} - u = 0, \tag{9}$$

$$\dot{y} - v = 0, \tag{10}$$

$$\dot{u} + Ax + 2Dxy = 0, \quad f_x = Ax + 2Dxy, \tag{11}$$

$$\dot{v} + By + Dx^2 - Cy^2 = 0, \quad f_y = By + Dx^2 - Cy^2. \tag{12}$$

The group generator is found by the classical method for the invariance under Lie group transformations of Eqs. (9)–(12). However, the calculations are long and entail many interconnected determining equations. As all the known invariants are time independent, we assume that the second invariants are independent of t. Then the group generator is simplified by dropping the dependence on t in the coordinate functions. Consequently, the invariants found from the characteristic equations of this form of the group generator are independent of t in the classical method. We do not investigate the possibility that a more general group generator could retain t-dependent coordinate functions but still produce the invariant E and a second invariant. A once-extended group generator in the four variables x,y,u,v acts on Eqs. (9)–(12). The resulting equations, which are given below, Eqs. (16)–(19), cannot be solved without an additional assumption: The coordinate functions for x and y are assumed to be quadratic functions of the velocities. The group generators are calculated and correspond to the invariants in Eqs. (4) and (5). The calculation contains many simple redundant relations. The possibility of determining the invariant from the group generator seems dim because the characteristic equations are so complicated; more so than for the three invariants of a charged particle in a helical magnetic field. Since the group generator is the same as that in Sec. III, we postpone its discussion until then.

III. LIE GROUP OF HÉNON–HEILES EQUATIONS: VECTOR FIELDS METHOD

The invariance of the Hénon–Heiles equations, which can be viewed as dynamical equations, can be analyzed by considering the commutator of differential operators. One operator is called the Vlasov operator here and the other is the group generator. The invariance condition is the vanishing of the commutator of these two operators or in other terminology the Lie bracket of two vector fields[15,16] is zero. The Vlasov operator is V (Ref. 19)

$$V = u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \dot{u}\frac{\partial}{\partial u} + \dot{v}\frac{\partial}{\partial v}, \tag{13}$$

where for $\dot{u}$ and $\dot{v}$ we substituted Eqs. (11) and (12), respectively. The group generator U is

$$U = U_x\frac{\partial}{\partial x} + U_y\frac{\partial}{\partial y} + U_u\frac{\partial}{\partial u} + U_v\frac{\partial}{\partial v}, \tag{14}$$

where the coordinate functions U_x, U_y, etc., are functions of x,y,u,v. The invariance condition

$$[V,U] = 0 \tag{15}$$

can be rewritten as

$$VU_x = U_u, \tag{16}$$

$$VU_y = U_v, \tag{17}$$

$$VU_u = -AU_x - 2DyU_x - 2DxU_y, \tag{18}$$

$$VU_v = -BU_y - 2DxU_x + 2CyU_y. \tag{19}$$

These equations are identical to the invariance relations found from the classical method if the coordinate functions are renamed U rather than η. The coordinate functions U_x and U_y are found from Eqs. (18) and (19) by eliminating U_u and U_v.

The solution of the Hénon–Heiles equations is equivalent to solving the corresponding Vlasov equation that consists of the Vlasov operator acting on a function of the invariant(s). Then, we surely have

$$VE = \{E,H\} = 0, \tag{20}$$

$$VG = \{G,H\} = 0, \tag{21}$$

where E is the energy, the first invariant, H is the Hamiltonian, and G is the second invariant when it exists. The Poisson bracket is indicated by $\{\ ,\ \}$. The invariance of the Hénon–Heiles equations under an infinitesimal Lie group is reflected in the invariance of the invariants under the Lie group generator:

$$UE = 0, \tag{22}$$

$$UG = 0. \tag{23}$$

From the Lie bracket (15) and the above relations, we see that the two vector fields are on an equal footing; that leads to

$$UG = \{G,\mathcal{G}\} = 0 \tag{24}$$

where $\mathcal{G}$ is the equivalent of the Hamiltonian H for the group generator. Then, by analogy with the relations for canonical variables in Hamiltonian dynamics, we find

$$U_x = \frac{\partial G}{\partial u}, \tag{25}$$

$$U_y = \frac{\partial G}{\partial v}, \tag{26}$$

$$U_u = -\frac{\partial G}{\partial x}, \tag{27}$$

$$U_v = -\frac{\partial G}{\partial y}, \tag{28}$$

where $G = \mathcal{G}$ is used here just as $H = E$ for this dynamical system which conserves energy. The four partial differential equations are easily integrated although Noether's theorem for higher symmetries could presumably be used.

The relations between the coordinate functions of the group generator and the second invariant simplify the determination of the second invariant once the group generator is known. On the other hand, one can pick the velocity dependence of the second invariant G; that choice restricts the velocity dependence of the group generator coordinate functions more than the initial assumption made in the earlier sections. In the vector fields method, the velocity dependence of the second invariant is first chosen to be quadratic (as well as linear) in the velocities. Then, from Eqs. (25) and (26), we see that U_x and U_y are linear in the velocities whereas the U_u and U_v may be quadratic in the velocities. The solution of the coordinate functions proceeds by first applying the integrability conditions on Eqs. (25)–(29) that are of the type

$$\frac{\partial U_x}{\partial x} = -\frac{\partial U_u}{\partial u} = \frac{\partial^2 G}{\partial x\, \partial u}. \tag{29}$$

This gives rise to a number of relations among the coefficients of each combination of the velocities. Once the coordinate functions are simplified, they are substituted into Eqs. (16) and (17) and the determining equations for these coefficients are found. The set of determining equations is redundant. The group generator is for a five-parameter group until the condition imposed by Eqs. (18) and (19) reduces the group to a two-parameter group. From the two-parameter group we can find two group generators. One group generator is the Vlasov operator itself which we exclude; the other is the group generator U. The group generators for the two cases are:

$$U_1 = v\frac{\partial}{\partial x} + u\frac{\partial}{\partial y} - (Ay + Dx^2 + Dy^2)\frac{\partial}{\partial u} - (Ax + 2Dxy)\frac{\partial}{\partial v}, \tag{30}$$

where $A = B$, $C = -D$, and

$$U_2 = \left(vx - 2uy + \frac{(4A-B)u}{2D}\right)\frac{\partial}{\partial x} + ux\frac{\partial}{\partial y} - \left(uv + 2Axy + 2Dxy^2 + Dx^3 + \frac{(4A-B)Ax}{2D}\right)\frac{\partial}{\partial u} + (u^2 - Ax^2 - 2Dx^2y) \times \frac{\partial}{\partial v}, \tag{31}$$

where $C = -6D$. One can verify that U_1G_1 and $U_2G_2 = 0$ by substituting invariants from Eqs. (4) and (5) and the group generators in the invariance relation (23).

The above approach to calculating the group generator is much more structured and consequently less prone to error than the classical method employed in Sec. II; however, the actual calculations are not much shorter and there are still redundant equations. To reduce the redundancy of the determining equations and the number of extra coefficients, we next try the direct method.

IV. DIRECT METHOD FOR DETERMINATION OF THE SECOND INVARIANT

By the direct method we calculate the second invariants quadratic in the velocities and also report attempted calculations of restricted forms of the second invariant quartic and sextic in the velocities. Once the second invariant is known, we find the group generator from Eqs. (25)–(29).

In the direct method the dependence of the invariant on velocities must be guessed. We assume a polynomial dependence on velocities, a natural first guess as the potential is a polynomial. Of course, we also know three invariants of this form have been discovered. Next, we use a property of polynomial invariants or constants of the motion reported by Thompson.[9] He showed that the polynomial invariants are either of even or odd degree in the velocities (momenta in his work). As a result, for invariants quadratic in the velocities we exclude terms linear in the velocities. If we had done that in Sec. III, the constraints imposed by Eqs. (25)–(29)

would have left the group generator unaltered and the calculation would have been simplified. The group generator is assumed to be

$$G=g(x,y)u^2-h(x,y)uv+i(x,y)v^2+j(x,y). \tag{32}$$

The direct method requires that

$$\frac{dG}{dt}=0, \tag{33}$$

where the coefficients of each combination of the velocities are set equal to zero separately. For our case that results in six partial differential equations, four of which, Eqs. (34)–(37), are trivially solved. The remaining two equations are solved by substitution of g, h, and i:

$$\frac{\partial g}{\partial x}=0, \tag{34}$$

$$\frac{\partial g}{\partial y}-\frac{\partial h}{\partial x}=0, \tag{35}$$

$$\frac{\partial h}{\partial y}-\frac{\partial i}{\partial x}=0, \tag{36}$$

$$\frac{\partial i}{\partial y}=0, \tag{37}$$

$$\frac{\partial j}{\partial x}+hf_y-2gf_x=0, \tag{38}$$

$$\frac{\partial j}{\partial y}+hf_x-2if_y=0. \tag{39}$$

The invariants found are the same as those found in Sec. II, G_1 in Eq. (4) and G_2 in Eq. (5). The calculations are much shorter than those done in Sec. II or III. The calculations in Sec. III could be shortened with the restriction of the second invariants to, say, even functions of velocities.

The last calculations are of possible invariants quartic and sextic in the velocities. The invariants are assumed to be of a restricted form and a more general form cannot be ruled out. The form is

$$G_3'=u^4/4+a_1u^2v^2+g(x,y)u^2-h(x,y)uv+i(x,y)v^2+j(x,y), \tag{40}$$

where a_1 is a constant and

$$G_4=u^6/6+d(x,y)u^4-e(x,y)u^3v+g(x,y)u^2-h(x,y)uv+j(x,y). \tag{41}$$

Here, G_3' is found to be the result in Eq. (6). From Eqs. (25)–(29) and (40) we find the group generator is

$$U_3=\left(u^3+Ax^2u+2Dx^2yu-\frac{Dx^3v}{3}\right)\frac{\partial}{\partial x}-\frac{Dx^3u}{3}\frac{\partial}{\partial y}+\left(-Axu^2-2Dxyu^2+Dx^2uv-A^2x^3+\frac{4ADx^3y}{3}+\frac{4D^2x^3y^2}{3}+\frac{D^2x^5}{3}\right)\frac{\partial}{\partial u}+\left(-Dx^2u^2+\frac{ADx^4}{3}+\frac{2D^2x^4y}{3}\right)\frac{\partial}{\partial v}. \tag{42}$$

No invariant was found for G_4 in Eq. (41).

V. DISCUSSION

The integrability and Lie point symmetries of the generalized Hénon–Heiles equations have been investigated by three approaches: the classical Lie group point transformation method, the vector fields method, and the direct method. The group generators have been calculated by all three methods for two different second invariants which are quadratic in velocity. For these equations the classical method is the most complicated and is long. A more structured method has been presented using vector fields but the shortest method has been shown to be the direct method. The Lie group generators are given for three separate values of the constants: (1) $A=B$, $C=-D$, (2) $C=-6D$, (3) $B=16A$, $C=-16D$. Although the three invariants are not new, the group generators are new and the methods in this form have not been applied to the Hénon–Heiles equations before.[6]

The approaches discussed here are suggested as alternatives to the Painlevé procedure for determining when a set of nonlinear, autonomous ordinary differential equations is integrable. The Painlevé procedure does not necessarily find all the integrable cases; in fact, a weaker criterion has been used in certain cases.[5] The Lie group method as presented here also does not necessarily find all integrable cases either because the velocity dependence of the invariant or of the group coordinate functions must be postulated. The possibility is left open that there are additional special cases for which invariants may exist.

ACKNOWLEDGMENT

Peter Leach is thanked for helpful discussions and for pointing out the Thompson reference.

[1] M. Hénon and C. Heiles, Astrophys. **69**, 73 (1964).
[2] Y. Aizawa and N. Saito, J. Phys. Soc. Jpn. **32**, 1636 (1972).
[3] P. Leach, J. Math. Phys. **22**, 679 (1981).
[4] Y. F. Chang, M. Tabor, and J. Weiss, J. Math. Phys. **23**, 531 (1982).
[5] A. Ramani, B. Dorizzi, and B. G. Grammaticos, Phys. Rev. Lett. **4**, 1539 (1982).
[6] L. S. Hall, Physica **8D**, 90 (1983).
[7] J. Weiss, Phys. Lett. A **102**, 329 (1984).
[8] M. Tabor, Nature **310**, 277 (1984).

[9] G. Thompson, J. Math. Phys. **25**, 3474 (1984).
[10] I. Marshall and S. Wojciechowski, J. Math. Phys. **29**, 1338 (1988).
[11] E. T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., London, 1937), 4th ed.
[12] A. Cohen, *An Introduction to the Lie Theory of One-Parameter Groups with Applications to the Solution of Differential Equations* (Heath, New York, 1911).
[13] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations*, Appl. Math. Sci. Vol. 13 (Springer-Verlag, New York, 1974).
[14] J. M. Hill, *Solution of Differential Equations by Means of One-Parameter Groups* (Pitman, Boston, 1982).
[15] P. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, New York, 1986).
[16] S. I. Sayegh and G. L. Jones, J. Phys. A: Math. Gen. **19**, 1793 (1986).
[17] L. V. Ovsiannikov, *Group Analysis of Differential Equations*, edited by W. F. Ames (Academic, New York, 1982) (translation).
[18] B. Abraham-Shrauner, J. Math. Phys. **26**, 1428 (1985).
[19] B. Abraham-Shrauner, J. Plasma Physics **32**, 197 (1984).
[20] B. Abraham-Shrauner, IEEE Trans. Plasma Sci. **PS-14**, 35 (1986).
[21] B. Abraham-Shrauner, Plasma Phys. **29**, 263 (1987).
[22] B. Abraham-Shrauner, in *Local and Global Methods in Dynamics*, Lec. Notes in Phys. Vol. 252 (Springer-Verlag, New York, 1985), p. 121.
[23] D. Roberts, J. Plasma Phys. **33**, 219 (1985).
[24] D. Roberts, J. Math. Phys. **26**, 1529 (1985).
[25] D. A. Roberts and B. Abraham-Shrauner, J. Plasma Phys. **38**, 335 (1987).
[26] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Sec. 2.701, p. 46.

Exponentially evolving densities

John M. Verosky
Mathematics Department, Western Washington University, Bellingham, Washington 98225

(Received 30 October 1989; accepted for publication 28 February 1990)

The notion of a conserved density is generalized to those densities whose integrals evolve exponentially. In the formal differential algebra used to find them, only a slight modification of the conserved density case is needed. Various examples for quasilinear evolution equations are given and an application to finding higher-order conserved densities by taking the Poisson bracket of exponentially evolving densities is indicated.

I. INTRODUCTION

Consider an evolutionary partial differential equation (pde) for a dependent variable u and independent variables t and x. Both u and x may be vectors but t is time and the equation takes the form

$$u_t = K,$$

where K is a function of u and its x derivatives but not of t or any t derivatives. A conserved density T is a function of t, x, u, and derivatives of u such that there is another similar function X satisfying

$$D_t T = D_x X,$$

where D_t and D_x are total t and x derivatives, respectively. It is assumed that boundary conditions for u are chosen so that the integral of T over space is constant in time:

$$D_t \int T\,dx = \int D_t T\,dx = \int D_x X\,dx = 0.$$

See Olver's book[1] for methods of finding conserved densities, and their properties, applications, and history. Conservation laws, as they are also known, are one of the major themes of mathematical physics and a thorough study of them is carried out in that book.

In this paper, this notion is generalized to the following:

$$D_t T = aT + D_x X$$

where a is a number, real or complex. We will call T an exponentially evolving density (EED) because the integral of T evolves exponentially with exponent a as the following shows:

$$D_t \int T\,dx = \int D_t T\,dx = \int (aT + D_x X)dx = a \int T\,dx.$$

That is, if we have

$$I(t) = \int T\,dx \text{ at time } t,$$

then we have

$$I(t) = I(0)e^{at}.$$

When a is a negative real number, then $I(t)$ tends to zero, and if a is pure imaginary, then $|I(t)|^2$ is conserved. Both kinds of quantities are useful.

II. RICH POISSON STRUCTURES

If the evolution system is a Hamiltonian system, then it has a Poisson bracket and the total t derivative of a density T, which does not depend on t explicitly, is given in terms of the Hamiltonian density H of the system and the bracket by

$$D_t T = \{T,H\}.$$

Again, see Olver's book for a thorough explanation of Hamiltonian pdes and their brackets. For an EED, we thus have

$$\{T,H\} = aT.$$

Suppose S is another EED with exponent b. Then the Jacobi identity says

$$\{\{T,S\},H\} + \{\{H,T\},S\} + \{\{S,H\},T\} = 0,$$

which reduces to

$$\{\{T,S\},H\} = (a+b)\{T,S\},$$

which means that $\{T,S\}$ is a new EED with exponent $a+b$. Hence, these densities form a closed subalgebra of the Poisson algebra of all densities. Note that if $b = -a$, then the new density is actually conserved so that new conserved densities can be produced in this manner. This is somewhat similar in spirit to Rosencrans' method of producing a conserved density out of two non-Hamiltonian symmetries.[2] Once a system has both a Hamiltonian structure and EEDs, one can expect a very rich Poisson algebra.

III. A SIMPLE EXAMPLE

First, we study the equation $u_t = f(u)u_x$. Let us start by seeking the most general EED of the form $T(u,u_x)$. Hence, we must expand

$$D_t T(u,u_x) = aT(u,u_x) + D_x X(u,u_x)$$

using the chain rule, and since the coefficient of u_{xx} must be zero, it follows that

$$f\frac{\partial T}{\partial u_x} = \frac{\partial X}{\partial u_x}.$$

Since f does not depend on u_x, we have $X = f(u)T + g(u)$, which is substituted into the rest of the expression. Finally, we get

$$\frac{\partial T}{\partial u_x} = g'u_x + \frac{(a + f'u_x)T}{f'u_x^{\,2}},$$

which is easily integrated for $g = 0$ to get what will be the prototype of all examples considered in this paper:

$$T = h(u)u_x \exp(-a/f'(u)u_x),$$

where h is an arbitrary function of u. Note that if $a = 0$, then T is simply $h(u)u_x$, which is a trivial conserved density since it is already an x derivative.

 0022-2488/90/071632-03$03.00

For $f(u) = u$ and h identically one, the density is

$$T = u_x \exp(-a/u_x)$$

and the equation $u_t = uu_x$ has a Hamiltonian structure given by

$$u_t = D_x E(u^3/6),$$

where E is the Euler operator or variational derivative $\partial/\partial u - D_x(\partial/\partial u_x) + \cdots$. The Poisson bracket is given by

$$\{A,B\} = (EA) D_x (EB).$$

A long calculation shows that the bracket of two of the above densities with exponents a and b, respectively, is

$$a^2 b^2 ((a-b) u_{xx}^3 / 2u_x^8) \exp(-(a+b)/u_x).$$

The important thing to note is that this new density is of order 2 although it was constructed out of two first-order densities, and so bracketing of EEDs can lead to higher-order ones, for which a direct search becomes prohibitively long and tedious algebraically as order increases. When $b = -a$, we get a higher-order conserved density

$$u_{xx}^3 / u_x^8,$$

and it is in fact easy to check that

$$D_t (u_{xx}^3 / u_x^8) = D_x (u u_{xx}^3 / u_x^8).$$

IV. EXAMPLE OF A QUASILINEAR DIAGONAL SYSTEM OF TWO EQUATIONS

Such a system is of the form

$$u_t = f u_x, \quad v_t = g v_x,$$

where f and g are functions of u and v. Its higher-order symmetries and conserved densities were first considered by Verosky.[3] A quick calculation shows that the density

$$T = h(u) u_x \exp\left(-a \Big/ \frac{\partial f}{\partial u} u_x\right)$$

(which is exactly the same as the one in the last section with f' replaced by $\partial f/\partial u$) is exponentially evolving with exponent a and arbitrary h if g can be expressed in terms of f as

$$g = (f f_{uv} - f_u f_v)/f_{uv}.$$

Let us call this expression $N(f)$. This calculation can be generalized to a diagonal system of three equations

$$u_t = f u_x, \quad v_t = g v_x, \quad w_t = h w_x,$$

where f, g, and h are functions of u, v, and w. The EED will have exactly the same form as above when both of the following hold:

$$g = (f f_{uv} - f_u f_v)/f_{uv}, \quad h = (f f_{uw} - f_u f_w)/f_{uw}.$$

The generalization to n variables is obvious.

Suppose that both $g = N(f)$ and $f = N(g)$ hold. There is then a symmetry in u and v and we get the following EED:

$$T = h(u) u_x \exp\left(-a \Big/ \frac{\partial f}{\partial u} u_x\right) + k(v) v_x \exp\left(-a \Big/ \frac{\partial g}{\partial v} v_x\right).$$

Expanding $f = N(N(f))$ and performing the derivatives shows what sort of f are possible. They are the solutions of the pde:

$$f_{uv} f_{uuvv} = f_{uvv} f_{vuu},$$

which is easy to solve. Its general solution is

$$f = p(u) q(v) + r(u) + s(v),$$

where p, q, r, and s are arbitrary twice differentiable functions of one variable. The corresponding g is

$$g = -r(u)s(v)/p(u)q(v) - p(u)r(u)/p'(u) + r(u) - q(v)s(v)/q'(v) + s(v),$$

which has the same general form of a product plus a sum of functions of u and v like f. It is not at all obvious that $N(g) = f$ but, on functions having the form that f does, $N^2 = 1$. Clearly, lots of examples can now be produced just by choosing p, q, r, and s, but finding a Hamiltonian structure for the resulting system is not easy. There is an example[4] of a diagonal quasilinear system that has no Hamiltonian structure but such systems are thought to be a singular case by a simple variable counting argument in the equations that need to be solved. A good question is to find all choices of p, q, r, and s that result in a Hamiltonian system. The forms of f and g are sufficiently limited to make this reasonable. In the next section, we give one example where there is a Hamiltonian structure.

V. A SYSTEM WITH BOTH HAMILTONIAN STRUCTURE AND EEDs

Consider a system describing two waves, one moving left and the other right, both with speeds given by the square of the total amplitude. This nonlinear wave equation looks like

$$u_t = (u+v)^2 u_x, \quad v_t = -(u+v)^2 v_x.$$

Both of the relations $g = N(f)$ and $f = N(g)$ hold as can be easily verified. The Hamiltonian structure is found by a change in variables

$$u = \frac{1}{2}\left(-\frac{1}{r} + q\right), \quad v = \frac{1}{2}\left(-\frac{1}{r} - q\right).$$

The system becomes

$$r_t = q_x, \quad q_t = (-1/3r^3)_x,$$

which has a Hamiltonian structure

$$\begin{pmatrix} r \\ q \end{pmatrix}_t = \begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix} \begin{pmatrix} \dfrac{\partial}{\partial r} \\ \dfrac{\partial}{\partial q} \end{pmatrix} \left(\frac{1}{6r^2} + \frac{q^2}{2}\right),$$

which transforms to one in the uv variables with a nonlinear Hamiltonian operator A:

$$\begin{pmatrix} u \\ v \end{pmatrix}_t = A \begin{pmatrix} \dfrac{\partial}{\partial u} \\ \dfrac{\partial}{\partial v} \end{pmatrix} \frac{(2u^2 - 2uv - 2v^2)}{3}.$$

If the Jacobian of the change of variables qr to uv is expressed in terms of u and v, then we get

$$J=\begin{pmatrix} u_r & u_q \\ v_r & v_q \end{pmatrix}=\begin{pmatrix} (u+v)^2 & 1 \\ (u+v)^2 & -1 \end{pmatrix},$$

and the Hamiltonian operator is

$$A=J\begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix}J^T,$$

where T denotes transpose. This change of variables is the one that Dubrovin and Novikov[5] proved exists when one has a nonlinear Hamiltonian operator A. Ironically, we went backwards here because it is easier to find a change of coordinates where the structure has a linear Hamiltonian operator rather than to look for the nonlinear Hamiltonian structure directly.

The Poisson bracket of two densities S and T is

$$\{S,T\}=(E_uS,E_vS)A\begin{pmatrix} E_uT \\ E_vT \end{pmatrix},$$

where E_u and E_v are the Euler operators or variational derivatives with respect to u and v and may be written as

$$E_u=\frac{\partial}{\partial u}-D_x\frac{\partial}{\partial u_x},\quad E_v=\frac{\partial}{\partial v}-D_x\frac{\partial}{\partial v_x}$$

for the first-order densities under consideration. We forego an explicit calculation because of its length. If the analogy with the example in Sec. III holds, a second-order EED will be the bracket of two first-order ones. An extremely rich Poisson structure can be expected and, in theory, bracketing the EEDs of opposite exponentials can be used to construct higher-order conserved densities in a way alternate to Olver and Nutku's[6] use of recursion operators for separable Hamiltonian systems (of which this is an example since in the coordinates which the Hamiltonian operator is linear the Hamiltonian function satisfies $H_{rr}/H_{qq}=1/r^4$ which is separable). An interesting question is: Which of the diagonal systems with $f=N(g)$, $g=N(f)$ and which have Hamiltonian structures are separable Hamiltonian systems?

ACKNOWLEDGMENTS

Part of this work was done at Heriot–Watt University in Edinburgh, Scotland. Many thanks for their wonderful hospitality and to SERC for financial support.

[1] P. J. Olver, "Applications of Lie groups to differential equations," in *Graduate Texts in Mathematics* (Springer-Verlag, New York, 1986), Vol. 107.

[2] S. I. Rosencrans, "Conservation laws generated by pairs of non-Hamiltonian symmetries," J. Diff. Eq. **43**, 305 (1982).

[3] J. M. Verosky, "Higher-order symmetries for the equations of isentropic gas dynamics," J. Math. Phys. **25**, 884 (1984).

[4] J. M. Verosky, "*Tu* commutativity for quasilinear evolution equations," Phys. Lett. A **119**, 25 (1986).

[5] B. A. Dubrovin and S. A. Novikov, "Hamiltonian formalism of one dimensional systems of hydrodynamic type and the Bogolykov–Whitham averaging method," Sov. Math Dokl. **27**, 665 (1983).

[6] P. J. Olver and Y. Nutku, "Hamiltonian structures for systems of hyperbolic conservation laws," J. Math. Phys. **29**, 1610 (1988).

Equivariant harmonic maps into homogeneous spaces

Caio J. C. Negreiros
Departamento de Matemática, IMECC–UNICAMP, C.P. 6101, 13081 Campinas, SP, Brazil

(Received 16 August 1989; accepted for publication 21 February 1990)

This paper is about harmonic maps from closed Riemann surfaces into homogeneous spaces such as flag manifolds and loop groups. It contains the construction of a family of new examples of harmonic maps from $T^2 = S^1 \times S^1$ into $F(n)$ or $\Omega(U(n))$ that are not holomorphic with respect to any almost complex structure on $F(n)$ or $\Omega(U(n))$, where $F(n)$ is the quotient of $U(n)$ by any maximal torus and $\Omega(\mathrm{u}(n))$ consists of $f: S^1 \to U(n)$ smooth such that $f(1) = I$.

I. INTRODUCTION

In this paper we study harmonic maps that are equivariant with respect to an S^1 action from $T^2 = S^1 \times S^1$ into the full flag manifold $F(n) = U(n)/T$, where T is any maximal torus of $U(n)$ and $F(n)$ is equippped with a large class of left-invariant metrics that includes the Kähler ones and the Killing form metric. In Ref. 1 we studied only $F(n)$ equipped with the Killing form metric. We also discuss the extension of these results to the Loop group $\Omega(U(n))$.

Often interesting examples of solutions to nonlinear problems are found by examining an equivariant case. The assumption of equivariance under a continuous group action whose orbits have codimension one in the domain manifold reduces a partial differential equation to an ordinary differential equation, and by using the theorem of existence and uniqueness of solutions of ordinary differential equations, we can produce lots of solutions of our problem.

We also know that the critical points of a functional on a space of maps are difficult to treat in general. The energy functionals whose critical points are the harmonic maps are easier to analyze from the point of view of computations than, for example, the Yang–Mills functional, but they share some of their important properties like conformed invariance in the domain manifold, bubbling-of phenomena, etc. This paper was inspired somehow by the fact that we can obtain, in a more or less standard way, Yang–Mills connections of $S^2 \times S^2$ that are not instantons. See Ref. 2 for possible connections with the subject of that paper.

In the case of S^4, according to results of Atiyah[3] and Donaldson,[4] we have a natural 1–1 correspondence between instanton connections on S^4 and holomorphic S^2 into Loop groups. Recently, L. M. Sibner, R. J. Sibner, and K. Uhlenbeck[5] announced the existence of SU(2) Yang–Mills connections on S^4 that are not instantons.

Now if one wants to understand the problem of harmonic maps into nonsymmetric spaces like the Loop group with the (symplectic) Kähler structure, it is natural to start this study with harmonic maps into full flag manifolds, since such manifolds model the geometry of the Loop group in finite dimensions. See Refs. 6 or 7 for more details.

In Sec. II we state some basic facts about maps into flag manifolds and describe a precise set of left-invariant metrics in such manifolds with which this paper is concerned.

In Sec. III we recall the expressions for the harmonic and holomorphic maps equations in terms of projection operators as in Refs. 1 or 8 and derive topological restrictions for a totally isotropic map $\phi: T^2 \to F(n)$ to be holomorphic with respect to a nonintegrable almost complex structure on $F(n)$.

In Sec. IV we construct a series of new examples of harmonic maps $\phi: T^2 \to F(n)$ that are not holomorphic with respect to any almost complex structure on $F(n)$, where $F(n)$ is equipped with a large class of left-invariant metrics. Then using Ref. 9 we see how these maps generate two-tori into $\Omega(U(n))$.

The content of this paper was originated during the period of my doctoral thesis.[10]

II. SOME BASIC FACTS ABOUT MAPS INTO FLAG MANIFOLDS

A flag manifold is a homogeneous space G/T, where G is a compact Lie group and T is any maximal torus. We denote by $F(n)$ the flag manifold with $G = U(n)$ and

$$T = \underbrace{U(1) \times \cdots \times U(1)}_{n \text{ times}}.$$

The Killing form of $U(n)$ is a positive-definite inner product $\langle , \rangle$ on the Lie algebra $u(n)$, and one has the decomposition

$$u(n) = p \oplus \underbrace{u(1) \oplus \cdots \oplus u(1)}_{n \text{ times}}.$$

If $\mathscr{G}$ is a Lie algebra over **R** and p is a subspace of $\mathscr{G}$, p is called a Lie triple system if given $X, Y, Z \in p$ then $[X,[Y,Z]] \in p$. Let us recall the following result due to E. Cartan.

Theorem 2.1: Let G be a Lie group and H be a closed subgroup of G, but let $M = G/H$ be a symmetric space. Let $\mathscr{G} = p \oplus h$, where $\mathscr{G}$ is the Lie aglebra of G and h is the Lie algebra of H. Let s be a Lie triple system contained in p. Put $S = \exp(s)$. Then S has a natural differentiable struture in which it is a totally geodesic submanifold of M satisfying $S_{p_0} = s$. On the other hand, if S is a totally geodesic submanifold of M and $p_0 \in S$, then the subspace $s = S_{p_0}$ of $\mathscr{G}$ is a Lie triple system.

Proof: See Ref. 11.

As a consequence of Theorem 2.1 we see that $F(n)$ cannot be a symmetric space.

We have $p = \Sigma_{s \in S} E_s$, where $S \subseteq N^*$ is the set of roots and E_s is the root-space corresponding to $s \in S$. We have

 0022-2488/90/071635-08$03.00

$$p \underset{\mathbb{R}}{\otimes} \mathbb{C} = \sum_{s \in S'} E_s,$$

where S' is the subset of complementary roots.

A T-invariant almost complex structure on $F(n)$ corresponds to a T-invariant endomorphism J of p with $J^2 = -I$. Such endomorphisms correspond to some decomposition of $S = S^+ \oplus S^-$, where

$$S^{-1} = \{ -\alpha; \alpha \in S^+ \}$$

and

$$p \underset{\mathbb{R}}{\otimes} \mathbb{C} = p_{(1,0)} \oplus p_{(0,1)} = \Big(\sum_{s \in S^+} E_s \Big) \oplus \Big(\sum_{s \in S^-} E_s \Big).$$

The almost complex structure is integrable precisely when S^+ is the set of positive roots with respect to a choice of fundamental Weyl chamber D in $u(1) \oplus \cdots \oplus u(1)$.

A general (T-invariant) almost complex structure is specified by whether or not it agrees with J on each $E_s \oplus E_{-s}$, so there are $2^{|S^+|}$ possibilities, but only $n!$ [order of the Weyl group of $U(n)$] are integrable.

Now we define a family of left-invariant metrics on $F(n)$, namely: Let A and B in p and consider the inner product

$$\langle A,B \rangle_{a = (a^{ij})} = \sum_{i,j} \operatorname{tr}(a^{ij} E_i A E_j B^*),$$

where

$$E_i = i \begin{array}{c} \begin{array}{cccccc} & & & i & & \end{array} \\ \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ & 0 & & & & \\ & & \ddots & & & \\ & & & 1 & & 0 \\ & & & & 0 & \\ & & & & & \ddots \\ 0 & & & & & 0 \end{pmatrix} \end{array},$$

$$a = (a^{ij}), \quad a^{ij} = a^{ji} > 0.$$

If we restrict our almost complex structures in $F(n)$ to the integrable ones, we can see that

$$g_{(a_1,\ldots,a_n,a_1 + a_2,\ldots,a_{n-2} + a_{n-1},\ldots,a_1 + \cdots + a_{n-1})}$$

give all left-invariant Kähler metrics on $F(n)$. See Refs. 12, 10, 13, or 14 for more details. It is worthwhile to point out that if we consider $F(n)$ with the normal metric induced from the natural bi-invariant metric on $U(n)$, it is not a Kähler manifold.

Let $\underline{\mathbb{C}}^n$ denotes the trivial holomorphic vector bundle $M^2 \times \mathbb{C}^n$ over M^2.

We use extrinsic differential geometry and think of ϕ: $M^2 \to F(n)$ as a map or a subbundle of $\mathbb{C}^n$ via the pullback of tautological defined vector bundles on $F(n)$. Note that we also think of $F(n)$ as the set of n-tuples $(L_1,\ldots,L_n)$. Here L_i is a one-dimensional subbundle of $\underline{\mathbb{C}}^n$, L_i is perpendicular of L_j if $i \neq j$, and $L_1 \oplus \cdots \oplus L_n = \underline{\mathbb{C}}^n$. Then the tautologously defined vector bundles on $F(n)$ have as fibers over a flag $(L_1,\ldots,L_n)$ the vector spaces $L_1,\ldots,L_n$, respectively.

As usual, we identify a smooth map $\phi: M^2 \to \mathbb{C}P^{n-1}$ with a subbundle ϕ of $\mathbb{C}^n$ of rank one that has fiber at $x \in M$ given by $\phi_x = T_{\phi(x)}$ where T is the tautological line bundle over $\mathbb{C}P^{n-1}$; i.e., $\underline{\phi} = \phi^*(T)$. Any subbundle $\underline{\phi}$ of $\underline{\mathbb{C}}^n$ inherits a metric denoted by $\langle , \rangle_\phi$ and a connection denoted by D_ϕ, from the flat metric and connection ∂ on $\underline{\mathbb{C}}^n$.

Explicitly,

$$\langle V,W \rangle_\phi = \langle V,W \rangle, \quad \forall V,W \in \phi_x, \quad x \in M,$$

and

$$(D_\phi)_Z W = \Pi_\phi(\partial_Z W), \quad W \in \Gamma(\phi), \quad Z \in T(M)^{1,10}.$$

Here $\Pi_{\underline{\phi}}: \underline{\mathbb{C}}^n \to \underline{\phi}$ denotes the Hermitian projection in the subbundle $\underline{\phi}$.

Note that we always describe $F(n)$ in terms of the natural embedding $F(n) \hookrightarrow \mathbb{C}P^{n-1} \times \cdots \times \mathbb{C}P^{n-1}$. So ϕ: $M^2 \to F(n)$ is described as $\phi = (\Pi_1,\ldots,\Pi_n)$ where Π_i: $M^2 \to \mathbb{C}P^{n-1}$ and $\Pi_i \Pi_j = \delta_{ij} \Pi_i$, $\Pi_i^* = \Pi_i$.

Now let $g: M^2 \to U(n) \cdot g$; this can be thought of as $g = (X_1,\ldots,X_n)$, where X_i is a matrix with n rows and one column and $gg^* = I$. We introduce the orthogonal projections $\Pi_1(g) = X_1 X_1^*,\ldots,\Pi_n(g) = X_n X_n^*$. Hence we must have $\Pi_1 + \cdots + \Pi_{(n)} = I$ and $\Pi_i \Pi_j = 0$ if $i \neq j$ and $\Pi_i^2 = \Pi_i$ since $g^* g = I$ and $X_i^* X_j = \delta_{ij}$.

Alternatively, we can think of Π_i as $\Pi_i(g) = g E_i g^*$, where E_i is already defined. Gauge transformations act on g as

$$(X_1,\ldots,X_n) \to (X_1 h_1,\ldots,X_n h_n),$$

where

$$\begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix} \in U(n).$$

The Gauge potential is

$$A_u = \begin{pmatrix} X_1^* \partial_u (X_1) & & \\ & \ddots & \\ & & X_n^* \partial_n (X_n) \end{pmatrix}$$

and the covariant derivative is

$$D_u(X_1,\ldots,X_n) = (\Pi_1(\partial_u(X_1)),\ldots,\Pi_n(\partial_u(X_n))),$$

where $u = \partial / \partial Z$ or $\partial / \partial \bar{Z}$. By composing g with $\Pi: U(n) \to U(n)/T = F(n)$, we can think of $\phi = \Pi \circ g$: $M^2 \to F(n)$ as $\phi = (\Pi_1,\ldots,\Pi_n)$. Then each such ϕ determines tautologously defined vector bundles $\tilde{\Pi}_1,\ldots,\tilde{\Pi}_n$ over M^2. Let $\partial \Pi_i / \partial x = \partial_\partial \Pi_i / \partial x$ be the covariant derivative of Π_i with respect to x. We call the partial second fundamental forms of ϕ the maps

$$A_x^{ij} = \Pi_i(A_x^j) = \Pi_i \frac{\partial \Pi_j}{\partial x} \quad \text{if } i \neq j.$$

Note that $A_x^{ij} \in \operatorname{Hom}(\Pi_j, \Pi_i)$ and $\Sigma_j A_x^{ij}$ is the second fundamental of the span of Π_i.

Now if we think of M^2 as a complex one-dimensional manifold, then we define

$$\frac{\partial \Pi_i}{\partial Z} = \frac{1}{2} \Big(\frac{\partial \Pi_i}{\partial x} - \sqrt{-1} \, \frac{\partial \Pi_i}{\partial y} \Big)$$

and

$$\frac{\partial \Pi_i}{\partial \bar{Z}} = \frac{1}{2}\left(\frac{\partial \Pi_i}{\partial x} + \sqrt{-1}\,\frac{\partial \Pi_i}{\partial y}\right).$$

We also define

$$A^{i}_{Z} = \sum_{i(\neq j)} A^{ij}_{\bar{Z}},\ A^{i}_{\bar{Z}} = \sum_{i(\neq j)} A^{ij}_{Z},$$

where

$$A^{ij}_{\bar{Z}} = \Pi_i \frac{\partial \Pi_j}{\partial Z} \quad \text{and} \quad A^{ij}_{Z} = \Pi_i \frac{\partial \Pi_j}{\partial Z}.$$

III. HARMONIC AND HOLOMORPHIC MAPS INTO FLAG MANIFOLDS

We now study the energy integral in terms of projection operators and write down the Euler–Lagrange equations for our variational problems.

Definition 3.1: Given a smooth map $\phi = (\Pi_1,...,\Pi_n)$: $M^2 \to (F(n) = U(n)/_T, g_{a=(a^{ij})})$, where $\Pi_i = \phi E_i \phi^*$, we define the energy of ϕ as

$$E(\phi) = \frac{1}{2}\sum_{i=1}^{n}\int_{M^2}\left(\left\langle\frac{\partial \Pi_i}{\partial Z},\frac{\partial \Pi_i}{\partial Z}\right\rangle_{g_a} + \left\langle\frac{\partial \Pi_i}{\partial \bar{Z}},\frac{\partial \Pi_i}{\partial \bar{Z}}\right\rangle\right)V_g$$

$$= \sum_{i,j=1}^{n}\int_{M} a^{ij}\langle A^{ij}_{\mu}, A^{ij}_{\mu}\rangle V_g$$

$$= \sum_{i,j=1}^{n}\int_{M^2} a^{ij}\,\mathrm{tr}(A^{ij}_{\mu}A^{ji}_{\mu})V_g,$$

where $\mu = Z$ or $\bar{Z}$.

Proposition 3.2: Let $\phi = (\Pi_1,...,\Pi_n): M^2 \to (F(n), g_a)$ be any smooth map. Then

$$(\delta E)(\delta\phi(q)) = -2\,\mathrm{Re}\left\{\int_{M^2}\left\langle a_{ij}A^{ij}_{\mu}, \frac{\partial q}{\partial \mu}\right\rangle V_f\right\}$$

$$= -2\,\mathrm{Re}\left\{\left\langle\left\langle \mathscr{A}^{a}_{\mu}, \frac{\partial q}{\partial \mu}\right\rangle\right\rangle\right\},$$

where $\mu = Z$ or $\bar{Z}$, $\langle\langle\ \rangle\rangle$ denotes the L^2-Hilbert inner product, and $\mathscr{A}^{a}_{\mu} = (\mathscr{A}^{ij}_{\mu})$, where $\mathscr{A}^{ij}_{\mu} = a^{ij}A^{ij}$.

Proof: By definition

$$E = \int_{M^2} a^{ij}\left\langle \Pi_i \frac{\partial \Pi_j}{\partial \mu}, \Pi_i \frac{\partial \Pi_j}{\partial \mu}\right\rangle V_g.$$

Then

$$\delta E = 2\,\mathrm{Re}\left\{\int_{M^2}\langle a^{ij}A^{ij}_{\mu}, \delta(A^{ij}_{\mu})\rangle V_g\right.$$

$$= 2\,\mathrm{Re}\left\{\int_{M^2} a^{ij}\left\langle A^{ij}_{\mu}, [A^{ij}_{\mu}, q] - \Pi_i \frac{\partial q}{\partial \mu}\right\rangle V_g\right\}$$

$$= -2\,\mathrm{Re}\left\{\int_{M^2} a^{ij}\langle A^{ji}_{\bar{\mu}}, A^{ij}_{\mu}, q\rangle\right\}$$

$$-2\,\mathrm{Re}\left\{\in_{M^2}\left\langle a^{ij}A^{ij}_{\mu}, \frac{\partial q}{\partial \bar{\mu}}\right\rangle V_g\right\}$$

$$= -2\mathrm{Re}\left\langle\left\langle {}^{a}_{\mu}, \frac{\partial q}{\partial \mu}\right\rangle\right\rangle \quad \text{since } \mathrm{Re}\{a^{ij}\langle [A^{ij}_{\mu}, A^{ji}_{\bar{\mu}}], q\rangle\} = 0.$$

Corollary 3.3: Let $\phi = (\Pi_1,...,\Pi_n): M^2 \to (F(n), g_a)$, where g_a is any left-invariant metric on $F(n)$. Then ϕ is harmonic if and only if

$$\frac{\partial}{\partial x}(\mathscr{A}^{a}_{x}) + \frac{\partial}{\partial y}(\mathscr{A}^{a}_{y}) = 0$$

if and only if

$$\frac{\partial}{\partial Z}(\mathscr{A}^{a}_{\bar{Z}}) + \frac{\partial}{\partial \bar{Z}}(\mathscr{A}^{a}_{Z}) = 0.$$

Proof: By the above and according to the fundamental lemma of the calculus of variations, we have that ϕ is harmonic if and only if

$$\mathrm{Re}\left\{\frac{\partial}{\partial \bar{Z}}(\mathscr{A}^{a}_{z})\right\} = \frac{\partial}{\partial x}(\mathscr{A}^{a}_{x}) + \frac{\partial}{\partial y}(\mathscr{A}^{a}_{y}) = 0.$$

Equivalently, we see that

$$\frac{\partial}{\partial \bar{Z}}(\mathscr{A}^{a}_{Z}) + \frac{\partial}{\partial Z}(\mathscr{A}^{a}_{\bar{Z}}) = 2\left(\frac{\partial}{\partial x}(\mathscr{A}^{a}_{x}) + \frac{\partial}{\partial y}(\mathscr{A}^{a}_{y})\right).$$

Now let $[1,n] = \{x \in \mathbb{Z};\ 1 \leqslant x \leqslant n\}$. Consider $D = \{(i,i);\ 1 \leqslant i \leqslant n\}$ and S^+ to be a partition of $([1,n]\times[1,n] - D)$ containing $(n^2-n)/2$ elements such that if $(i,j)\in S^+$ then $(j,i)\notin S^+$. We denote S^- as the complement of S^+ in $([1,n]\times[1,n] - D)$. We call S^+ a positive system in $[1,n]$.

Let E^0 and $\bar{E}$ denote the ∂ and $\bar{\partial}$ energy, respectively, defined by

$$E^{0}_{S^+}(\phi) = \sum_{(i,j)\in S^+}\int_{M^2}|A^{ij}_{Z}|^2 V_g$$

and

$$\bar{E}_{S^+}(\phi) = \sum_{(i,j)\in S^+}\int_{M^2}|A^{ij}_{\bar{Z}}|^2 V_g.$$

Therefore $\phi = (\Pi_1,...,\Pi_n): M^2 \to F(n)$ is holomorphic with respect to the almost complex structure determined by S^+ if and only if

$$\bar{E}_{S^+}(\phi) = \sum_{(i,j)\in S^+}\int_{M^2}|A^{ij}_{\bar{Z}}|^2 V_g = 0$$

if and only if $A^{ij}_{\bar{Z}} = 0\ \forall (i,j)\in S^+$.

Definition 3.4: Let $\phi = (\Pi_1,...,\Pi_n): M^2 \to (F(n), g_a)$ be a harmonic map. Note that ϕ is called totally isotropic if $[\mathscr{A}_Z, \mathscr{A}_{\bar{Z}}]_p = 0$, where $[\mathscr{A}_Z, \mathscr{A}_{\bar{Z}}]_p$ denotes the off diagonal part of the $n\times n$ matrix $[\mathscr{A}_Z, \mathscr{A}_{\bar{Z}}]$.

Now by using the Koszul–Malgrange theorem, we can prove that if $\phi = (\Pi_1,...,\Pi_n): M^2 \to (F(n), g_a)$ is a totally isotropic map, then $\mathscr{A}^{ij}_{\mu} \in \mathrm{Hom}(\Pi_j, \Pi_i) \cong \Pi_j^* \otimes \Pi_i$ is a holomorphic section of the line bundle $\Pi_j^* \otimes \Pi_i$ over M^2 when the total space of such bundle has a suitable complex structure. See Ref. 14 for the details of such fact.

We now prove a result for harmonic maps $\phi = (\Pi_1,...,\Pi_n): T^2 = S^1\times S^1 \to (F(n), g_a)$ that consists of a purely topological restriction for ϕ to be holomorphic with respect to some nonintegrable almost complex structure on $F(n)$.

Proposition 3.5: Let $\phi = (\Pi_1,...,\Pi_n): T^2 \to (F(n), g_a)$ be a totally isotropic map, holomorphic with respect to some nonintegrable almost complex structure on $F(n)$ but not with respect to any integrable one. Then $c_1[\Pi_1] = \cdots = c_1[\Pi_n] = 0$, where $c_1[\Pi_i]$ denotes the first Chern number of Π_i.

Proof: We give the proof for $n = 3$, and for arbitrary the proof is similar.

Without loss of generality we can assume, say, that $\mathscr{A}_Z^{12} \neq 0$, $\mathscr{A}_Z^{31} \neq 0$, and $\mathscr{A}_Z^{23} \neq 0$, otherwise ϕ would be holomorphic with respect to some integrable almost complex structure.

But $T(T^2) \otimes \Pi_j^* \otimes \Pi_i$ has a holomorphic section if and only if $-c_1[T^2] + c_1[\Pi_1] - c_1[\Pi_j] = c_1[\Pi_i] - c_1[\Pi_j] \geqslant 0$.

Therefore

$$c_1[\Pi_1] \geqslant c_1[\Pi_2],$$

$$c_1[\Pi_2] \geqslant c_1[\Pi_3],$$

and

$$c_1[\Pi_3] \geqslant c_1[\Pi_1],$$

i.e., $c_1[\Pi_1] = c_1[\Pi_2] = c_1[\Pi_3]$. But $\Pi_1 + \Pi_2 + \Pi_3$ is equal to the trivial bundle over T^2, hence $c_1[\Pi_1] + c_1[\Pi_2] + c_1[\Pi_3] = 3c_1[\Pi_1] = 0$. Therefore $c_1[\Pi_1] = c_1[\Pi_2] = c_1[\Pi_3] = 0$.

IV. EQUIVARIANT HARMONIC MAPS INTO FLAG MANIFOLDS

In this paragraph, we will study harmonic maps that are equivariant with respect to an S^1 action on the space of harmonic maps from $S^1 \times \mathbb{R}$ to $F(n)$ in the sense of Palais.[15] Such equivariant harmonic maps will provide new examples of equivariant tori $T^2 = S^1 \times S^1$ into $F(n)$ or more generally equivariance harmonic $T^2 = S^1 \times S^1$ into the loop group $\Omega(U(n))$.

Often interesting examples of solutions to nonlinear problems are found by examining an equivariant case. The assumption of equivariance under a continuous group action whose orbits have codimension one in the domain manifold reduces a partial differential equation to an ordinary differential equation; then we essentially use the theorem of existence and uniqueness of solutions of ordinary differential equations.

Now let us recall some useful facts from the general theory of equivariant harmonic maps.

Let G be a compact, connected group of isometries of M. An immersion f: $N \to M$ is called G-invariant if there exists a smooth action of G on N such that $g \cdot f = f \cdot g$, $\forall g \in G$. The submanifold f is said to be minimal if its mean curvature vector field vanishes identically.

Definition 4.1: By an equivariant variation of a G-invariant submanifold f: $N \to M$, we mean a differentiable variation f_t: $N \to M$, $-\epsilon < t < \epsilon$, $f_0 = f$, through submanifolds such that $g \cdot f_t = f_t \cdot g$ for all $g \in G$ and all t. We recall the following useful result of Hsiang and Lawson.[16]

Theorem 4.2: Let N be a compact manifold and f: $N \to M$ be a G-invariant submanifold of M. Then f:$N \to M$ is minimal if and only if the volume of N is stationary with respect to all compactly supported equivariant variations.

These results have a close relationship with the theory of harmonic maps if one recalls that every minimal surface is harmonic in some conformal structure. Furthermore, if ϕ:$(M,g) \to (N,h)$ is a nonconstant, harmonic and conformal, then it is a minimal branched immersion.

Let G be a group acting on Map(M,N) and let E: Map$(M,N) \to \mathbb{R}$ be a C^1 function invariant under the G action. A symmetric point of Map(M,N) is an element of the set $\Sigma = \{\phi \text{Map}(M,N); g \cdot \phi = \phi \ \forall g \in G\}$ of points fixed under the action of G. The principle of symmetric criticality states that in order for a symmetric point ϕ to be a cricital point, it suffices that it be a critical point of $E|_\Sigma$. Furthermore, Σ is a totally geodesic submanifold of Map(M,N). This principle is very useful in our present case, since the energy is invariant under the circle action because of the cyclic property of the trace. See Ref. 15 for more details.

Now we study the differential equations found in Ref. 17, adapted to our nonsymmetric case.

Consider ρ:$S^1 \to U(n)$ given by $\rho(\exp(\sqrt{-1}\,\theta)) = \exp(A\theta)$, where A is some fixed matrix in $u(n)$ and we also assume $\exp(2\pi A) = I$.

Let d/dt be a basis of $\mathbb{R}$ and consider $d\rho(0)(d/dt) = A \in u(n)$. Assume further that the set of equivariant harmonic maps

$$F_\rho = \{\phi \in C^\infty(S^1 \times \mathbb{R};\ F(n));$$

$$\phi(\exp(\sqrt{-1}\,\theta),t) = \rho(\exp(\sqrt{-1}\,\theta)) \cdot f(t),$$

where

$$f(t) = f_1(t),\ldots,f_n(t);$$

$$f_i^2(t) = f_i(t), \quad f_i(t) \cdot f_j(t) = 0$$

if $i \neq j$ and $f_i^*(t)$, and $\Sigma_i f_i(t) = I$, for all $\exp(\sqrt{-1}\,\theta) \in S^1\}$ is nonempty.

Note that $U(n)$ acts on $F(n)$ by conjugation:

$$U(n) \times F(n) \to F(n),$$

$$(A,X) \to AXA^{-1}.$$

Let ϕ: $(\Pi_1,\ldots,\Pi_n)$: $S^1 \times \mathbb{R} \to F(n)$ given by

$$\phi(\exp(\sqrt{-1}\,\theta),t) = (\Pi_1(\theta,t),\ldots,\Pi_n(\theta,t)) = \exp(A\theta) \cdot f(t).$$

So $\Pi_i(\theta,t) = \exp(A\theta) f_i(t) \exp(-A\theta)$.

Now by studying special cases of a general second-order ordinary differential equation, we will construct examples of harmonic maps $\phi = (\Pi_1,\ldots,\Pi_n)$: $T^2 = S^1 \times S^1 \to F(n)$ that are not holomorphic with respect to any almost complex structure on $F(n)$, where $F(n)$ is equipped with any left-invariant metric defined in Sec. II.

Consider a local chart $U \subseteq \mathbb{R}^2$ for a Riemann surface M^2 and B_1, B_2 in $u(n)$ such that $[B_1,B_2] = 0$. Then we can define locally the following map:

$$\tilde{U} \xrightarrow{\tilde{\phi}} U(n),$$

$$(x,y) \mapsto \exp(B_1 x + B_2 y).$$

We have seen that $\tilde{\phi}$ induces a map $\phi = (\Pi_1,\ldots,\Pi_n)$: $U \to F(n)$ given by $\Pi_i = \tilde{\phi} E_i \tilde{\phi}^* = \exp(B_1 x + B_2 y) \cdot E_i \cdot \exp(-B_1 x - B_2 y)$. We can prove the following.

Lemma 4.3: Let $\phi = (\Pi_1,\ldots\Pi_n)$: $U \to F(n)$ given by

$$\Pi_i = \exp(B_1x + B_2y)\cdot E_i\cdot\exp(-B_1x - B_2y),$$
where B_1,B_2 are in $u(n)$ and $[B_1,B_2]=0$. Then

$$\mathscr{A}_x = \sum_{i\neq j} a^{ij}\exp(B_1x+B_2y)E_iB_1E_j\exp(-B_1x-B_2y),$$

$$\mathscr{A}_y = \sum_{\substack{i,j\\ i\neq j}} a^{ij}\exp(B_1x+B_2y)E_iB_2E_j \times\exp(-B_1x-B_2y).$$

Proof: We will prove the expression for $\mathscr{A}_x$, and the one for $\mathscr{A}_y$ is proved similarly. We have

$$a^{ij}A_x^{ji} = a^{ji}\Pi_j\frac{\partial\Pi_i}{\partial x} = a^{ji}\Pi_j(B_1\exp(B_1x+B_2y)\cdot E_i \cdot\exp(-B_1x-B_2y) - \exp(B_1x+B_2y)E_iB_1 \times\exp(-B_1x-B_2y)).$$

But $B_1\cdot B_2 = B_2\cdot B_1$ so we have

$$a^{ji}\Pi_j(\exp(B_1x+B_2y)(B_1E_i - E_iB_1)\exp(-B_1x-B_2y))$$
$$= a^{ji}\exp(B_1x+B_2y)E_j\cdot\exp(-B_1x-B_2y) \times\exp(B_1x+B_2y)\cdot[B_1,E_i]\exp-(B_1x-B_2y)$$
$$= a^{ji}\exp(B_1x+B_2y)E_1B_1E_i\cdot\exp(-B_1x-B_2y).$$

Now we can find the Euler–Lagrange equations for the equivariant maps defined above.

Proposition 4.4: Let $\phi = (\Pi_1,\dots,\Pi_n):U\to F(n)$ be a smooth map such that $\Pi_i = \exp(B_1x+B_2y)\cdot E_i\cdot\exp(-B_1x-B_2y)$, where B_1 and B_2 are in $u(n)$ and $[B_1,B_2]=0$. Then ϕ is harmonic if and only if

$$\sum_{\substack{i,j\\ i\neq j}} a^{ij}E_i([B_1,\operatorname{diag} B_1] + [B_2,\operatorname{diag} B_2])E_j = 0,$$

where $\operatorname{diag}(B_i)$ denotes the diagonal part of B_i, $i=1,2$.

Proof: According to Corollary 3.3 ϕ is harmonic if and only if

$$\frac{\partial}{\partial x}(\mathscr{A}_x^a + \frac{\partial}{\partial y}(\mathscr{A}_y^a) = 0.$$

Hence let us compute $(\partial/\partial x)(\mathscr{A}_x^a)$ and $(\partial/\partial y)(\mathscr{A}_y^a)$. We have

$$\frac{\partial}{\partial x}(\mathscr{A}_x^a) = a^{ij}B_1\exp(B_1x+B_2y)E_iB_1E_j\exp(-B_1x-B_2y)$$
$$- a^{ij}\exp(B_1x+B_2y)E_iB_1E_j\cdot B_1\exp(-B_1x-B_2y)$$
$$\times\exp(B_1x-B_2y)\left[B_1,\sum_{\substack{i,j\\ i\neq j}} a^{ij}E_iB_1E_j\right]\exp(-B_1x-B_2y)$$
$$= \exp(B_1x+B_2y)a^{ij}E_i[B_1,\operatorname{diag} B_1]E_j\exp(-B_1x-B_2y).$$

Simlarly, we prove that

$$\frac{\partial}{\partial y}(\mathscr{A}_y^a) = \exp(B_1x+B_2y)a^{ij}E_i[B_2,\operatorname{diag} B_2]E_j \times\exp(-B_1x-B_2y).$$

Theorem 4.5: Let

$$\phi = (\Pi_1,\dots,\Pi_n):\ \frac{\mathbb{R}^2}{\alpha\mathbb{Z}\oplus\beta\mathbb{Z}} = T^2\to(F(n),g_{a=(a^{ij})})$$

be an equivariant map defined as in Lemma 4.3, where

$$\Pi_i = \exp(B_1x+B_2y)E_i\cdot\exp(-B_1x-B_2y),$$

and B_1,B_2 are in $u(n)$ with $[B_1,B_2]=0$. Furthermore, assume $E_iB_kE_j\neq 0$ for some $1\leqslant i\neq j\leqslant n$, $k=1$ or 2, and that

$$\sum_{\substack{i,j\\ i\neq j}} a^{ij}E_i([B_1,\operatorname{diag} B_1] + [B_2,\operatorname{diag} B_2])E_j = 0.$$

Then ϕ is harmonic with respect to the metric $g_{a=(a^{ij})}$ but is not holomorphic with respect to any almost complex structure on $F(n)$.

Proof: According to our hypothesis and Proposition 4.4., ϕ is harmonic. On the other hand, $\mathscr{A}_Z^{ij} = \mathscr{A}_x^{ij} + \sqrt{-1}\,\mathscr{A}_y^{ij}$ and $\mathscr{A}_Z^{ji} = A_x^{ji} + \sqrt{-1}\,\mathscr{A}_y^{ji}$ are both nozero according to our hypothesis and Lemma 4.3. Therefore, according to the holomorphic map equations in Sec. III, ϕ is not holomorphic with respect to any almost complex structure on $F(n)$.

The result above allows us to construct several examples of harmonic and nonholomorphic maps from T^2 to $(F(n),g_a)$. For example, let

$$f:\mathbb{R}\to U(n),$$
$$t\mapsto\exp(Bt),$$

where

$$B = \begin{pmatrix} 0 & \alpha\sqrt{-1} & 0 & 0 & \cdots & 0\\ \alpha\sqrt{-1} & 0 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & \beta\sqrt{-1} & \cdots & 0\\ 0 & 0 & \beta\sqrt{-1} & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \in u(n),$$

and α and β are nonzero real numbers.

Then

$$\exp(Bt) = I + Bt + \frac{(Bt)^2}{2!} + \cdots + \frac{(Bt)^n}{n!} + \cdots.$$

Hence

$$f(t)\exp(Bt)=\begin{pmatrix}\cos\alpha t & 0 & 0 & 0 & \cdots & 0\\ 0 & \cos\alpha t & 0 & 0 & \cdots & 0\\ 0 & 0 & \cos\beta t & \cdots & 0 & \\ 0 & 0 & 0 & \cos\beta t & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0\end{pmatrix}+\sqrt{-1}\begin{pmatrix}0 & \sin\alpha t & 0 & 0 & \cdots & 0\\ \sin\alpha t & 0 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & \sin\beta t & \cdots & 0\\ 0 & 0 & \sin\beta t & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0\end{pmatrix}.$$

Let us consider as the first set of examples the case $B_1=B_2=B$, where α and β are nonzero real numbers.

Now let us consider $\tilde{\phi}$: $\mathbb{R}^1\to U(n)$ given by $\tilde{\phi}(x,y)=\exp(Bx+By)$. Then $\tilde{\phi}$ induces a map:

$$\phi:\frac{\mathbb{R}^2}{(2\pi/\alpha)\mathbb{Z}\oplus(2\pi/\alpha)\mathbb{Z}}\to F(n)$$

given by

$$\phi\left(x+\frac{2\Pi}{\alpha}n,y+\frac{2\Pi}{\beta}m\right)=\tilde{\phi}(x,y,(E_1,...,E_n)\tilde{\phi}^*(x,y)=\exp(Bx+By)(E_1,...,E_n)\exp(-Bx-By).$$

Since diag $B=0$, according to Theorem 4.5, ϕ is harmonic with respect to any left-invariant metric on $F(n)$ but is not holomorphic with respect to any almost complex structure on $F(n)$, since $E_1BE_2=E_2BE_1=\alpha\sqrt{-1}\neq 0$.

$$B=\begin{pmatrix}0 & \alpha_1\sqrt{-1} & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0\\ \alpha_1\sqrt{-1} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0\\ 0 & 0 & 0 & \alpha_2\sqrt{-1} & & 0 & \cdots & 0 & \cdots & 0\\ 0 & 0 & \alpha_2\sqrt{-1} & 0 & \ddots & & \cdots & 0 & \cdots & 0\\ & & & & & & 0 & \alpha_k\sqrt{-1} & \cdots & 0\\ 0 & 0 & 0 & 0 & \cdots & 0 & \alpha_k\sqrt{-1} & 0 & \cdots & 0\\ & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & 0\\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & & & 0\end{pmatrix}\in u(n),$$

such that $2k\leqslant n$.

Another family of harmonic with respect to any left-invariant metric on $F(n)$ as defined in Sec. II but nonholomorphic maps is given by

$$B_1=\begin{pmatrix}0 & \alpha\sqrt{-1} & 0 & 0 & \cdots & 0\\ \alpha\sqrt{-1} & 0 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & \beta\sqrt{-1} & \cdots & 0\\ 0 & 0 & \beta\sqrt{-1} & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0\end{pmatrix}$$

and

$$B_2\begin{pmatrix}0 & \beta\sqrt{-1} & 0 & 0 & \cdots & 0\\ \beta\sqrt{-1} & 0 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & \alpha\sqrt{-1} & \cdots & 0\\ 0 & 0 & \alpha\sqrt{-1} & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0\end{pmatrix},$$

where α and β are nonzero real numbers such that $\alpha/\beta\in Q$. Then B_1 and B_2 are in $u(n)$, $[B_1,B_2]=0$, and furthermore there exists $\gamma\in\mathbb{R}$ such that $\alpha\cdot\gamma$ and $\beta\cdot\gamma$ are integers.

Now let us consider

$$\phi:\frac{\mathbb{R}^2}{2\Pi\gamma(\mathbb{Z}\oplus\mathbb{Z})}\to F(n)$$

given by

$$\begin{aligned}\phi(x&+2\Pi n\gamma,y+2\Pi m\gamma)\\ &=\tilde{\phi}(x,y)(E_1,...,E_n)\phi^*(x,y)\\ &=\exp(B_1x+B_2y)(E_1,...,E_n)\exp)-B_1x-B_2y).\end{aligned}$$

But $\mathrm{diag}(B_1)=\mathrm{diag}\ (B_2)=0$. Then again, using Theorem 4.5, we see that ϕ is harmonic with respect to all left-invariant metrics defined in Sec. II but not holomorphic, since $E_1B_1E_2=E_2B_1E_1=\alpha\sqrt{-1}\neq 0$.

We can generalize this example by taking $B_1 \in u(n)$ of the following form:

$$B_1 = \begin{pmatrix} B_1^1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & B_2^1 & & \vdots & & \vdots \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ & & & B_1^k & & \vdots \\ \vdots & \vdots & & & \ddots & \\ 0 & 0 & & & 0 & 0 \end{pmatrix},$$

where

$$B_1^i = \begin{pmatrix} 0 & \alpha_i\sqrt{-1} & 0 & 0 \\ \alpha_i\sqrt{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_i\sqrt{-1} \\ 0 & 0 & \beta_i\sqrt{-1} & 0 \end{pmatrix},$$

and

$$B_2 = \begin{pmatrix} B_2^1 & & & & \\ & \ddots & & & \\ & & B_2^k & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix},$$

where

$$B_2^i = \begin{pmatrix} 0 & \beta_i\sqrt{-1} & 0 & 0 \\ \beta_i\sqrt{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_i\sqrt{-1} \\ 0 & 0 & \alpha_i\sqrt{-1} & 0 \end{pmatrix},$$

such that $2k \leqslant n$. Furthermore, we also assume that $\alpha_1/\beta_1 = \cdots = \alpha_k/\beta_k$ are all rational numbers.

It would be interesting to look for another way of generating harmonic with respect to all left-invariant metrics defined in Sec. II but not holomorphic with respect to any almost complex structure on $F(n)$. Another class of harmonic but holomorphic maps are the Eells–Wood maps. See Refs. 12 or 14 for more details.

It would be nice to understand the stability of the maps that we have built in this paper with respect to the family of left-invariant metrics defined in Sec. II.

We notice that Lemma 5.4 in Ref. 14 would be true in this case, and to use this lemma in a profitable way, it would be necessary only to understand the stability of such maps when $F(n)$ is equipped with Kähler metrics.

Now let us show how these examples above provide examples of harmonic and nonholomorphic maps from T^2 into $\Omega(U(n))$, where $\Omega(U(n))$ is equipped with thes usual symplectic Kähler metric. To see this, we sketch the holomorphic and totally geodesic embedding of $F(n)$ into $\Omega(U(n))$ according to Ref. 9.

Let us now recall some basic facts about $\Omega(U(n))$. See Ref. 7 for many more details. Let $\Omega(U(n)) = \{f: S^1 \to U(n)$ smooth such that $f(1) = I\}$. We can put a group structure in $\Omega(U(n))$ defining $(f\cdot g)(e^{\sqrt{-1}\,\theta}) = f(e^{\sqrt{-1}\,\theta}\cdot g(e^{\sqrt{-1}\,\theta})$.

The simplest case is when $G = U(1)$. Then $\Omega(U(1))$ has components indexed according to the winding number, and each component can be identified with the space of functions $f: S^1 \to \mathbf{R}$ such that $f(1) = 0$. The Fourier series of such a function is

$$\phi = \sum_{n=-\infty}^{\infty} a_n z_n, \quad a_{-n} = \bar{a}_n, \quad \sum_n a_n = 0,$$

therefore the coefficients a_n, for $n > 0$, determine ϕ completely. Hence each component of $\Omega(U(1))$ becomes a complex vector space of infinite dimension.

For non-Abelian G, $\Omega(G)$ is not linear any more. However, it still is an infinite-dimensional manifold, and we can again use Fourier series to introduce complex coordinates. We know that $\Omega(G)_I$ is equal to $\Omega(g)$ and can be represented in Fourier series as

$$\phi = \sum_{n=-\infty}^{\infty} a_n z^n, \quad a_{-n} = -a_n^*, \quad \sum_n a_n = 0,$$

where $a_n \in g_{\mathbf{C}}$. If $G = U(m)$, then $a_n \in \mathbf{C}_m$ and a_n^* is the transpose conjugate matrix. So $\Omega(G)_1$ becomes an infinite-dimensional complex vector space.

Now we can define several almost complex structures on $\Omega(G)$, namely if $\phi \in \Omega(G)_I$, we define

$$J(\phi) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \alpha_n \sqrt{-1}\, a_n Z^n,$$

where $\alpha_n = \pm 1$ and $\alpha_{-n} = \mp 1$. The almost complex structure obtained by making $\alpha_n = 1$, $\forall n > 0$, is integrable and is called the canonical almost complex structure.

The next point is to define a canonical Kähler structure on $\Omega(G)$. To define a Hermitian metric on $\Omega(G)$ is again enough to define in $\Omega(G)_I$ and translate it via the group action. There are several natural left-invariant metrics on $\Omega(G)$ (see Ref. 7 for more details), but it seems to be the most natural when given with respect to the Fourier coefficients by $\Sigma_{n>0}\, n\, \mathrm{tr}(a_n a_n^*)$ where a_n is seen as a matrix. An important reason for this metric be the natural one to be considered relies on the fact that it is Kähler. The symplectic form associated is given by

$$(\phi,\psi) = \frac{1}{2\pi}\int_0^{2\pi} \langle \phi'(\theta), \psi(\theta)\rangle d\theta,$$

where $\langle\ ,\ \rangle$is given by the Killing form metric and $\phi'(\theta) = (d\phi/d\theta)(\theta)$.

Now let us recall the natural totally geodesic and holomorphic embedding of $F(n)$ into $\Omega(U(n))$ as in Ref. 9.

Let $\Gamma = \mathrm{Hom}(S^1, G)$ the subgroup formed by closed geodesics. Clearly G acts on Γ by conjugation. Furthermore, each connected component is a G-orbit, i.e., is of the form $\{g\gamma g^{-1}, g \in G$ and fixed $\gamma \in \Gamma\}$.

We know that $\gamma(\exp(\sqrt{-1}\,t)) = \exp(t\xi)$ for some $\xi \in g$ such that $\exp(2\pi\xi) = I$. We note that $g\gamma g^{-1}(\exp(\sqrt{-1}\,t)) = g\exp(t\xi)g^{-1} = \exp(t\,\mathrm{Ad}(g)\xi)$. Therefore the G-orbit of γ is of the form $\mathrm{Ad}(g)\xi = G/H$ where $H = \{g \in G; \mathrm{Ad}(g)\xi = \xi\}$, i.e., H is the centralizer of a torus.

Then we can define the embedding

$$\psi: G/H \to \Omega(G),$$
$$gH \mapsto g\gamma g^{-1}.$$

If we put on G/H the pulled back Kähler metric of $\Omega(G)$ and consider compactible almost complex structures, we can see that ψ is totally geodesic and holomorphic.

In the case we are primarily interested in this note it is enough to consider $G = U(n)$ and

$$\xi = \begin{pmatrix} \sqrt{-1}\lambda_1 & & & \\ & \sqrt{-1}\lambda_2 & & 0 \\ & & \ddots & \\ 0 & & & \sqrt{-1}\lambda_n \end{pmatrix},$$

where $\lambda_i \neq \lambda_j$ if $i \neq j$, $\lambda_i \in \mathbb{Z}$ so $\exp(2\pi\xi) = I$.

Let $g_{a=(a^{ij})}$ be the Kähler metric obtained pullingback via ψ the natural Kähler metric on $\Omega(U(n))$ constructed above. Now let ϕ: $T^2 \to (F(n), g_{a=(a^{ij})})$ be a harmonic but not holomorphic map with respect to any almost complex structure on $F(n)$. Now since ψ: $(F(n), g_a) \to \Omega(U(n))$, the Kähler metric is totally geodesic and holomorphic we have that $\phi = \psi \circ \phi$: $T^2 \to \Omega(U(n))$ is harmonic but not holomorphic.

ACKNOWLEDGMENTS

I want to express my gratitude to my thesis advisor Professor Karen Uhlenbeck for her deep advice and criticism. The author also wants to thank the referee for the useful comments.

[1] C. J. C. Negreiros, "Equivariant harmonic maps into flag manifolds," Internal Report No. 46, University of Campinas (1987).
[2] H. Urakawa, "Equivariant theory of Yang–Mills connections over Riemannian manifolds of cohomogeneity one," Indiana Univ. Math. **37**, 753 (1988).
[3] M. F. Atiyah, "Instantons in two and four manifolds," Commun. Math. Phys. **93**, 437 (1984).
[4] S. K. Donaldson, "Instantons and geometric invariant theory," Commun. Math. Phys. **93**, (1984).
[5] L. M. Sibner, R. J. Sibner, and K. Uhlenbeck, "Solutions of Yang–Mills equations which are not self-dual," preprint (1989).
[6] D. S. Freed, "The geometry of loop groups," J. Differential Geom. **28**, 223 (1988).
[7] A. N. Pressley and G. B. Segal, *Loop Groups* (Oxford U.P., City, 1987).
[8] K. Uhlenbeck, "Harmonic maps into Lie groups (Classical solutions of the Chiral model)," to appear in J. Differential Geom.
[9] J. Rawnsley, "Embedding flag manifolds into loop groups," manuscripted notes, University of Warwick (1983).
[10] C. J. C. Negreiros, "Harmonic maps from compact Riemann surfaces into flag manifolds," doctoral thesis, University of Chicago (1987).
[11] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, City, 1978).
[12] J. Eells and J. C. Wood, "Harmonic maps from surfaces to complex projective spaces," Advances in Math. **49**, 217 (1983).
[13] M. A. Guest, "The geometry of maps between generalized flag manifolds," J. Differential Geom. **25**, 223 (1987).
[14] C. J. C. Negreiros, "Some remarks about harmonic maps into flag manifolds," Indiana Univ. Math. **37**, 617 (1988).
[15] R. S. Palais, "The principle of symmetric criticality," Commun. Math. Phys. **69**, 19 (1979).
[16] W. Y. Hsiang and H. B. Lawson, Jr., "Minimal submanifolds of low cohomogeneity," J. Differential Geom. **5**, 1 (1971).
[17] K. Uhlenbeck, "Equivariant harmonic maps into spheres," *Lecture Notes in Mathematics*, Vol. 949 (Springer, New York).

Geometry of linear pairs for self-dual gauge fields

C Roiesnel and A. Chakrabarti
Centre de Physique Théorique de l'Ecole Polytechnique, Laboratoire Propre du C.N.R.S. n°14, 91128 Palaiseau Cedex, France

(Received 2 November 1989; accepted for publication 7 March 1990)

A linear pair for self-dual gauge fields is constructed for the metric $ds^2 = g_{z\bar z}\, dz\, d\bar z + g_{y\bar y}\, dy\, d\bar y$. It is shown that for consistency $g_{z\bar z}$ and $g_{y\bar y}$, apart from a possible overall conformal factor, are given in terms of two Liouville fields of equal and opposite curvatures. The null surface corresponding to the pair and the homogeneous solutions, playing a fundamental role, are constructed explicitly. The five-dimensional space of $y,\bar y,z,\bar z$ and the spectral parameter λ is studied. The proper transformation of λ corresponding to holomorphic ones of y and z is found. Known monopole, instanton, and (quasi)periodic solutions are all shown to emerge systematically as particular cases of our formalism. As examples of new possibilities, the case of accelerated observers and that of cosmic string backgrounds are presented.

I. INTRODUCTION

Linear pairs à la Zakharov and his co-workers[1-3] furnish a supple and powerful method for explicit construction of self-dual gauge fields. This formalism is closely related to the language of twistors.[4,5] Linear pairs were adapted to the construction of monopoles by Forgacs *et al.*[6,7] They used Cartesian and cylindrical coordinates for the pair. Hyperbolic and spherical coordinates have been used successfully to construct instantons[8,9] containing monopoles as limits and periodic[10] and quasiperiodic[11] solutions. In the general ADHM method[12] the final nonlinear constraint cannot be resolved except in very particular cases. Really complete and explicit solutions with unlimited ranges of indices have, up to now, been obtained for higher Atiyah–Ward classes only for restricted symmetries.[8-11] One may, for example, have axial symmetry along with the possibility of periodicity in time. In this context the proper choice of coordinates, adapted to the symmetry in question, is vital. Then independence with respect to one or more coordinates can be implemented from the beginning.

Here we present the linear pair for a class of metrics from which the interesting particular cases emerge directly and systematically. Thus different approaches are encompassed and unified. Moreover, our formalism leads to a deeper understanding of the geometry of self-duality in four dimensions. The privileged class of metrics, permitting consistent construction of a linear pair, is shown to be given, apart from a conformal factor, in terms of *two Liouville fields of equal and opposite curvatures*. This is the basic result leading to the rest. The geometry is elucidated by constructing the null surface associated to the pair. The homogeneous solutions, the kernels of the mappings induced by the operators D_1 and D_2 of the pair, are given explicitly. Their role is fundamental. The five-dimensional space formed by the coordinates $y,\bar y,z,\bar z$ and the spectral parameter λ is studied to better situate the null surface. A transformation of λ accompanying holomorphic ones of y and z (with conjugates ones for $\bar y$ and $\bar z$) is shown to lead to a better understanding of the general structure.

All the solitonic solutions cited above are extracted systematically from the general case as particular ones. The case of an accelerated observer is also treated.

Different choices of coordinates in locally Euclidean space do not exhaust the content of our formalism. As an example we indicate how the metric of a class of cosmic strings fits into this framework. Other possibilities should be looked for.

Our formalism seems to point out the minimal link (through equal and opposite curvatures) between two-dimensional conformal properties leading to nontrivial results in four dimensions. This aspect should be explored in broader context not limited to self-dual gauge fields.

II. THE LINEAR SYSTEM

We work in complexified four-dimensional space, i.e., we regard the space–time coordinates x_0, x_1, x_2, and x_3 as complex variables. Let the metric be

$$ds^2 = g_{z\bar z}\, dz\, d\bar z + g_{y\bar y}\, dy\, d\bar y, \tag{2.1}$$

with

$$z = x_3 + ix_0, \quad \bar z = x_3 - ix_0,$$
$$y = x_1 + ix_2, \quad \bar y = x_1 - ix_2, \tag{2.2}$$

and

$$g_{\mu\bar\mu} = g_{\mu\bar\mu}(y,z,\bar y,\bar z) = (g^{\mu\bar\mu})^{-1} \quad (\mu = y,z). \tag{2.3}$$

We coin the term "biconformal" for this class of four-dimensional metrics, for reasons that will become clear later on. The self-dual Yang–Mills equations,

$$F_{\mu\nu} = ({}_*F)_{\mu\nu} \equiv \tfrac{1}{2}\sqrt{|g|}\,\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}, \tag{2.4}$$

take a very convenient form in biconformal metrics:

$$F_{yz} = F_{\bar y\bar z} = 0, \tag{2.5}$$
$$g^{z\bar z}F_{z\bar z} + g^{y\bar y}F_{y\bar y} = 0. \tag{2.6}$$

They can be viewed as two curvatureless conditions on the yz and the $\bar y\,\bar z$ planes and a third constraining equation. An associated linear system is easily constructed by requiring that the three self-dual equations be encompassed into a single one. Define

$$\rho = (g^{y\bar y}/g^{z\bar z})^{1/2}, \tag{2.7}$$

and

 0022-2488/90/071643-09$03.00

$$A_1 = \lambda\rho^{-1}A_{\bar z} + A_y, \tag{2.8a}$$

$$A_2 = -\lambda\rho A_{\bar y} + A_z, \tag{2.8b}$$

where λ is a complex parameter. Then it is possible to find two linear differential operators D_1 and D_2 such that

$$F_{12} = D_1A_2 - D_2A_1 + i[A_1,A_2] \tag{2.9}$$

$$= F_{yz} + \lambda^2F_{\bar y\bar z} - \lambda\rho^{-1}(F_{z\bar z} + \rho^2F_{y\bar y}), \tag{2.10}$$

for all values of $\lambda\in\mathbb{C}$. We get

$$D_1 = \lambda\rho^{-1}\partial_{\bar z} + \partial_y - [\lambda\rho^{-1}\,\partial_{\bar z}\ln\rho + \partial_y\ln\rho]\lambda\,\partial_\lambda, \tag{2.11a}$$

$$D_2 = -\lambda\rho\,\partial_{\bar y} + \partial_z + [-\lambda\rho\,\partial_{\bar y}\ln\rho + \partial_z\ln\rho]\lambda\,\partial_\lambda. \tag{2.11b}$$

The self-dual potentials A_1 and A_2, which from Eq. (2.10) can be interpreted as curvatureless on some complex two-dimensional surface Σ, must be of the following form:

$$A_1 = i(D_1\psi)\psi^{-1}, \tag{2.12a}$$

$$A_2 = i(D_2\psi)\psi^{-1}, \tag{2.12b}$$

where ψ belongs to the complexified gauge group $G^{\mathbb{C}}$. These are precisely the linearized equations of Belavin and Zakharov[2] generalized to biconformal metrics.

III. THE ROLE OF TWO LIOUVILLE FIELDS

There is an implicit assumption in our derivation of the linear pair. Equations (2.12) hold true if and only if the operators D_1 and D_2 commute for any value of the spectral parameter λ:

$$\forall\lambda\in\mathbb{C},\quad [D_1,D_2] = 0. \tag{3.1}$$

This consistency condition does not hold for an arbitrary biconformal metric. A direct calculation yields the constraints:

$$\partial^2_{yz}\ln\rho = \partial^2_{\bar y\bar z}\ln\rho = 0, \tag{3.2}$$

$$\rho^{-1}\,\partial^2_{z\bar z}\ln\rho - \rho\,\partial^2_{y\bar y}\ln\rho = 0. \tag{3.3}$$

Equation (3.2) implies

$$\rho = \rho_2(z,\bar z)/\rho_1(y,\bar y), \tag{3.4}$$

which means that the metric, up to an overall conformal factor, is Kählerian. Then Eq. (3.3) becomes

$$\rho_2^{-2}\,\partial^2_{z\bar z}\ln\rho_2 = -\rho_1^{-2}\,\partial^2_{y\bar y}\ln\rho_1 = K, \tag{3.5}$$

for some constant K. We recognize Liouville's equation. Therefore a full parametrization of the admissible biconformal metrics reads:

$$\rho = \frac{1+K\omega_1\overline{\omega}_1}{1-K\omega_2\overline{\omega}_2}\left(\frac{\partial_z\,\omega_2\,\partial_{\bar z}\,\overline{\omega}_2}{\partial_y\omega_1\,\partial_{\bar y}\,\overline{\omega}_1}\right)^{1/2}, \tag{3.6}$$

where $\omega_1(y)$, $\omega_2(z)$, $\overline{\omega}_1(\bar y)$, and $\overline{\omega}_2(\bar z)$ are independent arbitrary holomorphic functions. Hence two-dimensional surfaces with opposite constant curvature play a special role in the geometry of self-duality. From now on, by biconformal metric, we mean this restricted class of metrics.

IV. THE NULL SURFACES

There is a standard technique to solve any system of first-order, linear partial differential equations, which is the method of characteristics. Its application to our linear pair is straightforward. Let the components of two vector fields $\mathbf{V}_1(y,z,\bar y,\bar z,\lambda)$ and $\mathbf{V}_2(y,z,\bar y,\bar z,\lambda)$ be given by the coefficient functions of the differential operators D_1 and D_2. These vector fields $\mathbf{V}_1$ and $\mathbf{V}_2$ define the tangent planes to some (complex) two-dimensional surface Σ called the characteristic surface of the differential system. The parametric equations of Σ depend upon two complex variables α and β:

$$\begin{aligned}
&\frac{\partial y}{\partial\alpha} = 0,\quad \frac{\partial y}{\partial\beta} = 1,\\
&\frac{\partial\bar y}{\partial\alpha} = -\lambda\rho,\quad \frac{\partial\bar y}{\partial\beta} = 0,\\
&\frac{\partial z}{\partial\alpha} = 1,\quad \frac{\partial z}{\partial\beta} = 0,\\
&\frac{\partial\bar z}{\partial\alpha} = 0,\quad \frac{\partial\bar z}{\partial\beta} = \lambda\rho^{-1},\\
&\frac{\partial\lambda}{\partial\alpha} = -\lambda(\lambda\rho\,\partial_{\bar y}\ln\rho - \partial_z\ln\rho),\\
&\frac{\partial\lambda}{\partial\beta} = -\lambda(\lambda\rho^{-1}\partial_{\bar z}\ln\rho + \partial_y\ln\rho).
\end{aligned} \tag{4.1}$$

Up to arbitrary integration constants, we get:

$$\begin{aligned}
&y = \beta,\\
&z = \alpha,\\
&\bar y = \bar y(\alpha),\quad \frac{d\bar y}{d\alpha} = -\lambda\rho,\\
&\bar z = \bar z(\beta),\quad \frac{d\bar z}{d\beta} = \lambda\rho^{-1},\\
&\lambda = \lambda(\alpha,\beta),\quad \frac{d}{d\alpha}(\lambda\rho^{-1}) = \frac{d}{d\beta}(\lambda\rho) = 0,
\end{aligned} \tag{4.2}$$

which imply (on Σ):

$$\lambda = \left(-\frac{d\bar y}{d\alpha}\cdot\frac{d\bar z}{d\beta}\right)^{1/2}, \tag{4.3}$$

$$\rho = \left(-\frac{d\bar y}{d\alpha}\Big/\frac{d\bar z}{d\beta}\right)^{1/2}. \tag{4.4}$$

It follows that the Riemannian metric on Σ is

$$\begin{aligned}
ds^2 &= g_{z\bar z}\,dz\,d\bar z + g_{y\bar y}\,dy\,d\bar y\\
&= \left(g_{z\bar z}\frac{d\bar z}{d\beta} + g_{y\bar y}\frac{d\bar y}{d\alpha}\right)d\alpha\,d\beta\\
&= 0
\end{aligned} \tag{4.5}$$

which means that the characteristic surface of the linear pair is a "null" surface or a "light" surface in Euclidean and Minkowski terminology, respectively.

V. THE HOMOGENEOUS SOLUTIONS

We still need to solve for $\bar y(\alpha)$ and $\bar z(\beta)$ to get the explicit equation of the null surface. We shall use the general parametrization of the ratio ρ in Eq.. (3.6). On Σ we have

$$\begin{aligned}
\rho^2 &= -\frac{d\bar y/d\alpha}{d\bar z/d\beta}\\
&= \left(\frac{1+K\omega_1(\beta)\overline{\omega}_1(\alpha)}{1-K\omega_2(\alpha)\overline{\omega}_2(\beta)}\right)^2\frac{\partial_\alpha\omega_2}{\partial_\beta\omega_1}\cdot\frac{\partial_\beta\overline{\omega}_2}{\partial_\alpha\overline{\omega}_1}\left(\frac{d\bar y/d\alpha}{d\bar z/d\beta}\right),
\end{aligned} \tag{5.1}$$

which holds true if and only if

$$\frac{\partial_\alpha \omega_2 \partial_\beta \overline{\omega}_2}{(1 - K\omega_2(\alpha)\overline{\omega}_2(\beta))^2} = -\frac{\partial_\beta \omega_1 \partial_\alpha \overline{\omega}_1}{(1 + K\omega_1(\beta)\overline{\omega}_1(\alpha))^2}. \tag{5.2}$$

Equation (5.2) states that $(-\overline{\omega}_1,\omega_1)$ and $(\omega_2,\overline{\omega}_2)$ define the same solution of the Liouville's equation

$$e^{-2\Phi}\partial^2_{\alpha\beta}\Phi = -K, \tag{5.3}$$

whose general solution is

$$\Phi(\alpha,\beta) = \ln \frac{(\partial_\alpha p\, \partial_\beta q)^{1/2}}{1 - Kp(\alpha)q(\beta)}, \tag{5.4}$$

where $p(\alpha)$ and $q(\beta)$ are two arbitrary holomorphic functions of the independent complex variables α and β. The invariance group of Liouville's equation is the group of linear fractional transformations. Indeed the solution is invariant, $\Phi = \Phi'$, under the homographic transformations

$$p \to p' = (ap + b)/(Kcp + d), \tag{5.5}$$

$$q \to q' = (dq + c)/(Kbq + a), \tag{5.6}$$

where a, b, c, and d are arbitrary complex constants with $ad - Kbc \neq 0$. It follows that Eq. (5.2) is solved by

$$\overline{\omega}_1 = (-a\omega_2 + c)/(Kb\omega_2 - d), \tag{5.7}$$

$$\overline{\omega}_2 = (a\omega_1 - b)/(Kc\omega_1 - d). \tag{5.8}$$

These expressions provide an explicit parametric equation of the null surfaces together with

$$\lambda = \left(-\frac{d\bar{y}}{d\alpha}\cdot\frac{d\bar{z}}{d\beta}\right)^{1/2} = \left(\frac{\partial_z \omega_2}{\partial_{\bar{y}}\overline{\omega}_1}\cdot\frac{\partial_y \omega_1}{\partial_{\bar{z}}\overline{\omega}_2}\right)^{1/2}\left(-\frac{d\overline{\omega}_1}{d\omega_2}\cdot\frac{d\overline{\omega}_2}{d\omega_1}\right)^{1/2}. \tag{5.9}$$

In fact, the null surfaces Σ are parametrized by the four complex numbers $(a,b,c,d)\in\mathbb{C}^4$ modulo any complex number $\zeta \neq 0$

$$\Sigma(a,b,c,d) \equiv \Sigma(\zeta a,\zeta b,\zeta c,\zeta d). \tag{5.10}$$

Hence the space of null surfaces forms a complex projective three-dimensional manifold CP^3. Although CP^3 is compact and cannot be completely spanned by three complex coordinates, it is useful to introduce a three-dimensional coordinate system. When $d \neq 0$, we can solve for a/d, b/d, and c/d as functions of $y,z,\bar{y},\bar{z}$, and λ:

$$\frac{a}{d} = \frac{\lambda\Omega + K\overline{\omega}_1\overline{\omega}_2\overline{\Omega}}{\lambda K\omega_1\omega_2\Omega - \overline{\Omega}},\quad \frac{b}{d} = \frac{\lambda\omega_1\Omega - \overline{\omega}_2\overline{\Omega}}{\lambda K\omega_1\omega_2\Omega - \overline{\Omega}},\quad \frac{c}{d} = \frac{\lambda\omega_2\Omega + \overline{\omega}_1\overline{\Omega}}{\lambda K\omega_1\omega_2\Omega - \overline{\Omega}}, \tag{5.11}$$

with $\Omega = (\partial_y \omega_1 \partial_z \omega_2)^{-1/2}$ and $\overline{\Omega} = (\partial_{\bar{y}}\overline{\omega}_1\partial_{\bar{z}}\overline{\omega}_2)^{-1/2}$.

By definition, these functions are constant on a null surface. These independent normal coordinates to Σ provide an implicit equation of the null surface. It follows that the general solution to the homogeneous linear pair,

$$D_1X = D_2X = 0, \tag{5.12}$$

is an arbitrary algebraic function of a,b,c, and d, homogeneous to degree zero:

$$X = X\left(\frac{a}{d},\frac{b}{d},\frac{c}{d}\right). \tag{5.13}$$

It is also convenient to express the null surface equations in terms of complex two-spinors. Equations (5.7) and (5.8) can be written as

$$\begin{bmatrix} b \\ c \end{bmatrix} = (q)\begin{bmatrix} a \\ d \end{bmatrix}, \tag{5.14}$$

where the complex (2×2) matrix (q) reads:

$$(q) = \frac{1}{1 + K^2\omega_1\overline{\omega}_1\omega_2\overline{\omega}_2} \times \begin{bmatrix} \omega_1(1 - K\omega_2\overline{\omega}_2) & \overline{\omega}_2(1 + K\omega_1\overline{\omega}_1) \\ \omega_2(1 + K\omega_1\overline{\omega}_1) & -\overline{\omega}_1(1 - K\omega_2\overline{\omega}_2) \end{bmatrix}. \tag{5.15}$$

Any point $x \equiv (y,z,\bar{y},\bar{z})$ in four-complex-dimensional space with biconformal metric can be represented by the matrix (q) that generalizes the quaternion representation of Euclidean space:

$$\begin{bmatrix} y & \bar{z} \\ z & -\bar{y} \end{bmatrix}. \tag{5.16}$$

Thus, many features of the Atiyah–Ward formulation of self-dual gauge fields in Euclidean metric[4,5] can be generalized to biconformal metrics.

VI. TRANSFORMATIONS OF THE SPECTRAL PARAMETER

The structure of the kernels of the mappings induced by D_1 and D_2, as given in (5.11), suggests the transformation

$$\Lambda = \lambda\frac{\Omega}{\overline{\Omega}} = \lambda\left[\frac{\dfrac{d\overline{\omega}_1}{d\bar{y}}\cdot\dfrac{d\overline{\omega}_2}{d\bar{z}}}{\dfrac{d\omega_1}{dy}\,\dfrac{d\omega_2}{dz}}\right]^{1/2}, \tag{6.1}$$

giving, setting $d = 1$,

$$a = \frac{\Lambda + K\overline{\omega}_1\overline{\omega}_2}{\Lambda K\omega_1\omega_2 - 1},\quad b = \frac{\Lambda\omega_1 - \overline{\omega}_2}{\Lambda K\omega_1\omega_2 - 1},\quad c = \frac{\Lambda\omega_2 + \overline{\omega}_1}{\Lambda K\omega_1\omega_2 - 1}. \tag{6.2}$$

Moving over to the coordinates (ω_1,ω_2) from (y,z) (with C.F. = conformal factor)

$$ds^2 = (\text{C.F.})[(1 + K\omega_1\overline{\omega}_1)^{-2}d\omega_1\, d\overline{\omega}_1 + (1 - K\omega_2\overline{\omega}_2)^{-2}d\omega_2\, d\overline{\omega}_2], \tag{6.3}$$

and

$$\mathscr{D}_1 = \left(\frac{d\omega_1}{dy}\right)^{-1} \quad D_1 = \partial\omega_1 + \Lambda\frac{1 - K\omega_2\overline{\omega}_2}{1 + K\omega_1\overline{\omega}_1}\partial\overline{\omega}_2 - \frac{K(\overline{\omega}_1 + \Lambda\omega_2)}{1 + K\omega_1\overline{\omega}_1}\Lambda\,\partial_\Lambda, \tag{6.4}$$

$$\mathscr{D}_2 = \left(\frac{d\omega_2}{dy}\right)^{-1} \quad D_2 = \partial\omega_2 + \Lambda\frac{1 + K\omega_1\overline{\omega}_1}{1 - K\omega_2\overline{\omega}_2}\partial\overline{\omega}_1 - \frac{K(\overline{\omega}_2 + \Lambda\omega_1)}{1 - K\omega_2\overline{\omega}_2}\Lambda\,\partial_\Lambda. \tag{6.5}$$

For $K=0$,

$$\mathscr{D}_1 = \partial\omega_1 + \Lambda\,\partial\overline{\omega}_2,$$
$$\mathscr{D}_2 = \partial\omega_2 - \Lambda\,\partial\overline{\omega}_1. \tag{6.6}$$

Thus, starting with Cartesian coordinates, the effects of the holomorphic transformations $(y,z)\to(\omega_1(y),\,\omega_2(z))$ on D_1 and D_2 (that now have ∂_λ present) are absorbed through (6.1).

Even for $K\neq 0$, (6.1) helps in understanding the structure of the formalism. Corresponding to $\mathscr{D}_1$ and $\mathscr{D}_2$, define

$$\mathscr{A}_1 = \left(\frac{d\omega_1}{dy}\right)^{-1} A_1 = \left(\frac{d\omega_1}{dy}\right)^{-1}[A_y + \lambda\rho^{-1}A_z] \tag{6.7}$$

$$= A_{\omega_1} + \Lambda\frac{1-K\omega_2\overline{\omega}_2}{1+K\omega_1\overline{\omega}_1}A_{\omega_2} \tag{6.8}$$

$$\left(A_{\omega_2} = \left(\frac{d\overline{\omega}_2}{d\overline{z}}\right)^{-1} A_z \text{ etc.}\right).$$

Similarly,

$$\mathscr{A}_2 = \left(\frac{d\omega_2}{dz}\right)^{-1} A_2 = A_{\omega_2} - \Lambda\frac{1+K\omega_1\overline{\omega}_1}{1-K\omega_2\overline{\omega}_2}A_{\omega_1}. \tag{6.9}$$

The square roots of derivatives are eliminated throughout the formalism thanks to (6.1). This can be compared to the passage to regular gauge in certain classes of instanton solutions.[13] (But in practice, it is sometimes better to stay with λ. This will be illustrated in the following sections.)

Starting with F_{12} of (2.9), defining

$$\mathscr{F}_{12} = \left(\frac{d\omega_1}{dy}\frac{d\omega_2}{dz}\right)^{-1} F_{12}$$
$$= \mathscr{D}_1\mathscr{A}_2 - \mathscr{D}_2\mathscr{A}_1 + i[\mathscr{A}_1,\mathscr{A}_2]$$
$$= F_{\omega_1\omega_2} + \Lambda^2 F_{\overline{\omega}_1\overline{\omega}_2} + \Lambda\left\{\frac{1+K\omega_1\overline{\omega}_1}{1-K\omega_2\overline{\omega}_2}F_{\omega_1\overline{\omega}_1} + \frac{1-K\omega_2\overline{\omega}_2}{1+K\omega_1\overline{\omega}_1}F_{\omega_2\overline{\omega}_2}\right\}, \tag{6.10}$$

the coefficients of $1,\Lambda,\Lambda^2$ equated to zero gives self-duality corresponding to (6.3). The parametrization of the null surface of Sec. IV is now

$$\omega_1 = u,\quad \omega_2 = v,$$
$$\overline{\omega}_1 = \frac{-av+c}{Kbv-d},\quad \overline{\omega}_2 = -\left(\frac{-au+b}{Kcu-d}\right)\quad (Kbc\neq ad),$$
$$\Lambda = (Kbc-ad)/(Kcu-d)(Kbv-d). \tag{6.11}$$

Now

$$d\overline{\omega}_1 = -\frac{(Kbc-ad)}{(Kbv-d)^2}dv,\quad d\overline{\omega}_2 = \frac{(Kbc-ad)}{(Kcu-d)^2}du, \tag{6.12}$$

and

$$(1+K\omega_1\overline{\omega}_1)(1-K\omega_2\overline{\omega}_2)^{-1} = (Kcu-d)(Kbv-d)^{-1}. \tag{6.13}$$

Inserting these in (6.3),

$$ds^2 = 0. \tag{6.14}$$

For $b=0$, $c=0$ the solution is independent of K. [Henceforth always $d=1$ as in (6.2).]

With $\mathscr{D}_1$ and $\mathscr{D}_2$ already fixed by (6.4) and (6.5) one can complete, in the five-dimensional complex space $(\omega_1,\omega_2,\overline{\omega}_1,\overline{\omega}_2,\Lambda)$, the list of coordinates and derivatives such that

$$[X_i,X_j]=0,\quad [\mathscr{D}_i,\mathscr{D}_j]=0,$$
$$[\mathscr{D}_i,X_j]=\delta_{ij}\quad (i,j=1,\ldots,5). \tag{6.15}$$

A possible choice of the X's is

$$X_1=\omega_1,\quad X_2=\omega_2,\quad X_3=a^{-1}b,$$
$$X_4=a^{-1}c,\quad X_5=\ln a. \tag{6.16}$$

Along with $\mathscr{D}_1$ and $\mathscr{D}_2$ the other derivatives can be compactly written on defining

$$f_1=\frac{\Lambda K\omega_1\omega_2-1}{1-K\omega_2\overline{\omega}_2},\quad f_2=\frac{\Lambda K\omega_1\omega_2-1}{1+K\omega_1\overline{\omega}_1}, \tag{6.17}$$

as

$$-a^{-1}\mathscr{D}_3 = K\omega_2\overline{\omega}_1 f_1\,\partial\overline{\omega}_1 + f_2\,\partial\overline{\omega}_2 + K\overline{\omega}_1 f_1 f_2\,\partial_\Lambda,$$
$$a^{-1}\mathscr{D}_4 = f_1\,\partial\overline{\omega}_1 - K\omega_1\overline{\omega}_2 f_2\,\partial\overline{\omega}_2 + K\overline{\omega}_2 f_1 f_2\,\partial_\Lambda, \tag{6.18}$$
$$-\mathscr{D}_5 = f_1(\overline{\omega}_1\,\partial\overline{\omega}_1 + \Lambda\,\partial_\Lambda) + f_2(\overline{\omega}_2\,\partial\overline{\omega}_2 + \Lambda\,\partial_\Lambda) - af_1f_2\,\partial_\Lambda.$$

[The $\mathscr{D}$'s corresponding to other choices such as $X_3=b$, $X_4=c$, $X_5=a$ can easily be obtained from (6.18).] The null surface (6.11) corresponds to

$$dX_3=0,\quad dX_4=0,\quad dX_5=0. \tag{6.19}$$

Evidently,

$$dX_1=0=dX_2, \tag{6.20}$$

again gives, in (6.3), $ds^2=0$. The operator $\mathscr{D}_5$ dilates a, b, and c.

$$\mathscr{D}_5(a,b,c)=(a,b,c). \tag{6.21}$$

For $K=0$, $f_1=f_2=-1$,

$$a=-\Lambda,\quad b=\overline{\omega}_2-\Lambda\omega_1,\quad c=-(\overline{\omega}_1+\Lambda\omega_2). \tag{6.22}$$

Redefining the X's of (6.16), for $K=0$, using simple linear combinations as

$$X_1=\tfrac{1}{2}(\omega_1+\Lambda^{-1}\overline{\omega}_2),\quad X_2=\tfrac{1}{2}(\omega_2-\Lambda^{-1}\overline{\omega}_1),$$
$$X_3=\tfrac{1}{2}(\omega_1-\Lambda^{-1}\overline{\omega}_2),\quad X_4=\tfrac{1}{2}(\omega_2+\Lambda^{-1}\overline{\omega}_1), \tag{6.23}$$
$$X_5=\ln\Lambda,$$

helps to display certain features conveniently. The corresponding derivatives are

$$\mathscr{D}_1=\partial\omega_1+\Lambda\,\partial\overline{\omega}_2,\quad \mathscr{D}_2=\partial\omega_2-\Lambda\,\partial\overline{\omega}_1,$$
$$\mathscr{D}_3=\partial\omega_1-\Lambda\,\partial\overline{\omega}_2,\quad \mathscr{D}_4=\partial\omega_2+\Lambda\,\partial\overline{\omega}_1, \tag{6.24}$$
$$\mathscr{D}_5=\overline{\omega}\,\partial\overline{\omega}_1+\overline{\omega}_2\,\partial\overline{\omega}_2+\Lambda\,\partial_\Lambda.$$

Evidently, like the pair $(\mathscr{D}_1,\mathscr{D}_2)$, the pair $(\mathscr{D}_3,\mathscr{D}_4)$ provides another formulation of self-duality. Like

$$dX_3=dX_4=dX_5=0, \tag{6.25}$$

the set

$$dX_1=dX_2=dX_5=0 \tag{6.26}$$

provides another null surface. Parallel possibilities are evidently obtained on interchanging throughout (y,z) and $(\overline{y},\overline{z})$ or (ω_1,ω_2) and $(\overline{\omega}_1,\overline{\omega}_2)$.

One can envisage further transformations of the spectral parameter. Set, for example,

$$\chi = ((1 + K\omega_1\overline{\omega}_1)/(1 - K\omega_2\overline{\omega}_2))^{\pm 1}\Lambda. \tag{6.27}$$

For the upper sign,

$$\mathscr{D}_1 = \partial\omega_1 + \chi((1 - K\omega_2\overline{\omega}_2)/(1 + K\omega_1\overline{\omega}_1))^2\partial\overline{\omega}_2,$$
$$\mathscr{D}_2 = \partial\omega_2 - \chi\,\partial\overline{\omega}_1 + 2K\,[\overline{\omega}_2/(1 - K\omega_2\overline{\omega}_2) - \chi\omega_1/(1 + K\omega_1\overline{\omega}_1)]\chi\,\partial_\chi. \tag{6.28}$$

For the lower sign, ∂_χ is absent in $\mathscr{D}_2$. Different representations can be helpful in solving for different types of seed solutions to be briefly presented in the following sections.

It is instructive to note how solutions are restricted on imposing symmetries and the corresponding transformations of the spectral parameter. Suppose we require independence with respect to the phase σ given by

$$\omega_2/\overline{\omega}_2 = e^{i2\sigma}. \tag{6.29}$$

Setting $\Lambda = \widetilde{\Lambda}e^{-i\sigma}$,

$$a = e^{-i\sigma}\frac{\widetilde{\Lambda} + K\overline{\omega}_1|\omega_2|}{\widetilde{\Lambda}K\omega_1|\omega_2| - 1}, \quad b = e^{-i\sigma}\frac{\widetilde{\Lambda}\omega_1 - |\omega_2|}{\widetilde{\Lambda}K\omega_1|\omega_2| - 1},$$
$$c = \frac{\widetilde{\Lambda}|\omega_2| + \overline{\omega}_1}{\widetilde{\Lambda}K\omega_1|\omega_2| - 1}. \tag{6.30}$$

Expressing $\mathscr{D}_1$ and $\mathscr{D}_2$ in terms of $\widetilde{\Lambda}$, $|\omega_2|$ and σ one sees that for $\partial_\sigma \approx 0$, in (5.12), one must have

$$X = X(b/a,c). \tag{6.31}$$

If one demands independence also with respect to the phase φ, where

$$\omega_1/\overline{\omega}_1 = e^{i2\varphi}, \tag{6.32}$$
$$\Lambda = \widehat{\Lambda}e^{-i(\sigma+\varphi)}, \tag{6.33}$$

setting only the combination

$$\frac{bc}{a} = \frac{(\widehat{\Lambda}|\omega_1| - |\omega_2|)(\widehat{\Lambda}|\omega_2| + |\omega_1|)}{(\widehat{\Lambda} + K|\omega_1||\omega_2|)(\widehat{\Lambda}K|\omega_1||\omega_2| - 1)} \tag{6.34}$$

now leads to solutions satisfying $\partial\sigma \approx 0$, $\partial\varphi \approx 0$, with $\widehat{\Lambda}$ as the spectral parameter. Such constraints are extremely important in constructing soliton solutions to be described below.

VII. APPLICATIONS TO SOLITON SOLUTIONS

We now apply the general formalism to construction of solitons. Suppressing derivations, one complete set of prescriptions is presented. Known important particular cases[6–11] are shown to emerge from this systematically. We also point out new interesting possibilities. (Background material can be found in Refs. 1–11.)

A. SU(2) gauge fields

Define

$$M = p^{-1/2}\begin{vmatrix} p & 0 \\ q & 1 \end{vmatrix}, \tag{7.1}$$

with p a real function and q, in general, a complex function of $(y,\bar{y},z,\bar{z})$. The ansatz for gauge potentials is

$$A_\mu = (i\,\partial_\mu M)M^{-1} \quad (\mu = y,z),$$
$$A_{\bar{\mu}} = (i\,\partial_{\bar{\mu}}M^{\dagger^{-1}})M^\dagger \quad (\bar{\mu} = \bar{y},\bar{z}) \tag{7.2}$$

Let

$$G = M^\dagger M. \tag{7.3}$$

The self-duality constraints reduce to (with $G_y \equiv \partial_y G$ etc.)

$$g^{y\bar{y}}(G_yG^{-1})_{\bar{y}} + g^{z\bar{z}}(G_zG^{-1})_{\bar{z}} = 0. \tag{7.4}$$

Using D_1, D_2 of (2.11) introduce the linear pair for the 2×2 matrix ψ as

$$D_1\psi = (G_yG^{-1})\psi,$$
$$D_2\psi = (G_zG^{-1})\psi, \tag{7.5}$$

with the condition

$$\psi(\lambda = 0) = G. \tag{7.6}$$

Choose a "seed solution" G_0 and construct the corresponding ψ_0. Prescriptions for a class of ψ_0 will be given later. The ansatz for iteration is (with 2×2 matrices R_n and the unit matrix I);

$$\psi_n = (I + R_n/(\lambda - \mu_n))\psi_{n-1} \quad (n = 1,2,...). \tag{7.7}$$

Here, R_n and μ_n are independent of λ. (The poles can also be introduced simultaneously and additively.)

The poles $(\mu_1,...,\mu_n)$ cannot be chosen arbitrarily. For consistency, the residue of each double pole on the left of (7.5) must vanish. Here, the homogeneous solutions of Sec. V play a fundamental role. It can be shown that if

$$D_jH(y,\bar{y},z,\bar{z},\lambda) = 0 \quad (j = 1,2), \tag{7.8}$$

then μ_m satisfying, for some constant c_m.

$$H(y,\bar{y},z,\bar{z},\mu_m) = c_m, \tag{7.9}$$

provide the right choice of poles. (This can be demonstrated, more generally, for kernels of mappings in arbitrary dimensions.) One need not write and solve separately "pole equations" as is usually done. The solutions of (7.8) not only give the poles but relate them to the geometry of the null surface. From (5.11), define

$$a(\mu) = (\mu\Omega + K\overline{\omega}_1\overline{\omega}_2\overline{\Omega})/(\mu K\omega_1\omega_2\Omega - \overline{\Omega}),$$
$$b(\mu) = (\mu\omega_1\Omega - \overline{\omega}_2\overline{\Omega})/(\mu K\omega_1\omega_2\Omega - \overline{\Omega}), \tag{7.10}$$
$$c(\mu) = (\mu\omega_2\Omega + \overline{\omega}_1\overline{\Omega})/(\mu K\omega_1\omega_2\Omega - \overline{\Omega}).$$

One must have

$$H_n(a(\mu_n),b(\mu_n),c(\mu_n)) = \text{const.} \quad (n = 1,2,...). \tag{7.11}$$

The proper choice of H at each stage is crucial in assuring desirable properties of the solutions. With the poles so chosen, the power of the formalism manifests itself by yielding explicit algebraic solutions of the matrices R. Define the row vectors

$$m_n = M_n(a(\mu_n),b(\mu_n),c(\mu_n))\psi_{n-1}^{-1}(\mu_n) \quad (n = 1,2,...), \tag{7.12}$$

the row vectors M being functions of a, b, and c only and $\psi(\mu) = \psi(\lambda = \mu)$. Proper choices of the vectors M_n is again crucial. Now

$$R_n = (\mu_n + \bar{\mu}_n^{-1})\frac{(G_{n-1}\cdot m_n^\dagger)\otimes m_n}{(m_n\cdot G_{n-1}\cdot m_n^\dagger)} \quad (n = 1,2,...), \tag{7.13}$$

where $G_n = \psi_n(\lambda = 0)$. The iterative structure of G is

$$G_n = G_{n-1} - (1+\mu_n^{-1}\bar{\mu}_n^{-1})\frac{(G_{n-1}\cdot m_n^\dagger)\otimes(m_n\cdot G_{n-1})}{(m_n\cdot G_{n-1}\cdot m_n^\dagger)}. \tag{7.14}$$

Finally, to assure unimodularity, take

$$G_n(\text{phys}) = \left(\prod_{k=1}^{n} \mu_k \bar{\mu}_k\right)^{1/2} G_n. \tag{7.15}$$

From G one obtains p and q and hence A_μ and $A_{\bar{\mu}}$.

B. Seed solutions

Suppose we set

$$G_0 = \text{diag}(e^{\overline{f(z)}+f(z)}, \epsilon e^{-f(z)-\overline{f(z)}}), \tag{7.16}$$

where $f(z)$ is some suitably chosen function of z and [to make (7.15) work] $\epsilon = \pm 1$ for an even and odd number of poles, respectively. This evidently satisfies (7.4). Setting

$$\psi_0 = \text{diag}(e^h, \epsilon e^{-h}), \tag{7.17}$$

from (7.5)

$$D_1 h = 0, \quad D_2 h = \partial_z f. \tag{7.18}$$

Let $f(z) = \mathscr{F}_1(\omega_2(z))$. From (5.10), for $\lambda = 0$,

$$a = K\overline{\omega}_1\overline{\omega}_2, \quad b = \overline{\omega}_2, \quad c = -\overline{\omega}_1, \tag{7.19}$$

so that $\overline{\omega}_2 = b = -a/Kc \quad (K \neq 0)$.

Hence

$$h = \mathscr{F}_1(\omega_2) + \mathscr{F}_2(b, -a/c), \tag{7.20}$$

such that

$$\mathscr{F}_2(\lambda = 0) = \overline{\mathscr{F}}_1 = \bar{f}$$

satisfies (7.18) and also

$$\psi_0(\lambda = 0) = G_0. \tag{7.21}$$

For $K = 0$ one can consider $\mathscr{F}_2(b,a)$. Correct choices of $\mathscr{F}_2$, will be indicated for particular cases to follow. Other choices of G_1 are evidently possible. The foregoing example suffices to illustrate the technique.

C. SU(N) Gauge fields

The same linear pair (the D's) and homogeneous solutions (a,b,c) are good for higher-dimensional gauge groups, such as SU(N). But now G and ψ are $(N\times N)$ matrices and M of (7.1) is now $(N\times N)$ lower triangular. The parametrization of M and G and the extraction of A_μ, $A_{\bar{\mu}}$ are all now more complicated. Without detailed discussion of such topics we briefly indicate some new features that arise.[7,8]

At each step one can now introduce up to $(N-1)$ row vectors

$$m_n^{(l)} = M_n^{(l)}(a(\mu_n), b(\mu_n), c(\mu_n))\psi_{n-1}^{-1}(\mu_n), \tag{7.22}$$
$$l = 1,2,\ldots,l_n \leqslant N-1.$$

The matrix (7.13) is now generalized to

$$R_n = (\mu_n + \bar{\mu}_n^{-1})\{(G_{n-1}\cdot m_n^{(i_n)})\otimes m_n^{(j_n)\dagger}\}(J_n^{-1})_{i_n j_n}, \tag{7.23}$$

where

$$(J_n)_{i_n j_n} = m_n^{(i_n)}\cdot G_{n-1}\cdot m_n^{(j_n)^\dagger}. \tag{7.24}$$

This leads to

$$G_n = G_{n-1} - (1+\mu_n\bar{\mu}_n^{-1})(G_{n-1}\cdot m_n^{(i_n)^\dagger}) \otimes (m_n^{(j_n)}\cdot G_{n-1})(J^{-1})_{i_n j_n}, \tag{7.25}$$

and (7.15) now becomes

$$G_{n_{(\text{phys.})}} = \left(\prod_{k=1}^{n} (\mu_k\bar{\mu}_k)^{l_k/N}\right) G_n. \tag{7.26}$$

For different choices of l_n at each step one has different solutions. This is an extra possibility arising beyond SU(2). For SU(2) at each step $l_n = 1$. One can generalize (7.16) to

$$G_0 = \text{diag}(\epsilon_1 e^{k_1(f+\bar{f})}, \quad \epsilon_2 e^{k_2(f+\bar{f})}, \ldots, \epsilon_N e^{K_N(f+\bar{f})}), \tag{7.27}$$

and set

$$\psi_0 = \text{diag}(\epsilon_1 e^{k_1 h}, \epsilon_2 e^{k_2 h}, \ldots, \epsilon_N e^{k_N h}), \tag{7.28}$$

where the ϵ's are ± 1 and

$$k_1 + k_2 + \cdots + k_N = 0.$$

The prescription (7.20) for h remains the same. Due to the complications mentioned for SU(N) at the beginning, the solutions are best discussed in terms of the "superpotential".[7,8] But details are beyond our scope in this paper. Spherically symmetric SU(N) solutions have been constructed without using linear pairs.[13] But they can also be obtained in this framework. This has been illustrated for SU(3).[7,8]

VIII. PARTICULAR CASES

The results of Sec. VII are now shown to yield systematically interesting particular cases.

Case 1: Hyperbolic coordinates: Let

$$\omega_1 = y, \quad \omega_2 = e^{-2z}, \quad K = 1, \tag{8.1}$$

and

$$y\tan(\vartheta/2)e^{i\varphi}, \quad z = \tfrac{1}{2}(\eta + i\tau). \tag{8.2}$$

Then from (2.1) and (3.6),

$$ds^2 = (\text{C.F.})[d\tau^2 + d\eta^2 + \sinh^2\eta(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2)], \tag{8.3}$$
$$(0 \leqslant \eta < \infty, \quad -\pi \leqslant \tau \leqslant \pi, \quad 0 \leqslant \vartheta \leqslant \pi, \quad 0 \leqslant \varphi < 2\pi).$$

For

$$\text{C.F.} = (\cosh\eta + \cos\tau)^{-2}, \tag{8.4}$$

one has the flat Euclidean space[18] and for

$$\text{C.F.} = (\cosh\eta)^{-2}, \tag{8.5}$$

one has the de Sitter space.

Expressing D_1 and D_2 in terms of $(\tau,\eta,\vartheta,\varphi)$ is straightforward.[8] The convenient combinations of the homogeneous solutions turn out to be[10]

$$B_1(\lambda) = \frac{a}{c} = [(\lambda e^\eta - \bar{y})/(\lambda - \bar{y}e^\eta)]e^{i\tau},$$
$$B_2(\lambda) = b = [(\lambda y e^\eta + 1)/(\lambda y + e^\eta)]e^{i\tau}, \tag{8.6}$$
$$B_3(\lambda) = \frac{a}{b} = (\lambda e^\eta - \bar{y})/(\lambda y e^\eta + 1).$$

For solutions independent of τ and φ only the combination $B_1 B_2^{-1}$ can appear (in the poles and the row vectors M_n)

along with the redefinition $\lambda \to \lambda e^{-i\varphi}$.

Choosing

$$G_0 = \mathrm{diag}(e^{\alpha(z+\bar{z})}, \epsilon e^{-\alpha(z+\bar{z})}), \qquad (8.7)$$

for SU(2),

$$\psi_0 = \mathrm{diag}(e^{\alpha h}, \epsilon e^{-\alpha h}), \qquad (8.8)$$

with

$$h = z - \tfrac{1}{4}\ln(B_1(\lambda)B_2(\lambda)). \qquad (8.9)$$

(The criteria for specific choices of h are hermiticity constraints for G[7,8].) The generalization to SU(N) is direct. This formulation has been used extensively[8,10] to construct instantons.

Case 2. Spherical coordinates: This can be very conveniently and fruitfully treated[8] as a scaling limit ($z \to z/\alpha$; $\alpha \to \infty$) of the previous case showing how monopoles emerge as limits of instanton sequences. But it is instructive to extract it directly from our general formalism. Let

$$\omega_1 = y, \quad \omega_2 = (z-1)/(z+1), \quad K = 1, \qquad (8.10)$$

and

$$y = \tan(\vartheta/2)e^{i\varphi}, \quad z = (r + it), \qquad (8.11)$$

($0 \leqslant r < \infty$, $-\infty < t < \infty$). Then, normalizing the C.F.,

$$ds^2 = [dt^2 + dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta d\varphi^2)], \qquad (8.12)$$

the standard spherical line element.

One can easily express D_1, D_2 in spherical coordinates. Note that from (3.6) and (8.10)

$$\rho = (1 + y\bar{y})/(z + \bar{z}) = (2r\cos^2(\vartheta/2))^{-1}, \qquad (8.13)$$

which is static. This simplicity is lost after (6.1) leading to $\mathscr{D}_1$ and $\mathscr{D}_2$ of (6.4) and (6.5). [This is what we meant in the comment following (6.9). Though (6.1) elucidates the general structure, it is not always advisable to implement it in constructing explicit solutions.] The combinations well adapted for this case are[10]

$$B_1(\lambda) = (a+c)/(a-c) = r[(\lambda + \bar{y})/(\lambda - \bar{y})] + it,$$

$$B_2(\lambda) = (b+1)/(b-1) = r[(\lambda y - 1)/(\lambda y + 1)] + it, \qquad (8.14)$$

$$B_3(\lambda) = (a-c)/(b-1) = (\lambda - \bar{y})/(\lambda y + 1).$$

These can be obtained also as scaling limits of (8.6). Choose, for SU(2),

$$G_0 = \mathrm{diag}(e^r, \epsilon e^{-r}) \quad (r = \tfrac{1}{2}(z + \bar{z})),$$

$$\psi_0 = \mathrm{diag}(e^h, \epsilon e^{-h}), \qquad (8.15)$$

where

$$h = \tfrac{1}{2}z - \tfrac{1}{2}(B_1(\lambda) + B_2(\lambda)). \qquad (8.16)$$

This formulation has been used to construct monopoles[8] and (quasi)periodic solutions.[10,11] The combination $(B_1 - B_2)$ along with $\lambda \to \lambda e^{-i\varphi}$ can be used for static, axially symmetric solutions. The roles of B_1 and B_2 in time-dependent solutions, however, are of particular interest.

Case 3. Accelerated observer: In the Minkowski line element, setting

$$x_0 = \alpha^{-1}e^{\alpha\zeta}\sinh\alpha\sigma,$$

$$x_3 = \alpha^{-1}e^{\alpha\zeta}\cosh\alpha\sigma, \qquad (8.17)$$

one has

$$ds^2 = -dx_0^2 + dx_3^2 + dx_1^2 + dx_2^2$$

$$= e^{2\alpha\zeta}(-d\sigma^2 + d\zeta^2) + dx_1^2 + dx_2^2. \qquad (8.18)$$

An accelerated observer[14] along a trajectory $\zeta = \text{const}$ with a proper acceleration $\alpha e^{-\alpha\zeta}$ has a proper time proportional to σ ($e^{\alpha\zeta}\sigma$). Hence he will see solutions independent of σ as static. To study self-dual gauge fields from the point of view of such an observer,[15] making explicit the role of the observer's proper time, consider the Euclidean section obtained through

$$x_0 \to -ix_0, \quad \sigma \to -i\sigma. \qquad (8.19)$$

Normalizing α to 1,

$$ds^2 = e^{2\zeta}(d\sigma^2 + d\zeta^2) + dx_1^2 + dx_2^2$$

$$= du^2 + u^2\,d\sigma^2 + dv^2 + v^2\,d\varphi^2 \qquad (8.20)$$

$$= e^{z+\bar{z}}\,dz\,d\bar{z} + e^{y+\bar{y}}\,dy\,d\bar{y}, \qquad (8.21)$$

where we have set

$$x_3 + ix_0 = ue^{i\sigma} = e^z, \quad x_1 + ix_2 = ve^{i\varphi} = e^y. \qquad (8.22)$$

This corresponds to (3.6) with

$$\omega_1 = e^y, \quad \omega_2 = e^z, \quad k = 0, \qquad (8.23)$$

with a corresponding choice of C.F. Now

$$a = -\lambda e^{1/2(\bar{y}+\bar{z}-y-z)} = -\lambda e^{-i(\sigma+\varphi)},$$

$$b = -\lambda e^{1/2(\bar{y}+\bar{z}+y-z)} + e^z = (-\lambda v + u)e^{-i\sigma}, \qquad (8.24)$$

$$c = -\lambda e^{1/2(\bar{y}+\bar{z}-y+z)} - e^y = -(\lambda u + v)e^{-i\varphi}.$$

Also with $\rho = e^{1/2(z+\bar{z}-y-\bar{y})} = u/v$,

$$\mathrm{D}_1 = \partial_y + \lambda\rho^{-1}\,\partial_{\bar{z}} + \tfrac{1}{2}(1 - \lambda\rho^{-1})\lambda\,\partial_\lambda, \qquad (8.25)$$

$$D_2 = \partial_z - \lambda\rho\partial_{\bar{y}} + \tfrac{1}{2}(1 + \lambda\rho)\lambda\,\partial_\lambda,$$

for

$$\lambda = 0, \quad a = 0, \quad b = e^z, \quad c = -e^y. \qquad (8.26)$$

Hence, for SU(2), choose

$$G_0 = \mathrm{diag}(e^{\beta(z+\bar{z})}, \epsilon e^{-\beta(z+\bar{z})}), \qquad (8.27)$$

$$\psi_0 = \mathrm{diag}(e^h, \epsilon e^{-h}), \qquad (8.28)$$

where

$$h = \beta z + \mathscr{F}(b,a) \qquad (8.29)$$

with

$$e^{\mathscr{F}(\lambda=0)} = e^{\beta z}. \qquad (8.30)$$

The relation of these coordinates with those of case 1 has been given elsewhere.[9,15] Instanton and monopoles have been studied.

Case 4. Cylindrical coordinates: This can be treated as a scaling limit of case 3, permitting again the exaction of monopoles as limits of instantons.[9] To extract it directly from our formalism set

$$\omega_1 = e^y, \quad \omega_2 = z, \quad k = 0. \qquad (8.31)$$

Suitably choosing the C.F. one has

$$ds^2 = dz\,d\bar{z} + e^{\nu+\bar{\nu}} dy\,d\bar{y}. \tag{8.32}$$

For

$$x_3 + ix_0 = z, \quad e^{\nu} = ve^{i\varphi}, \tag{8.33}$$

one gets the standard form

$$ds^2 = dx_0^2 + dx_3^2 + dv^2 + v^2\,d\varphi^2. \tag{8.34}$$

Now

$$a = -\lambda e^{-i\varphi}, \quad b = -\lambda v + \bar{z}, \quad c = -(\lambda z + v)e^{-i\varphi}. \tag{8.35}$$

The combination

$$c/a + b = -\lambda v + v/\lambda + (z + \bar{z}), \tag{8.36}$$

free from x_0 and φ can be used in the pole equations to construct static axially symmetric monopoles.[6,7]

Case 5. Gauge fields around cosmic strings: If instead of (8.34) one takes[16]

$$ds^2 = dx_0^2 + dx_3^2 + dv^2 + B^2 v^2\,d\varphi^2 \quad (0 < B < 1), \tag{8.37}$$

one gets the Euclidean section ($x_0 \to -ix_0$) of the metric of a straight cosmic string of linear mass density

$$\mu = \tfrac{1}{4}(1 - B). \tag{8.38}$$

It can be cast in the form[17]

$$ds^2 = dz\,d\bar{z} + e^{-4V} dy\,d\bar{y}, \tag{8.39}$$

where

$$V = \mu \ln(y\bar{y}), \quad \text{or } e^{-4V} = (y\bar{y})^{-4\mu}. \tag{8.40}$$

Setting

$$\frac{d\omega_1}{dy} = y^{-4\mu}, \quad \omega_2 = z, \quad k = 0, \tag{8.41}$$

one gets the link with our formalism. Multiple parallel strings are obtained[17] by setting

$$V = \sum_{l=1}^{n} \mu_l \ln\{(y - C_l)(\bar{y} - \bar{C}_l)\}, \tag{8.42}$$

where the mass density and the intersection with the 1-2 plane are given, for the lth string, by μ_l and C_l, respectively. Now

$$\frac{d\omega_1}{dy} = \prod_{l=1}^{n} (y - C_l)^{-4\mu_l}, \quad \omega_2 = z, \quad k = 0. \tag{8.43}$$

A continuous cloud of density can also be envisaged,[17] and one can choose a suitable interior solution. For thin strings (8.40), (8.42) the metric is singular on the string axes.

In applying our formalism to construct self-dual gauge fields in such a background (say that of a single string to start with) the real problem, as always in the linear pair approach, will be to fix the domains of the parameters consistent with desirable regularity and boundary conditions. Topological aspects should then be analyzed carefully. For nonsingular solutions one can consider the exterior of a thick string. Such solutions will be studied elsewhere. One should note, however, that a well-defined self-dual gauge field (the signature being Euclidean) has zero $T_{\mu\nu}$. So it does not disturb the gravitational background. So one has effectively a solution of the total gravitational Yang–Mills system.

Gauge field solutions in various other curved spaces are known.[8] Here we just point out how the special case of parallel strings fits into our formalism. De Sitter space was already mentioned in case 1. More generally, conformally flat Robertson–Walker spaces can be considered in our framework. One point should, however, be noted. Self-dual solutions in four dimensions are formally independent of an overall conformal factor in the metric. But the singularities of the conformal factor, limiting the domain of the coordinates, can have important consequences. It has been shown[8] how, in case 1, a close study of the region $\eta \to \infty$ restricts the parameter α of (8.7) to integer values, for the gauge potentials to be finite and free of branch cuts.

IX. REMARKS

In Ref. 1 the metric tensor depends, from the very beginning, only on *two* coordinates. This is not suitable for our purpose. The solution of Refs. 10 and 11 depend already on three coordinates (t,r,ϑ). So we have started with a formalism where all four coordinates can enter into play on an equal footing. Then we have shown how to impose different types of symmetry restrictions when desired. Thus we have generalized the metric tensor in one direction. Then we have derived the necessary restriction in another direction imposed by the integrability condition of the linear pair. This leads us to locally conformally flat metrics given by (2.1) and (3.6). Even for conformally flat metrics all choices of coordinates cannot be directly implemented in linear pairs. We have given the most general formalism possible. Such a constraint leaves however a rich structure. It includes, for example, all the cases of Sec. VIII. The B's of (8.6) and (8.14) played a basic role in the construction of certain classes of aperiodic, periodic, and quasiperiodic instantons.[10,11] They were, however, discovered in a groping fashion, the homographic forms being found for the first time in this context. Now we have the very satisfactory knowledge that they arise from the invariance properties of our Liouville fields. The fruitfulness of varied use of non-Cartesian coordinates needs no new demonstration (see references to our previous papers). Now we have the geometrical insight of the unified basis relating them all. New possibilities have been opened up.

[1] V. A. Belinskii and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)]; **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].

[2] A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).

[3] L. L. Chau, "General Integrability and Equations of Motion in Physics: A Unifying View," UCD-87-38. This review contains further references.

[4] R. S. Ward, Phys. Lett. A **61**, 81 (1977); Commun. Math. Phys. **79**, 317 (1981), Phys. Lett. B **102**, 136 (1981).

[5] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).

[6] P. Forgacs, Z. Horvath, and L. Palla, in *Monopoles in Quantum Field Theory*, Proceedings of the Monopole Meeting, Trieste, 1981, edited by N. S. Craigie, P. Goddard and W. Nahm (World Scientific, Singapore, 1981).

[7] P. Forgacs, Z. Horvath, and L. Palla, Nucl. Phys. B **221**, 235 (1983), **229**, 77 (1983).

[8] A. Chakrabarti, Fortschr. Phys. **35**, 1 (1987). This review indicates some previous sources.

[9] A. Chakrabarti, J. Math. Phys. **27**, 340 (1986).

[10] A. Chakrabarti, Phys. Rev. D **35**, 696 (1987).

[11] A. Chakrabarti, Phys. Rev. D **38**, 3219 (1988).

[12]M. F. Atiyah, V. G. Drinfeld, N. J. Hitchin, and Yu. I. Manin, Phys. Lett. A **65**, 185 (1978).
[13]A. Chakrabarti and C. Roiesnel, Phys. Rev. D **39**, 2381 (1989).
[14]N. D. Birrel and P. C. W. Davis, *Quantum Fields in Curved Space* (Cambridge U.P., London, 1981). Extensive references to original sources are in Sec. 4.5.
[15]A. Chakrabarti, "Acceleration, Instantons, Dyons and Fermions" (1985, unpublished).
[16]A. Vilenkin, Phys. Rep. **121**, 263 (1985).
[17]P. S. Letelier, Class. Quant. Grav. **4**, L75 (1987).

Parabose algebras and subgraph polynomials

Dirk Graudenz[a)]

II. Institut für Theoretische Physik, Universität Hamburg, Luruper Chaussee 149, D-2000 Hamburg 50, Federal Republic of Germany

(Received 20 November 1989; accepted for publication 28 February 1990)

A graph theoretical method is proposed for the calculation of inner products in the Fock spaces of parabose algebras. For this purpose, a new class of polynomials associated with finite graphs is introduced. The obtained results can be generalized to the parafermi case.

I. INTRODUCTION

There have been suggestions for quantization procedures that lead to particle statistics different from the usual Bose and Fermi types. A particular example that has received much interest is parastatistics.[1–7] The basic commutation relations (CRs) of parabose or parafermi operators are not bilinear, but trilinear. The algebras are classified by a number $p \in \mathbb{N}$, the parabose or parafermi order. Green has given a convenient decomposition of parabose and parafermi operators into components that obey bilinear CRs (cf. Ref. 1 and the subsequent section).

The unusual trilinear CRs are sometimes inconvenient in practical calculations. For the calculation of inner products in the Fock space, one could use the relation between the creation and annihilation operators

$$\begin{aligned} a_k a_l^+ a_m^+ &= a_m^+ a_l^+ a_k + a_m^+ a_k a_l^+ \\ &\quad - a_l^+ a_k a_m^+ + 2\delta_{km} a_l^+ \end{aligned} \tag{1.1}$$

to move the annihilation operators to the right until they act upon the vacuum vector. This method works, but the formulas soon become untractable. Another relation was derived in Ref. 7, Eq. (7.26), where the calculation of inner products was reduced to a study of certain representations of the permutation group. But this result does not provide explicit formulas either.

In this paper we describe a very simple graph theoretical algorithm for the calculation of inner products. In a first step, one has to determine certain polynomials g_σ by a graph theoretical method. The result is then given by

$$\langle\langle i_1 \cdots i_n \mathbb{I} j_1 \cdots j_n \rangle\rangle = \sum_{\sigma \in S_n} g_\sigma(p) \delta_{i, j\circ\sigma}, \tag{1.2}$$

where $\mathbb{I} i_1 \cdots i_n \rangle\rangle$ is an n-particle state in which the modes $i_1 \cdots i_n$ are occupied, p is the parabose order, and S_n is the symmetric group.

The outline is as follows. In Sec. II, we review some basic properties of parabose algebras and fix our notation. Furthermore, we describe a tensor space representation for the creation operators of the Green's decomposition. In Sec. III we introduce a class of polynomials that are related to the subgraphs of an arbitrary finite graph. These subgraph polynomials have some interesting properties. Section IV gives an application of these polynomials to the calculation of certain coefficients that appear in the tensor space representations. Then, we describe our method for the calculation of inner products in the Fock space. Here the subgraph polynomials enter again. Finally, we add a few remarks concerning the parafermi case.

II. PARABOSE ALGEBRAS, THE GREEN'S DECOMPOSITION, AND A TENSOR SPACE REPRESENTATION

In this section, we fix our notation and give a short review of some properties of parabose algebras. Furthermore, we describe the Green's decomposition and the commutation relations of the Green's components. For simplicity, we consider systems with a finite number of degrees of freedom. Our considerations apply equally well to the case of parafermi algebras. Since the changes are mainly signs that have to be removed from the formulas, we concentrate on the parabose case and mention the necessary changes in Sec. IV.

Let $(a_i^+, a_i)_{i=1,\dots,R}$ be a set of parabose creation and annihilation operators satisfying the following trilinear commutation relations:[1,5]

$$[a_k, \{a_l, a_m\}] = 0, \tag{2.1}$$

$$[a_k, \{a_l^+, a_m^+\}] = 2\delta_{km} a_l^+ + 2\delta_{kl} a_m^+, \tag{2.2}$$

$$[a_k, \{a_l^+, a_m\}] = 2\delta_{kl} a_m. \tag{2.3}$$

Denote the algebra that is spanned by the operators a_i^+ and a_i by $\mathfrak{U}$. Here, $\mathfrak{U}$ acts on a Fock space $\mathfrak{F}_{\mathfrak{U}}$ with vacuum $\mathbb{I}\Omega\rangle\rangle$ that is annihilated by the a_i. The Fock space is spanned by vectors of the form

$$\begin{aligned} \mathbb{I} i\rangle\rangle := \mathbb{I} i_1 \cdots i_n\rangle\rangle := a_{i_1}^+ \cdots a_{i_n}^+ \mathbb{I}\Omega\rangle\rangle, \\ i: \{1,\dots,n\} \to \{1,\dots,R\}. \end{aligned} \tag{2.4}$$

If $p \in \mathbb{N}$, the inner product $\langle\langle \cdot \mathbb{I} \cdot \rangle\rangle$ is positive definite.[5]

A convenient representation for $\mathfrak{U}$ is given by the Green's decomposition[1,5]

$$a_i := \sum_{\alpha=1}^{p} a_{i\alpha}, \quad a_i^+ := \sum_{\alpha=1}^{p} a_{i\alpha}^+, \tag{2.5}$$

where the Green's components obey the anomalous CRs

$$\begin{aligned} &[a_{i\alpha}, a_{j\alpha}] = 0, \quad [a_{i\alpha}, a_{j\alpha}^+] = \delta_{ij}, \\ &\{a_{i\alpha}, a_{j\beta}\} = 0, \quad \{a_{i\alpha}, a_{j\beta}^+\} = 0, \quad \text{for } \alpha \neq \beta. \end{aligned} \tag{2.6}$$

If we define

[a)] Work supported by Studienstiftung des deutschen Volkes.

$$\epsilon_{AB} := \begin{cases} 1, & \text{if } A = B, \\ -1, & \text{if } A \neq B, \end{cases} \quad \delta_{AB} := \begin{cases} 1, & \text{if } A = B, \\ 0, & \text{if } A \neq B, \end{cases} \tag{2.7}$$

the CRs of the Green's components read

$$a_{i\alpha} a_{j\beta} = \epsilon_{\alpha\beta} a_{j\beta} a_{i\alpha},$$
$$a_{i\alpha} a_{j\beta}^{+} = \epsilon_{\alpha\beta} a_{j\beta}^{+} a_{i\alpha} + \delta_{i\alpha j\beta} \mathbf{1}. \tag{2.8}$$

Let $\mathfrak{B}$ be the algebra that is spanned by the $a_{i\alpha}^{+}$ and $a_{i\alpha}$. $\mathfrak{B}$ acts on a Fock space $\mathfrak{F}_{\mathfrak{B}}$ with vacuum $\mathbf{I}\Omega\rangle\rangle$. Here $\mathfrak{F}_{\mathfrak{B}}$ is spanned by the vectors

$$\mathbf{I} i^{\alpha}\rangle\rangle := \mathbf{I} i_1^{\alpha_1},\ldots,i_n^{\alpha_n}\rangle\rangle := a_{i_1\alpha_1}^{+} \cdots a_{i_n\alpha_n}^{+} \mathbf{I}\Omega\rangle\rangle. \tag{2.9}$$

Obviously, $\mathfrak{U} \subset \mathfrak{B}$ and $\mathfrak{F}_{\mathfrak{U}} \subset \mathfrak{F}_{\mathfrak{B}}$.

Finally, we define the subalgebras $\mathfrak{U}^{+} \subset \mathfrak{U}$ and $\mathfrak{B}^{+} \subset \mathfrak{B}$ that are spanned by the creation operators $\{a_i^{+}\}$ and $\{a_{i\alpha}^{+}\}$, respectively.

Now, we give explicit representations for the subalgebras $\mathfrak{U}^{+}$ and $\mathfrak{B}^{+}$ on a tensor space. Let T be an R-dimensional $\mathbb{C}$-vector space with basis $(e_i)_{i \in \{1,\ldots,R\}}$, and

$$T^{p} := \bigoplus_{i=1}^{p} T \tag{2.10}$$

the vector space with basis $(e_{i\alpha})_{i \in \{1,\ldots,R\},\ \alpha \in \{1,\ldots,p\}}$.

Furthermore, define the tensor spaces

$$T^{(R)} := \bigoplus_{n \in \mathbb{N}} \bigotimes_{i=1}^{n} T, \quad T^{(pR)} := \bigoplus_{n \in \mathbb{N}} \bigotimes_{i=1}^{n} T^{p} \tag{2.11}$$

with basis

$$|i\rangle := |i_1 \cdots i_n\rangle := e_{i_1} \otimes \cdots \otimes e_{i_n},$$
$$|\Omega\rangle := 1 \in \bigotimes_{i=1}^{0} T := \mathbb{C},$$
$$\rceil i^{\alpha}\rangle := \rceil i_1^{\alpha_1} \cdots i_n^{\alpha_n}\rangle := e_{i_1\alpha_1} \otimes \cdots \otimes e_{i_n\alpha_n}, \tag{2.12}$$
$$\rceil\Omega\rangle := 1 \in \bigotimes_{i=1}^{0} T^{p} := \mathbb{C},$$

respectively. For these tensor spaces, we assume the canonical inner product

$$\langle i | j\rangle := \delta_{ij} \quad \text{and} \quad \langle i^{\alpha} \rceil j^{\beta}\rangle := \delta_{i\alpha j\beta}. \tag{2.13}$$

We define a representation of $\mathfrak{B}^{+}$ on $T^{(pR)}$ as in Ref. 6 by

$$a_{i\alpha}^{+} \rceil\Omega\rangle := \rceil i^{\alpha}\rangle, \tag{2.14}$$
$$a_{i\alpha}^{+} \rceil j_1^{\beta_1} \cdots j_n^{\beta_n}\rangle := \sum_{h=0}^{n} \left(\prod_{l=1}^{h} \epsilon_{\alpha\beta_l} \right) \times | j_1^{\beta_1} \cdots j_h^{\beta_h} i^{\alpha} j_{h+1}^{\beta_{h+1}} \cdots j_n^{\beta_n}\rangle.$$

It is a straightforward calculation to check the CRs

$$a_{i\alpha}^{+} a_{j\beta}^{+} = \epsilon_{\alpha\beta} a_{j\beta}^{+} a_{i\alpha}^{+}. \tag{2.15}$$

A general formula for tensors of this kind is given by

$$a_{i_1\alpha_1}^{+} \cdots a_{i_n\alpha_n}^{+} \rceil\Omega\rangle = \sum_{\sigma \in S_n} v(\sigma,\alpha) \rceil i^{\alpha} \circ \sigma\rangle, \tag{2.16}$$

where

$$(i^{\alpha} \circ \sigma)_k := i_{\sigma(k)}^{\alpha_{\sigma(k)}}. \tag{2.17}$$

Here, S_n is the permutation group of the set $\{1,\ldots,n\}$, $v(\sigma,\alpha)$ is a sign ± 1, $v(1,\alpha) = 1$, and if σ' is a permutation that differs from σ by a transposition of the kth and $(k+1)$th element, then

$$v(\sigma',\alpha) = \epsilon_{\alpha_k \alpha_{k+1}} v(\sigma,\alpha), \tag{2.18}$$

that is, the exchange of neighboring indices i and j with $\alpha \neq \beta$ generates a minus sign.

We obtain a representation of $\mathfrak{U}^{+}$ on $T^{(pR)}$ if we sum over the parabose index. For $\rceil\Psi\rangle \in T^{(pR)}$, define

$$a_i^{+} \rceil\Psi\rangle := \sum_{\alpha=1}^{p} a_{i\alpha}^{+} \rceil\Psi\rangle. \tag{2.19}$$

We note that we do not have a representation for the $a_{i\alpha}$ on $T^{(pR)}$. If we define $b_{i\alpha}$ to be the adjoint operator of $a_{i\alpha}^{+}$ with respect to the inner product $\langle \cdot \rceil \cdot \rangle$, the $a_{i\alpha}^{+}$ and $b_{i\alpha}$, in general, do *not* satisfy the relation

$$b_{i\alpha} a_{j\beta}^{+} = \epsilon_{\alpha\beta} a_{j\beta}^{+} b_{i\alpha} + \delta_{i\alpha j\beta}. \tag{2.20}$$

We define a projection

$$\Pi: T^{(pR)} \to T^{(R)}, \quad \rceil i_1^{\alpha_1} \cdots i_n^{\alpha_n}\rangle \overset{\Pi}{\mapsto} |i_1 \cdots i_n\rangle, \tag{2.21}$$

such that Π "forgets" parabose indices, and a map

$$\varphi: \mathfrak{F}_{\mathfrak{B}} \to T^{(pR)}, \quad a_{i_1\alpha_1}^{+} \cdots a_{i_n\alpha_n}^{+} \mathbf{I}\Omega\rangle\rangle \overset{\varphi}{\mapsto} a_{i_1\alpha_1}^{+} \cdots a_{i_n\alpha_n}^{+} \rceil\Omega\rangle. \tag{2.22}$$

Since $\mathfrak{F}_{\mathfrak{U}} \subset \mathfrak{F}_{\mathfrak{B}}$, we may regard φ as a map $\varphi: \mathfrak{F}_{\mathfrak{U}} \to T^{(pR)}$, too. If ι is the injection of $\mathfrak{F}_{\mathfrak{U}}$ into $\mathfrak{F}_{\mathfrak{B}}$, we may summarize the spaces and maps in the following diagram:

$$\mathfrak{F}_{\mathfrak{U}} \xrightarrow{\iota} \mathfrak{F}_{\mathfrak{B}} \xrightarrow{\varphi} T^{(pR)} \xrightarrow{\Pi} T^{(R)} \quad \text{with states}$$
$$\{\mathbf{I}\cdot\rangle\rangle\} \quad \{\mathbf{I}\cdot\rangle\rangle\} \quad \{\rceil\cdot\rangle\} \quad \{|\cdot\rangle\}. \tag{2.23}$$

It is obvious that Wick's theorem can be generalized to the case of the Green's decomposition of parabose operators. If it is assumed that a certain set of operators has only $\mathbb{C}$-number commutators and every operator A may be decomposed in a creation part $A^{(+)}$ and an annihilation part $A^{(-)}$, then a normal product :...: may be defined such that the "+" components stand to the left of the "−" components. For Bose operators there is no sign rule, and for Fermi operators an additional minus sign has to be introduced for every exchange of two Fermi operators. In the parabose case, the sign rule is slightly more complicated since here an additional minus sign is needed only if two operators with different parabose indices are exchanged. We conclude that the vacuum expectation value of an odd number of such operators vanishes and that for an even number of operators the following formula is valid:

$$\langle\langle \Omega \mathbf{I} A_{1\alpha_1} \cdots A_{2n\alpha_{2n}} \mathbf{I}\Omega\rangle\rangle = \sum_{\text{pairings}} \eta_{ij\alpha} \underbrace{A_{i_1\alpha_{i_1}} A_{j_1\alpha_{j_1}}} \cdots \underbrace{A_{i_n\alpha_{i_n}} A_{j_n\alpha_{j_n}}}, \tag{2.24}$$

where

$$\underbrace{A_i A_j} = \langle\langle \Omega \mathbf{I} A_i A_j \mathbf{I}\Omega\rangle\rangle, \quad \eta_{ij\alpha} = v(\sigma,\alpha),$$

and

$$\sigma := \begin{pmatrix} 1 & 2 & \cdots & 2n-1 & 2n \\ i_1 & j_1 & \cdots & i_n & j_n \end{pmatrix}. \tag{2.25}$$

The only nonvanishing contraction of the $a_{i\alpha}$ and $a_{i\alpha}^{+}$ is

$$\underbrace{a_{i\alpha} a_{j\beta}^{+}} = \langle\langle \Omega \mathbf{I} a_{i\alpha} a_{j\beta}^{+} \mathbf{I}\Omega\rangle\rangle = \langle\langle i^{\alpha} \mathbf{I} j^{\beta}\rangle\rangle = \delta_{i\alpha j\beta}. \tag{2.26}$$

The maps ι, φ, and Π are linear, but φ and Π are, in general, not isometries. We want to find out the metrical relations between the various spaces. For this purpose, we now develop the tools for the calculation of the map $\Pi\circ\varphi\circ\iota$.

III. SUBGRAPH POLYNOMIALS

This section is a short excursion into the realm of graph theory. It will provide us with a class of polynomials that arises if the properties of the maps ι, φ, and Π in (2.23) are studied. But these polynomials are perhaps also interesting from a graph theoretical point of view.

We assign to each graph G a polynomial $f_G(p,q)\in K[p,q]$, where K is an arbitrary field. To be definite, we will work with $K=\mathbb{C}$. Here, $f_G(p,q)$ will be called a *subgraph polynomial.*

An undirected graph $G=(V(G), E(G))$ is given by a finite set $V(G)$ of vertices and a set $E(G) \subset \{\{a,b\}|a,b\in V(G)\}$ of edges.[8]

We define $\sharp G := \sharp E(G)$ to be the number of edges of G and $\flat G$ to be the number of components of G, where a vertex $a\in V(G)$ that is not an element of any edge of G is counted as a single component.

A graph H is called a *subgraph* of G if $V(H)=V(G)$ and $E(H)\subset E(G)$. We write $H\leqslant G$. Any graph G has $2^{\sharp G}$ subgraphs.

We define

$$f_G(p,q) := \sum_{H\leqslant G} p^{\flat H} q^{\sharp H}. \tag{3.1}$$

Example: The subgraph polynomial of the graph defined in Fig. 1 is

$$f_G(p,q) = p^4 + 4qp^3 + (q^3+6q^2)p^2 + (q^4+3q^3)p, \tag{3.2}$$

as the interested reader may check by drawing all subgraphs and applying definition (3.1).

Now we state some properties of subgraph polynomials that simplify their calculation.

A. Factorization property

A graph G *splits* into G_1 and G_2, $G=G_1\dot\cup G_2$, if

$$V(G)=V(G_1)\dot\cup V(G_2), \quad E(G)=E(G_1)\dot\cup E(G_2), \tag{3.3}$$

where "$\dot\cup$" denotes the disjoint union. Then

$$\sharp G = \sharp G_1 + \sharp G_2, \quad \flat G = \flat G_1 + \flat G_2. \tag{3.4}$$

Since

$$\sum_{H\leqslant G_1\dot\cup G_2} f(H) = \sum_{\substack{H_1\leqslant G_1\\ H_2\leqslant G_2}} f(H_1\dot\cup H_2), \tag{3.5}$$

we obtain

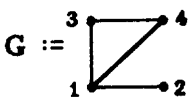


FIG. 1. A connected and one-edge reducible graph.

$$f_{G_1\dot\cup G_2}(p,q) = \sum_{\substack{H_1\leqslant G_1\\ H_2\leqslant G_2}} p^{\flat H_1+\flat H_2} q^{\sharp H_1+\sharp H_2} = f_{G_1}(p,q)\cdot f_{G_2}(p,q). \tag{3.6}$$

B. One-edge reducible graphs

A graph G is called one-edge reducible, if the number of components increases (by one) if an edge $k=\{a,b\}\in E(G)$, say, is removed from G. If $k\in E(G)$ and $G'=(V(G),E(G)\setminus\{k\})$, we write $G=G'\oplus k$ and say that G is k reducible if G is reduced if k is removed. Then

$$\sharp(G'\oplus k) = \sharp G' + 1, \quad \flat(G'\oplus k) = \flat G' - 1. \tag{3.7}$$

The set of subgraphs $H\leqslant G'\oplus k$ is divided into two disjoint classes: Those subgraphs that contain k and those that do not. Therefore,

$$\sum_{H\leqslant G'\oplus k} f(H) = \sum_{H\leqslant G'} f(H) + \sum_{H\leqslant G'} f(H\oplus k). \tag{3.8}$$

If $H\leqslant G'$, k is not an edge of G' and $G'\oplus k$ is k reducible, then

$$\sharp(H\oplus k) = \sharp H + 1, \quad \flat(H\oplus k) = \flat H - 1. \tag{3.9}$$

Therefore

$$\begin{aligned} f_{G'\oplus k}(p,q) &= \sum_{H\leqslant G'} p^{\flat H} q^{\sharp H} + \sum_{H\leqslant G'} p^{\flat(H\oplus k)} q^{\sharp(H\oplus k)} \\ &= f_{G'}(p,q) + \sum_{H\leqslant G'} p^{\flat H-1} q^{\sharp H+1} \\ &= (1+q/p) f_{G'}(p,q) \\ &\qquad (\text{if } G'\oplus k \text{ is } k \text{ reducible}). \end{aligned} \tag{3.10}$$

We note that $f_G(p,-p)=0$ if G is one-edge reducible.

C. Isomorphic graphs

Two graphs are called *isomorphic* ($G_1\approx G_2$) if they differ only in the numbering of their vertices. Then,

$$f_{G_1}(p,q) = f_{G_2}(p,q). \tag{3.11}$$

A general strategy for the calculation of subgraph polynomials is to remove all edges k from a graph G for which G is k reducible. If the number of these edges is c, this contributes a factor of $(1+q/p)^c$. We are then left with a graph $G'=G_1\dot\cup\ldots\dot\cup G_d$ that splits into d irreducible and connected graphs G_i. Their subgraph polynomials will be easier to calculate compared with $f_G(p,q)$. Then,

$$f_G(p,q) = \left(1+\frac{q}{p}\right)^c \prod_{i=1}^{d} f_{G_i}(p,q). \tag{3.12}$$

A graph G may have many isomorphic subgraphs. Then the third of the above mentioned properties is helpful.

We apply (3.12) to the graph G defined in Fig. 1. Here, G is $\{1,2\}$ reducible. Therefore

$$G = G'\oplus k, \quad G' = \nabla\,. = (\nabla)\dot\cup(\cdot), \tag{3.13}$$

so

$$f_G(p,q) = (1+q/p) f_{\nabla}(p,q) f_{\cdot}(p,q). \tag{3.14}$$

Now

$$f_{\nabla}(p,q) = (p^2 + 3qp + q^3 + 3q^2)p, \quad f_{.}(p,q) = p. \tag{3.15}$$

Therefore,

$$f_G(p,q) = (1 + q/p)\cdot(p^2 + 3qp + q^3 + 3q^2)p\cdot p, \tag{3.16}$$

which is the result (3.2) already mentioned above.

Now we describe an application of subgraph polynomials. Let $V(G) = \{1,...,n\}$ and

$$E_{\alpha\beta} := -(q\delta_{\alpha\beta} + 1). \tag{3.17}$$

For $\alpha: V(G) \to \{1,...,p\}$ and a graph $G = (V(G),E(G))$, let

$$v(G,\alpha) := \prod_{\{a,b\}\in E(G)} E_{\alpha_a\alpha_b} \tag{3.18}$$

$$= (-1)^{\#G} \prod_{\{a,b\}\in E(G)} (q\delta_{\alpha_a\alpha_b} + 1) \tag{3.19}$$

and

$$g_G(p,q) := \sum_{\alpha_1=1}^{p} \cdots \sum_{\alpha_n=1}^{p} v(G,\alpha). \tag{3.20}$$

We want to calculate $g_G(p,q)$. First, we multiply all factors $(q\delta_{\alpha\beta} + 1)$ in (3.19) resulting in a sum of $2^{\#G}$ terms. Each of these terms is related to a subgraph $H \lessdot G$ in the sense that a factor $q\delta_{\alpha\beta}$ is represented by an edge in H. So,

$$v(G,\alpha) = (-1)^{\#G} \sum_{H\lessdot G} q^{\#H} \prod_{\{a,b\}\in E(H)} \delta_{\alpha_a\alpha_b}. \tag{3.21}$$

Now the sums in (3.20) may be calculated:

$$g_G(p,q) = (-1)^{\#G} \sum_{\alpha_1=1}^{p} \cdots \sum_{\alpha_n=1}^{p} \sum_{H\lessdot G} q^{\#H}$$

$$\times \prod_{\{a,b\}\in E(H)} \delta_{\alpha_a\alpha_b}. \tag{3.22}$$

For fixed H the sum over the α's of the term

$$\prod_{\{a,b\}\in E(H)} \delta_{\alpha_a\alpha_b} \tag{3.23}$$

gives a factor $p^{\flat H}$ since those α_a that belong to a certain component of H are constrained by the Kronecker δ's to the same value. Finally, we arrive at

$$g_G(p,q) = (-1)^{\#G} \sum_{H\lessdot G} p^{\flat H} q^{\#H} = (-1)^{\#G} f_G(p,q). \tag{3.24}$$

In the sequel, we need a special class of graphs that are related to the symmetric group. Let $\sigma\in S_n$ be a permutation and

$$F_\sigma := \{(i,j) \mid i > j \wedge \sigma^{-1}(i) < \sigma^{-1}(j)\} \tag{3.25}$$

the set of inversions of σ^{-1}. Then, with each σ, we may associate a graph G_σ with

$$V(G_\sigma) = \{1,...,n\}, \quad E(G_\sigma) = \{\{i,j\} \mid (i,j)\in F_\sigma\}. \tag{3.26}$$

For example, if

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}, \quad \sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}, \tag{3.27}$$

then

$$F_\sigma = \{(2,1),(3,1),(4,1),(4,3)\} \tag{3.28}$$

and G_σ is the graph from Fig. 1.

IV. APPLICATION OF SUBGRAPH POLYNOMIALS TO THE GREEN'S DECOMPOSITION

Now we return to the tensor space representations. We calculate the map $\Pi\circ\varphi\circ\iota$ if the result is expanded in terms of canonical basis vectors in $T^{(R)}$. Then, we use this expansion for the calculation of inner products in the Fock space $\mathfrak{F}_{\mathfrak{U}}$.

By using (2.16) and (2.19) we obtain

$$\prod a_{i_1}^+ \cdots a_{i_n}^+ \rrbracket\Omega\rangle = \prod \sum_{\alpha_1=1}^{p} \cdots \sum_{\alpha_n=1}^{p} \sum_{\sigma\in S_n} v(\sigma,\alpha) \rrbracket i^{\alpha}\circ\sigma\rangle \tag{4.1}$$

$$= \sum_{\sigma\in S_n} g_\sigma(p) | i\circ\sigma\rangle, \tag{4.2}$$

where we have defined

$$g_\sigma(p) := \sum_{\alpha_1=1}^{p} \cdots \sum_{\alpha_n=1}^{p} v(\sigma,\alpha). \tag{4.3}$$

Note that we may obtain $v(\sigma,\alpha)$ in the following way. We start with the sequence $\alpha_1...\alpha_n$ and move these symbols around until we reach the sequence $\alpha_{\sigma(1)}...\alpha_{\sigma(n)}$. For each exchange of two different indices, we change sign, starting with 1, and finally obtain $v(\sigma,\alpha)$. Now the symbol α_i stands in the $\sigma^{-1}(i)$th position in the final sequence. Two symbols α_i and α_j have to be exchanged, if $i > j$ and position $(i) <$ position (j), that is $\sigma^{-1}(i) < \sigma^{-1}(j)$, so $(i,j)\in F_\sigma$. This exchange results in a factor $\epsilon_{\alpha(i)\alpha(j)}$. So, we get the general formula

$$v(\sigma,\alpha) = \prod_{(i,j)\in F_\sigma} \epsilon_{\alpha_i\alpha_j} = \prod_{\{a,b\}\in E(G_\sigma)} \epsilon_{\alpha_a\alpha_b}. \tag{4.4}$$

Now, we use (3.17), where we set $q = -2$, such that $E_{\alpha\beta} = \epsilon_{\alpha\beta}$. Then, (3.18) and (3.24) imply

$$\begin{aligned} v(\sigma,\alpha) &= v(G_\sigma,\alpha), \\ g_\sigma(p) &= g_{G_\sigma}(p) = (-1)^{\#G_\sigma} f_{G_\sigma}(p,-2) \\ &= \operatorname{sgn}(\sigma) f_{G_\sigma}(p,-2). \end{aligned} \tag{4.5}$$

A simple formulation is due to Kastening:[9] Given the permutation σ, multiply the terms $(2\delta_{\alpha(i)\alpha(j)} - 1)$ where each factor stands for an inversion of σ^{-1}. Then sum over the parabose indices to obtain g_σ.

Now that we know the coefficients of the expansion (4.2), we can calculate inner products in the Fock space. Applying Wick's theorem, we obtain

$$\begin{aligned} \langle\langle i^{\alpha} \| j^{\beta} \rangle\rangle &:= \langle\langle i_1{}^{\alpha_1} \cdots i_n{}^{\alpha_n} \| j_1{}^{\beta_1} \cdots j_n{}^{\beta_n} \rangle\rangle \\ &= \langle\langle \Omega \| a_{i_n\alpha_n} \cdots a_{i_1\alpha_1} a_{j_1\beta_1}^+ \cdots a_{j_n\beta_n}^+ \| \Omega \rangle\rangle \\ &= \sum_{\sigma\in S_n} v(\sigma,\beta) \underbrace{a_{i_1\alpha_1} a_{j_{\sigma(1)}\beta_{\sigma(1)}}^+}_{} \cdots \underbrace{a_{i_n\alpha_n} a_{j_{\sigma(n)}\beta_{\sigma(n)}}^+}_{} \\ &= \sum_{\sigma\in S_n} v(\sigma,\beta) \langle i^{\alpha} \rrbracket j^{\beta}\circ\sigma\rangle \quad \text{by (2.13)} \\ &= \langle i^{\alpha} \rrbracket a_{j_1\beta_1}^+ \cdots a_{j_n\beta_n}^+ \rrbracket \Omega\rangle \quad \text{by (2.16)} \\ &= \langle i^{\alpha} \rrbracket \varphi(\| j^{\beta}\rangle\rangle) \end{aligned} \tag{4.6}$$

and

$$\langle\langle i_1^{\alpha_1}\cdots i_m^{\alpha_m} \rrbracket j_1^{\beta_1}\cdots j_n^{\beta_n}\rangle\rangle = 0 \quad \text{if } m \neq n. \tag{4.7}$$

By linearity, we obtain for arbitrary $\rrbracket\Psi\rangle\rangle \in \mathfrak{F}_{\mathfrak{P}}$,

$$\langle\langle i_1^{\alpha_1}\cdots i_n^{\alpha_n} \rrbracket \Psi\rangle\rangle = \langle i_1^{\alpha_1}\cdots i_n^{\alpha_n} \rrbracket \varphi \rrbracket \Psi\rangle\rangle. \tag{4.8}$$

A short calculation yields for all $\rrbracket\Psi\rangle \in T^{(pR)}$,

$$\sum_{\alpha_1=1}^{p} \cdots \sum_{\alpha_n=1}^{p} \langle i^{\alpha} \rrbracket \Psi\rangle = \langle i | \Pi \rrbracket \Psi\rangle. \tag{4.9}$$

The inner product in $\mathfrak{F}_{\mathrm{II}}$ is calculated in the following way:

$$\begin{aligned}\langle\langle i_1\cdots i_n \rrbracket \Psi\rangle\rangle &= \sum_{\alpha} \langle\langle i^{\alpha} \rrbracket \iota \rrbracket \Psi\rangle\rangle \\ &= \sum_{\alpha} \langle i^{\alpha} \rrbracket \varphi\circ\iota \rrbracket \Psi\rangle\rangle \quad \text{by (4.8)} \\ &= \langle i | \Pi\circ\varphi\circ\iota \rrbracket \Psi\rangle\rangle \quad \text{by (4.9)},\end{aligned} \tag{4.10}$$

such that

$$\begin{aligned}\langle\langle i_1\cdots i_n \rrbracket j_1\cdots j_n\rangle\rangle &= \langle i_1\cdots i_n | \Pi\circ\varphi\circ\iota \rrbracket j_1\cdots j_n\rangle\rangle \quad \text{by (4.10)} \\ &= \langle i_1\cdots i_n | \sum_{\sigma\in S_n} g_\sigma(p) | j\circ\sigma\rangle \quad \text{by (4.1)} \\ &= \sum_{\sigma\in S_n} g_\sigma(p)\,\delta_{i,j\circ\sigma}.\end{aligned} \tag{4.11}$$

Equation (4.10) shows how the metrical properties of the tensor spaces are related.

Finally, we prove that $\Pi\circ\varphi\circ\iota$ is an injective map on $\mathfrak{F}_{\mathrm{II}}$. We calculate the kernel of this map. Let $\rrbracket\Psi\rangle\rangle \in \mathfrak{F}_{\mathrm{II}}$ with $\Pi\circ\varphi\circ\iota \rrbracket \Psi\rangle\rangle = 0$. For every $\rrbracket i_1\cdots i_n\rangle\rangle$, we obtain using (4.10)

$$\langle\langle i_1\cdots i_n \rrbracket \Psi\rangle\rangle = \langle i_1\cdots i_n | \Pi\circ\varphi\circ\iota \rrbracket \Psi\rangle\rangle = \langle i_1\cdots i_n | 0\rangle = 0. \tag{4.12}$$

Since the $\rrbracket i_1 \ldots i_n\rangle\rangle$ span $\mathfrak{F}_{\mathrm{II}}$ and $\langle\langle \cdot \rrbracket \cdot\rangle\rangle$ is not degenerate, we conclude that $\rrbracket\Psi\rangle\rangle = 0$. Therefore, $\Pi\circ\varphi\circ\iota$ is injective.

We conclude with a remark concerning the parafermi case. If ϵ_{AB} is defined as

$$\epsilon_{AB} := \begin{cases} -1, & \text{if } A = B, \\ 1, & \text{if } A \neq B, \end{cases} \tag{4.13}$$

in contrast to (2.7), the formulas of Sec. II hold for the parafermi case. Then, $g_\sigma(p)$ in (4.5) is changed to

$$g_\sigma(p) = f_{G_\sigma}(p, -2), \tag{4.14}$$

and the results (4.10), (4.11), and (4.12) still hold.

V. CONCLUSIONS

We have derived a graph theoretical method for the calculation of inner products in the Fock space of parabose and parafermi algebras. Essential for this method is the introduction of subgraph polynomials that are associated with permutations.

We think that there are many additional interesting questions. Since the subgraph polynomials are intimately related to coefficients that arise in the representation of an algebra, they should possess a rich structure of algebraic properties. Furthermore, they are related to graph theory. An interesting question would be whether these polynomials fix a graph up to isomorphisms. A generalization of the contents of this paper to an infinite number of degrees of freedom (field theory) should be straightforward.

ACKNOWLEDGMENTS

I would like to thank C. F. v. Weizsäcker, who brought some of these problems to my attention and Th. Görnitz, B. Kastening, and D. Michelsen for interesting and stimulating discussions. I am grateful to D. Buchholz and G. Kramer for reading the manuscript.

[1] H. S. Green, Phys. Rev. **90**, 270 (1953).
[2] D. V. Volkov, Sov. Phys. JETP **36**, 1107 (1959).
[3] O. W. Greenberg and A. M. L. Messiah, Phys. Rev. **138**, 1155 (1965).
[4] K. Drühl, R. Haag, and J. E. Roberts, Comm. Math. Phys. **18**, 204 (1970).
[5] Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (Springer-Verlag, Berlin, 1982), and references therein.
[6] C. F. v. Weizsäcker, *Aufbau der Physik* (DTV-Verlag, München, 1988).
[7] S. Doplicher, R. Haag, and J. E. Roberts, Comm. Math. Phys. **35**, 49 (1974).
[8] M. Aigner, *Combinatorial Theory* (Springer-Verlag, Berlin, 1979).
[9] B. Kastening (private communication).

Hannay's angles outside the adiabatic range: Coupled linear oscillators near resonance

Jair Koiller
Laboratorio Nacional de Computação Cientifica, Rua Lauro Muller, 455 Rio de Janeiro, Brazil 22290 and Instituto de Matemática da UFRJ, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, Brazil 21944

(Received 30 May 1989; accepted for publication 21 February 1990)

The simple harmonic 2-degree of freedom (dof) oscillator with the potential $V=\frac{1}{2}[(1+r\sin\epsilon t)q_1^2+(1-r\sin\epsilon t)q_2^2+2r\cos\epsilon t q_1 q_2]$ is considered. It is shown that a coupling parameter $k\sim r/\epsilon$ determines the behavior of the system after a loop $0\leqslant t\leqslant 2\pi/\epsilon$ in parameter space. Somewhat unexpected mode conversion and phase corrections occur. An optical model is outlined.

I. INTRODUCTION

Consider the slowly varying $2\frac{1}{2}$-degree of freedom (dof) oscillator:

$$T=\tfrac{1}{2}(p_1^2+p_2^2),\quad V=\tfrac{1}{2}(aq_1^2+cq_2^2+2bq_1q_2),$$
$$a=1+r\sin(\epsilon t),\quad c=1-r\sin(\epsilon t),\quad b=r\cos(\epsilon t). \tag{1.1}$$

The frequencies of the "frozen" ($t=\text{const}$) systems are constant,

$$\omega_{1,2}^2=1\pm r, \tag{1.2}$$

and the angle, say between the ω_2 mode and the q_1 axis, is

$$\phi=\epsilon t/2-\pi/4. \tag{1.3}$$

As the slow time variable $s=\epsilon t$ varies between zero and 2π, the parameters loop around the degeneracy point corresponding to $r=0$. For small, but fixed r, the following is conventional wisdom among those working in the subjects of the Berry phase[1] and Hannay angles:[2] as $\epsilon\to 0$, the actions I_j are adiabatic invariants and the angle variables θ_j evolve just with the dynamical phases $\theta_j=\theta_j(0)+\int_0^t\omega_j\,dt$ (in other words, Berry's one-form vanishes). However, as a result of (1.3), after the circuit the phases are in *opposition* to those of the oscillator in which a, b, c are maintained with their original values.[3,4]

On the other hand, for ϵ fixed and $r\to 0$ (approach to resonance) the limit system is the isotropic oscillator, with constant parameters $a=c=1$, $b=0$, for which no such opposition of phase results. The aim of this article is merely to indicate that quite different types of behavior may occur depending on a coupling parameter $k\sim r/\epsilon$. Actually, this would be no surprise to experts in linear mode conversion.[5,6]

II. "SCISSORED" EQUATIONS OF MOTION IN ACTION ANGLE VARIABLES

Let a time-dependent canonical transformation be applied to (p_1,p_2,q_1,q_2) space via

$$(p_1,p_2)^t=R(P_1,P_2)^t,\quad (q_1,q_2)^t=R(Q_1,Q_2)^t,$$
$$R=\begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}, \tag{2.1}$$

where $\phi=\phi(t)$ is the angle between the ω_1 mode and the q_1 axis. Furthermore, let

$$Y_j=P_j/\sqrt{\omega_j},\quad X_j=\sqrt{\omega_j}Q_j,$$
$$I_j=\tfrac{1}{2}(X_j^2+Y_j^2),\quad \theta_j=\arctan(X_j/Y_j). \tag{2.2}$$

The Hamiltonian $H=T+V$ for (1.1) becomes[7]

$$K=\omega_1I_1+\omega_2I_2-\left(p_1\frac{\partial q_1}{\partial t}+p_2\frac{\partial q_2}{\partial t}\right),$$

where p_j, q_j are functions of θ_1, θ_2, I_1, I_2, t via (2.1) and (2.2). After a long, but straightforward calculation one obtains

$$K=\omega_1I_1+\omega_2I_2+\frac{\dot\omega_1}{\omega_1}I_1\sin(\theta_1)\cos(\theta_1)+\frac{\dot\omega_2}{\omega_2}I_2\sin(\theta_2)\cos(\theta_2)+\dot\phi\sqrt{I_1I_2}\times\left[\frac{\omega_1+\omega_2}{\sqrt{\omega_1\omega_2}}\sin(\theta_2-\theta_1)+\frac{\omega_1-\omega_2}{\sqrt{\omega_1\omega_2}}\sin(\theta_1+\theta_2)\right]. \tag{2.3}$$

Formula (2.3) holds for any time-dependent variations of a, b, c [yielding the corresponding functions $\omega_j(t)$ and $\phi(t)$]. In our case $\dot\omega_j\equiv 0$ and $\dot\phi\equiv\epsilon/2$, so that

$$K=\sqrt{1+r}\,I_1+\sqrt{1-r}\,I_2+\frac{\epsilon}{2}\sqrt{I_1I_2}\times\left\{\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\sin(\theta_2-\theta_1)+\frac{\sqrt{1+r}-\sqrt{1-r}}{(1-r^2)^{1/4}}\sin(\theta_1+\theta_2)\right\}. \tag{2.4}$$

It is customary in resonance problems to introduce the slowly varying combination of phases as one of the dynamic variables and we proceed accordingly. Let

$$\alpha=\theta_1-\theta_2,\quad \lambda=\theta_1+\theta_2,$$
$$2J=I_1-I_2,\quad 2I=I_1+I_2,$$

so that

$$K=I(\sqrt{1+r}+\sqrt{1-r})+J(\sqrt{1+r}-\sqrt{1-r})-\frac{\epsilon}{2}\sqrt{I^2-J^2}\left(\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\right)\sin(\alpha)$$

$$+\frac{\epsilon}{2}\sqrt{I^2-J^2}\left(\frac{\sqrt{1+r}-\sqrt{1-r}}{(1-r^2)^{1/4}}\right)\sin(\lambda). \qquad (2.5)$$

For $r \ll 1$, $\epsilon \ll 1$, the "scissored" Hamiltonian K_{sc} obtained by dropping the last term in (2.5) gives a good approximation to the motion.[8] For K_{sc}, I is a constant of motion and one is led to the one-degree-of-freedom system with the Hamiltonian

$$M(J,\alpha) = J(\sqrt{1+r}-\sqrt{1-r}) - (\epsilon/2)\sqrt{I^2-J^2} \times [(\sqrt{1+r}+\sqrt{1-r})/(1-r^2)^{1/4}]\sin(\alpha). \qquad (2.6)$$

In the next section we show that $J(t)$, $\alpha(t)$ can be obtained in closed form just using elementary functions. Note that the approximate dynamics for λ is then governed by

$$\dot\lambda = (\sqrt{1+r}+\sqrt{1-r}) - (\epsilon/2)[I/\sqrt{I^2-J^2(t)}] \times [(\sqrt{1+r}+\sqrt{1-r})/(1-r^2)^{1/4}]\sin\alpha(t).$$

Now, since $\theta_1 = \frac{1}{2}(\alpha+\lambda)$, $\theta_2 = \frac{1}{2}(\lambda-\alpha)$, it follows that

$$\dot\theta_1 \cong \sqrt{1+r} - \frac{\epsilon}{4}\frac{I-J}{\sqrt{I^2-J^2}}\frac{\sqrt{1-r}+\sqrt{1+r}}{(1-r^2)^{1/4}}\sin(\alpha),$$

$$\dot\theta_2 \cong \sqrt{1-r} - \frac{\epsilon}{4}\frac{I+J}{\sqrt{I^2-J^2}}\frac{\sqrt{1-r}+\sqrt{1+r}}{(1-r^2)^{1/4}}\sin(\alpha). \qquad (2.7)$$

[The true solutions are solutions of (2.7) with an $O(r\epsilon)$, zero-average, fast periodic function added.[9]]

In analogy with the Hannay angles,[2] we define

$$\Delta\theta_1 \doteq -\frac{\epsilon}{4}\frac{\sqrt{1-r}+\sqrt{1+r}}{(1-r^2)^{1/4}}\int_0^{2\pi/\epsilon}\frac{I-J}{\sqrt{I^2-J^2}}\sin\alpha(t)dt,$$

$$\Delta\theta_2 \doteq -\frac{\epsilon}{4}\frac{\sqrt{1-r}+\sqrt{1+r}}{(1-r^2)^{1/4}}\int_0^{2\pi/\epsilon}\frac{I+J}{\sqrt{I^2-J^2}}\sin\alpha(t)dt. \qquad (2.8)$$

Using the energy parameter M in (2.6) we obtain

$$\Delta\theta_1 = \frac{1}{2}\int_0^{2\pi/\epsilon}\frac{M-J(t)(\sqrt{1+r}-\sqrt{1-r})}{I+J(t)}dt,$$

$$\Delta\theta_2 = \frac{1}{2}\int_0^{2\pi/\epsilon}\frac{M-J(t)(\sqrt{1+r}-\sqrt{1-r})}{I-J(t)}dt. \qquad (2.8')$$

We now show that $J(t)$ can be computed in closed form.

III. CLOSED FORM SOLUTION OF THE REDUCED SYSTEM: PHASE PORTRAIT

The equations of motion for (2.6) are

$$\dot J = \frac{\epsilon}{2}\sqrt{I^2-J^2}\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\cos\alpha, \qquad (3.1a)$$

$$\dot\alpha = \sqrt{1+r}-\sqrt{1-r}+\frac{\epsilon}{2}\frac{J}{\sqrt{I^2-J^2}} \times \left(\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\right)\sin\alpha. \qquad (3.1b)$$

From (3.1a) and (2.6) we obtain

$$\dot J = \pm\frac{\epsilon}{2}\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\sqrt{I^2-J^2} \times \sqrt{1-\frac{4(1-r^2)^{1/2}}{(\sqrt{1+r}+\sqrt{1-r})^2\epsilon^2}\frac{1}{I^2-J^2}[M-J(\sqrt{1+r}-\sqrt{1-r})]^2},$$

where the $\pm$ indicates the sign of $\cos\alpha$ in the corresponding region. This expression simplifies to

$$\dot J = \pm\sqrt{\frac{\epsilon^2}{4}\frac{(\sqrt{1+r}+\sqrt{1-r})^2}{(1-r^2)^{1/2}}(I^2-J^2) - [M-J(\sqrt{1+r}-\sqrt{1-r})]^2}$$

or

$$\dot J = \pm\sqrt{-(aJ^2+bJ+c)},$$

$$a = (\sqrt{1+r}-\sqrt{1-r})^2 + \frac{\epsilon^2}{4}\frac{(\sqrt{1+r}+\sqrt{1-r})^2}{(1-r^2)^{1/2}},$$

$$b = -2M(\sqrt{1+r}-\sqrt{1-r}),$$

$$c = M^2 - \frac{\epsilon^2}{4}\frac{(\sqrt{1+r}+\sqrt{1-r})^2}{(1-r^2)^{1/2}}I^2. \qquad (3.2)$$

Equations (3.2) are readily integrated as

$$\arccos[1-2((J-J^-)/(J^+-J^-))] = \pm\sqrt{a}t + \text{const}, \qquad (3.3)$$

where $J^+ > J^-$ are the roots of $aJ^2+bJ+c=0$. Again, we stress that the plus sign is taken for $\cos\alpha > 0$. Before going further is it perhaps best to sketch the phase portraits of (2.6) in $\alpha \in (-\infty,\infty)$, $J \in (-I,I)$. The equilibrium points for (3.1) are

$$C_1\colon \alpha = \pi/2, \quad J = -(k/\sqrt{1+k^2})I,$$
$$C_2\colon \alpha = 3\pi/2, \quad J = (k/\sqrt{1+k^2})I, \qquad (3.4)$$

where

$$k = 2\left(\frac{\sqrt{1+r}-\sqrt{1-r}}{\sqrt{1+r}+\sqrt{1-r}}\right)\frac{(1-r^2)^{1/4}}{\epsilon} \sim \frac{r}{\epsilon} + O\left(\frac{r^3}{\epsilon}\right). \qquad (3.5)$$

These equilibria are centers since in both cases the trace of the Jacobian matrix is zero, while the determinant is equal to

$$\omega^2 = \frac{\epsilon^2}{4}\frac{(\sqrt{1+r}+\sqrt{1-r})^2}{(1-r^2)^{1/2}}(1+k^2). \qquad (3.6)$$

Although there are no saddles, the points $J = \pm I$, $\alpha = n\pi$ give rise to the separatrices C^+, C^-, as indicated in Fig. 1. Coordinates of points P, Q are given by

$$P = \left(\alpha = \frac{3\pi}{2}, J = I\frac{k^2-1}{k^2+1}\right),$$

$$Q = \left(\alpha = \frac{\pi}{2}, J = I\frac{1-k^2}{1+k^2}\right).$$

For very large k the islands in Fig. 1 are squeezed to a very thin region [Fig. 2(a)], whereas for $k \to 0$ the ribbon region is squeezed into a polygonal, as indicated in Fig. 2(b) and (c).

IV. LIMITS $k \to 0$ AND $k \to \infty$

For $k = 0$ ($r \to 0$, ϵ fixed) we obtain, in the limit, a collection of cells separated by a polygonal line [Fig. 2(c)]. Take, for instance, $0 < \alpha < \pi$. We obtain

$$J(t) = \sqrt{I^2 - (M/\epsilon)^2}\,\sin(\epsilon t + \text{const}),$$

$$\sin\alpha(t) = \frac{-(M/\epsilon)}{\sqrt{I^2 - [I^2 - (M/\epsilon)^2]\sin^2(\epsilon t + \text{const})}},$$

$$(M<0) \qquad (4.1)$$

for the closed trajectories in the cell. We calculate (2.8′). Here

$$\langle\Delta\theta_{1,2}\rangle = \frac{1}{2}\int_0^{2\pi/\epsilon} \frac{M\,dt}{I \pm \sqrt{I^2 - (M/\epsilon)^2}\,\sin(\epsilon t)}. \qquad (4.2)$$

Changing variables to $z = \tan(\epsilon t/2)$ we obtain

$$\langle\Delta\theta_{1,2}\rangle = \frac{M}{\epsilon I}\int_{-\infty}^{\infty} \frac{dz}{z^2 \pm 2\sqrt{1-(M/\epsilon I)^2} + 1} = \pi \qquad (4.2')$$

by a simple residue calculation. Note that the calculation (4.2′) is exact since the last term in (2.5) vanishes for $r = 0$.

Let us now compute $\Delta\theta_{1,2}$ along the polygonal lines ($M = 0$). The speed along the horizontal segments is infinite, so that we may compute only over the vertical segments. The result $\langle\Delta\theta_{1,2}\rangle = \pi$ is valid for all values of M, so that it carries over to $M = 0$.

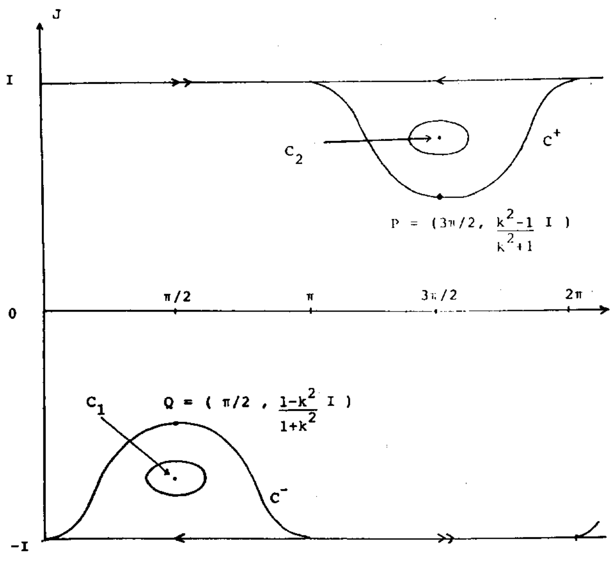
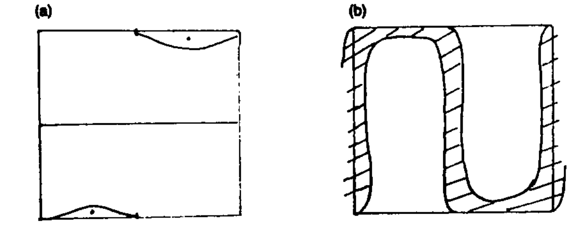


FIG. 1. Phase portrait of the reduced system (sketch).



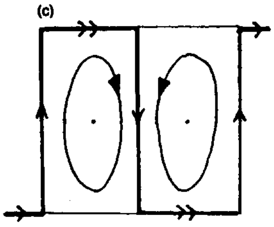


FIG. 2. Limiting phase portraits. (a) $k \to \infty$, (b) $k \to 0$, and (c) $k = 0$.

We now fix r and let $\epsilon \to 0$ (the adiabatic limit, without resonance). In the overwhelmingly dominating ribbon region, we can safely approximate $J \sim$ const for trajectories sufficiently bounded away from the limiting lines $J = \pm I$. One can also approximate $\dot{\alpha} = \sqrt{1-r} - \sqrt{1-r}$. Changing variables to $s = \epsilon t$ we obtain

$$(\Delta\theta_1)_{\text{geo}} \cong -\frac{1}{4}\frac{\sqrt{1+r}-\sqrt{1-r}}{(1-r^2)^{1/4}}\sqrt{\frac{I-J_0}{I+J_0}} \times \lim_{\epsilon\to 0}\int_0^{2\pi}\sin\left(\frac{(\sqrt{1+r}-\sqrt{1-r})}{\epsilon}s\right)ds = 0. \qquad (4.3)$$

What happens on the orbits trapped in the infinitesimally thin islands? It is easy to compute $\Delta\theta_1$ and $\Delta\theta_2$ for the equilibrium points (Fig. 1). For instance, by inserting $C_2: \alpha = 3\pi/2$, $J = kI/\sqrt{1+k^2}$ into (2.8) we obtain

$$\Delta\theta_1 = \frac{\pi}{2}\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\left[\frac{1-k/\sqrt{1+k^2}}{1+k/\sqrt{1+k^2}}\right]^{1/2},$$

$$\Delta\theta_2 = \frac{\pi}{2}\frac{\sqrt{1+r}+\sqrt{1-r}}{(1-r^2)^{1/4}}\left[\frac{1+k/\sqrt{1+k^2}}{1-k/\sqrt{1+k^2}}\right]^{1/2}. \qquad (4.4)$$

Thus

$$\lim_{\epsilon\to 0}\Delta\theta_1(C_2) = 0, \quad \lim_{\epsilon\to 0}\Delta\theta_2(C_2) = \infty. \qquad (4.5)$$

Therefore, the main conclusion of the paper is the following.

All intermediary values between (4.2′) and (4.5) will occur as the coupling parameter k ranges in $(0,\infty)$. Furthermore, the action variables may vary considerably, as depicted in Figs. 1 and 2.

Detailed predictions along more general trajectories will not be done here (although it is not a very difficult calculation). This study would be useful for comparison with real experiments.

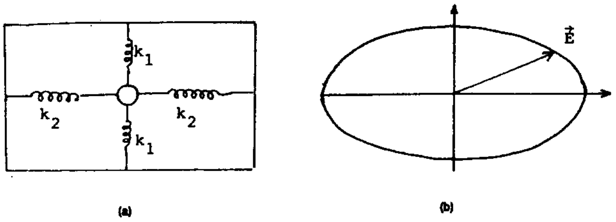


FIG. 3. (a) mechanical analog and (b) optical analog.

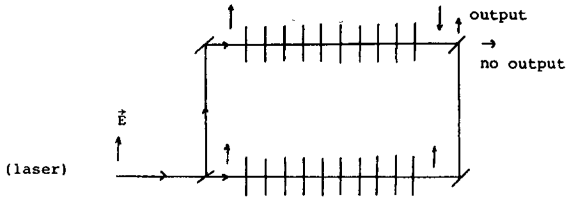


FIG. 4. A gedanken experiment.

V. AN OPTICAL MODEL

The following comments are due to Chiao.[10]

System (1.1)–(1.3) represents the mechanical system consisting of a two-dimensional anisotropic simple harmonic oscillator (2D-SHO), fixed to a "box" on a table that is slowly rotated through 180° around the vertical axis. Now, the elliptical polarization of light is clearly an analog of the mechanical motion of the 2D-SHO; in this optical analog, the major axis of the ellipse of polarization is slowly rotated around an axis parallel to the direction of propagation (Fig. 3).

In the adiabatic limit of infinitesimally small rate of rotation, after a 180° rotation there is a sign change in the amplitude of the light (this sign change is a special case of the Pancharatnam's phase[11]). To study the deviation from adiabaticity, the following experiment could be tried (Fig. 4).

A Mach–Zehnder interferometer has two stacks of polaroid sheets inserted into its two arms. Let these sheets be ideal and in order to keep the intensities balanced, let there be an equal number of sheets in both arms. To approach the adiabatic limit, let the number of sheets be large. In one arm, the polarization axis is forced by the sheets to slowly rotate through 180°. This is done by arranging the angle between adjacent sheets to be small, but of constant sign. In the other arm the polarization is forced by the sheets to alternate back and forth, so that the net rotation is 0°. This is done by arranging the angle between adjacent sheets to be the same as in the first stalk, but of an alternating sign. The optical path lengths through the two stalks are equal by arrangement; the output intensities are also equal.

In the adiabatic limit, there is a sign change between the two arms of the interferometer, so that the light would come out in the unexpected exit port, as indicated in Fig. 4. However, if the eccentricity of the polarization ellipse is small enough (compared with the number of polaroid sheets), then light will come out from both ports.

In principle, an elaboration of this article could predict the rate between the two outcomes. A certain amount of statistics may be necessary because of the dependency on initial phases. We also remark that it seems worthwhile to study the quantum mechanical version of (1.1)–(1.3).

Chiao[10] also pointed out that there may be other physical realizations of (1.1)–(1.3), e.g., optical fibers with a graded refractive index[12] and microwave waveguides.

ACKNOWLEDGMENTS

I would like to thank Professor J. E. Marsden and Professor Alan Weinstein for sponsoring a visit to MSRI and Professor R. Y. Chiao for his comments, especially pertaining to Sec. V, and his encouragement.

I would like to thank Professor A. I. Neishtadt for the following remarks. The "scissored" system can be considered as a linear system, coming back from the action angle coordinates I_j, θ_j to the Cartesian variables $Y_j = (2I_j)^{1/2}\cos\theta_j$, $X_j = (2I_j)^{1/2}\sin\theta_j$. The "scissored" Hamiltonian is

$$M = \tfrac{1}{2}(1+r)^{1/2}(X_1^2+Y_1^2) + \tfrac{1}{2}(1-r)(X_2^2+Y_2^2) + (\epsilon/4)\{[(1+r)^{1/2}+(1-r)^{1/2}]/(1-r^2)^{1/4}\} \times (X_2Y_1 - X_1Y_2).$$

The "scissored" system is linear, time independent, and possesses the integral

$$4I = X_1^2 + Y_1^2 + X_2^2 + Y_2^2.$$

It is possible to describe the trajectories of M by means of the theory of linear Hamiltonian systems; the results are entirely equivalent. Also, the Hamiltonian M may be derived from H by means of the normal forms approach. The first step is the introduction of the variables X_j, Y_j instead of p_j, q_j, yielding a Hamiltonian $K(X,Y,\epsilon)$. "Scissoring" means taking the normal form up to terms of order three (two in X, Y plus one in ϵ) for the resonance 1:1. This normal form contains, by definition, the variables $z_j = X_j + iY_j$, $\bar{z}_j = X_j - iY_j$ only in the combinations $z_j\bar{z}_j$, $z_1\bar{z}_2$, $z_2\bar{z}_1$; it is just M.

This work was performed while visiting the Mathematical Sciences Research Institute, May–June 1989, under DOE Grant No. BE-SG03-88ER25604; travel was financed by the Brazilian agency FAPERJ.

[1] M. V. Berry, J. Phys. A: Math. Gen. **18**, 15 (1985).
[2] J. H. Hannay, J. Phys. A: Math. Gen. **18**, 221 (1985).
[3] J. Koiller, Sem. Bras. Analise **26**, 325 (1987).
[4] R. Montgomery, Commun. Math. Phys. **120** (2), 269 (1988).
[5] R. Grimshaw and J. S. Allen, Stud. Appl. Math. **61**, 55 (1979).
[6] R. O. Dendy, J. Math. Phys. **29**, 2202 (1988).
[7] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).
[8] V. I. Arnol'd, Ed., *Dynamical Systems III* (Springer, Berlin, 1988).
[9] J. A. Morrison, Siam J. Appl. Math. **16** (5), 1024 (1968).
[10] R. Y. Chiao, personal communications (1987, 1989).
[11] M. V. Berry, preprint, 1987.
[12] R. J. Black and A. Ankiewicz, Am. J. Phys. **53**, 554 (1985).

The linearized gauge field propagator in an inhomogeneous medium

P. Tang and L. Wilets
Department of Physics, FM-15, University of Washington, Seattle, Washington 98195

(Received 15 August 1989; accepted for publication 28 February 1990)

In an inhomogeneous medium with $\mu = \kappa^{-1}$, the linearized gauge field propagator is presented in terms of spherical harmonics for the Coulomb gauge. Explicit expressions for the propagator are given in the case of the MIT bag for both the Coulomb and Lorentz gauges. Gauge invariance is shown explicitly for the one gluon interaction in the MIT bag. This work supplements and corrects the earlier work by Bickeböller, Goldflam, and Wilets [J. Math. Phys. **26**, 1810 (1985)].

I. INTRODUCTION

The linearized gauge field propagator of quantum chromodynamics is identical to the Maxwell Green's function of electrodynamics. It is the propagator that is used to calculate one gluon exchange matrix elements. In chromodielectric soliton models of QCD,[1] the propagator must be calculated in an inhomogeneous chromodielectric medium with a dielectric function $\kappa(\mathbf{r})$ and magnetic susceptibility $\mu = 1/\kappa$, in units where $c = 1$. The MIT bag model is an important special case of chromodielectric models. There are many other applications of Maxwell propagators in inhomogeneous media and we hope that this paper will be useful to workers in various fields.

Construction of the Maxwell propagator in terms of vector spherical harmonics has been fraught with difficulties. Even in the case of homogeneous media ($\kappa = 1$), the expansion was not formulated correctly until 1979 when Johnson, Howard, and Dudley[2] solved the problem. Bickeböller, Goldflam, and Wilets[3] (BGW) presented a formulation for the general inhomogeneous $\kappa(\mathbf{r})$. Their work, however, contains an error in the definition of the transverse current. We show here how that error can be simply rectified, so that the propagator can be calculated correctly utilizing their results.

The present authors discovered the error by noting that for the MIT bag, $\kappa = \theta(R - r)$, the appropriate boundary conditons could not be satisfied for the electric (TM) modes. The gluon field in the MIT bag has been studied by several authors,[4–6] but in Refs. 4 and 5 it is studied only for special transitions; no general propagator is given. In Ref. 6 a propagator in Feynman gauge is proffered, but we find that it does not satisfy the boundary condition $\mathbf{r}\times\mathbf{B} = 0$ (see the Appendix). Here we give the general propagator in a Coulomb gauge and in a Lorentz gauge, and gauge invariance is shown explicitly in one gluon interaction.

The three-vector current can always be decomposed into a transverse and a longitudinal component, although this is not unique: One can always add to one and subtract from the other a term of the form $\nabla\phi$, where $\nabla^2\phi = 0$. There is yet a different problem here. The complement of the transverse current that is required in medium is not necessarily longitudinal. We do not have a separation into transverse and longitudinal currents.

We present below the correct definition of the transverse current $\mathbf{J}_{t\kappa}$ for a general dielectric $\kappa(\mathbf{r})$ and the transverse delta function $\overleftrightarrow{\delta}_{t\kappa}(\mathbf{r},\mathbf{r}')$ that projects $\mathbf{J}$ onto $\mathbf{J}_t$. By operating on the BGW propagator from the right with $\overleftrightarrow{\delta}_{t\kappa}$, one can construct the proper propagator. Complete results are given for the case of a spherically symmetric dielectric function and the special case of the MIT bag.

II. THE COULOMB GAUGE IN MEDIUM

The medium is assumed to be color neutral so that the propagator is diagonal in the color indices. In what follows, we will drop reference to color.

Maxwell's equations, with $c = 1$ and $\mu = \kappa^{-1}$, are given by

$$\partial^\mu \kappa(\mathbf{r})[\partial_\mu A_\nu - \partial_\nu A_\mu] = J_\nu. \tag{2.1}$$

We work in the Coulomb gauge defined by

$$\nabla\cdot\kappa\mathbf{A} = 0. \tag{2.2}$$

The $\nu = 0$ component of Eq. (2.1) yields

$$-\nabla\kappa\cdot\nabla A_0 = J_0(\mathbf{r},t). \tag{2.3}$$

The time–time component of the Green's function, G^{00}, defined by

$$A_0(\mathbf{r},t) = \int d^3r'\, G^{00}(\mathbf{r},\mathbf{r}')J_0(\mathbf{r}',t), \tag{2.4}$$

satisfies the equation

$$-\nabla\kappa\cdot\nabla G^{00}(\mathbf{r},\mathbf{r}') = \delta^3(\mathbf{r}-\mathbf{r}'). \tag{2.5}$$

Note that G^{00} is instantaneous.

Consideration of the $\nu = i$ components of Eq. (2.1) leads to

$$\kappa\,\partial_t^2\mathbf{A} - \nabla^2\kappa\mathbf{A} + \nabla\times(\kappa\mathbf{A}\times\nabla\ln\kappa) = \mathbf{J} - \kappa\nabla\,\partial_t A_0 \equiv \mathbf{J}_t. \tag{2.6}$$

The transverse current defined by (2.6) can be expressed in terms of $\mathbf{J}$ using the time–time Green's function:

$$\mathbf{J}_t(\mathbf{r},t) = \mathbf{J}(\mathbf{r},t) - \kappa(\mathbf{r})\nabla\int d^3r'\, G^{00}(\mathbf{r},\mathbf{r}')\partial_t J_0(\mathbf{r}',t). \tag{2.7}$$

 0022-2488/90/071661-09$03.00

Using current conservation, $\partial_t J_0 + \nabla\cdot\mathbf{J} = 0$, and partial integration, we obtain

$$\mathbf{J}_t(\mathbf{r},t) = \mathbf{J}(\mathbf{r},t) - \kappa(\mathbf{r})\nabla\int d^3r'\,(\nabla' G^{00}(\mathbf{r},\mathbf{r}'))\cdot\mathbf{J}(\mathbf{r}',t). \tag{2.8}$$

It is clear that $\mathbf{J}_t$ is indeed transverse since, using (2.5), we find

$$\begin{aligned}\nabla\cdot\mathbf{J}_t &= \nabla\cdot\mathbf{J} + \int d^3r'(\nabla'\delta^3(\mathbf{r}-\mathbf{r}'))\cdot\mathbf{J}(\mathbf{r}',t)\\ &= \nabla\cdot\mathbf{J} - \nabla\cdot\mathbf{J} = 0.\end{aligned} \tag{2.9}$$

Note that $\mathbf{J} - \mathbf{J}_t$ is not necessarily longitudinal since

$$\nabla\times(\mathbf{J}-\mathbf{J}_t) = (\nabla\kappa)\times\nabla\int d^3r'(\nabla' G^{00}(\mathbf{r},\mathbf{r}'))\cdot\mathbf{J}(\mathbf{r}',t) \tag{2.10}$$

does not vanish indentically for κ not a constant.

We now define the transverse vector delta function $\overleftrightarrow{\delta}_{t\kappa}(\mathbf{r},\mathbf{r}')$ by

$$\delta^{ij}_{t\kappa}(\mathbf{r},\mathbf{r}') = \delta^{ij}\delta^3(\mathbf{r}-\mathbf{r}') - \kappa(\mathbf{r})\partial^i\partial'^j G^{00}(\mathbf{r},\mathbf{r}'); \tag{2.11}$$

we define the second part of the right-hand side as

$$U^{ij}(\mathbf{r},\mathbf{r}') = \partial^i\partial'^j G^{00}(\mathbf{r},\mathbf{r}').$$

The homogeneous case $\kappa = 1$ is of special interest. Since $G^{00}_{\kappa=1} = 1/4\pi|\mathbf{r}-\mathbf{r}'|$ we find

$$\delta^{ij}_{t1}(\mathbf{r},\mathbf{r}') = \delta^{ij}\delta^3(\mathbf{r}-\mathbf{r}') - \partial^i\partial'^j\frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}. \tag{2.12}$$

More on this in the next section.

Note that $\overleftrightarrow{\delta}_{t1}$ is a projection operator onto the space of transverse vectors. Specifically, for any transverse vector $\mathbf{V}_t$, where $\nabla\cdot\mathbf{V}_t = 0$, we have

$$\int d^3r'\,\delta^{ij}_{t1}(\mathbf{r},\mathbf{r}')V^j_t(\mathbf{r}') = V^i_t(\mathbf{r}). \tag{2.13}$$

For general κ, $\delta^{ij}_{t\kappa}$ does not have this property. In particular,

$$\overleftrightarrow{\delta}_{t\kappa}\cdot\overleftrightarrow{\delta}_{t1} = \overleftrightarrow{\delta}_{t\kappa}, \tag{2.14a}$$

$$\overleftrightarrow{\delta}_{t1}\cdot\overleftrightarrow{\delta}_{t\kappa} = \overleftrightarrow{\delta}_{t\kappa}. \tag{2.14b}$$

Here and below there are implicit integrations.

Let us Fourier transform the time dependence of $\mathbf{J}(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$ to $\mathbf{J}(\mathbf{r},\omega)$ and $\mathbf{A}(\mathbf{r},\omega)$, and let κ be time-independent. The Green's function corresponding to Eq. (2.6) satisfies

$$-[\nabla^2 + \omega^2 + \nabla\times(\nabla\ln\kappa)\times]\kappa\overleftrightarrow{G}_\kappa(\mathbf{r},\mathbf{r}') = \overleftrightarrow{\delta}_{t\kappa}(\mathbf{r},\mathbf{r}'), \tag{2.15}$$

where $\overleftrightarrow{G}\to G^i_j$. Here we indicate explicitly the κ-dependence of $\overleftrightarrow{G}_\kappa$.

BGW[3] erroneously solved Eq. (2.15) with $\overleftrightarrow{\delta}_{t1}$ instead of $\overleftrightarrow{\delta}_{t\kappa}$

$$-[\nabla^2 + \omega^2 + \nabla\times(\nabla\ln\kappa)\times]\kappa\overleftrightarrow{G}_{\mathrm{BGW}}(\mathbf{r},\mathbf{r}') = \overleftrightarrow{\delta}_{t1}(\mathbf{r},\mathbf{r}'). \tag{2.16}$$

If we operate on Eq. (2.16) on the right by $\overleftrightarrow{\delta}_{t\kappa}$ and use Eq. (2.14b), we find that this

$$\overleftrightarrow{G}_{\mathrm{BGW}}\cdot\overleftrightarrow{\delta}_{t\kappa} = \overleftrightarrow{G}_\kappa \tag{2.17}$$

satisfies the Green's function equation (2.15), or, alternatively,

$$\mathbf{A} = \overleftrightarrow{G}_\kappa\cdot\mathbf{J} = \overleftrightarrow{G}_{\mathrm{BGW}}\cdot\overleftrightarrow{\delta}_{t\kappa}\cdot\mathbf{J}. \tag{2.18}$$

The interaction between two currents, $\mathbf{J}_1$ and $\mathbf{J}_2$, is

$$-\mathbf{J}_1\cdot\mathbf{A}_2 = -\mathbf{J}_1\cdot\overleftrightarrow{G}_\kappa\cdot\mathbf{J}_2. \tag{2.19}$$

III. CONSTRUCTION OF PROPAGATORS

The time-independent equation for the scalar Green's function $G(\mathbf{r},\mathbf{r}')\equiv G^{00}(\mathbf{r},\mathbf{r}')$ is

$$\nabla\cdot\kappa(\mathbf{r})\nabla G(\mathbf{r},\mathbf{r}') = -\delta^3(\mathbf{r}-\mathbf{r}'). \tag{3.1}$$

It can be solved[3] to yield

$$\begin{aligned}G(\mathbf{r},\mathbf{r}') &= C_{\alpha\alpha'}\frac{1}{r_<}j^\alpha_{lm}(r_<)Y_{lm}(\Omega_<)\frac{1}{\sqrt{\kappa(r_<)}}\\ &\quad\times\frac{1}{r_>}n^{\alpha'}_{l'm'}(r_>)Y^*_{l'm'}(\Omega_>)\frac{1}{\sqrt{\kappa(r_>)}}\\ &\equiv C_{\alpha\alpha'}u^{\alpha'}_{l'm'}(r_<)v^\alpha_{lm}(r_>)Y_{l'm'}(\hat r_<)Y^*_{lm}(\hat r_>),\end{aligned} \tag{3.2}$$

where j,n satisfy the equations

$$\left\{\left(-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}\right)\delta_{ll'}\delta_{mm'} + W_{LM}(r)\langle lm|Y_{LM}|l'm'\rangle\right\}\times\begin{Bmatrix}j^\alpha_{l'm'}(r)\\ n^{\alpha'}_{l'm'}(r)\end{Bmatrix} = 0. \tag{3.3}$$

The repeated index summation convention is employed throughout unless otherwise noted. The set $\{\alpha\}$ of solutions that are regular at the origin is given by $\{j^\alpha_{lm}\}$ and the set $\{\alpha'\}$ that is regular at infinity by $\{n^{\alpha'}_{l'm'}\}$. If $\kappa(\mathbf{r})$ goes asymptotically to a constant greater than zero, the corresponding boundary conditions are

$$j^\alpha_{lm}(r)\sim r^{l+1}\delta^\alpha_{lm},\quad \text{for } r\to 0, \tag{3.4a}$$

$$n^{\alpha'}_{l'm'}(r)\sim\frac{1}{r^{l'}}\delta^{\alpha'}_{l'm'},\quad \text{for } r\to\infty. \tag{3.4b}$$

The function

$$W(\mathbf{r})\equiv\tfrac14|\nabla\ln\kappa(\mathbf{r})|^2 + \tfrac12\nabla^2\ln\kappa(\mathbf{r}) \tag{3.5}$$

has been expanded in spherical harmonics:

$$W(\mathbf{r}) = W_{LM}(r)Y_{LM}(\Omega). \tag{3.6}$$

From the continuity equation for G and the discontinuity equation for the derivative, the coefficients $C_{\alpha\alpha'}$ satisfy the linear algebraic equations

$$\{j^\alpha_{lm}(r)n^{\alpha'}_{l'm'}(r) - (-1)^{m+m'}j^\alpha_{l'-m'}(r)n^{\alpha'}_{l-m}(r)\}\times C_{\alpha\alpha'} = 0, \tag{3.7a}$$

$$\left\{\left(\frac{d}{dr}j^\alpha_{lm}(r)\right)n^{\alpha'}_{l'm'}(r) - (-1)^{m+m'}j^\alpha_{l'-m'}(r)\times\left(\frac{d}{dr}n^{\alpha'}_{l-m}(r)\right)\right\}C_{\alpha\alpha'} = \delta_{l,l'}\delta_{m,m'} \tag{3.7b}$$

at any radius r. The factors $(-1)^{m+m'}$ that were omitted in Ref. 3 contribute for nonaxially symmetric κ.

Since we already have available G^{ij}_{BGW} in explicit vector harmonic form from Ref. 3, we can construct G^{ij}_κ from (2.17) with U^{ij}_κ. After putting (3.2) into the definition of U^{ij}_κ, we have

$$U^{ij}_{\kappa}(\mathbf{r},\mathbf{r}') = C_{\alpha\alpha'}\left[\sqrt{\frac{l'+1}{2l'+1}}\,u^{+}_{l',m'}(r_<)\mathscr{Y}_{l',l'+1,m'}(\hat{r}_<) + \sqrt{\frac{l'}{2l'+1}}\,u^{-}_{l',m'}(r_<)\mathscr{Y}_{l',l'-1,m'}(\hat{r}_<)\right]^{i_<}$$
$$\times\left[\sqrt{\frac{l+1}{2l+1}}\,v^{+}_{l,m}(r_>)\mathscr{Y}^{*}_{l,l+1,m}(\hat{r}_>) + \sqrt{\frac{l}{2l+1}}\,v^{-}_{l,m}(r_>)\mathscr{Y}^{*}_{l,l-1,m}(\hat{r}_>)\right]^{i_>} + \frac{\delta(r-r')}{\sqrt{\kappa(\mathbf{r})\kappa(\mathbf{r}')}}\frac{1}{r^2}$$
$$\times\left[\sqrt{\frac{l+1}{2l+1}}\,\mathscr{Y}_{l,l+1,m} - \sqrt{\frac{l}{2l+1}}\,\mathscr{Y}_{l,l-1,m}\right]^{i}\left[\sqrt{\frac{l+1}{2l+1}}\,\mathscr{Y}^{*}_{l,l+1,m} - \sqrt{\frac{l}{2l+1}}\,\mathscr{Y}^{*}_{l,l-1,m}\right]^{j}, \tag{3.8}$$

where we define

$$\binom{u}{v}^{+}_{l,m} \equiv \left(\frac{l}{r} - \frac{d}{dr}\right)\binom{u}{v}_{l,m}, \qquad \binom{u}{v}^{-}_{l,m} \equiv \left(\frac{l+1}{r} + \frac{d}{dr}\right)\binom{u}{v}_{l,m}. \tag{3.9}$$

We now have

$$\overleftrightarrow{G}_{\kappa} = \overleftrightarrow{G}_{\mathrm{BGW}} - \overleftrightarrow{G}_{\mathrm{BGW}}\cdot\kappa\overleftrightarrow{U}_{\kappa}. \tag{3.10}$$

For the case $\kappa = 1$ everywhere, we have

$$\delta^{ij'}_{t1}(\mathbf{r}-\mathbf{r}') = \frac{\delta(r-r')}{r^2}\left\{\mathscr{Y}^{i}_{llm}(\Omega)\mathscr{Y}^{*i'}_{llm}(\Omega') + \left(\sqrt{\frac{1}{2l+1}}\,\mathscr{Y}^{i}_{l,l+1,m}(\Omega) + \sqrt{\frac{l+1}{2l+1}}\,\mathscr{Y}^{i}_{l,l-1,m}(\Omega)\right)\right.$$
$$\left.\times\left(\sqrt{\frac{l}{2l+1}}\,\mathscr{Y}^{*i'}_{l,l+1,m}(\Omega') + \sqrt{\frac{l+1}{2l+1}}\,\mathscr{Y}^{*i'}_{l,l-1,m}(\Omega')\right)\right\} - \sqrt{l(l+1)}\,\frac{r_<^{l-1}}{r_>^{l+2}}\,\mathscr{Y}^{i_<}_{l,l-1,m}(\Omega_<)\mathscr{Y}^{*i_>}_{l,l+1,m}(\Omega_>). \tag{3.11}$$

IV. TENSOR PROPAGATOR FOR SPHERICAL κ

In the spherically symmetric κ case, $\kappa(\mathbf{r}) \equiv \kappa(r)$, we can first construct transverse $\overline{G}^{\,ij}_{\mathrm{BGW}} = \kappa G^{\,ij}_{\mathrm{BGW}}$ according to Ref. 3. For $r < r'$, we have

$$\overline{G}^{\,ii'}_{\mathrm{BGW}}(\mathbf{r},\mathbf{r}',\omega) = \frac{1}{r}j_{Mj}(r,\omega)\mathscr{Y}^{i}_{jjm}(\Omega)\frac{1}{r'}a_{Mj}(r',\omega)\mathscr{Y}^{*i'}_{jjm}(\Omega') + \left(-i\nabla\times\frac{1}{r}j_{Ej}(r,\omega)\mathscr{Y}_{jjm}(\Omega)\right)^{i}$$
$$+\left(-i\nabla\times\frac{1}{r}z^{<}_{Ej}(r,\omega)\mathscr{Y}_{jjm}(\Omega)\right)^{i}\left(-i\nabla\times\frac{1}{r'^{(j+1)}}\mathscr{Y}^{*}_{jjm}(\Omega')\right)^{i'}$$
$$\equiv j_{jl}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)a_{jl'}(r',\omega)\mathscr{Y}^{*i'}_{jl'm}(\Omega') + z^{<}_{jl}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\,\frac{-\sqrt{j(2j+1)}}{r'^{(j+2)}}\,\mathscr{Y}^{*}_{j,j+1,m}(\Omega')^{i'}, \tag{4.1a}$$

and for $r > r'$

$$\overline{G}^{\,ii'}_{\mathrm{BGW}}(\mathbf{r},\mathbf{r}',\omega) = \frac{1}{r}n_{Mj}(r,\omega)\mathscr{Y}^{i}_{jjm}(\Omega)\frac{1}{r'}b_{Mj}(r',\omega)\mathscr{Y}^{*i'}_{jjm}(\Omega') + \left(-i\nabla\times\frac{1}{r}n_{Ej}(r,\omega)\mathscr{Y}_{jjm}(\Omega)\right)^{i}\left(-i\nabla'\times\frac{1}{r'}b_{Ej}(r',\omega)\mathscr{Y}^{*}_{jjm}(\Omega')\right)^{i'}$$
$$+\left(-i\nabla\times\frac{1}{r}z^{>}_{Ej}(r,\omega)\mathscr{Y}_{jjm}(\Omega)\right)^{i}\left(-i\nabla\times r'^{j}\mathscr{Y}^{*}_{jjm}(\Omega')\right)^{i'}$$
$$\equiv n_{jl}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)b_{jl'}(r',\omega)\mathscr{Y}^{*i'}_{jl'm}(\Omega') + z^{>}_{jl}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\sqrt{(j+1)(2j+1)}\,r'^{j-1}\mathscr{Y}^{*}_{j,j-1,m}(\Omega')^{i'}, \tag{4.1b}$$

where the various (j,n,z) functions, generically f_{Ml}, f_{El}, are obtained from the following differential equations:

$$L_1\begin{Bmatrix} j_{Ml} \\ n_{Ml}\end{Bmatrix} = 0, \quad L_2\begin{Bmatrix} j_{El} \\ n_{El}\end{Bmatrix} = 0;$$
$$L_1 z^{>,<}_{Ml} = 0, \quad L_2 z^{>,<}_{El} = \frac{1}{2l+1}\begin{Bmatrix} r^{-l} \\ r^{l+1}\end{Bmatrix}; \tag{4.2}$$

the operators are defined by

$$L_0 \equiv -\omega^2 + \frac{l(l+1)}{r^2} - \frac{d^2}{dr^2},$$
$$L_1 \equiv L_0 + (\ln\kappa)'\frac{d}{dr} + (\ln\kappa)'', \tag{4.3}$$
$$L_2 \equiv L_0 + (\ln\kappa)'\frac{d}{dr};$$

for any transverse function, the relation between $\{f_{jlm}\}$ and $\{f_{Mlm}, f_{Elm}\}$ is[7]

$$f_{j,j+1,m}(r,\omega) = \left(\frac{j}{2j+1}\right)^{1/2}\left(\frac{d}{dr} - \frac{j}{r}\right)\frac{1}{r}f_{Ejm}(r,\omega),$$
$$f_{jjm}(r,\omega) = \frac{1}{r}f_{Mjm}(r,\omega),$$
$$f_{j,j-1,m}(r,\omega) = \left(\frac{j+1}{2j+1}\right)^{1/2}\left(\frac{d}{dr} + \frac{j+1}{r}\right)\frac{1}{r}f_{Ejm}(r,\omega). \tag{4.4}$$

For the spherically symmetric κ considered here, the electric (TM) modes do not couple to the magnetic (TE) modes. Because of the continuity condition for $G^{\,ii'}_{\mathrm{BGW}}(r,r,\Omega,\Omega')$ and the discontinuity of derivative implied by (2.16), it follows that $a(r), b(r)$ for any r satisfy the equations

$$j_{Ml}a_{Ml'} - n_{Ml}b_{Ml'} = 0, \tag{4.5a}$$
$$j_{El}a_{El'} - n_{El}b_{El'} + z^{<}_{El}\frac{1}{r^{l'}} - z^{>}_{El}\,r^{l'+1} = 0, \tag{4.5b}$$
$$\left(\frac{d}{dr}j_{Ml}\right)a_{Ml'} - \left(\frac{d}{dr}n_{Ml}\right)b_{Ml'} = \delta_{ll'}, \tag{4.5c}$$

$$\left(\frac{d}{dr}j_{El}\right)a_{El'}-\left(\frac{d}{dr}n_{El}\right)b_{El'}+\left(\frac{d}{dr}z^{<}_{El}\right)\frac{1}{r^{l'}}-\left(\frac{d}{dr}z^{>}_{El}\right)r^{l'+1}=0. \tag{4.5d}$$

We can solve as a function of r for

$$a_{Ml}=-\frac{n_{Ml}}{W(j_{Ml},n_{Ml})},$$

$$b_{Ml}=-\frac{j_{Ml}}{W(j_{Ml},n_{Ml})},$$

$$a_{El}=\frac{W(n_{El},z^{<}_{El})/r^{l}-W(n_{El},z^{>}_{El})r^{l+1}}{W(j_{El},n_{El})},$$

$$b_{El}=\frac{W(j_{El},z^{<}_{El})/r^{l}-W(j_{El},z^{>}_{El})r^{l+1}}{W(j_{El},n_{El})}; \tag{4.6}$$

here we define the Wronskian $W(f,g)\equiv fg'-f'g$. From (4.2), it can be shown that all the $W(j,n)$ are proportional to $\kappa(r)$.

From (3.8) we can construct $\overset{\leftrightarrow}{\delta}_{t\kappa}$ for spherically symmetric $\kappa(r)$:

$$\begin{aligned}\delta^{ii'}_{t\kappa}(\mathbf{r},\mathbf{r}')=\frac{\delta(r-r')}{r^2}&\Bigg\{\mathscr{Y}^{i}_{llm}(\Omega)\mathscr{Y}^{i'*}_{llm}(\Omega')+\left(\sqrt{\frac{l}{2l+1}}\mathscr{Y}^{i}_{l,l+1,m}(\Omega)+\sqrt{\frac{l+1}{2l+1}}\mathscr{Y}^{i}_{l,l-1,m}(\Omega)\right)\\&\times\left(\sqrt{\frac{l}{2l+1}}\mathscr{Y}^{i'*}_{l,l+1,m}(\Omega')+\sqrt{\frac{l+1}{2l+1}}\mathscr{Y}^{i'*}_{l,l-1,m}(\Omega')\right)\Bigg\}-\frac{\kappa(r)}{2l+1}\left[\sqrt{l+1}\,u^{+}_{l}(r_<)\mathscr{Y}^{i<}_{l,l+1,m}(\Omega_<)\right.\\&\left.+\sqrt{l}\,u^{-}_{l}(r_<)\mathscr{Y}^{i<}_{l,l-1,m}(\Omega_<)\right]\left[\sqrt{l+1}\,v^{+}_{l}(r_>)\mathscr{Y}^{*i>}_{l,l+1,m}(\Omega_>)+\sqrt{l}\,v^{-}_{l}(r_>)\mathscr{Y}^{*i>}_{l,l-1,m}(\Omega_>)\right],\end{aligned} \tag{4.7}$$

the relations between j,u and n,v are defined in (3.2). Here all the magnetic and electric modes have no m,m' and l,l' coupling and there is no α coupling.

With (4.7) and (4.1), we can integrate (2.17) and obtain transverse $\overline{G}^{ii'}_{\kappa}=\kappa G^{ii'}_{\kappa}$ in explicit vector harmonic form. In the following, the various functions f_{jl} $(f=j,n,z,a,b)$ are related to the corresponding f_{El} or f_{Ml} according to (4.4). For $r<r'$

$$\begin{aligned}\overline{G}^{ii'}_{\kappa}(\mathbf{r},\mathbf{r}',\omega)&=j_{jj}(r,\omega)\mathscr{Y}^{i}_{jjm}(\Omega)a_{jj}(r',\omega)\mathscr{Y}^{*i'}_{jjm}(\Omega')\\&+\mathscr{H}^{jl}_{1}(r,r',\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\left(\sqrt{\frac{j}{2j+1}}\mathscr{Y}^{*i'}_{j,j+1,m}(\Omega')+\sqrt{\frac{j+1}{2j+1}}\mathscr{Y}^{*i'}_{j,j-1,m}(\Omega')\right)\\&+\mathscr{H}^{jl}_{2}(r,r',\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\left[\sqrt{\frac{j+1}{2j+1}}v^{+}_{j}(r')\mathscr{Y}^{*i'}_{j,j+1,m}(\Omega')+\sqrt{\frac{j}{2j+1}}v^{-}_{j}(r')\mathscr{Y}^{*i'}_{j,j-1,m}(\Omega')\right]\\&+\mathscr{H}^{jl}_{3}(r,r',\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\left[\sqrt{\frac{j+1}{2j+1}}u^{+}_{j}(r')\mathscr{Y}^{*i'}_{j,j+1,m}(\Omega')+\sqrt{\frac{j}{2j+1}}u^{-}_{j}(r')\mathscr{Y}^{*i'}_{j,j-1,m}(\Omega')\right].\end{aligned} \tag{4.8}$$

Here $\mathscr{H}_1,\mathscr{H}_2,\mathscr{H}_3$ are given

$$\mathscr{H}^{jl}_{1}(r,r',\omega)=j_{jl}(r)\left(\sqrt{\frac{j}{2j+1}}a_{j,j+1}(r')+\sqrt{\frac{j+1}{2j+1}}a_{j,j-1}(r')\right)-z^{<}_{jl}(r)\frac{j}{r'^{j+2}}, \tag{4.9a}$$

$$\begin{aligned}\mathscr{H}^{jl}_{2}(r,r',\omega)=&-n_{jl}(r)\int_0^r d\tilde{r}_2\left[\sqrt{\frac{j+1}{2j+1}}b_{j,j+1}(r_2)u^{+}_{j}(r_2)+\sqrt{\frac{j}{2j+1}}b_{j,j-1}(r_2)u^{-}_{j}(r_2)\right]\\&-j_{jl}(r)\int_r^{r'} d\tilde{r}_2\left[\sqrt{\frac{j+1}{2j+1}}a_{j,j+1}(r_2)u^{+}_{j}(r_2)+\sqrt{\frac{j}{2j+1}}a_{j,j-1}(r_2)u^{-}_{j}(r_2)\right]\\&-z^{>}_{jl}(r)\int_0^r d\tilde{r}_2\sqrt{j(j+1)}\,r_2^{j-1}u^{-}_{j}(r_2)+z^{<}_{jl}(r)\int_r^{r'}d\tilde{r}_2\sqrt{j(j+1)}\frac{1}{r_2^{j+2}}u^{+}_{j}(r_2),\end{aligned} \tag{4.9b}$$

$$\begin{aligned}\mathscr{H}^{jl}_{3}(r,r',\omega)=&-j_{jl}(r)\int_{r'}^{\infty}d\tilde{r}_2\left[\sqrt{\frac{j+1}{2j+1}}a_{jj+1}(r_2)v^{+}_{j}(r_2)+\sqrt{\frac{j}{2j+1}}a_{jj-1}(r_2)v^{-}_{j}(r_2)\right]\\&+z^{<}_{jl}(r)\int_{r'}^{\infty}d\tilde{r}_2\frac{\sqrt{j(j+1)}}{r_2^{j+2}}v^{+}_{j}(r_2);\end{aligned} \tag{4.9c}$$

here $l=j\pm1$ and $d\tilde{r}_2=dr_2\,r_2^2\kappa(r_2)$. We have not been able to simplify the electric modes ($\mathscr{H}$) further, but the magnetic modes of $G^{ii'}_{\kappa}(\mathbf{r},\mathbf{r}',\omega)$, the first term on the rhs of (4.8), using (4.6), can be written as

$$-\frac{j_{Mj}(r,\omega)}{\kappa(r)}\mathscr{Y}^{i}_{jjm}(\Omega)\frac{n_{Mj}(r',\omega)}{\kappa(r')}\mathscr{Y}^{*i'}_{jjm}(\Omega').$$

Similarly we can obtain $G^{ii'}_{\kappa}(\mathbf{r},\mathbf{r}',\omega)$ for $r>r'$, and it is symmetric in $\mathbf{r}$ and $\mathbf{r}'$.

V. THE PROPAGATOR IN MIT BAG IN COULOMB AND LORENTZ GAUGES

The usual MIT bag model boundary conditions can be obtained from the statement

$$\kappa(\mathbf{r}) = \begin{cases} 1, & r<R \\ \kappa_v, & r>R \end{cases} \tag{5.1}$$

in the limit $\kappa_v \to 0$. This gives $\mathbf{D}=0$ (except for $l=0$) and $\mathbf{H}=0$ for $r>R$. From the usual boundary conditions that $\mathbf{E}_{\parallel}$, $\mathbf{D}_{\perp}$, $\mathbf{H}_{\parallel}$, and $\mathbf{B}_{\perp}$ are continuous, this implies $\mathbf{r}\cdot\mathbf{E}=0$ and $\mathbf{r}\times\mathbf{B}=0$ as $r\to R$ inside the bag.

First we solve the gluon propagator in Coulomb gauge with the method given above. It is not difficult to solve the scalar Green's function equation (3.1). Here we are interested only in $r,r'<R$, and we neglect the terms of higher order in κ_v (Ref. 8):

$$G_{\mathrm{MIT}}(\mathbf{r},\mathbf{r}',\omega) = \sum_{l=1} \frac{r_<^l}{2l+1}\left[\frac{1}{r_>^{l+1}} + \frac{l+1}{l}\frac{r_>^l}{R^{2l+1}}\right] Y_{lm}(\Omega_<)Y^*_{lm}(\Omega_>) + \frac{1}{4\pi}\left[\frac{1}{r_>} + \frac{1-\kappa_v}{\kappa_v R}\right]. \tag{5.2}$$

Although the last term is infinite, it is independent of r and r' and does not contribute to the transverse delta function.

From definitions (3.9), we have

$$(r^l)^+ = 0,\quad (r^l)^- = (2l+1)r^{l-1},$$

$$(r^{-(l+1)})^+ = \frac{2l+1}{r^{l+2}},\quad (r^{-(l+1)})^- = 0. \tag{5.3}$$

After putting these into (3.8), the transverse δ function is given by

$$\begin{aligned}\delta^{ii'}_{t,\mathrm{MIT}}(\mathbf{r},\mathbf{r}') = \frac{\delta(r-r')}{r^2}&\Bigg\{\mathcal{Y}^i_{llm}(\Omega)\mathcal{Y}^{i'*}_{llm}(\Omega') + \left(\sqrt{\frac{l}{2l+1}}\,\mathcal{Y}^i_{l,l+1,m}(\Omega) + \sqrt{\frac{l+1}{2l+1}}\,\mathcal{Y}^i_{l,l-1,m}(\Omega)\right)\\ &\times\left(\sqrt{\frac{l}{2l+1}}\,\mathcal{Y}^{i'*}_{l,l+1,m}(\Omega') + \sqrt{\frac{l+1}{2l+1}}\,\mathcal{Y}^{i'*}_{l,l-1,m}(\Omega')\right)\Bigg\}\\ &-\sum_{l=1}^{\infty} r_<^{l-1}\mathcal{Y}^{i_<}_{l,l-1,m}\left[\sqrt{l(l+1)}\,\frac{1}{r_>^{l+2}}\mathcal{Y}^*_{l,l+1,m}(\Omega_>) + \frac{(1+l)r_>^{l-1}}{R^{2l+1}}\mathcal{Y}^*_{l,l-1,m}(\Omega_>)\right]^{i_>}.\end{aligned} \tag{5.4}$$

We see that the transverse δ function for the MIT bag is different from that for free space only in the electric mode. Note that $\overleftrightarrow{\delta}_{t,\mathrm{MIT}}$ is quite simple, so we solve (2.15) directly instead of using (2.17). The Green's function for the MIT bag is then

$$\begin{aligned}G^{ii'}_{\mathrm{MIT}}(\mathbf{r},\mathbf{r}',\omega) = &-\omega\,[j_l(x_<)\mathcal{Y}_{llm}(\Omega_<)]^{i_<}[\tilde{n}_{Ml}(x_>)\mathcal{Y}^*_{llm}(\Omega_>)]^{i_>}\\ &+\omega\,[\nabla\times j_l(x_<)\mathcal{Y}_{llm}(\Omega_<)]^{i_<}\,[\nabla\times\tilde{n}_{El}(x_>)\mathcal{Y}^*_{llm}(\Omega_>)]^{i_>}\\ &+\omega\frac{1}{2l+1}\,[\nabla\times x_<^l\,\mathcal{Y}_{llm}(\Omega_<)]^{i_<}\left[\nabla\times\left(\frac{1}{x_>^{l+1}} - \frac{x_>^l}{X^{2l+1}}\right)\mathcal{Y}^*_{llm}(\Omega_>)\right]^{i_>}.\end{aligned} \tag{5.5}$$

We define $x'=\omega r'$, $x=\omega r$, $X=\omega R$; j_l,n_l are spherical Bessel functions; $\tilde{n}_{El},(x\tilde{n}_{Ml})'$ vanish at $r=R$ to satisfy boundary conditions $\mathbf{B}_{\parallel}=0$ (see the Appendix) and $\mathbf{E}_{\perp}=0$ at $r=R$, respectively. Thus

$$\tilde{n}_{Ml}(x) = n_l(x) - C_{Ml}\,j_l(x),$$

$$\tilde{n}_{El}(x) = n_l(x) - C_{El}\,j_l(x), \tag{5.6}$$

with

$$C_{Ml} = (Xn_l(X))'/(Xj_l(X))',\quad C_{El} = n_l(X)/j_l(X). \tag{5.7}$$

Using relation (4.4), G can be written in the more explicit form

$$\begin{aligned}G^{ii'}_{\mathrm{MIT}}(\mathbf{r},\mathbf{r}',\omega) = &-\omega[j_l(x_<)\mathcal{Y}_{llm}(\Omega_<)]^{i_<}[\tilde{n}_{Ml}(x_>)\mathcal{Y}^*_{llm}(\Omega_>)]^{i_>}\\ &-\omega\left[\sum_{l'=l\pm1} j_{ll'}(x_<)\mathcal{Y}_{ll'm}(\Omega_<)\right]^{i_<}\left[\sum_{l'=l\pm1}(n_{ll'}(x_>) - C_{El}\,j_{ll'}(x_>))\mathcal{Y}^*_{ll'm}(\Omega_>)\right]^{i_>}\\ &+\omega\sqrt{l+1}\,x_<^{l-1}\mathcal{Y}_{l,l-1,m}(\Omega_<)^{i_<}\left[\frac{\sqrt{l}}{x_>^{l+2}}\mathcal{Y}^*_{l,l+1,m}(\Omega_>) + \frac{\sqrt{l+1}\,x_>^{l-1}}{X^{2l+1}}\mathcal{Y}^*_{l,l-1,m}(\Omega_>)\right]^{i_>},\end{aligned} \tag{5.8}$$

where we define

$$\binom{j}{n}_{l,l-1} \equiv \sqrt{\frac{l+1}{2l+1}}\binom{j}{n}_{l-1},$$

$$\binom{j}{n}_{l,l+1} \equiv -\sqrt{\frac{l}{2l+1}}\binom{j}{n}_{l+1}. \tag{5.9}$$

To see the gauge invariance explicitly, we now solve the linearized gluon propagator for the MIT bag in the Lorentz gauge. In the Lorentz gauge, the gluon field A^{μ} satisfies the wave equation

$$-(\nabla^2+\omega^2)A^{\mu}=j^{\mu} \tag{5.10}$$

inside the bag. It also has to satisfy the gauge condition $\partial_{\mu}\mathbf{A}^{\mu}=0$ in the bag, and the boundary conditions $\mathbf{E}_{\perp}=0,\mathbf{B}_{\parallel}=0$ at $r=R$. This gives four conditions, becaue the last equation has two components.

We construct $A^{\mu}=A^{\mu}_{I}+A^{\mu}_{H}$ to satisfy all requirements. The inhomogeneous part A^{μ}_{I} can be chosen to be a particular free space solution. We choose the free space Feynman gauge propagator to construct our inhomogeneous part $A^{\mu}_{I}=A^{\mu}_{F}$. It has the local property $G^{\mu}_{F\nu}(\mathbf{r},t;\mathbf{r}',t')=g^{\mu}_{\nu}G_F(\mathbf{r}-\mathbf{r}',t-t')$. After Fourier transformation to time, it can be written

$$G^{\mu}_{F\nu}(\omega,\mathbf{r},\mathbf{r}')$$
$$=-g^{\mu}_{\nu}\omega\sum_{lm}j_l(\omega r^<)n_l(\omega r^>)Y_{lm}(\Omega)Y_{lm}(\Omega')^*, \tag{5.11}$$

$$G^{00}_F(\omega,\mathbf{r},\mathbf{r}')$$
$$=-\omega\sum_{lm}j_l(\omega r^<)n_l(\omega r^>)Y_{lm}(\Omega)Y_{lm}(\Omega')^*,$$
$$\overleftrightarrow{G}_F(\omega,\mathbf{r},\mathbf{r}')$$
$$=-\omega\sum_{ll'm}j_{l'}(\omega r^<)n_{l'}(\omega r^>)\mathcal{Y}_{ll'm}(\Omega)\mathcal{Y}_{ll'm}(\Omega')^*. \tag{5.12}$$

We can see that $A^{\mu}_F(\mathbf{r},t)=\int d^3r'\,dt'G^{\mu}_{F\nu}(\mathbf{r}-\mathbf{r}',t-t^1)\,j^{\nu}(\mathbf{r}',t')$ satisfies the wave equation (5.10). The gauge condition,

$$\partial_{\mu}A^{\mu}_F(\mathbf{r},t)$$
$$=\int d^3r'\,dt'\,G_F(\mathbf{r}-\mathbf{r}',t-t')\partial_{\mu}'j^{\mu}(\mathbf{r},t')=0,$$

is satisfied because of the local property of the propagator and current conservation.

The homogeneous part A^{μ}_H satisfies the homogeneous equation

$$-(\nabla^2+\omega^2)A^{\mu}_H=0. \tag{5.13}$$

The most general solution to (5.13) can be written as

$$A^{\mu}_H(\mathbf{r},\omega)$$
$$=\sum_{lm}\Big(b^0_l j_l(\omega r)Y_{lm}(\Omega),i\sum_{l'}b_{ll'}j_{l'}(\omega r)\mathcal{Y}_{ll'm}(\Omega)\Big), \tag{5.14}$$

where all $(b^0_l,b_{ll'})$ are constants. We also write the current in spherical harmonics

$$j^{\mu}(\mathbf{r},\omega)=\sum_{lm}\Big(c^0_l(\omega r)Y_{lm}(\Omega),i\sum_{l'}c_{ll'}(\omega r)\mathcal{Y}_{ll'm}(\Omega)\Big). \tag{5.15}$$

Current conservation requires

$$\omega c^0_l(\omega r)+\sqrt{\frac{l+1}{2l+1}}\Big(\frac{d}{dr}+\frac{l+2}{r}\Big)c_{l,l+1}(\omega r)$$
$$+\sqrt{\frac{l}{2l+1}}\Big(-\frac{d}{dr}+\frac{l-1}{r}\Big)c_{l,l-1}(\omega r)=0 \tag{5.16a}$$

inside the bag, and

$$\sqrt{\frac{l+1}{2l+1}}c_{l,l+1}(\omega R)-\sqrt{\frac{l}{2l+1}}c_{l,l-1}(\omega R)=0 \tag{5.16b}$$

on the boundary. With (5.12) and (5.15), we have A^{μ}_F on the boundary

$$A^{\mu}_F(\mathbf{R},\omega)$$
$$=\sum_{lm}\Big(\xi^0_l n_l(\omega R)Y_{lm}(\Omega),i\sum_{l'}\xi_{ll'}n_{l'}(\omega R)\mathcal{Y}_{ll'm}(\Omega)\Big), \tag{5.17}$$

where constants ξ's are defined as

$$\xi^0_l\equiv-\omega\int_0^R r^2\,dr\,j_l(\omega r)c^0_l(\omega r),$$
$$\xi_{ll'}\equiv-\omega\int_0^R r^2\,dr\,j_{l'}(\omega r)c_{ll'}(\omega r). \tag{5.18}$$

After angular decomposition, we obtain four sets of algebraic equations for the $(b^0_l,b_{ll'})$, by putting (5.14) and (5.18) into the equations for the four conditions. For the gauge condition, we have

$$b^0_l+\sqrt{\frac{l+1}{2l+1}}b_{l,l+1}+\sqrt{\frac{l}{2l+1}}b_{l,l-1}=0. \tag{5.19a}$$

For the boundary condition of the $\mathbf{E}$ field,

$$\sqrt{\frac{l+1}{2l+1}}(n_{l+1}(\omega R)\xi_{l,l+1}+j_{l+1}(\omega R)b_{l,l+1})$$
$$-\sqrt{\frac{l}{2l+1}}(n_{l-1}(\omega R)\xi_{l,l-1}+j_{l-1}(\omega R)b_{l,l-1})$$
$$-(n_l(\omega R)'\xi^0_l+j_l(\omega R)'b^0_l)=0. \tag{5.19b}$$

For $\mathbf{B}$ field boundary conditions, we have two equations

$$\sqrt{\frac{l}{2l+1}}(n_l(\omega R)\xi_{l,l+1}+j_l(\omega R)b_{l,l+1})$$
$$-\sqrt{\frac{l+1}{2l+1}}(n_l(\omega R)\xi_{l,l-1}+j_l(\omega R)b_{l,l-1})=0, \tag{5.19c}$$

$$[Rn_l(\omega R)]'\xi_{l,l}+[Rj_l(\omega R)]'b_{l,l}=0. \tag{5.19d}$$

We see that the algebraic equations are decoupled in l and the magnetic mode coefficients b_{ll} decouple from electric mode coefficients $(b^0_l,b_{l\pm1})$. From the current conservation relations (5.16), the first three sets of equations for the three sets of electric modes coefficients can be proved to be only two sets of irreducible equations. Thus, there are infinite sets of solutions in the Lorentz gauge. In Ref. 4, DeGrand and Jaffe state that the solution for $\mathbf{A}$ given in their (B.4,B.6) is unique because it satisfies the boundary condition (B.2). In fact, in the Lorentz gauge, this is not sufficient to fix the gauge. Here we fix the gauge by the special choice $b^0_l=0$. This determines all the $b_{ll'}$, and we obtain A^{μ}. With $G^{\mu\nu}=\delta A^{\mu}/\delta j_{\nu}$, we have the linearized gluon propagator in the Lorentz gauge

$$G_0^0(\omega,\mathbf{r},\mathbf{r}') = -\omega \sum_{lm} j_l(\omega r^<) n_l(\omega r^>) Y_{lm}(\Omega^<) Y_{lm}(\Omega^>)^*$$

$$\overleftrightarrow{G}(\omega,\mathbf{r},\mathbf{r}') = -\omega \sum_{l'} j_{l'}(\omega r^<) n_{l'}(\omega r^>) \mathscr{Y}_{ll'm}(\Omega^<) \mathscr{Y}_{ll'm}(\Omega^>)^* + \omega C_{Ml} j_l(\omega r) \mathscr{Y}_{llm}(\Omega) j_l(\omega r') \mathscr{Y}_{llm}(\Omega')^*$$
$$+ \omega C_{El} \left(\sqrt{\frac{l}{2l+1}} j_{l+1} \mathscr{Y}_{l,l+1} - \sqrt{\frac{l+1}{2l+1}} j_{l-1} \mathscr{Y}_{l,l-1} \right)^< \times \left(\sqrt{\frac{l}{2l+1}} j_{l+1} \mathscr{Y}_{l,l+1} - \sqrt{\frac{l+1}{2l+1}} j_{l-1} \mathscr{Y}_{l,l-1} \right)^{>*} \quad (5.20)$$

here the C_{Ml}, C_{El} are the same as in the Coulomb gauge. The sum over l and m is implicit. We can see that the magnetic mode of the propagator in the Lorentz gauge is exactly the same as in the Coulomb gauge.

VI. GAUGE INVARIANCE

We begin this section with a review[9] of gauge invariance, but for an inhomogeneous static dielectric medium ($\kappa = \mu^{-1}$). The inhomogeneous Maxwell's equations are

$$\partial^\mu \kappa(\mathbf{r}) F_{\mu\nu} = J_\nu, \quad (6.1a)$$

and the homogeneous equations are

$$\epsilon^{\mu\nu\rho\sigma} \partial_\mu F_{\rho\sigma} = 0, \quad (6.1b)$$

where

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}. \quad (6.2a)$$

Introduction of the four-vector potential $A^\mu = (A_0, \mathbf{A})$, with

$$\mathbf{B} = \nabla\times\mathbf{A}, \quad \mathbf{E} = -\nabla A_0 - \partial_t \mathbf{A}, \quad (6.2b)$$

guarantees the satisfaction of the homogeneous equations. The A^μ must then satisfy [cf. Eq. (2.1)]

$$\partial^\mu \kappa(\mathbf{r}) [\partial_\mu A_\nu - \partial_\nu A_\mu] = J_\nu. \quad (2.1)$$

The A^μ are not unique. Note that $F_{\mu\nu}$ remains invariant under the transformation

$$A_0 \to A_0' = A_0 + \partial_t \chi,$$
$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} - \nabla\chi. \quad (6.3)$$

Restrictions on the A^μ determine a choice of gauge. The Coulomb gauge condition is

$$\nabla\cdot\kappa\mathbf{A}_C = 0. \quad (6.4)$$

The Lorentz gauge condition is

$$\nabla\cdot\kappa\mathbf{A}_L + \kappa\,\partial_t A_{0L} = 0. \quad (6.5)$$

Different potentials obtained from the same current are connected by a gauge transformation. We now display the gauge transformation that relates the Coulomb and Lorentz gauges. Let A^μ_L be any Lorentz gauge potential. Then choose χ_{CL} to satisfy

$$\nabla\cdot\kappa\nabla\chi_{CL} = \kappa\,\partial_t A_{0L}. \quad (6.6)$$

The transformed potential

$$A'_{0C} = A_{0L} + \partial_t \chi_{CL},$$
$$\mathbf{A}'_C = \mathbf{A}_L - \nabla\chi_{CL}, \quad (6.7)$$

clearly satisfies the Coulomb gauge condition. We see that if $\partial_t A_{0L} = 0$, χ_{CL} can be chosen zero. When χ_{CL} is zero, the two gauges have the same A^μ field. Note that A_0 is the longitudinal electric part, so $A_0 = 0$ in the magnetic (TE) mode. That is why for static current, or for the magnetic (TE) mode, the two gauges can have the same A^μ field and propagator as we discussed in Sec. V.

Second, we discuss uniqueness. First consider free space. In the Coulomb gauge, if two solutions A^μ, A'^μ satisfy the wave equation, the gauge transformation function χ_C between them must satisfy $\nabla^2\chi_C = 0$. The only solution that is regular everywhere is a constant. So the solution for the Coulomb gauge in free space is unique up to a constant. In Sec. V, we found that the solution in the MIT bag is also unique up to a constant. In the Lorentz gauge (Fourier transformed in time), the gauge transformation function satisfies $(\omega^2 + \nabla^2)\chi_L = 0$. If ω is not zero there will be an infinity of nontrivial regular solutions, e.g., $j_l(\omega r) Y_{lm}$. In Sec. V, we displayed this for the MIT bag.

Now, let us look into the Feynman gauge ($G^{\mu\nu} = g^{\mu\nu} G$), which is one of the Lorentz gauges. It was proffered in Ref. 6 for the MIT bag. In free space, we have exhibited that the Feynman solution (5.11) automatically satisfies the Lorentz gauge condition. But in the dielectric medium, this is not possible. As a special example, consider the MIT bag. If the propagator is written in the Feynman form,[6]

$$G^{\mu\nu}_{\rm MIT} = g^{\mu\nu} \left[G_F + \sum_{lm} c_l j_l(\omega r) j_l(\omega r') Y_{lm}(\Omega) Y^*_{lm}(\Omega') \right], \quad (6.8)$$

there is only one parameter for each l mode. It is impossible to satisfy the four conditions that give three irreducible algebraic equations for each l mode as given in Sec. V, and we find that the solution in Ref. 6 does not satisfy the boundary condition $\mathbf{r}\times\mathbf{B} = 0$ (see the Appendix). The problem comes from forcing the homogeneous part to be of the Feynman form: Although the Feynman form $g^{\mu\nu} j_l(\omega r) j_l(\omega r') Y_{lm}(\Omega) Y^*_{lm}(\Omega')$ is a solution of the homogeneous equation $-(\omega^2 + \nabla^2) G^{\mu\nu}_H = 0$, it is not the most general one; the mode constant c_l could depend on μ and ν.

Now let us look at gauge invariance for one gluon interaction in the MIT bag, with the propagator given in both Coulomb and Lorentz gauges in Sec. V. The magnetic (TE) modes in both gauges will surely have the same contribution since they have the same propagator. The OGE magnetic mode matrix element is given by

$$H_M = -\omega \int c_{ll}(\omega r^<) j_l(\omega r^<) (n_l(\omega r^>)$$
$$- C_{Ml} j_l(\omega r^>)) c_{ll}(\omega r^>). \quad (6.9)$$

The contribution of electric (TM) modes will now be proved to be the same in both gauges. With the current conservation equations (5.16), we have the electric mode Hamiltonian for both Coulomb and Lorentz gauges:

$$H_E = \omega \int \left(\sqrt{\frac{l}{2l+1}} j_{l+1}(\omega r_<) c_{l,l+1}(\omega r_<) - \sqrt{\frac{l+1}{2l+1}} j_{l-1}(\omega r_<) c_{l,l-1}(\omega r_<) \right)$$
$$\times \left(\sqrt{\frac{l}{2l+1}} n_{l+1}(\omega r_>) c_{l,l+1}(\omega r_>) - \sqrt{\frac{l+1}{2l+1}} n_{l-1}(\omega r_>) c_{l,l-1}(\omega r_>) \right)$$
$$- \omega C_{El} \int \left(\sqrt{\frac{l}{2l+1}} j_{l+1}(\omega r_<) c_{l,l+1}(\omega r_<) - \sqrt{\frac{l+1}{2l+1}} j_{l-1}(\omega r_<) c_{l,l-1}(\omega r_<) \right)$$
$$\times \left(\sqrt{\frac{l}{2l+1}} j_{l+1}(\omega r_>) c_{l,l+1}(\omega r_>) - \sqrt{\frac{l+1}{2l+1}} j_{l-1}(\omega r_>) c_{l,l-1}(\omega r_>) \right)$$
$$+ \int \frac{r^2\, dr}{\omega^2} \left(\sqrt{\frac{l}{2l+1}} c_{l,l-1}(\omega r) - \sqrt{\frac{l+1}{2l+1}} c_{l,l+1}(\omega r) \right)^2. \tag{6.10}$$

All the notations $c_{ll'}$ for the current above depend on the initial and final states of the quark, which have been suppressed for simplification. Here we calculate the charge and current density for quark states $(\kappa, \kappa', \mu, \mu')$. In the MIT bag, the quark wave function can be written

$$\psi = \begin{pmatrix} u_\kappa(r) \\ i\sigma_r v_\kappa(r) \end{pmatrix} \mathscr{Y}_{\kappa\mu}(\Omega). \tag{6.11}$$

The charge density is

$$\rho(\mathbf{r},\omega) = \psi^\dagger_{\kappa'\mu'} \psi_{\kappa\mu} = c_{lm}(\omega r, \kappa\kappa'\mu\mu') Y_{lm}(\Omega), \tag{6.12}$$

where

$$c_{lm}(\omega r, \kappa\kappa'\mu\mu') = (-1)^{1/2-\mu} (u_\kappa u_{\kappa'} + v_\kappa v_{\kappa'}) \sqrt{\frac{(2l_\kappa+1)(2l_{\kappa'}+1)(2j_\kappa+1)(2j_{\kappa'}+1)(2l+1)}{4\pi}}$$
$$\times \begin{pmatrix} l_\kappa & l_{\kappa'} & l \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} j_\kappa & j_{\kappa'} & l \\ -\mu & \mu' & m \end{pmatrix} \begin{Bmatrix} j_\kappa & j_{\kappa'} & l \\ l_{\kappa'} & l_\kappa & \frac{1}{2} \end{Bmatrix}. \tag{6.13}$$

The current density is

$$\mathbf{j}(\mathbf{r},\omega) = \psi^\dagger_{\kappa'\mu'} \boldsymbol{\alpha} \psi_{\kappa\mu} = i c_{ll'm}(\omega r, \kappa\kappa'\mu\mu') \mathscr{Y}_{ll'm}(\Omega), \tag{6.14}$$

where[10]

$$c_{ll'm}(\omega r, \kappa\kappa'\mu\mu') = (-1)^{\mu+l+l'} \sqrt{\frac{(2l_\kappa+1)(2l'_\kappa+1)(2j_\kappa+1)(2j_{\kappa'}+1)(2l+1)(2l'+1)}{\pi}}$$
$$\times \Bigg[3(-1)^{l_{\kappa'}+j_\kappa} (v_\kappa u_{\kappa'} + u_\kappa v_{\kappa'}) \sum_{J'} (2J'+1) \begin{pmatrix} l' & 1 & J' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_\kappa & l_{\kappa'} & J' \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times \begin{Bmatrix} l & J' & 1 \\ 1 & 1 & l' \end{Bmatrix} \begin{Bmatrix} j_\kappa & j_{\kappa'} & l \\ l_\kappa & l_{\kappa'} & J' \\ \frac{1}{2} & \frac{1}{2} & 1 \end{Bmatrix} - \frac{1}{2} (-1)^{j_\kappa + j_{\kappa'} + l + 1/2} (v_\kappa u_{\kappa'} - u_\kappa v_{\kappa'})$$
$$\times \begin{pmatrix} l' & 1 & l \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_{\kappa'} & l & l_\kappa \\ 0 & 0 & 0 \end{pmatrix} \begin{Bmatrix} l_\kappa & l & \frac{1}{2} \\ j_{\kappa'} & l_{\kappa'} & l \end{Bmatrix} \Bigg]. \tag{6.15}$$

VII. NUMERICAL IMPLEMENTATION

We have calculated the one gluon matrix elements for the MIT bag in the Coulomb gauge and in a particular Lorentz gauge. We also compared our results with the work of Wroldsen and Myher,[5] who work in a Lorentz gauge without solving for the gluon propagator. We found three-way agreement for the following $q_1 q_2 \to q_1' q_2'$ elements: $ss \to ss$, $sp_{1/2} \to sp_{1/2}$, $p_{1/2} p_{1/2} \to p_{1/2} p_{1/2}$, $sp_{3/2} \to sp_{3/2}$, $p_{3/2} p_{3/2} \to p_{3/2} p_{3/2}$, $sp_{1/2} \to p_{1/2} s$, $sp_{3/2} \to p_{3/2} s$.

We also used the Perry's propagator given in Ref. 6 and the same charge and current density given in (6.14),(6.15) to calculate the matrix elements; we found disagreement except for the electric monopole terms. This is consistent with our arguments above.

VIII. CONCLUSION

In the Coulomb gauge, we have obtained the scalar (time–time) and tensor (space–space) gauge field propagator in a general dielectric medium $\kappa(\mathbf{r})$ $(\mu = \kappa^{-1})$, supplementing and correcting the results of Ref. 1. We found that there is a subtle error in the transverse current of Ref. 1. A general transverse tensor projector is formulated utilizing the scalar propagator, which is $\kappa(\mathbf{r})$-dependent. We recover the usual transverse dyadic delta function when $\kappa = 1$ everywhere. We have found a formula to construct the correct tensor propagator from the tensor propagator in Ref. 1 and the transverse projector. The general transverse projector is calculated explicitly in terms of vector spherical harmonics. We also have corrected a sign error in the scalar propagator that appears for nonaxially symmetric media. In the case of a spherically symmetric medium, explicit expressions for the tensor propagator are given. In the MIT bag model, the propagator is calculated in both Coulomb and Lorentz gauges, and gauge invariance is shown explicitly for one gluon interaction.

The gauge field propagator is very useful for obtaining OGE matrix elements in various bag models. We also plan to use it to obtain the nonlocal self-energy in a chromodielectric model[2] by solving the Schwinger–Dyson equation with the (nonlocal) gluon propagator.

ACKNOWLEDGMENTS

We wish to thank Dr. Fl. Stancu, Dr. F. Myhrer, Dr. A. G. Williams, and Dr. G. Krein for useful discussions.

This work was supported in part by the U. S. Department of Energy.

APPENDIX: BOUNDARY CONDITION FOR THE MAGNETIC MODE

In the magnetic (TE) mode, the vector field is

$$\mathbf{A}_{\mathrm{TE}}(\mathbf{r},\omega) = -i\omega \sum_{lm} f_{ll}(\omega r)\mathscr{Y}_{ll}(\Omega), \tag{A1}$$

for which $\mathbf{r}\cdot\mathbf{A} = 0$ holds everywhere; it automatically satisfies the boundary condition of $\mathbf{r}\cdot\mathbf{E}_{r=R} = 0$. Since it does not couple to the electric modes, its boundary condition can be considered separately. The remaining constraint for the magnetic mode is

$$\mathbf{r}\times\mathbf{B}|_{r=R} = 0. \tag{A2}$$

The transverse part of current which couples to this mode can be written as

$$\mathbf{j}_{\mathrm{TE}}(\mathbf{r},\omega) = i\sum_{lm} c_{ll}(\omega r)\mathscr{Y}_{ll}(\Omega), \tag{A3}$$

which satisfies $\mathbf{r}\cdot\mathbf{j} = 0$; the c_{ll} is defined in (6.15). For either formulations of the propagator G^i_j, namely $-\omega g^i_j j_l(\omega r_<)\tilde{n}_l(\omega r_<)Y_{lm}(\Omega)Y^*_{lm}(\Omega')$ in Ref. 6 and $-\omega j_l(\omega r_<)\tilde{n}_l(\omega r_<)\mathscr{Y}_{llm}(\Omega)\mathscr{Y}^*_{llm}(\Omega')$ in Eqs. (5.5) and (5.20), one has $\mathbf{A}_{TE} = \int G\cdot\mathbf{j}_{TE}$ in the form of (A1). (One should note that for some special transitions, e.g., $ss - ss$, $ss - p_{1/2}p_{1/2},\ldots$, only the magnetic mode contributes to $\mathbf{A}$ since the total current is transverse.) The f_{ll} in (A1) can be written as

$$f_{ll}(\omega r) = \tilde{n}_l(\omega r)\int_0^r dr'\, r'^2 j_l(\omega r)c_{ll}(\omega r') + j_l(\omega r)\int_r^R dr'\, r'^2\tilde{n}_l(\omega r')c_{ll}(\omega r'). \tag{A4}$$

The boundary condition (A2) can be written

$$\begin{aligned}\mathbf{r}\times(\nabla\times\mathbf{A}) &= \hat{k}(r_i\,\partial_k A_i - r_i\,\partial_i A_k)\\ &= \hat{k}(\partial_k(r_iA_i) - (\partial_k r_i)A_i - r_i\,\partial_i A_k)\\ &= -\mathbf{A} - r\frac{\partial}{\partial r}\mathbf{A} = -\frac{\partial}{\partial r}(r\mathbf{A});\end{aligned} \tag{A5}$$

In the third step, $\mathbf{r}\cdot\mathbf{A} = 0$ has been used. So in radial form, the boundary condition is

$$\frac{\partial}{\partial r}(rf_{ll}(\omega r))|_{r=R} = 0. \tag{A6}$$

inserting f_{ll} from (A4) into (A5) we have

$$\begin{aligned}\frac{\partial}{\partial r}(rf_{ll}(\omega r)) &= \frac{\partial}{\partial r}[r\tilde{n}_l(\omega r)]\left[\int_0^r dr'\, r'^2 j_l(\omega r')c_{ll}(\omega r')\right] + r\tilde{n}_l(\omega r)r^2 j_l(\omega r)c_{ll}(\omega r) + \frac{\partial}{\partial r}[rj_l(\omega r)]\\ &\times\left[\int_r^R dr'\, r'^2\tilde{n}_l(\omega r')c_{ll}(\omega r')\right] - rj_l(\omega r)r^2\tilde{n}_l(\omega r)c_{ll}(\omega r) \underset{r\to R}{\to} \left[\int_0^R dr'\, r'^2 j_l(\omega r')c_{ll}(\omega r')\right]\frac{\partial}{\partial R}(R\tilde{n}_l(\omega R)).\end{aligned} \tag{A7}$$

This requires

$$\frac{\partial}{\partial R}(R\tilde{n}_l(\omega R)) = 0. \tag{A8}$$

Here $\tilde{n}$ is given by (5.6) for the Coulomb gauge, or by (5.20) for our Lorentz gauge. One finds in both cases

$$\tilde{n}_l(\omega r) = n_l(\omega r) - \frac{(Rn_l(\omega R))'}{(Rj_l(\omega R))'}j_l(\omega r), \tag{A9}$$

which does satisfy (A8). However, Perry[6] gives

$$\tilde{n}_l(\omega r) = n_l(\omega r) - \frac{(n_l(\omega R))'}{(j_l(\omega R))'}j_l(\omega r), \tag{A10}$$

which gives, for the lhs of (A8),

$$\frac{\partial}{\partial R}(R\tilde{n}_l(\omega R)) = -\frac{1}{R^2 j(\omega R)'}\neq 0, \tag{A11}$$

which does not satisfy the boundary condition.

[1] Cf. G. Fai, R. J. Perry, and L. Wilets, Phys. Lett B **208**, 1 (1988); G. Krein, P. Tang, L. Wilets, and A. G. Williams, Phys. Lett. B **212**, 362 (1988); G. Krein, P. Tang, L. Wilets, and A. G. Williams, submitted to Phys. Rev. D.
[2] W. A. Johnson, A. Q. Howard, and D. G. Dudley, Radio Sci. **14**, 961 (1979).
[3] M. Bickeböller, R. Goldflam, and L. Wilets, J. Math. Phys. **26**, 1810 (1985).
[4] T. A. DeGrand and R. L. Jaffe, Ann. Phys. **100**, 425 (1976).
[5] J. Wroldsen and F. Myhrer, Z. Phys. C **25**, 59 (1984).
[6] R. J. Perry, Phys. Rev. D **28**, 2099 (1983).
[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957), Chap. 5.
[8] Cf. L. Wilets, *Nontopological Solitons* (World Scientific, Singapore, 1989), Chap. 6.
[9] Cf. J. D. Jackson, *Classical Electrodynamics* (Wiley, New York), Chaps. 6 and 11.
[10] Fl. Stancu, P. Tang, and L. Wilets (private communications).

Refined asymptotic expansion for the partition function of unbounded quantum billiards

F. Steiner and P. Trillenberg
II. Institut für Theoretische Physik, Universität Hamburg, Luruper Chaussee 149, 2000 Hamburg 50, Federal Republic of Germany

(Received 7 August 1989; accepted for publication 21 February 1990)

This paper presents a refined asymptotic expansion for the partition function $\Theta(t) = \mathrm{Tr}\, e^{t\Delta}$ of quantum billiards in the unbounded regions $\{0<x, 0<yx^\mu<1\}$, $\mu>0$, and $\{0<ye^{|x|}<1\}\subset\mathbb{R}^2$, where Δ is the Dirichlet Laplacian. Simon [Ann. Phys. **146**, 209 (1983); J. Funct. Anal. **53**, 84 (1983)] determined the leading divergence of the trace of the heat kernel for the first class of systems. Standard techniques are combined for the evaluation of Θ for bounded region billiards with results by Van den Berg [J. Funct. Anal. **71**, 279 (1987)] for "horn-shaped regions" using an optimized way of dividing the region into "narrow" and "wide" parts to determine the first three terms in the asymptotic expansion of Θ. Results are also stated for bounded regions with cusps that can be obtained by the same method. As an application, the spectral staircase of the strongly chaotic billiard system defined in the region $\{0<xy<2, x>0\}$, which has been discussed in connection with the Riemann ζ function and the search for quantum chaos is considered.

I. INTRODUCTION

Billiards (i.e., a point particle of mass m sliding freely in a region $\Omega\subset\mathbb{R}^2$, but being prevented from leaving the region by an infinitely high potential wall) form a large class of systems providing examples of all kinds of regular and chaotic motion. If we put $\hbar=1$ and $m=\frac{1}{2}$, their quantum Hamiltonian is $-\Delta$ (where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the Dirichlet Laplacian for Ω), and *Schrödinger's equation* reads:

$$(\Delta+\lambda)\psi(z)=0. \tag{1}$$

[We denote points $(x,y)\in\mathbb{R}^2$ by z]. This is also Helmholtz's equation that describes vibrating membranes.

If $|\Omega|<\infty$, the spectrum of (1) is purely discrete; let λ_n be the nth eigenvalue of the Schrödinger equation (1), then the *partition function* $\Theta(t)$ of the billiard is defined as

$$\Theta(t) := \mathrm{Tr}\, e^{t\Delta} = \sum_{n=1}^{\infty} e^{-\lambda_n t} = \int_\Omega d^2z\, G_\Omega(t|z,z), \quad t>0, \tag{2}$$

where $G_\Omega(t|z,z')$ is Green's function of $\Delta-(\partial/\partial t)$ (Dirichlet heat kernel). The partition function of quantum billiards (or membranes) has attracted extensive attention because its behavior for $t\searrow 0$ is intimately connected with the geometry of the region considered:

$$\Theta(t) \sim \frac{|\Omega|}{4\pi t} - \frac{L(\partial\Omega)}{8\sqrt{\pi t}} + \frac{1}{6}, \quad t\searrow 0. \tag{3}$$

The last result holds for simply connected domains with area $|\Omega|$ and smooth boundary $\partial\Omega$ of length $L(\partial\Omega)$. Three more terms are given in Refs. 1 and 2; corrections for boundaries with edges and cusps are discussed in Refs. 3 and 2, respectively.

Apart from the question of inferring the billiard geometry from Eq. (3) ("Can one hear the shape of a drum?", [Refs. 1 and 4], Eq. (3) can be used to obtain an asymptotic expression for $\langle N(\lambda)\rangle$, $\lambda\to\infty$, with $\langle\cdots\rangle$ denoting an averaging process and $N(\lambda) := \#\{n|\lambda_n\leqslant\lambda\}$ ("spectral staircase").

Under certain assumptions, billiards in unbounded regions also have a purely discrete spectrum (see Refs. 5 and 6), but even if $\Theta(t)<+\infty$ $\forall t>0$ (for a criterion see again Ref. 5), the expansion (3) has to be modified.

In Ref. 5 Davies derived bounds on $\Theta(t)$ for very general domains $\Omega\subset\mathbb{R}^n$; however, they reveal only the first term of the asymptotic expansion of $\Theta(t)$. The next term can be obtained by an approximation due to Van den Berg[7] in the case of "horn-shaped" regions:

$$\mathbb{R}^2\supset\Omega \text{ horn-shaped:} \Leftrightarrow \Omega = \{-g(x)\leqslant y\leqslant f(x)\},$$
$$f,g:\mathbb{R}\to\mathbb{R}_+ \text{ decreasing for } x>0, \text{ increasing for } x<0$$
$$\text{with } \lim_{|x|\to\infty} f(x) = \lim_{|x|\to\infty} g(x) = 0.$$

In general, his method is precise up to order $O(1/\sqrt{t})$, which suffices for two terms in the examples studied here. His approximation for the partition function is [with $h(x) := f(x)+g(x)$]:

$$\frac{1}{\sqrt{4\pi t}}\int_{-\infty}^{\infty} dx \sum_{n=1}^{\infty} \exp\left(-\frac{n^2\pi^2 t}{h^2(x)}\right),$$

and it can be obtained by replacing the trace in $\mathrm{Tr}\, e^{t\Delta}$ by a classical phase space integral *after* having applied a Born–Oppenheimer approximation [i.e., having introduced potentials $V_n(x) := n^2\pi^2/f^2(x)$ given by the eigenvalues of the one-dimensional billiard problem for x fixed].

We shall—in addition to hornshapeness—assume that the functions f,g are convex on $(-\infty,0)$, $(0,\infty)$, and deter-

 0022-2488/90/071670-07$03.00

mine the third term of the asymptotic expansion of $\Theta(t)$ for the above-mentioned regions by a combination of Van den Berg's method with standard techniques for bounded-region billiards. In summary, our method can be described as follows.

(i) Integration of $G_\Omega(t|zz)$ over the region $\Omega(t) := \{z \in \Omega | \exists \text{ circle } K : z \in K \subset \Omega;\ \text{diam } (K) = t^{1/2-\epsilon}\}$ $(\epsilon > 0)$ with classical methods. The boundary condition influences (roughly speaking) only the points within a distance of $\sqrt{t}$ from $\partial\Omega$, therefore most of the points of the above mentioned set are not affected by the boundary. For them we can use the free heat kernel with small corrections ("Kac's principle of not feeling the boundary."[4,7] To obtain an upper bound, Lemma 3 in Ref. 7 can be used; for the lower bound error terms vanishing as $t \searrow 0$ stronger than any power of t are required. They can be supplied by inscribing certain squares Q into Ω and using $G_\Omega(t|zz) \geqslant G_Q(t|zz)$.

(ii) Integration of $G_\Omega(t|zz)$ over the region $\Omega \setminus \Omega(t)$ with Van den Berg's formula,[7] quoted in Eq. (10) below. The points of this domain lie between two nearly parallel parts of $\partial\Omega$ ("horn") with distance smaller than $t^{1/2-\epsilon}$, and they are strongly affected by both of these parts.

Our paper is organized as follows: In Sec. II we state our results for unbounded regions in terms of three theorems. In Sec. III we give the proofs of Theorems 1–3. Performing the integration mentioned under (ii) we find that the approximation is precise up to order $t^{-\epsilon}$; in the first example there is a competing error term of order $t^{-(1/2-\epsilon)}$ [due to $\Omega(t)$] requiring the choice $\epsilon = \frac{1}{4}$; a similar situation (error terms of order $t^{-\epsilon}$ and $t^{-(1/\mu)(1/2-\epsilon)}$) occurs in the second example. In the third example (Theorem 3) any choice $\epsilon > 0$ is allowed. In Sec. IV we present two results for regions with cusps, which can be obtained by calculations very similar to the ones performed in Sec. III, and which generalize results by Stewartson and Waechter.[2] Finally, in Sec. V we use the results from Sec. III to obtain information on a "smoothed" spectral staircase for the billiard region $\{0 \leqslant xy \leqslant 2\} \cap \mathbb{R}^2_+$ and comment on a remark by Berry[8] on a hypothetical connection of the spectrum of this system with the imaginary parts of the nontrivial zeroes of the Riemann ζ function.

II. RESULTS FOR UNBOUNDED REGIONS

Our results are formulated in the following three theorems.

Theorem 1: Let Δ be the Dirichlet Laplacian for $\Omega = \{(x,y) \in \mathbb{R}^2_+ \ |0 \leqslant xy \leqslant 1\}$, then for $t \searrow 0$,

$$\Theta(t) = \text{Tr}\, e^{t\Delta} = -\frac{\log t}{4\pi t} + \frac{A}{4\pi t} - \frac{B}{8\sqrt{\pi t}} + O(t^{-1/4}), \tag{4}$$

where

$$A = 1 + \gamma - 2\log(2\pi) = -2.0985\cdots,$$
$$B = -4[\pi^{3/2}/\Gamma^2(\tfrac{1}{4})] = -1.6944\cdots, \tag{5}$$

where γ denotes Euler's constant.

For an early reference about this system, see, e.g., Ref. 9. Note that Van den Berg's method,[7,10] suffices to determine A, though he did not exploit this. The coefficient A could also be found by a careful Dirichlet–Neumann bracketing[11] following the lines of Simon's proof 3 in Ref. 12.

There is no obvious interpretation of the first two terms (as there is for the classical case.[1,2,4]) Here, B is (in a short but very informal phrasing) the difference between the length of the hyperbola and the length of two half-axes, and the remaining factor is that one that is expected for perimeter corrections.

Theorem 2: Let Δ be the Dirichlet Laplacian for $\Omega = \{(x,y) \in \mathbb{R}^2_+ \ |0 \leqslant yx^\mu \leqslant 1\}$, $1 \neq \mu > 0$, then for $t \searrow 0$,

$$\Theta(t) = \text{Tr}\, e^{t\Delta} = \frac{1}{t^{(1/2)(1+\mu)}} \frac{\Gamma(1+\mu/2)\zeta(\mu)}{2\pi^{\mu+1/2}} - \frac{H_\mu - 1}{8\sqrt{\pi t}} + \left(\mu \leftrightarrow \frac{1}{\mu}\right) + O(t^{-1/2(\mu+1)} + t^{-\mu/2(\mu+1)}), \tag{6}$$

where

$$H_\mu := \int_1^\infty (\sqrt{1 + x^{-2(\mu+1)}} - 1)\, dx$$
$$= 1 - \sqrt{2} + 2^{\beta - 1/2}\Gamma(\tfrac{1}{2} + \beta)\mathscr{P}^{-1/2-\beta}_{-1/2-\beta}(\sqrt{2}),$$
$$\beta = \mu/2(\mu + 1).$$

Let $\zeta(\mu)$ denote the Riemann zeta function and $\mathscr{P}^a_b(z)$ the associated Legendre functions of the first kind.

Note that the limit $\mu \to 1$ in Eq. (6) exists and leads back to Theorem 1. Theorem 2 refines the expansion given in Ref. 13 by Simon.

Theorem 3: Let Δ be the Dirichlet Laplacian for $\Omega = \{(x,y) \in \mathbb{R}^2 | 0 \leqslant y \leqslant e^{-|x|}\}$, then for $t \searrow 0$,

$$\Theta(t) = \text{Tr}\, e^{t\Delta} = \frac{|\Omega|}{4\pi t} + \frac{2\log t}{8\sqrt{\pi t}} - \frac{B'}{8\sqrt{\pi t}} + O(t^{-\epsilon}), \tag{7}$$

(for any ϵ with $0 < \epsilon < \frac{1}{2}$), where

$$B' = 2(3\log 2 - 1 + \sqrt{2} - \log(1 + \sqrt{2}) - \gamma) = 2.0701\cdots. \tag{8}$$

The last billiard system was introduced in Ref. 7, and Van den Berg determined the two leading terms. Note that the first term is analogous to the standard situation, whereas the second and the third term lack any simple interpretation.

Unfortunately, we have not been able to determine the t^0 term in the expansions (4), (6), and (7), which in the case of bounded regions reflects a topological feature of the billiard region. As we already mentioned, the $1/\sqrt{t}$ contribution in Theorems 1 and 2 is due only to length contributions. This could indicate that the subsequent terms follow the standard pattern given in Ref. 2.

III. PROOFS

Proof of Theorem 1: The proof is given in three steps corresponding to the integrations over $\Omega \setminus \Omega(t)$, $\Omega(t)$, Ω_3: For t given with $0 < t < 1$ let the sets $\Omega(t)$, Ω_i be as follows:

$$\Omega(t)=\{(x,y)\,|\,x<t^{-1/4}\wedge y<t^{-1/4}\}\cap\Omega,$$

$$\Omega_1(t)=\{(x,y)\,|\,\mathrm{dist}(z,\{xy=1\})<\tfrac{1}{4}t^{1/4}\}\cap\Omega(t),$$

$$\Omega_3=\{(x,y)\,|\,x,y<\tfrac{1}{2}\}\cap\Omega,$$

$$\Omega_2(t)=\{(x,y)\,|\,\mathrm{dist}(z,\{xy=0\})<\tfrac{1}{4}t^{1/4}\}\cap\Omega(t)\setminus\Omega_3.$$

[We will sometimes omit the t dependence of $\Omega_1(t),\Omega_2(t)$; the sets are illustrated in Fig. 1.]

Step 1: integration over $\Omega\setminus\Omega(t)$:

$$\int_{\Omega\setminus\Omega(t)} d^2z\, G_\Omega(t\,|zz) = -\frac{\tfrac{1}{2}\log t}{4\pi t}+\frac{\gamma-2\log(2\pi)}{4\pi t}+\frac{1}{t^{3/4}\sqrt{4\pi}}+O(t^{-1/4}),\quad t\searrow 0. \tag{9}$$

Proof of Eq. (9): Theorem 3 in Ref. 7 states for $f:[x_0,\infty)\to\mathbb{R}_+$ decreasing and $D\subset\mathbb{R}^2$ with $D\cap\{x\geqslant x_0\}=\{(x,y)\in\mathbb{R}^2|0\leqslant y\leqslant f(x),\,x\geqslant x_0\}=D'$,

$$-\frac{f(x_0)}{4\sqrt{\pi t}}\leqslant\int_{D'} d^2z\, G_D(t\,|zz)-\frac{1}{\sqrt{4\pi t}}\int_{x_0}^{\infty}dx\,\Theta(t,x)\leqslant\frac{f(x_0)}{8\sqrt{\pi t}}, \tag{10}$$

where $\Theta(t,x):=\Sigma_{n=1}^{\infty}\exp(-\pi^2n^2t/f^2(x))$. We choose $x_0=t^{-1/4}$ and $f(x)=1/x$. The integral over $\Theta(t,x)$ can be carried out after use of the transformation formula of the Jacobi theta function:

$$\begin{aligned}\int_{t^{-1/4}}^{\infty}dx\sum_{n=1}^{\infty}\exp(-\pi^2n^2x^2t) &=\frac{1}{\sqrt{\pi t}}\int_{t^{1/4}\sqrt{\pi}}^{\infty}dx\sum_{n=1}^{\infty}\exp(-\pi n^2x^2)\\ &=\frac{1}{\sqrt{\pi t}}\int_{t^{1/4}\sqrt{\pi}}^{1}dx\,\frac{1}{2}\left(\frac{1+2\sum_{n=1}^{\infty}\exp(-\pi n^2/x^2)}{x}-1\right)\\ &\quad+\frac{1}{\sqrt{\pi t}}\int_{1}^{\infty}dx\sum_{n=1}^{\infty}\exp(-\pi n^2x^2)\\ &=-\frac{\tfrac{1}{4}\log(t\pi^2)}{\sqrt{4\pi t}}-\frac{1-t^{1/4}\sqrt{\pi}}{2\sqrt{\pi t}}\\ &\quad+\frac{1}{\sqrt{4\pi t}}\left(1+\frac{\gamma}{2}-\tfrac{1}{2}\log(4\pi)\right)\\ &\quad-\frac{1}{\sqrt{\pi t}}\int_0^{t^{1/4}\sqrt{\pi}}\frac{dx}{x}\sum_{n=1}^{\infty}\exp\left(-\frac{\pi n^2}{x^2}\right).\end{aligned} \tag{11}$$

Here we have employed

$$\int_1^{\infty}\frac{du}{u}(1+\sqrt{u})\sum_{n=1}^{\infty}e^{-n^2\pi u}=1+\frac{\gamma}{2}-\tfrac{1}{2}\log(4\pi). \tag{12}$$

This can be obtained from

$$\zeta(z)=\frac{\pi^{z/2}}{\Gamma(z/2)}\left\{\frac{1}{z(z-1)}+\int_1^{\infty}\frac{du}{u}\sum_{n=1}^{\infty}e^{-n^2\pi u}(u^{(1-z)/2}+u^{z/2})\right\} \tag{13}$$

(c.f. Ref. 14, p. 21) in the limit $z\to 1$.

For x with $0<x<\pi/\sqrt{\log 2}$,

$$\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi}{x^2}\right)\leqslant 2e^{-\pi/x^2}\leqslant 2x^3e^{-\pi/2x^2},$$

which gives the required bound for the last term in (11).

Step 2: integration over $\Omega(t)\setminus\Omega_3$:

Let $H_1=\int_1^{\infty}(\sqrt{1+1/x^4}-1)dx$, then

$$\int_{\Omega(t)\setminus\Omega_3}d^2z\, G_\Omega(t\,|zz)=\frac{|\Omega(t)\setminus\Omega_3|}{4\pi t}-\frac{2t^{-1/4}+(H_1-\tfrac{3}{2})}{4\sqrt{\pi t}}+O(t^{-1/4}),\quad t\searrow 0. \tag{14}$$

Proof of Eq. (14): Van den Berg[7] showed for regions with R-smooth boundary (for $z_0\in\partial\Omega$ there is a circle K with radius R and $\partial K\cap\Omega=\{z_0\}$): Let $z\in\Omega_1\cup\Omega_2$ and $\delta=\mathrm{dist}(z,\partial\Omega)$, then

$$G_\Omega(t\,|zz)\leqslant\frac{1}{4\pi t}\left\{1-e^{-\delta^2/t}+\frac{4\delta}{R}e^{-\delta^2/t}+4\frac{t}{R^2}\right\}. \tag{15}$$

For $z\in\Omega(t)\setminus(\Omega_1\cup\Omega_2\cup\Omega_3)$ we employ

$$G_\Omega(t\,|zz)\leqslant 1/4\pi t. \tag{16}$$

We choose $R=2$ and integrate:

$$\int_{\Omega(t)\setminus\Omega_3}d^2z\, G_\Omega(t\,|zz)\leqslant\frac{|\Omega(t)\setminus\Omega_3|}{4\pi t}-\frac{1}{4\pi t}\int_{\Omega_1\cup\Omega_2}e^{-\delta^2/t}d^2z+\frac{1}{2\pi t}\int_{\Omega_1\cup\Omega_2}\delta e^{-\delta^2/t}d^2z+\frac{|\Omega(t)|}{4\pi}.$$

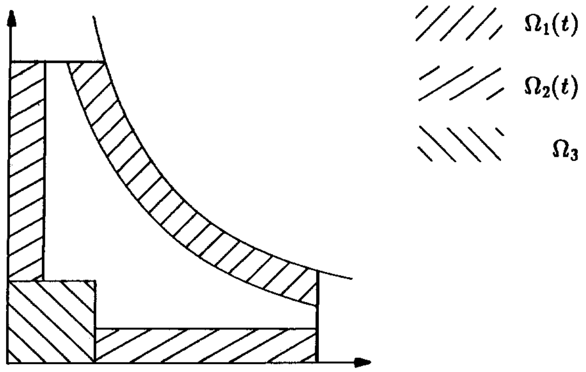


FIG. 1. The sets used in the proof of Theorem 1 are shown.

Let $\delta(z) = \mathrm{dist}(z,\{xy=1\})$ and $s(z)$ be the arc length between $(1,1)$ and the projection of z onto the hyperbola. Since for $C = \{0<\delta<\frac{1}{4}t^{1/4}, s_1<s<s_2\}$:

$$\int_C e^{-\delta^2/t}\,d^2z = \int_{s_1}^{s_2} ds \int_0^{(1/4)t^{1/4}} d\delta(1+\kappa\delta)e^{-\delta^2/t}$$
$$= (s_2-s_1)\frac{\sqrt{\pi t}}{2}\left(1-\mathrm{erfc}\left(\frac{1}{4t^{1/4}}\right)\right) + \frac{t}{2}(1-e^{-1/16\sqrt{t}})\int_{s_1}^{s_2}\kappa\,ds \tag{17}$$

(κ: curvature of the hyperbola) and

$$\int_C \delta e^{-\delta^2/t}\,d^2z = \int_{s_1}^{s_2} ds\int_0^{(1/4)t^{1/4}} d\delta(1+\kappa\delta)\delta e^{-\delta^2/t}$$
$$= \frac{t}{2}(s_2-s_1)(1-e^{-1/16\sqrt{t}}) + t^{3/2}\int_{s_1}^{s_2}\kappa\,ds \times\left\{\frac{\sqrt{\pi}}{4}\left(1-\mathrm{erfc}\left(-\frac{1}{16\sqrt{t}}\right)\right) - \frac{t^{-1/4}}{8}e^{-1/16\sqrt{t}}\right\}, \tag{18}$$

it is easily seen from (18) that $\int_{\Omega_1}\delta e^{-\delta^2/t}\,d^2z = O(t^{3/4})$. Furthermore from (17):

$$\sqrt{\pi t}\left(1-\mathrm{erfc}\left(\frac{1}{4t^{1/4}}\right)\right)\int_1^{t^{1/4}}\sqrt{1+x^{-4}}\,dx \leqslant \int_{\Omega_1(t)} e^{-\delta^2/t}\,d^2z \leqslant \sqrt{\pi t}\int_1^{t^{1/4}+(1/4)t^{1/4}}\sqrt{1+x^{-4}}\,dx + \frac{\pi t}{4},$$

and using $\int_1^x\sqrt{1+u^{-4}}\,du = x-1+H_1+O(1/x^3)$, $x\to\infty$, one finds

$$\int_{\Omega_1(t)} e^{-\delta^2/t}\,d^2z = \sqrt{\pi t}\,(t^{-1/4}-1+H_1) + O(t^{3/4}).$$

Since the integration over Ω_2 is trivial, it is clear that

$$\limsup_{t\searrow 0} t^{1/4}\left(\int_{\Omega(t)\setminus\Omega_3} d^2z\,G_\Omega(t|zz) - \left\{\frac{|\Omega(t)-\Omega_3|}{4\pi t} - \frac{2t^{-1/4}+(H_1-\frac{3}{2})}{4\sqrt{\pi t}}\right\}\right) < +\infty. \tag{19}$$

To obtain a lower bound we integrate the following inequalities:

$z\in\Omega_1(t)\cup\Omega_2(t)$

$$\Rightarrow G_\Omega(t|zz) \geqslant \frac{1-e^{-\delta^2/t}}{4\pi t} - \frac{1}{\pi t}\exp\left(-\frac{1}{32\sqrt{t}}\right), \tag{20}$$

$z\in\Omega(t)\setminus(\Omega_1(t)\cup\Omega_2(t)\cup\Omega_3(t))$

$$\Rightarrow G_\Omega(t|zz) \geqslant \frac{1}{4\pi t} - \frac{1}{\pi t}\exp\left(-\frac{1}{32\sqrt{t}}\right). \tag{21}$$

The first one can be taken from Ref. 15 after a trivial geometric consideration showing that for $z_0\in\partial\Omega(t)\cap\partial\Omega$ there is a square $Q\subset\Omega$, with one side centered at z_0, tangential to $\partial\Omega$ whose sides have length $l=\frac{1}{2}t^{1/4}$. For this bound the convexity of $y=1/x$ is indispensible. If $z\in\Omega(t)\setminus(\Omega_1\cup\Omega_2\cup\Omega_3)$, there is a square $Q'\subset\Omega$ with center z and sidelength $l/\sqrt{2}$ and again[15] establishes the bound.

Integrating we find:

$$\int_{\Omega(t)\setminus\Omega_3} d^2z\,G_\Omega(t|zz) \geqslant \frac{|\Omega(t)\setminus\Omega_3|}{4\pi t} - \frac{1}{4\pi t}\int_{\Omega_1(t)\cup\Omega_2(t)} e^{-\delta^2/t}\,d^2z - \frac{|\Omega(t)|}{4\pi}\exp\left(-\frac{1}{32\sqrt{t}}\right)$$
$$\geqslant \frac{|\Omega(t)\setminus\Omega_3|}{4\pi t} - \frac{1}{4\sqrt{\pi t}}\left(2t^{-1/4}+\left(H_1-\frac{3}{2}\right)\right) - \frac{1}{4\pi} - \frac{|\Omega(t)|}{4\pi}\exp\left(-\frac{1}{32\sqrt{t}}\right),$$

and therefore

$$-\infty < \liminf_{t\searrow 0} t^{1/4}\left(\int_{\Omega(t)-\Omega_3} d^2z\,G_\Omega(t|zz) - \left\{\frac{|\Omega(t)-\Omega_3|}{4\pi t} - \frac{2t^{-1/4}+(H_1-\frac{3}{2})}{4\sqrt{\pi t}}\right\}\right).$$

Together with (19) this completes the proof of (14).

Step 3: integration over Ω_3:

$$0 \leqslant \int_{\Omega_3} d^2z\,G_\Omega(t|zz) - \left(\frac{|\Omega_3|}{4\pi t} - \frac{1}{8\sqrt{\pi t}}\right) \leqslant \frac{1}{\pi}, \tag{22}$$

for $0<t<\frac{1}{10}$.

Proof of Eq. (22): Let $G_Q(t|zz')$ be Green's function for the unit square, and $G_1(t|zz')$ Green's function in the case of Dirichlet boundary conditions on $\{xy=0\}$. Then

$$\int_{\Omega_3} d^2z\,G_Q(t|zz) = \left\{\frac{1}{2}\sum_{n=1}^{\infty} e^{-\pi^2 n^2 t}\right\}^2 \geqslant \frac{1}{4}\left(\frac{1}{\sqrt{4\pi t}} - \frac{1}{2}\right)^2$$

(for $0<t<1/\pi$) by virtue of the Jacobi theta function transformation formula. Furthermore

$$\int_{\Omega_3} d^2z\,G_1(t|zz) = \frac{1}{4\pi t}\left(\int_0^{1/2} dx(1-e^{-x^2/t})\right)^2 \leqslant \frac{|\Omega_3|}{4\pi t} - \frac{1}{8\sqrt{\pi t}} + \frac{1}{\pi}$$

for $0<t<\frac{1}{10}$ (because of $\mathrm{erfc}\,x \leqslant 2/x\sqrt{\pi}$).

The proof of Theorem 1 is obtained simply by adding the results of the three steps, Eqs. (9), (14), and (22). The function H_1 will be determined in the Appendix. □

The proofs of the other theorems are very much like the proof of Theorem 1, therefore they will not be given in great detail.

Proof of Theorem 2: The sets considered for the proof of Theorem 1 have to be replaced by

$\Omega(t) = \{(x,y)\,|\,x<t^{-1/2(1+\mu)}\wedge y<t^{-1/2(1+1/\mu)}\}\cap\Omega,$

$$\begin{aligned}\Omega_1(t) = {} & [\{(x,y)|\mathrm{dist}((x,y);\{x^\mu y=1\}) \\ & <\tfrac14 t^{\mu/2(\mu+1)}, x\geqslant y\}\cap\Omega(t)] \\ & \cup[\{(x,y)|\mathrm{dist}((x,y);\{x^\mu y=1\}) \\ & <\tfrac14 t^{1/2(\mu+1)}, x<y\}\cap\Omega(t)],\end{aligned}$$

$$\Omega_3 = \{(x,y)|x,y<\tfrac12\}\cap\Omega,$$

$$\begin{aligned}\Omega_2(t) = {} & [\{(x,y)|\mathrm{dist}((x,y);\{y=0\}) \\ & <\tfrac14 t^{\mu/2(\mu+1)}, x\geqslant y\}\cap\Omega(t)] \\ & \cup[\{(x,y)|\mathrm{dist}((x,y);\{x=0\}) \\ & <\tfrac14 t^{1/2(\mu+1)}, x<y\}\cap\Omega(t)]\setminus\Omega_3.\end{aligned}$$

Note that the thickness of the strips $\Omega_1(t)$, $\Omega_2(t)$ jumps at $\{x=y\}$ (in contrast to the situation shown in Fig. 1). In a first step the integration over $\Omega\setminus\Omega(t)$ has to be performed using Van den Berg's inequality (10). For $x_0=t^{-1/2(\mu+1)}$ the integral

$$\begin{aligned}&\int_{x_0}^\infty dx \sum_{n=1}^\infty e^{-n^2\pi^2 t x^{2\mu}} \\ &= \frac{1}{t^{1/2\mu}}\cdot\frac{\Gamma(1+1/2\mu)\zeta(1/\mu)}{\pi^{1/\mu}} \\ &\quad -\frac{1/(1-\mu)}{\sqrt{4\pi t}}x_0^{-(\mu-1)}+\frac{x_0}{2}-\frac{(\pi t)^{-1/2\mu}}{2\mu} \\ &\quad\times\int_{x_0^{-2\mu}/\pi t}^\infty du\, u^{-(1/2)(1/\mu+1)}\sum_{n=1}^\infty e^{-n^2\pi u}\end{aligned} \tag{23}$$

is required that is evaluated with the help of (13). Modifying (17) and (18) one finds

$$\int_{\Omega_1\cap\{x>y\}} d^2z\, e^{-\delta^2/t} = \frac{\sqrt{\pi t}}{2}(t^{-1/2(\mu+1)}-1+H_\mu)+O(t^{1-1/2(\mu+1)})$$

$$\int_{\Omega_1\cap\{x>y\}} d^2z\,\delta e^{-\delta^2/t} = O(t^{1-1/2(\mu+1)})$$

[with $H_\mu := \int_1^\infty(\sqrt{1+1/x^{2(\mu+1)}}-1)dx$]. Since $|\Omega(t)\cup\{x\geqslant y\}| = O(t^{(\mu-1)/2(\mu+1)})$ the integration of (15), (16), (20), and (21) gives

$$\begin{aligned}&\int_{\Omega(t)\setminus\Omega_3} d^2z\, G(t|xy,xy) \\ &= \frac{|\Omega(t)\setminus\Omega_3|}{4\pi t}-\left(\frac{2t^{-1/2(\mu+1)}-\frac32+H_\mu}{8\sqrt{\pi t}}\right. \\ &\quad \left.+O(t^{1/2-1/2(\mu+1)})\right)-\left(\mu\leftrightarrow\frac1\mu\right) \\ &\quad +O(t^{-1/2(\mu+1)}+t^{-\mu/2(\mu+1)}),\quad t\searrow 0.\end{aligned} \tag{24}$$

Since

$$\begin{aligned}|\Omega(t)| &= 1+\int_1^{t^{-1/2(\mu+1)}}\frac{dx}{x^\mu}+\left(\mu\leftrightarrow\frac1\mu\right) \\ &= 1+\left(\frac{x_0^{1-\mu}}{1-\mu}-\frac{1}{1-\mu}\right)+\left(\mu\leftrightarrow\frac1\mu\right),\end{aligned}$$

one obtains the proof of Theorem 2 by dividing (23) for μ and $1/\mu$, respectively, by $\sqrt{4\pi t}$ and adding (24). □

Proof of Theorem 3: It suffices to consider $\{x\geqslant 0\}\cap\Omega$. Let $x_0=\frac12(\epsilon-1)\log t$, then with (12):

$$\begin{aligned}&\int_{x_0}^\infty dx\sum_{n=1}^\infty e^{-n^2\pi^2 t e^{2x}} \\ &= \frac{1}{\sqrt{4\pi}}(t^{-\epsilon/2}-\sqrt{\pi})+\frac{\epsilon}{4}\log t \\ &\quad +\frac14(\gamma-2\log 2+2)-\frac{1}{\sqrt\pi}\int_0^{t^{\epsilon/2}}\frac{dy}{y^2}\sum_{n=1}^\infty e^{-n^2/y^2}.\end{aligned} \tag{25}$$

Denoting

$$\Omega(t) = \{(x,y)|t^{(1/2)(1-\epsilon)}<x<x_0\}\cap\Omega,$$

$$\begin{aligned}\Omega_1(t) = {} & \{(x,y)|\mathrm{dist}((x,y),ye^x=1) \\ & <\tfrac14 t^{(1/2)(1-\epsilon)}\}\cap\Omega(t),\end{aligned}$$

$$\Omega_2(t) = \{(x,y)|0<y<\tfrac14 t^{(1/2)(1-\epsilon)}\}\cap\Omega(t),$$

one finds

$$\begin{aligned}\int_{\Omega_1} d^2z\, e^{-\delta^2/t} = {} & \frac{\sqrt{\pi t}}{2}\left(-\frac{\epsilon}{2}\log t+\log 2-1+\sqrt2\right. \\ & \left.-\log(1+\sqrt2)\right)+O(t^{(1-\epsilon)/2}),\end{aligned}$$

$$\int_{\Omega_1} d^2z\,\delta e^{-\delta^2/t} = O(t\log t),$$

and concludes with (15), (16), (20), and (21):

$$\begin{aligned}&\int_{\Omega(t)} d^2z\, G_\Omega(t|zz) \\ &= \frac{|\Omega(t)|}{4\pi t}-\frac{1}{8\sqrt{\pi t}}((\epsilon-1)\log t+\log 2-1+\sqrt2 \\ &\quad -\log(1+\sqrt2))+O(t^{\epsilon/2}),\quad t\searrow 0.\end{aligned} \tag{26}$$

Together with

$$\begin{aligned}&\int_{\Omega\cap\{0<x<t^{(1-\epsilon)/2}\}} d^2z\, G_\Omega(t|zz) \\ &= \frac{|\Omega\cap\{0\leqslant x\leqslant t^{(1-\epsilon)/2}\}|}{4\pi t}+O(t^{-\epsilon}),\quad t\searrow 0,\end{aligned} \tag{27}$$

Eqs. (25) and (26) establish Theorem 3. □

IV. RESULTS FOR REGIONS WITH CUSPS

In Ref. 2, Stewartson and Waechter considered a class of regions Ω with cusps: for $a,k_1,k_2\in\mathbb{R}_+$ let

$$\Omega\cap\{x\geqslant 0\} = \{-k_1(x-a)^\mu\leqslant y\leqslant k_2(x-a)^\mu, a\geqslant x\geqslant 0\}.$$

They gave the contribution resulting from the vicinity of the point $(a,0)$ using exactly Van den Berg's approximation, but without error bounds.

Theorem 4 will give their expansion for the partition function including error bounds, and Theorem 5 will give a similar result for a larger group of systems. Since the proofs of Theorems 4 and 5 follow exactly the lines of Sect. III, we omit them and state merely the results:

Theorem 4: Let $\Omega\subset\mathbf{R}^2$ be bounded, and for some $a,b\in\mathbf{R}_+$ $\partial\Omega\cap\{x<a/2\}$ be smooth, R-smooth and $\Omega\cap\{x\geqslant 0\}=\{(x,y)\,|\,0\leqslant y\leqslant b(a-x)^\mu\wedge 0\leqslant x\leqslant a\}, \mu>1$. Then for $t\searrow 0$:

$$\Theta(t)=\operatorname{Tr}e^{t\Delta}=\frac{|\Omega|}{4\pi t}-\frac{L(\partial\Omega)}{8\sqrt{\pi t}}+\frac{\Gamma(\frac{1}{2}+1/2\mu)\zeta(1+1/\mu)}{4\pi\mu b^{1/\mu}t^{1/2-1/2\mu}}+O(t^{-\epsilon}),\quad \forall\epsilon>0. \tag{28}$$

For the proof of Theorem 4 one requires

$$\int_0^c dx\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2 t}{b^2(c-x)^{2\mu}}\right)=\frac{bc^{\mu+1}/(\mu+1)}{\sqrt{4\pi t}}-\frac{c}{2}+\frac{t^{1/2\mu}b^{-1/\mu}}{2\mu\sqrt{\pi}}\Gamma\left(\frac{1}{2}+\frac{1}{2\mu}\right)\times\zeta\left(1+\frac{1}{\mu}\right)+O(t^r),\quad t\searrow 0, \tag{29}$$

which holds for $\mu>1$ and any $r\in\mathbf{R}$, and can be derived with the help of Eq. (13). Equation (29) is used in connection with (10) for the integration in the vicinity of the point $(a,0)$. Theorem 4 agrees with a result from Ref. 2 with $\mu=2n$, $n\in\mathbf{N}$, apart from the error bounds. These bounds permit the extension of the Theorem to a larger group of systems:

Theorem 5: Let $\Omega\subset\mathbf{R}^2$ be bounded, $\partial\Omega\cap\{x<a/2\}$ be smooth and R-smooth and $\Omega\cap\{x\geqslant 0\}=\{(x,y)\,|\,0\leqslant y\leqslant f(x)\wedge 0\leqslant x\leqslant a\}$, where $f:[0,a]\to\mathbf{R}_+$ is strictly decreasing and convex with $f(x)=b(a-x)^\mu+O((a-x)^{\mu+\alpha})$ and $f'(x)=\mu b(a-x)^{\mu-1}+O((a-x)^{\mu+\alpha-1})$ $(\mu>1, \alpha>0)$, then for $t\searrow 0$,

$$\Theta(t)=\operatorname{Tr}e^{t\Delta}=\frac{|\Omega|}{4\pi t}-\frac{L(\partial\Omega)}{8\sqrt{\pi t}}+\frac{\Gamma(\frac{1}{2}+1/2\mu)\zeta(1+1/\mu)}{4\pi\mu b^{1/\mu}t^{1/2-1/2\mu}}+O(g(t)),\quad t\searrow 0,$$

where $g(t)=t^{-\epsilon}$ $(\epsilon>0)$ if $\frac{1}{2}-(\alpha+1)/2\mu<0$ and $g(t)=t^{-(1/2-(\alpha+1)/2\mu)}$ else.

The proof follows again the pattern of Sec. III, if (29) is replaced by

$$\int_0^a dx\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2 t}{f^2(x)}\right)=\int_0^a dx\left(\frac{f(x)}{\sqrt{4\pi t}}-\frac{1}{2}\right)+\frac{t^{1/2\mu}b^{-1/\mu}}{2\mu\sqrt{\pi}}\Gamma\left(\frac{1}{2}+\frac{1}{2\mu}\right)\times\zeta\left(1+\frac{1}{\mu}\right)+O(t^{(\alpha+1)/2\mu}),\quad t\searrow 0.$$

V. APPLICATION: ASYMPTOTIC EXPANSION FOR $\langle N(\lambda)\rangle$

In general, $N(\lambda)$ does not possess an asymptotic expansion for $\lambda\to\infty$ (see, e.g., Ref. 16), which could be evaluated knowing an expansion of $\Theta(t)$ for $t\searrow 0$. Brownell[17] introduced so-called $\tilde{O}$ bounds that lead to the following result.

Theorem 6 (Brownell): Let $0<\lambda_1\leqslant\lambda_2\leqslant\cdots$ and $N(\lambda)=\Sigma_{\lambda_i<\lambda}1$ satisfy $\int_0^\infty\lambda^{-r_0}|dN(\lambda)|<\infty$ for some $r_0>0$ and let

$$\int_0^\infty dN(\lambda)e^{-\lambda t}=\sum_{i=1}^{k}(t^{-r_i}(c_i+c_i'\log t))+O(t^{-r_{k+1}}),$$

with $r_{k+1}<r_k<\cdots<r_1$. Then

$$N(\lambda)=\sum_{i=1}^{k}\frac{\lambda^{r_i}}{\Gamma(r_i+1)}\times\left[-c_i'\log\lambda+\left(c_i+c_i'\left(\psi(r_i)+\frac{1}{r_i}\right)\right)\right]+\tilde{O}(\lambda^{r_{k+1}}\log\lambda), \tag{30}$$

where ψ denotes the digamma function and $\tilde{O}$ refers to "log Gaussian error estimates": Let F be of bounded variation over every finite interval, where it is continuous and $\int_0^\infty y^{-r_0}|dF(y)|<+\infty$ for some $r_0\geqslant 0$; let either $f_r(y)=y^r\log y$ or $f_r(y)=y^r, r>0$, then

$$F(y)=\tilde{O}(f_r(y))\Leftrightarrow$$

$$\forall\rho>0\,\exists M_\rho:\left|\int_b^\infty\exp\left(-\frac{1}{2}\rho^2\left(\log\frac{v}{y}\right)^2\right)dF(y)\right|\leqslant M_\rho f_r(v).$$

(We have added the case $c_i'\neq 0$, which can easily be handled mimicking Brownell's original proof.)

If an averaging procedure $\langle\cdots\rangle$ defined by

$$\langle F(v)\rangle:=\left|\int_0^\infty\exp\left(-\frac{1}{2}\rho^2\left(\log\frac{v}{y}\right)^2\right)dF(y)\right|, \tag{31}$$

is applied to (30) the $\tilde{O}$ estimates can be replaced by ordinary error estimates:

$$\langle N(\lambda)\rangle=\sum_{i=1}^{k}\frac{\lambda^{r_i}}{\Gamma(r_i+1)}\times\left[-c_i'\log\lambda+\left(c_i+c_i'\left(\psi(r_i)+\frac{1}{r_i}\right)\right)\right]+O(\lambda^{r_{k+1}}\log\lambda)\},\quad\lambda\to\infty. \tag{32}$$

For the significance of the "smoothed" level density $d\langle N(\lambda)\rangle/d\lambda$ see Ref. 1, Chap. VII. For its application in the field of chaotic systems see Refs. 18 and 19. Applying (31) to the system in Theorem 1, one finds

$$\langle N(\lambda)\rangle\sim\frac{\lambda\log\lambda}{4\pi}+\frac{a}{4\pi}\lambda-\frac{b}{4\pi}\sqrt{\lambda}+O(\lambda^{1/4}\log\lambda), \tag{33}$$

for $\lambda\to\infty$, where

$$a=2(\gamma-\log(2\pi))=-2.5213\cdots,$$

$$b=-4(\pi^{3/2}/\Gamma^2(\tfrac{1}{4}))=-1.6944\cdots.$$

The staircase function for the quantum billiard in the region

$\{0 \leqslant xy \leqslant k\} \cap \mathbb{R}^2_+$ can be found by replacing λ by $k\lambda$ in the rhs of Eq. (33). For example, for $k = 2$,

$$\langle N_{k=2}(\lambda) \rangle = \frac{\lambda \log \lambda}{2\pi} + \frac{(a + \log 2)}{2\pi} \lambda - \frac{b\sqrt{2}}{4\pi} \sqrt{\lambda} + O(\lambda^{1/4} \log \lambda), \quad \lambda \to \infty. \tag{34}$$

Berry[8] pointed out that the leading term in the last expression coincides with the first term of the asymptotic expansion for the number of the nontrivial zeroes of the Riemann zeta function with imaginary part less than λ:

$$\langle N_{\text{Riemann}}(\lambda) \rangle = \frac{\lambda \log \lambda}{2\pi} - \frac{(1 + \log(2\pi))}{2\pi} \lambda + O(\lambda^0), \quad \lambda \to \infty. \tag{35}$$

The lack of time-reversal invariance, which a hypothetical system whose eigenvalues are given by the imaginary parts of the Riemann zeroes is believed to be subject to (see Ref. 8), could be enforced upon a billiard system by introducing a magnetic field, thus establishing an Aharonov–Bohm billiard (c.f. Ref. 20). Note, however, that the terms we determined are not affected by a magnetic field of the Aharonov–Bohm type, i.e., the above formulas also hold for Aharonov–Bohm quantum billiards.

ACKNOWLEDGMENT

This work was supported by Deutsche Forschungsgemeinschaft under Contract No. DFG-Ste 241/4-2.

APPENDIX

In Theorems 1 and 2 the integral

$$H_\mu = \int_1^\infty dx \left(\sqrt{1 + x^{-2(\mu+1)}} - 1\right)$$

is required. It can be evaluated in terms of a hypergeometric function:

$$\begin{aligned} H_\mu &= \int_1^\infty dx \left(\sqrt{1 + x^{-2(\mu+1)}} - 1\right) \\ &= 1 - \sqrt{2} + (\mu + 1) \int_1^\infty \frac{dx}{x^{\mu+1}\sqrt{x^{2(\mu+1)} + 1}} \\ &= 1 - \sqrt{2} + \frac{1}{2} \int_1^\infty \frac{dz}{\sqrt{1+z}} z^{-\mu/[2(\mu+1)] - 1} \\ &= 1 - \sqrt{2} + \frac{1}{1 + 2\beta} \\ &\quad \times {}_2F_1\left(\frac{1}{2}, \frac{1}{2} + \beta; \frac{3}{2} + \beta; -1,\right), \end{aligned}$$

with $\beta = \mu/2(\mu + 1)$ (formula 3.194.1 in Ref. 21). If $\mu = 1$, i.e., $\beta = \frac{1}{4}$:

$$\begin{aligned} {}_2F_1\left(\frac{1}{2}, \frac{3}{4}; \frac{7}{4}; -1\right) &= -4\left(\frac{\pi}{2}\right)^{1/2} \Gamma\left(\frac{7}{4}\right) \left(\frac{1}{\Gamma(\frac{1}{4})} - \frac{1}{\Gamma(\frac{3}{4})\Gamma(\frac{1}{2})}\right) \\ &= -3 \frac{\pi^{3/2}}{\Gamma^2(\frac{1}{4})} + \frac{3}{\sqrt{2}} \end{aligned}$$

(cf. Ref. 14, p. 40) and therefore

$$H_1 = 1 - 2(\pi^{3/2}/\Gamma^2(\tfrac{1}{4})).$$

If $\mu \neq 1$:

$$\begin{aligned} {}_2F_1\left(\frac{1}{2}, \frac{1}{2} + \beta; \frac{3}{2} + \beta; -1\right) &= 2^{1/2+\beta} \Gamma\left(\frac{3}{2} + \beta\right) \mathscr{P}^{-1/2-\beta}_{-1/2-\beta}(\sqrt{2}) \end{aligned}$$

(Ref. 14, p. 52) and thus

$$H_\mu = 1 - \sqrt{2} + 2^{\beta - 1/2} \Gamma(\beta + \tfrac{1}{2}) \mathscr{P}^{-1/2-\beta}_{-1/2-\beta}(\sqrt{2}).$$

[1] H. P. Baltes and E. R. Hilf, *Spectra of Finite Systems* (Bibliographisches Institut, Zürich, 1976).

[2] K. Stewartson and R. T. Waechter, Proc. Cambridge Philos. Soc. **69**, 353 (1971).

[3] H. P. McKean, and I. M. Singer, J. Diff. Geom. **1**, 43 (1967).

[4] M. Kac, Am. Math. Mon. **73**, 1 (1966).

[5] E. B. Davies, Math. Z. **188**, 245 (1985).

[6] F. Rellich, in, *Studies and Essays* (Interscience, New York, 1948).

[7] M. Van den Berg, J. Funct. Anal. **71**, 279 (1987).

[8] M. V. Berry, in, *Quantum Chaos and Statistical Nuclear Physics*, Proceedings Cuernavaca, Lecture Notes in Physics, edited by T. H. Seligman and H. Nishioka (Springer, New York, 1986), Vol. 263, pp. 1–17.

[9] H. Tamura, Nagoya Math. J. **60**, 7 (1976).

[10] M. Van den Berg, J. Funct. Anal. **58**, 150 (1984).

[11] B. Simon, Adv. Math. **30**, 268 (1978).

[12] B. Simon, Ann. Phys. **146**, 209 (1983).

[13] B. Simon, J. Funct. Anal. **53**, 84 (1983).

[14] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), 3rd ed.

[15] M. Van den Berg, J. Math. Phys. **22**, 2452 (1981), Commun. Math. Phys. **92**, 525 (1984).

[16] B. Simon, *Functional Integration and Quantum Physics* (Pure and applied mathematics, a series of monographs and textbooks) (Academic, New York, 1979).

[17] F. H. Brownell, J. Math. Mech. **6**, 119 (1957).

[18] R. Aurich, and F. Steiner, Physica D **32**, 451 (1988); R. Aurich, M. Sieber, and F. Steiner, Phys. Rev. Lett. **61**, 483 (1988); R. Aurich, and F. Steiner, Physica D **39**, 169 (1989).

[19] M. Sieber, and F. Steiner, DESY-preprint 89-093; Physica D (in press).

[20] M. V. Berry, J. Phys. A **19**, 2281 (1986); F. Steiner, Fortschr. Phys. **35**, 87 (1987).

[21] I. S. Gradshteyn, and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).

Exact solution of the n-dimensional Dirac–Coulomb equation

M. K. F. Wong
Fairfield University, Fairfield, Connecticut 06430

(Received 15 August 1989; accepted for publication 21 February 1990)

An exact solution of the n-dimensional Dirac–Coulomb equation is obtained with the radial wave function containing only one term of a confluent hypergeometric function. It is of the same form as the solutions to the Schrödinger and Klein–Gordon equations with a Coulomb potential in n dimensions.

I. INTRODUCTION

It is clear that the radial solutions of the three-dimensional Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential are of the same form, and can be treated in a uniform way.[1] The purpose of this paper is to show that the n-dimensional Coulomb problem for the three equations retains the same property. The solutions to the n-dimensional Schrödinger and Klein–Gordon equations with a Coulomb potential have been obtained by Nieto.[2] We shall show in this paper that the n-dimensional Dirac–Coulomb equation has an exact solution, which is, moreover, in the same form as the Schrödinger and Klein–Gordon equations for the radial wave function.

Aside from the theoretical significance of the result, there is at least one application for our result. This is the so-called large N method. In the realistic three-dimensional case, if the potential is other than Coulomb, but (is) say, spherically symmetric, an exact solution may not exist in analytic form. However, it might still be possible to calculate the eigenvalue of E by the large N method.[3]

To extend the Dirac–Coulomb equation from three-dimensional to n-dimensional space, we follow mainly the structure set up by Joseph[4] and Coulson and Joseph.[5] They obtained the necessary quantum numbers to label the n-dimensional Dirac–Coulomb equation, but did not obtain the wave function solution explicitly. We complete their work by using the transformation S, which in the three-dimensional case was obtained by Biedenharn,[6] Wong and Yeh,[7] and Su[8] to solve the Dirac–Coulomb equation.

II. n-DIMENSIONAL DIRAC-COULOMB EQUATION AND ITS SOLUTION

The n-dimensional Dirac–Coulomb equation we wish to solve is

$$H\psi=\left(c\rho_1\sum_{i=1}^{n}p_i\sigma_{i,n+1}+\rho_3 mc^2-\frac{Ze^2}{r}\right)\psi, \tag{2.1}$$

where we have put $\hbar=1$. In what follows, we use mainly the notation of Joseph[4] and Coulson and Joseph.[5] Let

$$\sigma_{ij}=-(i/2)[\beta_i,\beta_j], \tag{2.2}$$

with

$$\beta_i\beta_j+\beta_j\beta_i=2\delta_{ij}. \tag{2.3}$$

The n-dimensional space is separated into "spherical" coordinates with r the radial distance and $(n-1)$ angular variables $\alpha_1,\alpha_2,\ldots,\alpha_{n-1}$. We have

$$\begin{aligned}
x_1&=r\sin\alpha_{n-1}\cdots\sin\alpha_2\cos\alpha_1,\\
x_2&=r\sin\alpha_{n-1}\cdots\sin\alpha_1,\\
x_3&=r\sin\alpha_{n-1}\cdots\cos\alpha_2,\\
&\vdots\\
x_n&=r\cos\alpha_{n-1}.
\end{aligned} \tag{2.4}$$

From (2.4) it is easily seen that

$$x_1^2+x_2^2+\cdots+x_n^2=r^2. \tag{2.5}$$

As usual, we define the orbital angular momentum $\mathscr{L}_{ij}$ to be

$$\mathscr{L}_{ij}=x_ip_j-x_jp_i,\quad p_i=-i\frac{\partial}{\partial x_i}. \tag{2.6}$$

They form the generators of SO(n). We select the representation with one number only in each row of the Gel'fand pattern, labeled by ℓ_n for SO(n), etc. Then each state in the irreducible representation of SO(n) is completely specified by $(\ell_n,\ell_{n-1},\ldots,\ell_3,\ell_2)$, with ℓ_2 = integers and ℓ_m, $m\geqslant 3$ nonnegative integers, $\ell_m\geqslant\ell_{m-1}\geqslant|\ell_2|$. The Casimir invariant for SO(n): $\Sigma_{i<j}^n\mathscr{L}_{ij}^2$ has the value $\ell_n(\ell_n+n-2)$. The eigenfunctions of these irreducible representations are the generalized spherical harmonics, which can be found explicitly in Alcaras and Ferreira.[9]

The next step is to choose the correct invariants in the n-dimensional Dirac–Coulomb equation so that the labels are easily identified corresponding to nonnegative integers, $\ell_n,\ell_{n-1},\ldots$, etc. (except ℓ_2). This is done by working with the invariant K_n:

$$K_n=\rho_3(\mathrm{L}_n+(n-1)/2), \tag{2.7}$$

where

$$\mathrm{L}_n=\sum_{m=2}^{n}L_m, \tag{2.8}$$

$$L_m=\sum_{i=1}^{m-1}\sigma_{im}\mathscr{L}_{im}. \tag{2.9}$$

It has been shown by Joseph[4] that the Casimir invariant

$$\mathscr{L}_n^2=\sum_{i<j}^{n}\mathscr{L}_{ij}^2$$

is equal to

$$\mathscr{L}_n^2=\mathrm{L}_n(\mathrm{L}_n+n-2). \tag{2.10}$$

However,

$$\mathscr{L}_n^2=\ell_n(\ell_n+n-2), \tag{2.11}$$

therefore,

 0022-2488/90/071677-04$03.00

$$\mathrm{L}_n = \ell_n, \tag{2.12}$$

or

$$-(\ell_n + n - 2). \tag{2.13}$$

The eigenfunctions for the angular variables in the Dirac–Coulomb equation corresponding to (2.12) and (2.13) are denoted by $\phi^{(a)}$ and $\phi^{(b)}$, respectively. We shall show that these are the wave functions used by Joseph.[4] But first we shall add another interpretation for these wave functions as follows. From (2.7) we have, using k_n as the eigenvalue of K_n,

$$k_n^2 = (L_n + (n-1)/2)^2. \tag{2.14}$$

If we use (2.12) *we have*

$$k_n^2 = (\ell_n + (n-1)/2)^2. \tag{2.15}$$

If we use (2.13), we have

$$\begin{aligned} k_n^2 &= [-\ell_n - n + 2 + (n-1)/2]^2 \\ &= [(\ell_n - 1) + (n-1)/2]^2. \end{aligned} \tag{2.16}$$

When $n = 3$, we have $\ell = j + \widetilde{\omega}/2$ and

$$k_3 = \widetilde{\omega}(j + \tfrac{1}{2}) = \widetilde{\omega}(\ell - \widetilde{\omega}/2 + \tfrac{1}{2}). \tag{2.17}$$

So (2.15) corresponds to $\widetilde{\omega} = -1$, i.e.,

$$k_3 = -(\ell + 1), \tag{2.18}$$

and (2.16) corresponds to $\widetilde{\omega} = +1$, i.e.,

$$k_3 = (\ell - 1) + 1 = \ell. \tag{2.19}$$

In (2.15) and (2.16), we can take ℓ_n to be a nonnegative integer, thus obtaining, from (2.16),

$$k_n = (\ell_n - 1) + (n-1)/2, \tag{2.20}$$

and from (2.15),

$$k_n = -[\ell_n + (n-1)/2]. \tag{2.21}$$

This will agree with the conventional labeling for the $n = 3$ case, where $\widetilde{\omega} = \pm 1$, with $\widetilde{\omega} = +1$ for (2.20) and $\widetilde{\omega} = -1$ for (2.21). In the n-dimensional case, according to Joseph, we use $\phi^{(a)}$ and $\phi^{(b)}$ with $\phi^{(a)}$ having "$\ell_n - 1$" and $\phi^{(b)}$ having "ℓ_n" as labels. So the (a) and (b) states differ from each other by one unit of angular momentum, with $\phi^{(b)}$ having a value one unit larger than $\phi^{(a)}$. This classification is in agreement with the one given by Joseph.[4] What we have added is that instead of (a) and (b), we could use $\widetilde{\omega} = \pm 1$ with (a) corresponding to $\widetilde{\omega} = +1$ and (b) corresponding to $\widetilde{\omega} = -1$.

The corresponding coupling through the generalized "Clebsch–Gordan" coefficients is given by Joseph as follows:

$$\begin{aligned} &\phi^{(a)}_{\ell_n}(\alpha_1,\dots,\alpha_{n-1}) \\ &\quad = (2\ell_n + n - 2)^{-1/2}\Big\{(\ell_n + \ell_{n-1} + n - 2)^{1/2} \\ &\qquad \times \psi_{\ell_n,\ell_{n-1}}(\alpha_{n-1})\phi^{(a)}_{\ell_{n-1}}(\alpha_1,\dots,\alpha_{n-2}) \\ &\qquad + (\ell_n - \ell_{n-1})^{1/2}\psi_{\ell_n,\ell_{n-1}+1}(\alpha_{n-1}) \\ &\qquad \times \phi^{(b)}_{\ell_{n-1}}(\alpha_1,\dots,\alpha_{n-2})\Big\}, \end{aligned} \tag{2.22}$$

$$\begin{aligned} &\phi^{(b)}_{\ell_n}(\alpha_1,\dots,\alpha_{n-1}) \\ &\quad = (2\ell_n + n - 2)^{-1/2}\Big\{(\ell_n - \ell_{n-1})^{1/2}\psi_{\ell_n,\ell_{n-1}}(\alpha_{n-1}) \\ &\qquad \times \phi^{(a)}_{\ell_{n-1}}(\alpha_1,\dots,\alpha_{n-2}) - (\ell_n + \ell_{n-1} + n - 2)^{1/2} \\ &\qquad \times \psi_{\ell_n,\ell_{n-1}+1}(\alpha_{n-1})\phi^{(b)}_{\ell_{n-1}+1}(\alpha_1,\dots,\alpha_{n-2})\Big\}, \end{aligned} \tag{2.23}$$

where $\psi_{\ell_n,\ell_{n-1}}(\alpha_{n-1})$ is part of the spherical harmonics obtained, for example, by Alcaras and Ferreira[9] and denoted by them as G.

Incidentally, in the n-dimensional case one can also define an equivalent total angular momentum operator

$$J_{ij} = \mathscr{L}_{ij} + \tfrac{1}{2}\sigma_{ij}. \tag{2.24}$$

Then,

$$J_n^2 = \sum_{i<j}^{n} J_{ij}^2 \tag{2.25}$$

is a constant of the motion. However, its eigenvalue is complicated [see (2.26) below] and thus we shall not use it. For the state (b), we get

$$J_n^2 = \ell_n^2 + \ell_n(n-1) + n(n-1)/8. \tag{2.26}$$

In what follows, we shall write the eigenvalue of K_n as

$$\begin{aligned} k_n &= k_n(\widetilde{\omega},\ell_n) \\ &= \begin{cases} -[\ell_n + (n-1)/2], & (2.27) \\ +[(\ell_n - 1) + (n-1)/2], & (2.28) \end{cases} \end{aligned}$$

with $\widetilde{\omega} = +1$ corresponding to the state (a) in (2.28) and $\widetilde{\omega} = -1$ corresponding to the state (b) in (2.27). If $\ell_n = 0$, then only the (b) state exists; the (a) state does not exist. An irreducible representation is then completely specified by the two quantum numbers ℓ_n and $\widetilde{\omega}$, where ℓ_n is a nonnegative integer and $\widetilde{\omega}$ is either $+1$ or -1. Then the solution of the Dirac–Coulomb equation (after the transformation S) can be written as

$$\psi' = \begin{pmatrix} iR(r)\phi_{\ell_n\widetilde{\omega}}(\alpha_1,\dots,\alpha_{n-1}) \\ Q(r)\widehat{A}_{n+1}\phi_{\ell_n\widetilde{\omega}}(\alpha_1,\dots,\alpha_{n-1}) \end{pmatrix}, \tag{2.29}$$

where $\phi_{\ell_n\widetilde{\omega}}$ is either $\phi^{(a)}_{\ell_n}$ for $\widetilde{\omega} = +1$ of (2.22) or $\phi^{(b)}_{\ell_n}$ for $\widetilde{\omega} = -1$ of (2.23). Let

$$\widehat{A}_{n+1} = \frac{A_{n+1}}{\|A_{n+1}\|}, \tag{2.30}$$

$$A_{n+1} = \sum_{i=1}^{n} x_i\sigma_{i,n+1}. \tag{2.31}$$

The Dirac–Coulomb equation in spherical coordinates is

$$\begin{aligned} H\psi = \Big[&-c\rho_1 i\widehat{A}_{n+1}\Big(\frac{\partial}{\partial r} - \frac{\rho_3 K_n - (n-1)/2}{r}\Big) \\ &+ \rho_3 mc^2 - \frac{Ze^2}{r}\Big]\psi. \end{aligned} \tag{2.32}$$

The transformed wave function is

$$\psi' = S\psi, \tag{2.33}$$

where

$$S=\cosh\theta/2-\rho_2(\sinh\theta/2)\hat{A}_{n+1}, \tag{2.34}$$

with

$$\tanh\theta=-Ze^2/cK_n. \tag{2.35}$$

It is then easy to calculate

$$E\psi'=\mathrm{SHS}^{-1}\psi', \tag{2.36}$$

obtaining two coupled equations for $R(r)$ and $Q(r)$;

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}\end{bmatrix}\begin{pmatrix} R(r) \\ Q(r)\end{pmatrix}=E\begin{pmatrix} R(r) \\ Q(r)\end{pmatrix}, \tag{2.37}$$

$$M_{11}=mc^2\cosh\theta+c\left[\sinh\theta\left(\frac{d}{dr}+\frac{(n-1)}{2r}\right)\right]-\frac{Ze^2}{r}, \tag{2.38}$$

$$M_{12}=-\left\{mc^2\sinh\theta+c\left[\cosh\theta\left(\frac{d}{dr}+\frac{(n-1)}{2r}\right)-\frac{K_n}{r}\right]\right\}, \tag{2.39}$$

$$M_{21}=mc^2\sinh\theta+c\left[\cosh\theta\left(\frac{d}{dr}+\frac{(n-1)}{2r}\right)+\frac{K_n}{r}\right], \tag{2.40}$$

$$M_{22}=-\left\{mc^2\cosh\theta+c\left[\sinh\theta\left(\frac{d}{dr}+\frac{(n-1)}{2r}\right)+\frac{Ze^2}{r}\right]\right\}. \tag{2.41}$$

Defining

$$\gamma_n=(K_n^2-Z^2e^4/c^2)^{1/2}, \tag{2.42}$$

we obtain

$$c\left[\frac{d}{dr}+\frac{(n-1)}{2r}\right]R+\left(\frac{c\tilde{\omega}\gamma_n}{r}\right)R-\left(\frac{E\tilde{\omega}Ze^2}{c\gamma_n}\right)R=\left(mc^2+\frac{E\,|k_n|}{\gamma_n}\right)Q, \tag{2.43}$$

$$c\left[\frac{d}{dr}+\frac{(n-1)}{2r}\right]Q-\left(\frac{c\tilde{\omega}\gamma_n}{r}\right)Q+\left(\frac{E\tilde{\omega}Ze^2}{c\gamma_n}\right)Q=\left(mc^2-\frac{E\,|k_n|}{\gamma_n}\right)R. \tag{2.44}$$

From (2.43) and (2.44) we obtain the second-order equations for $R(r)$ and $Q(r)$,

$$\left[\frac{d^2}{dr^2}+\frac{(n-1)}{r}\frac{d}{dr}+\frac{E^2-m^2c^4}{c^2}+\frac{2EZe^2}{c^2r}-\frac{(\gamma_n^2+\tilde{\omega}\gamma_n-n^2/4+n-\frac{3}{4})}{r^2}\right]R(r)=0, \tag{2.45}$$

$$\left[\frac{d^2}{dr^2}+\frac{(n-1)}{r}\frac{d}{dr}+\frac{E^2-m^2c^4}{c^2}+\frac{2EZe^2}{c^2r}-\frac{(\gamma_n^2-\tilde{\omega}\gamma_n-n^2/4+n-\frac{3}{4})}{r^2}\right]Q(r)=0. \tag{2.46}$$

To solve for R, let us change the variable from r to ρ, with

$$\rho=2r(m^2c^4-E^2)^{1/2}/c=2\mu r. \tag{2.47}$$

We then have

$$\left[\frac{d^2}{d\rho^2}+\frac{(n-1)}{\rho}\frac{d}{d\rho}-\frac{1}{4}+\frac{EZe^2}{\rho c(m^2c^4-E^2)^{1/2}}-\frac{(\gamma_n^2+\tilde{\omega}\gamma_n-n^2/4+n-\frac{3}{4})}{\rho^2}\right]R(\rho)=0. \tag{2.48}$$

Equation (2.48) is entirely similar in form with the radial equation of the Klein–Gordon equation with a Coulomb potential in n dimensions. According to Nieto,[2] Eq. (5.4), this equation is

$$\left(\frac{d^2}{d\rho^2}+\frac{(n-1)}{\rho}\frac{d}{d\rho}-\frac{1}{4}+\frac{\lambda}{\rho}-\frac{[\ell(\ell+n-2)-\gamma^2]}{\rho^2}\right)R(\rho)=0. \tag{2.49}$$

Nieto has obtained the solution for (2.49) in terms of generalized Laguerre polynomials, which are also equivalent to confluent hypergeometric functions. Thus in order to obtain the solution of (2.48), we only have to translate Nieto's notation into our notation. This is done as follows:

Nieto's notation		our notation
ℓ	$\to$	γ_n,
λ	$\to$	$EZe^2/c(m^2c^4-E^2)^{1/2}$,
γ^2	$\to$	$n\gamma_n-2\gamma_n-\tilde{\omega}\gamma_n+n^2/4-n+\frac{3}{4}$, (always $\geqslant 0$ for $n\geqslant 3$).

(2.50)

Thus Nieto's s, which is defined as

$$s=-(n-2)/2+\{[\ell+(n-2)/2]^2-\gamma^2\}^{1/2}, \tag{2.51}$$

is defined in the same way by us, with the translation given by (2.50).

Thus the final expression for $R(\rho)$ is

$$R(\rho)=N\exp(-\rho/2)\rho^sL_{n'}^{(2s+n-2)}(\rho), \tag{2.52}$$

where

$$n'=\text{non-negative integer}=\lambda-(n-1)/2-s, \tag{2.53}$$

and the generalized Laguerre polynomial is connected with the confluent hypergeometric function as follows:

$$L_n^{(\alpha)}(t)=\sum_{j=0}^{n}\binom{n+\alpha}{n-j}\frac{(-t)^j}{j!}=\binom{n+\alpha}{n}{}_1F_1(-n,\alpha+1,t). \tag{2.54}$$

The expression for $Q(\rho)$ is obtained from $R(\rho)$ by changing $\tilde{\omega}$ to $-\tilde{\omega}$.

From (2.53) and (2.51), one obtains the energy spectrum for the bound state of the n-dimensional Dirac–Coulomb equation:

$$E=\pm mc^2\left[1+\frac{Z^2e^4}{c^2(n'+\frac{1}{2}+\tilde{\omega}/2+\gamma_n)^2}\right]^{-1/2}, \tag{2.55}$$

which agrees with Coulson and Joseph using the second-order Dirac–Coulomb equation in n dimensions.

The normalization constants are obtained as follows. First we write

$$R(r) = C(S_\pm) e^{-\rho/2} \rho^{S_\pm} \times {}_1F_1(S_\pm - \lambda + (n-1)/2, 2S_\pm + n - 1, \rho), \tag{2.56}$$

$$Q(r) = C(S_\mp) e^{-\rho/2} \rho^{S_\mp} \times {}_1F_1(S_\mp - \lambda + (n-1)/2, 2S_\mp + n - 1, \rho), \tag{2.57}$$

where

$$S_- = \gamma_n - (n-3)/2, \tag{2.58}$$

$$S_+ = \gamma_n - (n-1)/2. \tag{2.59}$$

The upper sign is for $\tilde{\omega} = -1$, the lower sign for $\tilde{\omega} = +1$.

By putting $r = 0$, we obtain from Eqs. (2.43) and (2.44) the following relation:

$$C(S_-) = C(S_+) \frac{(mc^2 + \tilde{\omega} E |k_n|/\gamma_n)}{2\mu c (2\gamma_n + 1)}. \tag{2.60}$$

The angular parts of the wave functions are already normalized. The radial parts are normalized according to the following condition:

$$\int_0^\infty (R^2(r) + Q^2(r)) r^{n-1}\, dr = 1. \tag{2.61}$$

From (2.61) we obtain

$$C(S_+) = \frac{-\tilde{\omega}}{(2\gamma_n - 1)!} \left[\frac{(2\mu)^n (\lambda + \gamma_n - 1)! \gamma_n |mc^2 - \tilde{\omega} E |k_n|/\gamma_n|}{4\lambda E |k_n| (\lambda - \gamma_n)!} \right]^{1/2}. \tag{2.62}$$

In deriving (2.62), use has been made of the following identity:

$$(\mu^2 c^2/\gamma_n^2)(\gamma_n^2 - \lambda^2) = m^2 c^4 - E^2 k_n^2/\gamma_n^2. \tag{2.63}$$

[1] O. L. de Lange, J. Math. Phys. **30**, 858 (1989).
[2] M. M. Nieto, Am. J. Phys. **47**, 1067 (1979).
[3] S. Atag, J. Math. Phys. **30**, 696 (1989), and references therein.
[4] A. Joseph, Rev. Mod. Phys. **39**, 829 (1967).
[5] C. A. Coulson and A. Joseph, Rev. Mod. Phys. **39**, 838 (1967).
[6] L. C. Biedenharn, Phys. Rev. **126**, 845 (1962).
[7] M. K. F. Wong and H. Y. Yeh, Phys. Rev. D **25**, 3396 (1982).
[8] J. Y. Su, Phys. Rev. A **32**, 3251 (1985).
[9] J. A. Alcaras and P. L. Ferreira, J. Math. Phys. **6**, 578 (1965).

Extended Hilbert space approach to few-body problems

Yu. A. Kuperin, K. A. Makarov, S. P. Merkuriev, A. K. Motovilov,[a] and B. S. Pavlov
Department of Theoretical Physics, Leningrad University, Leningrad 199034, USSR

(Received 30 August 1989; accepted for publication 3 January 1990)

A general formulation of the quantum scattering theory for a system of few particles, which have an internal structure, is given. Due to freezing out the internal degrees of freedom in the external channels, a certain class of energy-dependent potentials is generated. By means of potential theory, a modified Faddeev equation is derived both in external and internal channels. The Fredholmity of these equations is proven and this is what provides a sound basis for solving the addressed scattering problem.

I. INTRODUCTION

This paper is concerned with the treatment of low-energy quantum scattering for few particles with internal structure. Problems of this kind arise in describing hadron–hadron as well as nucleus–nucleus scattering[1–5] and similarly in constructing NN potentials in the baglike approaches.[6] There exist, however, no appropriate mathematical methods which may be applied for a rigorous study of the wavefunction properties. Already in the two-body problem there is no self-adjoint (s.a.) Hamiltonian, which could be generated by a time-dependent unitary group of operators responsible for the evolution of the system. In the three-body case, the original Faddeev equations are also not directly applicable due to the lack of an underlying s.a. Hamiltonian.

In the present paper, we overcome these difficulties by means of the extension theory using an auxiliary Hilbert space corresponding to the internal degrees of freedom.[7–9] In the special extended Hilbert space we construct the total s.a. Hamiltonian. After eliminating the internal channels, we propose modified Faddeev equations for the components of external-channel Green functions. Using well-known methods,[10,11] we prove that these equations represented in configuration space are of Fredholm type. Due to this property the equations provide the justification of treating the three-body scattering problem for particles interacting via energy-dependent potentials. Our modified Faddeev equations in differential form may also be used in an efficient manner for numerical calculations.

II. TWO-BODY PROBLEM

We will consider here the following special case of the general situation.[7,9,12] Let us assume that the dynamics of the external degrees of freedom are given by the s.a. Hamiltonian $h^{\rm ex}$, which is defined by

$$h^{\rm ex}u=(-\Delta+v(x))u \tag{1}$$

in the Hilbert space $\mathscr{H}^{\rm ex}=L^2(\mathbb{R}^3)$. The potential $v(x)$ represents a so-called peripheral interaction (e.g., a meson-exchange potential) of strongly interacting particles and it will be assumed to decrease rapidly and be sufficiently smooth.

We shall also separate the two-body configuration space $\mathbb{R}^3$ into the two domains $\Omega^\pm$ such that $\mathbb{R}^3=\Omega^-\cup\Omega^+$. Let Ω^- be the part of the space $\mathbb{R}^3$ where the coordinate x is bounded. Physically, the compact domain Ω^- may be interpreted as the region of reaction (or where clusters overlap) and the domain $\Omega^+=\mathbb{R}^3\setminus\Omega^-$ as the region where the particles move "asymptotically free." The common boundary γ of the domains $\Omega^\pm$ will in this situation be a surface, where the phase transition between internal and external channels takes place.

In our model we shall restrict the s.a. Hamiltonian $h^{\rm ex}$ to the symmetric operator h_0 with the domain $\mathscr{D}(h_0)=C^\infty_{0,\gamma}$, where $C^\infty_{0,\gamma}$ is the class of infinitely differentiable functions, which vanish together with all derivative in the neighborhood of the surface γ. Then the Hermitian conjugate operator h_0^* has a nontrivial boundary form $J^{\rm ex}$, namely,

$$\begin{aligned}J^{\rm ex}(u,w)&=\langle h_0^*u,w\rangle-\langle u,h_0^*w\rangle\\&=\lim_{\delta\to0}\left[\int_{\gamma_\delta^+}dS(\partial_n u\overline{w}-u\partial_n\overline{w})\right.\\&\qquad\left.-\int_{\gamma_\delta^-}dS(\partial_n u\overline{w}-\partial_n\overline{w}u)\right],\end{aligned} \tag{2}$$

where ∂_n is the normal derivative on the surfaces

$$\gamma_\delta^\pm=\{x\in\Omega^\pm:\ \ \mathrm{dist}(x,\gamma^\pm)=\delta\}.$$

Now we assume that the dynamics of the internal degrees of freedom without connection to the external channel $\mathscr{H}^{\rm ex}$ is given by an arbitrary s.a. operator A acting in some Hilbert space $\mathscr{H}^{\rm in}$. In order to "switch on" the interaction between channels $\mathscr{H}^{\rm ex}$ and $\mathscr{H}^{\rm in}$, one must restrict the operator A to some symmetric operator A_0 and construct all s.a. extensions of the operator $h_0\oplus A_0$ in the direct sum $\mathscr{H}^{\rm ex}\oplus\mathscr{H}^{\rm in}$. The important question of the model is the following: How to construct the boundary form $J^{\rm in}$ for an arbitrary s.a. operator A? The general answer was obtained in Ref. 7. Namely, the symmetric restriction of the Hamiltonian should be made in terms of its Cayley transform $U=(A-i)/(A+i)$. For this purpose let us consider the special isometric restriction $U_0=U\restriction\mathscr{H}^{\rm in}\ominus U^*\theta$, where θ is a generative element[3] of the operator A. The symmetric restriction A_0 can be obtained as the inverse Cayley transformation of the isometry U_0. Hence, the operator A_0 has deficiency indices (1,1) and the domain $\mathscr{D}(A_0^*)$ of its adjoint can be described in terms of von Neumann's theory[3]: $\mathscr{D}(A_0^*)=\mathscr{D}(\overline{A}_0)+\mathscr{L}(\theta,U^*\theta)$; here $\overline{A}_0$ is the closure of

[a] Permanent address: Department of Mathematics, Arkhangelsk Institute for Lumber Technique, Arkhangelsk, USSR.

A_0 and $\mathscr{L}(\theta, U^*\theta)$ is the span of deficiency elements θ and $U^*\theta$. It is convenient to introduce some new basis in $\mathscr{L}$: $w^+ = \frac{1}{2}(U^*\theta + \theta), w^- = (1/2i)(U^*\theta - \theta)$. In accordance with von Neumann's representation, an arbitrary vector u can be decomposed as

$$u = \tilde{u} + \xi^+ w^+ + \xi^- w^-, \quad u \in \mathscr{D}(A_0^*), \quad \tilde{u} \in \mathscr{D}(\overline{A}_0), \tag{3}$$

where $\xi^{\pm}(u)$ are the so-called boundary values of element u (Ref. 7). In terms of $\xi^{\pm}$ the boundary form of the operator A_0^* may be written as

$$\begin{aligned} J^{\text{in}}(u,f) &= \langle A_0^* u, f\rangle - \langle u, A_0^* f\rangle \\ &= \xi^-(u)\,\overline{\xi^+(f)} - \xi^+(u)\,\overline{\xi^-(f)}. \end{aligned} \tag{4}$$

It should be noted that (4) is an abstract variant of (2).

After the preparation of the boundary forms J^{ex} and J^{in} of the operators h_0^* and A_0^* the next step is to construct an s.a. extension h of the operator $h_0 \oplus A_0$, acting in the direct sum $\mathscr{H}^{\text{ex}} \oplus \mathscr{H}^{\text{in}}$. In accordance with our general method one should impose on γ such boundary conditions that make the sum of the boundary forms vanish, i.e., $J^{\text{ex}} + J^{\text{in}} = 0$.

It can be shown that all such nulllifying conditions may only be of two types, namely:

$$\begin{pmatrix} u^+ \\ \partial_n u^- \\ \\ \xi^+ \end{pmatrix} = \begin{pmatrix} & \mathfrak{S} & & \varphi^+ \\ & & & \varphi^- \\ & & & \\ \langle\cdot,\varphi^+\rangle & \langle\cdot,\varphi^-\rangle & & \alpha_B \end{pmatrix} \begin{pmatrix} \partial_n u^+ \\ u^- \\ \\ \xi^- \end{pmatrix} \tag{5}$$

and

$$\begin{pmatrix} \partial_n u^+ \\ u^- \\ \\ \xi^+ \end{pmatrix} = \begin{pmatrix} & \mathfrak{S} & & \varphi^+ \\ & & & \varphi^- \\ & & & \\ -\langle\cdot,\varphi^+\rangle & -\langle\cdot,\varphi^-\rangle & & \alpha_B \end{pmatrix} \begin{pmatrix} u^+ \\ \partial_n u^- \\ \\ \xi^- \end{pmatrix}. \tag{6}$$

Here, $u^{\pm}$ and $\partial_n u^{\pm}$ are boundary values of u and $\partial_n u$ on the bilateral surface $\gamma^{\pm}$ and the functions $\varphi^{\pm} \in L^2(\gamma^{\pm})$ are parameters of the model; they generate functionals

$$\langle u, \varphi^{\pm}\rangle = \int_{\gamma^{\pm}} dS\, u\, \overline{\varphi^{\pm}}, \quad u \in L^2(\gamma^{\pm}). \tag{7}$$

Finally, $\mathfrak{S}$ is a 2×2 arbitrary s.a. matrix given on the surface γ and α_B is an arbitrary real number.

Let us denote by u_α, $\alpha = 0,1$ the external ($\alpha = 0$) and internal ($\alpha = 1$) channel wavefunctions. Then we can study their properties on the basis of the two-channel Schrödinger equation:

$$\begin{aligned} &(h - z)\,\mathscr{U} = 0, \quad x \in \mathbb{R}^3 \setminus \gamma, \\ &\mathscr{U} = (u_0, u_1), \end{aligned} \tag{8}$$

with the appropriate boundary conditions (5) or (6).

Let us suppose now for simplicity that the external channel wavefunctions u_0 are smooth on the surface γ, i.e., $u_0^+ = u_0^- = u_0$ and that the matrix $\mathfrak{S}$ has the special structure $\mathfrak{S} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$; let us also put $\alpha_B = 0$. In this special case the boundary conditions, which one should add to (8), are of the form

$$[\partial_n u_0]_\gamma = -\varphi \xi^-(u_1), \tag{9}$$

$$\xi^+(u_1) = \langle u_0, \varphi\rangle, \tag{10}$$

where $[\partial_n u_0]_\gamma = \partial_n u_0^- - \partial_n u_0^+$ and $\varphi = \varphi^- = \varphi^+$.

We want to emphasize that the operator h, which is the total Hamiltonian in the two-body system with internal structure, is an s.a. operator and hence the boundary value problem (8),(9),(10) is mathematically correctly defined. It should also be noted that in our model we are able to simulate an arbitrary complicated internal structure of particles due to the general nature of the internal s.a. operator A.

On the other hand, one can now operate in the external channel $\mathscr{H}^{\text{ex}}$ only. For this purpose one must solve the boundary conditions (9),(10) by excluding the internal ingredients $\xi^{\pm}(u_1)$. This procedure is based on the following linear relations[7]:

$$\xi^- = \Delta(z)\xi^+, \tag{11}$$

where $\Delta(z)$ is the Schwartz integral of the spectral measure of the s.a. operator A,

$$\begin{aligned} \Delta(z) &\equiv \langle (I + zA)(A - z)^{-1}\theta, \theta\rangle \\ &= \int_{\mathbb{R}} \frac{1 + \lambda z}{\lambda - z}\, d\langle E_\lambda^A \theta, \theta\rangle. \end{aligned} \tag{12}$$

Taking into account (11) we obtain from (9) and (11) the following energy-dependent boundary conditions in the external space $\mathscr{H}^{\text{ex}}$:

$$[\partial_n u_0]_\gamma = -\Delta(z)\varphi\langle u_0, \varphi\rangle. \tag{13}$$

In accordance with (8) the external component u_0 obeys the equation

$$(h_0^* - z)u_0 = 0, \quad x \in \mathbb{R}^3 \setminus \gamma. \tag{14}$$

In order to obtain the differential equations in the whole configuration space $\mathbb{R}^3$ it is convenient to use the quasipotential approach (see, e.g., Ref. 11). Let us consider the quasipotential $w(z)$ acting on the function u in accordance with the rule

$$w(z)u = -\delta_\gamma \Delta(z)\varphi\langle u, \varphi\rangle. \tag{15}$$

Here, $\delta_\gamma \mu$ is the distribution, usually called the simple layer,[14] that acts on the set of sufficiently smooth functions f in the following way:

$$\langle \delta_\gamma \mu, f\rangle = \int_\gamma dS\, \mu \bar{f}. \tag{16}$$

In terms of the quasipotential $w(z)$ the boundary-value problem (13) and (14) may be written as

$$(h^{\text{ex}} + w(z) - z)u_0 = 0, \tag{17}$$

where the variable x now runs over the whole configuration space $\mathbb{R}^3$.

One can show that (17) is equivalent to the boundary-value problem (13),(14).

At the end of this section we add the following remarks.

(1) As it was stated above, a mathematically correct formulation of the two-body problem with energy-dependent interactions can be achieved only in terms of an s.a. operator h acting in the sum $\mathscr{H}^{\text{ex}} \oplus \mathscr{H}^{\text{in}}$ of internal and external channels. On the contrary, there exists no s.a. Hamiltonian corresponding to (17) or, what is the same, to the boundary-value problem (13),(14). This means that in

terms of the external space $\mathscr{H}^{ex}$ only the well-defined total S-matrix cannot be constructed.

(2) As it follows from (12) and (15), the energy dependence of the potentials cannot be arbitrary. It is given by the Schwartz integral $\Delta(z)$, which is real on the real axis and is an analytical function on the half-plane $\operatorname{Im} z > 0$ with the positive imaginary part $\operatorname{Im} \Delta(z) > 0$. It can be shown that such kind of interactions ensure the analyticity and unitarity of the appropriate total scattering matrix.

(3) In our model the quasipotentials $w(z)$ are separable of rank 1. The generalization to any arbitrary rank of $w(z)$ is trivial. For this purpose one should increase the dimension of the deficiency subspaces $\mathfrak{N} = \{\theta\}$, $\mathfrak{N}^* = \{U^*\theta\}$ and change in a self-consistent way the functionals φ, $\langle \cdot,\varphi \rangle$ by arbitrary bounded operators B,B^*.

III. THREE-BODY HAMILTONIAN

We consider in this section a system of three particles having a nontrivial internal structure. To describe this system, we use in the external configuration space $\mathbb{R}^6$ usual relative coordinates x_α, y_α, $\alpha = 1,2,3$, which we combine into the six-vector $X = x_\alpha \oplus y_\alpha$ (Ref. 15). Every pair x_α, y_α fixes an orthogonal coordinate system in $\mathbb{R}^6$.

Let $\Gamma_\alpha = \gamma_\alpha \times \mathbb{R}^3_{y_\alpha}$ be the cylinders in $\mathbb{R}^6$ and $\Gamma \equiv \underset{\alpha}{U}\, \Gamma_\alpha$. An example of an external configuration space (for the one-dimensional case) is represented in Fig. 1.

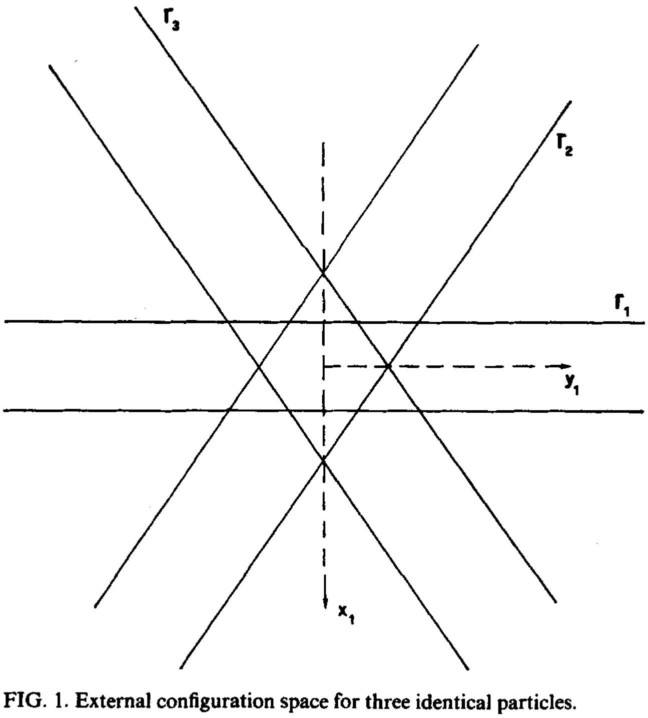


FIG. 1. External configuration space for three identical particles.

A total s.a. Hamiltonian H governing the dynamics of internal and external degrees of freedom will be an important object in the three-body analysis.

We start by considering the two-particle Hamiltonian H_α in the six-dimensional external configuration space:

$$H_\alpha = h_\alpha \otimes I_{y_\alpha} + I_\alpha \otimes (-\Delta_{y_\alpha}). \tag{18}$$

Here, h_α is the s.a. two-body Hamiltonian defined in Sec. II, I_{y_α} and I_α are the unit operators, in the spaces $L^2(\mathbb{R}^3_{y_\alpha})$ and $\mathscr{H}_\alpha = \mathscr{H}^{ex}_\alpha \oplus \mathscr{H}^{in}_\alpha$, respectively, and $-\Delta_{y_\alpha}$ is the Laplacian defined on its natural domain $W^2_2(\mathbb{R}^3_{y_\alpha})$.

The operator H_α is essentially s.a. on the domain

$$\mathscr{D}(H_\alpha) = \mathscr{D}(h_\alpha) \otimes W^2_2(\mathbb{R}^3_{y_\alpha}). \tag{19}$$

The closure $\overline{H}_\alpha$ of this operator is the s.a. operator, which will be denoted by the same symbol H_α.

The domain $\mathscr{D}(H_\alpha)$ also may be described in terms of boundary conditions. Namely, let $\mathscr{U} = \{u_0, u_\alpha\} \in \mathscr{D}(H_\alpha)$. Then the external component u_0 is a W^2_2 smooth function outside the neighborhood of Γ_α continued on Γ_α. The internal component $u_\alpha \in \mathfrak{H}^{in}_\alpha = \mathscr{H}^{in}_\alpha \otimes L^2(\mathbb{R}^3_{y_\alpha})$ can be decomposed into the sum

$$u_\alpha = \tilde{u}_\alpha + \xi^+_\alpha(y_\alpha) w^+_\alpha + \xi^-_\alpha(y_\alpha) w^-_\alpha, \quad \tilde{u}_\alpha \in \mathscr{D}(H^{in}_{\alpha_0}), \tag{20}$$

where $w^\pm_\alpha$ are the deficiency elements[13] of the symmetric operator A_{α_0} which is the restriction of the s.a. operator A_α and

$$H^{in}_{\alpha_0} = A_{\alpha_0} \otimes I_{y_\alpha} + I_\alpha \otimes (-\Delta_{y_\alpha}). \tag{21}$$

The functions $\mathscr{U} \in \mathscr{D}(H_\alpha)$ satisfy the boundary conditions

$$[\partial_n u_0]_{\Gamma_\alpha} = -\varphi_\alpha(x_\alpha) \xi^-_\alpha(y_\alpha), \tag{22}$$

$$\xi^+_\alpha(y_\alpha) = \langle u_0, \varphi_\alpha \rangle (y_\alpha), \tag{23}$$

where

$$\langle u_0, \varphi_\alpha \rangle (y_\alpha) = \int_{\gamma_\alpha} dx_\alpha\, u_0(X)\, \overline{\varphi_\alpha(x_\alpha)}. \tag{24}$$

It should be noted that the boundary conditions (22),(23) are essentially of two-body character [see (9) and (10)]. The only difference is that the $\xi^\pm_\alpha(y_\alpha)$ are now functions of the variable $y_\alpha \in \mathbb{R}^3_{y_\alpha}$.

We are now ready to construct the total three-body Hamiltonian H. Let us consider in the space

$$\mathfrak{H} = L^2(\mathbb{R}^6) \oplus \sum_{\alpha=1}^{3} \oplus \mathfrak{H}^{in}_\alpha$$

symmetric operator H_0,

$$H_0 \mathscr{U} = \begin{cases} (-\Delta_X + \sum_\alpha v_\alpha(x_\alpha)) u_0, \\ H^{in}_{\alpha_0} u_\alpha, \quad \alpha = 1,2,3, \end{cases} \tag{25}$$

on the domain

$$\mathscr{D}(H_0) = C^\infty_0(\mathbb{R}^6 \setminus \Gamma) \oplus \sum_\alpha \oplus \mathscr{D}(H^{in}_{\alpha_0}).$$

Any s.a. extension H of the operator H_0 is a total three-body Hamiltonian describing the whole dynamics in both external and internal channels. In accordance with the von Neumann theory[13] all such extensions can be obtained by the extension of the operator H_0 on its deficiency subspaces. So we shall extend the domain $\mathscr{D}(H_0)$ to the linear set $\mathscr{D}(\tilde{H}_0)$ in the following manner:

$$\mathscr{D}(\tilde{H}_0) = \begin{cases} u_0 = \tilde{u}_0 + \sum_\alpha R_0(-1)\rho_\alpha, \\ \qquad \tilde{u}_0 \in C_0^\infty(\mathbb{R}^6 \setminus \Gamma), & (26) \\ u_\alpha = \tilde{u}_\alpha + \xi_\alpha^+ w_\alpha^+ + \xi_\alpha^- w_\alpha^-, \\ \qquad \xi_\alpha^\pm \in W_2^2(\mathbb{R}^3_{y_\alpha}). & (27) \end{cases}$$

Here, $R_0(z) = (H^{\text{ex}} - z)^{-1}$ is the resolvent of the s.a. operator $H^{\text{ex}} = -\Delta_X + \Sigma_\alpha v_\alpha$ and ρ_α are the densities of the simple layer potentials given on cylinders $\Gamma_\alpha, \alpha = 1,2,3$. The operator $\tilde{H}_0$ on the domain $\mathscr{D}(\tilde{H}_0)$ obtained in this way is the nonsymmetric restriction of the adjoint operator H_0^*. Now one must restrict the operator $\tilde{H}_0$ to some symmetric one. For this purpose it is sufficient to impose the boundary conditions (22),(23). In terms of the densities ρ_α these conditions may be written as

$$\rho_\alpha(X) = -\varphi_\alpha(x_\alpha)\xi^-(y_\alpha), \tag{28}$$

$$\xi_\alpha^+(y_\alpha) = \left\langle \sum_{\beta=1}^3 R_0(-1)\rho_\beta, \varphi_\alpha \right\rangle (y_\alpha). \tag{29}$$

As a result we restrict the domain $\mathscr{D}(H_0)$ to the linear set $\mathscr{D}(\tilde{H})$ by means of conditions (28),(29). The symmetric operator that is such a restriction of $\tilde{H}_0$ on the domain $\mathscr{D}(\tilde{H})$ will be called $\tilde{H}$.

Let us now collect the important facts about the operator $\tilde{H}$.

Theorem 1: The operator $\tilde{H}$ on the domain

$$\mathscr{D}(\tilde{H}), \overline{\mathscr{D}(\tilde{H})} = \mathfrak{H} = L^2(\mathbb{R}^6) \oplus \sum_\alpha \mathfrak{H}_\alpha^{\text{in}}$$

is symmetric and bounded from below.

The proof of this statement is given in Appendix A.

The last step is now an extension of the symmetric operator $\tilde{H}$ to the s.a. operator H obeying the following conditions.

(1) On the domain $\mathscr{D}(H)$ the translation-invariant boundary conditions (22) and (23) must be kept.

(2) The Hamiltonian H must be bounded from below.

For this purpose we shall choose the Friedrichs extension H of the operator $\tilde{H}$ (Ref. 13). On the domain $\mathscr{D}(H)$, which can be described as usual,[13] the action of H is given by

$$H\mathscr{U} = \begin{cases} H^{\text{ex}}u_0, \\ -\Delta_{y_\alpha}u_\alpha + A_\alpha\tilde{u}_\alpha - \xi_\alpha^+ w^- + \xi_\alpha^- w_\alpha^+, \\ \mathscr{U} = (u_0, u_\alpha), \quad \alpha = 1,2,3, \end{cases} \tag{30}$$

with the boundary conditions (22),(23).

IV. RESOLVENT EQUATIONS

This section deals with the Fredholm-type equations for the resolvent $R(z)$ of the s.a. Hamiltonian H. As in the case of energy-independent interactions,[16] these equations provide the basis for the three-body scattering problem.

All the results of this section can be extended to the N-body case for arbitrary N.

First we shall derive differential equations for the resolvent components $R_{ab}(z)$ corresponding to the decomposition of $\mathfrak{H}$ into the sum (24),

$$R(z) = \{R_{ab}(z)\}, \quad a,b = 0,1,2,3. \tag{31}$$

Here the indices a,b stand for the external $(a,b=0)$ and internal $(a,b = 1,2,3)$ subspaces $\mathfrak{H}^{\text{ex}} = L^2(\mathbb{R}^6)$ and $\mathfrak{H}_\alpha^{\text{in}}$, $\alpha = 1,2,3$.

Because $R(z)$ is the resolvent of the s.a. operator H it satisfies the usual relation

$$R_{ab}^*(z) = R_{ba}(\bar{z}). \tag{32}$$

We shall introduce the following notations. Let F be an arbitrary element of $\mathfrak{H}$ and $\mathscr{U} = R(z)F$, i.e., $F = \{f_a\}$, $a = 0,1,2,3$,

$$u_a = \sum_{b=0}^3 R_{ab}(z) f_b. \tag{33}$$

Then due to (27) and (33) one gets

$$\xi_\alpha^\pm = \sum_{b=0}^3 \mathscr{C}_{ab}^\pm(z) f_b, \quad \alpha = 1,2,3, \tag{34}$$

where $\mathscr{C}_{ab}^\pm$ are the operators that act from $\mathfrak{H}^{\text{ex}}$ in $L^2(\mathbb{R}^3_{y_\alpha})$ at $b=0$ and from $\mathfrak{H}_\alpha^{\text{in}}$ in $L^2(\mathbb{R}^3_{y_\alpha})$ at $b \neq 0$. The relation (34) can be considered as the definition of these operators.

Let $\tilde{R}_{ab}(z)$ denote the operators

$$\tilde{R}_{ab}f_b = (R_{ab} - w_\alpha^+ \mathscr{C}_{ab}^+ - w_\alpha^- \mathscr{C}_{ab}^-) f_b,$$
$$\alpha = 1,2,3, \quad b = 0,1,2,3. \tag{35}$$

Then using the identity

$$(H - z)R(z)F = F, \tag{36}$$

one can obtain the set of equations for the kernels of the operators $R_{ab}(z)$ and $\mathscr{C}_{ab}^\pm(z)$:

$$(H^{\text{ex}} - z)R_{0b}(z) = \delta_{0b}I_0, \tag{37}$$

$$A_\alpha\tilde{R}_{ab} - w^-\mathscr{C}_{ab}^+ + w_\alpha^+\mathscr{C}_{ab}^- - (\Delta_{y_\alpha} + z)R_{ab}(z) = \delta_{ab}I_\alpha, \tag{38}$$

with the following boundary conditions:

$$[\partial_n R_{0b}]_{\Gamma_\alpha} = -\varphi_\alpha \mathscr{C}_{ab}^-, \tag{39}$$

$$\mathscr{C}_{ab}^+ = \langle R_{0b}, \varphi_\alpha \rangle|_{\Gamma_\alpha}. \tag{40}$$

Equations (37) and (38) representing the set of differential equations for the external R_{0b} and internal R_{ab} components of the resolvent $R(z)$ serve as background for the construction of the Faddeev equations.

We shall rewrite the conditions (39) and (40) in terms of the internal Hamiltonians A_α. For this purpose we use the relation:

$$\mathscr{C}_{ab}^- \cdot = Q_\alpha(z)\mathscr{C}_{ab}^+ \cdot + \delta_{ab}\langle (A_\alpha - i)(H_\alpha^{\text{in}} - z)^{-1} \cdot, \theta_\alpha \rangle, \tag{41}$$

which can be obtained by arguments analogous to the two-body case[7] [see (11)]. Here,

$$H_\alpha^{\text{in}} = A_\alpha \otimes I_{y_\alpha} + I_\alpha \otimes (-\Delta_{y_\alpha}) \tag{42}$$

and $Q_\alpha(z)$ is the generalization of the Schwartz integral in the three-body configuration space:

$$Q_\alpha(z) = \langle (I + (\Delta_{y_\alpha} + z)A_\alpha)(H_\alpha^{\text{in}} - z)^{-1}\theta_\alpha, \theta_\alpha \rangle. \tag{43}$$

This operator, in accordance with (18), can be realized as the integral operator having the kernel

$$Q_\alpha(y_\alpha - y'_\alpha,z) = \frac{1}{2\pi i}\oint_{\mathscr{L}_\alpha} d\lambda\, \Delta_\alpha(\lambda)\, r_0^{(\alpha)}(y_\alpha - y'_\alpha, z-\lambda). \tag{44}$$

Here, $r_0^{(\alpha)}(z) = (h_\alpha^{\rm ex} - z)^{-1}$ is the resolvent of the two-body s.a. Hamiltonian $h_\alpha^{\rm ex}$; $\Delta_\alpha(\lambda)$ is the two-body Schwartz integral, and the counter $\mathscr{L}_\alpha$ encircles the spectrum of A_α.

The operators $\mathscr{C}_{ab}^{\pm}$ can now be excluded from (39) and (40) by virtue of relation (41):

$$[\partial_n R_{0b}]_{\Gamma_\alpha} = -\varphi_\alpha\{Q_\alpha(z)\langle R_{0b}\cdot,\varphi_\alpha\rangle + \delta_{ab}\langle (A_\alpha - i)(H_\alpha^{\rm in} - z)\cdot,\theta_\alpha\rangle\}. \tag{45}$$

If the internal channel Hamiltonians A_α have the point spectra $\sigma_p(A_\alpha) = \{z_s^\alpha\}$ only, then the kernels Q_α should be written in the form

$$Q_\alpha(y_\alpha - y'_\alpha,z) = \sum_s (1+(z_s^\alpha)^2)\langle \mathscr{P}_s^\alpha\theta_\alpha,\theta_\alpha\rangle \times r_0^{(\alpha)}(y_\alpha - y'_\alpha, z - z_s^\alpha), \tag{46}$$

where $\mathscr{P}_s^\alpha$ are spectral projectors of the operators A_α.

Notice that such kind of internal Hamiltonians are used for describing internal channels, e.g., with quark confinement.[5,6]

V. FADDEEV EQUATIONS

The study of the total resolvent $R(z)$ can be reduced to considering the external-channel component $R_{00}(z)$ only. In order to see this, Eqs. (37)–(41) should be used. Namely, let the component $R_{00}(z)$ be known. From (40) we can get $\mathscr{C}_{\alpha_0}^{+}(z)$ for substitution into (41) to yield $\mathscr{C}_{\alpha_0}^{-}(z)$. Then from (38) one can obtain $\widetilde{R}_{\alpha_0}(z)$. It gives the components $R_{\alpha_0}^{(z)}(z)$, $\alpha = 1,2,3$. Then, in accordance with (32) the components $R_{\alpha_0}(z)$, $\alpha = 1,2,3$ will also be known. The diagram in Fig. 2 illustrates this procedure. Thus we shall now deal with $R_{00}(z)$ which, for simplicity, is denoted by $G(z)$. It should be noticed that $G(z)$ is the so-called Krein's quasiresolvent[17] and it has corresponding properties.

By Eq. (45) the kernel $G(X,X',z)$ of the quasiresolvent $G(z)$ obeys the boundary conditions

$$[\partial_n G(X,X',z)]_{\Gamma_\alpha} = -\varphi_\alpha(x_\alpha)Q_\alpha(z)\langle G(z)\cdot,\varphi_\alpha\rangle. \tag{47}$$

As in the two-body case, these conditions can be written in terms of quasipotentials

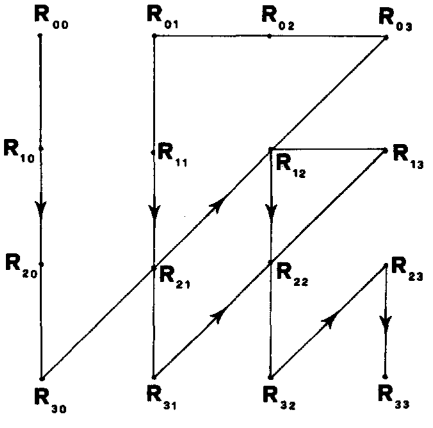


FIG. 2. Diagram for the reconstruction of the resolvent components R_{ab}, $a,b = 0,1,2,3$.

$$W_\alpha(z)\mathscr{U} = \delta_{\Gamma_\alpha} V_\alpha(z)\mathscr{U}, \tag{48}$$

where $V_\alpha(z)$ is the integral operator in $L^2(\Gamma_\alpha)$ with the kernel

$$V_\alpha(X,X',z) = -\varphi_\alpha(x_\alpha)Q_\alpha(y_\alpha - y'_\alpha,z)\,\overline{\varphi_\alpha(x'_\alpha)}. \tag{49}$$

In accordance with (37), (47), and (48) we obtain the following equation:

$$\left(H^{\rm ex} + \sum_{\alpha=1}^{3} W_\alpha(z) - z\right)G(X,X',z) = \delta(X-X'). \tag{50}$$

To derive an integral equation for Krein's quasiresolvent one can use the usual procedure. Namely, applying the operator $R_0(z)$ to (50) we obtain the resolvent identity for $G(z)$:

$$G(z) = R_0(z) - R_0(z)\sum_{\alpha=1}^{3} W_\alpha(z)G(z). \tag{51}$$

In accordance with this equation of Lippman–Schwinger type, the operator $G(z)$ may be reproduced explicitly in terms of generalized operators

$$M_\alpha(z) = W_\alpha(z)G(z) \tag{52}$$

by the relation

$$G(z) = R_0(z) - R_0(z)\sum_\alpha M_\alpha(z). \tag{53}$$

Thus we reduce the problem of investigating the quasiresolvent $G(z)$ to the study of the operators $M_\alpha(z)$.

The next problem is to derive the Faddeev equations from Eq. (53). Applying the operators $W_\alpha(z)$ to (53) one can write this equation in the form

$$(I + W_\alpha R_0)M_\alpha = W_\alpha R_0 - W_\alpha R_0 \sum_{\beta\neq\alpha} M_\beta. \tag{54}$$

Following Faddeev's method we have to invert the operator $I + W_\alpha R_0$. This inversion may be done explicitly in terms of the two-body operator $G_\alpha = (H_\alpha - z)^{-1}$, which is the resolvent of the s.a. operator H_α. The following formula can be verified:

$$(I + W_\alpha R_0)W_\alpha G_\alpha = W_\alpha R_0. \tag{55}$$

This relation yields in a straightforward way the equations

$$M_\alpha(z) = W_\alpha G_\alpha(z) - W_\alpha G_\alpha(z)\sum_{\beta\neq\alpha} M_\beta(z), \tag{56}$$

which have the structure of Faddeev equations.

Nevertheless, to be convinced that these equations are the Faddeev ones, one must prove the following statement.

Theorem 2: Let μ_α be the densities of the simple layer potentials $M_\alpha(z) = \delta_{\Gamma_\alpha}\mu_\alpha$ and $\mu = (\mu_1,\mu_2,\mu_3)$. Then the following can be proven.

(1) Equations (56) rewritten in terms of densities μ_α:

$$\mu(z) = \mu_b(z) + B(z)\mu(z) \tag{57}$$

are of Fredholm type and B^n, $n > N_{\max}$ with sufficiently large $N_{\max}$ is a compact operator in an appropriate Banach space.

(2) Equations (56) or (57) are spectral equivalent to

the original Schrödinger equation with the s.a. Hamiltonian H.

The proof of the first statement proceeds in a standard way.[10,11] Nevertheless, for the reader's convenience we sketch it in Appendix B.

The second statement of the theorem is much more delicate in contrast to the case of energy-independent interactions. Namely, we must show that the homogeneous equations (57) have a nontrivial solution, if and only if $z \in \sigma_p(H)$, where $\sigma_p(H)$ is the point spectrum of the s.a. operator H.

Let μ be the solution of the homogeneous equations

$$\mu_\alpha(z) = -V_\alpha(z) G_\alpha(z) \sum_{\beta \neq \alpha} \mu_\beta(z). \tag{58}$$

Consider the function

$$u_0 = R_0(z) \sum_\beta \mu_\beta,$$

which is evidently the simple layer potential given on the hypersurface $\Gamma = U_\alpha \Gamma_\alpha$ and hence it satisfied the equation

$$(H^{\text{ex}} - z) u_0(X) = 0, \quad X \notin \Gamma. \tag{59}$$

To find the appropriate boundary conditions one must apply the operator $I + V_\alpha R_0$ to Eq. (58). Taking into account (55) and the properties[18] of simple layer potentials $[\partial_n u_0]_{\Gamma_\alpha} = -\mu_\alpha$, we find the boundary conditions

$$[\partial_n u_0]_{\Gamma_\alpha} = V_\alpha(z) u_0. \tag{60}$$

By means of iterations of (58) one can achieve that $\mu_0 \in W_2^2(\mathbb{R}^6 \setminus \Gamma)$ both at $\operatorname{Im} z \neq 0$ and at $z = E \pm i0$, $E \in \mathbb{R}$ and furthermore that $\langle u_0, \varphi_2 \rangle \in W_2^2(\mathbb{R}^3_{y_\alpha})$. Now we shall express the internal functions μ_α in terms of the external component u_0. To this end one must take into account the representation u_α in the form (28) and relations (22) as well as (21), which state the connection between $\xi_\alpha^\pm$ and u_0:

$$\xi_\alpha^+ = \langle u_0, \varphi_\alpha \rangle, \tag{61}$$

$$\xi_\alpha^- = Q_\alpha(z) \xi_\alpha^+. \tag{62}$$

The functions $\tilde{u}_\alpha$ may then be found as the solution of the equation

$$\begin{aligned}(-\Delta_{y_\alpha} + A_\alpha - z)\tilde{u}_\alpha &= \xi_\alpha^+(y_\alpha) w_\alpha^- - \xi_\alpha^-(y_\alpha) w_\alpha^+ + (\Delta_{y_\alpha} \\ &\quad + z)(\xi_\alpha^-(y_\alpha) w_\alpha^- + \xi_\alpha^+(y_\alpha) w_\alpha^+).\end{aligned} \tag{63}$$

By virtue of Eq. (61) the functions $\xi_\alpha^+ \in W_2^2(\mathbb{R}^3_{y_\alpha})$ and hence $\xi_\alpha^- \in W_2^2(\mathbb{R}^3_{y_\alpha})$. This means that $\mathscr{U} = \{u_0, u_\alpha\} \in \mathscr{D}(H)$ and, in accordance with (59)–(63), $\mathscr{U}$ is an eigenvector of the s.a. operator H:

$$(H - z)\mathscr{U} = 0. \tag{64}$$

This equation implies that $\mathscr{U} = 0$ if $z \notin \sigma_p(H)$ and hence $u_0 = 0$. In other words we have proven that Eq. (57) has a unique solution, if $z \notin \sigma_p(H)$.

On the contrary let $\mathscr{U}$ be an eigenvector of the Hamiltonian H. Then one must repeat the derivation of (56) for densities $\mu_\alpha = -V_\alpha u_0$ which obey Eqs. (58).

Hence we have proven that the Faddeev equations (58) are "spectral equivalent" to the original Schrödinger equation (64).

Consequently, the Fredholm alternative may be applied to (57) and the properties of densities μ_α may be investigated and its completeness in the whole space $\mathfrak{H}$ established using the methods of Refs. 19 and 20.

VI. INTERNAL-CHANNEL FADDEEV EQUATIONS

In this section, we shall derive Faddeev equations for the boundary values $\mathscr{C}_{ab}^\pm(z)$ of the resolvent components $R_{ab}(z)$.

First, one observes that due to the separability of the quasipotentials $V_\alpha(z)$, Eq. (56) can be written in terms of kernels of the operators $\mathscr{C}_{\alpha_0}^-(z)$,

$$\mathscr{C}_{\alpha_0}^-(y_\alpha, X', z) = Q_\alpha(y_\alpha - y'_\alpha, z) \langle G(z) \cdot, \varphi_\alpha \rangle, \tag{65}$$

which by (40), (41), and (42) are the boundary values of $R_{ab}(z)$. In order to obtain equations for the operators $\mathscr{C}_{\alpha_0}^-$ it is advantageous to pick out coefficients at the φ_α in (56) using the relations (48) and (49). Due to this procedure one gets

$$\mathscr{C}_{\alpha_0}^-(z) = D_\alpha^-(z) - \sum_{\beta \neq \alpha} D_{\alpha\beta}^-(z) \mathscr{C}_{\beta_0}^-(z). \tag{66}$$

Here, the operators $D_\alpha^-(z)$ are average value of the resolvent

$$D_\alpha^-(z) = \langle G_\alpha(z), \varphi_\alpha \rangle \tag{67}$$

and

$$D_{\alpha\beta}^-(z) = \langle D_\alpha^-(z), \varphi_\beta \rangle. \tag{68}$$

Equations (66) represent a set of integral equations in internal subspaces for Faddeev-type operators $\mathscr{C}_{\alpha_0}^-(z)$.

Let us note that there exists an additional effect, which can be taken into account. Namely, owing to the separability of the potentials $V_\alpha(z)$ Eqs. (66) are three dimensional in contrast to the five-dimensional equations (56). This property can be important for numerical calculations.

VII. DIFFERENTIAL FADDEEV EQUATIONS FOR COMPONENTS

The Faddeev differential equations are known to be useful for numerical calculations in nuclear physics.[15,20] Let us discuss similar equations for three particles interacting via energy-dependent potentials. For their derivation we define the Faddeev components $G^\alpha(z)$ of the external-channel quasiresolvent $G(z)$ in the usual form:

$$G^\alpha(z) = \delta_{\alpha_1} R_0(z) - R_0(z) W_\alpha(z) G(z). \tag{69}$$

By this definition the quasiresolvent $G(z)$ is the sum of its components

$$G(z) = \sum_{\alpha=1}^{3} G^\alpha(z). \tag{70}$$

Applying the "operator" $H^{\text{ex}} + W_\alpha(z) - z$ to Eq. (69) we obtain the differential equations

$$\begin{aligned}&(H^{\text{ex}} + W_\alpha(z) - z) G^\alpha(X, X', z) \\ &\quad = \delta_{\alpha_1} \delta(X - X') - W_\alpha(z) \sum_{\beta \neq \alpha} G^\beta(X, X', z)\end{aligned} \tag{71}$$

to be fulfilled by the components $G^\alpha(z)$.

By Eq. (47) the Faddeev components $G^\alpha(z)$ obey the boundary conditions

$$[\partial_n G^\alpha]_{\Gamma_\alpha} = -\varphi_\alpha Q_\alpha(z) \sum_\beta \langle G^\beta(z)\cdot,\varphi_\alpha\rangle. \qquad (72)$$

Here we have used the following property of $G^\alpha(z)$:

$$[\partial_n G^\beta]_{\Gamma_\alpha} = 0, \quad \alpha\neq\beta, \qquad (73)$$

which can be obtained by virtue of the usual properties of simple layer potentials.[18]

If Z is real, i.e., $Z = E \pm i0$, the differential equations (71) together with the boundary conditions described in Refs. 19 and 20 define a unique solution for the wave functions.

The methods of Ref. 15 may be used to solve this boundary-value problem numerically.

At the end we would like to point out that all results and statements about equations for the Faddeev components $G^\alpha(z)$ and for the operators $M_\alpha(z)$ can be rigorously obtained only on the basis of equations for the total resolvent $R(z)$ or, in other words, on the basis of the s.a. Hamiltonian H.

VIII. DISCUSSION

In this paper, we have presented a new approach toward a mathematically correct study of the scattering theory for few-body systems with energy-dependent potentials. The main result is that the treating of such systems in usual configuration space is inconsistent from an operator point of view. We have demonstrated that an energy dependence of the potentials is generated by the internal structure of the interacting particles. This energy dependence, however, turns out not to be arbitrary, since it is given by some class of operator-valued R functions, including, in particular, Schwartz integrals as described above.

The main effect incorporated in our scheme is the possibility to separate the contributions from two-body and three-body forces. We remark that also many-body forces can be included into our consideration without a drastic change of the formulation.

ACKNOWLEDGMENTS

We are indebted to L. D. Faddeev for encouragement and fruitful discussions and to W. Plessas for some essential remarks, which improved the content of this manuscript.

APPENDIX A: PROOF OF THEOREM 1

Let us prove that

$$\mathscr{D}(H_0)\subset\mathscr{D}(\widetilde{H})\subset\mathscr{D}(H_0^*). \qquad (A1)$$

Let $\xi_\alpha^-(y_\alpha)$ be smooth functions with compact supports from $L^2(\mathbb{R}^3_{y_\alpha})$ and define the densities ρ_α and the functions $\xi_\alpha^+(y_\alpha)$ by (28) and (29).

The representations (28) and (29) give us a possibility to reconstruct the components u_a, $a = 0,1,2,3$ with arbitrary elements $\tilde{u}_\alpha\in\mathscr{D}(H^{\rm in}_{\alpha_0})$ and $\tilde{u}_0\in C_0^\infty(\mathbb{R}^6\setminus\Gamma)$. The element $\mathscr{U} = \{u_a\}$ defined in such manner, belongs to the domain $\mathscr{D}(H_0^*)$. To prove this it is sufficient to verify that the internal components $u_\alpha\in\mathscr{D}(H^{\rm in*}_{\alpha_0})$ or, what is the same, that the functions $\xi_\alpha^+\in W_2^2(\mathbb{R}^3_{y_\alpha})$.

The simple layer potential $R_0\rho_\beta$ given as a function on the cylinder Γ_α, $\alpha\neq\beta$ is not a smooth function in the neighborhood of the set $\Gamma_0 = \cap_\alpha \Gamma_\alpha$. If $\alpha=\beta$ then the smoothness of $R_0\rho_\beta$ is defined by the smoothness of its density ρ_β. Nevertheless, the average $\langle R_0\rho_\beta,\varphi_\alpha\rangle(y_\alpha)$ on the section $\{y_\alpha = \text{const}\}$ of the cylinder Γ_α is the W_2^2 smooth function. The proof of the last statement is based on the following local representation:

$$R_0\rho_\beta(X) = \tfrac{1}{2}\,\mathrm{dist}(X,\Gamma_\beta)\rho_\beta(S_\beta) + \widetilde{R}_0(X),$$

which takes place for the smooth ρ_β. Here, S_β is the projection of the point X on the cylinder Γ_β and $\widetilde{R}_0(X)$ is a smooth function. In fact, one must verify the smoothness of the average $\langle \mathrm{dist}(\cdot,\Gamma_\beta),\varphi_\alpha\rangle(y_\alpha)$ in the neighborhood of the set Γ_0, considered as a function of the distance between the plane $\{y_\alpha = \text{const}\}$ and the set Γ_β. It can be done immediately. Thus we have proved that (A1) is true.

As a consequence of (A1), the closure of $\mathscr{D}(\widetilde{H})$ coincides with the total Hilbert space $\mathfrak{H}$.

To prove the symmetry of the operator $\widetilde{H}$, one must calculate the boundary form

$$J(\mathscr{U},\mathscr{V}) = \langle\widetilde{H}\mathscr{U},\mathscr{V}\rangle - \langle\mathscr{U},\widetilde{H}\mathscr{V}\rangle. \qquad (A2)$$

In order to estimate the contribution of the external channel operator into the total boundary form (A2), it is convenient to make such calculations for a system of smooth "parallel surfaces":

$$\Gamma^1_\delta = \{X:\ \ \mathrm{dist}(X,\Gamma) = \delta\}, \quad \delta>0,$$
$$\Gamma^0_{6\delta} = \{X:\ \ \mathrm{dist}(X,\Gamma_0) = 6\delta\}, \quad \delta>0,$$

in the limit $\delta\to 0$. The integration over the pieces $\Gamma^1_\delta \cap\{X:\ \ \mathrm{dist}(X,\Gamma_\alpha)\geqslant 6\delta\}$ gives the sum of the integrals over the cylinders Γ_α when $\delta\to 0$. The contribution from the integration over $\Gamma^0_{6\delta}\cap\{X:\ \ \mathrm{dist}(X,\Gamma_0)\geqslant\delta\}$, $\delta\to 0$, vanishes, if the simple layer potentials $R_0\rho_\alpha$ were generated by smooth densities. It takes place because such simple layer potentials $R_0\rho_\alpha$ have both uniformly bounded values and bounded normal derivatives on the surface $\Gamma^0_{6\delta}$.

Thus the calculation of the boundary form corresponding to the external channel can be reduced to the evaluation of the contribution from every cylinder Γ_α:

$$J(u_0,v_0) = \sum_{\alpha=1}^{3} \lim_{\delta\to 0} \int_{\Gamma^\delta_\alpha = \gamma_\alpha\times\mathbb{R}^3_{y_\alpha}} dS_\alpha(\partial_n u_0\bar{v}_0 - \partial_n\bar{v}_0 u_0).$$

The external components $u_0,v_0\in\mathscr{D}(\widetilde{H})$ satisfy the boundary conditions (22),(23), in the form (28),(29). So, the contribution into the external boundary form can be written as

$$\lim_{\delta\to 0}\int_{\Gamma^\delta_\alpha} dS_\alpha([\partial_n u_0]_{\Gamma_\alpha}\bar{v}_0 - u_0[\partial_n\bar{v}_0]_{\Gamma_\alpha})$$
$$= \int_{\Gamma_\alpha} dS_\alpha\Bigg[\overline{\Big(\sum_\beta R_0\rho_\beta(v_0)\Big)}\varphi_\alpha\xi_\alpha^-(u_\alpha) - \sum_\beta R_0\rho_\beta\ \overline{(u_0)\xi_\alpha^-(v_\alpha)\varphi_\alpha}\Bigg]$$
$$= -\int dy_\alpha(\xi_\alpha^-(u_\alpha)\ \overline{\xi_\alpha^+(v_\alpha)} - \xi_\alpha^+(u_\alpha)\ \overline{\xi_\alpha^-(v_\alpha)}).$$

This coincides with the contribution from the boundary form of the operator $H^{\rm in*}_{\alpha_0}$ acting in the internal channel $\mathfrak{H}^{\rm in}_\alpha$.

Hence, the total boundary form (A2) is equal to zero. In other words, the operator $\tilde{H}$ is the symmetric one.

To prove that the operator $\tilde{H}$ is bounded from below consider its quadratic form on the domain $\mathscr{D}(\tilde{H})$ [We omit here, for simplicity, all energy-independent potentials $v_\alpha(x_\alpha)$ in the external-channel Hamiltonian H^{ex}.]:

$$\langle \tilde{H}\mathscr{U},\mathscr{U}\rangle = \langle -\Delta_X u_0,u_0\rangle + \sum_\alpha \langle (A^*_{\alpha_0} - \Delta_{y_\alpha})u_\alpha,u_\alpha\rangle.$$

Integrating by parts

$$\langle \tilde{H}\mathscr{U},\mathscr{U}\rangle = \|\nabla u_0\|^2_{L^2(\mathbf{R}^6)} - \sum_\alpha \int_{\Gamma_\alpha} dS_\alpha [\partial_n u_0]_{\Gamma_\alpha}\bar{u}_0 + \sum_\alpha \left\{\langle A^*_{\alpha_0}u_\alpha,u_\alpha\rangle + \|\nabla u_\alpha\|^2_{L^2(\mathbf{R}^3_{y_\alpha})}\right\}, \tag{A3}$$

one can show that the boundary terms in (A3) can be estimated by the Dirichlet form of the operators $-\Delta_X$ and $-\Delta_{y_\alpha}$ and also by the norms of the elements from the external and internal channels.

First, let us estimate the quadratic form of the operator $A^*_{\alpha_0}$. In the representation

$$u_\alpha = \tilde{u}_\alpha + \xi^+_\alpha w^+_\alpha + \xi^-_\alpha w^-_\alpha, \tag{A4}$$

the boundary values $\xi^\pm_\alpha$ are not arbitrary but are connected by the relation

$$\xi^-_\alpha\langle\theta_\alpha,\theta_\alpha\rangle = \langle (A_\alpha - i)u_\alpha,\theta_\alpha\rangle - \xi^+_\alpha\langle A_\alpha\theta_\alpha,\theta_\alpha\rangle. \tag{A5}$$

This formula can be derived immediately from the condition[7] $\theta_\alpha \perp (A-i)\tilde{u}_\alpha$, where θ_α is the generative element of the Hamiltonian A_α. By the relation (A5) one can estimate the L^2-norm of the boundary vector ξ^-_α:

$$\|\xi^-_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})} \leqslant C\left[\|\xi^+_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})} + \|u_\alpha\|_{\mathfrak{H}^{\mathrm{in}}_\alpha}\right]. \tag{A6}$$

By means of decomposition (A4) we have

$$\|\tilde{u}_\alpha\|_{\mathfrak{H}^{\mathrm{in}}_\alpha} \leqslant C\left[\|u_\alpha\|_{\mathfrak{H}^{\mathrm{in}}_\alpha} + \|\xi^+_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})} + \|\xi^-_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})}\right] \leqslant C\left[\|\xi^+_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})} + \|u_\alpha\|_{\mathfrak{H}^{\mathrm{in}}_\alpha}\right]. \tag{A7}$$

By the equality

$$A^*_{\alpha_0}u_\alpha = A_\alpha\tilde{u}_\alpha - \xi^+_\alpha w^-_\alpha + \xi^-_\alpha w^+_\alpha,$$

and under the assumption that the operator A_α is bounded, one has

$$|\langle A^*_{\alpha_0}u_\alpha,u_\alpha\rangle| \leqslant C\left[\|u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_\alpha} + \|\xi^+_\alpha\|^2_{L^2(\mathbf{R}^3_{y_\alpha})}\right]. \tag{A8}$$

Then, taking into account (23) one can estimate the quadratic form (A8) in terms of the external element

$$\|\xi^+_\alpha\|_{L^2(\mathbf{R}^3_{y_\alpha})} \leqslant \|\varphi_\alpha\|_{L^2(\gamma_\alpha)}\|u_0\|_{L^2(\Gamma_\alpha)}. \tag{A9}$$

Therefore,

$$\left|\sum_\alpha \langle A^*_{\alpha_0}u_\alpha,u_\alpha\rangle\right| \leqslant C\left[\sum_\alpha \left(\|u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_\alpha} + \|u_0\|^2_{L^2(\Gamma_\alpha)}\right)\right]. \tag{A10}$$

By the condition (22) and the relations (A6),(A9) we obtain

$$\left|\sum_\alpha \int_{\Gamma_\alpha} dS_\alpha [\partial_n u_0]_{\Gamma_\alpha}\bar{u}_0\right| \leqslant C\left[\sum_\alpha \left(\|u_0\|^2_{L^2(\Gamma_\alpha)} + \|\xi^-_\alpha\|^2_{L^2(\mathbf{R}^3_{y_\alpha})}\right)\right] \leqslant C\sum_\alpha \left(\|u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_\alpha} + \|u_0\|^2_{L^2(\Gamma_\alpha)}\right). \tag{A11}$$

Using the embedding theorem

$$\int_\Gamma |u_0|^2\, dS \leqslant C\left(\delta \int |\nabla u_0|^2\, dX + \frac{1}{\delta}\int |u_0|^2\, dX\right), \quad \delta > 0, \tag{A12}$$

and collecting together (A10)–(A12) we have

$$\langle \tilde{H}\mathscr{U},\mathscr{U}\rangle \geqslant (1-\delta C)\|\nabla u_0\|^2_{L^2(\mathbf{R}^6)} + \sum_\alpha \|\nabla u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_2} - \frac{C}{\delta}\|u_0\|^2_{L^2(\mathbf{R}^6)} - C\sum_\alpha \|u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_\alpha}.$$

If $\delta C < 1$ then

$$\langle \tilde{H}\mathscr{U},\mathscr{U}\rangle \geqslant \max\left\{\frac{C}{\delta},C\right\}\left[\|u_0\|^2_{\mathfrak{H}^{\mathrm{ex}}} + \sum_\alpha \|u_\alpha\|^2_{\mathfrak{H}^{\mathrm{in}}_\alpha}\right] = -C\|\mathscr{U}\|^2_{\mathfrak{H}}.$$

It means that the operator $\tilde{H}$ is bounded from below. So, Theorem 1 is completely proved.

Let us make some comments about essential points of the proof. The most important question concerns the presence of three-body forces in the model. From the geometrical point of view such kind forces are connected with the boundary conditions, which may be stated on the manifold $\Gamma_0 = \cap_\alpha \Gamma_\alpha$. The deficiency subspaces of the operator H_0, corresponding to the manifold Γ_0, are parametrized by simple layer densities belonging to the Sobolev class $W_2^{-3/2}$. In order to conserve the pair character of the boundary conditions (28),(29) we do not include such deficiency elements into the domain $\mathscr{D}(\tilde{H})$.

APPENDIX B: PROOF OF THEOREM 2

Let us check the Fredholm nature of Eq. (57). Although the potentials $W_\alpha(z)$ are energy-dependent integral operators for variable y_α [see (48) and (49)] the kernels $V_\alpha G_\alpha(S,X',z)$ have both standard analytical properties in the variable z and standard asymptotical behavior in the variables S,X', which are typical for analogous kernels in the potential model[20] and in the boundary conditions model.[10,11]

The representation (18) of the operator H_α ensures validity of the relation

$$V_\alpha G_\alpha(z) = \frac{1}{2\pi i}\oint_{\mathscr{S}_\alpha} d\lambda\, v_\alpha g_\alpha(\lambda) r_0^{(\alpha)}(\lambda - z), \tag{B1}$$

where $g(z)$ is the generalized Green's function of the "operator" from Eq. (17) and $V(z)$ is the integral operator in the representation (15) such that

$$w(z)\cdot = \delta_\gamma v(z), \tag{B2}$$

$$v(x,x',z) = -\varphi(x)\Delta(z)\overline{\varphi(x')}. \tag{B3}$$

The Hamiltonian h of the two-particle system has both a discrete and continuous spectrum. We shall consider the situation when the Hamiltonian h has one bound state χ of energy $-\varkappa^2$. The corresponding decomposition of the kernel $vg(z)$ looks as follows

$$vg(x,x',z) = -\varphi(x)\chi(x')[(z+\varkappa^2)N_0]^{-1} + vg^c(x,x',z), \tag{B4}$$

where

$$N_0^2 = -\frac{d}{dz}[\Delta^{-1}(z) - \langle g_0(z)\varphi,\varphi\rangle]|_{z=-\varkappa^2}, \tag{B5}$$

and $g_0(z) = (h^{\rm ex} - z)^{-1}$ is the resolvent of the s.a. operator $h^{\rm ex}$.

The decomposition (B4) leads to the following representation of the kernel $V_\alpha G_\alpha(z)$:

$$V_\alpha G_\alpha = V_\alpha G_\alpha^d + V_\alpha G_\alpha^c, \tag{B6}$$

where

$$V_\alpha G_\alpha^d(S,X',z) = -\varphi_\alpha(s_\alpha)\chi_\alpha(x'_\alpha)N_{0\alpha}^{-1}r_0^{(\alpha)} \times (y_\alpha - y'_\alpha, z+\varkappa^2) \tag{B7}$$

and $S=\{s_\alpha,y_\alpha\}$, $s_\alpha\in\gamma_\alpha$, $y_\alpha\in\mathbb{R}^3_{y_\alpha}$. The kernel $V_\alpha G_\alpha^c$ describes the contribution from the continuous spectrum.

We shall need the asymptotic of the kernel $V_\alpha G_\alpha^c(S,X',z)$ when $x'_\alpha\to\infty$. This asymptotic can be obtained by the saddle-point method[21] from the asymptotic of the corresponding kernel vg^c related to the two-body problem,

$$vg^c(x,x',z) \underset{x'\to\infty}{\sim} -\varphi(x) < \psi_p^{(+)},\varphi > \frac{\exp\{i\sqrt{z}|x'|\}}{d(z)4\pi|x'|}. \tag{B8}$$

Here, $\psi_p^{(+)}$ is the wavefunction of the continuous spectrum corresponding to the operator $h^{\rm ex}$, $p = -\sqrt{z}\hat{x}'$ and

$$d(z) = \Delta^{-1}(z) - \langle g_0(z)\varphi,\varphi\rangle. \tag{B9}$$

Then the asymptotic of the kernel $V_\alpha G_\alpha^c(S,X',z)$ when $x'_\alpha\to\infty$ looks as follows:

$$V_\alpha G_\alpha^c(S,X',z) \underset{x'_\alpha\to\infty}{\sim} \varphi_\alpha(s_\alpha)C_0(z) \times\exp\{i\sqrt{z}L_{\alpha_0}\}L_{\alpha_0}^{-5/2}\mathcal{F}_\alpha \times(z\cos^2\omega_\alpha,\hat{X}'). \tag{B10}$$

Here, L_{α_0} is the eikonal[20] that corresponds to the propagation of the ray from the point $\{0,y_\alpha\}$ to the point $\{x'_\alpha,y_\alpha\}$:

$$L_{\alpha_0}(y_\alpha,X') = [|x'|^2 + (y_\alpha - y'_\alpha)^2]^{1/2}. \tag{B11}$$

The function $\mathcal{F}$ looks like

$$\mathcal{F}_\alpha(z,\hat{X}') = -d_\alpha^{-1}(z)\langle\psi_{p_\alpha}^{(+)},\varphi_\alpha\rangle, \tag{B12}$$

where $p_\alpha = -\sqrt{z}\hat{x}'_\alpha$ and

$$C_0(z) = (1/4\pi)(i(\sqrt{z}/2\pi))^{3/2},$$
$$\cos\omega_\alpha = |x'_\alpha|/L_{\alpha_0}. \tag{B13}$$

Using the relations obtained above one can notice that the asymptotic and analytic properties of the kernels $V_\alpha G_\alpha$ coincide with those of the usual three-body problem with energy-independent potentials.[20] Hence, one can prove the Fredholm property for Eq. (57) using the techniques, proposed in Refs. 10 and 11. Namely, the kernels V_α and G_α have the singularities $|y_\alpha - y'_\alpha|^{-1}$ and $|S_\alpha - S'_\beta|^{-4}$, respectively. Consequently, a restriction of the operators $V_\alpha G_\alpha$ to any bounded part of the surface $\Gamma = \cup_\alpha \Gamma_\alpha$ leads to a compact operator, because $\dim\Gamma_\alpha = 5$. On the other hand, because of the slow decrease of the kernels $V_\alpha G_\alpha(S,S',z)$ at infinity, the operator $B(z)$ is not compact. However, due to the Faddeev structure of the operator $B(z)$, the arguments of the kernels $V_\alpha G_\alpha$ are located on different cylinders Γ_α and Γ_β, $\alpha\neq\beta$ and during the iteration procedure the products of operators $V_{\alpha_1}G_{\alpha_1}V_{\alpha_2}G_{\alpha_2}\cdots V_{\alpha_n}G_{\alpha_n}$ only occur and $\alpha_i \neq\alpha_{i+1}$.

Using the representation (B6) one can pick up terms of the following types in the kernels of the operators $B^n(z)$:

(1) The products $V_{\alpha_1}G_{\alpha_1}^c\cdots V_{\alpha_j}G_{\alpha j}^d\cdots V_{\alpha_n}G_{\alpha_n}$ which contain not less than one operator $V_{\alpha_j}G_{\alpha_j}^d$ corresponding to the discrete spectrum of the two-body subsystem Hamiltonian;

(2) The product $V_{\alpha_1}G_{\alpha_1}^c\cdots V_{\alpha_n}G_{\alpha_n}^c$ which contains operators corresponding to the continuous spectrum only.

Due to the existence of the exponentially decreasing eigenfunctions $\chi_\beta(x'_\beta)$ in the kernel $V_\beta G_\beta^d$ the asymptotics of the first-type kernels (which have $V_{\alpha_1}G_{\alpha_1}^c$ and $V_{\alpha_n}G_{\alpha_n}^c$ at the beginning and at the end of product, respectively) is described in terms of spherical waves in $\mathbb{R}^6$.

The asymptotic of the second-type kernels can be investigated by the stationary-phase method.[21] Due to the existence of the phase with the eikonal L_{β_0} in the asymptotics (B10) of the kernel $V_\beta G_\beta^c$, the phase of the integrand is defined by the sum of distances between the points located on the cylinders $\Gamma_{\alpha_1}, \Gamma_{\alpha_2}, \cdots, \Gamma_{\alpha_n}$ axes $\{0,y_{\alpha_i}\}$, $y_{\alpha_i}\in\mathbb{R}^3$.

Minimalization of such a sum in the framework of the stationary-phase method gives the length of the trajectory of the ray that goes from the point $S'_{\alpha_n} = \{0,y'_{\alpha_n}\}$ to the point $S_{\alpha_1} = \{0,y_{\alpha_1}\}$ by reflecting on the cylinders $\Gamma_{\alpha_2},\ldots,\Gamma_{\alpha_{n-1}}$ axes. As a result one can obtain the description of the asymptotics in terms of the corresponding eikonal.[20] If such a process is impossible, i.e., the corresponding minimalization leads to a spherical eikonal $|S| + |S'|$, the kernel $V_{\alpha_1}G_{\alpha_1}^c\cdots V_{\alpha_n}G_{\alpha_n}^c$ asymptotically turns in a product of spherical waves in $\mathbb{R}^6$ in variables S and S' when S and $S'\to\infty$. Maximal number $N_{\max}$ of possible reflections on the cylinders axes is defined by the angles between these axes. Thus the nth power $B^n(z)$ of the operator $B(z)$ is a compact operator in a proper Banach space when $n > N_{\max}$. Hence, the Fredholm alternative applies to Eq. (57). The first statement of Theorem 2 is proved.

[1] H. Feshbach, Ann. Phys. **5**, 357 (1958); **19**, 287 (1962).
[2] R. F. Dashen, F. E. Low, and I. J. Muzinich, Ann. Phys. **102**, 1 (1976).
[3] E. W. Schmid and G. Spitz, Z. Phys. A **321**, 581 (1985).
[4] R. L. Jaffe and F. E. Low, Phys. Rev. D **19**, 2105 (1979).
[5] Yu. A. Simonov, Phys. Lett. B **107** (1981); Yad. Fiz. **36**, 722 (1982).
[6] I. M. Narodetskii, *Few-Body Problems in Physics*, edited by L. D. Faddeev and T. L. Kopaleishvili (World Scientific, Singapore, 1985).
[7] B. S. Pavlov, Teor. Mat. Fiz. **59**, 345 (1984).
[8] B. S. Pavlov and M. D. Faddeev, Zap. Nauch. Semin. LOMI **126**, 159 (1983).
[9] Yu. A. Kuperin, K. A. Makarov, and B. S. Pavlov, Teor. Mat. Fiz. **63**, 78 (1985); **69**, 100 (1986).

[10]S. P. Merkuriev and A. K. Motovilov, Lett. Math. Phys. **7**, 497 (1983); Preprint Inst. des Sciences Nucléaires de Grenoble ISN 83.41, 1983.
[11]A. A. Kvitsinsly, Yu. A. Kuperin, S. P. Merkuriev, A. K. Motovilov, and S. L. Yakovlev, Elem. Part. Atom. Nucl. **17**, 267 (1986) (in Russian).
[12]Yu. A. Kuperin, K. A. Makarov, S. P. Merkuriev, B. S. Pavlov, and A. K. Motovilov, ITP-Budapest-Rep. No. 441, Budapest, 1986.
[13]M. S. Birman and M. Z. Solomjak, *The Spectral Theory of Selfadjoint Operators in the Hilbert Space* (Leningrad U.P., Leningrad, 1980) (in Russian).
[14]V. S. Vladimirov, *Distributions in Mathematical Physics* (Nauka, Moscow, 1979) (in Russian).
[15]S. P. Merkuriev, C. Gignoux, and A. Laverne, Ann. Phys. **99**, 30 (1976).
[16]L. D. Faddeev, Trudy Mat. In-ta Akad. Nauk. SSSR **69**, 1 (1963).
[17]M. G. Krein, Dokl. Akad. Nauk. **52**, 657 (1946).
[18]V. I. Smirnov, *A Course of Higher Mathematics* (Pergamon, Oxford, 1964), Vol. 4.
[19]S. P. Merkuriev, Ann. Phys. **130**, 395 (1980).
[20]S. P. Merkuriev and L. D. Faddeev, *Quantum Scattering Theory for Few-Body Systems* (Nauka, Moscow, 1985) (in Russian).
[21]M. V. Fedorjuk, *A Saddle-Point Method* (Nauka, Moscow, 1977) (in Russian).

Low-frequency moments in inverse scattering theory

George Dassios
Division of Applied Mathematics, Department of Chemical Engineering, University of Patras and Institute of Chemical Engineering and High Temperature Chemical Processes, GR 261 10 Patras, Greece

(Received 6 October 1989; accepted for publication 21 February 1990)

The low-frequency moments of the scattering amplitude are utilized in order to identify the capacity, the center, and the orientation of an acoustically soft scatterer.

Let the closed, connected, and smooth surface S describe the boundary of an acoustically soft scatterer, whose exterior is denoted by V. The scatterer is excited by the plane incident wave

$$\Phi(\mathbf{r},t) = e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \tag{1}$$

where $\mathbf{k}$ stands for the propagation vector and ω is the angular frequency. Suppressing the harmonic time dependence $\exp(-i\omega t)$ and introducing the time-independent total field

$$\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} + u(\mathbf{r}), \tag{2}$$

with $u(\mathbf{r})$ being the scattered wave, we arrive at the following direct scattering problem.[1] We need to find the field $\Psi(\mathbf{r})$ that solves the boundary value problem

$$(\Delta + k^2)\Psi(\mathbf{r}) = 0, \quad \mathbf{r}\in V, \tag{3}$$

$$\Psi(\mathbf{r}) = 0, \quad \mathbf{r}\in S, \tag{4}$$

$$\partial_r u(\mathbf{r}) - iku(\mathbf{r}) = O(1/r^2), \quad r\to +\infty. \tag{5}$$

Utilizing the analyticity of the total field at the point $k = 0$[2] we can introduce the low-frequency expansion[3,2]

$$\Psi(\mathbf{r}) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!}\Phi_n(\mathbf{r}), \tag{6}$$

and reduce the problem (3)–(5) into a sequence of boundary value problems for the low-frequency coefficients Φ_n, $n = 0,1,2,\ldots$ that can be solved iteratively. Specificly, the coefficient Φ_n has to satisfy the following exterior problem:[4]

$$\Delta\Phi_n(\mathbf{r}) = n(n-1)\Phi_{n-2}(\mathbf{r}), \quad \mathbf{r}\in V, \tag{7}$$

$$\Phi_n(\mathbf{r}) = 0, \quad \mathbf{r}\in S, \tag{8}$$

$$\Phi_n(\mathbf{r}) = (\hat{\mathbf{k}}\cdot\mathbf{r})^n - \frac{1}{4\pi}\sum_{\rho=0}^{n-1}\binom{n}{\rho+1}\int_S |\mathbf{r}-\mathbf{r}'|^\rho \times \partial_{n'}\Phi_{n-\rho-1}(\mathbf{r}')ds(\mathbf{r}') + O(1/r), \quad r\to\infty. \tag{9}$$

The normalized scattering amplitude $g(\mathbf{r},\mathbf{k})$ that appears in the asymptotic expansion

$$u(\mathbf{r}) = g(\hat{\mathbf{r}},\hat{\mathbf{k}})(e^{ikr}/ikr) + O(1/r^2), \quad r\to+\infty, \tag{10}$$

and assumes the integral representation

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}}) = -\frac{ik}{4\pi}\int_S e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}\partial_{n'}\Psi(\mathbf{r}')ds(\mathbf{r}'), \tag{11}$$

yields the following low-frequency expansion:[4]

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \sum_{n=0}^{\infty}\frac{(ik)^{n+1}}{n!}\sum_{\rho=0}^{n}\binom{n}{\rho}(-1)^{\rho+1}M_{n-\rho}^{(\rho)}(\hat{\mathbf{r}}), \tag{12}$$

where

$$M_{n-\rho}^{(\rho)}(\hat{\mathbf{r}}) = \frac{1}{4\pi}\int_S (\hat{\mathbf{r}}\cdot\mathbf{r}')^\rho\,\partial_{n'}\Phi_{n-\rho}(\mathbf{r}')dS(\mathbf{r}') \tag{13}$$

stands for the low-frequency moment of order ρ generated by the $(n-\rho)$ coefficient.

The inverse scattering problem we want to consider here concerns the recovery of information about the shape and the orientation of S from the knowledge of the scattering amplitude.

As it is seen from (12) all the information about the shape and the orientation of the scatterer is hidden in the low-frequency moments $M_{n-\rho}^{(\rho)}(\mathbf{r})$ and we want to investigate the particular information carried by each one of the first few of them.

As it is well known,[5,2]

$$M_0^{(0)}(\hat{\mathbf{r}}) = C, \tag{14}$$

where C denotes the electrostatic capacity of the scatterer. Many estimates for the capacity of a body can be found in Ref. 6.

The low-frequency moment of the first order generated by the Φ_0 coefficient is given by

$$M_0^{(1)}(\hat{\mathbf{r}}) = \hat{\mathbf{r}}\cdot\left[\frac{1}{4\pi}\int_s \mathbf{r}'\,\partial_{n'}\Phi_0(\mathbf{r}')ds(\mathbf{r}')\right]. \tag{15}$$

The expression in brackets defines a fixed vector that is nothing else but the "center" of the scatterer with respect to the surface measure induced by the normal derivative of the electrostatic potential Φ_0. Consequently, the center of the scatterer is given by the vector

$$\mathbf{r}_0 = \frac{1}{4\pi}\int_s \mathbf{r}'\,\partial_{n'}\Phi_0(\mathbf{r}')ds(\mathbf{r}'). \tag{16}$$

If we choose $\mathbf{r}_0$ to be the origin of the coordinates, then $M_0^{(1)}(\mathbf{r})$ vanishes. Furthermore, if the scatterer has inversion symmetry ($\mathbf{r}\in S$ implies $-\mathbf{r}\in S$) then

$$M_0^{(2n+1)}(\hat{\mathbf{r}}) = 0, \quad n = 0,1,2,\ldots, \tag{17}$$

while

$$M_0^{(2n)}(\hat{\mathbf{r}}) \neq 0, \quad n = 0,1,2,\ldots.$$

For the low-frequency moment of the second order that is generated by the potential Φ_0 we obtain

$$M_0^{(2)}(\hat{\mathbf{r}}) = \hat{\mathbf{r}}\otimes\hat{\mathbf{r}}:\left[\frac{1}{4\pi}\int_s \mathbf{r}'\otimes\mathbf{r}'\,\partial_{n'}\Phi_0(\mathbf{r}')ds(\mathbf{r}')\right]. \tag{18}$$

The quantity inside the brackets is a real symmetric tensor and it is identified with the inertia tensor of the scatterer with

 0022-2488/90/071691-02$03.00

respect to the surface measure

$$dm_0(\mathbf{r}) = \partial_n \Phi_0(\mathbf{r}) ds(\mathbf{r}). \tag{19}$$

Consequently, the diagonalization of

$$M_0 = \frac{1}{4\pi} \int_s \mathbf{r} \otimes \mathbf{r} \, dm_0(\mathbf{r}) \tag{20}$$

will reveal the principal directions of the scatterer.

Therefore, if the first three low-frequency moments generated by the leading low-frequency approximation are known, then (i) a first estimate of the volume of the scatterer (via its capacity[6]) can be found, (ii) the center of the scatterer can be evaluated, and (iii) the orientation of the scatterer (through its principal axes) can be specified. Once these three characteristics are obtained, we choose a coordinate system with its origin at the center of the scatterer and with axes along its principal directions. This particular choice of coordinate system could be the starting point for applying the more elaborate inverse scattering techniques of Angell *et al.*,[7] Colton and Monk,[8] or Kirsch and Kress[9] that will provide the detailed characteristics of the scatterer. Of course, the above steps require a first approximation to the position of the scatterer in space, but this can be accomplished (as Sleeman has indicated[10]) by using a least-squares search[11] to localize the scatterer.

It is of interest to note that for the case of an ellipsoid centered at the origin[4]

$$\partial_n \Phi_0(\mathbf{r}) = 2\left(\sqrt{\alpha_1^2 - \mu^2}\sqrt{\alpha_1^2 - \nu^2} \times \int_0^{+\infty} \frac{dx}{\sqrt{x+\alpha_1^2}\sqrt{x+\alpha_2^2}\sqrt{x+\alpha_3^2}}\right)^{-1}, \tag{21}$$

where $\alpha_1 > \alpha_2 > \alpha_3 > 0$ are the three semiaxes of the ellipsoid and μ, ν stand for the angular ellipsoidal coordinates, and

$$M_0 = 2[\alpha_1^2 \hat{\mathbf{x}}_1 \otimes \hat{\mathbf{x}}_1 + \alpha_2^2 \hat{\mathbf{x}}_2 \otimes \hat{\mathbf{x}}_2 + \alpha_3^2 \hat{\mathbf{x}}_3 \otimes \hat{\mathbf{x}}_3] \times \left(3\int_0^{+\infty} \frac{dx}{\sqrt{x+\alpha_1^2}\sqrt{x+\alpha_2^2}\sqrt{x+\alpha_3^2}}\right)^{-1}. \tag{22}$$

Using a method essentially equivalent to the one developed in Ref. 12, we can show that for the case of an ellipsoid $M_0^{(0)}$ and $M_0^{(2)}$, or C and M_0, are enough to uniquely identify the size (three semiaxes) and the orientation (three Euler angles) of the scatterer. Evidently, the restriction of our requirements to just two low-frequency moments generated by the electrostatic potential for the case of an ellipsoid reflects the simplicity and the symmetry of this particular scatterer. On the other hand, we observe that the electrostatic potential Φ_0 is the only low-frequency coefficient that enters every term of the expansion (12) for the scattering amplitude. It seems that it is possible to recover the geometry of "reasonable" scatterers from a knowledge of all the low-frequency moments generated by Φ_0, i.e., we can state the following conjecture.

Conjecture: If all the low-frequency moments $\{M_0^{(n)}(r)\}_{n=0}^{\infty}$ are known on the unit sphere then the surface S of the scatterer can be recovered.

ACKNOWLEDGMENTS

The author wants to thank Professor B. D. Sleeman for fruitfull discussions during the Ross Priory Workshop on Inverse Problems, in July of 1989.

The present work was partially supported by the Research Institute of Chemical Engineering and High-Temperature Chemical Processes.

[1] R. E. Kleinman, A.R.M.A. **18**, 205–229 (1965).
[2] R. E. Kleinman and T. B. A. Senior, in *Low- and High-Frequency Asymptotics*, edited by V. K. Varadan and V. V. Varadan (North-Holland, Amsterdam, 1986) pp. 1–70.
[3] G. Dassios, SIAM J. Appl. Math. **42**, 272–280 (1982).
[4] G. Dassios, SIAM J. Appl. Math. **38**, 373–381 (1980).
[5] J. B. Keller, R. E. Kleinman, and T. B. A. Senior, J.I.M.A. **9**, 14–22 (1972).
[6] L. E. Payne, "Isoperimetric Inequatities and their Applications", SIAM Rev. **9**, 453–488 (1967).
[7] T. S. Angell, R. E. Kleinman, B. Kok, and G. F. Roach, Wave Motion **11**, 185–200 (1989).
[8] D. Colton and P. Monk, SIAM J. Appl. Math. **45**, 1039–1053 (1985); and **46**, 506–523 (1986).
[9] A. Kirsch and R. Kress, in *Boundary Elements* IX, edited by C. Brebbia, W. Wendland, and G. Kuhn (Springer, Heidelberg, 1987) pp. 3–18.
[10] B. D. Sleeman, Proceedings of the Workshop on Inverse Problems and Imaging, University of Strathcyde, Glasgow, (July 1989).
[11] W. A. Imbriale, and R. Mittra, IEEE Trans. Antennas Propag. **AP-18**, 633–642 (1970).
[12] G. Dassios, J. Math. Phys. **28**, 2858–2862 (1987).

Summational invariants in systems of mass points and rigid bodies

Wolfgang Huber[a)]
Institut für Theoretische Physik der Universität Göttingen, Bunsenstrasse 9, D-3400 Göttingen, West Germany

(Received 18 October 1988; accepted for publication 7 March 1990)

For systems of rigid bodies and mass points, the general form of summational invariants will be derived using some results of Amigó and Reeh [J. Math. Phys. **24**, 1594 (1983)].

I. INTRODUCTION AND RESULTS

Recently, Amigó and Reeh derived the structure of summational invariants of a system of mass points.[1] Under reasonable physical conditions these are of the form

$$f_i(\mathbf{x},\mathbf{p},t) = A_1(\mathbf{p}^2/2m_i) + \mathbf{B}_1\cdot\mathbf{p} + \mathbf{B}_2(\mathbf{x}\times\mathbf{p}) + \mathbf{B}_3\cdot\mathbf{r}_i + A_2(\mathbf{p}\cdot\mathbf{r}_i/m_i) + A_3(\mathbf{r}_i^2/m_i) + k_i, \tag{1}$$

with $\mathbf{r}_i := m_i\mathbf{x} - \mathbf{p}t$ in the Galilei invariant case. Only the constant k_i may depend on the particle under consideration (see the Appendix). Under reasonable conditions the expressions $\sim\mathbf{r}_i^2$ and $\sim\mathbf{r}_i\cdot\mathbf{p}$ appear only in the case of the very special interaction potential $V_{ij} := \alpha/|\mathbf{r}_i - \mathbf{r}_j|^2$ between all particles of the system (for more details cf. Ref. 1). In this paper I shall study a system of rigid bodies and mass points. The subsystem of the mass points is assumed to obey the conditions on the results of Ref. 1 (i.e., the interaction between the mass points has to be sufficiently "strong"). These will be used to derive the general form of summational invariants of this system in the frame of Galilei invariant mechanics (for a general introduction into this field see Refs. 1 and 2 and the literature cited there).

Consider a system of n rigid bodies and m $(m > 1)$ mass points. If one denotes the coordinates and the momentum of the center of mass of the ith rigid body by $\mathbf{x}_i(t)$ and $\mathbf{p}_i(t)$, the angular momentum relative to its principal axis system by $\mathbf{s}_i$, and its Euler angles by $\Psi_i := (\psi_i,\theta_i,\phi_i)$ $(i = 1,\dots,n)$ one can describe a state of this system by a point

$$(\mathbf{x}_1(t),\mathbf{p}_1(t),\, \mathbf{s}_1(t),\Psi_1(t),\dots,\mathbf{x}_{n+m}(t),\mathbf{p}_{n+m}(t)),$$

in the phase space Γ, where the indices $j = n+1,\dots,n+m$ are related to the mass points.

For the systems considered here it is assumed that the interaction between the particles decreases sufficiently fast such that the following assumptions hold (for more details see Ref. 3).

A. Asymptotic condition

Let $\mathbf{x}_k^{\mathrm{ex}}(t)$ and $\mathbf{p}_k^{\mathrm{ex}}$ describe the free motion and $\mathbf{s}_i^{\mathrm{ex}}$ and $\Psi_i^{\mathrm{ex}}(t)$ the free rotation of a particle. Then there are orbits $(\mathbf{x}_1(t),\dots,p_{n+m}(t))$ in the phase space of the system such that

$$|\mathbf{x}_k(t) - \mathbf{x}_k^{\mathrm{ex}}(t)| + |\mathbf{p}_k(t) - \mathbf{p}_k^{\mathrm{ex}}| \to 0 \quad (t \to \pm\infty),$$

$$|\mathbf{s}_i(t) - \mathbf{s}_i^{\mathrm{ex}}| + |\Psi_i(t) - \Psi_i^{\mathrm{ex}}(t)| \to 0 (t \to \pm\infty),$$

$$\left(i = 1,\dots,n;\ k = 1,\dots,n+m;\ \mathrm{ex} = \begin{Bmatrix}\text{in } t\to -\infty\\ \text{out } t\to +\infty\end{Bmatrix}\right).$$

B. Asymptotic completeness

The set of all (asymptotically) free incoming states is equal to the set of all free outgoing states and to the set of all free states up to sets of measure 0.

C. Cluster condition

For any of the particles one can find initial conditions (i.e., for $t \to -\infty$) such that it never interacts with the rest of the system.

A function,

$$F(\mathbf{x}_1(t),\, \mathbf{p}_1(t),\, \mathbf{s}_1(t),\Psi_1(t),\dots,\mathbf{x}_{n+m}(t),\, \mathbf{p}_{n+m}(t);\, t),$$

on the phase space of the system is called *summational invariant* if there are functions

$$f_i(\mathbf{x}(t),\mathbf{p}(t),\, \mathbf{s}(t),\Psi(t);t) \text{ resp. } f_j(\mathbf{x}(t),\mathbf{p}(t);t)$$

corresponding to the particles of the system with

$$\begin{aligned}
F(\mathbf{x}_1^{\mathrm{in}}(t),\dots,\mathbf{p}_{n+m}^{\mathrm{in}};t)
&= \sum_{i=1}^{n} f_i(\mathbf{x}_i^{\mathrm{in}},\mathbf{p}_i^{\mathrm{in}},\mathbf{s}_i^{\mathrm{in}},\Psi_i^{\mathrm{in}};t) + \sum_{j=n+1}^{n+m} f_j(\mathbf{x}_j^{\mathrm{in}},\mathbf{p}_j^{\mathrm{in}};t)\\
&= \sum_{i=1}^{n} f_i(\mathbf{x}_i^{\mathrm{out}},\mathbf{p}_i^{\mathrm{out}},\mathbf{s}_i^{\mathrm{out}},\Psi_i^{\mathrm{out}};t) + \sum_{j=n+1}^{n+m} f_j(\mathbf{x}_j^{\mathrm{out}},\mathbf{p}_j^{\mathrm{out}};t)\\
&= F(\mathbf{x}_1^{\mathrm{out}}(t),\dots,\mathbf{p}_{n+m}^{\mathrm{out}};t),
\end{aligned}$$

and $d/dt f_k = 0$ $(k = 1,\dots,n+m)$ for all asymptotic orbits. The f_k may be different for different particles. The function F is not necessarily conserved along the actual path of the interacting system.

From Sec. I C it follows that the addition of further particles to a certain system does not increase the number of independent invariants. Therefore, it is sufficient to consider all possible two-particle subsystems and derive the structure of their summational invariants. The invariants of the subsystem of the mass points are known according to Ref. 1 and are of the form (1). Now the notation is slightly altered: the indices "ex" are dropped. The mass points coordinates before resp. after the interaction are labeled by the index 1 resp. 3, the rigid bodies by 2 resp. 4. The equation under consideration obtains the form

$$f_1(\mathbf{x}_1,\mathbf{p}_1,t) + f_2(\mathbf{x}_2,\mathbf{p}_2,\mathbf{s}_2,\Psi_2,t) = f_1(\mathbf{x}_3,\mathbf{p}_3,t) + f_2(\mathbf{x}_4,\mathbf{p}_4,\mathbf{s}_4,\Psi_4,t)$$

---

a) Permanent address: Institut für Theoretische Physik III, Universität Düsseldorf, Universitätsstr. 1, Geb. 25.32, D-4000 Düsseldorf 1, West Germany.

 0022-2488/90/071693-05$03.00

for all incoming and outgoing states related by a realizable scattering process, with f_1 of the form (1).

To answer the question whether there are, besides the ten classical conservation laws, additional summational invariants that show up only on the manifold of physically realizable scattering events U, one has to examine carefully its structure that is determined by the interaction. If there is no interaction at all, any function that is constant for free orbits is a summational invariant of the system. Therefore to get reasonable results one has to find physically realistic assumptions on the interaction. The first idea one might have is to postulate that U contains a full open subset of M, the scattering manifold that is constrainted by the ten classical conservation laws. But for general summational invariants depending on positions and momenta (and Euler angles and angular momenta) this is impossible (a detailed discussion of this problem is given in Ref. 2). So one has to extract the essential features from a not-to-ill behaved interaction.

One should expect that there is scattering. The scattering should be continuous in the sense that small changes of the ingoing states should lead to small changes of the outgoing states (at least on an open subset of U). There should exist scattering events that, keeping the outgoing momentum (angular momentum) fixed, result in a set of outgoing angular momenta (momenta), i.e., the outgoing linear and angular momenta may be varied separately by changes of the ingoing states. And there should be scattering events that change the energy of rotation. These assumptions form the essence of the propositions of the central theorems of this paper.

To deal with the difficulties that arise from the complicated motion of a rigid body (even in the free case), two simplifications will be made: As indicated by the title of this paper, systems consisting purely of rigid bodies will not be discussed, but the case that a subsystem of "auxiliary" mass points with well-known summational invariants is added. This allows the reduction to a one-body problem. By virtue of the cluster condition the interaction between the rigid bodies no longer affects the form of the summational invariants related to them. In this sense the mass points carry the "effective" interaction of the rigid bodies (if there is more than one). Second, the dependence of the f_i on the Euler angles will be dropped. Thus one circumvents the difficulties that are induced by the fact that there is no further *a priori* conservation law related to the angle variables and the angular momenta analogous to the conservation of the center of mass. With respect to the rigid body one arrives at a situation similar to the case of pure momentum dependence in the mechanics of mass points.

To sum it up: We are interested in the solution of the following equation:

$$f_1(\mathbf{x}_1,\mathbf{p}_1,t)+f_2(\mathbf{x}_2,\mathbf{p}_2,\mathbf{s}_2,t)$$
$$=f_1(\mathbf{x}_3,\mathbf{p}_3,t)+f_2(\mathbf{x}_4,\mathbf{p}_4,\mathbf{s}_4,t), \qquad (2)$$

for all states related by a realizable scattering event.

As a first step we consider a system that is constrained to a plane, the angular momenta of all particles being parallel to the planes normal vector. Due to the geometry of this system the conditions on the interaction have a considerably simple physical content, namely the existence of at least one nontrivial scattering event (1.1) and continuity of the scattering in the neighborhood of "forward scattering" in the space of the angular momenta $\mathbf{s}$ (1.2).

Theorem 1: Consider a system of n rigid bodies and m $(m>1)$ mass points obeying the conditions stated above. The scattering is assumed to be constrained to a plane. Let the interaction between a mass point and a rigid body in any two-particle subsystem fulfill the following conditions:

(1.1) For almost all incoming angular momenta $\mathbf{s}_2$ exists a realizable scattering process with $\mathbf{p}_4-\mathbf{p}_2\neq 0$.

(1.2) The scattering is continuous in the sense that for almost all incoming angular momenta $\mathbf{s}_2$ one can cover almost an open neighborhood of $\mathbf{s}_4=\mathbf{s}_2$ with scattering events.

(1.3a) $f_2\in C^1$, if $A_2=A_3=0$[A_2, A_3 from (1) for f_1].

(1.3b) $f_2\in C^2$, if A_2 or $A_3\neq 0$.

Then f_2 has the following structure:

$$f_2(\mathbf{x},\mathbf{p},\mathbf{s},t)=A_1[\mathbf{p}^2/2m_2+\mathbf{s}^2/2\theta_2]+\mathbf{B}_1\cdot\mathbf{p}$$
$$+\mathbf{B}_2(\mathbf{x}\times\mathbf{p}+\mathbf{s})+\mathbf{B}_3\cdot\mathbf{r}_4+k_2,$$

where $\mathbf{r}_2:=m_2\mathbf{x}-\mathbf{p}t$. Only the constant k_2, beside the mass m_2 and the moment of inertia θ_2, may depend on the particle considered.

For a system that is not subjected to any *a priori* constraints, the conditions on the interaction have to be stronger. Conditions (2.1) and (2.2) secure that the outgoing linear and angular momenta may be varied sufficiently independent of each other. Condition (2.3) postulates, in analogy to (1.2), the continuity of the scattering in the neighborhood of "forward scattering" with respect to the angular momenta $\mathbf{s}$.

Theorem 2: Consider a system of n rigid bodies and m $(m>1)$ mass points obeying the conditions stated above. Let the interaction between a rigid body and a mass point in any two-particle subsystem obey the following conditions:

(2.1) For almost all incoming angular momenta $\mathbf{s}_2$ one can reach at least three linear independent vectors $\mathbf{p}^*:=\mathbf{p}_4-\mathbf{p}_2$, depending on $\mathbf{s}_2$, by realizable scattering processes.

(2.2) For almost all incoming angular momenta $\mathbf{s}_2$ one can find a $\mathbf{p}^*\neq 0$ with $\mathbf{p}^*\times\mathbf{s}_2\neq 0$ such that one can reach three linear independent outgoing angular momenta $\mathbf{s}_4$ by realizable scattering events, keeping $\mathbf{s}_2$ and $\mathbf{p}^*$ fixed.

(2.3) The scattering is continuous in the sense that for almost all values of the amount of the incoming angular momenta one can cover almost an open neighborhood of $|\mathbf{s}_4|=|\mathbf{s}_2|$ by realizable scattering events.

(2.4a) Let $f_2\in C^1$, if $A_2=A_3=0$.

(2.4b) Let $f_2\in C^2$, if A_2 or $A_3\neq 0$. Assume that there exists at least one scattering event which changes the energy of rotation and the momentum of the rigid body.

Then f_2 has the following structure:

$$\mathbf{f}_2(\mathbf{x},\mathbf{p},\mathbf{s},t)=A_1\left[\frac{\mathbf{p}^2}{2m_2}+\sum_{j=1}^{3}\frac{(s^j)^2}{2\theta_2^j}\right]+\mathbf{B}_1\cdot\mathbf{p}$$
$$+\mathbf{B}_2(\mathbf{x}\times\mathbf{p}+\mathbf{s})+\mathbf{B}_3\cdot\mathbf{r}_2+k_2,$$

with $\mathbf{r}_2:=m_2\mathbf{x}-\mathbf{p}t$. Only the constant k_2, beside the mass m_2 and the moment of inertia θ_2^j, may depend on the particle. As a result of this paper, it turns out that there are no further summational invariants of a system of rigid bodies and mass points besides the *a priori* known ones. The two additional invariants for systems with $1/r^2$ interaction potentials do not occur in the case of interaction with rigid bodies.

II. PREPARATIONS FOR THE PROOF OF THE THEOREMS

The explicit time dependence of the f_k may be removed by the transformations: $(\mathbf{x},\mathbf{p},\mathbf{s},t) \to (\mathbf{r},\mathbf{p},\mathbf{s})$ and $(\mathbf{x},\mathbf{p},t) \to (\mathbf{r},\mathbf{p})$. There are continuously differentiable functions $\hat{f}(\mathbf{r},\mathbf{p},\mathbf{s})$ and $\hat{f}(\mathbf{r},\mathbf{p})$ with

$$\hat{f}(\mathbf{r},\mathbf{p},\mathbf{s}) = f(\mathbf{x},\mathbf{p},\mathbf{s},t) \quad \text{and} \quad \hat{f}_j(\mathbf{r},\mathbf{p}) = f_j(\mathbf{x},\mathbf{p},t),$$

where $\mathbf{r} = m\mathbf{x} - \mathbf{p}t$.

Now the problem can be reformulated as follows: We are interested in the solution of the functional equation (identifying $\hat{f}_k$ with f_k for convenience):

$$f_1(\mathbf{r}_1,\mathbf{p}_1) + f_2(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = f_1(\mathbf{r}_3,\mathbf{p}_3) + f_2(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{3}$$

whenever $(\mathbf{r}_1,\mathbf{p}_1,...,\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4) \in U \subset M$. The ten *a priori* conservation laws can be used to replace the expressions in the coordinates of the mass points by those in coordinates of the rigid body. There are two possibilities.

(i) The expressions $\sim \mathbf{r}\cdot\mathbf{p}$ resp. $\sim \mathbf{r}^2$ are not invariants of the subsystem of the mass points (i.e., $A_2 = A_3 = 0$). With

$$F(\mathbf{r},\mathbf{p},\mathbf{s}) := f_2(\mathbf{r},\mathbf{p},\mathbf{s}) - A_1\left[\frac{\mathbf{p}^2}{2m_2} + \sum_{j=1}^{3} \frac{(s^j)^2}{2\theta_2^j}\right] - \mathbf{B}_1\mathbf{p} - \mathbf{B}_2\cdot[(1/m_2)\mathbf{r}\times\mathbf{p} + \mathbf{s}] - \mathbf{B}_2\cdot\mathbf{r},$$

one obtains from (3) by insertion:

$$F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{4}$$

whenever $(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2,\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4) \in V$, where V denotes the Galilei invariant subset of U of realizable in and out states of the rigid body.

(ii) The expressions $\sim \mathbf{r}\cdot\mathbf{p}$ resp. $\sim \mathbf{r}^2$ are invariants of the subsystem of the mass points (i.e., A_2 or $A_3 \neq 0$). If there is no function $g(\mathbf{r},\mathbf{p},\mathbf{s})$ of the coordinates of the rigid body such that the following equation holds:

$$(1/m_1)(ar_1^2 + b\mathbf{r}_1\cdot\mathbf{p}_1) + g(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = (1/m_1)(ar_3^2 + b\mathbf{r}_3\cdot\mathbf{p}_3) + g(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{5}$$

whenever $(\mathbf{r}_1,...,\mathbf{s}_4) \in U$, for any choice of $a \neq 0$ or $b \neq 0$, then the terms $\sim \mathbf{r}^2$ and $\sim \mathbf{r}\cdot\mathbf{p}$ cannot be summational invariants of the whole system. Thus A_2 and A_3 have to vanish and one is led back to case (*i*).

Now the problem to find a solution of (3) is reduced to solve (4), which depends only on the coordinates of the rigid body, and to show that there is no $g(\mathbf{r},\mathbf{p},\mathbf{s})$ that solves (5) for any nontrivial choice of a and b (i.e., a or $b \neq 0$).

A. Some conclusions from Galilei invariance

Since the angular momentum $\mathbf{s}_2$ of the rigid body is measured relative to its fixed center-of-mass system it is invariant under translations and boosts. As the states considered are free asymptotic states, the momenta and angular momenta do not depend on time. It is suggested that we examine the behavior of a solution of (4) under infinitesimal Galilei transformations. Applying (i) an infinitesimal translation, (ii) an infinitesimal boost, (iii) an infinitesimal rotation, and (iv) an infinitesimal time shift one obtains,

$$\text{(i)} \quad \nabla_r F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = \nabla_r F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{6}$$

$$\text{(ii)} \quad \nabla_p F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = \nabla_p F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{7}$$

$$\text{(iii)} [(\mathbf{p}_2 - \mathbf{p}_4)\times\nabla_p + (\mathbf{r}_2 - \mathbf{r}_4)\times\nabla_r + \mathbf{s}_2\times\nabla_s] \times F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = \mathbf{s}_4\times\nabla_s F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{8}$$

$$\text{(iv)} \quad \mathbf{p}_2\cdot\nabla_r F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = \mathbf{p}_4\cdot\nabla_r F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{9}$$

if $(\mathbf{r}_2,...,\mathbf{s}_4) \in V$.

The same steps [except (iii)] applied to (3) lead to

$$m_1\nabla_p f_1(\mathbf{r}_1,\mathbf{p}_1) + m_2\nabla_p f_2(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = m_1\nabla_p f_1(\mathbf{r}_3,\mathbf{p}_3) + m_2\nabla_p f_2(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{10}$$

$$m_1\nabla_r f_1(\mathbf{r}_1,\mathbf{p}_1) + m_2\nabla_r f_2(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2), \quad m_1\nabla_r f_1(\mathbf{r}_3,\mathbf{p}_3) + m_2\nabla_r f_2(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{11}$$

$$\mathbf{p}_1\cdot\nabla_r f_1(\mathbf{r}_1,\mathbf{p}_1) + \mathbf{p}_2\cdot\nabla_r f_2(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2), \quad \mathbf{p}_3\cdot\nabla_r f_1(\mathbf{r}_3,\mathbf{p}_3) + \mathbf{p}_4\cdot\nabla_r f_2(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4), \tag{12}$$

if $(\mathbf{r}_1,...,\mathbf{s}_4) \in U$.

III. PROOF OF THE THEOREMS

The proof will be carried out in two steps. First it will be shown that a solution of (4) is, in case of Theorem 1, a continuously differentiable function $k(\lambda,\mu,\mathbf{n})$ (see below) that is constant in any fixed inertial frame and is a constant k in the case of Theorem 2. The second step will lead to the conclusion that there is no solution $g(\mathbf{r},\mathbf{p},\mathbf{s})$ of (5) for any nontrivial choice of (a,b).

In the case of Theorem 1, consider any arbitrary inertial system and let $\mathbf{n}$ be the normal vector of the scattering plane. Due to the geometry of the problem, the expressions $\mathbf{r}_2\cdot\mathbf{n} = \mathbf{r}_4\cdot\mathbf{n} =: \lambda$ and $\mathbf{p}_2\cdot\mathbf{n} = \mathbf{p}_4\cdot\mathbf{n} =: \mu$ are constants of motion. Therefore any function $k = k(\lambda,\mu,\mathbf{n})$ is a solution of (4) and is constant for any given inertial system. Let $P(\mathbf{s}_2)$ be the set of all $\mathbf{p}^* := \mathbf{p}_4 - \mathbf{p}_2$ that can be reached by realizable scattering events, keeping $\mathbf{s}_2$ fixed. Also, P is invariant under boosts and translations. Given any solution F of (4), inserting (6) into (9) yields

$$(\mathbf{p}_2 - \mathbf{p}_4)\cdot\nabla_r F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = -\mathbf{p}^*\cdot\nabla_r F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) = 0, \tag{13}$$

for all states related by a realizable scattering process. Equation (13) is Galilei invariant, thus one has in the lab system of the rigid body $\mathbf{p}^*\nabla_r F(0,0,\mathbf{s}_2) = 0$ for all $\mathbf{p}^* \in P(\mathbf{s}_2)$. Due to (1.1) there is for almost all $\mathbf{s}_2$ at least one $\mathbf{p}^* \neq 0$. Therefore, $P(\mathbf{s}_2)$ contains a circle in the scattering plane consisting of all $\mathbf{p}^{*\prime}$ obtained by rotation of the experimental setup or, equivalently, of the lab system of the rigid body about $\mathbf{n}$. Thus one may deduce $\nabla_r F(0,0,\mathbf{s}_2) \sim \mathbf{n}$ for almost all $\mathbf{s}_2$. Boost and translation yield, due to the invariance of $P(\mathbf{s}_2)$:

$$\nabla_r F(\mathbf{r},\mathbf{p},\mathbf{s}_2) \sim \mathbf{n},$$

for all $\mathbf{r},\mathbf{p}$ and almost all $\mathbf{s}_2$. The continuity of $\nabla_r F$ extends the statement to all $\mathbf{s}$. Thus we have, instead of (4),

$$F(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) := G(\lambda,\mathbf{p}_2,\mathbf{s}_2) = G(\lambda,\mathbf{p}_4,\mathbf{s}_4) = F(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4),$$

if $(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2,\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4)\in V$, $\lambda = \mathbf{n}\cdot\mathbf{r}_2 = \mathbf{n}\cdot\mathbf{r}_4$.

Performing an infinitesimal rotation about $\mathbf{n}$, (8) yields, after multiplication with $\mathbf{n}$,

$$\mathbf{n}\cdot[(\mathbf{p}_2-\mathbf{p}_4)\times\nabla_p G(\lambda,\mathbf{p}_2,\mathbf{s}_2)] = \mathbf{n}\cdot[\mathbf{p}^*\times\nabla_p G(\lambda,\mathbf{p}_2,\mathbf{s}_2)] = 0. \tag{14}$$

According to (1.1), $P(\mathbf{s}_2)$ contains at least a circle in the scattering plane. Due to (14), the vector in the brackets has to lie in the scattering plane. This is possible only if $\nabla_p G(\lambda,\mathbf{p},\mathbf{s}_2)\sim\mathbf{n}$ for all $\mathbf{p}$ and all $\mathbf{s}_2$, due to continuity. Defining $G(\lambda,p,s) =: H(\lambda,\mu,\mathbf{s})$ one arrives at

$$H(\lambda,\mu,\mathbf{s}_2) = H(\lambda,\mu,\mathbf{s}_4),$$

if $(\mathbf{r}_2,...,\mathbf{s}_4)\in V$. Taking into account that $\mathbf{s}_2 = s_2\cdot\mathbf{n}$ one has according to (1.2),

$$H(\lambda,\mu,s_2\cdot\mathbf{n}) = H(\lambda,\mu,s_4\cdot\mathbf{n}),$$

for all $s_4\in U(s_2)$, due to continuity. Thus $\partial_s H(\lambda,\mu,s\cdot\mathbf{n}) = 0$ almost everywhere. Continuity of the first derivatives of H yields,

$$H(\lambda,\mu,s\cdot\mathbf{n}) = H(\lambda,\mu,\mathbf{n}). \tag{15}$$

For any given inertial system $H(\lambda,\mu,\mathbf{n})$ is a constant.

Consider now the case of Theorem 2. Let $P(\mathbf{s}_2)$ be the set of all $\mathbf{p}^* = \mathbf{p}_4 - \mathbf{p}_2$ that can be obtained by realizable scattering events, keeping $\mathbf{s}_2$ fixed. Consider any solution of (4). As in the case of Theorem 1 one gets in the lab system of the rigid body, taking (6) and (9) into account,

$$\mathbf{p}^*\cdot\nabla_r F(0,0,\mathbf{s}_2) = 0.$$

According to (2.1), P contains at least three linear independent $\mathbf{p}^*$. Then P is invariant under boosts and translations. Thus

$$\nabla_r F(\mathbf{r},\mathbf{p},\mathbf{s}_2) = 0, \tag{16}$$

for all $\mathbf{s}_2$, due to continuity of $\nabla_r F$. With $F(\mathbf{r},\mathbf{p},\mathbf{s}) =: G(\mathbf{p},s)$, we have,

$$G(\mathbf{p}_2,\mathbf{s}_2) = G(\mathbf{p}_4,\mathbf{s}_4),$$

for all states that are related by realizable scattering events. Using (7) and (8) one obtains,

$$(\mathbf{p}_2-\mathbf{p}_4)\times\nabla_p G(\mathbf{p}_2,\mathbf{s}_2) + \mathbf{s}_2\times\nabla_s G(\mathbf{p}_2,\mathbf{s}_2) = \mathbf{s}_4\times\nabla_s G(\mathbf{p}_4,\mathbf{s}_4). \tag{17}$$

Consider a pair $(\mathbf{s}_2,\mathbf{p}^*)$ for which the assumption (2.2) is fulfilled. There is, specifically, an $\mathbf{s}_4\neq 0$. Multiplication with $\mathbf{n} := \mathbf{s}_4/|\mathbf{s}_4|$ yields,

$$\mathbf{n}\cdot(-\mathbf{p}^*\times\nabla_p + \mathbf{s}_2\times\nabla_s)G(\mathbf{p}_2,\mathbf{s}_2) = 0. \tag{18}$$

By virtue of (2.2) there are at least three linear independent $\mathbf{n}$ that fulfill (18). Therefore one gets, by multiplying with $\mathbf{s}_2$,

$$\nabla_p G(\mathbf{p}_2,\mathbf{s}_2)\cdot(\mathbf{p}^*\times\mathbf{s}_2) = 0. \tag{19}$$

Because of the invariance of $\mathbf{n}$, $\mathbf{p}^*$, and $\mathbf{s}_2$ under boosts, Eq. (19) holds for all $\mathbf{p}_2$. Consider the lab system of the rigid body. One has,

$$\nabla_p G(0,\mathbf{s}_2)\cdot(\mathbf{p}^*\times\mathbf{s}_2) = 0. \tag{20}$$

This equation is valid for all $\mathbf{p}^{*\prime}$ that one gets by rotation of the relative frame about the direction of $\mathbf{s}_2$. These $\mathbf{p}^{*\prime}$ lie on a surface or a cone about the direction of $\mathbf{s}_2$.

Due to (20), $\nabla_p G(\mathbf{0},s_2)$ has to be perpendicular to the plane that is spanned by the $\mathbf{s}_2\times\mathbf{p}^{*\prime}$. That is possible only if $\nabla_p G(0,\mathbf{s}_2)\sim\mathbf{s}_2$. For all $\mathbf{p}_2$ the statement follows because of the invariance of $\mathbf{p}^*\times\mathbf{s}_2$ under boosts. Thus one ends up with,

$$\nabla_p G(\mathbf{p},\mathbf{s}_2) =: g(\mathbf{p},\mathbf{s}_2)\cdot\mathbf{s}_2,$$

for all $\mathbf{p}$ and all $\mathbf{s}_2$, due to continuity. Insertion into (4) yields,

$$g(\mathbf{p}_2,\mathbf{s}_2)\cdot\mathbf{s}_2 = g(\mathbf{p}_4,\mathbf{s}_4)\cdot\mathbf{s}_4,$$

for all $(\mathbf{p}_2,\mathbf{s}_2,\mathbf{p}_4,\mathbf{s}_4)$ that are related by realizable scattering events. According to (2.2) there are $\mathbf{s}_4$ with $\mathbf{s}_2\times\mathbf{s}_4\neq 0$. Thus one has

$$g(\mathbf{p},\mathbf{s}_2) = 0,$$

for almost all $\mathbf{s}_2$. By continuity this extends to all $\mathbf{s}$. With $G(\mathbf{p},\mathbf{s}) := H(s)$ we arrive at

$$H(\mathbf{s}_2) = H(\mathbf{s}_4),$$

for all $(\mathbf{s}_2,\mathbf{s}_4)$ that are related by realizable scattering events. Insertion into (17), multiplication with $\mathbf{s}_2$, and (2.2) yield,

$$\mathbf{s}_2\times\nabla_s H(\mathbf{s}_2) = 0$$

for almost all $\mathbf{s}_2$. Now one has $\nabla_s H(\mathbf{s}) =: h(\mathbf{s})\cdot\mathbf{s}$ for all $\mathbf{s}$ due to continuity and therefore

$$\hat{H}(s_2^2) = \hat{H}(s^2),$$

for all s out of a neighborhood of $|s_2|$ due to continuity. Thus $\hat{H}(s^2)$ is piecewise constant. The assertion $\hat{H} = k = \text{const}$ follows by continuity from $\partial_s\hat{H}(s^2) = 0$ almost everywhere. Now we end up with,

$$F(\mathbf{r},\mathbf{p},\mathbf{s}) = k = \text{const}.$$

To complete the proof of the theorems one has to consider the following two alternatives.

(i) If the expressions $\sim\mathbf{r}\cdot\mathbf{p}$ and $\sim\mathbf{r}^2$ are not invariants of the subsystem of the mass points (i.e., $A_2 = A_3 = 0$) the assertion of the theorems follows by insertion into the definition of $F(\mathbf{r},\mathbf{p},\mathbf{s})$.

(ii) In the case that the expressions $\sim\mathbf{r}\cdot\mathbf{p}$ or $\sim\mathbf{r}^2$ are invariants of the subsystem of the mass points (i.e., A_2 or $A_3\neq 0$) one has to show for any nontrivial choice of a and b that there is no function $g\in C^2$ that solves (5). Then A_2 and A_3 have to vanish and the assertion follows along the line of case (i). Proof by contradiction: consider any solution g of (5). Taking (11) into account one obtains

$$m_2\cdot\nabla_r g(\mathbf{r}_2,\mathbf{p}_2,\mathbf{s}_2) - 2a\cdot\mathbf{r}_2 - b\cdot\mathbf{p}_2 = m_2\cdot\nabla_r g(\mathbf{r}_4,\mathbf{p}_4,\mathbf{s}_4) - 2a\cdot\mathbf{r}_4 - b\cdot\mathbf{p}_4,$$

if $(r_2,...,s_4)\in V$, keeping in mind that $\mathbf{r}_1 - \mathbf{r}_3 = \mathbf{r}_4 - \mathbf{r}_2$ and $\mathbf{p}_1 - \mathbf{p}_3 = \mathbf{p}_4 - \mathbf{p}_2$. This is an equation of type (4) in each component. Using the results derived up to now one gets,

$$\text{lhs} =: K = \text{rhs}, \tag{21}$$

with $K = \mathbf{K}(\lambda,\mu,n)$ in case of Theorem 1 and $K = \text{const}$ in case of Theorem 2. Applying an infinitesimal time translation to (5), inserting into (21), and collecting the expressions in K yields,

$$(1/m_1)[2a\cdot(\mathbf{r}_1\cdot\mathbf{p}_1-\mathbf{r}_3\cdot\mathbf{p}_3)+b\cdot(\mathbf{p}_1^2-\mathbf{p}_3^2)]$$
$$+(1/m_2)[2a\cdot(\mathbf{r}_2\cdot\mathbf{p}_2-\mathbf{r}_4\cdot\mathbf{p}_4)+b\cdot(\mathbf{p}_2^2-\mathbf{p}_4^2)]$$
$$=(1/m_2)(\mathbf{p}_4-\mathbf{p}_2)\cdot\mathbf{K}=(1/m_2)\cdot\mathbf{p}^*\cdot\mathbf{K}, \qquad (22)$$

if $(\mathbf{r}_1,...,\mathbf{s}_4)\in U$.

Consider the case of Theorem 1. Assume $a\neq 0$. Performing an infinitesimal translation (22) yields,

$$2a\cdot(p_i^j+p_2^j-p_3^j-p_4^j)=0=p^{*i}\cdot(\nabla_r)^j K^i(\lambda,\mu,\mathbf{n}) \qquad (23)$$

($j=1,2,3$; summation convention). Now one has

$$0=n^j\cdot p^{*i}\cdot\partial_\lambda K^i(\lambda,\mu,\mathbf{n})$$
$$=:n^j\cdot p^{*i}\cdot L^i(\lambda,\mu,\mathbf{n}) \quad (j=1,2,3).$$

and therefore $\mathbf{p}^*\cdot\mathbf{L}(\lambda,\mu,\mathbf{n})=0$ for all $(\mathbf{r}_2,...,\mathbf{s}_4)\in V$.

By performing an infinitesimal time transition one obtains from (22),

$$2a\cdot\left[\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}\right]=\mu\cdot\mathbf{p}^*\cdot\mathbf{L}(\lambda,\mu,\mathbf{n})=0,$$

for all realizable scattering events. This is a contradiction to (1.2), which states that there are scattering events that change the translation energy. Assume now $a=0$, $b\neq 0$. Equation (22) yields

$$b\cdot\left[\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}\right]=\frac{1}{m_2}\cdot\mathbf{p}^*\cdot\mathbf{K}(\lambda,\mu,\mathbf{n}), \qquad (24)$$

for all $(r_1,...,s_4)\in U$. Consider a scattering event with $\mathbf{p}^*\neq 0$ in the lab system of the rigid body. Then all other scattering events are realizable that can be obtained by rotating the experimental setup, or the lab system of the rigid body about the direction of $\mathbf{n}$. The lhs of (24) is invariant under these rotations, whereas $\mathbf{p}^*$ describes a circle in the scattering plane. Thus $\mathbf{K}$ has to have constant product with all vectors of constant length in a plane. This is possible only if $\mathbf{K}$ is perpendicular to the plane. The result is

$$\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}=0,$$

for all realizable scattering events, in contradiction to (1.2).

Consider the case of Theorem 2. Let $a\neq 0$. An infinitesimal time shift yields for (22),

$$2a\cdot\left[\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}\right]=0,$$

for all $(\mathbf{p}_1,...,\mathbf{p}_4)$ that are related by realizable scattering events. This is a contradiction to (2.4b). Now let $a=0$, $b\neq 0$. One obtains from (22):

$$b\cdot\left[\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}\right]=\frac{1}{m_2}(\mathbf{p}_4-\mathbf{p}_2)\cdot\mathbf{K},$$

for all realizable scattering events. Consider a scattering with $\mathbf{p}_4-\mathbf{p}_2\neq 0$. Due to the Galilei invariance of the set of all physically realizable scattering events, all scattering events are possible that one obtains by rotating the experimental setup. The corresponding vectors $\mathbf{p}_4'-\mathbf{p}_2'$ lie on the surface of a sphere. The lhs is invariant under these rotations. The constant vector $\mathbf{K}$ has to have constant product with all vectors of constant length on a sphere. Therefore one has $\mathbf{K}=0$. Thus one obtains,

$$b\cdot\left[\frac{\mathbf{p}_1^2}{m_1}+\frac{\mathbf{p}_2^2}{m_2}-\frac{\mathbf{p}_3^2}{m_1}-\frac{\mathbf{p}_4^2}{m_2}\right]=0,$$

in contradiction to (2.4b). It follows that $a=b=0$ and the proof is completed as in case (i).

ACKNOWLEDGMENTS

I thank H. Reeh and W. Grieger for many fruitful discussions and B. Schmittmann for a critical reading of this manuscript.

APPENDIX

One of the results of Amigó and Reeh is the following theorem.

Consider a system of n mass points of nonvanishing mass with central force interaction such that scattering in a neighborhood of forward scattering between each two particles occurs and depends there continuously on the impact parameter ρ. If the interaction has infinite range, it is assumed to decrease such that for the scattering angle θ in the lab systems: $|\theta|\leqslant\text{const}/\rho^2$, $|\partial\theta/\partial\rho|\leqslant\text{const}/\rho^3$ for large ρ. Assume the f_i to be locally L^1. Then

$$f_i(\mathbf{x},\mathbf{p},t)=A_1(\mathbf{p}^2/2m_i)+\mathbf{B}_1\cdot\mathbf{p}+\mathbf{B}_2\cdot(\mathbf{x}\times\mathbf{p})+\mathbf{B}_3\cdot\mathbf{r}_i$$
$$+A_2\cdot\frac{\mathbf{p}\cdot\mathbf{r}_i}{m_i}+A_3\cdot\frac{\mathbf{r}_i^2}{m_i}+k_i,$$

with $\mathbf{r}_i:=m_i\mathbf{x}-\mathbf{p}t$ in the Galilei invariant case and constant A_j,B_j,k_i.

[1]J. M. Amigó and H. Reeh, J. Math. Phys. **24**, 1594 (1983).
[2]M. Requardt, J. Math. Phys., **28**, 1827 (1987).
[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (New York, 1979).

Equations of state and plane-autonomous systems in Bianchi V imperfect fluid cosmology

Alan A. Coley
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada

(Received 12 July 1989; accepted for publication 7 March 1990)

A new general approach for investigating imperfect fluid cosmological models is introduced in which the equations of state are completely "dimensionless." Such equations of state are then utilized to reduce the Einstein field equations governing Bianchi V imperfect fluid cosmologies to a plane-autonomous system of equations, thus enabling the qualitative behavior of these cosmological models to be analyzed in a straightforward manner. The resulting plane-autonomous system is investigated. Finally, exact solutions of the Bianchi V imperfect fluid field equations in the case when the equations of state take on a particularly simple form are discussed.

I. INTRODUCTION

In a recent paper[1] (hereafter referred to as paper I) Bianchi V imperfect fluid cosmology was investigated. [For brevity, we will adopt the notation that an equation or reference in paper I will be referred to using a label I]. It is of interest to study cosmological models with a richer structure, both geometrically and physically, than the standard perfect fluid Friedmann–Robertson–Walker (FRW) models. Bianchi V models are of particular interest since they are sufficiently complex (e.g., the Einstein tensor has off-diagonal terms) while, at the same time, they are a simple generalization of the negative-curvature FRW models. Cosmological models that include viscosity have been investigated in an attempt to explain the currently observed highly isotropic matter distribution (I1–I3) and the high entropy per baryon in the present state of the Universe (I4, I5), and in order to further study the nature of the initial singularity (I6) and the formation of galaxies (I3). Models that include heat conduction have also been studied in spatially homogeneous cosmologies (in particular, see I7). The motivation and background for this research is discussed in more detail in Ref. I.

In MacCallum[2] a general class of Bianchi models were studied [all class A models, and the set of class B with $n_\alpha^\alpha = 0$ ($\alpha = 1,2,3$)]. In this class (that contains the Bianchi V models) the general exact (two-parameter) orthogonal perfect fluid solution is known up to quadratures[3]. Collins[4] has investigated a certain subclass of this class of models whose equations reduce to an autonomous system and are therefore susceptible to a qualitative analysis utilizing geometric techniques. More precisely, Collins studied a subclass of perfect fluid, nonrotating, spatially homogeneous cosmological models with equation of state $p = (\gamma - 1)\rho$ and zero cosmological constant. In particular, this subclass includes the (not necessarily LRS) Bianchi V models (see Fig. 3, in Ref. 4). Later, this subclass was extended to include perfect fluid LRS Bianchi models (again including type V models) with tilt.[5]

Here, we shall use the techniques and notation of Refs. 2–5 to reduce the differential equations governing the Bianchi V imperfect fluid cosmological models under consideration to a plane-autonomous system of equations.

More precisely, in this paper we will investigate a class of phenomenological equations of state (for the pressure and coefficients of bulk and shear viscosity) in imperfect fluid cosmological models. This general class of equations of state is characterized by the fact that completely dimensionless quantities are inter-related (i.e., the equations of state are "dimensionless"). It is noted that this class includes as special cases all the most commonly considered equations of state. This procedure amounts to introducing a new approach for dealing with equations of state in cosmology, an approach that is quite general, but for illustrative purposes we restrict our attention to Bianchi V cosmologies. The feature of this class of greatest interest here is that equations of state of this type are the most general under which the resulting Einstein field equations reduce to a plane-autonomous system.

The analysis will consequently enable us to write the Bianchi V imperfect fluid field equations as a plane-autonomous system. This in turn will enable us to analyze the qualitative behavior of these cosmological models in a straightforward manner. The plane-autonomous system is studied further in the case that the equations of state are of a special (power law) form; the resulting system in a particularly simple subcase is displayed in the final section for illustration.

In Sec. IV we shall look for exact solutions of the Bianchi V imperfect fluid field equations in the case when the equations of state take on the simple form $p = (\gamma - 1)\rho$, $\zeta = \zeta_0\theta$, and $\eta = \eta_0\theta$ [see Eqs. (4.1)]. Exact solutions will of course be very useful in concert with any qualitative analysis. A simple, *general* first integral of the field equations is found. Using this first integral it is then shown that the field equations reduce to a single, second-order, ordinary differential equation for a single variable. In the particular case of $\gamma = 2$ (stiff matter), a simple (albeit unphysical) solution is exhibited.

II. THE MODELS

We shall study LRS Bianchi type V spatially homogeneous cosmology, where the metric is given by

$$ds^2 = -dt^2 + a^2(t)dx^2 + b^2(t)e^{2x}(dy^2 + dz^2), \qquad (2.1)$$

 0022-2488/90/071698-06$03.00

in which the source of the gravitational field is a viscous fluid with heat conduction, so that the energy–momentum tensor is given by

$$T_{ab} = (\rho + \bar{p})u_a u_b + \bar{p} g_{ab} - 2\eta\sigma_{ab} + q_a u_b + u_a q_b, \tag{2.2}$$

with

$$\bar{p} = p - \zeta\theta, \tag{2.3}$$

where p is the thermodynamic pressure and ζ and η are the coefficients of bulk and shear viscosity, respectively, thereby allowing dissipative processes to be included in the models.

The Einstein field equations for a comoving fluid then yield an equation defining the energy density (I.8a),

$$\rho = \frac{\dot{b}^2}{b^2} + 2\frac{\dot{a}\dot{b}}{ab} - \frac{3}{a^2}, \tag{2.4}$$

an equation that defines the only nonzero component of the heat conduction vector q_a (I.8b),

$$q_1 = 2[\dot{b}/b - \dot{a}/a], \tag{2.5}$$

and the remaining nontrivial equations (I.8c) and (I.8d),

$$\frac{1}{a^2} - \frac{\dot{b}^2}{b^2} - 2\frac{\ddot{b}}{b} = \bar{p} - \frac{4}{3}\eta\left[\frac{\dot{a}}{a} - \frac{\dot{b}}{b}\right], \tag{2.6}$$

$$\frac{1}{a^2} - \frac{\ddot{a}}{a} - \frac{\ddot{b}}{b} - \frac{\dot{a}\dot{b}}{ab} = \bar{p} - \frac{2}{3}\eta\left[\frac{\dot{b}}{b} - \frac{\dot{a}}{a}\right]. \tag{2.7}$$

We recall, that for the Bianchi V models under consideration,

$$\sigma^2 = \tfrac{1}{3}[\dot{a}/a - \dot{b}/b]^2, \tag{2.8}$$

$$\theta = \dot{a}/a + 2(\dot{b}/b), \tag{2.9}$$

and

$${}^3R = -6a^{-2}, \tag{2.10}$$

where ${}^3R$ denotes the Ricci curvature of the hypersurfaces of homogeneity. (Note that when $a/b = \text{const}$, the heat conduction vector and the shear are consequently both zero—this case is discussed in some detail in Ref. I.) We also recall the identity (the "generalized Friedmann equation" or "first integral"),

$$\theta^2 = 9/a^2 + 3\sigma^2 + 3\rho, \tag{2.11}$$

where use has been made of (2.10). By adding Eq. (2.6) to two times Eq. (2.7), we obtain [using (2.11)] the Raychaudhuri equation:

$$\dot{\theta} = -\tfrac{1}{3}\theta^2 - 2\sigma^2 - \tfrac{1}{2}(\rho + 3\bar{p}). \tag{2.12}$$

The second independent equation we shall write as

$$\dot{\sigma} = -2\eta\sigma - \sigma\theta, \tag{2.13}$$

which is obtained by subtracting Eq. (2.7) from Eq. (2.6). Finally, from the conservation law ($T^{ab}{}_{;b}u_a = 0$), we find that

$$\dot{\rho} = -(\rho + p)\theta + \zeta\theta^2 + 4\eta\sigma^2 + (4/\sqrt{3}) \times \sigma[\tfrac{1}{3}\theta^2 - \sigma^2 - \rho]. \tag{2.14}$$

Now, we define the new variables β and x, and the new time coordinate Ω, as follows:

$$\beta \equiv \frac{2}{\theta}\left[\frac{\dot{a}}{a} - \frac{\dot{b}}{b}\right], \quad \beta = 2\sqrt{3}\,\frac{\sigma}{\theta}, \tag{2.15a}$$

so that β measures the rate of shear in terms of the expansion,

$$x \equiv 3\rho/\theta^2, \tag{2.15b}$$

so that x measures the dynamical importance of the matter content, and

$$\ell \equiv e^{-\Omega}, \quad \frac{d\Omega}{dt} = -\frac{1}{3}\theta, \tag{2.15c}$$

where ℓ is the representative length scale with $\theta = 3\dot{\ell}/\ell$. Therefore, using Eqs. (2.15) and Eq. (2.12), we can write Eq. (2.13) as

$$\frac{d\beta}{d\Omega} = \frac{1}{2}\beta\left[4 - \beta^2 - x - \frac{9p}{\theta^2} + \frac{9\zeta}{\theta} + 12\frac{\eta}{\theta}\right], \tag{2.16}$$

and we can write Eq. (2.14) as

$$\frac{dx}{d\Omega} = x[1 - x - \beta^2] + \beta\left[2x - 2 + \frac{\beta^2}{2}\right] + 9\frac{p}{\theta^2}(1-x) - 9\frac{\zeta}{\theta}(1-x) - 3\frac{\eta}{\theta}\beta^2 \tag{2.17}$$

[where we have used Eq. (2.12) for $\bar{p}$]. Finally, we note that from (2.11) we are only interested in the region

$$\beta^2 + 4x \leqslant 4, \quad x \geqslant 0. \tag{2.18}$$

III. EQUATIONS OF STATE

In order to complete the system of equations we need to specify three equations of state for p, ζ, and η. In principle, these equations of state can be derived from kinetic theory[6–8]. For example, Collins and Stewart[9] considered a class of nonrotating Bianchi models (that included Bianchi type V's) with shear viscosity (but no bulk viscosity) in which $\eta = \frac{1}{5}pt_{\text{coll}}$, where the harmonic mean of the collision times for the various reactions, t_{coll}, is assumed to be given by $t_{\text{coll}} = 1/\sqrt{2}n\Sigma$, where n is the number density and Σ is the mean total scattering cross section (related to the temperature by a suitable approximate relationship). Subject to some additional, physically motivated assumptions, Collins and Stewart[9] concluded from a qualitative analysis that for arbitrary initial conditions the shear anisotropy could be arbitrarily large now, and that the Universe need not have been in thermal equilibrium during the early stages. These conclusions are relevant in determining whether strong dissipative mechanisms in the early Universe (such as neutrino viscosity) could produce the observed highly isotropic matter distribution.[10,11]

However, in practice, it is necessary to specify phenomenological equations of state subject to a set of thermodynamical laws.[12] Of course, specification of p, ζ, and η requires special conditions for which there may be no physical foundations. This specification should be subject to physical constraints such as p, ζ, and η should tend to zero as the density tends to zero and must be subject to the energy conditions. It goes without saying that the behavior of the fluid (e.g., it's asymptotic behavior) depends on the assumptions made on the form of these physical quantities. We also note that in writing down the energy–momentum tensor for a viscous fluid with heat conduction in the form of Eq. (2.2) we have assumed that $\pi_{ab} = -2\eta\sigma_{ab}$, where π_{ab} is the tensor of anisotropic stress. This assumption (the "viscosity assumption") is valid whenever the anisotropy is small (i.e., $|\pi_{ab}/p| \ll 1$).

There are some equations of state that are commonly used that, although not widely applicable, are obtained as a result of approximate estimates for particular fluids. The barotropic equation of state, $p = (\gamma - 1)\rho$ is often assumed. Here, $1 \leqslant \gamma \leqslant 2$ is necessary for the existence of local mechanical stability and for the speed of sound in the fluid to be no greater than the speed of light. Belinskii and Khalatnikov[13,14] consider viscous fluids in which the viscosity coefficients depend on powers of the energy density. It is argued that this approach will be valid whenever the kinetic coefficients that arise at a higher order of approximation will be proportional to the energy taken to a power greater than the one characterizing the coefficients of ζ and η. Consequently, this approach ought to be valid (at least) near the initial singularity when the energy density is very small. Moreover, it is argued that the qualitative picture ought not to change substantially from that obtained from this approach.[13]

As noted above, in order to complete the system of equations three equations of state must be given, specifying p, ζ, and η in terms of the other physical quantities. Since we are considering a viscous fluid with heat conduction, in general all physical quantities depend on two independent thermodynamical variables, one of which will be chosen as ρ and the second of which will be denoted by X (e.g., temperature or entropy density), viz.,

$$\begin{aligned} p &= p(\rho, X), \\ \zeta &= \zeta(\rho, X), \\ \eta &= \eta(\rho, X). \end{aligned} \tag{3.1}$$

As also noted earlier, in principle these equations can be obtained from kinetic theory, but in practice phenomenological equations of state need to be assumed. In addition, we also recall the variables β and x occurring in Eqs. (2.16) and (2.17), viz.,

$$\beta = 2\sqrt{3}(\sigma/\theta), \tag{3.2a}$$

$$x = 3\rho/\theta^2, \tag{3.2b}$$

and note that, firstly, β and x are dimensionless, and, secondly, in the absence of viscosity and with $p = (\gamma - 1)\rho$, Eqs. (2.16) and (2.17) form a plane-autonomous system in β and x.

Here, we are going to consider equations of state of the following form:

$$\begin{aligned} p/\theta^2 &= F(\beta, x), \\ \zeta/\theta &= G(\beta, x), \\ \eta/\theta &= H(\beta, x). \end{aligned} \tag{3.3}$$

Let us argue in favor of these equations:

(i) First, Eqs. (3.3) are completely dimensionless equations since, as noted above, β and x are dimensionless, and the ratios p/θ^2, ζ/θ, and η/θ are dimensionless. It can be argued that dimensionless equations of state are the most physically natural. In particular, it might be expected that such equations will be valid whenever the physics is scale invariant. Scale-invariant solutions in classical hydrodynamics have been a fruitful source of models for physical systems having no intrinsic units of length, mass, or time. Moreover, this situation might be especially pertinent in the qualitative analysis that we intend to carry out, where it will be of interest to study physical systems that have no intrinsic scale in an asymptotic sense.

We note that our particular "choice" of dimensionless physical quantities (3.3) is to some extent arbitrary, and the choice has been made for convenience. However, Eqs. (3.3) are independent of this choice. For example, if σ is nonzero, and if we assume that $p/\sigma^2 = f(\beta, x)$ and $\eta/\sigma = h(\beta, x)$, then

$$p/\theta^2 = [\sigma^2/\theta^2] f(\beta, x) = F(\beta, x),$$

and

$$\eta/\theta = [\sigma/\theta] h(\beta, x) = H(\beta, x).$$

In addition, $p/\rho = h(\beta, x)$ implies that $p/\theta^2 = (\rho/\theta^2) h(\beta, x) = H(\beta, x)$.

(ii) Second, Eqs. (3.3) are the *most general* equations of state such that Eqs. (2.16) and (2.17) reduce to a plane-autonomous system, enabling us to study the viscous models under consideration qualitatively in a straightforward manner. In general, it may be possible for the system of equations under investigation to reduce to an autonomous system of dimension greater than two even if Eqs. (3.3) are not assumed. However, it is strongly suggested by Eqs. (2.16) and (2.17) that equations of state (3.3) are clearly the most natural in any attempted reduction to an autonomous system, and, moreover, from the above comments Eqs. (3.3) are perhaps suggested by dimensional considerations.

(iii) Next, since we are considering spatially homogeneous models it is natural for all the physical quantities ρ, p, ζ, η (etc.) and the kinematical quantities θ and σ to depend only on t, so that p/θ^2, ζ/θ, η/θ, β, and x are function of t alone, and the equations of state can be considered in the form (3.3) in all generality.

(iv) Equations (3.3) are completely general for physical systems in which β and x can be regarded as independent thermodynamical variables.

(v) The most commonly considered equations of state are of the form (3.3). For example, the barotropic equation of state $p = (\gamma - 1)\rho$ is equivalent to $p/\theta^2 = (\gamma - 1)\rho/\theta^2 = \frac{1}{3}(\gamma - 1)x$, and is consequently of the form (3.3), where F is simply given by $F(\beta, x) = \frac{1}{3}(\gamma - 1)x$. Also, $\zeta = \zeta_0 \rho^{1/2}$ and $\eta = \eta_0 \rho^{1/2}$ are equivalent to $\zeta/\theta = \zeta_0[\rho/\theta^2]^{1/2}$ and $\eta/\theta = \eta_0[\rho/\theta^2]^{1/2}$, which are simple examples of Eqs. (3.3) with $G(\beta, x) = (\zeta_0/\sqrt{3})x^{1/2}$ and $H(\beta, x) = (\eta_0/\sqrt{3})x^{1/2}$. In particular, Belinskii and Khalatnikov[13] have studied viscous fluid models in which the equations of state are asymptotically of this form. In addition, since these "common" equations of state (particularly the barotropic equation of state) are derived from kinetic theory, it can be argued that there is some kinetic theoretical basis for Eqs. (3.3).

(vi) Finally, FRW models can be written in terms of a plane-autonomous system of equations in (ρ,θ) space. Murphy[15] included bulk viscosity in isotropic and spatially homogeneous cosmologies, and it can be shown that the plane-autonomous character of the resulting field equations can be retained if the bulk viscosity dissipation is modeled by means of an equation of state $\bar{p} = \bar{p}(\rho,\theta)$ (where $\zeta = -\partial\bar{p}/\partial\theta$). It is known that FRW cosmological models are structurally unstable.[5] Golda *et al.*[16] have shown that if the equation of state $\bar{p} = (\gamma-1)\rho - \zeta\theta$ with $\zeta(\rho) = \zeta_0(\rho)^m$ is assumed, then the *only* possible solutions that are structurally stable are those with $m = \frac{1}{2}$ [that is, those in which $\zeta/\theta = \zeta_0(\rho/\theta^2)^{1/2}$, which is of the form of (3.3)].

It should be noted once again that the analysis, and, in particular, the discussion above, is quite general, and is equally applicable in all Bianchi-type models. For illustration, we are considering only Bianchi V imperfect fluid models here; the analysis will be extended elsewhere.

No qualitative analysis can be undertaken unless F, G, and H are further specified. Here we shall assume for simplicity that F, G, and H are independent of β. This, of course, still enables us to write the equations as a plane-autonomous system. In addition, if it is possible for ρ and θ to be regarded as the two independent thermodynamical variables (recall the baryon conservation law in the form $\dot{n} + n\theta = 0$, where n is the particle number density), then this assumption is the special case guaranteeing that the equations of state are dimensionless. Finally, in general, this will always be possible if all the quantities of interest are functions of t only, as is expected in the spatially homogeneous models under consideration.

Moreover, for simplicity we shall consider the case when $F(x)$, $G(x)$, and $H(x)$ are functions that depend on a power of the argument; namely,

$$p/\theta^2 = p_0 x^l, \tag{3.4a}$$

$$\zeta/\theta = \zeta_0 x^m, \tag{3.4b}$$

$$\eta/\theta = \eta_0 x^n, \tag{3.4c}$$

where l, m, and n are constants. Such equations may be valid, at least in an approximate sense, and ought to be applicable in a qualitative analysis. In addition, these equations are consistent with the "common" examples alluded to above. Using Eqs. (3.4), Eqs. (2.16) and (2.17) reduce to a plane-autonomous system. [We note that since ${}^3R<0$ [Eq. (2.10)] it follows that $\theta^2>0$ [Eq. (2.11)] and $\dot{\theta}<0$ [Eq. (2.12)] imply that if $\theta_0>0$ (at present) then $\theta>0$ for all t; hence all quantities in equations (3.4) are well defined. Care must be taken in extending this analysis to Bianchi IX models in which ${}^3R>0$ since θ is no longer always positive.]

IV. EXACT SOLUTIONS

In a series of papers cosmological models have been examined in which the condition $\sigma^2/\theta^2 = \text{const.}$ is assumed (I11, I12, and I16–21), and Bali[17] has investigated Bianchi I viscous fluid cosmology with magnetic field under the assumption $\eta = \eta_0\theta$ and has found an exact solution (in which $\lim_{t\to 0}\sigma/\theta = 0$). In this section we shall investigate exact solutions of Eqs. (2.6) and (2.7) with equations of state given by (3.4). In particular, we shall consider the simple case in which $l = 1$, $m = 0$, and $n = 0$ in Eqs. (3.4), i.e.,

$$p = (\gamma-1)\rho,\quad \zeta = \zeta_0\theta,\quad \eta = \eta_0\theta \tag{4.1}$$

(where $\gamma = 3p_0 + 1$), in which first integrals of Eqs. (2.6) and (2.7) can be obtained by the method of decomposable operators of Maartens and Nel.[18] Exact solutions will be extremely useful in combination with any possible qualitative analysis.

Using equations of state (4.1) [and employing Eqs. (2.3), (2.4), and (2.9)], Eqs. (2.6) and (2.7) become, respectively,

$$-\frac{2\ddot{b}}{b} + \left[\zeta_0 + \frac{4}{3}\eta_0\right]\frac{\dot{a}^2}{a^2} + \left[-\gamma + 4\zeta_0 - \frac{8}{3}\eta_0\right]\frac{\dot{b}^2}{b^2} + \left[-2(\gamma-1) + 4\zeta_0 + \frac{4}{3}\eta_0\right]\frac{\dot{a}\dot{b}}{ab} + \frac{1}{a^2}(3\gamma-2) = 0, \tag{4.2}$$

and

$$-\frac{\ddot{a}}{a} - \frac{\ddot{b}}{b} + \left[\zeta_0 - \frac{2}{3}\eta_0\right]\frac{\dot{a}^2}{a^2} + \left[(1-\gamma) + 4\zeta_0 + \frac{4}{3}\eta_0\right]\frac{\dot{b}^2}{b^2} + \left[(1-2\gamma) + 4\zeta_0 - \frac{2}{3}\eta_0\right]\frac{\dot{a}\dot{b}}{ab} + \frac{1}{a^2}(3\gamma-2) = 0. \tag{4.3}$$

These equations constitute two independent (coupled, nonlinear, second-order, ordinary) differential equations for a and b. Multiplying Eq. (4.2) by the constant α, and Eq. (4.3) by the constant β, and adding, yields the equation

$$[-\beta]\frac{\ddot{a}}{a} + [-2\alpha-\beta]\frac{\ddot{b}}{b} + \frac{\dot{a}^2}{a^2}[(\alpha+\beta)\zeta_0 + \tfrac{2}{3}(2\alpha-\beta)\eta_0] + \frac{\dot{b}^2}{b^2}[\beta(1-\gamma) - \alpha\gamma + 4(\alpha+\beta)\zeta_0 + \tfrac{4}{3}(\beta-2\alpha)\eta_0] + \frac{\dot{a}\dot{b}}{ab}[(-2\alpha)(\gamma-1) + (1-2\gamma)\beta + 4(\alpha+\beta)\zeta_0 + \tfrac{2}{3}(2\alpha-\beta)\eta_0] + (1/a^2)(3\gamma-2)(\alpha+\beta) = 0. \tag{4.4}$$

Using the method of decomposable differential operators[18], we can find a first integral of this equation [and hence Eqs. (4.2) and (4.3)] whenever the following algebraic equation is satisfied:

$$\beta(\alpha+\beta)(4\alpha+\beta)(\gamma-2) + (\alpha+\beta)(2\alpha-\beta)^2[\zeta_0 + \tfrac{4}{3}\eta_0] = 0. \tag{4.5}$$

The solutions of this equation are:

(i) $\alpha+\beta=0$, and (ii) α

$$=\{2(\zeta_0+\tfrac{4}{3}\eta_0)^{-1}[(\xi_0+\tfrac{4}{3}\eta_0-(\gamma-2)) \pm\sqrt{(2-\gamma)(2+3\zeta_0+4\eta_0-\gamma)}\,]\}\beta \quad (\zeta_0+\tfrac{4}{3}\eta_0\neq 0).$$

(i) Taking $\alpha+\beta=0$ [i.e., subtracting Eq. (4.3) from (4.2)] yields the *general* first integral:

$$\frac{d}{dt}\left(\frac{a}{b}\right)+K\left(\frac{a}{b}\right)(ab^2)^{-(2\eta_0+1)} \tag{4.6}$$

(where K is an integration constant). Defining the new time coordinate τ by

$$\frac{d\tau}{dt}=(ab^2)^{-(2\eta_0+1)}, \tag{4.7}$$

Eq. (4.6) integrates to

$$(a/b)=C\exp(K\tau). \tag{4.8}$$

(ii) We define $\Sigma:=\frac{1}{2}[1+k\pm\sqrt{k(3+k)}\,]$, where $k=(2-\gamma)/(\zeta_0+\frac{4}{3}\eta_0)$ is non-negative (since $\gamma\leqslant 2$ and $\zeta_0>0$ and $\eta_0>0$) ensuring two real values for Σ (Σ_+ and Σ_-). Taking $\alpha=\Sigma\beta$ yields the *general* first integral(s);

$$\frac{d}{dt}\left\{a^{[-(\Sigma+1)\zeta_0-(2/3)(2\Sigma-1)\eta_0]}b^{[-4(\Sigma+1)\zeta_0-(4/3)(1-2\Sigma)\eta_0+\gamma(1+\Sigma)-(4\Sigma^2+2\Sigma+1)](2\Sigma+1)^{-1}}\frac{d}{dt}(ab^{2\Sigma+1})\right\}$$
$$=(3\gamma-2)(\Sigma+1)a^{[-(\Sigma+1)\zeta_0-(2/3)(2\Sigma-1)\eta_0-1]}b^{[-4(\Sigma+1)\zeta_0-(4/3)(1-2\Sigma)\eta_0+\gamma(1+\Sigma)+2\Sigma](2\Sigma+1)^{-1}}. \tag{4.9}$$

Defining the new variable $B:=e^{K\tau}b^{2(\Sigma+1)}$, and using the general first integral given by Eq. (4.8) in terms of the time coordinate τ, Eq. (4.9) yields the following differential equation for B:

$$\frac{B''}{B}+p\left(\frac{B'}{B}\right)^2+q\left(\frac{B'}{B}\right)=\overline{C}e^{r\tau}B^s, \tag{4.10}$$

where the constants $\overline{C}$, p, q, r, and s are given by

$$\overline{C}=(3\gamma-2)(\Sigma+1)C^{4\eta_0},$$
$$\frac{r}{K}=\frac{2(2\Sigma-1)}{(\Sigma+1)}\eta_0-\frac{2}{1+\Sigma},$$
$$s=6\eta_0/(\Sigma+1)+2/(1+\Sigma), \tag{4.11}$$
$$p=\frac{-(5+2\Sigma)(\zeta_0+\frac{4}{3}\eta_0)+(\gamma-2)}{2(2\Sigma+1)}-1,$$
$$\frac{q}{K}=\frac{-(2\Sigma+3)(2\Sigma-1)(\zeta_0+\frac{4}{3}\eta_0)+(\gamma-2)}{2(2\Sigma+1)}.$$

In the above, a prime denotes differentiation with respect to τ. Equation (4.10) is a (single) second-order, ordinary differential equation for the (single) variable B.

(iii) Let us consider the case $\gamma=2$ (corresponding to stiff matter) separately. In this case $k=0$ and $\Sigma=\frac{1}{2}$ [this case corresponds to a double root for α/β in Eq. (4.5)]. Taking $2\alpha-\beta=0$ when $\gamma=2$ [i.e., adding twice Eq. (4.3) to Eq. (4.2)] yields the first integral:

$$a^2(ab^2)^{3\zeta_0/2-1}\frac{d}{dt}\left[(ab^2)^{-3\zeta_0/2}\frac{d}{dt}(ab^2)\right]=6. \tag{4.12}$$

Again, employing the first integral $a=Ce^{K\tau}b$ and defining the new variable X by $e^X=b^3e^{K\tau}$, in terms of the time coordinate τ defined by (4.7) Eq. (4.12) reduces to a "simple" second-order differential equation for X, which we can attempt to solve in order to obtain a solution of the Bianchi V imperfect fluid field equations in the particular case of stiff matter. Alternatively, defining X by $e^X=B$, when $\gamma=2$ ($\Sigma=\frac{1}{2}$) Eq. (4.10) becomes

$$X''-[\tfrac{3}{2}\zeta_0+2\eta_0](X')^2=\overline{C}\exp\{[4\eta_0+\tfrac{4}{3}]X-\tfrac{4}{3}K\tau\}. \tag{4.13}$$

If a solution for X is found to this second-order differential equation, a and b are then obtained by

$$a=Ce^{(1/3)(X+2K\tau)},$$
$$b=e^{(1/3)(X-K\tau)}. \tag{4.14}$$

We note the simple solution

$$X=X_0+[K/(3\eta_0+1)]\tau, \tag{4.15}$$

to Eq. (4.13), where the constant X_0 satisfies

$$\overline{C}\exp[(4\eta_0+\tfrac{4}{3})X_0]=-[(3\zeta_0+4\eta_0)/2(3\eta_0+1)^2]K^2, \tag{4.16}$$

whence

$$a=C\exp[\tfrac{1}{3}X_0+[(2\eta_0+1)/(3\eta_0+1)]K\tau],$$
$$b=\exp[\tfrac{1}{3}X_0-[\eta_0/(3\eta_0+1)]K\tau], \tag{4.17}$$

and t and τ are related by

$$(t_0-t)=C_1e^{C_2\tau}, \tag{4.18}$$

where

$$C_1=\frac{-(3\eta_0+1)}{K(2\eta_0+1)}\times\left[\frac{-(3\zeta_0+4\eta_0)K^2C^{4/3}}{12(3\eta_0+1)^2}\right]^{(2\eta_0+1)/(4\eta_0+4/3)}, \tag{4.19a}$$

and

$$C_2=(2\eta_0+1)K/(3\eta_0+1). \tag{4.19b}$$

Unfortunately, a straightforward calculation using Eq. (2.4) shows that this solution is unphysical since it leads to a negative energy density.

V. PLANE-AUTONOMOUS SYSTEMS AND DISCUSSION

Collins[4,5] was the first to use geometric techniques of standard differential equations theory, analyzing both nonrotating and tilting Bianchi-type models in the case of a perfect fluid source. Roy and Prakesh[19] derived some results for viscous fluid models of Petrov type D, under the unphysical assumption of constant shear and constant ζ. Belinskii and Khalatnikov[13,14] were the first to consider the qualitative behavior of spatially homogeneous viscous fluid cosmological models in any generality. In particular, they investigated viscous fluid Bianchi I models with barotropic equation of state $p=(\gamma-1)\rho$ and in which the viscosity coefficients depend (only) on the powers of the energy density, in which case the field equations reduce to a plane-autonomous system.

In this article we have introduced a new approach for dealing with equations of state in Bianchi-type cosmologies and we have shown by way of illustration that the field equations in LRS Bianchi V imperfect fluid cosmologies can be written as a plane-autonomous system, facilitating a qualitative analysis of such cosmological models. This work therefore generalizes the previous results in nonrotating and (LRS) tilting perfect fluid Bianchi-type (including type V) models[4,5], and in viscous fluid Bianchi I models,[13] to the imperfect Bianchi-V case.

As noted above, Eqs. (2.16) and (2.17) reduce to a plane-autonomous system when Eqs. (3.3) or (3.4) are employed. For illustration, if we consider the equations of state in the form

$$p/\theta^2=\tfrac{1}{3}(\gamma-1)x, \tag{5.1a}$$

$$\zeta/\theta=\zeta_0 x^{1/2}, \tag{5.1b}$$

$$\eta/\theta=\eta_0 x^{1/2}, \tag{5.1c}$$

then Eqs. (2.16) and (2.17) reduce to

$$\frac{d\beta}{d\Omega}=\frac{1}{2}\beta\,[4-\beta^2-(3\gamma-2)x+3x^{1/2}[3\zeta_0+4\eta_0]], \tag{5.2}$$

and

$$\frac{dx}{d\Omega}=x[(3\gamma-2)(1-x)-\beta^2]+\beta\left[2x-2+\frac{\beta^2}{2}\right] - 3x^{1/2}[3(1-x)\zeta_0+\eta_0\beta^2]. \tag{5.3}$$

We shall analyze the qualitative nature of Bianchi V imperfect fluid cosmological models in a future paper.

ACKNOWLEDGMENTS

I would like to thank Barry Collins, Charles Curry, and John Wainwright for helpful discussions. I would also like to thank Janice Smith for her excellent typing of the manuscript.

This work was supported, in part, by the Natural Sciences and Engineering Research Council of Canada through an operating grant to the author.

[1] A. A. Coley, Gen. Relativ. Grav. **22**, 3 (1990) (paper I).
[2] M. A. H. MacCallum, Commun. Math. Phys. **20**, 57 (1971).
[3] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969); K. Rosquist, and R. T. Jantzen, Phys. Rep. **166**, 89 (1988); B. K. Nayak and G. B. Bhuyan, Gen. Relativ. Grav. **18**, 79 (1986).
[4] C. B. Collins, Commun. Math. Phys. **23**, 137 (1971).
[5] C. B. Collins, Commun. Math. Phys. **39**, 131 (1974).
[6] J. M. Stewart, *Nonequilibrium Relativistic Kinetic Theory* (Springer, Heidelberg, 1971).
[7] W. A. VanLeeuwen, A. J. Kox, and S. R. de Groot, Physica A **79**, 233 (1975).
[8] N. Caderni and R. Fabri, Phys. Lett. B **69**, 508 (1977).
[9] C. B. Collins and J. M. Stewart, Mon. Not. R. Astron. Soc. **153**, 419 (1971).
[10] C. W. Misner, Nature **214**, 40 (1967).
[11] C. W. Misner, Astrophys. J. **151**, 431 (1968).
[12] A. A. Coley, Phys. Lett. A **137**, 235 (1989).
[13] V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976).
[14] V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **45**, 1 (1977).
[15] G. L. Murphy, Phys. Rev. D **8**, 423 (1973).
[16] Z. Golda, M. Heller, and M. Szydlowski, Astrophys. Space Sci. **90**, 313 (1983).
[17] R. Bali, Astrophys. Space Sci. **116**, 81 (1985).
[18] R. Maartens and S. D. Nel, Commun. Math. Phys. **59**, 273 (1978).
[19] S. R. Roy and S. Prakash, J. Phys. A **9**, 261 (1976).

Conformally Ricci-flat viscous fluids

B. O. J. Tupper
Department of Mathematics and Statistics, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada

(Received 16 October 1989; accepted for publication 28 February 1990)

It is shown that if the gradient of the conformal scalar in a conformally Ricci-flat space-time is parallel to an eigenvector (timelike, spacelike, or null) of the stress-energy tensor, then acceptable solutions, without restrictions on the physical and kinematical quantities, of the Einstein field equations for a viscous fluid with heat conduction may be found.

I. INTRODUCTION

If $\bar{g}_{ab}$ is the metric tensor of a space-time $\bar{S}$ that satisfies the Einstein field equations (EFE's) for vacuum, i.e., $\bar{R}_{ab} = 0$, and if g_{ab} is the metric tensor of a space-times that is conformal to $\bar{S}$, i.e.,

$$g_{ab} = e^{2U}\bar{g}_{ab}, \quad U = U(x^c), \tag{1.1}$$

then the Einstein tensor of S is given by

$$G_{ab} = -2U_aU_b - 2U_{a;b} - g_{ab}(U^cU_c - 2U^c{}_{;c}), \tag{1.2}$$

where $U_a = U_{,a}$ and the semicolon denotes covariant differentiation with respect to the Christoffel symbols formed from g_{ab}. The space-time S is termed conformally Ricci-flat.

Given the expression (1.2), it is interesting to ask the following. What types of matter distribution have stress-energy tensors that can be written in the form

$$T_{ab} = -2U_aU_b - 2U_{a;b} - g_{ab}(U^cU_c - 2U^c{}_{;c}) \tag{1.3}$$

and be realistic in that they satisfy the necessary energy conditions? In other words, what matter distributions can be admitted by conformally Ricci-flat space-times? This question has been answered in part by Van den Bergh[1] who has discussed perfect fluid and Einstein–Maxwell space-times with stress-energy tensor of the form (1.3). Carot and Mas[2] have investigated conformally Ricci-flat space-times corresponding to a distribution of viscous fluid with heat conduction, i.e., the stress-energy tensor (1.3) is of the form

$$T_{ab} = (\mu + p)u_au_b + pg_{ab} - 2\eta\sigma_{ab} + q_au_b + q_bu_a, \tag{1.4}$$

where u, p, u_a, σ_{ab}, q_a, and $\eta(\geqslant 0)$ are, respectively, the density, pressure, fluid velocity vector, shear tensor, heat-conduction vector, and shear-velocity coefficient. Note that $\sigma_{ab}u^b = q_au^a = 0$.

In order to show that the stress-energy tensor given by (1.3) may be of the form (1.4), it is assumed in Ref. 2 that the gradient U_a is proportional to the velocity vector of the fluid, i.e.,

$$U_a = \gamma u_a. \tag{1.5}$$

As a consequence, it is found that the viscous fluid must be either shear-free or collapsing. Assumption (1.5) is unnecessarily restrictive, as is clear from the fact that the FRW cosmological models, which are conformally flat and thus conformally Ricci-flat, have been shown[3] to satisfy the field equations for a viscous heat-conducting fluid. In this imperfect fluid interpretation of the FRW models, the shear is nonzero and the expansion is positive, so it cannot correspond to the situation in which U_a is proportional to u_a. In fact, in the $k = 0$ FRW models, the conformal factor e^{2U} is a function of t only, so that U_a is proportional to the timelike eigenvector of T_{ab}. In the case of the perfect fluid interpretation of the FRW model, the timelike eigenvector is parallel to u_a, but in the viscous fluid interpretation, which requires a tilting four-velocity, the timelike eigenvector lies in the two-space spanned by u_a and q_a.

In this paper, we show that the assumption that U_a is parallel to the timelike, spacelike, or null eigenvector of T_{ab} given by (1.4) leads to conformally Ricci-flat viscous fluid solutions without the restrictions mentioned earlier. In Sec. II we give a brief discussion of the eigenvectors of T_{ab}, and in the subsequent sections, we present examples of viscous fluid solutions obtained by this technique.

II. EIGENVECTORS OF T_{ab}

A thorough investigation of the eigenvectors of the stress-energy tensor (1.4) has been made by Kolassis *et al.*[4] Here we will mention only those results that are relevant to our problem.

We write $q_a = Qe_a$, where $Q = (q_aq^a)^{1/2}$ is the magnitude of q_a and e_a is the unit spacelike vector in the direction of q_a, and we assume that q_a (equivalently e_a) is an eigenvector of the shear tensor σ_{ab}, i.e.,

$$\sigma_{ab}e^b = \lambda e_a. \tag{2.1}$$

This assumption implies that there exist no shear velocities between neighborhood surface elements orthogonal to the direction of the heat flux.[4] Defining the quantity X by

$$2X = \mu + p - 2\eta\lambda, \tag{2.2}$$

we note that the dominant energy condition is satisfied if

$$\mu \geqslant X \geqslant Q. \tag{2.3}$$

The eigenvectors of T_{ab} are

$$t_a = u_a + Q^{-1}(X - \sqrt{X^2 - Q^2})e_a \quad \text{(timelike)}, \tag{2.4}$$

$$s_a = u_a + Q^{-1}(X + \sqrt{X^2 - Q^2})e_a \quad \text{(spacelike)}, \tag{2.5}$$

together with two other spacelike eigenvectors in the two-space orthogonal to u_a and e_a. If $Q = 0$, then u_a and e_a are each eigenvectors of T_{ab} and if, in addition, $X = 0$, the corresponding eigenvalues are equal so that any linear combination is an eigenvector. In this case, T_{ab} is of Segré type ((1,1)1 1). If T_{ab} is of Segré type (2 1 1), or one of its degen-

0022-2488/90/071704-03$03.00

eracies (viscous fluid solutions of this type do exist[5,6]), there exists a null eigenvector $u_a \pm e_a$ that requires that $X = \pm Q$.

If U_a is an eigenvector of G_{ab} given by (1.2), then U_a must be an eigenvector of $U_{a;b}$, i.e.,

$$U_{a;b} U^b = \omega U_a \tag{2.6}$$

or

$$(U_b U^b)_{,a} = \omega U_a, \tag{2.7}$$

where ω is a scalar function satisfying $\omega_{[b} U_{a]} = 0$. In particular, if U_a is a null vector, then $\omega = 0$, and we find that

$$U^a{}_{;a} = \tfrac{1}{2}(X - \mu) = \tfrac{1}{4}(-\mu + p - 2\eta\lambda). \tag{2.8}$$

By contracting (1.3) and (1.4), we find that

$$6U^a{}_{;a} = -\mu + 3p, \tag{2.9}$$

and from Eqs. (2.8) and (2.9) we see that when U_a is a null eigenvector of T_{ab}, the viscous fluid must satisfy the condition

$$\mu + 3p + 6\eta\lambda = 0. \tag{2.10}$$

III. EXAMPLES

We now present a number of examples, known and new, of conformally Ricci-flat space-times that satisfy the EFE's with T_{ab} given by (1.4) and for which U_a is an eigenvector of T_{ab}.

(i) We start with a case in which U_a is timelike. Consider the Einstein–de Sitter universe with metric written in conformal coordinates, i.e.,

$$ds^2 = t^4(-dt^2 + dx^2 + dy^2 + dz^2). \tag{3.1}$$

Two types of viscous fluid solution corresponding to this model are known,[3] namely *radial solutions*, for which the spatial component of the tilting four-velocity u^a is in the radial direction of spherical polar coordinates, and *axial solutions*, for which the spatial component of u^a is in the direction of one of the Cartesian spatial directions. We will consider an axial solution with u^a and q^a given by

$$\begin{aligned} u^a &= t^{-2}(\cosh\phi, \sinh\phi, 0, 0), \\ q^a &= -Qt^{-2}(\sinh\phi, \cosh\phi, 0, 0), \end{aligned} \tag{3.2}$$

where $\phi = \phi(t)$ is a scalar function. Using (3.1) and (3.2), the solution of the EFE's for viscous fluid is

$$\begin{aligned} &\mu = 12t^{-6}\cosh^2\phi, \quad p = 4t^{-6}\sinh^2\phi, \\ &Q = 12t^{-6}\cosh\phi\sinh\phi, \quad \eta\dot{\phi} = -6t^{-4}\sinh\phi, \end{aligned} \tag{3.3}$$

so that $\cosh\phi$ must be decreasing for $\eta \geqslant 0$.

The eigenvalue λ is given by $\lambda = \tfrac{3}{2}t^{-2}\dot{\phi}\sinh\phi$, so that

$$X = 6t^{-4}(\cosh^2\phi + \sinh^2\phi), \tag{3.4}$$

and the timelike eigenvector t_a, given by (2.4), takes the form

$$t_a = t^2 \operatorname{sech}\phi(-1, 0, 0, 0), \tag{3.5}$$

which, since $e^U = t^2$, is parallel to $U_a = (2t^{-1}, 0, 0, 0)$. Note also that Eq. (2.6) is satisfied with $\omega = 6t^{-6}$.

(ii) As a second example in which U_a is timelike, we look for a conformally Ricci-flat space-time that is not conformally flat. We start with a special case of the Kasner vacuum solution with metric

$$ds^2 = -dt^2 + t^{-2/3}dx^2 + t^{4/3}(dy^2 + dz^2), \tag{3.6}$$

and look for a viscous fluid solution with $e^U = t^m$, i.e., we use the metric

$$\begin{aligned} ds^2 = t^{2m}(&-dt^2 + t^{-2/3}dx^2 + t^{4/3}dy^2 \\ &+ t^{4/3}dz^2). \end{aligned} \tag{3.7}$$

Taking u^a and q^a to be given by

$$\begin{aligned} u^a &= t^{-m}(\cosh\phi, t^{1/3}\sinh\phi, 0, 0), \\ q^a &= -Qt^{-m}(\sinh\phi, t^{1/3}\cosh\phi, 0, 0), \end{aligned} \tag{3.8}$$

where $\phi = \phi(t)$, the EFE's lead to

$$\begin{aligned} &\mu = \tfrac{1}{3}m(3m+2)t^{-2-2m}(2\cosh^2\phi + 1), \\ &p = \tfrac{1}{9}mt^{-2-2m}[2(3m+2)\cosh^2\phi - (15m-2)], \\ &Q = \tfrac{2}{3}m(3m+2)t^{-2-2m}\sinh\phi\cosh\phi, \\ &\eta Y = \tfrac{1}{3}mt^{-2-m}[3m+5-(3m+2)\cosh^2\phi], \end{aligned} \tag{3.9}$$

where $Y = \dot{\phi}\sinh\phi - t^{-1}\cosh\phi$. This solution satisfies the dominant energy condition if $m > 0$ and satisfies $\eta \geqslant 0$ provided that $\cosh\phi$ is decreasing and $\cosh^2\phi \geqslant (3m+5)(3m+2)^{-1}$. If $m \leqslant \tfrac{2}{3}$, then $p \geqslant 0$ for any value of $\cosh\phi$.

The eigenvalue λ is given by $\lambda = \tfrac{2}{3}t^{-m}Y$, so that

$$X = \tfrac{1}{3}m(3m+2)t^{-2-2m}(\cosh^2\phi + \sinh^2\phi), \tag{3.10}$$

and the timelike eigenvector t_a is

$$t_a = t^m \operatorname{sech}\phi(-1, 0, 0, 0), \tag{3.11}$$

which is parallel to $U_a = (mt^{-1}, 0, 0, 0)$.

(iii) We now turn to the case in which U_a is a spacelike eigenvector of T_{ab} and consider the Bertotti–Robinson non-null electrovac space-time with metric

$$\begin{aligned} ds^2 = (a^2/r^2)(&-dt^2 + dr^2 + r^2\,d\theta^2 \\ &+ r^2\sin^2\theta\,d\psi^2). \end{aligned} \tag{3.12}$$

This space-time is conformally flat with $e^U = ar^{-1}$ and satisfies the EFE's for a viscous fluid with

$$\begin{aligned} &u^a = a^{-1}r(\cosh\phi, \sinh\phi, 0, 0), \\ &e^a = a^{-1}r(\sinh\phi, \cosh\phi, 0, 0), \\ &\mu = 3p = \eta Y = a^2, \quad Q = 0, \end{aligned} \tag{3.13}$$

where $\phi = \phi(r)$ and $Y = a^{-1}(r\phi'\cosh\phi - \sinh\phi) \geqslant 0$ for $\eta \geqslant 0$. The eigenvalue $\lambda = \tfrac{2}{3}Y$, so that $X = 0$. This is an example of the special case described in Sec. II, the Segré type of T_{ab} being $((1,1)(1\,1))$. Any linear combination of u_a and e_a is an eigenvector; in this case

$$\begin{aligned} U_a &= (0, -r^{-1}, 0, 0) \\ &= a^{-1}(u_a\sinh\phi + e_a\cosh\phi). \end{aligned} \tag{3.14}$$

(iv) As a second spacelike example consider a space-time conformal to the Schwarzschild vacuum solution, i.e.,

$$ds^2 = e^{2U}[-(1-(2m/r))dt^2 + (1-(2m/r))^{-1}dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\psi^2], \tag{3.15}$$

where $U = U(r)$. We look for viscous fluid solutions with this metric for a suitable choice of U. The need to satisfy the dominant energy condition imposes strong restrictions on U; we choose

$$e^U = (1-(2m/r))^{1/2}r^{-1}, \tag{3.16}$$

and the dominant energy condition is satisfied, provided that

$$r > \sqrt{6}m. \tag{3.17}$$

The EFE's are satisfied with

$$\begin{aligned}
u^a &= [(r-2m)^{-1}r^2\cosh\phi, r\sinh\phi, 0, 0],\\
q^a &= -Q[(r-2m)^{-1}r^2\sinh\phi, r\cosh\phi, 0, 0],\\
\mu &= (r-2m)^{-2}[6m^2\cosh^2\phi + (r^2-9m^2)],\\
p &= \tfrac{1}{3}(r-2m)^{-2}[6m^2\cosh^2\phi + r^2 - 24mr + 45],\\
\eta Y &= -\tfrac{1}{3}[3m^2\cosh^2\phi - (r-2m)^2],\\
Y &= \tfrac{1}{3}[r\phi'\cosh\phi - (r+m)(r-2m)^{-1}\sinh\phi],
\end{aligned} \tag{3.18}$$

where $\phi = \phi(r)$. Note that $\eta \geqslant 0$ if $\phi > 0$, $\phi' < 0$, and $\cosh^2\phi \geqslant \frac{1}{3}m^{-2}(r-2m)^2$. The eigenvalue $\lambda = 2Y$ and X has the value

$$X = 6m^2(r-2m)^{-2}(\cosh^2\phi + \sinh^2\phi), \tag{3.19}$$

and the spacelike eigenvector s_a of T_{ab} is given by

$$s_a = u_a + \coth\phi\, e^a = (0, -r^{-1}\operatorname{csch}\phi, 0, 0), \tag{3.20}$$

and is parallel to $U_a = [0, -(r-3m)/(r(r-2m)), 0, 0]$.

(v) As our final example, we look at the case in which U_a is parallel to a null eigenvector of T_{ab}. We again use the metric of a space-time conformal to the Schwarzschild vacuum solution, i.e., the metric (3.15), but we now take $U = U(r,t)$. In fact, we specify

$$\dot{U} = -1, \quad U' = -(1-(2m/r))^{-1}, \tag{3.21}$$

so that $U_a U^a = 0$. The relations (3.21) integrate to give

$$e^U = e^{-(r+t)}(r-2m)^{2m}. \tag{3.22}$$

With this choice of U, the dominant energy condition is satisfied for all $r \geqslant 2m$ provided that $r > m^{1/2}$, i.e., we must have $m > \frac{1}{4}$.

The EFE's are satisfied with

$$\begin{aligned}
u^a &= [(1-2m/r)^{-1/2}e^{-U}\cosh\phi,\\
&\qquad (1-2m/r)^{1/2}e^{-U}\sinh\phi, 0, 0],\\
q^a &= -Q[(1-2m/r)^{-1/2}e^{-U}\sinh\phi,\\
&\qquad (1-2m/r)^{1/2}e^{-U}\cosh\phi, 0, 0],\\
\mu &= 2(1-2m/r)^{-1}e^{-2U}[(1-m/r^2)e^{2\phi}\\
&\qquad + (2/r)(1-2m/r)],\\
p &= \tfrac{2}{3}(1-2m/r)^{-1}e^{-2U}[(1-m/r^2)e^{2\phi}\\
&\qquad - (4/r)(1-2m/r)],\\
Q &= 2(1-2m/r)^{-1}(1-m/r^2)e^{-2U}e^{2\phi},\\
\eta Y &= -\tfrac{1}{3}(1-2m/r)^{-1}e^{-2U}[(1-m/r^2)e^{2\phi}\\
&\qquad - (1/r)(1-2m/r)],\\
Y &= \tfrac{1}{3}(1-2m/r)^{-1/2}e^{-U}[\phi'(1-2m/r)\cosh\phi\\
&\qquad + (\dot{\phi} - 1/r + 3m/r^2)\sinh\phi],
\end{aligned} \tag{3.23}$$

where $\phi = \phi(t,r)$ must be chosen so that η and p are positive.

The eigenvalue $\lambda = 2Y$ and we find that $X = Q$, as expected, and that Eq. (2.10) is satisfied. The null eigenvector, $u_a + e_a$, of T_{ab} is parallel to U_a, i.e.,

$$u_a + e_a = (1-2m/r)^{1/2}e^U e^{-\phi}U_a. \tag{3.24}$$

IV. CONCLUSION

We have shown that conformally Ricci-flat solutions of the EFE's for viscous fluid with heat conduction satisfying the dominant energy condition do exist without imposing any particularly stringent restrictions on the behavior of the viscous fluid. Of the examples presented here, the FRW viscous models (i) have been extensively discussed in the literature and provide interesting cosmological models as do models such as (ii). Example (iv) shows that static viscous fluid solutions exist while example (v) appears to represent a collapsing distribution of viscous fluid exterior to the Schwarzschild singularity.

ACKNOWLEDGMENT

This research was supported by the Natural Sciences and Engineering Research Council of Canada through an operating grant to the author.

[1] N. Van den Bergh, Gen. Relativ. Gravit. **18**, 649 (1986).
[2] J. Carot and Ll. Mas, J. Math. Phys. **27**, 2336 (1986).
[3] A. A. Coley and B. O. J. Tupper, Astrophys. J. **280**, 26 (1984).
[4] C. A. Kolassis, N. O. Santos, and D. Tsoubelis, Class. Quantum Grav. **5**, 1329 (1988).
[5] B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 849 (1983).
[6] A. A. Coley and B. O. J. Tupper, J. Math Phys. **30**, 649 (1989).

Ricci collineation vectors in fluid space-times

M. Tsamparlis
Department of Physics, Division of Mechanics, Astronomy and Astrophysics, University of Athens, Zografos, Athens, Greece

D. P. Mason
Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of the Witwatersrand, P. O. Wits 2050, Johannesburg, South Africa

(Received 6 July 1989; accepted for publication 17 January 1990)

The properties of fluid space-times that admit a Ricci collineation vector (RCV) parallel to the fluid unit four-velocity vector u^a are briefly reviewed. These properties are expressed in terms of the kinematic quantities of the timelike congruence generated by u^a. The cubic equation derived by Oliver and Davis [Ann. Inst. Henri Poincaré **30**, 339 (1979)] for the equation of state $p = p(\mu)$ of a perfect fluid space-time that admits an RCV, which does not degenerate to a Killing vector, is solved for physically realistic fluids. Necessary and sufficient conditions for a fluid space-time to admit a spacelike RCV parallel to a unit vector n^a orthogonal to u^a are derived in terms of the expansion, shear, and rotation of the spacelike congruence generated by n^a. Perfect fluid space-times are studied in detail and analogues of the results for timelike RCVs parallel to u^a are obtained. Properties of imperfect fluid space-times for which the energy flux vector q^a vanishes and n^a is a spacelike eigenvector of the anisotropic stress tensor π_{ab} are derived. Fluid space-times with anisotropic pressure are discussed as a special case of imperfect fluid space-times for which n^a is an eigenvector of π_{ab}.

I. INTRODUCTION

A space-time admits a Ricci collineation vector (RCV) v^a if

$$\mathscr{L}_v R_{ab} = 0, \tag{1.1}$$

where R_{ab} is the Ricci tensor and $\mathscr{L}_v$ denotes the Lie derivative along v^a. In this paper, we will consider fluid space-times and we will investigate the properties of RCVs $\eta^a = \eta u^a$, parallel to the fluid unit four-velocity vector u^a:

$$u_a u^a = -1, \; \eta = (-\eta_a \eta^a)^{1/2} > 0$$

and spacelike RCVs, $\xi^a = \xi n^a$, orthogonal to u^a:

$$n_a n^a = +1, \; n_a u^a = 0, \; \xi = (\xi_a \xi^a)^{1/2} > 0.$$

We will express the necessary and sufficient conditions for a fluid space-time to admit a timelike RCV parallel to u^a and a spacelike RCV parallel to n^a in terms of the kinematic quantities of the timelike congruence of world-lines generated by u^a and the expansion, shear, and rotation of the spacelike congruence generated by n^a, respectively.

Previous work on RCVs has been undertaken by Oliver and Davis[1,2] who gave necessary and sufficient conditions for a matter space-time to admit an RCV, $\eta^a = \eta u^a$, with $u^a = u^a_D$ where u^a_D is the dynamic four-velocity. The dynamic four-velocity u^a_D is the timelike eigenvector of the energy-momentum tensor:

$$T^{ab} u_{Db} = -\mu_D u^a_D. \tag{1.2}$$

It is characterized by the property that an observer with four-velocity u^a_D measures vanishing energy flux:

$$q^a_D = 0. \tag{1.3}$$

We will briefly review and extend the results of Oliver and Davis. We will then establish corresponding results for spacelike RCVs, $\xi^a = \xi n^a$, orthogonal to u^a by making use of the theory of spacelike congruences.[3–5]

A conservation law, valid for any RCV, was established by Collinson.[6] If v^a is an RCV, then it can be verified that

$$(R^{ab} v_b)_{;a} = 0, \tag{1.4}$$

and if Einstein's field equations,

$$R_{ab} = T_{ab} + (\Lambda - \tfrac{1}{2}T) g_{ab}, \tag{1.5}$$

are satisfied, then

$$[(T^{ab} + (\Lambda - \tfrac{1}{2}T) g^{ab}) v_b]_{;a} = 0. \tag{1.6}$$

Equation (1.6) plays an important part in the following theory as one of the necessary and sufficient conditions for a space-time to admit an RCV, v^a.

An outline of the paper is as follows. Fluid space-times that admit a timelike RCV, $\eta^a = \eta u^a$, are considered in Sec. II. The results of Oliver and Davis[1,2] are extended to the case in which $u^a \neq u^a_D$. We investigate which equations of state of the form $p = p(\mu)$ are permitted in perfect fluid space-times that admit an RCV, ηu^a, which does not degenerate to a Killing vector (KV); Oliver and Davis[2] have shown that the choice of equation of state is quite restrictive.

In Sec. III, necessary and sufficient conditions for a perfect fluid space-time to admit an RCV parallel to $n^a (n_a u^a = 0, n_a n^a = +1)$ are derived. These conditions are expressed in terms of the expansion and shear of the spacelike congruence of curves generated by n^a. Analogues of the properties of timelike RCVs parallel to u^a are investigated.

In Sec. IV, necessary and sufficient conditions for an imperfect fluid space-time to admit a spacelike RCV parallel to n^a are given. The special case in which n^a is a spacelike eigenvector of the anisotropic stress tensor π_{ab}, which in-

 0022-2488/90/071707-16$03.00

cludes fluids with anisotropic pressure, is considered in detail.

Finally, concluding remarks are made in Sec. V.

The notation and conventions of Ellis[7,8] will be followed throughout.[9]

II. TIMELIKE RICCI COLLINEATION VECTORS PARALLEL TO u^a

This section will be concerned with fluid space-times that admit a timelike RCV, $\eta^a = \eta u^a$.

The total energy-momentum tensor T_{ab} of the fluid may be decomposed with respect to u^a as

$$T_{ab} = \mu u_a u_b + p h_{ab} + 2q_{(a}u_{b)} + \pi_{ab}, \tag{2.1}$$

where μ is the total energy density measured by an observer with four-velocity u^a, q^a is the energy flux relative to $u^a (q_a u^a = 0)$, p is the isotropic pressure, and π_{ab} is the trace-free anisotropic stress tensor ($\pi_{ab} = \pi_{ba}$, $\pi_{ab}u^b = 0$, $\pi^a{}_a = 0$). We will first derive necessary and sufficient conditions for any fluid space-time to admit an RCV parallel to u^a and then we will consider the special cases of a perfect fluid ($q^a = 0$, $\pi^{ab} = 0$) and an imperfect fluid with $q^a = 0$ but $\pi^{ab} \neq 0$. This latter case is equivalent to considering $u^a = u^a_D$.

A. Imperfect fluid space-times

Theorem 2.1: If Einstein's field equations (1.5) are satisfied, then a fluid space-time with energy-momentum tensor (2.1) admits an RCV, $\eta^a = \eta u^a$, if and only if

$$\text{(i)}\quad h^c_a h^d_b \dot{\pi}_{cd} = -(\mu - p + 2\Lambda)\sigma_{ab} + \tfrac{2}{3}[\pi^{cd}\sigma_{cd} + q^c(\dot{u}_c - (\log\eta)_{,c})]h_{ab} - 2q_{(a}[\dot{u}_{b)} - (\log\eta)_{,b)} - (\log\eta)^{\cdot}u_{b)}] - \tfrac{2}{3}\theta\pi_{ab} - 2\sigma_{c(a}\pi_{b)}{}^c - \omega_{c(a}\pi_{b)}{}^c, \tag{2.2}$$

$$\text{(ii)}\quad h^b_a \dot{q}_b = -\tfrac{1}{2}(\mu + 3p - 2\Lambda)[\dot{u}_a - (\log\eta)_{,a} - \theta u_a] + (q^b{}_{;b} + q^b(\log\eta)_{,b})u_a - (\theta/3 + \log\eta)^{\cdot}) \times q_a - q^b\sigma_{ba} - q^a\omega_{ba}, \tag{2.3}$$

$$\text{(iii)}\quad [\eta(q^a + \tfrac{1}{2}(\mu + 3p - 2\Lambda)u^a)]_{;a} = 0, \tag{2.4}$$

where θ is the rate-of-expansion, σ_{ab} is the rate-of-shear tensor, and ω_{ab} is the vorticity tensor of the timelike congruence generated by u^a.

Proof: From the definition of the Lie derivative it follows that

$$\mathscr{L}_{\eta u} R_{ab} = \eta[\dot{R}_{ab} + 2u^c R_{c(a}(\log\eta)_{,b)} + 2R_{c(a}u^c{}_{;b)}], \tag{2.5}$$

which, using Einstein's field equations (1.5), may be rewritten as

$$\mathscr{L}_{\eta u} R_{ab} = \eta[\tfrac{1}{2}(\dot{\mu} + 3\dot{p})u_a u_b + \tfrac{1}{2}(\dot{\mu} - \dot{p})h_{ab} + 2(\mu + p)\dot{u}_{(a}u_{b)} + 2\dot{q}_{(a}u_{b)} + 2q_{(a}\dot{u}_{b)} + \dot{\pi}_{ab} - (\mu + 3p - 2\Lambda)u_{(a}(\log\eta)_{,b)} - 2q_{(a}(\log\eta)_{,b)} + (\mu - p + 2\Lambda)u_{(a;b)} + 2(q_t u_{(a} + \pi_{t(a})u^t{}_{;b)}]. \tag{2.6}$$

Suppose first that ηu^a is an RCV. Then (1.1) holds and the right-hand side of (2.6) vanishes. By contracting (2.6) in turn with $u^a u^b$, $u^a h^b_c$, h^{ab}, and $h^a_c h^b_d - \tfrac{1}{3}h^{ab}h_{cd}$ and by using the expansion

$$u_{a;b} = \sigma_{ab} + (\theta/3)h_{ab} + \omega_{ab} - \dot{u}_a u_b, \tag{2.7}$$

we obtain, respectively,

$$\dot{\mu} + 3\dot{p} + 2(\mu + 3p - 2\Lambda)(\log\eta)^{\cdot} = 0, \tag{2.8}$$

$$h^b_a \dot{q}_b = -\tfrac{1}{2}(\mu + 3p - 2\Lambda)[\dot{u}_a - (\log\eta)_{,a} - (\log\eta)^{\cdot}u_a] - (\theta/3 + (\log\eta)^{\cdot})q_a - q^b\sigma_{ba} - q^b\omega_{ba}, \tag{2.9}$$

$$\dot{\mu} - \dot{p} + \tfrac{4}{3}q^a(\dot{u}_a - (\log\eta)_{,a}) + \tfrac{2}{3}(\mu - p + 2\Lambda)\theta + \tfrac{4}{3}\pi^{ab}\sigma_{ab} = 0, \tag{2.10}$$

and Eq. (2.2).

We will also require the energy conservation equation along a fluid particle world line, which follows from Einstein's field equations:

$$\dot{\mu} = -(\mu + p)\theta - \pi_{ab}\sigma^{ab} - q^a{}_{;a} - q_a\dot{u}^a. \tag{2.11}$$

Condition (2.2) was derived directly in the decomposition of (2.6). In order to determine condition (2.3), we first obtain an expression for $(\mu + 3p - 2\Lambda)(\log\eta)^{\cdot}$ by eliminating $\dot{\mu}$ and $\dot{p}$ from (2.8). Substituting from (2.11) for $\dot{\mu}$ into (2.10) gives

$$\dot{p} = -\tfrac{1}{3}(\mu + 5p - 4\Lambda)\theta + \tfrac{1}{3}\pi^{ab}\sigma_{ab} + \tfrac{4}{3}q^a(\dot{u}_a - (\log\eta)_{,a}) - q^a{}_{;a} - q_a\dot{u}^a, \tag{2.12}$$

and using (2.11) for $\dot{\mu}$ and (2.12) for $\dot{p}$, Eq. (2.8) becomes

$$(\mu + 3p - 2\Lambda)(\log\eta)^{\cdot} = (\mu + 3p - 2\Lambda)\theta + 2q^b{}_{;b} + 2q^b(\log\eta)_{,b}. \tag{2.13}$$

Condition (2.3) is derived immediately from (2.9) and (2.13).

In order to derive condition (2.4), we observe that (2.8) may be written as

$$(\mu + 3p - 2\Lambda)^{\cdot} + 2(\mu + 3p - 2\Lambda)(\log\eta)^{\cdot} = 0. \tag{2.14}$$

If (2.13) is used to replace one of the terms $(\mu + 3p - 2\Lambda)(\log\eta)^{\cdot}$ in (2.14), then (2.14) becomes

$$(\mu + 3p - 2\Lambda)_{,a}\eta u^a + (\mu + 3P - 2\Lambda) \times (\eta_{,a}u^a + \eta u^a{}_{;a}) + 2(\eta q^a)_{;a} = 0, \tag{2.15}$$

from which (2.4) follows directly.

Conditions (2.2)–(2.4) are therefore necessary conditions if ηu^a is an RCV.

Conversely, suppose that conditions (2.2)–(2.4) are satisfied. Then if (2.2) for $\dot{\pi}_{ab}$ and (2.3) for $\dot{q}^a$ are substituted into (2.6), and (2.7) is used to expand $u_{a;b}$ and $u_{[;b}$, (2.6) becomes

$$\begin{aligned}\mathscr{L}_{\eta u} R_{ab} = \tfrac{1}{2}\eta [(\dot{\mu} + 3\dot{p} + 2(\mu + 3p - 2\Lambda)\theta + 4q^c{}_{;c} \\ + 4q^c(\log \eta)_{,c})u_a u_b \\ + (\dot{\mu} - \dot{p} + \tfrac{2}{3}(\mu - p + 2\Lambda)\theta \\ + \tfrac{4}{3}q^c(\dot{u}_c - (\log \eta)_{,c}) + \tfrac{4}{3}\pi^{cd}\sigma_{cd})h_{ab}].\end{aligned} \tag{2.16}$$

Now, $\dot{\mu}$ is given by the energy conservation equation (2.11). In order to obtain an expression for $\dot{p}$, we first observe that (2.4) can be expanded as

$$\dot{\mu} + 3\dot{p} + (\mu + 3p - 2\Lambda)\theta + (\mu + 3p - 2\Lambda)(\log \eta)^{\cdot} + 2q^c{}_{;c} + 2(\log \eta)_{,c}q^c = 0. \tag{2.17}$$

But, by contracting (2.3) with u^a, (2.13) is again obtained and by eliminating $(\mu + 3p - 2\Lambda)(\log \eta)^{\cdot}$ from (2.17), we find that

$$\dot{\mu} + 3\dot{p} + 2(\mu + 3p - 2\Lambda)\theta + 4q^c{}_{;c} + 4q^c(\log \eta)_{,c} = 0. \tag{2.18}$$

On substituting from (2.11) for $\dot{\mu}$ into (2.18), Eq. (2.12) for $\dot{p}$ is again derived. By using (2.11) for $\dot{\mu}$ and (2.12) for $\dot{p}$ it is easily verified that the coefficients of $u_a u_b$ and h_{ab} in (2.16) vanish and therefore ηu^a is an RCV. ∎

It is easily verified that condition (2.4) is the conservation law (1.6) with $v_b = \eta u_b$. Conditions (2.2) and (2.3) may be regarded as propagation equations for q^a and π_{ab} along a fluid particle world line.

B. Perfect fluid space-times

The following result of Oliver and Davis[2] for a perfect fluid space-time is obtained directly from Theorem 2.1 by setting $\pi_{ab} = 0$ and $q^a = 0$.

Theorem 2.2: If Einstein's field equations (1.5) are satisfied, then a perfect fluid space-time admits an RCV, $\eta^a = \eta u^a$, if and only if

(i) $(\mu - p + 2\Lambda)\sigma_{ab} = 0,$ (2.19)

(ii) $(\mu + 3p - 2\Lambda)(\dot{u}_a - (\log \eta)_{,a} - \theta u_a) = 0,$ (2.20)

(iii) $[(\mu + 3p - 2\Lambda)\eta u^a]_{;a} = 0.$ (2.21)

∎

Conditions (2.19) and (2.20) may be rewritten alternatively as

$$\text{either } \mu - p + 2\Lambda = 0 \text{ or } \sigma_{ab} = 0, \tag{2.22a,b}$$

$$\text{either } \mu + 3p - 2\Lambda = 0 \text{ or } \dot{u}_a = (\log \eta)_{,a} + \theta u_a. \tag{2.23a,b}$$

Although a conformal Killing vector (CKV) is not necessarily an RCV, a special conformal Killing vector (SCKV), v^a, is defined by

$$\mathscr{L}_v g_{ab} = 2\psi g_{ab}, \quad \psi_{;ab} = 0, \tag{2.24}$$

is an RCV (Ref. 10). The necessary and sufficient conditions for a fluid space-time to admit a CKV, $\eta^a = \eta u^a$, are[1,11,12]

$$\sigma_{ab} = 0, \tag{2.25}$$

$$\dot{u}_a = (\log \eta)_{,a} + (\theta/3)u_a, \tag{2.26}$$

and the conformal factor ψ satisfies

$$\psi = \eta\theta/3. \tag{2.27}$$

Conditions (2.25) and (2.26) are purely kinematic. If a perfect fluid space-time can admit an SCKV parallel to u^a and ηu^a is both an SCKV and RCV then it may appear that (2.23b) and (2.26) are inconsistent because the factor multiplying θ in each equation is different. This is not the case. For, with the aid of Einstein's field equations it can be shown[13] that if a perfect fluid space-time admits an SCKV, ηu^a, then

$$\text{either } \theta = 0 \text{ or } \mu + 3p - 2\Lambda = 0. \tag{2.28}$$

When $\theta = 0$, (2.23b) and (2.26) agree. When $\theta \neq 0$, the SCKV belongs to the subset of RCVs for which $\mu + 3p - 2\Lambda = 0$ and (2.23b) does not apply.

A material curve in a fluid is a curve that always consists of the same fluid particles and therefore it moves with the fluid as the fluid evolves; it is sometimes said to be "frozen-in" to the fluid.

Theorem 2.3: Vortex lines are material lines in a perfect fluid space-time that admits an RCV, $\eta^a = \eta u^a$, if $\mu + 3p - 2\Lambda \neq 0$ and also if $\mu + 3p - 2\Lambda = 0$ provided $\mu + \Lambda \neq 0$.

Proof: It follows from (2.23) that if $\mu + 3p - 2\Lambda \neq 0$ then

$$\dot{u}_a = -(\log(1/\eta))_{,b} h^b_{\ a}, \tag{2.29}$$

and therefore η^{-1} is an acceleration potential. If $\mu + 3p - 2\Lambda = 0$ and $\mu + \Lambda \neq 0$ then, since the fluid is a perfect fluid,[8]

$$r = \exp\left(\int_{p_0}^{p} \frac{dp}{p + \mu}\right) \propto (\mu + \Lambda)^{-1/2} \tag{2.30}$$

is an acceleration potential. Ellis[8] has shown that, in a rotational fluid which admits an acceleration potential, vortex lines are material lines. ∎

This result is of interest because it shows that vortex lines may be material lines in a perfect fluid even though the fluid does not possess an equation of state of the form $p = p(\mu)$; the vortex lines are material lines because of a symmetry property of the flow.

The following theorem is due to Oliver and Davis,[2] but generalized to include a nonzero cosmological constant. It follows directly from Theorem 2.2 by verifying with the aid of the energy and momentum conservation equations for a perfect fluid,

$$\dot{\mu} + (\mu + p)\theta = 0, \tag{2.31}$$

$$(\mu + p)\dot{u}_a = -p_{,b} h^b_{\ a}, \tag{2.32}$$

that conditions (2.20)–(2.22) are satisfied. We state the theorem here in order to compare it with the corresponding result derived in Sec. III for an RCV orthogonal to u^a.

Theorem 2.4: Consider a perfect fluid space-time. If Einstein's field equations are satisfied and if the equation of state of the fluid is

$$p = \mu + 2\Lambda, \quad \mu + \Lambda \neq 0, \tag{2.33}$$

then

$$\eta^a = u^a/(\mu + \Lambda)^{1/2} \tag{2.34}$$

is an RCV. ∎

The next theorem is similar to that established by Oliver and Davis[1] for a CKV without the restrictions $\mu + 3p - 2\Lambda \neq 0$ and $p \neq \mu + 2\Lambda$.

Theorem 2.5: Consider a perfect fluid space-time such that $\mu + 3p - 2\Lambda \neq 0$ and $p \neq \mu + 2\Lambda$. If Einstein's field equations are satisfied and $\eta^a = \eta u^a$ is an RCV and

$$\dot{u}^a = 0, \tag{2.35}$$

then

either (i) $\theta = 0$ and u^a is a Killing vector (KV),

or (ii) $\theta \neq 0$ and $\omega = 0$ and $\theta_{,b} h^b_{\ a} = 0$.

Proof: Suppose first that $\theta = 0$. Since $p \neq \mu + 2\Lambda$ we have $\sigma_{ab} = 0$ and since $\dot{u}_a = 0$ a direct calculation using (2.7) gives

$$u_{(a;b)} = (\theta/3) h_{ab} = 0. \tag{2.36}$$

Hence, u^a is a KV.

Second, suppose that $\theta \neq 0$. Since $\mu + 3p - 2\Lambda \neq 0$ and $\dot{u}^a = 0$ it follows from (2.23) that

$$\theta u_a = -(\log \eta)_{,a} \tag{2.37}$$

and therefore

$$u_{[a} \theta_{,b]} + \theta u_{[a;b]} = 0. \tag{2.38}$$

Projecting on (2.38) with $h^a_{\ r} h^b_{\ s}$ gives

$$\theta \omega_{rs} = 0, \tag{2.39}$$

and since $\theta \neq 0$ it follows that $\omega = 0$. Projecting on (2.38) with $u^a h^{bc}$ and noting that $\dot{u}_a = 0$ gives

$$\theta_{,b} h^{bc} = 0, \tag{2.40}$$

which establishes the theorem. ∎

The foregoing result is another example of a shear-free perfect fluid for which it is necessary that $\omega\theta = 0$.

We outline the proof of the following theorem due to Oliver and Davis[2] and then consider the solution of the differential equation (2.41) below for the equation of state $p = p(\mu)$. It is an extension of the result (2.27) and (2.28) that if a perfect fluid space-time can admit an SCKV parallel to u^a then the SCKV is necessarily a KV unless $\mu + 3p - 2\Lambda = 0$. A corresponding result for a spacelike RCV orthogonal to u^a will be obtained in Sec. III.

Theorem 2.6 (Oliver and Davis[2]): If Einstein's field equations are satisfied and if a perfect fluid space-time with equation of state $p = p(\mu)$ admits an RCV, $\eta^a = \eta u^a$, then either η^a degenerates to a KV or

$$(\mu + p) \frac{dp}{d\mu} = \frac{1}{3} (\mu + 5p - 4\Lambda). \tag{2.41}$$

Proof: We suppose that

$$(\mu + p) \frac{dp}{d\mu} \neq \frac{1}{3} (\mu + 5p - 4\Lambda) \tag{2.42}$$

and show that $\eta_{(a;b)} = 0$. Since $p = \mu + 2\Lambda$ and $p = -\frac{1}{3}\mu + \frac{2}{3}\Lambda$ are particular solutions of (2.41) it follows that if (2.42) holds, then $\mu - p + 2\Lambda \neq 0$ and $\mu + 3p - 2\Lambda \neq 0$. Hence (2.22b) and (2.23b) are satisfied and a direct calculation with the aid of (2.22b) and (2.23b) gives

$$(\eta u_{(a})_{;b)} = (\eta/3)(h_{ab} - 3u_a u_b)\theta. \tag{2.43}$$

We must now obtain θ which can be derived from (2.21), (2.23b) contracted with u^a, the energy conservation equation for a perfect fluid (2.31), and the equation of state $p = p(\mu)$ covariantly differentiated along a fluid particle world line:

$$\dot{\mu} + 3\dot{p} + (\mu + 3p - 2\Lambda)(\theta + (\log \eta)^{\cdot}) = 0, \tag{2.44}$$

$$(\log \eta)^{\cdot} = \theta, \tag{2.45}$$

$$\dot{\mu} + (\mu + p)\theta = 0, \tag{2.46}$$

$$\dot{p} = \frac{dp}{d\mu} \dot{\mu}. \tag{2.47}$$

By eliminating $\dot{\mu}$, $\dot{p}$, and $(\log \eta)^{\cdot}$ from (2.44)–(2.47), we obtain

$$\left[(\mu + p) \frac{dp}{d\mu} - \frac{1}{3} (\mu + 5p - 4\Lambda)\right]\theta = 0, \tag{2.48}$$

and (2.42) implies that $\theta = 0$. Hence, from (2.43), ηu^a is a KV, which establishes the theorem. ∎

We have observed that $p = \mu + 2\Lambda$ and $p = -\frac{1}{3}\mu + \frac{2}{3}\Lambda$ are particular solutions of the differential equation (2.41). Consider now the general solution of (2.41). For simplicity we will take $\Lambda = 0$; (2.41) then reduces to

$$\frac{dp}{d\mu} = \frac{\mu + 5p}{3(\mu + p)}. \tag{2.49}$$

If $\Lambda \neq 0$, the change of variables

$$\bar{p} = p - \Lambda, \quad \bar{\mu} = \mu + \Lambda \tag{2.50}$$

reduces (2.41) to the homogeneous form (2.49) in $\bar{p}$ and $\bar{\mu}$ and the following method of solution would also apply. The right-hand side of (2.49) is a homogeneous function of degree 0 in p and μ and we therefore make the standard transformation from (p,μ) to (v,μ) where v is defined by

$$p = v\mu. \tag{2.51}$$

Equation (2.49) becomes

$$\mu \frac{dv}{d\mu} = \frac{(3v + 1)(1 - v)}{3(1 + v)}. \tag{2.52}$$

The variables are separable in (2.52); its solution is

$$(1 + 3v)/(1 - v)^3 \mu^2 = 1/\alpha, \tag{2.53}$$

where α is a constant. The solution (2.53) was derived by Oliver and Davis.[2] On transforming back to (p,μ), (2.52) becomes

$$(3p + \mu)/(\mu - p)^3 = 1/\alpha. \tag{2.54}$$

We will assume that $\mu > 0$. Consider first solutions with $\alpha < 0$. From (2.54), if $\alpha < 0$ then either $p > \mu$ of $p < -\mu/3$. If $p > \mu$, then by rewriting (2.49) as

$$\frac{dp}{d\mu} = 1 + \frac{2(p - \mu)}{3(p + \mu)}, \tag{2.55}$$

we see that $1 < dp/d\mu \leqslant \frac{5}{3}$. If $-\mu < p < -\mu/3$, then by rewriting (2.49) as

$$\frac{dp}{d\mu} = -\frac{1}{3} + \frac{2(p + \frac{1}{3}\mu)}{p + \mu}, \tag{2.56}$$

we see that $-\infty < dp/d\mu < -\frac{1}{3}$ and if $-\infty < p < -\mu$ then by rewriting (2.49) as

$$\frac{dp}{d\mu} = \frac{5}{3} - \frac{4\mu}{3(\mu+p)} \tag{2.57}$$

we see that $\frac{5}{3} < dp/d\mu < \infty$. Thus if $\alpha < 0$, either $dp/d\mu > 1$ and therefore the speed of sound relative to the fluid exceeds the speed of light or $dp/d\mu < -\frac{1}{3}$ and the fluid is unstable against mechanical perturbations.[14] We will therefore not consider further solutions with $\alpha < 0$.

Consider next solutions with $\alpha \geqslant 0$. This includes the particular solutions $p = \mu$ and $p = -\mu/3$ that correspond to $\alpha = 0$ and $\alpha = \infty$, respectively. From (2.54), if $\alpha \geqslant 0$ then $-\mu/3 \leqslant p \leqslant \mu$. From (2.49) and (2.55) we see that for $-\mu/5 \leqslant p \leqslant \mu$ we have $0 \leqslant dp/d\mu \leqslant 1$, and from (2.49) and (2.55) it follows that for $-\mu/3 \leqslant p \leqslant -\mu/5$ we have $-\frac{1}{3} \leqslant dp/d\mu \leqslant 0$. The case $\alpha \geqslant 0$ therefore includes all solutions that satisfy the physically reasonable conditions,

$$\mu > 0, \quad p \geqslant 0, \quad 0 \leqslant \frac{dp}{d\mu} \leqslant 1, \tag{2.58}$$

as well as all solutions which satisfy simply $0 \leqslant dp/d\mu \leqslant 1$.

To obtain the solutions for $\alpha \geqslant 0$ we rewrite (2.54) as the following cubic equation for p:

$$p^3 - 3\mu p^2 + 3(\alpha + \mu^2)p + \mu(\alpha - \mu^2) = 0. \tag{2.59}$$

The transformation

$$p = P + \mu \tag{2.60}$$

takes (2.59) to the reduced form

$$P^3 + 3\alpha P + 4\alpha\mu = 0. \tag{2.61}$$

In general, for the cubic equation

$$x^3 + \beta x + \gamma = 0, \tag{2.62}$$

where β and γ are constants, the discriminant D is defined as

$$D = -4\beta^3 - 27\gamma^2. \tag{2.63}$$

If β and γ are real and if $D \geqslant 0$ then there are three real roots while if $D < 0$ there is one real root and two complex conjugate roots. For the cubic equation (2.61),

$$D = -108\alpha^2(\alpha + 4\mu^2). \tag{2.64}$$

When $\alpha = 0$, $D = 0$ and from (2.61), $P = 0$, i.e., $p = \mu$, is a multiple root of multiplicity three. When $\alpha > 0$, $D < 0$ and there is one real root and two complex conjugate roots. To obtain this real root let

$$P = 2\sqrt{\alpha}\,Q, \quad \alpha > 0, \tag{2.65}$$

and (2.61) becomes

$$4Q^3 + 3Q = -(2\mu/\sqrt{\alpha}). \tag{2.66}$$

If we let

$$Q = \sinh\phi \tag{2.67}$$

and use the identity

$$4\sinh^3\phi + 3\sinh\phi = \sinh 3\phi \tag{2.68}$$

then (2.66) reduces to

$$\sinh 3\phi = -(2\mu/\sqrt{\alpha}). \tag{2.69}$$

Hence

$$\phi = -\tfrac{1}{3}\sinh^{-1}(2\mu/\sqrt{\alpha}) \tag{2.70}$$

and by transforming back using (2.67), (2.65), and (2.60) we find that

$$p = \mu - 2\sqrt{\alpha}\sinh\left[\tfrac{1}{3}\sinh^{-1}(2\mu/\sqrt{\alpha})\right], \quad \alpha > 0. \tag{2.71}$$

Equation (2.71) can be written equivalently as

$$p = \mu - 2\sqrt{\alpha}\sinh\left[\frac{1}{3}\ln\left(\frac{2\mu}{\sqrt{\alpha}} + \left(1 + \frac{4\mu^2}{\alpha}\right)^{1/2}\right)\right], \quad \alpha > 0. \tag{2.72}$$

We now consider some properties of the equation of state (2.71). With the aid of the series expansions[15]

$$\sinh^{-1}x = x - x^3/6 + O(x^5) \text{ as } x \to 0, \tag{2.73}$$

$$\sinh x = x + x^3/6 + O(x^5) \text{ as } x \to 0, \tag{2.74}$$

it can be verified that

$$p = -(\mu/3) + 8(\mu^3/\alpha) + O(\mu^5) \text{ as } \mu \to 0, \ \alpha > 0. \tag{2.75}$$

The asymptotic behavior of p for large μ is most easily determined from (2.72). For large μ,

$$p \doteq \mu - 2\sqrt{\alpha}\sinh\left[\ln((4\mu/\sqrt{\alpha})^{1/3})\right], \tag{2.76}$$

and therefore for large μ,

$$p \doteq \mu\left[1 - (2\sqrt{\alpha}/\mu)^{2/3}\right]. \tag{2.77}$$

Hence,

$$p = \mu + O(\mu^{1/3}) \text{ as } \mu \to \infty. \tag{2.78}$$

The derivative $dp/d\mu$ is most easily calculated using (2.49) with p given by (2.71):

$$\frac{dp}{d\mu} = \frac{3\mu - 5\sqrt{\alpha}\sinh\left[\frac{1}{3}\sinh^{-1}(2\mu/\sqrt{\alpha})\right]}{3(\mu - \sqrt{\alpha}\sinh\left[\frac{1}{3}\sinh^{-1}(2\mu/\sqrt{\alpha})\right])}. \tag{2.79}$$

Asymptotic expressions for $dp/d\mu$ can be determined using (2.75) and (2.78). It is easily verified that

$$\left.\frac{dp}{d\mu}\right|_{\mu=0} = -\frac{1}{3}, \tag{2.80}$$

which is independent of α ($\alpha > 0$) and

$$\frac{dp}{d\mu} = 1 + O(\mu^{-2/3}) \text{ as } \mu \to \infty. \tag{2.81}$$

Graphs of p plotted against μ for a selection of values of α in the range $0 \leqslant \alpha \leqslant \infty$ are presented in Fig. 1. The family of curves is bounded by the straight-line graphs $p = \mu$ ($\alpha = 0$) and $p = -\mu/3$ ($\alpha = \infty$). The pressure p decreases from zero at $\mu = 0$ and is negative for sufficiently small values of μ for each $\alpha > 0$ in agreement with (2.75) and (2.80). The pressure eventually increases with increasing μ and becomes positive for sufficiently large μ for each $0 \leqslant \alpha < \infty$.

Graphs of $dp/d\mu$ plotted against μ for the same values of α as used in Fig. 1 are presented in Fig. 2. The family of curves is bounded by the straight lines $dp/d\mu = 1$ ($\alpha = 0$) and $dp/d\mu = -\frac{1}{3}$ ($\alpha = \infty$). For each $0 < \alpha < \infty$, $dp/d\mu$ increases monotonically from $-\frac{1}{3}$ at $\mu = 0$ to $+1$ at $\mu = \infty$, consistent with (2.80) and (2.81). For sufficiently small values of μ, $dp/d\mu < 0$ for each $\alpha > 0$ and the fluid is unstable to mechanical perturbations. However, for sufficiently large values of μ, $0 < dp/d\mu \leqslant 1$ for each $0 \leqslant \alpha < \infty$; the fluid is stable against mechanical perturbations and the speed of sound

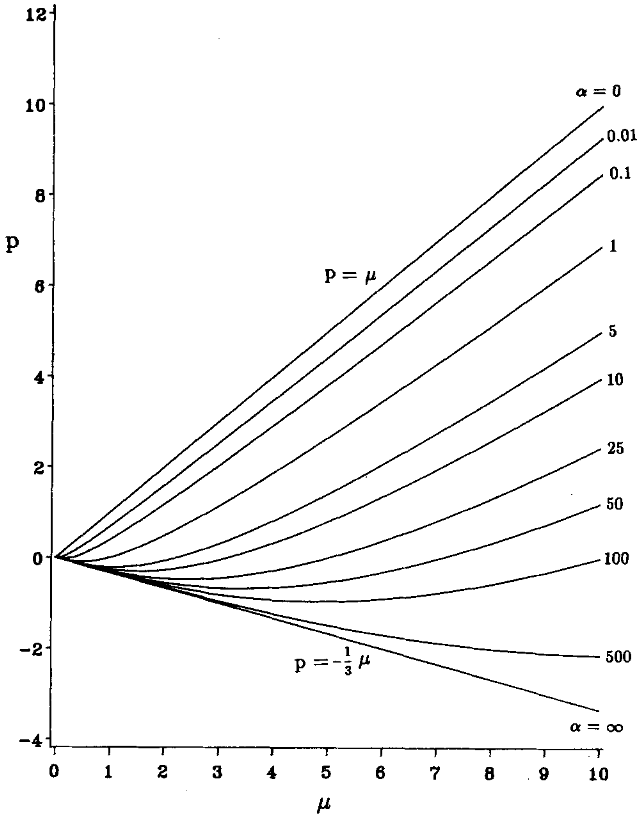


FIG. 1. The pressure p, given by Eq. (2.71), plotted against μ for $\alpha = 0$, 0.01, 0.1, 1, 5, 10, 25, 50, 100, 500, and ∞. The straight-line graphs, $\alpha = 0$ and $\alpha = \infty$, correspond to the equations of state $p = \mu$ and $p = -\mu/3$, respectively.

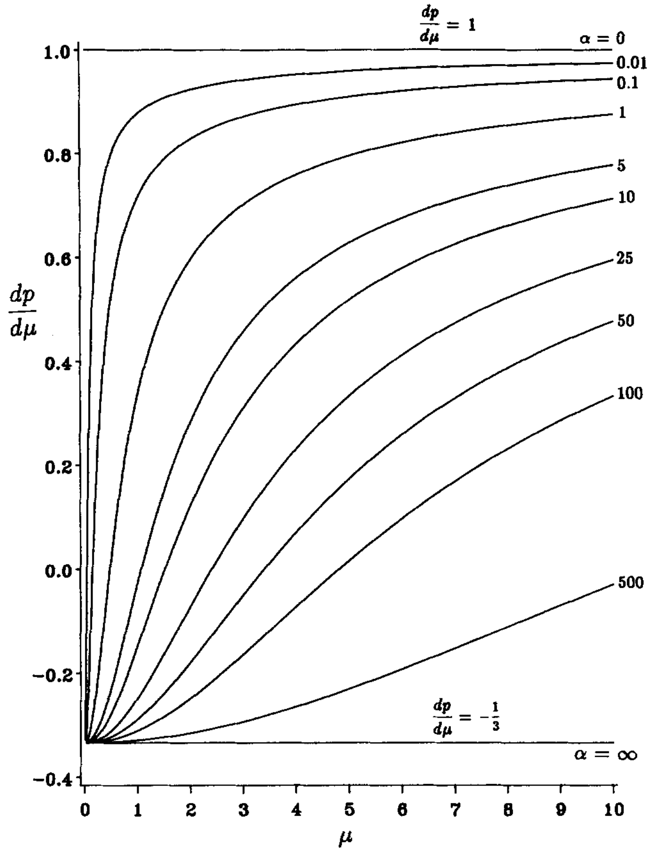


FIG. 2. The pressure gradient $dp/d\mu$, given by Eq. (2.79), plotted against μ for $\alpha = 0$, 0.01, 0.1, 1, 5, 10, 25, 50, 100, 500, and ∞. The straight-line graphs, $\alpha = 0$ and $\alpha = \infty$, correspond to the equations of state $p = \mu$ and $p = -\mu/3$, respectively.

relative to the fluid, $(dp/d\mu)^{1/2}$, does not exceed the speed of light.

The curves $p = 0$ and $dp/d\mu = 0$ in the (α,μ) plane for $\alpha > 0$ and $\mu > 0$ are plotted in Fig. 3. For a given value of $\alpha > 0$, the values of μ for which $p > 0$ and $dp/d\mu > 0$ can be determined from Fig. 3. We conclude that the only solution of (2.49) which satisfies the physically reasonable conditions (2.58) for all values of $\mu > 0$ is $p = \mu$ corresponding to $\alpha = 0$, although all solutions with $0 < \alpha < \infty$ eventually satisfy conditions (2.58) for sufficiently large μ.

C. Fluid space-times with $q^a=0$ but $\pi^{ab}\neq 0$

Consideration of a fluid space-time with $q^a = 0$ but $\pi^{ab} \neq 0$ is equivalent to taking $u^a = u^a_D$ where u^a_D is the dynamic four-velocity. The following theorem follows directly from Theorem 2.1.

Theorem 2.7: If Einstein's field equations (1.5) are satisfied then a fluid space-time with energy-momentum tensor (2.1), with $q^a = 0$ but $\pi^{ab} \neq 0$, where q^a and π^{ab} are measured relative to the four velocity u^a, admits an RCV, $\eta^a = \eta u^a$, if and only if

$$\text{(i)}\quad h^c_a h^d_b \dot{\pi}_{cd} = -(\mu - p + 2\Lambda)\sigma_{ab} + \tfrac{1}{3}(\pi^{cd}\sigma_{cd})h_{ab} - \tfrac{2}{3}\theta\pi_{ab} - 2\sigma_{c(a}\pi_{b)}{}^{c} - 2\omega_{c(a}\pi_{b)}{}^{c}, \tag{2.82}$$

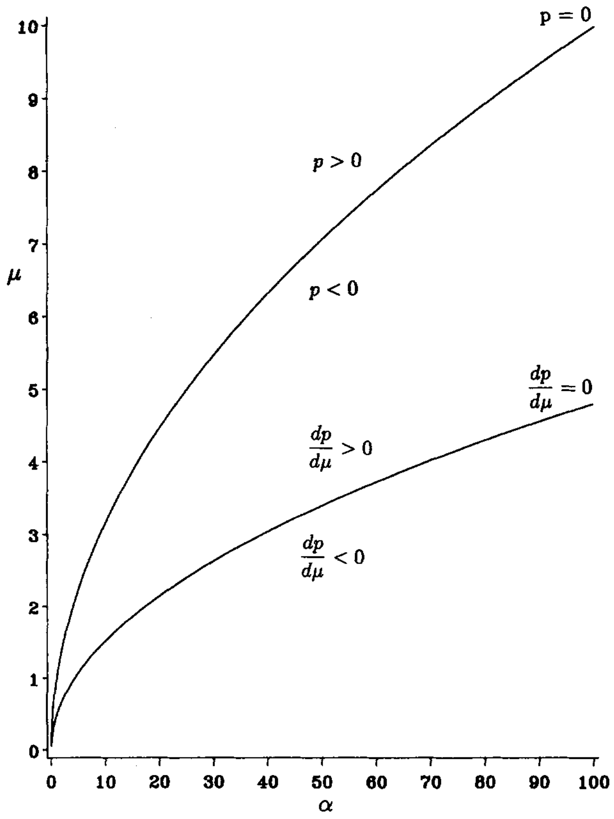


FIG. 3. The curves $p = 0$ and $dp/d\mu = 0$ in the (α,μ) plane for $\alpha > 0$ and $\mu > 0$, where p and $dp/d\mu$ are given by Eqs. (2.71) and (2.79), respectively.

(ii) $(\mu+3p-2\Lambda)[\dot{u}_a-(\log\eta)_{,a}-\theta u_a]=0,$ (2.83)

(iii) $[(\mu+3p-2\Lambda)\eta u^a]_{;a}=0,$ (2.84)

where θ, σ_{ab}, and ω_{ab} are the rate of expansion, the rate-of-shear tensor and the vorticity tensor of the timelike congruence generated by u^a. ■

Equation (2.83) can be written alternatively as (2.23a,b). A nonzero π_{ab} enters into (2.82) to (2.84) only through condition (2.82); (2.83) and (2.84) are the same as for a perfect fluid space-time. If $\mu+3p-2\Lambda\neq 0$ then η^{-1} is an acceleration potential and if the fluid is rotational then vortex lines are material lines in the fluid. This is a consequence of the symmetry of the flow and is not due to a physical property of the fluid; in general a fluid with $\pi_{ab}\neq 0$ [or even a perfect fluid if $p\neq p(\mu)$] does not admit an acceleration potential and the vortex lines are not generally material lines.

The phenomenological equation of state

$$\pi_{ab}=-\lambda\sigma_{ab},\quad \lambda\geqslant 0, \tag{2.85}$$

where λ is the coefficient of shear viscosity, is necessary if the rate of entropy production is never negative.[7,8] We now establish the following result which in some ways corresponds to Theorem 2.6 for a perfect fluid.

Theorem 2.8: If Einstein's field equations are satisfied and the fluid space-time admits an RCV, $\eta^a=\eta u^a$, and if

$$p=p(\mu)\ \text{ but }\ \frac{dp}{d\mu}\neq-\frac{1}{3}, \tag{2.86}$$

$$q^a=0, \tag{2.87}$$

$$\pi_{ab}=-\lambda\sigma_{ab},\quad \lambda>0, \tag{2.88}$$

where q^a and π_{ab} are measured relative to u^a and

$$\text{either }\ 3(\mu+p)\frac{dp}{d\mu}=\mu+5p-2\Lambda \tag{2.89}$$

$$\left(\text{excluding }\frac{dp}{d\mu}=-\frac{1}{3}\right)$$

$$\text{or }\ \theta=0, \tag{2.90}$$

then $\sigma_{ab}=0$ and the fluid has a perfect fluid energy-momentum tensor. For the case $\theta=0$, the RCV reduces to a KV.

Proof: Since it is assumed that $dp/d\mu\neq-\frac{1}{3}$ it follows that $\mu+3p-2\Lambda\neq 0$ and therefore (2.23b) is satisfied. Equations (2.44), (2.45), and (2.47) again hold but in place of (2.46) the energy conservation equation now takes the form

$$\dot{\mu}+(\mu+p)\theta+\pi_{ab}\sigma^{ab}=0. \tag{2.91}$$

By eliminating $\dot{\mu}$, $\dot{p}$, and $(\log\eta)^{\cdot}$ from (2.44), (2.45), (2.47), and (2.91) we obtain

$$\left[\mu+5p-4\Lambda-3(\mu+p)\frac{dp}{d\mu}\right]\theta=\left(1+3\frac{dp}{d\mu}\right)\pi_{ab}\sigma^{ab}, \tag{2.92}$$

and since $\pi_{ab}=-\lambda\sigma_{ab}$ it follows that

$$\left[\mu+5p-4\Lambda-3(\mu+p)\frac{dp}{d\mu}\right]\theta=-2\lambda\sigma^2\left(1+3\frac{dp}{d\mu}\right), \tag{2.93}$$

where $\sigma^2=\frac{1}{2}\sigma_{ab}\sigma^{ab}$. Now, $\lambda\neq 0$ and $dp/d\mu\neq-\frac{1}{3}$ and therefore if either (2.89) or (2.90) is satisfied it follows from (2.93) that $\sigma^2=0$ and hence $\sigma_{ab}=0$.

When $\sigma_{ab}=0$ and (2.23b) is satisfied, (2.43) is again valid and therefore if $\theta=0$ the RCV reduces to a KV. ■

An equation of state that satisfies (2.86) and (2.89) is $p=\mu+2\Lambda$. If a fluid space-time with equation of state $p=\mu+2\Lambda$ admits an RCV, $\eta^a=\eta u^a$, and if the fluid is viscous ($\lambda>0$) and $q^a=0$ then it will be shear-free and have a perfect fluid energy-momentum tensor. In comparison, from (2.19), a perfect fluid space-time which admits an RCV, $\eta^a=\eta u^a$, need not necessarily be shear free when $p=\mu+2\Lambda$.

III. SPACELIKE RICCI COLLINEATION VECTORS ORTHOGONAL TO u^a: PERFECT FLUID SPACE-TIMES

We now consider the properties of fluid space-times that admit a spacelike RCV, ξ^a, orthogonal to u^a:

$$\xi^a=\xi n^a,\quad n_a n^a=+1,\quad n_a u^a=0. \tag{3.1}$$

For an imperfect fluid, the theory of spacelike RCVs orthogonal to u^a is more complex than that for timelike RCVs parallel to u^a, one reason being that whereas $q_a u^a=0$ and $\pi_{ab}u^b=0$, in general $q_a n^a\neq 0$ and $\pi_{ab}n^b\neq 0$. Hence, instead of first considering the most general case of an imperfect fluid and then specializing to a perfect fluid, as we did for timelike RCVs in Sec. II, we will consider first perfect fluid space-times in this section and then consider the more complex case of imperfect fluid space-times in Sec. IV.

We will express the necessary and sufficient conditions for a fluid space-time to admit an RCV parallel to n^a in terms of the rotation, expansion, and shear of the spacelike congruence generated by n^a. To measure the deformation of the congruence generated by n^a at any given point P an observer with four-velocity w^a orthogonal to n^a at P must be specified. Since $n_a u^a=0$, an observer comoving with the fluid with four-velocity u^a may be employed at P and in the subsequent theory a comoving observer, u^a, will always be used. Once the observer has been specified at any one point of the congruence, the observers employed at all other points along the congruence cannot be arbitrarily assigned; their four-velocities must satisfy a transport law derived by Greenberg.[3-5] It follows from the Greenberg transport law that if a comoving observer with 4-velocity u^a is chosen at any one given point P then the observers employed at all other points along the congruence can be comoving observers with four-velocity u^a if and only if

$$h^a_b\overset{*}{n}{}^b=\overset{*}{u}{}^a-(n_b\overset{*}{u}{}^b)n^a, \tag{3.2}$$

where an overhead star denotes covariant differentiation along an integral curve of n^a; for example,

$$\overset{*}{A}{}^a=A^a{}_{;b}n^b. \tag{3.3}$$

In the following theory it will not be required to employ comoving observers all along the congruence and (3.2) need not hold; only the observer at the given point P will be comoving. It can be shown that (3.2) is the necessary and sufficient condition for the integral curves of $n^a(n_a u^a=0,$

$n_a n^a = +1$) to be material curves in the fluid[4,5] and therefore comoving observers u^a can be employed all along the n-congruence if and only if the curves of the congruence are material curves. For the following theory it will be convenient to define

$$N^a = h^a_b \dot{n}^b - \overset{*}{\dot{u}}{}^a + (n_b \overset{*}{\dot{u}}{}^b) n^a. \tag{3.4}$$

The rotation tensor $\mathscr{R}_{ab}$, the expansion $\mathscr{E}$, and the shear tensor $\mathscr{S}_{ab}$ of the spacelike congruence generated by n^a as measured by an observer with four-velocity u^a are

$$\mathscr{R}_{ab} = p^c_a p^d_b n_{[c;d]}, \tag{3.5}$$

$$\mathscr{E} = p^{ab} n_{a;b}, \tag{3.6}$$

$$\mathscr{S}_{ab} = p^c_a p^d_b n_{(c;d)} - \tfrac{1}{2}\mathscr{E} p_{ab}, \tag{3.7}$$

where p^{ab} is the projection tensor that projects onto the two-space orthogonal to u^a and n^a:

$$p^{ab} = g^{ab} + u^a u^b - n^a n^b; \quad p^{ab} u_b = 0, \quad p^{ab} n_b = 0. \tag{3.8}$$

The covariant derivative of n_a can be decomposed as

$$n_{a;b} = \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab} + \overset{*}{n}_a n_b - \dot{n}_a n_b + u_a (n^t u_{t;b}) + (n^t \dot{u}_t) u_a u_b - (n_t \overset{*}{\dot{u}}{}^t) u_a n_b, \tag{3.9}$$

where $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ are measured by an observer with four-velocity u^a.

The following theorem corresponds directly with Theorem 2.2.

Theorem 3.1: If Einstein's field equations (1.5) are satisfied, then a perfect fluid space-time admits an RCV, $\xi^a = \xi n^a (n_a n^a = +1, n_a u^a = 0)$ if and only if

(i) $(\mu + 3P - 2\Lambda)\omega_{at} n^t = \tfrac{1}{2}(\mu - p + 2\Lambda) N_a$, (3.10)

(ii) $(\mu - p + 2\Lambda)\mathscr{S}_{ab} = 0$, (3.11)

(iii) $(\mu - p + 2\Lambda)[\overset{*}{n}_a + (\log \xi)_{,a} - \tfrac{1}{2}\mathscr{E} n_a] = 0$, (3.12)

(iv) $(\mu - p + 2\Lambda)(\tfrac{1}{2}\mathscr{E} + n_t \dot{u}^t) = 0$, (3.13)

(v) $[(\mu - p + 2\Lambda)\xi n^a]_{;a} = 0$, (3.14)

where N^a is defined by (3.4) and $\mathscr{E}$ and $\mathscr{S}_{ab}$ are the expansion and shear of the spacelike congruence generated by n^a as measured by an observer with the fluid unit four-velocity u^a.

Proof: From the definition of the Lie derivative,

$$\mathscr{L}_{\xi n} R_{ab} = \xi\,[\overset{*}{R}_{ab} + 2n^c R_{c(a} (\log \xi)_{,b)} + 2R_{c(a} n^c{}_{;b)}]. \tag{3.15}$$

With the aid of Einstein's field equations for a perfect fluid, (3.15) may be rewritten as

$$\mathscr{L}_{\xi n} R_{ab} = \xi\,[\tfrac{1}{2}(\overset{*}{\mu} + 3\overset{*}{p}) + \tfrac{1}{2}(\overset{*}{\mu} - \overset{*}{p}) h_{ab} + 2(\mu + p)(\overset{*}{\mu}_{(a} u_{b)} - u_{(a} u^t{}_{;b)} n_t) + (\mu - p + 2\Lambda)(n_{(a;b)} + n_{(a} (\log \xi)_{,b)})]. \tag{3.16}$$

Suppose first that ξn^a is an RCV. Then (1.1) is satisfied. The right-hand side of (3.16) is therefore zero and by contracting it in turn with $u^a u^b$, $u^a n^b$, $u^a p^{bc}$, $n^a n^b$, $n^a p^{bc}$, p^{ab}, and $p^{ac} p^{bd} - \tfrac{1}{2} p^{ab} p^{cd}$ the following seven equations are derived:

$$\overset{*}{\mu} + 3\overset{*}{p} + 2(\mu + 3p - 2\Lambda) n_a \dot{u}^a = 0, \tag{3.17}$$

$$(\mu - p + 2\Lambda)((\log \xi)^{\cdot} + u_a \overset{*}{n}{}^a) = 0, \tag{3.18}$$

$$\tfrac{1}{2}(\mu - p + 2\Lambda) h^b_a \dot{n}_b - (\mu + p)(\overset{*}{\dot{u}}_a - n_b \overset{*}{\dot{u}}{}^b) n_a) + \tfrac{1}{2}(\mu + 3p - 2\Lambda) p^b_a n^t u_{t;b} = 0, \tag{3.19}$$

$$\overset{*}{\mu} - \overset{*}{p} + 2(\mu - p + 2\Lambda)(\log \xi)^* = 0, \tag{3.20}$$

$$(\mu - p + 2\Lambda) p^b_a [\overset{*}{n}_b + (\log \xi)_{,b}] = 0, \tag{3.21}$$

$$\overset{*}{\mu} - \overset{*}{p} + (\mu - p + 2\Lambda)\mathscr{E} = 0, \tag{3.22}$$

$$(\mu - p + 2\Lambda)\mathscr{S}_{ab} = 0. \tag{3.23}$$

We will also require the momentum conservation equation for a perfect fluid, (2.32), contracted with n^a:

$$\overset{*}{p} + (\mu + p) n_a \dot{u}^a = 0. \tag{3.24}$$

The momentum conservation equation (2.32) followed from Einstein's field equations.

(i) Condition (3.10) is derived from (3.19). We have

$$n^t u_{t;b} = 2n^t u_{[t;b]} + \overset{*}{u}_b = -2\omega_{bt} n^t - (n_t \dot{u}^t) u_b + \overset{*}{u}_b, \tag{3.25}$$

and by substituting from (3.25) into (3.19), (3.10) follows directly.

(ii) Condition (3.11) is given by (3.23).

(iii) To derive (3.12), we first expand (3.21) and use (3.18); this gives

$$(\mu - p + 2\Lambda)[\overset{*}{n}_a + (\log \xi)_{,a} - (\log \xi)^* n_a] = 0. \tag{3.26}$$

But by subtracting (3.22) from (3.20) it follows that

$$(\mu - p + 2\Lambda)(\log \xi)^* = \tfrac{1}{2}(\mu - p + 2\Lambda)\mathscr{E}, \tag{3.27}$$

and by substituting from (3.27) into (3.26), (3.12) is immediately derived.

(iv) To derive (3.13), we first substitute (3.24) for $\overset{*}{p}$ into (3.17) to obtain

$$\overset{*}{\mu} = (\mu - 3p + 4\Lambda) n_a \dot{u}^a. \tag{3.28}$$

By replacing $\overset{*}{p}$ and $\overset{*}{\mu}$ in (3.22) by (3.24) and (3.28), respectively, (3.13) is obtained.

(v) Consider the final condition (3.14). Substitute (3.24) and (3.28) into (3.20); this gives

$$(\mu - p + 2\Lambda)(\log \xi)^* = -(\mu - p + 2\Lambda) n_a \dot{u}^a, \tag{3.29}$$

and subtract (3.29) from twice (3.27) to obtain

$$(\mu - p + 2\Lambda)(\log \xi)^* = (\mu - p + 2\Lambda)(\mathscr{E} + n_a \dot{u}^a). \tag{3.30}$$

But from (3.6),

$$\mathscr{E} + n_a \dot{u}^a = n^a{}_{;a}, \tag{3.31}$$

and therefore (3.30) becomes

$$(\mu - p + 2\Lambda)(\log \xi)^* = (\mu - p + 2\Lambda) n^a{}_{;a}. \tag{3.32}$$

If one of the terms $(\mu - p + 2\Lambda)(\log \xi)^*$ in (3.20) is replaced by (3.32) then (3.20) may be written as

$$(\mu - p + 2\Lambda)_{,a} \xi n^a + (\mu - p + 2\Lambda)(\xi_{,a} n^a + \xi n^a{}_{;a}) = 0, \tag{3.33}$$

from which (3.14) follows directly.

Hence, if $\xi^a = \xi n^a$ is an RCV then conditions (3.10)–(3.14) are satisfied.

Conversely, suppose that (3.10)–(3.14) are satisfied and Einstein's field equations hold.

Using (3.9) for $n_{(a;b)}$, (3.11) and (3.12) for $(\mu - p + 2\Lambda)(\log \xi)_{,a}$ Eq. (3.16) becomes

$$\begin{aligned}\mathscr{L}_{\xi n} R_{ab} = \xi \, [& \tfrac{1}{2}(\overset{*}{\mu} + 3\overset{*}{p}) u_a u_b + \tfrac{1}{2}(\overset{*}{\mu} - \overset{*}{p}) h_{ab} \\ & + \tfrac{1}{2}(\mu - p + 2\Lambda)(\mathscr{E} h_{ab} - 2u_{(a} N_{b)}) \\ & - (\mu + 3p - 2\Lambda) u_{(a} (u^t{}_{;b)} n_t - \overset{*}{u}_{b)})].\end{aligned} \tag{3.34}$$

Further, by using (3.25) for $n^t u_{t;b}$ and (3.10) for $(\mu + 3p - 2\Lambda)\omega_{at} n^t$ and by replacing $\mathscr{E}$ by $n_t \dot{u}^t$ with the aid of (3.13), (3.34) reduces to

$$\begin{aligned}\mathscr{L}_{\xi n} R_{ab} = \xi \, [& \tfrac{1}{2}(\overset{*}{\mu} + 3\overset{*}{p} + 2(\mu + 3p - 2\Lambda) n_t \dot{u}^t) u_a u_b \\ & + \tfrac{1}{2}(\overset{*}{\mu} - \overset{*}{p} - 2(\mu - p + 2\Lambda) n_t \dot{u}^t) h_{ab}].\end{aligned} \tag{3.35}$$

Now, $\overset{*}{p}$ is expressed in terms of $n_t \dot{u}^t$ through (3.24). To obtain $\overset{*}{\mu}$ in terms of $n_t \dot{u}^t$ we use the remaining condition (3.14), which may be expanded as

$$\overset{*}{\mu} - \overset{*}{p} + 2(\mu - p + 2\Lambda)(\log \xi)^* = 0. \tag{3.36}$$

But if (3.12) is contracted with n^a we obtain, with the aid of (3.13),

$$(\mu - p + 2\Lambda)(\log \xi)^* = -(\mu - p + 2\Lambda) n_t \dot{u}^t \tag{3.37}$$

and therefore (3.36) becomes

$$\overset{*}{\mu} - \overset{*}{p} - 2(\mu - p + 2\Lambda) n_t \dot{u}^t = 0. \tag{3.38}$$

By eliminating $\overset{*}{p}$ from (3.38) using (3.24), Eq. (3.28) for $\overset{*}{\mu}$ is again derived. It is easily verified with the aid of (3.24) and (3.28) that the right-hand side of (3.35) vanishes and therefore $\xi^a = \xi n^a$ is an RCV. ∎

It is easily verified that (3.14) is the conservation law (1.6) for the special case of a perfect fluid and $v_b = \xi n_b$. Conditions (3.11)–(3.13) may be written alternatively as

either $p = \mu + 2\Lambda$

$$\text{or} \begin{cases} \mathscr{S}_{ab} = 0, & (3.39\text{a,b}) \\ \overset{*}{n}_a = -(\log \xi)_{,a} + \tfrac{1}{2}\mathscr{E} n_a, & (3.40\text{a,b}) \\ n_c \dot{u}^c = -\tfrac{1}{2}\mathscr{E}. & (3.41\text{a,b}) \end{cases}$$

One of the necessary and sufficient conditions for a fluid space-time to admit a CKV, $\xi^a = \xi n^a (n_a u^a = 0, n_a n^a = +1)$, is[5]

$$n_c \dot{u}^c = +\tfrac{1}{2}\mathscr{E}. \tag{3.42}$$

Equation (3.42) is purely kinematic. If a perfect fluid space-time can admit an SCKV orthogonal to u^a and ξn^a is both an SCKV and an RCV it may appear that (3.41b) and (3.42) are inconsistent because of the difference in sign. This is not the case because it can be shown[13,16] using Einstein's field equations that if a perfect fluid space-time admits an SCKV, $\xi^a = \xi n^a$, then

$$\text{either } \xi = 0 \text{ or } p = \mu + 2\Lambda. \tag{3.43}$$

When $\mathscr{E} = 0$, (3.41b) and (3.42) agree. When $\mathscr{E} \neq 0$, the SCKV belongs to the subset of RCVs for which $p = \mu + 2\Lambda$ and therefore (3.41b) does not apply.

The following theorem corresponds to Theorem 2.3 for an RCV parallel to u^a.

Theorem 3.2: Suppose that a perfect fluid space-time admits an RCV, $\xi^a = \xi n^a (n_a n^a = +1, n_a u^a = 0)$ and that Einstein's field equations are satisfied.

(a) $\omega = 0$. Then either $p = \mu + 2\Lambda$ or the integral curves of n^a are material curves in fluid.

(b) $\omega \neq 0$.

(i) If $p = \mu + 2\Lambda$ but $\mu + \Lambda \neq 0$ then $n^a = \pm \omega^a/\omega$ and the vortex lines are material lines in the fluid.

(ii) If the integral curves of n^a are material curves and $\mu + 3p - 2\Lambda \neq 0$, then $n^a = \pm \omega^a/\omega$.

(iii) If $n^a = \pm \omega^a/\omega$, then the vortex lines are material lines if $p \neq \mu + 2\Lambda$ and also if $p = \mu + 2\Lambda$ provided $\mu + \Lambda \neq 0$.

Proof: All of the results are established from (3.10).

(a) $\omega = 0$. It follows from (3.10) that when $\omega = 0$, either $p = \mu + 2\Lambda$ or $N^a = 0$. When $N^a = 0$ the integral curves of n^a are material curves in the fluid.[4,5]

(b) $\omega \neq 0$.

(i) If $p = \mu + 2\Lambda$ then $\mu + 3p - 2\Lambda = 4(\mu + \Lambda) \neq 0$. Hence, from (3.10),

$$\omega_{at} n^t = 0, \tag{3.44}$$

and since $\omega_{at} = \eta_{atrs} \omega^r u^s$ we find by contracting (3.44) with $\eta^{abcd} \omega_c u_d$ that

$$n^a = [(\omega_t n^t)/\omega^2] \omega^a. \tag{3.45}$$

Since both $n^a \neq 0$ and $\omega^a \neq 0$ it follows that $n^a = \pm \omega^a/\omega$. Also, since the fluid is a perfect fluid and $p = \mu + 2\Lambda$,

$$r = \exp\left(\int_{p_0}^{p} \frac{dp}{p + \mu} \right) \propto (\mu + \Lambda)^{1/2} \neq 0 \tag{3.46}$$

is an acceleration potential and therefore the vortex lines are material lines in the fluid.[8]

(ii) If the integral curves of n^a are material curves then $N^a = 0$ (Refs. 4 and 5). Hence, since $\mu + 3p - 2\Lambda \neq 0$, (3.44) is again obtained from (3.10) and therefore $n^a = \pm \omega^a/\omega$.

(iii) If $n^a = \pm \omega^a/\omega$ and $p \neq \mu + 2\Lambda$ then $N^a = 0$ and the vortex lines are material lines. If $p = \mu + 2\Lambda$ and $\mu + \Lambda \neq 0$ then from (i) vortex lines are material lines. ∎

The following theorem corresponds to theorem 2.4 for an RCV parallel to u^a.

Theorem 3.3: Consider a perfect fluid space-time. If Einstein's field equations are satisfied and if the equation of state of the fluid is

$$p = \mu + 2\Lambda, \tag{3.47}$$

and if

$$\mu + \Lambda \neq 0 \tag{3.48}$$

then

(a) $\omega = 0$. Any vector orthogonal to u^a is RCV.

(b) $\omega \neq 0$. A vector orthogonal to u^a is an RCV if and only if it is parallel to ω^a.

Proof: The only condition of Theorem 3.1 not identically satisfied when $p = \mu + 2\Lambda$ is (3.10) which reduces to

$$(\mu + \Lambda) \omega_{at} n^t = 0. \tag{3.49}$$

(a) $\omega = 0$. If $\omega = 0$ then (3.49) is identically satisfied and any vector $\xi^a = \xi n^a$ orthogonal to u^a is an RCV.

(b) $\omega \neq 0$. Suppose first that $\xi n^a (n_a u^a = 0, n_a n^a = +1)$ is an RCV. Then, if $\mu + \Lambda \neq 0$ it follows from

(3.49) that $\omega_{at}n^t = 0$ and therefore (3.45) is satisfied; thus n^a is parallel to ω^a.

Conversely, suppose that $n^a = \omega^a/\omega$. Then since $\omega_{at}n^t = 0$, (3.49) is satisfied and therefore ξn^a is an RCV. ∎

As a simple example of Theorem 3.3, consider the Gödel universe which is a rotational perfect fluid space-time that satisfies[7,8]

$$p = \mu + 2\Lambda, \quad \mu + \Lambda = \omega^2 \neq 0. \tag{3.50}$$

Thus any vector parallel to ω^a is an RCV of the Gödel metric and any RCV orthogonal to u^a admitted by the Gödel universe must be parallel to ω^a. For the Gödel universe[5]

$$\omega^a = (\alpha/\sqrt{2})\delta_3^a, \quad \alpha = \text{const}. \tag{3.51}$$

Any vector of the form $\xi\delta_3^a$ is therefore an RCV of the Gödel metric. It is well known that δ_3^a is a KV of the Gödel metric which is a special case of an RCV.

We now consider perfect fluid space-times with $p \neq \mu + 2\Lambda$. The following theorem corresponds to Theorem 2.5 for an RCV parallel to u^a.

Theorem 3.3: Consider a perfect fluid space-time such that $p \neq \mu + 2\Lambda$ and $\mu + 3p - 2\Lambda \neq 0$. If Einstein's field equations are satisfied and $\xi^a = \xi n^a (n_a u^a = 0, n_a n^a = +1)$ is an RCV such that

$$\overset{*}{n}{}^a = 0, \tag{3.52}$$

and if the integral curves of n^a are material curves in the fluid, or equivalently when $\omega \neq 0$ if $n^a = \omega^a/\omega$, then

either (i) $\mathscr{E} = 0$ and n^a is a KV,

or (ii) $\mathscr{E} \neq 0$ and $\mathscr{R}_{ab} = 0$ and $\mathscr{E}_{,b}p_a^b = 0$.

Proof: (i) $\mathscr{E} = 0$. Using (3.9) for $n_{a;b}$ and (3.25) it can be verified that

$$n_{(a;b)} = \tfrac{1}{2}\mathscr{E}p_{ab} + \mathscr{S}_{ab} + \overset{*}{n}_{(a}n_{b)} - u_{(a}N_{b)} - 2u_{(a}\omega_{b)t}n^t - (n_t\dot{u}^t)u_a u_b. \tag{3.53}$$

But since $p \neq \mu + 2\Lambda$ it follows from (3.11) and (3.13) that $\mathscr{S}_{ab} = 0$ and $n_t\dot{u}^t = -\tfrac{1}{2}\mathscr{E} = 0$. Also, if the integral curves of n^a are material curves then $N^a = 0$ and from (3.10), $\omega_{at}n^t = 0$. [Alternatively, when $\omega \neq 0$, if $n^a = \omega^a/\omega$ then $\omega_{at}n^t = 0$ and from (3.10), $N^a = 0$.] Since $\mathscr{E} = 0$ and $\overset{*}{n}{}^a = 0$, (3.53) reduces to $n_{(a;b)} = 0$ and therefore n^a is a KV.

(ii) $\mathscr{E} \neq 0$. Since $p \neq \mu + 2\Lambda$ and $\overset{*}{n}{}^a = 0$, it follows from (3.12) that

$$\mathscr{E}n_a = 2(\log\xi)_{,a}, \tag{3.54}$$

and therefore

$$n_{[a}\mathscr{E}_{,b]} + \mathscr{E}n_{[a;b]} = 0. \tag{3.55}$$

Projecting on (3.55) with $p_c^a p_d^b$ gives

$$\mathscr{E}\mathscr{R}_{cd} = 0 \tag{3.56}$$

and therefore $\mathscr{R}_{ab} = 0$ since $\mathscr{E} \neq 0$. Projecting on (3.55) with $n^a p_c^b$ and using $\overset{*}{n}{}^a = 0$ yields

$$\mathscr{E}_{,b}p_c^b = 0. \quad ∎ \tag{3.57}$$

If a perfect fluid space-time can admit an SCKV orthogonal to u^a then, since[5]

$$\psi = \tfrac{1}{2}\xi\mathscr{E} \tag{3.58}$$

where ψ is the conformal factor, it follows from (3.43) that the SCKV is necessarily a KV unless $p = \mu + 2\Lambda$. We now consider the extension of this result to an RCV. The following theorem corresponds to Theorem 2.6 for an RCV parallel to u^a.

Theorem 3.4: Suppose that Einstein's field equations are satisfied. If a perfect fluid space-time with equation of state $\mu = \mu(p)$ admits an RCV, $\xi^a = \xi n^a (n_a u^a = 0, n_a n^a = +1)$ and if the integral curves of n^a are material curves in the fluid, or equivalently when $\omega \neq 0$ if $n^a = \omega^a/\omega$, then either ξ^a degenerates to a KV or

$$(\mu + p)\frac{d\mu}{dp} = 3p - \mu - 4\Lambda. \tag{3.59}$$

Proof: We suppose that

$$(\mu + p)\frac{d\mu}{dp} \neq 3p - \mu - 4\Lambda. \tag{3.60}$$

and show that $\xi_{(a;b)} = 0$. Since $\mu = p - 2\Lambda$ and $\mu = -3p + 2\Lambda$ are particular solutions of (3.59), it follows that if (3.60) holds then $\mu - p + 2\Lambda \neq 0$ and $\mu + 3p - 2\Lambda \neq 0$. Since $\mu - p + 2\Lambda \neq 0$, Eqs. (3.39b), (3.40b), and (3.41b) are satisfied. If the integral curves of n^a are material curves then $N^a = 0$ and from (3.10), since $\mu + 3p - 2\Lambda \neq 0$, we have $\omega_{at}n^t = 0$. [If, alternatively, when $\omega \neq 0$, $n^a = \omega^a/\omega$ then $\omega_{at}n^t = 0$ and from (3.10), since $\mu - p + 2\Lambda \neq 0$, we have $N^a = 0$.] Hence with the aid of (3.53) a direct calculation gives

$$(\xi n_{(a})_{;b)} = \tfrac{1}{2}\xi\mathscr{E}(h_{ab} + u_a u_b). \tag{3.61}$$

It remains to determine $\mathscr{E}$. Equation (3.14) when expanded is

$$\overset{*}{\mu} - \overset{*}{p} + (\mu - p + 2\Lambda)\big((\log\xi)^* + n^a{}_{;a}\big) = 0. \tag{3.62}$$

But (3.40b) contracted with n^a gives

$$(\log\xi)^* = \tfrac{1}{2}\mathscr{E}, \tag{3.63}$$

and also from (3.6) and (3.41b) we have

$$n^a{}_{;a} = \tfrac{1}{2}\mathscr{E}. \tag{3.64}$$

Equation (3.62) therefore becomes

$$\overset{*}{\mu} - \overset{*}{p} + (\mu - p + 2\Lambda)\mathscr{E} = 0. \tag{3.65}$$

Also, (3.24) and (3.41b) give

$$\overset{*}{p} - \tfrac{1}{2}(\mu + p)\mathscr{E} = 0. \tag{3.66}$$

Finally, $\mu = \mu(p)$ covariantly differentiated along an integral curve of n^a yields

$$\overset{*}{\mu} = \frac{d\mu}{dp}\overset{*}{p}. \tag{3.67}$$

Equations (3.65), (3.66), and (3.67) are three homogeneous equations for $\overset{*}{\mu}$, $\overset{*}{p}$, and $\mathscr{E}$. On eliminating $\overset{*}{\mu}$ and $\overset{*}{p}$ we find that

$$\left[(\mu + p)\frac{d\mu}{dp} + \mu - 3p + 4\Lambda\right]\mathscr{E} = 0, \tag{3.68}$$

and therefore (3.60) implies that $\mathscr{E} = 0$. Hence, from (3.61), ξn^a is a KV which establishes the theorem. ∎

We have noted that $\mu = p - 2\Lambda$ and $\mu = -3p + 2\Lambda$ are particular solutions of the differential equation (3.59).

We now consider the general solution of (3.59). We will assume that $\mu + p \neq 0$ and for simplicity we will take $\Lambda = 0$. Equation (3.59) reduces to

$$\frac{d\mu}{dp} = \frac{3p-\mu}{p+\mu}. \tag{3.69}$$

If $\Lambda \neq 0$ the change of variables (2.50) reduces (3.59) to the homogeneous form (3.69) in $\bar{p}$ and $\bar{\mu}$ and the following method of solution would still apply. The right-hand side of (3.69) is a homogeneous function of degree 0 in μ and p. We therefore make the standard transformation from (μ,p) to (v,p) where

$$\mu = vp. \tag{3.70}$$

Equation (3.69) becomes

$$p\frac{dv}{dp} = \frac{(1-v)(3+v)}{1+v}. \tag{3.71}$$

The variables are separable in (3.71) and its solution is

$$p^2(1-v)(v+3) = \beta, \tag{3.72}$$

where β is a constant. Expressed in terms of μ and p, (3.72) is

$$(p-\mu)(\mu+3p) = 3\beta, \tag{3.73}$$

which may be rewritten as the following quadratic equation for p:

$$3p^2 - 2\mu p - (\mu^2 + \beta) = 0. \tag{3.74}$$

Whereas the differential equation (2.49) gave rise to a cubic equation for p, (3.69) has produced a quadratic equation for p. The two solutions of (3.69) are

$$p_{\pm} = \frac{\mu}{3} \pm \frac{2}{3}\left(\mu^2 + \frac{3}{4}\beta\right)^{1/2}. \tag{3.75}$$

When $\mu = 0$,

$$p_{\pm}(0) = \pm(\beta/3)^{1/2} \tag{3.76}$$

and therefore for p to be real when $\mu = 0$ we require $\beta \geqslant 0$. We will consider only $\beta \geqslant 0$ and hence from (3.73), either $p \geqslant \mu$ or $p \leqslant -\mu/3$. The only solution that satisfies the physically reasonable condition, $p = 0$ when $\mu = 0$, is obtained when $\beta = 0$ and is $p = \mu$.

Graphs of p_+ and p_- plotted against μ for a selection of values of $\beta \geqslant 0$ are presented in Fig. 4. The solution p_+ is positive for all $\beta > 0$ and all $\mu \geqslant 0$ and this family of graphs is bounded below by the straight line $p_+ = \mu(\beta = 0)$. The solution p_- is negative for all $\beta > 0$ and all $\mu \geqslant 0$ and this family of graphs is bounded above by the straight line $p_- = -\frac{1}{3}\mu(\beta = 0)$.

From (3.75) we have

$$\frac{dp_{\pm}}{d\mu} = \frac{1}{3} \pm \frac{2\mu}{3(\mu^2 + \frac{3}{4}\beta)^{1/2}}, \tag{3.77}$$

and

$$\left.\frac{dp_{\pm}}{d\mu}\right|_{\mu=0} = \frac{1}{3}, \tag{3.78}$$

which is independent of β. From (3.69) it follows that if $\mu \geqslant 0$ then $dp/d\mu \geqslant 0$ if either $p \geqslant \mu/3$, which is satisfied only by the p_+ solution, or $p \leqslant -\mu$ which can be attained only by the p_- solution. From (3.75) we see that

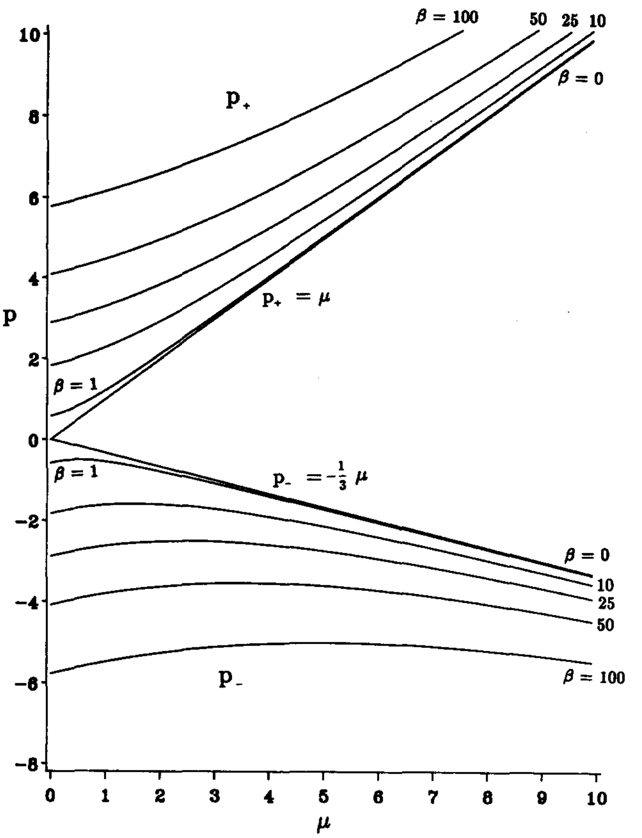


FIG. 4. The pressures p_+ and p_-, given by Eq. (3.75), plotted against μ for $\beta = 0$, 1, 10, 25, 50, and 100. The two straight-line graphs for $\beta = 0$ correspond to the equations of state $p_+ = \mu$ and $p_- = -\mu/3$.

$p_- \leqslant -\mu$ if $\mu \leqslant \frac{1}{2}\beta^{1/2}$; thus $dp_-/d\mu \geqslant 0$ if $0 \leqslant \mu \leqslant \frac{1}{2}\beta^{1/2}$.

Graphs of $dp_{\pm}/d\mu$ plotted against μ for a selection of values of $\beta \geqslant 0$ are presented in Fig. 5. The family of graphs of $dp_+/d\mu$ is bounded by the straight lines $dp_+/d\mu = 1(\beta = 0)$ and $dp_+/d\mu = \frac{1}{3}(\beta = \infty)$ and therefore the fluid is stable against mechanical perturbations $(dp_+/d\mu > 0)$ and the speed of sound relative to the fluid does not exceed the speed of light $(dp_+/d\mu < 1)$ for all $\mu \geqslant 0$ and $\beta \geqslant 0$. The family of graphs of $dp_-/d\mu$ is bounded by the straight lines $dp_-/d\mu = -\frac{1}{3}(\beta = 0)$ and $dp_-/d\mu = \frac{1}{3}(\beta = \infty)$. We have observed that $dp_-/d\mu \geqslant 0$, and therefore the fluid is stable against mechanical perturbations, if for given β, $0 \leqslant \mu \leqslant \frac{1}{2}\beta^{1/2}$. For this range of μ, the speed of sound relative to the fluid v_s does not exceed the speed of light:

$$v_s = \left(\frac{dp_-}{d\mu}\right)^{1/2} \leqslant \frac{1}{\sqrt{3}}. \tag{3.79}$$

We conclude that p_+ satisfies the physically reasonable conditions (2.58) for all $\beta \geqslant 0$ and $\mu \geqslant 0$, but if we insist that $p(0) = 0$, then we must take $\beta = 0$ and $p = \mu$. The solution p_- is negative for all $\beta \geqslant 0$ and $\mu \geqslant 0$ [except that $p_-(0) = 0$ when $\beta = 0$] and $dp_-/d\mu < 0$ and the fluid is unstable to mechanical perturbations, when $\mu > \frac{1}{2}\beta^{1/2}$. When $0 \leqslant \mu \leqslant \frac{1}{2}\beta^{1/2}$, $0 \leqslant dp_-/d\mu \leqslant \frac{1}{3}$ and the fluid is stable to mechanical perturbations and the speed of sound relative to the fluid does not exceed the speed of light.

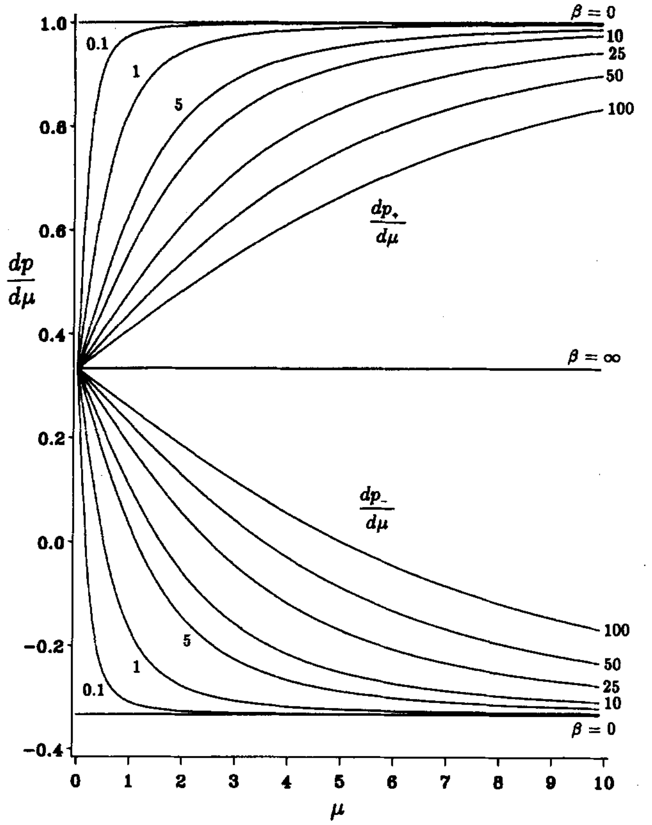


FIG. 5. The pressure gradients $dp_{\pm}/d\mu$, given by Eq. (3.77), plotted against μ for $\beta = 0, 0.1, 1, 5, 10, 25, 50, 100$, and ∞. The two straight-line graphs for $\beta = 0$ correspond to the equations of state $p_+ = \mu$ and $p_- = -\mu/3$.

IV. SPACELIKE RICCI COLLINEATION VECTORS ORTHOGONAL TO u^a: IMPERFECT FLUID SPACE-TIMES

When considering the properties of RCVs $\xi^a = \xi n^a (n_a u^a = 0, n_a n^a = +1)$ admitted by imperfect fluid space-times in which either or both $q^a \neq 0$ and $\pi_{ab} \neq 0$ it is convenient to decompose q^a and π_{ab} with respect to n^a. If X^a and Y_{ab} are any vector and second-order tensor orthogonal to u^a on all indices, then

$$X^a = (n_t X^t) n^a + p^{at} X_t, \tag{4.1}$$

$$Y_{ab} = (Y_{st} n^s n^t) n_a n_b + n_a p^t_b Y_{st} n^s + n_b p^s_a Y_{st} n^t + \tfrac{1}{2}(Y_{st} p^{st}) p_{ab} + (p^s_a p^t_b - \tfrac{1}{2} p^{st} p_{ab}) Y_{st}, \tag{4.2}$$

where p^{ab} is defined by (3.8). Thus q^a and π_{ab} may be decomposed with respect to n^a as

$$q^a = \nu n^a + Q^a, \tag{4.3}$$

$$\pi_{ab} = \gamma(n_a n_b - \tfrac{1}{2} p_{ab}) + 2P_{(a} n_{b)} + D_{ab}, \tag{4.4}$$

where

$$\nu = q_t n^t, \tag{4.5}$$

$$\gamma = \pi_{st} n^s n^t, \tag{4.6}$$

$$Q^a = p^{at} q_t, \tag{4.7}$$

$$P^a = p^{as} \pi_{st} n^t, \tag{4.8}$$

$$D_{ab} = (p^s_a p^t_b - \tfrac{1}{2} p^{st} p_{ab}) \pi_{st}. \tag{4.9}$$

We have

$$Q_a u^a = 0, \quad Q_a n^a = 0, \quad P_a u^a = 0, \quad P_a n^a = 0;$$
$$D_{(ab)} = D_{ab}, \quad D_{ab} u^b = 0, \quad D_{ab} n^b = 0, \quad D^a{}_a = 0. \tag{4.10}$$

The necessary and sufficient conditions for an imperfect fluid space-time to admit an RCV, $\xi^a = \xi n^a$, can be expressed in terms of propagation equations for ν, γ, Q^a, P^a, and D_{ab} along the integral curves of n^a. The analysis is more complex than that for an RCV parallel to u^a as presented in Theorem 2.1. One reason for this is that n^a, unlike u^a, is not in general orthogonal to q_a or π_{ab}. Since $\pi_{ab} u^b = 0$, u^a is a timelike eigenvector of π_{ab} with zero eigenvalue. This suggests that the simpler case in which $q^a = 0$ and n^a is a spacelike eigenvector of π_{ab} should first be considered. This will be done in the next subsection which corresponds to Sec. III C for an RCV parallel to u^a. We will then state without derivation the more complex theorem for a fluid space-time with $q^a \neq 0$ and general π_{ab}.

A. Fluid space-times with $q^a = 0$ and n^a an eigenvector of π_{ab}

Suppose that n^a is a spacelike eigenvector of π_{ab}:

$$\pi_{ab} n^b = \gamma n_a, \tag{4.11}$$

where γ is given by (4.6). Then from (4.8),

$$P^a = 0 \tag{4.12}$$

and (4.4) reduces to

$$\pi_{ab} = \gamma(n_a n_b - \tfrac{1}{2} p_{ab}) + D_{ab}. \tag{4.13}$$

An important example of (4.11) is a fluid with anisotropic pressure. If $q^a = 0$, the energy-momentum tensor of a fluid with anisotropic pressure is[13]

$$T^{ab} = \mu u^a u^b + p_\parallel s^a s^b + p_\perp p^{ab}(s), \tag{4.14}$$

where s^a is a spacelike unit vector orthogonal to $u^a (s_a s^a = +1,\ s_a u^a = 0)$, $p_\parallel$ and $p_\perp$ denote the fluid pressure parallel and perpendicular to s^a, respectively, and

$$p^{ab}(s) = g^{ab} + u^a u^b - s^a s^b. \tag{4.15}$$

When necessary the projection tensor p^{ab} defined by (3.8) will be denoted by $p^{ab}(n)$ to distinguish from $p^{ab}(s)$. Equation (4.14) may be rewritten in the standard form (2.1) with $q^a = 0$ and

$$p = \tfrac{1}{3}(p_\parallel + 2p_\perp), \tag{4.16}$$

$$\pi_{ab} = (p_\perp - p_\parallel)(\tfrac{1}{3} h_{ab} - s_a s_b). \tag{4.17}$$

There are two cases of interest.

(i) $n^a = \pm s^a$. If $n^a = \pm s^a$ then using (4.17),

$$\pi_{ab} n^b = \tfrac{2}{3}(p_\parallel - p_\perp) n_a \tag{4.18}$$

and therefore n^a is a spacelike eigenvector of π_{ab} with

$$\gamma = \tfrac{2}{3}(p_\parallel - p_\perp). \tag{4.19}$$

Further, it is easily verified from (4.9) that

$$D_{ab} = 0. \tag{4.20}$$

(ii) $n_a s^a = 0$. If n^a is orthogonal to s^a then

$$\pi_{ab} n^b = \tfrac{1}{2}(p_\perp - p_\parallel) n_a, \tag{4.21}$$

and therefore n^a is a spacelike eigenvector of π_{ab} with eigenvalue

$$\gamma = \tfrac{1}{2}(p_\perp - p_\parallel). \tag{4.22}$$

A direct calculation using (4.9) gives

$$D_{ab} = (p_\parallel - p_\perp)(s_a s_b - \tfrac{1}{2} p_{ab}(n)) \tag{4.23}$$

where $p_{ab}(n)$ is given by (3.8).

An example of a fluid with energy-momentum tensor of the form (4.14) is a plasma in a strong magnetic field. If the particle collision density is low, a strong magnetic field can cause the pressure along and perpendicular to the magnetic field to be unequal. The total energy-momentum tensor is[13]

$$T^{ab} = \mu u^a u^b + p_\parallel s^a s^b + p_\perp p^{ab}(s) + T^{ab}_{\mathrm{EM}}, \tag{4.24}$$

where $s^a = H^a/H$, H^a is the local magnetic field measured by u^a, and T^{ab}_{EM} is the electromagnetic energy–momentum tensor. We will take for T^{ab}_{EM} the Minkowski tensor. If the local electric field E^a vanishes the Minkowski tensor for a pure magnetic field is[8]

$$\begin{aligned} T^{ab}_{\mathrm{EM}} &= \tfrac{1}{2}\lambda H^2 u^a u^b + \tfrac{1}{6}\lambda H^2 h_{ab} + \lambda H^2(\tfrac{1}{3} h_{ab} - s_a s_b) \\ &= \tfrac{1}{2}\lambda H^2 u^a u^b - \tfrac{1}{2}\lambda H^2 s^a s^b + \tfrac{1}{2}\lambda H^2 p^{ab}(s), \end{aligned} \tag{4.25}$$

where λ is the magnetic permeability. Thus T^{ab}_{EM} also has the form (4.14) and the total energy-momentum tensor (4.24) can be written as

$$T^{ab} = \bar{\mu} u^a u^b + \bar{p}_\parallel s^a s^b + \bar{p}_\perp p^{ab}(s), \tag{4.26}$$

where

$$\bar{\mu} = \mu + \tfrac{1}{2}\lambda H^2, \tag{4.27}$$

$$\bar{p}_\parallel = p_\parallel - \tfrac{1}{2}\lambda H^2, \tag{4.28}$$

$$\bar{p}_\perp = p_\perp + \tfrac{1}{2}\lambda H^2. \tag{4.29}$$

The results (4.18) and (4.23) are valid with $p_\parallel$ and $p_\perp$ replaced by $\bar{p}_\parallel$ and $\bar{p}_\perp$.

We now outline the derivation of the following theorem that is established in a similar way to Theorems 2.1 and 3.1.

Theorem 4.1: If Einstein's field equations (1.5) are satisfied and $q^a = 0$ but $\pi_{ab} \neq 0$ and if $n^a(n_a n^a = +1, n_a u^a = 0)$ is a spacelike eigenvector of π_{ab}:

$$\pi_{ab} n^b = \gamma n_a, \tag{4.30}$$

then the fluid space-time admits an RCV, $\xi^a = \xi n^a$, if and only if

(i) $$(\mu + 3p - 2\Lambda)\omega_{at} n^t = \tfrac{1}{2}(\mu - p - \gamma + 2\Lambda) N_a + D_{at} N^t, \tag{4.31}$$

(ii) $$\begin{aligned} p^c_a p^d_b \overset{*}{D}_{cd} &= -(\mu - p - \gamma + 2\Lambda)\mathscr{S}_{ab} + (D^{cd}\mathscr{S}_{cd}) p_{ab} \\ &\quad - \mathscr{E} D_{ab} - 2\mathscr{S}_{t(a} D_{b)}{}^t - 2\mathscr{R}_{t(a} D_{b)}{}^t, \end{aligned} \tag{4.32}$$

(iii) $$(\mu - p + 2\gamma + 2\Lambda)[\overset{*}{n}_a + (\log \xi)_{,a} - (\mathscr{E} + n_t \dot{u}^t) n_a] = 0, \tag{4.33}$$

(iv) $$\overset{*}{\gamma} = -\tfrac{1}{2}(\mu - p + 2\gamma + 2\Lambda)(\mathscr{E} + 2 n_t \dot{u}^t) - \gamma \mathscr{E} + \tfrac{2}{3} D^{cd}\mathscr{S}_{cd}, \tag{4.34}$$

(v) $$[(\mu - p + 2\gamma + 2\Lambda)\xi n^a]_{;a} = 0, \tag{4.35}$$

where N^a and D_{ab} are defined by (3.4) and (4.9) and $\mathscr{E}$, $\mathscr{S}_{ab}$, and $\mathscr{R}_{ab}$ are the expansion, shear, and rotation of the spacelike congruence generated by n^a as measured by an observer with the fluid unit four-velocity u^a.

Proof: We give the main steps in the proof. Using Einstein's field equations (1.5) and (4.13) for π_{ab}, it can be verified that (3.15) becomes

$$\begin{aligned} \mathscr{L}_{\xi n} R_{ab} = \xi\, [&\tfrac{1}{2}(\overset{*}{\mu} + 3\overset{*}{p}) u_a u_b + \tfrac{1}{2}(\overset{*}{\mu} - \overset{*}{p} - \overset{*}{\gamma}) p_{ab} + \tfrac{1}{2}(\overset{*}{\mu} - \overset{*}{p} + 2\overset{*}{\gamma}) n_a n_b + 2(\mu + p - \tfrac{1}{2}\gamma)(\overset{*}{\dot{u}}_{(a} u_{b)} - u_{(a} u^t{}_{;b} n_t) \\ &+ 3\gamma \overset{*}{n}_{(a} n_{b)} + (\mu - p + 2\gamma + 2\Lambda) n_{(a} (\log \xi)_{,b} + (\mu - p - \gamma + 2\Lambda) n_{(a;b)} + 2 D_{t(a} n^t{}_{;b}]. \end{aligned} \tag{4.36}$$

Suppose first that ξn^a is an RCV. Then the right-hand side of (4.36) vanishes and by contracting it in turn with $u^a u^b$, $u^a n^b$, $u^a p^{bc}$, $n^a n^b$, $n^a p^{bc}$, p^{ab}, and $p^{ac} p^{bd} - \tfrac{1}{2} p^{ab} p^{cd}$ the following seven equations are derived:

$$\overset{*}{\mu} + 3\overset{*}{p} + 2(\mu + 3p - 2\Lambda) n_t \dot{u}^t = 0, \tag{4.37}$$

$$(\mu - p + 2\gamma + 2\Lambda)[(\log \xi)^{\cdot} - n_b \overset{*}{\dot{u}}{}^b] = 0, \tag{4.38}$$

$$\begin{aligned} &\tfrac{1}{2}(\mu - p + 2\Lambda) h_a^{\ b} \dot{n}_b - (\mu + p)(\overset{*}{\dot{u}}_a - (n_t \overset{*}{\dot{u}}{}^t) n_a) \\ &\quad + \tfrac{1}{2}(\mu + 3p - 2\Lambda) p^b_a n^t u_{t;b} - \tfrac{1}{2}\gamma N_a + D_{at} N^t = 0, \end{aligned} \tag{4.39}$$

$$\overset{*}{\mu} - \overset{*}{p} + 2\overset{*}{\gamma} + 2(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^* = 0, \tag{4.40}$$

$$(\mu - p + 2\gamma + 2\Lambda) p^b_a [\overset{*}{n}_b + (\log \xi)_{,b}] = 0, \tag{4.41}$$

$$\overset{*}{\mu} - \overset{*}{p} - \overset{*}{\gamma} + (\mu - p - \gamma + 2\Lambda)\mathscr{E} + 2 D^{cd}\mathscr{S}_{cd} = 0, \tag{4.42}$$

$$\begin{aligned} &p^c_a p^d_b \overset{*}{D}_{cd} + (\mu - p - \gamma + 2\Lambda)\mathscr{S}_{ab} - (D^{cd}\mathscr{S}_{cd}) p_{ab} \\ &\quad + \mathscr{E} D_{ab} + 2\mathscr{S}_{t(a} D_{b)}{}^t + 2\mathscr{R}_{t(a} D_{b)}{}^t = 0. \end{aligned} \tag{4.43}$$

The momentum conservation equation contracted with n^a will also be required. For a fluid with $q^a = 0$, the momentum conservation equation, which follows from Einstein's field equations, is[7,8]

$$(\mu + p)\dot{u}_a = -h_a^{\ b}(p_{,b} + \pi_b{}^c{}_{;c}). \tag{4.44}$$

If (4.44) is contracted with n^a and (4.13) is used for π^{ab} then we obtain

$$\overset{*}{p} + \overset{*}{\gamma} + (\mu + p + \gamma) n_c \dot{u}^c + \tfrac{2}{3}\gamma \mathscr{E} - D^{cd}\mathscr{S}_{cd} = 0. \tag{4.45}$$

(i) Condition (4.31) follows directly from (4.39) with the aid of (3.25).

(ii) Condition (4.32) is given by (4.43).

(iii) To derive condition (4.33) we first expand (4.41) and use (4.38); this gives

$$(\mu - p + 2\gamma + 2\Lambda)[\overset{*}{n}_a + (\log \xi)_{,a} - (\log \xi)^* n_a] = 0. \tag{4.46}$$

Now, $(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^*$ is given in terms of $\overset{*}{\mu}$, $\overset{*}{p}$, and $\overset{*}{\gamma}$ by (4.40). By solving (4.37), (4.42), and (4.45) for $\overset{*}{\mu}$, $\overset{*}{p}$, and $\overset{*}{\gamma}$ we find that

$$\begin{aligned}\overset{*}{\mu} = & -(\mu - p + \tfrac{1}{2}\gamma + 2\Lambda)\mathscr{E} \\ & -(\mu + p + \gamma)n_c\dot{u}^c - D^{cd}\mathscr{S}_{cd},\end{aligned} \tag{4.47}$$

$$\begin{aligned}\overset{*}{p} = & \tfrac{1}{3}(\mu - p + \tfrac{1}{2}\gamma + 2\Lambda)\mathscr{E} \\ & -\tfrac{1}{3}(\mu + 5p - \gamma - 4\Lambda)n_c\dot{u}^c + \tfrac{1}{3}D^{cd}\mathscr{S}_{cd},\end{aligned} \tag{4.48}$$

$$\begin{aligned}\overset{*}{\gamma} = & -\tfrac{1}{3}(\mu - p + 5\gamma + 2\Lambda)\mathscr{E} \\ & -\tfrac{2}{3}(\mu - p + 2\gamma + 2\Lambda)n_c\dot{u}^c + \tfrac{2}{3}D^{cd}\mathscr{S}_{cd},\end{aligned} \tag{4.49}$$

and by substituting for $\overset{*}{\mu}$, $\overset{*}{p}$, and $\overset{*}{\gamma}$ in (4.40), we obtain

$$\begin{aligned}&(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^* \\ &\quad = (\mu - p + 2\gamma + 2\Lambda)(\mathscr{E} + n_c\dot{u}^c).\end{aligned} \tag{4.50}$$

Condition (4.33) follows directly from (4.46) and (4.50).

(iv) Condition (4.34) is given by (4.49).

(v) Finally, consider (4.33). Equation (4.40) may be written as

$$(\mu - p + 2\gamma + 2\Lambda)^* + 2(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^* = 0. \tag{4.51}$$

But from (3.31) and (4.50), we have

$$(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^* = (\mu - p + 2\gamma + 2\Lambda)n^a{}_{;a}, \tag{4.52}$$

and by replacing one of the terms $(\mu - p + 2\gamma + 2\Lambda)(\log \xi)^*$ in (4.51) by (4.52) it follows that

$$\begin{aligned}&(\mu - p + 2\gamma + 2\Lambda)_{,a}\xi n^a \\ &\quad + (\mu - p + 2\gamma + 2\Lambda)(\xi_{,a}n^a + \xi n^a{}_{;a}) = 0,\end{aligned} \tag{4.53}$$

from which (4.35) is immediately obtained.

Hence, if ξn^a is an RCV then conditions (4.31)–(4.35) are satisfied.

Conversely, suppose that (4.31)–(4.33) hold and that Einstein's field equations are satisfied. We show that ξn^a is an RCV.

If (4.32) for $\overset{*}{D}_{ab}$, (4.33) for $(\mu - p + 2\gamma + 2\Lambda)(\log \xi)_{,a}$, and (4.34) for $\overset{*}{\gamma}$ are substituted into (4.36) and $n_{a;b}$ and $n_{t;b}$ are expanded using (3.9), then (4.36) becomes

$$\begin{aligned}\mathscr{L}_{\xi n}R_{ab} = \xi\,[&\tfrac{1}{2}(\overset{*}{\mu} + 3\overset{*}{p})u_au_b + \tfrac{1}{2}\{\overset{*}{\mu} - \overset{*}{p} + \tfrac{4}{3}(\mu - p + \tfrac{1}{2}\gamma + 2\Lambda)\mathscr{E} + \tfrac{2}{3}(\mu - p + 2\gamma + 2\Lambda)n_t\dot{u}^t + \tfrac{4}{3}D^{cd}\mathscr{S}_{cd}\}h_{ab} \\ & + (\mu + 3p - 2\Lambda)(\overset{*}{u}_{(a}u_{b)} - n^tu_{t;(a}u_{b)}) - (\mu - p - \gamma + 2\Lambda)N_{(a}u_{b)} - 2u_{(a}D_{b)t}N^t].\end{aligned} \tag{4.54}$$

With the aid of the identity (3.25) for $n^tu_{t;b}$ and (4.31), (4.54) simplifies to

$$\begin{aligned}\mathscr{L}_{\xi n}R_{ab} = \tfrac{1}{2}\xi\,[&\{\overset{*}{\mu} + 3\overset{*}{p} + 2(\mu + 3p - 2\Lambda)n_t\dot{u}^t\}u_au_b \\ & + \{\overset{*}{\mu} - \overset{*}{p} + \tfrac{4}{3}(\mu - p + \tfrac{1}{2}\gamma + 2\Lambda)\mathscr{E} \\ & + \tfrac{2}{3}(\mu - p + 2\gamma + 2\Lambda)n_t\dot{u}^t + \tfrac{4}{3}D^{cd}\mathscr{S}_{cd}\}h_{ab}].\end{aligned} \tag{4.55}$$

It remains to obtain expressions for $\overset{*}{\mu}$ and $\overset{*}{p}$. We first expand (4.35) to obtain

$$\overset{*}{\mu} - \overset{*}{p} + 2\overset{*}{\gamma} + (\mu - p + 2\gamma + 2\Lambda)((\log \xi)^* + n^a{}_{;a}) = 0. \tag{4.56}$$

But contraction of (4.33) with n^a gives again (4.50) and using (4.50) and also (3.31) for $n^a{}_{;a}$, (4.56) becomes

$$\overset{*}{\mu} - \overset{*}{p} + 2\overset{*}{\gamma} + 2(\mu - p + 2\gamma + 2\Lambda)(\mathscr{E} + n_t\dot{u}^t) = 0. \tag{4.57}$$

If (4.57), (4.34), and (4.45), which follows from Einstein's field equations, are solved for $\overset{*}{\mu}$ and $\overset{*}{p}$ then (4.47) and (4.48) are again obtained. By substituting from (4.47) and (4.48) for $\overset{*}{\mu}$ and $\overset{*}{p}$ into (4.55), it is readily verified that the right-hand side of (4.55) vanishes and therefore ξn^a is an RCV. ■

It is easily verified that (4.35) is the conservation law (1.6) for the special case of an imperfect fluid with $q^a = 0, \pi_{ab}n^b = \gamma n_a$ and $v_b = \xi n_b$. Unlike a perfect fluid, for which the necessary and sufficient conditions (3.10)–(3.14) do not depend on $\mathscr{R}_{ab}$, the rotation of the spacelike congruence enters through (4.32) when $D_{ab} \neq 0$.

Applications of Theorem 4.1 may be divided into two cases, $D_{ab} = 0$ and $D_{ab} \neq 0$. We first consider imperfect fluids with $D_{ab} = 0$.

Theorem 4.2: Suppose that an imperfect fluid space-time admits an RCV, $\xi^a = \xi n^a (n_an^a = +1, n_au^a = 0)$, that Einstein's field equations hold and that

$$T_{ab} = \mu u_au_b + ph_{ab} + \pi_{ab}, \tag{4.58}$$

where π_{ab} satisfies

$$\pi_{ab}n^b = \gamma n_a, \quad D_{ab} = (p^c_ap^d_b - \tfrac{1}{2}p^{cd}p_{ab})\pi_{cd} = 0. \tag{4.59}$$

(a) Then, either $p = \mu - \gamma + 2\Lambda$ or $\mathscr{S}_{ab} = 0$, (4.60)

where $\mathscr{S}_{ab}$ is the shear of the n-congruence as measured by an observer with 4-velocity u^a.

(b) $\omega = 0$. Then either $p = \mu - \gamma + 2\Lambda$ or the integral curves of n^a are material curves in the fluid.

(c) $\omega \neq 0$. (i) If $p = \mu - \gamma + 2\Lambda$ but $\mu + \Lambda - \frac{3}{4}\gamma \neq 0$, then $n^a = \pm\omega^a/\omega$.

(ii) If the integral curves of n^a are material curves in the fluid and $\mu + 3p - 2\Lambda \neq 0$ then $n^a \pm \omega^a/\omega$.

(iii) If $n^a = \pm\omega^a/\omega$ and if $p \neq \mu - \gamma + 2\Lambda$, then the vortex lines are material lines in the fluid.

Proof: (a) When $D_{ab} = 0$, (4.32) reduces to

$$(\mu - p - \gamma + 2\Lambda)\mathscr{S}_{ab} = 0 \tag{4.61}$$

and therefore either $p = \mu - \gamma + 2\Lambda$ or $\mathscr{S}_{ab} = 0$.

(b) When $\omega = 0$ and $D_{ab} = 0$, (4.31) reduces to

$$(\mu - p - \gamma + 2\Lambda)N_a = 0, \tag{4.62}$$

and hence either $p = \mu - \gamma + 2\Lambda$ or $N_a = 0$. When $N_a = 0$ the integral curves n^a are material curves.

(c) When $\omega \neq 0$ and $D_{ab} = 0$, (4.31) becomes

$$(\mu + 3p - 2\Lambda)\omega_{at} n^t = \tfrac{1}{2}(\mu - p - \gamma + 2\Lambda)N_a. \quad (4.63)$$

(i) If $p = \mu - \gamma + 2\Lambda$, then (4.63) reduces to

$$(\mu + \Lambda - \tfrac{3}{4}\gamma)\omega_{at} n^t = 0 \quad (4.64)$$

and hence if $\mu + \Lambda - \frac{3}{4}\gamma \neq 0$ then $\omega_{at} n^t = 0$ and therefore $n^a = \pm \omega^a/\omega$.

(ii) If the integral curves of n^a are material curves then $N^a = 0$ and therefore from (4.63), if $\mu + 3p - 2\Lambda \neq 0$, then $\omega_{at} n^t = 0$ and $n^a = \pm \omega^a/\omega$.

(iii) If $n^a = \pm \omega^a/\omega$ and if $p \neq \mu - \gamma + 2\Lambda$ then from (4.63), $N^a = 0$ and the integral curves of n^a are material curves. ■

As an application of Theorem 4.2, consider a fluid with anisotropic pressure and energy-momentum tensor of the form (4.14). This includes a fluid with anisotropic pressure produced by a pure magnetic field. If $n^a = \pm s^a$, where s^a is the preferred direction, then by (4.18) and (4.20), n^a is an eigenvector of π_{ab} and $D_{ab} = 0$.

Theorem 4.3: Suppose that a fluid space-time with energy-momentum tensor (4.14) admits an RCV, $\xi^a = \xi n^a$, with $n^a = \pm s^a (s_a u^a = 0,\ s_a s^a = +1)$.

(i) Then

$$\text{either } p_\parallel = \mu + 2\Lambda \text{ or } \mathscr{S}_{ab} = 0, \quad (4.65)$$

where $\mathscr{S}_{ab}$ is the shear of the spacelike congruence generated by n^a (equivalently s^a) as measured by an observer with four-velocity u^a.

(ii) If the anisotropic pressure is produced by a pure magnetic field and

$$\mu + p_\parallel + 2p_\perp + \lambda H^2 - 2\Lambda \neq 0, \quad (4.66)$$

then the magnetic field lines must coincide with the vortex lines if $\omega \neq 0$:

$$n^a = \pm H^a/H = \pm \omega^a/\omega. \quad (4.67)$$

Proof: (i) The result (4.65) follows directly from (4.60) using (4.16) for p and (4.19) for γ.

(ii) Ellis[8] has shown that if the local electric field $E^a = 0$, then the magnetic field lines are material lines in the fluid. This follows from the Maxwell equation governing the propagation of H^a along u^a: if the magnetic permeability λ is constant, then

$$h^a_b \dot{H}^b = u^a{}_{;b} H^b - \theta H^a, \quad (4.68)$$

and contraction of (4.68) with H^a gives

$$\dot{H} = H n_a \dot{u}^a - \theta H, \quad (4.69)$$

where $n^a = H^a/H$. It is readily verified using (4.68) and (4.69) that (3.2) is satisfied. The total energy-momentum tensor is (4.26), which is of the form (4.14), and the result (4.67) follows from Theorem 4.2(c) (ii) if

$$\bar{\mu} + 3\bar{p} - 2\Lambda \neq 0, \quad (4.70)$$

where, by (4.16), $\bar{p} = \frac{1}{3}(p_\parallel + 2p_\perp)$. Using (4.27)–(4.29) it can be checked that (4.70) is condition (4.66). ■

The result (4.65) is independent of $p_\perp$. When H^a is parallel to ω^a the charge density as measured by u^a, ϵ, must be nonzero. This follows from the Maxwell equation[8]

$$2\omega_a H^a = \epsilon. \quad (4.71)$$

Consider next imperfect fluids with $D_{ab} \neq 0$.

Theorem 4.4: Suppose that an imperfect fluid space-time admits an RCV, $\xi^a = \xi n^a (n_a n^a = +1, n_a u^a = 0)$, that Einstein's field equations are satisfied and that T_{ab} is given by (4.58) where $\pi_{ab} n^b = \gamma n_a$ and $D_{ab} \neq 0$.

(i) If $\omega = 0$ or if $n^a = \pm \omega^a/\omega$, then N^a is a spacelike eigenvector of D_{ab} and π_{ab} with eigenvalues $-\frac{1}{2}(\mu - p - \gamma + 2\Lambda)$ and $-\frac{1}{2}(\mu - p + 2\Lambda)$, respectively.

(ii) If $\omega \neq 0$ and $\mu + 3p - 2\Lambda \neq 0$ and if the integral curves of n^a are material curves in the fluid then $n^a = \pm \omega^a/\omega$.

Proof: (i) If $\omega = 0$ or if $n^a = \pm \omega^a/\omega$, then (4.31) reduces to

$$D_{ab} N^b = -\tfrac{1}{2}(\mu - p - \gamma + 2\Lambda)N_a. \quad (4.72)$$

Also, it follows by contracting (4.13) with N^a that

$$\pi_{ab} N^b = -\tfrac{1}{2}(\mu - p + 2\Lambda)N_a, \quad (4.73)$$

which establishes the results.

(ii) If the integral curves of n^a are material curves in the fluid then $N^a = 0$ and (4.31) reduces to

$$(\mu + 3p - 2\Lambda)\omega_{at} n^t = 0. \quad (4.74)$$

Hence, if $\mu + 3p - 2\Lambda \neq 0$ then $\omega_{at} n^t = 0$ and therefore $n^a = \pm \omega^a/\omega$. ■

The results of Theorem 4.4 apply to a fluid with anisotropic pressure in which n^a is orthogonal to the preferred direction s^a.

B. Fluid space-times with $q^a \neq 0$ and general π_{ab}

Finally, we state without derivation a set of necessary and sufficient conditions for a fluid space-time with $q^a \neq 0$ and general $\pi_{ab} \neq 0$ to admit an RCV, $\xi^a = \xi n^a$, orthogonal to u^a. These conditions are expressed in terms of the propagation equations for ν, γ, Q^a, P^a, and D^{ab} along the integral curves of n^a and the conservation law (1.6).

Theorem 4.5: If Einstein's field equations (1.5) are satisfied, then a fluid space-time with energy-momentum tensor (2.1) admits an RCV, $\xi^a = \xi n^a (n_a n^a = +1, n_a u^a = 0)$, if and only if

$$\begin{aligned}
\text{(i) } p^b_a \overset{*}{Q}_b = {} & -(\mu + 3p - 2\Lambda)\omega_{ab} n^b + \tfrac{1}{2}(\mu - p - \gamma + 2\Lambda)N_a + D_{ab}N^b + ((\log \xi)^{\cdot} - n_b \overset{*}{u}{}^b)P_a \\
& - \nu p^b_a(\overset{*}{n}_b + (\log \xi)_{,b}) - (n_b \dot{u}^b + \tfrac{1}{2}\mathscr{E})Q_a - Q^b \mathscr{S}_{ba} - Q^b \mathscr{R}_{ba}, \quad (4.75) \\
\text{(ii) } p^c_a p^d_b \overset{*}{D}_{cd} = {} & -(\mu - p - \gamma + 2\Lambda)\mathscr{S}_{ab} + (D^{cd}\mathscr{S}_{cd} + 2\omega_{cd} Q^c n^d + P^c(\overset{*}{n}_c + (\log \xi)_{,c}))p_{ab} \\
& - 4Q_{(a}\omega_{b)t} n^t - 2P_{(a} p^c_{b)}(\overset{*}{n}_c + (\log \xi)_{,c}) - \mathscr{E} D_{ab} - 2\mathscr{S}_{c(a} D_{b)}{}^c - 2\mathscr{R}_{c(a} D_{b)}{}^c, \quad (4.76)
\end{aligned}$$

$$(\mathrm{iii})\ p_a^b \overset{*}{P}_b = -\tfrac{1}{2}(\mu - p + 2\gamma + 2\Lambda)[\overset{*}{n}_a + (\log\xi)_{,a} - (\mathscr{E} + n_b \dot{u}^b)n_a - (n_b \overset{*}{u}^b - (\log\xi)^{\cdot})u_a] + [P^b(\log\xi)_{,b} + P^b{}_{;b} + \nu(\log(\xi\nu))^{\cdot} + \nu\theta]n_a - 2\nu\omega_{ab}n^b - (\tfrac{1}{2}\mathscr{E} + (\log\xi)^*)P_a - P^b\mathscr{S}_{ba} - P^b\mathscr{R}^{ba}, \tag{4.77}$$

$$(\mathrm{iv})\ \overset{*}{\nu} = \tfrac{1}{2}(\mu - p + 2\gamma + 2\Lambda)((\log\xi)^{\cdot} - n_a \overset{*}{u}^a) - \nu(n_a \dot{u}^a + (\log\xi)^*) + N_a P^a, \tag{4.78}$$

$$(\mathrm{v})\ \overset{*}{\gamma} = -\tfrac{1}{3}(\mu - p + 5\gamma + 2\Lambda)\mathscr{E} - \tfrac{2}{3}(\mu - p + 2\gamma + 2\Lambda)n_a \dot{u}^a + \tfrac{2}{3}D^{ab}\mathscr{S}_{ab} + \tfrac{4}{3}\omega_{ab}Q^a n^b - \tfrac{4}{3}\nu(\theta + (\log\xi\nu)^{\cdot}) - \tfrac{4}{3}P^a{}_{;a} + \tfrac{2}{3}P^a(\overset{*}{n}_a - (\log\xi)_{,a}), \tag{4.79}$$

$$(\mathrm{vi})\ [\xi(P^a + \nu u^a + \tfrac{1}{2}(\mu - p + 2\gamma + 2\Lambda)n^a)]_{;a} = 0, \tag{4.80}$$

where N^a is defined by (3.4), ν, γ, Q^a, P^a, and D^{ab} are defined in terms of q^a and π^{ab} by (4.5) to (4.9) and $\mathscr{E}$, $\mathscr{S}_{ab}$, and $\mathscr{R}_{ab}$ are the expansion, shear, and rotation of the spacelike congruence generated by n^a as measured by an observer with the fluid unit four-velocity u^a. ■

Theorem 4.5 can be established in a similar way to Theorem 4.1. It is readily verified that (4.80) is the conservation law (1.6).

V. CONCLUDING REMARKS

All of the properties derived in this paper are dynamic results because they were obtained with the aid of Einstein's field equations. They depend on the nature of the fluid through the energy-momentum tensor and the equation of state. We have seen that many of the properties of timelike RCVs parallel to u^a have direct analogues for spacelike RCVs orthogonal to u^a. The decomposition of π_{ab} and q^a with respect to n^a proved useful in obtaining necessary and sufficient conditions for a space-time to admit a spacelike RCV parallel to n^a. These conditions depend on the rotation tensor $\mathscr{R}_{ab}$, of the spacelike congruence generated by n^a only when the components D_{ab} or P^a of π_{ab} or Q^a or q^a are nonzero. The vorticity vector of the fluid, ω^a, plays an essential role in the properties of rotational fluid space-times that admit an RCV orthogonal to u^a. This is not unexpected because ω^a defines locally a preferred direction in a rotational fluid.

ACKNOWLEDGMENTS

We thank M. Nakayama for producing the graphs in Figs. 1–5.

DPM thanks the Foundation for Research Development of the Council for Scientific and Industrial Research, Pretoria, for an Operating Grant.

[1] D. R. Oliver and W. R. Davis, Gen. Relat. Gravit. **8**, 905 (1977).

[2] D. R. Oliver and W. R. Davis, Ann. Inst. Henri Poincaré **30**, 339 (1979).

[3] P. J. Greenberg, J. Math. Anal. Appl. **30**, 128 (1970).

[4] M. Tsamparlis and D. P. Mason, J. Math. Phys. **24**, 1577 (1983).

[5] D. P. Mason and M. Tsamparlis, J. Math. Phys. **26**, 2881 (1985).

[6] C. D. Collinson, Gen. Relat. Gravit. **1**, 137 (1970).

[7] G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–189.

[8] G. F. R. Ellis, in *Cargèse Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6, pp. 1–60.

[9] Latin indices run over the four coordinates of space-time. A semicolon denotes covariant differentiation with respect to the metric tensor, g_{ab}, of space-time [signature$(-+++)$]. An overhead dot denotes covariant differentiation along a fluid particle world line; for example

$$\dot{A}^a = A^a{}_{;b}u^b.$$

The projection tensor

$$h_{ab} = g_{ab} + u_a u_b$$

projects into the instantaneous rest space of an observer with four-velocity u^a. The Riemann curvature tensor is defined through the identity

$$A_{a;[bc]} = 2R_{tabc}A^t$$

and the Ricci tensor R_{ab} and the Ricci scalar R are defined, respectively, as

$$R_{ab} = R^t{}_{atb}, \quad R = R^a{}_a.$$

Units are used in which the speed of light in vacuum and Einstein's gravitational constant are both unity. Einstein's field equations are

$$R_{ab} + (\Lambda - \tfrac{1}{2}R)g_{ab} = T_{ab},$$

or equivalently

$$R_{ab} = T_{ab} + (\Lambda - \tfrac{1}{2}T)g_{ab}.$$

[10] L. K. Norris, L. H. Green, and W. R. Davis, J. Math. Phys. **18**, 1305 (1977).

[11] J. Ehlers, Akad. Wiss. Lit. (Mainz) Abh. Math. Naturwiss. Kl. **11**, 793 (1961).

[12] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).

[13] R. Maartens, D. P. Mason, and M. Tsamparlis, J. Math. Phys. **27**, 2987 (1986).

[14] The speed of sound relative to the fluid, v_s, is

$$v_s = \left(\frac{dp}{d\mu}\right)^{1/2}.$$

The condition $dp/d\mu > 0$ is required to ensure that v_s is real and therefore that the fluid is stable against mechanical perturbations and the condition $dp/d\mu < 1$ is required to ensure that v_s does not exceed the speed of light (Ref. 8, p. 19).

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980), pp. 34 and 51.

[16] L. Herrera, L. Jiménez, J. Leal, J. Ponce de Léon, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).

Transversal affine connection and quantization of constrained systems

Petr Hajicek
Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland
Karel V. Kuchar
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 16 November 1989; accepted for publication 14 February 1990)

The Dirac quantization of a finite-dimensional relativistic system with a quadratic super-Hamiltonian and linear supermomenta is investigated. In a previous work, the operator constraints were consistently factor-ordered in such a way that the resulting quantum theory was invariant under all relevant transformations of the classical theory. The method was based on a special choice of coordinates and gauge. Here, coordinate-independent methods are worked out and a quite general gauge is used. A new mathematical concept, the so-called "transversal affine connection," is introduced. This connection is not a *linear* connection and is associated with a *degenerate* metric. The corresponding curvature tensor is defined and its components are calculated. The formalism is used to reconstruct the operator constraints, clarify their geometric meaning, and calculate their commutators.

I. INTRODUCTION

In Ref. 1, which will be abbreviated by I below, we studied the Dirac constraint quantization of a finite-dimensional relativistic gauge system with a quadratic super-Hamiltonian and linear supermomenta. We required that certain classical symmetries of the system are preserved in the quantum theory. This requirement has implied a unique factor ordering of the constraint operators, which automatically satisfies the condition that the commutators do not produce more constraints.

The constraint operators contain terms that strongly resemble a curvature scalar or a covariant derivative in a Riemannian space. The nature of the system does not, however, admit an intrinsic nondegenerate metric that would define a linear connection. In I, we avoided discussing the geometrical structure that underlies the constraints and relied instead on suitable gauges and special coordinates. Our present aim is to reveal the geometry of the constraints by using coordinate-independent methods. We believe that the resulting formalism shall be useful in the quantum theory of general gauge systems.

The program of the paper is as follows. In Sec. II, we briefly introduce our model system. In Sec. III, we describe the geometrical structure imposed on the configuration space-time by the constraints. In Sec. IV, we introduce the concept of a transversal distribution and of transversal tensor fields. In Sec. V, we present some properties of Lie derivatives of these fields that are important later. The keystone of our geometrical theory, the so-called transversal affine connection, is defined and discussed in Sec. VI. It is a connection in a subbundle of the bundle of linear frames determined by a degenerate metric. As far as we know, such an object was not previously identified in literature. In Sec. VII, we introduce the corresponding curvature tensor, Ricci tensor, and curvature scalar, evaluate their components, and write down the Bianchi identities. Finally, in Sec. VIII, we demonstrate the usefulness of the formalism by directly calculating the commutation relations of the constraint operators. (In I, we did the calculation in special coordinates and gauge and then transformed to the general case.)

II. DESCRIPTION OF THE MODEL

We consider a finite-dimensional relativistic parametrized system with additional gauge degrees of freedom. Its phase space is a cotangent bundle over a manifold $\mathscr{M}$—the configuration space-time—of dimension $N+1$. We denote the coordinates in the configuration space by Q^A, $A=0,\dots,N$, and the components of the momenta by P_A. This structure is invariant under the contact transformation:

$$Q^{A'}=Q^{A'}(Q),\quad P_{A'}=\frac{\partial Q^B}{\partial Q^{A'}}P_B. \tag{1}$$

The constraints have the form

$$H=\tfrac{1}{2}G^{AB}(Q)P_AP_B+U^A(Q)P_A+V(Q), \tag{2}$$

$$H_\alpha=\Phi^A_\alpha(Q)P_A, \tag{3}$$

where $\alpha=1,\dots,\nu$. With respect to the contact transformations (1), $G^{AB}(Q)$ is a contravariant symmetric tensor field called "metric," $U^A(Q)$ and $\Phi^A_\alpha(Q)$ are vector fields, and $V(Q)$ is a scalar field, on $\mathscr{M}$. We assume that the gauge constraints H_α can be chosen in such a way that the ν fields $\Phi^A_\alpha(Q)$ determine ν *independent* vectors at each point Q of $\mathscr{M}$.

The vectors $\Phi^A_\alpha(Q)$ at a point Q span a ν-dimensional vector space $T_\parallel$ called longitudinal vector space. The covectors w_A at Q that annihilate $T_\parallel$ are called transversal covectors. They span an $(n+1)$-dimensional space $T^*_\perp$ called the transversal covector space at Q, with $n=N-\nu$. The metric $G^{AB}(Q)$ induces a nondegenerate metric on $T^*_\perp$ with the signature $(-,+,\dots,+)$.

The fields $G^{AB}(Q)$, $U^A(Q)$, $\Phi^A_\alpha(Q)$, and $V(Q)$ must be such that the Poisson algebra of the constraints closes:

$$\{H_\alpha,H_\beta\}=c^\gamma{}_{\alpha\beta}H_\gamma, \tag{4}$$

$$\{H,H_\alpha\}=C_\alpha H+(C^{\beta A}_\alpha P_A+B^\beta_\alpha)H_\beta. \tag{5}$$

The coefficients on the right-hand side of Eqs. (4) and (5)

 0022-2488/90/071723-10$03.00

(they are tensor fields on $\mathscr{M}$) are called structure functions.

The most general transformation that preserves the constraint hypersurface $\mathscr{C}$ reads

$$H' = e^{\Omega(Q)}H + (\Lambda^{\alpha A}P_A + \Lambda^\alpha)H_\alpha, \tag{6}$$

$$H_{\alpha'} = \Lambda^\beta_{\alpha'} H_\beta, \tag{7}$$

where $\Omega(Q)$, $\Lambda^\alpha(Q)$, and $\Lambda^\beta_{\alpha'}(Q)$ are arbitrary scalars and $\Lambda^{\alpha A}(Q)$ arbitrary vectors on $\mathscr{M}$ such that the transformation (7) is invertible:

$$\mathrm{Det}(\Lambda_\alpha{}^\beta(Q)) \neq 0, \quad \forall Q \in \mathscr{M}. \tag{8}$$

In I, we presented the Dirac constraint quantization of the system that is covariant under the transformations (6) and (7) of the constraints as well as under the point transformations (1) in the big phase space.

III. GEOMETRY OF CONFIGURATION SPACE-TIME

The constraints define a geometrical structure on the configuration space-time $\mathscr{M}$. Let us briefly describe this structure.

In I we showed that the vector fields $\partial_\alpha = \Phi^A_\alpha \partial_A$ are surface forming and we called the corresponding maximal surfaces "orbits." The tangent space to the orbit at a point Q of $\mathscr{M}$ is identical with $T_\parallel$. This is the geometrical structure in $\mathscr{M}$ determined by the linear constraints (3). The transformation (7) is equivalent to an invertible linear transformation of the vectors ∂_α, so the orbits are invariant with respect to (7). The transversal covector space, $T^*_\perp$ as a space of all covectors which annihilate Φ, is also invariant under (7).

To see what structure is determined by the quadratic constraint (2), we substitute (2) and (3) for the constraints into Eqs. (6) and (7), and thereby obtain the following transformation relations for the fields:

$$G'^{AB} = e^\Omega G^{AB} + \Lambda^{\alpha A}\Phi^B_\alpha + \Lambda^{\alpha B}\Phi^A_\alpha, \tag{9}$$

$$U'^A = e^\Omega U^A + \Lambda^\alpha \Phi^A_\alpha, \tag{10}$$

$$V' = e^\Omega V. \tag{11}$$

Hence, the fields G^{AB}, U^A, and V are not invariant by themselves; it is only the class $\{G^{AB}, U^A, V\}$ of the fields, whose elements are obtained by all transformations (9)–(11), that forms the geometric structure. The above transformations consist of the addition of arbitrary longitudinal terms as well as of the rescaling of all the fields by an arbitrary positive common factor. Observe that the signature of the metric G^{AB} on the transversal covector space is invariant with respect to the transformation (9), and it is thus a property of the whole class, whereas the signature of G^{AB} in the remaining directions is arbitrary.

In I, we saw that the Jacobi identity for the algebra (4) and (5) implies the following equation for C_α:

$$\{C_\alpha, H_\beta\} - \{C_\beta, H_\alpha\} = c^\gamma{}_{\alpha\beta} C_\gamma. \tag{12}$$

The structure functions C_α (called constraint cocycle) can thus be considered as nonholonomic components (in the frame Φ^A_α) of some form on each given orbit, and that form is closed. One can thus always transform C_α away, at least locally, by a suitable rescaling (6). The case in which all C_α vanish is particularly interesting. We will call it "the case of equidistant orbits," and use its (local) existence for proofs of some theorems.

Relation (4) implies that the fields G^{AB}, U^A, and V satisfy Eqs. (2.15)–(2.17) of I:

$$\mathscr{L}_\alpha G^{AB} = C_\alpha G^{AB} + C^{\beta A}_\alpha \Phi^B_\beta + C^{\beta B}_\alpha \Phi^A_\beta, \tag{13}$$

$$\mathscr{L}_\alpha U^A = C_\alpha U^A + B_\alpha{}^\beta \Phi^A_\beta, \tag{14}$$

$$\mathscr{L}_\alpha V = C_\alpha V; \tag{15}$$

here $\mathscr{L}_\alpha$ denotes the Lie derivative with respect to the vector field $\partial_\alpha = \Phi^A_\alpha \partial_A$, $\alpha = 1,\dots,\nu$. The geometrical meaning of Eqs. (13)–(15) can be described as follows. The vector field ∂_α defines an infinitesimal diffeomorphism, φ_α, along the orbits in $\mathscr{M}$. Let Q_1 be an arbitrary point and Q_2 be its image by φ_α; Q_1 and Q_2 are neighboring points on the same orbit. Let some particular fields G^{AB}, U^A, and V be given on $\mathscr{M}$ and let their values at Q_1 and Q_2 be G^{AB}_1, U^A_1, V_1 and G^{AB}_2, U^A_2, V_2, respectively. Then, the image of G^{AB}_1, U^A_1, V_1 by $\varphi_{\alpha *}$ lies in the class $\{G^{AB}_2, U^A_2, V_2\}$. We can say: the class field $\{G^{AB}(Q), U^A(Q), V(Q)\}$ defined on $\mathscr{M}$ by the constraints is "Lie-constant" along each orbit.

Suppose for a moment that the orbit space, $m = \mathscr{M}/$orbit, is a quotient manifold so that the projection π that sends each point of $\mathscr{M}$ to the orbit through that point is a submersion. The derivative of π at any given point Q of $\mathscr{M}$ defines a tensor algebra homeomorphism, $\pi_*(Q)$, of the algebra of purely contravariant tensors at Q to the corresponding algebra at $\pi(Q)$. The homeomorphism $\pi_*(Q)$ annihilates the ideal generated by the longitudinal vectors $\Phi^A_\alpha(Q)$, and satisfies the relation

$$\pi_*(\varphi_\alpha(Q))\varphi_{\alpha *} t = \pi_*(Q) t,$$

for any contravariant tensor t at Q. From Eqs. (13)–(15), it follows that

$$\begin{aligned}(\pi_*(Q_1)G^{AB}(Q_1), \pi_*(Q_1)U^A(Q_1), \pi_*(Q_1)V(Q_1))\\ = (\lambda\pi_*(Q_2)G^{AB}(Q_2), \lambda\pi_*(Q_2)U^A(Q_2),\\ \times\lambda\pi_*(Q_2)V(Q_2)),\end{aligned}$$

where λ is some number ($\lambda - 1$ is infinitesimal in this case). Thus, applying the map π_* to the fields G^{AB}, U^A, and V at each point Q of a given orbit, we obtain a whole subset of the conformal class, $\{g^{ab}, u^a, v\}$, of tensors at $\pi(Q)$. Each two elements of this class are related by the transformation

$$\{g'^{ab}, u'^a, v'\} = \{e^\omega g^{ab}, e^\omega u^a, e^\omega v\},$$

where ω is some real number. All metrics g^{ab} in the same class will be nondegenerate and will have the same signature that G^{AB} has on the space of transverse covectors. Under reasonable assumptions, the conformal classes defined in this way on m will be smooth in the sense that there will be a smooth representative field $\{g^{ab}, u^a, v\}$ on m. The geometry on m is invariant under the transformations (1), (6), and (7).

IV. TRANSVERSAL FIELDS

In Sec. III we have shown that the constraints induce two geometrical structures on the configuration space: a foliation by the orbits and a transversal conformal class of fields, $\{G^{AB}, U^A, V\}$. In I, we have chosen some particular representative of the class and constructed a corresponding differential operator—the operator constraint. Every such operator constraint must be invariant with respect to (1), and operator constraints obtained from different representatives must define the same quantum theory. In I, we introduced the so-called transversal distribution $T_\perp$,

$$T_\perp + T_\parallel = T_Q(\mathscr{M}).$$

There are clearly many such distributions, some of them integrable and some of them not. We will only require the distribution to be *differentiable*. In I, $T_\perp$ was associated with the metric G^{AB} defined by the fixed representative, and it was shown that the resulting quantum theory does not depend on its choice. Here, we select the transversal distribution $T_\perp$ independently of G^{AB}. The final result still does not depend on $T_\perp$, because we can always find a metric that is associated with $T_\perp$ in the way assumed in I.

We can consider the spaces $T_\perp$ and $T_\perp^*$ as being dual to each other. Indeed, let x_A be an arbitrary covector from $T_\perp^*$; then, x_A defines a form on $T_\perp$ by

$$\langle x_A, y^A \rangle = x_A y^A,$$

where y^A is an arbitrary vector from $T_\perp$. Moreover, there is no nonzero vector in $T_\perp$ that is annihilated by all covectors from $T_\perp^*$, because such a vector had to lie in $T_\parallel$.

We introduce the projection tensor $g^A{}_B$ associated with the distribution $T_\perp$ that satisfies the equations

$$g^A{}_B y^B = y^A, \quad \forall y^A \in T_\perp, \tag{16}$$

$$g^A{}_B \Phi^B = 0, \quad \forall \Phi^B \in T_\parallel, \tag{17}$$

and

$$g^A{}_B x_A = x_B, \quad \forall x_A \in T_\perp^*. \tag{18}$$

The requirement that the distribution $T_\perp$ be differentiable can be succinctly expressed as the requirement that the projector $g^A{}_B$ be a differentiable tensor field on $\mathscr{M}$. Observe that not all the derivatives of $g^A{}_B$ are independent, because

$$g^A{}_C g^D{}_B \partial_E g^C{}_D = 0. \tag{19}$$

Any metric G^{AB} and any vector potential U^A from the given class define a transversal metric, g^{AB}, and a transversal vector potential, u^A, by

$$g^{AB} = G^{CD} g^A{}_C g^B{}_D,$$

$$u^A = U^B g^A{}_B. \tag{20}$$

The quantities g^{AB} and u^A are invariant with respect to adding longitudinal terms to G^{AB} and U^A, but not with respect to a rescaling. Observe that g^{AB} and u^A lie in the class $\{G^{AB}, U^A, V\}$ because they differ from G^{AB} and U^A merely by longitudinal terms. In this way, a particular choice of representatives of the class $\{G^{AB}, U^A, V\}$ is associated with a fixed distribution. However, any representative of the form $\{g^{AB}, u^A, V\}$, i.e., one that is associated with some distribution, is special in two respects: (i) the metric g^{AB} is maximally degenerate (its zero space is ν-dimensional), and (ii) both g^{AB} and u^A lie in the same proper tensor subalgebra $\mathscr{T}(T_\perp, T_\perp^*)$, namely that generated by $T_\perp$ and $T_\perp^*$.

Since the metrics G^{AB} and g^{AB} related by Eq. (20) differ only by longitudinal terms, they induce the same metric on $T_\perp^*$; this metric is nondegenerate and hence

$$g^{AB} x_B \neq 0$$

for any nonzero $x_A \in T_\perp^*$. Further, Eq. (20) implies

$$g^{AB} x_B = G^{CD} g^A{}_C g^B{}_D x_B$$

$$= g^A{}_C (G^{CD} g^B{}_D x_B) \in T_\perp.$$

Thus, g^{AB} defines a linear isomorphism from $T_\perp^*$ to $T_\perp$. Let us denote by g_{AB} a tensor determined by the following properties of the associated map of vectors into covectors: on $T_\perp$, the map coincides with the inverse of the linear isomorphism g^{AB}; on $T_\parallel$, the map is zero. The covariant transversal metric g_{AB} satisfies

$$g_{AB} g^{BC} = g_A{}^C. \tag{21}$$

We also define

$$u_A = g_{AB} u^B = g_{AB} U^B.$$

After these preliminaries, we are ready to introduce an important concept of transversal tensor.

Definition 1: A tensor $t^{A\cdots B}{}_{C\cdots D}$ is called a "transversal tensor," if it satisfies the relation

$$t^{A\cdots B}{}_{C\cdots D} = t^{P\cdots Q}{}_{R\cdots S} g^A{}_P \cdots g^B{}_Q g^R{}_C \cdots g^S{}_D.$$

We see that g^{AB}, g_{AB}, u^A, u_A, and $g_A{}^B$ are transversal tensors. We shall denote the transversal tensors by lower case letters. Transversal tensors can be considered as elements of the tensor algebra $\mathscr{T}(T_\perp, T_\perp^*)$. We shall now construct the corresponding tensor algebra isomorphism as follows.

Let x_a^A be any basis of $T_\perp$ and let x_A^a be the dual basis in $T_\perp^*$ (called a "transversal frame"):

$$x_a^A x_A^b = \delta_a^b, \quad x_a^A x_B^a = g_B^A.$$

Then, any transversal tensor $\mathbf{t}$ determines a tensor in $\mathscr{T}(T_\perp, T_\perp^*)$ defined by

$$t^{a\cdots b}{}_{c\cdots d} = t^{A\cdots B}{}_{C\cdots D} x_A^a \cdots x_B^b x_c^C \cdots x_d^D.$$

The basis x_a^A can be chosen orthonormal ("transversal orthonormal frame"):

$$g_{AB} x_a^A x_b^B = \eta_{ab}, \quad \forall a,b = 0,\ldots,n,$$

where

$$\eta_{00} = +1, \quad \eta_{11} = \cdots = \eta_{nn} = -1.$$

Any tensor $\mathbf{T}$ can be projected into a unique transversal tensor $\mathbf{t}$:

$$t^{A\cdots B}{}_{C\cdots D} = g_P^A \cdots g_R^B g_C^S \cdots g_D^T T^{P\cdots R}{}_{S\cdots T}.$$

We will abbreviate this operation by the symbol P:

$$t^{A\cdots B}{}_{C\cdots D} = P T^{A\cdots B}{}_{C\cdots D}.$$

The concepts introduced in this section are vital.

V. LIE DERIVATIVE OF TRANSVERSAL FIELDS

In the formalism that we are going to develop, the Lie derivative of transversal fields with respect to the longitudinal vector fields will play an important role. To begin with, we prove the following.

Lemma 1: Let $t_{A\cdots B}$ be a transversal tensor field of type $(0,q)$. Then, $(\mathscr{L}_\alpha t)_{A\cdots B}$ is again a transversal tensor field.

Proof: The requirement that a tensor field be transversal can be expressed as follows:

$$t_{A\cdots B}\Phi^A_\beta = \cdots = t_{A\cdots B}\Phi^B_\beta = 0, \quad \forall\beta.$$

The Lie derivatives of these equations lead to

$$(\mathscr{L}_\alpha t_{A\cdots B})\Phi^A_\beta = -t_{A\cdots B}\mathscr{L}_\alpha\Phi^A_\beta$$
$$= -c^\gamma{}_{\alpha\beta}t_{A\cdots B}\Phi^A_\gamma = 0.$$

Q. E. D.

The Lie derivative of transversal tensor fields that have some contravariant indices will have a nonzero longitudinal part. We need to know an explicit form of this part for some tensor fields. Let us choose a basis of ν longitudinal vector fields Φ^A_α and denote the dual basis by Φ^α_A:

$$\Phi^A_\alpha\Phi^\beta_A = \delta^\beta_\alpha, \quad \Phi^A_\alpha\Phi^\alpha_B = \delta^A_B - g^A{}_B,$$
$$\Phi^\beta_A g^A{}_B = 0.$$

It is useful to introduce an abbreviation:

$$c_{A\alpha}{}^\beta = \Phi^\beta_B\mathscr{L}_\alpha g^B{}_A. \tag{22}$$

For a fixed longitudinal basis Φ^A_α, $c_{A\alpha}{}^\beta$ is a transversal covector. The transformation (7) of the longitudinal basis,

$$\Phi'^A_\alpha = \Lambda_\alpha{}^\beta\Phi^A_\beta, \tag{23}$$

induces a transformation of $c_{A\alpha}{}^\beta$:

$$c'_{A\alpha}{}^\beta = \Lambda^{-1}{}_\gamma{}^\beta(\Lambda_\alpha{}^\delta c_{A\delta}{}^\gamma - g^B{}_A\partial_B\Lambda_\alpha{}^\delta). \tag{24}$$

The quantities $c_{A\alpha}{}^\beta$ help us to define a kind of covariant derivative of longitudinal covectors in transversal directions. This derivative will appear in some important formulas. Let Ξ_α transform as

$$\Xi'_\alpha = \Lambda_\alpha{}^\beta\Xi_\beta$$

under the change (23) of the basis. Then, Eq. (24) implies immediately that $\Psi_{A\alpha}$ defined by

$$\Psi_{A\alpha} = g_A{}^B\partial_B\Xi_\alpha + c_{A\alpha}{}^\beta\Xi_\beta$$

transforms as

$$\Psi'_{A\alpha} = \Lambda_\alpha{}^\beta\Psi_{A\beta}.$$

As a result, $\Psi_{A\alpha}$ can be considered as a kind of a covariant derivative of Ξ_α.

The Lie derivatives of the transversal metrics and vector potentials are given by the following.

Lemma 2: It holds that

$$\mathscr{L}_\alpha u^A = c_\alpha u^A - c_\alpha{}^\beta\Phi^A_\beta, \tag{25}$$

$$\mathscr{L}_\alpha g^{AB} = c_\alpha g^{AB} + c^A{}_\alpha{}^\beta\Phi^B_\beta + c^B{}_\alpha{}^\beta\Phi^A_\beta, \tag{26}$$

$$\mathscr{L}_\alpha g^A{}_B = c_{B\alpha}{}^\beta\Phi^A_\beta, \tag{27}$$

$$\mathscr{L}_\alpha u_A = 0, \tag{28}$$

$$\mathscr{L}_\alpha g_{AB} = -c_\alpha g_{AB}, \tag{29}$$

where

$$c_\alpha = C_\alpha, \quad c_\alpha{}^\beta = u^A c_{A\alpha}{}^\beta, \quad c^A{}_\alpha{}^\beta = g^{AB}c_{B\alpha}{}^\beta. \tag{30}$$

Proof: Choose a transversal frame field x^A_a and denote the dual frame by x^a_A. The Lie derivative of x^A_a can be decomposed into its transversal and longitudinal parts:

$$\mathscr{L}_\alpha x^A_a = t_{\alpha a}{}^b x^A_b + l_{\alpha a}{}^\beta\Phi^A_\beta. \tag{31}$$

We have, on the one hand,

$$\mathscr{L}_\alpha(x^A_a x^b_A) = 0,$$

and on the other hand,

$$\mathscr{L}_\alpha(x^A_a x^b_A) = x^A_a\mathscr{L}_\alpha x^b_A + t_{\alpha a}{}^b.$$

According to Lemma 1, $\mathscr{L}_\alpha x^b_A$ must be transversal; therefore

$$\mathscr{L}_\alpha x^b_A = -t_{\alpha a}{}^b x^a_A. \tag{32}$$

Let us calculate the Lie derivative of $g^A{}_B$:

$$\mathscr{L}_\alpha g^A{}_B = \mathscr{L}_\alpha(x^A_a x^a_B)$$
$$= (t_{\alpha a}{}^b x^A_b + l_{\alpha a}{}^\beta\Phi^A_\beta)x^a_B + x^A_b(-t_{\alpha a}{}^b x^a_B)$$
$$= l_{\alpha a}{}^\beta\Phi^A_\beta x^a_B.$$

A comparison with (22) yields

$$l_{\alpha a}{}^\beta = c_{A\alpha}{}^\beta x^A_a. \tag{33}$$

For g^{AB}, the above relation gives

$$\mathscr{L}_\alpha g^{AB}$$
$$= \mathscr{L}_\alpha(\eta^{ab}x^A_a x^B_b)$$
$$= \eta^{ab}(t_{\alpha a}{}^c x^A_c + l_{\alpha a}{}^\beta\Phi^A_\beta)x^B_b$$
$$+ \eta^{ab}x^A_a(t_{\alpha b}{}^c x^B_c + l_{\alpha b}{}^\beta\Phi^B_\beta)$$
$$= \eta^{ab}t_{\alpha a}{}^c(x^A_c x^B_b + x^A_b x^B_c) + c^A{}_\alpha{}^\beta\Phi^B_\beta + c^B{}_\alpha{}^\beta\Phi^A_\beta.$$

As g^{AB} differs from G^{AB} only by longitudinal terms, the transversal parts of their Lie derivatives coincide. This proves the relation (26) and implies

$$t_\alpha{}^{ab}(x^A_a x^B_b + x^A_b x^B_a) = c_\alpha g^{AB}.$$

To find the Lie derivative of g_{AB}, we calculate

$$\mathscr{L}_\alpha g^A{}_B = \mathscr{L}_\alpha(g^{AC}g_{BC})$$
$$= g^{AC}\mathscr{L}_\alpha g_{BC} + (c_\alpha g^{AC} + c^A{}_\alpha{}^\beta\Phi^C_\beta + c^C{}_\alpha{}^\beta\Phi^A_\beta)g_{BC}.$$

Since this expression can contain only longitudinal terms and because $\mathscr{L}_\alpha g_{BC}$ is purely transversal, we obtain Eq. (29). The proof of the relations (25) and (28) is analogous. Q.E.D.

VI. TRANSVERSAL COVARIANT DERIVATIVE

We need some sort of covariant derivative to construct covariant and conformally covariant differential operators. In particular, for the second task we need some sort of scalar curvature. However, the metric is in general degenerate on the longitudinal space, and hence it determines a covariant derivative of transversal fields only in the transversal directions. It will turn out that we need to differentiate only transversal tensor fields; covariant derivatives of more general tensor fields thus need not be introduced. The derivative of the transverse fields in longitudinal directions will be specified by convenience.

Definition 2: For any vector field $\mathbf{X}$ on $\mathscr{M}$, the *transversal covariant derivative* $\nabla_{\mathbf{X}}$ is a map with the following properties:

(1) $\nabla_{\mathbf{X}}$ maps transversal tensor fields into transversal tensor fields of the same type.

(2) $\nabla_{\mathbf{X}}$ is linear in $\mathbf{X}$, linear in its argument, and it satisfies the Leibniz rule in its argument.

(3) For any two transversal vector fields $\mathbf{u}$ and $\mathbf{v}$,

(a) $\nabla_{\mathbf{u}} g_{AB} = 0$,

(b) $\nabla_{\mathbf{u}}\mathbf{v} - \nabla_{\mathbf{v}}\mathbf{u} = P[\mathbf{u},\mathbf{v}]$,

where $[\mathbf{u},\mathbf{v}]$ denotes the Lie bracket of the vector fields $\mathbf{u}$ and $\mathbf{v}$.

(4) For any longitudinal vector field Ψ and any transversal tensor field $\mathbf{t}$ of type (p,q),

$$\nabla_{\Psi}\mathbf{t} = P\mathscr{L}_{\Psi}\mathbf{t} - \tfrac{1}{2}(p-q)c_{\Psi}\mathbf{t},$$

where

$$c_{\Psi} = C_{\alpha}\Psi^{\alpha},$$

and Ψ^{α} is defined by $\Psi^{A} = \Psi^{\alpha}\Phi^{A}_{\alpha}$.

Let us discuss this definition. The covariant derivative can be described by the Ricci rotation coefficients, $\gamma^{a}{}_{bA}$, with respect to a given transversal frame x^{A}_{a};

$$X^{A}\gamma^{a}{}_{bA} = x^{a}_{B}\nabla_{X}x^{B}_{b}, \quad \forall \mathbf{X}. \tag{34}$$

If we know $\gamma^{a}{}_{bA}$, we can calculate the covariant derivative of any transversal tensor field $\mathbf{t}$ from the standard formula

$$\begin{aligned}\nabla_{\mathbf{X}} t^{A\cdots B}{}_{C\cdots D} &= X^{E}x^{A}_{a}\cdots x^{B}_{b}x^{c}_{C}\cdots x^{d}_{D}(\partial_{E}t^{a\cdots b}{}_{c\cdots d} \\ &\quad + \gamma^{a}{}_{eE}t^{e\cdots b}{}_{c\cdots d} + \cdots - \gamma^{e}{}_{cC}t^{a\cdots b}{}_{e\cdots d} - \cdots);\end{aligned} \tag{35}$$

here,

$$t^{a\cdots b}{}_{c\cdots d} = t^{A\cdots B}{}_{C\cdots D}x^{a}_{A}\cdots x^{b}_{B}x^{C}_{c}\cdots x^{D}_{d}.$$

[Equation (35) is equivalent to the requirements 1, 2, and 3.] The requirements 3a and 4 together with Eq. (29) imply that

$$\nabla_{\mathbf{X}} g_{AB} = 0, \quad \forall \mathbf{X}.$$

The covariant derivative of a transversal orthonormal frame is thus a rotation, i.e., the corresponding $\gamma_{b}{}^{a}{}_{A}$ satisfies

$$\gamma^{a}{}_{bA} = \eta^{ac}\gamma_{cbA},$$

where

$$\gamma_{abA} = -\gamma_{baA}. \tag{36}$$

If we introduce the notation

$$\gamma_{abc} = \gamma_{abA}x^{A}_{c}$$

and

$$\omega_{abc} = [\mathbf{x}_{a},\mathbf{x}_{b}]^{A}x_{cA}, \tag{37}$$

then the requirement 3b implies that

$$\omega_{abc} = \gamma_{cba} - \gamma_{cab} \tag{38}$$

and

$$\gamma_{bac} = \tfrac{1}{2}(\omega_{cab} + \omega_{bca} - \omega_{abc}). \tag{39}$$

We see that the requirements determine $\gamma^{a}{}_{bA}$ uniquely.

We observe that the transversal connection has, in general, a nonzero torsion. The torsion tensor, $T^{C}{}_{AB}$, is well defined for a *linear* connection $\nabla_{\mathbf{X}}$. The standard definition (Ref. 2, p. 133),

$$T^{C}{}_{AB}X^{A}Y^{B} = [\mathbf{X},\mathbf{Y}]^{C} - (\nabla_{\mathbf{X}}\mathbf{Y})^{C} + (\nabla_{\mathbf{Y}}\mathbf{X})^{C},$$

is in our case meaningful only for transversal vectors $\mathbf{X}$ and $\mathbf{Y}$; consequently, only the components $T^{C}{}_{AB}$ with transversal covariant indices make sense. According to requirement 3b,

$$T^{C}{}_{AB}x^{A}y^{B} = [\mathbf{x},\mathbf{y}]^{D}(\delta^{C}{}_{D} - g^{C}{}_{D}). \tag{40}$$

We see that $T^{C}{}_{AB}$ vanishes only if the distribution $T_{\perp}$ is integrable (holonomic).

So far, we have given the covariant derivative in terms of the Ricci rotation coefficients. It is useful to express it also by means of the transversal metric and the constraint cocycle. This is given by the following theorem.

Theorem 1: Let $\mathbf{t}$ be any transversal tensor field and let $\mathbf{X}$ be an arbitrary vector field. Then

$$\begin{aligned}\nabla_{\mathbf{X}} t^{A\cdots B}{}_{C\cdots D} &= P\partial_{\mathbf{X}} t^{A\cdots B}{}_{C\cdots D} + t^{E\cdots B}{}_{C\cdots D}\Gamma^{A}{}_{EF}X^{F} + \cdots \\ &\quad - t^{A\cdots B}{}_{E\cdots D}\Gamma^{E}{}_{CF}X^{F} - \cdots,\end{aligned} \tag{41}$$

where

$$\Gamma^{A}{}_{EF} = P\{^{A}{}_{EF}\} + g^{A}{}_{N}g^{M}{}_{E}g^{N}{}_{F,M} - \tfrac{1}{2}c_{E}g^{A}{}_{F} \tag{42}$$

and

$$\{^{A}{}_{EF}\} = \tfrac{1}{2}g^{AD}(g_{ED,F} + g_{FD,E} - g_{EF,D}). \tag{43}$$

Proof: Let us substitute

$$t^{a\cdots b}{}_{c\cdots d} = t^{A\cdots B}{}_{C\cdots D}x^{a}_{A}\cdots x^{b}_{B}x^{C}_{c}\cdots x^{D}_{d}$$

into Eq. (35). By differentiating, we obtain Eq. (41) with

$$\begin{aligned}\Gamma^{A}{}_{BC} &= g^{D}{}_{B}x^{A}_{a}x^{a}_{D,C} + \gamma^{A}{}_{BC} \\ &= -g^{A}{}_{D}x^{a}_{B}x^{D}_{a,C} + \gamma^{A}{}_{BC},\end{aligned} \tag{44}$$

and

$$\gamma^{A}{}_{BC} = x^{A}_{a}x^{b}_{B}\gamma^{a}{}_{bC}.$$

Let us set

$$\Gamma^{A}{}_{BC} = {}^{\perp}\Gamma^{A}{}_{BC} + {}^{\parallel}\Gamma^{A}{}_{BC},$$

where

$${}^{\perp}\Gamma^{A}{}_{BC} = g^{D}{}_{C}\Gamma^{A}{}_{BD}. \tag{45}$$

We have

$${}^{\perp}\Gamma^{A}{}_{BC} = {}^{\perp}\Gamma^{A}{}_{CB}.$$

Indeed, substituting for ${}^{\perp}\Gamma^{A}{}_{BC}$ from Eqs. (44) and (45), we obtain

$$\begin{aligned}{}^{\perp}\Gamma^{A}{}_{BC} - {}^{\perp}\Gamma^{A}{}_{CB} &= -g^{A}{}_{E}g^{F}{}_{C}x^{a}_{B}x^{E}_{a,F} + g^{F}{}_{C}\gamma^{A}{}_{BF} \\ &\quad + g^{A}{}_{E}g^{F}{}_{B}x^{a}_{C}x^{E}_{a,F} - g^{F}{}_{B}\gamma^{A}{}_{CF} \\ &= g^{A}{}_{E}x^{a}_{B}x^{b}_{C}([\mathbf{x}_{a},\mathbf{x}_{b}]^{E} \\ &\quad - x^{M}_{a}\nabla_{M}x^{E}_{b} + x^{M}_{b}\nabla_{M}x^{E}_{a}) \\ &= g^{A}{}_{E}x^{a}_{B}x^{b}_{C}S^{E}{}_{ab} = 0,\end{aligned}$$

because $S^{E}{}_{AB}$ is longitudinal in the upper index. The requirement 3a implies

$$g^{E}{}_{A}\nabla_{E}g_{BC} = 0,$$

which means that

$$Pg_{BC,A} = g_{EC}{}^{\perp}\Gamma^{E}{}_{AB} + g_{BE}{}^{\perp}\Gamma^{E}{}_{CA}.$$

As ${}^{1}\Gamma^{E}{}_{AB}$ is symmetric and transversal, we get

$$ {}^{1}\Gamma^{A}{}_{BC} = P\{{}^{A}{}_{BC}\}, $$

where the Christoffel symbol is calculated from the metric g_{AB} and g^{AB}. For the longitudinal component we have

$$ \begin{aligned} {}^{\parallel}\Gamma^{A}{}_{BC} &= \Phi^{\alpha}_{C}\Phi^{D}_{\alpha}g^{E}{}_{B}x^{A}_{a}x^{a}_{E,D} + \Phi^{\alpha}_{C}\gamma^{A}{}_{B\alpha} \\ &= \Phi^{\alpha}_{C}g^{E}{}_{B}x^{A}_{a}x^{a}_{E,\alpha} - \Phi^{\alpha}_{C}x^{A}_{a}\mathscr{L}_{\alpha}x^{a}_{B} - \tfrac{1}{2}g^{A}{}_{B}c_{C} \\ &= -\Phi^{\alpha}_{C}g^{E}{}_{B}g^{A}{}_{F}\Phi^{F}_{\alpha,E} - \tfrac{1}{2}g^{A}{}_{B}c_{C} \\ &= -g^{E}{}_{B}g^{A}{}_{F}\partial_{E}(\Phi^{\alpha}_{C}\Phi^{F}_{\alpha}) - \tfrac{1}{2}g^{A}{}_{B}c_{C} \\ &= g^{E}{}_{B}g^{A}{}_{F}g^{F}{}_{C,E} - \tfrac{1}{2}g^{A}{}_{B}c_{C}, \end{aligned} $$

because

$$ g^{F}{}_{C} + \Phi^{\alpha}_{C}\Phi^{F}_{\alpha} = \delta^{F}{}_{C}. $$

Adding both components of Γ, we obtain Eq. (42). Q.E.D.

VII. CURVATURE TENSOR

The curvature tensor, $R^{A}{}_{BCD}$, of a *linear* connection $\nabla_{\mathbf{X}}$ is defined by

$$ \nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}u^{A} - \nabla_{\mathbf{Y}}\nabla_{\mathbf{X}}u^{A} = R^{A}{}_{B\mathbf{XY}}u^{B} + \nabla_{[\mathbf{X},\mathbf{Y}]}u^{A}. \tag{46} $$

We can adopt this formula to our case: $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary vector fields, $\mathbf{u}$ is a transversal vector field, and

$$ R^{A}{}_{B\mathbf{XY}} = R^{A}{}_{BCD}X^{C}Y^{D}, $$

where $R^{A}{}_{BCD}$ is a tensor of type (1,3), transversal in the indices A and B. This means that the full curvature tensor can be defined for our kind of connection by Eq. (46). Notice that all covariant derivatives act on transversal tensor fields even if $\mathbf{X}$ and $\mathbf{Y}$ are not transversal; of course, the expressions like

$$ \nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}u^{A} $$

cannot then be rearranged into

$$ (\nabla_{\mathbf{X}}Y^{B})\nabla_{B}u^{A} + Y^{B}\nabla_{\mathbf{X}}\nabla_{B}u^{A}. $$

When expressing the covariant derivatives in Eq. (46) in terms of $\gamma^{a}{}_{bA}$, we find

$$ \begin{aligned} R^{a}{}_{b\mathbf{XY}} = X^{E}Y^{F}(&\partial_{E}\gamma^{a}{}_{bF} - \partial_{F}\gamma^{a}{}_{bE} \\ &+ \gamma^{a}{}_{dE}\gamma^{d}{}_{bF} - \gamma^{a}{}_{dF}\gamma^{d}{}_{bE}). \end{aligned} \tag{47} $$

Our next task is to express the components of the curvature tensor by means of the known fields: the metric and the structure functions.

Theorem 2: For any two longitudinal vectors Φ and Ψ,

$$ R^{a}{}_{b\Phi\Psi} = 0. \tag{48} $$

Proof: For any transversal covector $\mathbf{u}$, the requirement 5 yields

$$ \begin{aligned} \nabla_{\Phi}\nabla_{\Psi}\mathbf{u} &= \nabla_{\Phi}(\mathscr{L}_{\Psi}\mathbf{u} + \tfrac{1}{2}c_{\Psi}\mathbf{u}) \\ &= \mathscr{L}_{\Phi}\mathscr{L}_{\Psi}\mathbf{u} + \tfrac{1}{2}c_{\Phi}\mathscr{L}_{\Psi}\mathbf{u} + \tfrac{1}{2}(\partial_{\Phi}c_{\Psi})\mathbf{u} \\ &\quad + \tfrac{1}{2}c_{\Psi}\mathscr{L}_{\Phi}\mathbf{u} + \tfrac{1}{4}c_{\Psi}c_{\Phi}\mathbf{u}. \end{aligned} $$

Hence,

$$ \begin{aligned} &\nabla_{\Phi}\nabla_{\Psi}\mathbf{u} - \nabla_{\Psi}\nabla_{\Phi}\mathbf{u} \\ &\quad = (\mathscr{L}_{\Phi}\mathscr{L}_{\Psi} - \mathscr{L}_{\Psi}\mathscr{L}_{\Phi})\mathbf{u} + \tfrac{1}{2}(\partial_{\Phi}c_{\Psi} - \partial_{\Psi}c_{\Phi})\mathbf{u}. \end{aligned} $$

If we use Eq. (12), requirement 5, and the properties of the Lie derivative, we get

$$ \nabla_{\Phi}\nabla_{\Psi}\mathbf{u} - \nabla_{\Psi}\nabla_{\Phi}\mathbf{u} = \nabla_{[\Phi,\Psi]}\mathbf{u}. $$

Q.E.D.

The fact that the transversal connection is flat along the orbits enables us, among other things, to choose an orbit-parallel transversal frame.

Theorem 3: The following formula holds:

$$ \begin{aligned} R_{abc\alpha} = {}&\tfrac{1}{2}(\partial_{b}c_{\alpha} + c_{b\alpha}{}^{\beta}c_{\beta})g_{ac} \\ &- \tfrac{1}{2}(\partial_{a}c_{\alpha} + c_{a\alpha}{}^{\beta}c_{\beta})g_{bc}, \end{aligned} \tag{49} $$

where

$$ R_{abc\alpha} = g_{ad}R^{d}{}_{bEF}x^{E}_{c}\Phi^{F}_{\alpha}. $$

Proof: The right-hand side of Eq. (49) is a tensor—indeed, in the expression in the brackets, we have covariant derivatives of a longitudinal field in transversal directions. Let us choose an orbit-parallel transversal orthonormal frame x^{A}_{a} to calculate $R_{abc\alpha}$. Then,

$$ \gamma^{a}{}_{b\alpha} := \Phi^{E}_{\alpha}\gamma^{a}{}_{bE} = 0. \tag{50} $$

By using Eqs. (50) and (47), we obtain

$$ R^{a}{}_{bc\alpha} = [\Phi_{\alpha},\mathbf{x}_{c}]^{E}\gamma^{a}{}_{bE} - \gamma^{a}{}_{bc,\alpha}. $$

The Lie bracket can be written in the form

$$ [\Phi_{\alpha},\mathbf{x}_{a}] = \mathscr{L}_{\alpha}\mathbf{x}_{a} = \nabla_{\alpha}\mathbf{x}_{a} + \tfrac{1}{2}c_{\alpha}\mathbf{x}_{a} + c_{a\alpha}{}^{\beta}\Phi_{\beta}, $$

or

$$ \mathscr{L}_{\alpha}\mathbf{x}_{a} = \tfrac{1}{2}c_{\alpha}\mathbf{x}_{a} + c_{a\alpha}{}^{\beta}\Phi_{\beta}. \tag{51} $$

Equation (50) then yields

$$ R^{a}{}_{bc\alpha} = \tfrac{1}{2}c_{\alpha}\gamma^{a}{}_{bc} - \gamma^{a}{}_{bc,\alpha}. \tag{52} $$

By Eq. (39), $\gamma^{a}{}_{bc,\alpha}$ can be expressed in terms of $\omega_{ab}{}^{c}{}_{,\alpha}$. We obtain

$$ \begin{aligned} \omega_{ab}{}^{c}{}_{,\alpha} &= \mathscr{L}_{\alpha}([\mathbf{x}_{a},\mathbf{x}_{b}]^{A}x^{c}_{A}) \\ &= [\Phi_{\alpha},[\mathbf{x}_{a},\mathbf{x}_{b}]]^{A}x^{c}_{A} + [\mathbf{x}_{a},\mathbf{x}_{b}]^{A}\mathscr{L}_{\alpha}x^{c}_{A} \\ &= [\mathbf{x}_{a},\mathscr{L}_{\alpha}\mathbf{x}_{b}]^{A}x^{c}_{A} - [\mathbf{x}_{b},\mathscr{L}_{\alpha}\mathbf{x}_{a}]^{A}x^{c}_{A} \\ &\quad + [\mathbf{x}_{a},\mathbf{x}_{b}]^{A}\mathscr{L}_{\alpha}x^{c}_{A}. \end{aligned} $$

By comparing Eq. (51) with Eqs. (31) and (32), we conclude that

$$ \mathscr{L}_{\alpha}\mathbf{x}^{a} = -\tfrac{1}{2}c_{\alpha}\mathbf{x}^{a}. \tag{53} $$

The Lie derivatives are given by the expressions (51) and (53); from there,

$$ \begin{aligned} \omega_{ab}{}^{c}{}_{,\alpha} = {}&\tfrac{1}{2}c_{\alpha}\omega_{ab}{}^{c} + \tfrac{1}{2}(c_{\alpha,a} + c_{a\alpha}{}^{\beta}c_{\beta})\delta_{b}{}^{c} \\ &- \tfrac{1}{2}(c_{\alpha,b} + c_{b\alpha}{}^{\beta}c_{\beta})\delta_{a}{}^{c}. \end{aligned} $$

The formula (39) then gives Eq. (49). Q.E.D.

So far, we have calculated the mixed components of the curvature tensors by means of $\gamma^{a}{}_{bA}$, that is, by means of a transversal frame. The purely transversal components can be expressed in terms of the metric g_{AB} and the projection tensor $g^{A}{}_{B}$. Our next task is to derive these formulas.

Theorem 4: The transversal components of the curvature tensor can be expressed in terms of the transversal metric and projector as

$$PR^{A}{}_{BCD}=P\left[\tfrac{1}{2}g^{AT}(g_{BC,DT}+g_{DT,BC}-g_{BD,CT}-g_{CT,BD})\right.$$
$$+g^{T}{}_{S,R}(g^{AS}g^{R}{}_{C}\{_{TBD}\}-g^{AS}g^{R}{}_{D}\{_{TBC}\}-g^{AM}g^{R}{}_{C}g^{S}{}_{D}\{_{TMB}\}$$
$$+g^{AM}g^{R}{}_{D}g^{S}{}_{C}\{_{TMB}\}+g^{AM}g^{R}{}_{D}g^{S}{}_{B}\{_{TMC}\}-g^{AM}g^{R}{}_{C}g^{S}{}_{B}\{_{TMD}\})$$
$$\left.+\{^{R}{}_{BC}\}\{_{R}{}^{A}{}_{D}\}-\{^{R}{}_{BC}\}\{_{R}{}^{A}{}_{D}\}-\tfrac{1}{2}(c_{D,C}-c_{C,D})g^{A}{}_{B}\right]. \tag{54}$$

Proof: We calculate the curvature tensor directly from its definition (41). Let u^A be any transversal vector field, and let $\mathbf{X}$ and $\mathbf{Y}$ be any vector fields. Then

$$\nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}u^A-\nabla_{\mathbf{Y}}\nabla_{\mathbf{X}}u^A=[\mathbf{X},\mathbf{Y}]^B\nabla_B u^A+[g^A{}_S g^S{}_{R,E}u^R{}_{,F}-g^A{}_S g^S{}_{R,F}u^R{}_{,E}+g^A{}_S(\Gamma^S{}_{RF,E}-\Gamma^S{}_{RE,F})u^R+(\Gamma^A{}_{SE}\Gamma^S{}_{RF}-\Gamma^A{}_{SF}\Gamma^S{}_{RE})u^R]X^EY^F.$$

Since u^A is transversal, Eq. (19) yields

$$g^A{}_S g^S{}_{R,E}u^R=0.$$

By differentiating this equation with respect to Q^F, we get

$$g^A{}_S g^S{}_{R,E}u^R{}_{,F}=-g^A{}_{S,F}g^S{}_{R,E}u^R-g^A{}_S g^S{}_{R,EF}u^R.$$

In this way, we arrive at the formula

$$R^{A}{}_{BCD}=g^A{}_R g^S{}_B(\Gamma^R{}_{SF,E}-\Gamma^R{}_{SE,F}+g^R{}_{T,E}g^T{}_{S,F}-g^R{}_{T,F}g^T{}_{S,E})+\Gamma^A{}_{RE}\Gamma^R{}_{BF}-\Gamma^A{}_{RF}\Gamma^R{}_{BE}. \tag{55}$$

If we substitute for Γ from the expression (42) and then project, we obtain

$$PR^{A}{}_{BCD}=P[\partial_C(g^{AT}\{_{TMN}\}g^M{}_B g^N{}_D+g^A{}_M g^N{}_B g^M{}_{D,N}-\tfrac{1}{2}c_D g^A{}_B)+g^A{}_{T,C}g^T{}_{B,D}+\Gamma^A{}_{TC}\Gamma^T{}_{BD}-(CD)], \tag{56}$$

where the symbol (CD) denotes the preceding terms with the indices C and D interchanged. The right-hand side of Eq. (56) can be considerably simplified. To achieve this aim, we need some identities. First, the derivatives of the metric can be related to the Christoffel symbols by the familiar formula

$$g_{AB,C}=\{_{ABC}\}+\{_{BAC}\}. \tag{57}$$

Second, by differentiating Eq. (21), we obtain

$$g^A{}_E g^{EC}{}_{,B}=g^{AE}g^C{}_{E,B}-g^{AE}g^{CF}g_{EF,B}. \tag{58}$$

Third, Eq. (21) yields

$$P(g^A{}_{T,C}g^T{}_{B,D})=P(g^{AS}g_{ST,C}g^T{}_{B,D}+g_{ST}g^{AS}{}_{,C}g^T{}_{B,D}).$$

The last term vanishes by virtue of Eq. (19) because both indices T and B at $g^T{}_{B,D}$ are hit by a projector. This leads to the final identity,

$$P(g^A{}_{T,C}g^T{}_{B,D})=P[g^{AS}g^T{}_{B,D}(\{_{STC}\}+\{_{TSC}\})]. \tag{59}$$

By performing the derivative in Eq. (56) and using the identities (57), (58), and (59), we find

$$PR^{A}{}_{BCD}=P[g^{AT}\{_{TBD}\}_{,C}+g^T{}_{S,R}(g^{AS}g^R{}_C\{_{TBD}\}-g^{AM}g^R{}_C g^S{}_D\{_{TMB}\}+g^{AM}g^R{}_D g^S{}_B\{_{TMC}\})-\{^R{}_{BD}\}\{_R{}^A{}_C\}-\tfrac{1}{2}c_{D,C}g^A{}_B-(CD)].$$

The formula (54) then follows immediately. Q.E.D.

From the curvature tensor, we can derive the Ricci tensor

$$PR_{AB}=g^D{}_B R^C{}_{ACB},$$

and the curvature scalar

$$R=g^{AB}R^C{}_{ACB}.$$

From Eq. (54),

$$PR_{AB}=P\left[\tfrac{1}{2}g^{ST}(g_{BT,AS}+g_{AT,BS}-g_{AB,ST}-g_{ST,AB})+g^T{}_{S,R}(g^{RS}\{_{TAB}\}-g^R{}_B\{_{TA}{}^S\}-g^S{}_B\{_T{}^R{}_A\}\right.$$
$$+g^R{}_B\{_T{}^S{}_A\}-g^S{}_A\{_T{}^R{}_B\}+g^S{}_A g^R{}_B\{_T{}^M{}_M\})$$
$$\left.+\{^R{}_{AS}\}\{_R{}^S{}_B\}-\{^R{}_{AB}\}\{_R{}^S{}_S\}-\tfrac{1}{2}(c_{B,S}-c_{S,B})g^S{}_A\right], \tag{60}$$

and

$$R = g^{ST}g^{AB}(g_{BT,AS} - g_{AB,ST}) + 2g^T{}_{S,R}(g^{RS}\{{}_T{}^P{}_P\} - \{{}_T{}^{RS}\}) + \{{}_{RST}\}\{{}^{RST}\} - \{{}^{RT}{}_T\}\{{}_R{}^S{}_S\}. \tag{61}$$

The Bianchi identity for the curvature tensor of our connection cannot be written in the usual form with the torsion tensor (see, e.g., Ref. 2, p. 135), but we have the following.

Theorem 5: For any three vectors **X**, **Y**, and **Z**, the following identity holds:

$$\sigma\{\nabla_{\mathbf{Z}} R^A{}_{B\mathbf{X}\mathbf{Y}}\} = \sigma\{R^A{}_{B[\mathbf{X},\mathbf{Y}]\mathbf{Z}}\}, \tag{62}$$

where σ denotes the cyclic sum with respect to **X**, **Y**, and **Z**.

Proof: Apply $\nabla_{\mathbf{Z}}$ to Eq. (46) and take the cyclic sum with respect to **X**, **Y**, and **Z**:

$$\sigma\{(\nabla_{\mathbf{X}}\nabla_{\mathbf{Y}} - \nabla_{\mathbf{Y}}\nabla_{\mathbf{X}})\nabla_{\mathbf{Z}}u^A\} = \sigma\{(\nabla_{\mathbf{Z}}R^A{}_{B\mathbf{X}\mathbf{Y}})u^B + R^A{}_{B\mathbf{X}\mathbf{Y}}\nabla_{\mathbf{Z}}u^B + \nabla_{\mathbf{Z}}\nabla_{[\mathbf{X},\mathbf{Y}]}u^A\}.$$

By using Eq. (46) on the left-hand side, we find

$$\sigma\{(\nabla_{\mathbf{Z}}R^A{}_{B\mathbf{X}\mathbf{Y}})u^B + (\nabla_{\mathbf{Z}}\nabla_{[\mathbf{X},\mathbf{Y}]} - \nabla_{[\mathbf{X},\mathbf{Y}]}\nabla_{\mathbf{Z}})u^A\} = 0.$$

Another use of Eq. (46) leads to

$$\sigma\{(\nabla_{\mathbf{Z}}R^A{}_{B\mathbf{X}\mathbf{Y}} + R^A{}_{B\mathbf{Z}[\mathbf{X},\mathbf{Y}]})u^B + \nabla_{[\mathbf{Z},[\mathbf{X},\mathbf{Y}]]}u^A\} = 0.$$

However,

$$\sigma\{\nabla_{[\mathbf{Z},[\mathbf{X},\mathbf{Y}]]}u^A\} = 0$$

for any u^A because of the Jacobi identity. This yields Eq. (62). Q.E.D.

We write the Bianchi identity for a particular choice of vectors, namely,

$$\mathbf{X} = \mathbf{x}_c, \quad \mathbf{Y} = \mathbf{x}_d, \quad \mathbf{Z} = \Phi_\alpha,$$

where $\mathbf{x}_a$ is a transversal orthonormal frame parallel along the orbits:

$$\nabla_\alpha R^A{}_{Bcd} = -\nabla_c R^A{}_{Bd\alpha} + R^A{}_{B[\alpha,c]d} - (cd) + R^A{}_{B[c,d]\alpha}, \tag{63}$$

where

$$[\alpha,c] = [\Phi_\alpha,\mathbf{x}_c], \quad [c,d] = [\mathbf{x}_c,\mathbf{x}_d].$$

Equations (31), (33), and the requirement 5 imply

$$[\alpha,c] = \mathscr{L}_\alpha \mathbf{x}_c = \tfrac{1}{2}c_\alpha \mathbf{x}_c + c_{c\alpha}{}^\beta\Phi_\beta,$$

$$[c,d] = \omega_{cd}{}^e\mathbf{x}_e + \cdots = (\gamma^e{}_{dc} - \gamma^e{}_{cd})\mathbf{x}_e + \cdots,$$

where the points denote longitudinal terms. By substituting these relations into Eq. (63) and using Eq. (48), we get

$$\nabla_\alpha R^A{}_{Bcd} = -\nabla_c R^A{}_{Bd\alpha} + \tfrac{1}{2}c_\alpha R^A{}_{Bcd} - c_{c\alpha}{}^\beta R^A{}_{Bd\beta} + \gamma^e{}_{dc}R^A{}_{Be\alpha} - (cd).$$

If we multiply this relation by $x^a_A x^B_b$, take into account that the frame is parallel along the orbits, and use Eq. (34), we find that

$$\partial_\alpha R^a{}_{bcd} = c_\alpha R^a{}_{bcd} - (\tilde{\nabla}_c R^a{}_{bd\alpha} - (cd)), \tag{64}$$

where

$$\tilde{\nabla}_c R^a{}_{bd\alpha} = \partial_c R^a{}_{bd\alpha} + \gamma^a{}_{ec}R^e{}_{bd\alpha} - \gamma^e{}_{bc}R^a{}_{ed\alpha} - \gamma^e{}_{dc}R^a{}_{be\alpha} + c_{c\alpha}{}^\beta R^a{}_{bd\beta}.$$

Equation (64) is the desired relation. Observe that both sides of it transform as tensors if we rotate the frame by an amount that is constant along the orbits, or if we transform Φ_α by the transformation (7).

VIII. CONSTRAINT OPERATORS

If we assume that the orbit space is a quotient manifold, we have the conformal geometry with a nondegenerate metric, a vector field, and a scalar field. Such a geometry determines *uniquely* some differential operators in the following sense.

Lemma 1: There is a unique pair of differential operators $\mathbf{D}_1, \mathbf{D}_2$ on the space of scalar functions on $\mathscr{m}$ with the properties:

(A) $\mathbf{D}_1$ ($\mathbf{D}_2$) is formed solely from the field g^{ab} (g^{ab}, u^a) and its derivatives, each term having dimension -2 (-1), the leading part being

$$g^{ab}\partial_a\partial_b \quad (u^a\partial_a).$$

(B) $\mathbf{D}_1\varphi$ ($\mathbf{D}_2\varphi$) is a scalar for any scalar field φ.

(C) There is a number k for any dimension $n+1>1$ of the space $\mathscr{m}$ such that

$$\mathbf{D}_1{}'\varphi' = e^{(k+1)\Omega}\mathbf{D}_1\varphi, \quad \mathbf{D}_2{}'\varphi' = e^{(k+1)\Omega}\mathbf{D}_2\varphi,$$

where

$$\varphi' = e^{k\Omega}\varphi,$$

and $\mathbf{D}_1{}'$ ($\mathbf{D}_2{}'$) is formed from the Ω-rescaled field g^{ab} (g^{ab}, u^a).

Proof: From A and B, it follows that $\mathbf{D}_1$ ($\mathbf{D}_2$) must have the form

$$\mathbf{D}_1 = \nabla_a g^{ab}\nabla_b + g^a\nabla_a + g,$$

$$(\mathbf{D}_2 = u^a\nabla_a + u),$$

where g^a is a vector field formed from the metric and its first derivatives, g is a scalar field formed from the metric, its first and second derivatives, u is a scalar field formed from u^a, the metric, and their first derivatives, and ∇_a is the covariant derivative of the metric.

It is well known that g^a must vanish, and the only possible scalars g and u are given by

$$g = \xi R, \quad u = \zeta \operatorname{div}\mathbf{u},$$

where R is the curvature scalar of the metric, div **u** is the covariant divergence, and ξ and ζ are arbitrary reals. Then, the condition C uniquely determines the numbers k, ξ, and ζ:

$$k = (n-1)/4, \quad \xi = -(n-1)/4n, \quad \zeta = \tfrac{1}{2}. \tag{65}$$

Q.E.D.

This indicates that there is only one operator super-Hamiltonian **H** on $\mathscr{m}$ that is invariant with respect to all transformations (9)–(11) and whose coefficients are formed from the fields G^{AB}, U^A, and V, namely

$$\mathbf{H} = -\tfrac{1}{2}g^{ab}\nabla_a\nabla_b - \tfrac{1}{2}\xi R - iu^a\nabla_a - \tfrac{1}{2}i\nabla_a u^a + v.$$

What about the supermomenta constraints? On calculating the quotient manifold m, we have, in fact, solved these constraints "before quantizing." Indeed, the coordinates x^a on m are formed by n independent functions on $\mathscr{M}$ that are constant along the orbits: x^a form a complete solution of the system of differential equations

$$\Phi^A_\alpha \partial_A x^a = 0.$$

To find such a solution in interesting cases (especially if we are going to generalize our methods to field theories) is practically impossible. Moreover, the procedure described above is a mixture of the Dirac and covariant reduction methods, whereas we are interested in a pure Dirac method. Hence, its value is only in showing that the theory we are looking for is a unique one.

In I, we pursued the following strategy: we stayed in the "large" space and constructed the operator constraints from fixed classical representatives of the conformal transversal class $\{G^{AB}, U^A, V\}$ in such a way that the constructed quantum theory was independent of the choice of the representative. The theory of the transversal connection developed in the previous sections offers an alternative way.

The choice of the representative is associated with a particular transversal distribution $T_\perp$. Thus given the classical constraints and the distribution $T_\perp$, we calculate the transversal fields g^{AB}, u^A, and v (these are "algebraic" calculations). Then, we define the operator constraints as

$$\mathbf{H} = \tfrac{1}{2}\mathbf{L} + \mathbf{u} + v, \tag{66}$$

$$\mathbf{H}_\alpha = -i\Phi^A_\alpha \partial_A + ikc_\alpha, \tag{67}$$

where $\mathbf{L}$ is the transversal covariant Laplacian,

$$\mathbf{L} = -\nabla_A g^{AB}\nabla_B - \xi R,$$

and $\mathbf{u}$ is given by

$$\mathbf{u} = -(i/2)(\nabla_A u^A + u^A\nabla_A);$$

here, ∇_A is the transversal covariant derivative, R is the curvature scalar given by Eq. (61), and ξ is the factor (65). It follows that the above operators are scalar operators on scalar "wave functions" Ψ on $\mathscr{M}$, and that they have the conformal weight 1, when acting on functions with the confomal weight k of Eq. (65). Thus, the corresponding quantum theory will be covariant and conformally covariant. With the help of the developed formalism, we can show directly that the operator constraint algebra closes:

Theorem 6: It holds that

$$(1/i)[\mathbf{H}_\alpha, \mathbf{H}_\beta] = c^\gamma{}_{\alpha\beta}\mathbf{H}_\gamma, \tag{68}$$

where $c^\gamma{}_{\alpha\beta}$ are the structure functions of the classical algebra (4).

Proof: A simple calculation gives

$$[\mathbf{H}_\alpha, \mathbf{H}_\beta] = -[\Phi_\alpha, \Phi_\beta]^A\partial_A + k(\partial_\alpha C_\beta - \partial_\beta C_\alpha).$$

From Eqs. (5) and (12) we then obtain Eq. (68). Q.E.D.

To evaluate the other commutator is more laborious.

Theorem 7: It holds that

$$(1/i)[\mathbf{H}, \mathbf{H}_\alpha] = c_\alpha \mathbf{H} + \mathbf{c}_\alpha{}^\beta \mathbf{H}_\beta, \tag{69}$$

where

$$\mathbf{c}_\alpha{}^\beta = -(i/2)(c^A{}_\alpha{}^\beta\nabla_A + \nabla_A c^A{}_\alpha{}^\beta + c^A{}_\alpha{}^\gamma c^A{}_\gamma{}^\beta) + c_\alpha{}^\beta, \tag{70}$$

and c_α, $c^A{}_\alpha{}^\beta$, and $c_\alpha{}^\beta$ are the structure functions of the classical algebra given by Eqs. (30) and (22).

Proof: As the first step, we prove the relation

$$(1/i)[\mathbf{u}, \mathbf{H}_\alpha] = c_\alpha \mathbf{u} - c_\alpha{}^\beta \mathbf{H}_\beta. \tag{71}$$

A straightforward calculation yields

$$[\mathbf{u}, \mathbf{H}_\alpha] = [\Phi_\alpha, \mathbf{u}]^A\partial_A + ku^A(\partial_A c_\alpha) + (1/2)\partial_\alpha(\nabla_A u^A).$$

By substituting (25) for the Lie bracket, we get

$$(1/i)[\mathbf{u}, \mathbf{H}_\alpha] = c_\alpha \mathbf{u} - c_\alpha{}^\beta \mathbf{H}_\beta + \mathscr{A},$$

where

$$i\mathscr{A} = \tfrac{1}{2}\partial_\alpha(\nabla_A u^A) - \tfrac{1}{2}c_\alpha(\nabla_A u^A) + k(u^A\partial_A c_\alpha - c_\alpha{}^\beta c_\beta).$$

To show that $\mathscr{A} = 0$, we calculate $\partial_\alpha(\nabla_A u^A)$:

$$\partial_\alpha(\nabla_A u^A) = \nabla_\alpha(\nabla_A u^A) = \nabla_A\nabla_\alpha u^A + [\nabla_\alpha, \nabla_A]u^A.$$

We again use an orbit-parallel transversal orthonormal frame x^A_a. The divergence of any transversal vector v^A is given by

$$\nabla_A v^A = (\nabla_a v^A)x^a_A.$$

The requirement 4, Eq. (25), and the definition (46) of the curvature tensor imply

$$\partial_\alpha(\nabla_A u^A) = \nabla_A(\tfrac{1}{2}c_\alpha u^A) + [\Phi_\alpha, \mathbf{x}_a]^B(\nabla_B u^A)x^a_A - R^A{}_{BA\alpha}u^B.$$

Equation (49) gives

$$R^A{}_{BA\alpha} = x^a_B R^b{}_{ab\alpha} = \tfrac{1}{2}(n-1)(\partial_B c_\alpha + c_{B\alpha}{}^\beta c_\beta). \tag{72}$$

Using this and Eq. (51), we find

$$\partial_\alpha(\nabla_A u^A) = c_\alpha(\nabla_A u^A) + \tfrac{1}{2}(2-n)(\partial_A c_\alpha)u^A + c_{A\alpha}{}^\beta(\nabla_\beta u^A + \tfrac{1}{2}(1-n)c_\beta u^A).$$

Requirement 4 and Eq. (25) lead to the relation

$$\nabla_\beta u^A = \tfrac{1}{2}c_\alpha u^A,$$

and Eq. (30) finally yields

$$\partial_\alpha(\nabla_A u^A) = c_\alpha(\nabla_A u^A) + 2k(u^A\partial_A c_\alpha - c_\alpha{}^\beta c_\beta),$$

which is equivalent to $\mathscr{A} = 0$.

As the second step, we prove the relation

$$(1/i)[\mathbf{L}, \mathbf{H}_\alpha] = c_\alpha\mathbf{L} + 2(-ic^A{}_\alpha{}^\beta\nabla_A - (i/2)(\nabla_A c^A{}_\alpha{}^\beta) - (i/2)c^A{}_\alpha{}^\gamma c_{A\gamma}{}^\beta)\mathbf{H}_\beta. \tag{73}$$

By substituting for the operators, we obtain

$$[\mathbf{L}, \mathbf{H}_\alpha]\psi = -i\nabla_\alpha\nabla_A g^{AB}\nabla_B\psi + i\nabla_A g^{AB}\nabla_B\nabla_\alpha\psi - i\xi(\partial_\alpha R)\psi - ik(c_\alpha\nabla_A g^{AB}\nabla_B\psi - \nabla_A g^{AB}(\nabla_B c_\alpha\psi)).$$

We employ again an orbit-parallel transversal orthonormal frame. With the shorthand $\widetilde{\nabla}_A$ introduced in Sec. VII we can write

$$\nabla_A g^{AB}\nabla_B\psi = \widetilde{\nabla}_a\eta^{ab}\widetilde{\nabla}_b\psi = \eta^{ab}\widetilde{\nabla}_a\widetilde{\nabla}_b\psi$$

and

$$\tilde{\nabla}_\alpha \tilde{\nabla}_a v^b - \tilde{\nabla}_a \tilde{\nabla}_\alpha v^b = [\Phi_\alpha, \mathbf{x}_a]^B \tilde{\nabla}_B v^b + R^b{}_{c\alpha a} v^c.$$

By commuting the covariant derivatives we see that

$$\begin{aligned}
&\eta^{ab}\tilde{\nabla}_\alpha \tilde{\nabla}_a \tilde{\nabla}_b \psi - \eta^{ab}\tilde{\nabla}_a \tilde{\nabla}_b \tilde{\nabla}_\alpha \psi \\
&\quad = c_\alpha \eta^{ab}\tilde{\nabla}_a \tilde{\nabla}_b \psi + 2(c^a{}_\alpha{}^\beta \tilde{\nabla}_a + \tfrac{1}{2}\tilde{\nabla}_a c^a{}_\alpha{}^\beta \\
&\qquad + \tfrac{1}{2} c^a{}_\alpha{}^\gamma c^a{}_\gamma{}^\beta)\mathbf{H}_\beta \psi + 2k\eta^{ab}(\partial_a c_\alpha)(\partial_b \psi) \\
&\qquad - k(2c^a{}_\alpha{}^\beta \partial_a c_\beta + (\tilde{\nabla}_a c^a{}_\alpha{}^\beta)c_\beta + c^a{}_\alpha{}^\gamma c_{a\gamma}{}^\beta c_\beta)\psi,
\end{aligned}$$

and similarly,

$$\begin{aligned}
&kc_\alpha \eta^{ab}\tilde{\nabla}_a \tilde{\nabla}_b \psi - k\eta^{ab}\tilde{\nabla}_a(\tilde{\nabla}_b c_\alpha \psi) \\
&\quad = -k(\eta^{ab}\tilde{\nabla}_a \tilde{\nabla}_b c_\alpha)\psi - 2k\eta^{ab}(\partial_a c_\alpha)(\partial_b \psi).
\end{aligned}$$

Hence we have

$$\begin{aligned}
[\mathbf{L},\mathbf{H}_\alpha] = ic_\alpha \mathbf{L} + 2i(&-ic^A{}_\alpha{}^\beta \nabla_A - (i/2)(\nabla_A c^A{}_\alpha{}^\beta) \\
&- (i/2)c^A{}_\alpha{}^\gamma c_{A\gamma}{}^\beta)\mathbf{H}_\beta - i\mathscr{B},
\end{aligned}$$

where

$$\begin{aligned}
\mathscr{B} = \xi\partial_\alpha R - \xi c_\alpha R - k(&\eta^{ab}\tilde{\nabla}_a \tilde{\nabla}_b c_\alpha + 2c^a{}_\alpha{}^\beta \tilde{\nabla}_a c_\beta \\
&+ \tilde{\nabla}_a c^a{}_\alpha{}^\beta c_\beta + c^a{}_\alpha{}^\gamma c_{a\gamma}{}^\beta c_\beta).
\end{aligned}$$

However, $\mathscr{B} = 0$ due to Eq. (64). Q.E.D.

The transversal affine connection is thus useful not only in revealing the geometrical structure underlying the constraints. It also provides us with the mathematical formalism of covariant derivatives, their commutators, the curvature tensor, and the Bianchi identities that facilitates a direct calculation of the commutators of the operator constraints.

ACKNOWLEDGMENTS

We are grateful to Peter Thomi for checking the equations.

This work was partially supported by Tomalla–Stiftung and Schweizerischer Nationalfonds, and by NFS grants PHY85-03653 and PHY89-07937 to the University of Utah.

[1] P. Hajicek and K. V. Kuchar, Phys. Rev. D **41**, 1091 (1990).

[2] S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry* (Interscience, New York, 1963), Vol. 1.

The covariant linear oscillator and generalized realization of the dynamical SU(1,1) symmetry algebra

E. D. Kagramanov
Institute of Physics, Academy of Science of Azerbaijan, pr. Narimanova 33, 370073 Baku, USSR
R. M. Mir-Kasimov[a)]
International Center for Theoretical Physics, Trieste, Italy
Sh. M. Nagiyev
Institute of Physics, Academy of Science of Azerbaijan, pr. Narimanova 33, 370073 Baku, USSR

(Received 26 April 1989; accepted for publication 17 January 1990)

An exactly solvable problem for the finite-difference Schrödinger equation in the relativistic configurational space is considered. The appropriate finite-difference generalization of the factorization method is developed. The theory of new special functions "the relativistic Hermite polynomials," in which the solutions are expressed, is constructed.

I. INTRODUCTION: THE RELATIVISTIC CONFIGURATIONAL SPACE AND RELATIVISTIC QUANTUM MECHANICS

The concept of the relativistic configurational **r** space is based on the Fourier expansion over the "relativistic plane waves"

$$\langle \mathbf{r}|\mathbf{p}\rangle = ((p_0 - \mathbf{pn})/mc)^{-1 - ir(mc/\hbar)},$$
$$\mathbf{r} = r\mathbf{n}, \quad \mathbf{n} = 1, \quad 0 < r < \infty, \quad p_0 = \sqrt{m^2c^2 + \mathbf{p}^2}, \quad (1.1)$$

instead of the usual plane waves $e^{i(\mathbf{pr}/\hbar)}$ (Ref. 1; see also Ref. 2). The variable r is relativistic invariant and can be expressed in terms of eigenvalues of the Casimir operator of the Lorentz group $C = \mathbf{N}^2 - \mathbf{L}^2$ ($\mathbf{N}$,$\mathbf{L}$ are the boost and rotation generators):

$$C = (\hbar/mc)^2 + r^2, \quad (1.2)$$

where (1.1) is the generating function for the matrix elements of the principal series of the unitary irreducible Lorentz group.[3] In the nonrelativistic limit

$$|\mathbf{p}| \ll mc, \quad r \gg \hbar/mc, \quad (1.3)$$

the function $\langle \mathbf{r}|\mathbf{p}\rangle$ goes over into the usual plane wave

$$\langle \mathbf{r}|\mathbf{p}\rangle \to e^{i(\mathbf{pr}/\hbar)}. \quad (1.4)$$

The concept of the **r** space has firstly been introduced in the context of the quasipotential approach in the relativistic two-body problem[4–9] (the detailed list of references can be found in Ref. 7).

The quasipotential approach in the r representation possesses many features of the description of the interaction of two (many) relativistic particles (extended objects) via the action at a distance.[10] The quasipotential equation for the relativistic wave function $\Psi(\mathbf{p})$ has the form

$$\Psi(\mathbf{p}) = (2\pi)^3\delta(\mathbf{p}(-)\mathbf{q}) + \frac{1}{(2\pi)^3} G_q(p) \times \int V(\mathbf{p},\mathbf{k};E_q)\Psi(\mathbf{k})d\Omega_k, \quad (1.5)$$

where

$$G_q(p) = (2q_0 - 2p_0 + i\epsilon)^{-1},$$
$$\delta(\mathbf{p}(-)\mathbf{q}) = \sqrt{1 + \mathbf{q}^2/m^2c^2}\,\delta(\mathbf{p} - \mathbf{q}), \quad (1.6)$$
$$d\Omega_k = \frac{d\mathbf{k}}{\sqrt{1 + \mathbf{k}^2/m^2c^2}}.$$

The integration is carried over the mass shell of the particle with mass m, i.e., over the upper sheet of the hyperboloid

$$p_0^2 - \mathbf{p}^2 = m^2c^2 \quad (1.7)$$

(the p space of Lobachevsky).

Equation (1.5) has the absolute character with respect to the geometry of the momentum space, i.e., formally it does not differ from the nonrelativistic Schrödinger equation. We can derive Eq. (1.5) substituting the relativistic (non-Euclidean) expressions for the energy, volume element, and δ function by their nonrelativistic (Euclidean) analogs:

$$E_q = \mathbf{q}^2/2m \to q_0 = \sqrt{\mathbf{q}^2 + m^2c^2},$$
$$d\mathbf{k} \to d\Omega_k = d\mathbf{k}/\sqrt{1 + \mathbf{k}^2/m^2c^2}, \quad (1.8)$$
$$\delta(\mathbf{p} - \mathbf{q}) \to \delta(\mathbf{p}(-)\mathbf{q}).$$

As a consequence of this geometrical treatment, application of the Fourier transformation on the Lorentz group becomes natural. After performing this transformation in Eq. (1.5), we obtain the Schrödinger equation with the local potential in the relativistic **r** space

$$(H_0 + V(\mathbf{r}) - 2q_0)\Psi(\mathbf{r}) = 0. \quad (1.9)$$

The Hamiltonian operator H_0 is the differential-difference operator with the step equal to the Compton wavelength of the particle;

$$H_0 = 2mc^2 \operatorname{ch} \frac{i\hbar}{mc}\frac{\partial}{\partial r} + \frac{2i\hbar c}{r} \operatorname{sh} \frac{i\hbar}{mc}\frac{\partial}{\partial r} - \frac{\hbar^2}{m}\frac{\Delta_{\vartheta,\varphi}}{r^2} e^{(i\hbar/mc)(\partial/\partial r)}. \quad (1.10)$$

Taking into account the finite-difference character of Eq. (1.9) and the group-theoretical interpretation of the vector **r**, we can consider this scheme as the quantum me-

a) Permanent address: Laboratory of Theoretical Physics, Joint Institute of Nuclear Research, Head P. O. Box 79, SU-141980, Dubna, USSR.

 0022-2488/90/071733-06$03.00

i.e.,

$$\Psi_0 = C_0 e^{-\omega x^2/2}. \tag{2.14}$$

It is easily seen that the corresponding eigenvalue $E_0 = \omega/2$. Now, using the operator a^+, one subsequently constructs other eigenvectors Ψ_n corresponding to higher eigenvalues

$$E_n = \omega(h + \tfrac{1}{2}). \tag{2.15}$$

We have

$$\begin{aligned}\Psi_n &= C_n (a^+)^n \Psi_0 \\ &= C_n(-1)^n\left[e^{\omega x^2/2}\frac{d}{dx}e^{-\omega x^2/2}\right]e^{-\omega x^2/2} \\ &= C_n H_n(\sqrt{\omega}x)e^{-\omega x^2/2},\end{aligned} \tag{2.16}$$

where $H_n(\sqrt{\omega}x)$ are the Hermite polynomials given by the Rodrigues formula

$$H_n(\sqrt{\omega}x) = (-1)^n e^{\omega x^2}\left(\frac{d}{dx}\right)^n e^{-\omega x^2}. \tag{2.17}$$

III. ONE-DIMENSIONAL RELATIVISTIC CONFIGURATIONAL SPACE

The one-dimensional configurational x space is introduced by the Fourier expansion of the wave function $\Psi(p)$,

$$\Psi(x) = \frac{1}{\sqrt{2\pi}}\int d\Omega_p \langle x|p\rangle\Psi(p), \tag{3.1}$$

where

$$\langle x|p\rangle = (p_0 - p)^{-ix} \tag{3.2}$$

is the matrix element of the unitary representation of the one-dimensional Lorentz group. The momentum space in this case is the one-dimensional Lobachevsky space (the hyperbola)

$$p_0^2 - p^2 = 1, \tag{3.3}$$

embedded into the two-dimensional pseudo-Euclidean momentum space (p_0, p). In the hyperbolic coordinate system

$$p_0 = \operatorname{ch}\chi,\quad d\Omega_p = dp/p_0 = d\chi,\quad p = \operatorname{sh}\chi, \tag{3.4}$$

where $\chi = \ln(p_0 + p)$ is the rapidity, we have

$$\langle x|p\rangle = e^{ix\chi}. \tag{3.5}$$

These functions compose a complete and orthogonal system,

$$\frac{1}{2\pi}\int\langle x|p\rangle d\Omega_p\langle p|x'\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty}\langle x|\chi\rangle d\chi\langle\chi|x'\rangle = \delta(x - x'), \tag{3.6a}$$

$$\frac{1}{2\pi}\int\langle p|x\rangle dx\langle x|p'\rangle = \delta(p(-)p') = \delta(\chi - \chi'). \tag{3.6b}$$

The free energy and momentum operators are finite-difference operators

$$\hat{H}_0 = \operatorname{ch} i\frac{d}{dx},\quad p = -\operatorname{sh} i\frac{d}{dx}. \tag{3.7}$$

The plane wave (3.5) obeys the free relativistic finite-difference Schrödinger equation,

$$(\hat{H}_0 - p_0)\langle x|p\rangle = 0. \tag{3.8}$$

Using the formula $\operatorname{ch}\chi = 1 + 2\operatorname{sh}^2\chi/2$, we can introduce the relativistic "kinetic energy" operator $\hat{h}_0$:

$$\begin{aligned}\hat{h}_0 &= 2\operatorname{sh}^2\frac{i}{2}\frac{d}{dx} = \frac{\hat{k}^2}{2} = \hat{H}_0 - 1, \\ \hat{k} &= -2\operatorname{sh}\frac{i}{2}\frac{d}{dx},\quad k = 2\operatorname{sh}\frac{\chi}{2},\quad e = \frac{k^2}{2}.\end{aligned} \tag{3.9}$$

The Schrödinger equation takes the form

$$\begin{aligned}(\hat{k}/2 + V(x) - k^2/2)\Psi(x) &= (\hat{h}_0 + V(x) - e)\Psi(x) \\ &= (\hat{h} - e)\Psi(x) = 0,\end{aligned} \tag{3.10}$$

that is indistinguishable from the nonrelativistic equation. This reflects the absolute character of this approach mentioned above.

IV. THE RELATIVISTIC FACTORIZATION METHOD

Let us generalize the factorization method to the case of the relativistic finite-difference Schrödinger Eq. (3.10). We suppose that the ground-state wave function has again the form (2.2) and the energy e_0. Let us consider the finite-difference operators

$$\begin{aligned}M^{\pm} &= \mp\frac{i\alpha(x)}{\sqrt{2}}e^{\rho(x)}\hat{k}e^{-\rho(x)} \\ &= \pm\frac{2i\alpha(x)}{\sqrt{2}}e^{\rho(x)}\operatorname{sh}\frac{i}{2}\frac{d}{dx}e^{-\rho(x)}.\end{aligned} \tag{4.1}$$

In the finite-difference case we have to consider, instead of the commutator, a more-complicated expression (the generalized commutator)

$$[M^-, M^+]_\omega = M^- e^{a(x)}M^+ - M^+ e^{-a(x)}M^-, \tag{4.2}$$

where

$$a(x) = 2\operatorname{ch}\frac{i}{2}\frac{d}{dx}\rho(x) - 2\rho(x). \tag{4.3}$$

The direct calculation gives

$$\begin{aligned}[M^-, M^+]_\omega = &-2\alpha(x)\{\alpha_c(x)\operatorname{sh}(\rho_c(x) - \rho(x))\operatorname{ch}\rho_s(x) \\ &+ \alpha_s(x)\operatorname{ch}(\rho_c(x) - \rho(x))\operatorname{sh}\rho_s(x)\},\end{aligned} \tag{4.4}$$

where

$$\begin{aligned}\rho_c(x) &= \operatorname{ch} i\frac{d}{dx}\rho(x),\quad \rho_s(x) = \operatorname{sh} i\frac{d}{dx}\rho(x), \\ \alpha_c(x) &= \operatorname{ch}\frac{i}{2}\frac{d}{dx}\alpha(x),\quad \alpha_s(x) = \operatorname{sh}\frac{i}{2}\frac{d}{dx}\alpha(x).\end{aligned} \tag{4.5}$$

In the nonrelativistic limit

$$-2i\operatorname{sh}\frac{i}{2}\frac{d}{dx}\to\frac{d}{dx},\quad M^{\pm}\to a^{\pm},$$

$$[M^-, M^+]_\omega \to [a^-, a^+]$$

and the expression in the right-hand side of (4.4) becomes $-\partial^2\rho/\partial x^2$. It is natural to consider the situation with [cf. (2.8)]

$$[M^-, M^+]_\omega = \text{const}, \tag{4.6}$$

as the relativistic oscillator.

chanics on the covariant lattice. On the other hand, formally the manifold **r** is the three-dimensional Euclidean manifold. We can then think that this approach to the relativization of quantum mechanics is equivalent to introducing the differential-difference Schrödinger operator instead of the usual second-order differential operator of the nonrelativistic quantum mechanics. The formalism based on Eq. (1.9) carries many features of the nonrelativistic quantum mechanics. The scattering theory based on the partial phase shifts was built up. The approximations usually exploited in quantum mechanics were also constructed. In a number of important cases (the Coulomb field and the potential well) the differential-difference equation (1.9) can be exactly solved. With this purpose, the generalization of the theory of special functions that is based on the difference equations (recurrence relations) but not on the differential ones was developed.[8]

We stress the following important feature of this formalism. The rapidities χ or non-Euclidean distances in the Lobachevsky space are defined by the relation

$$\chi = \ln((p_0 + |\mathbf{p}|)/mc). \tag{1.11}$$

They are canonically conjugated to r in the sense of the relativistic Fourier transformation. As a consequence, the uncertainty relation holds,

$$\Delta r \cdot \Delta\chi \gtrsim \hbar/mc. \tag{1.12}$$

In this paper we consider the generalization of the problem of the one-dimensional harmonic oscillator for the relativistic configurational space. It is worthwhile to stress that the important problem of the harmonic oscillator perpetually attracts the attention of physicists from different points of view. It plays an important role in models describing relativistic objects with internal structure, strings, approaches allowing to circumvent the no-go theorems, and in particular to build up a new model exhibiting a generalization of supersymmetry, etc.[11–13] In contrast with another important case of the Coulomb potential,[14] which can be calculated as an input of the one-photon exchange, the relativistic generalization of the oscillator potential is not uniquely defined. We shall require the relativistic linear oscillator to possess the next properties: (a) exact solubility; (b) the correct nonrelativistic limit; (c) the minimization of the uncertainty relation; (d) the existence of the "nonrelativistic" dynamical symmetry group SU(1,1); (e) the symmetry between the descriptions in configurational and in momentum spaces.

As a starting point of our construction we use the finite-difference generalization of the well-known factorization method. This method was first employed by Dirac. Its systematic study was made by Infeld and Hall, the group-theoretical meaning was given by Moshinsky, Wolf, and other authors.[15]

II. THE QUANTUM-MECHANICAL FACTORIZATION METHOD

Let us consider the nonrelativistic one-dimensional Hamiltonian (we use in what follows the unit system $\hbar = m = c = 1$)

$$H = -\frac{1}{2}\frac{\partial^2}{\partial x^2} + V(x), \tag{2.1}$$

with the positive-definite wave function of the ground state

$$\Psi_0(x) = e^{-\rho(x)}, \tag{2.2}$$

and the energy E_0

$$H\Psi_0 = E\Psi_0. \tag{2.3}$$

We can express $V(x)$ in terms of $\rho(x)$ and E_0:

$$V(x) = \frac{1}{2}\left[\left(\frac{\partial\rho}{\partial x}\right)^2 - \frac{\partial^2\rho}{\partial x^2}\right] - E_0. \tag{2.4}$$

Taking this equation into account we can write down H in the factorized form

$$H - E_0 = a^+ a^-, \tag{2.5}$$

where

$$a^\pm = \frac{1}{\sqrt{2}}\left(\mp\frac{\partial}{\partial x} + \frac{\partial\rho}{\partial x}\right)$$
$$= \mp\frac{1}{\sqrt{2}} e^{\pm\rho(x)}\frac{\partial}{\partial x} e^{\mp\rho(x)}. \tag{2.6}$$

The $a^\pm$ operators obey the commutation relation

$$[a^-, a^+] = \frac{\partial^2\rho}{\partial x^2}. \tag{2.7}$$

For the harmonic oscillator the commutator is constant,

$$[a^-, a^+] = \omega = \text{const}, \tag{2.8}$$

and we have from (2.7)

$$\rho(x) = \omega x^2/2. \tag{2.9}$$

The creation and annihilation operators take the form

$$a^\pm = \frac{1}{\sqrt{2}}\left(\mp\frac{d}{dx} + \omega x\right) = \mp\frac{i}{\sqrt{2}}(\hat{p}_n \pm i\omega x), \tag{2.10}$$

where $\hat{p}_n$ is the nonrelativistic momentum operator $\hat{p}_n = -i(d/dx)$. The following relations:

$$\begin{aligned} &H = a^+a^- + (\omega/2) = \tfrac{1}{2}(a^+a^- + a^-a^+), \\ &Ha^+ = a^+(H+\omega), \quad Ha^- = a^-(H-\omega), \end{aligned} \tag{2.11}$$

are easily derived.

These relations give us the method for constructing the eigenvectors and eigenvalues of H. If Ψ is an eigenvector of H ($H\Psi = E\Psi$), the functions $a^+\Psi$ and $a^-\Psi$ [provided that they are nonzero and belong to $L^2(R)$] and new eigenvectors corresponding to the eigenvalues $E+\omega$ and $E-\omega$, respectively,

$$H(a^+\Psi) = a^+(H+\omega)\Psi = (E+\omega)a^+\Psi, \tag{2.12a}$$

$$H(a^-\Psi) = a^-(H-\omega)\Psi = (E-\omega)a^-\Psi. \tag{2.12b}$$

Since the operator H is positive definite, one can immediately find the lowest-energy eigenstate Ψ_0, as that one for which

$$a^-\Psi_0 = \frac{2}{\sqrt{2}} e^{-\omega x^2/2}\frac{d}{dx} e^{\omega x^2/2}\Psi_0 = 0, \tag{2.13}$$

It is easily seen that the solution of this equation coincides with the nonrelativistic oscillatory function $\rho(x) = \omega x^2/2$. In this case $a(x) = -\omega/4$, and $\alpha(x)$ has to be

$$\alpha(x) = [\cos(\omega x/2)]^{-1}. \tag{4.7}$$

We have [cf. (2.10)]

$$M^{\pm} = \pm \frac{2i}{\sqrt{2}} e^{\pm\omega/8}\left(\operatorname{sh}\frac{i}{2}\frac{d}{dx} \mp i\tan\frac{\omega x}{2}\operatorname{ch}\frac{i}{2}\frac{d}{dx}\right)$$
$$= \mp \frac{i}{\sqrt{2}} e^{\pm\omega/8}\left(\hat{k} \pm 2i\tan\frac{\omega x}{2}\operatorname{ch}\frac{i}{2}\frac{d}{dx}\right). \tag{4.8}$$

In the oscillator case the generalized commutator (4.2) is a combination of the conventional commutator and anticommutator

$$[M^-,M^+]_\omega = e^{-\omega/4}M^-M^+ - e^{\omega/4}M^+M^-$$
$$= \operatorname{ch}(\omega/4)[M^-,M^+]$$
$$-\operatorname{sh}(\omega/4)\{M^+,M^-\}. \tag{4.9}$$

It is easy to see that

$$[M^-,M^+]_\omega = 4\operatorname{sh}(\omega/4). \tag{4.10}$$

Using the Baker–Campbell–Hausdorff formula we can write $M^{\pm}$ in the form [cf. (2.10)]

$$M^{\pm} = -\frac{2i}{\sqrt{2}\cos\omega x/2}\operatorname{sh}\frac{i}{2}\left(\mp\frac{d}{dx} + \omega x\right)$$
$$= -\frac{2i}{\sqrt{2}\cos\omega x/2}\operatorname{sh}\frac{i}{2}a^{\pm}. \tag{4.11}$$

Then, defining the operator

$$\hat{\mathscr{D}} = -\frac{2i}{\cos\omega x/2}\operatorname{sh}\frac{i}{2}\frac{d}{dx}, \tag{4.12}$$

we write

$$M^{\pm} = \mp\frac{1}{\sqrt{2}}e^{\pm\omega x^2/2}\hat{\mathscr{D}}e^{\mp\omega x^2/2}. \tag{4.13}$$

The Hamiltonian of the relativistic linear oscillator is

$$\hat{h} = \tfrac{1}{2}\{M^+,M^-\}_\omega = \tfrac{1}{2}\{e^{\omega/4}M^+M^- + e^{-\omega/4}M^-M^+\}$$
$$= e^{\omega/4}M^+M^- + e_0 = 2\left(\frac{1}{\cos\omega x/2}\operatorname{ch}\frac{i}{2}\frac{d}{dx}\right)^2$$
$$-2\operatorname{ch}\frac{\omega}{4}, \tag{4.14}$$

where

$$e_0 = 2\operatorname{sh}(\omega/4). \tag{4.15}$$

It follows from (4.9) and (4.14) that

$$[M^+,\hat{h}]_{-\omega} = -2\operatorname{sh}(\omega/2)M^+, \tag{4.16a}$$
$$[\hat{h},M^-]_\omega = -2\operatorname{sh}(\omega/2)M^-, \tag{4.16b}$$

then M^+ and M^- are raising and lowering operators for energy levels:

$$\hat{h}\Psi_{n+1} = \hat{h}M^+\Psi_n = e_{n+1}\Psi_{n+1}, \tag{4.17a}$$
$$\hat{h}\Psi_{n-1} = \hat{h}M^-\Psi_n = e_{n-1}\Psi_{n-1}, \tag{4.17b}$$

where

$$e_{n+1} = e^{\omega/2}e_n + 2e^{\omega/4}\operatorname{sh}(\omega/2). \tag{4.18}$$

From (4.15) and (4.18) we derive by induction the formula for the energy spectrum:

$$e_n = 2(e^{[(2n+1)/4]} - \operatorname{ch}(\omega/4)). \tag{4.19}$$

V. WAVE FUNCTIONS

We look for solutions of the relativistic Schrödinger equation

$$\hat{h}\Psi_n(x) = e_n\Psi_n(x), \tag{5.1}$$

in the form

$$\Psi_n(x) = e^{-\omega x^2/2}h_n(x). \tag{5.2}$$

The functions $h_n(x)$ "the relativistic Hermite polynomials," obey the finite-difference equation

$$(\hat{K}^2 - e_n - 2\operatorname{ch}\omega/4)h_n(x), \tag{5.3}$$

where

$$\hat{K} = \frac{1}{\cos\omega x/2}e^{\omega x^2/2}\operatorname{ch}\frac{i}{2}\frac{d}{dx}e^{-\omega x^2/2}. \tag{5.4}$$

Using (4.13), (4.17a), and (4.18) we define $h_n(x)$ by the "relativistic Rodriguez formula":

$$h_n(x) = (-1/\sqrt{\omega})^n e^{\omega x^2}\hat{\mathscr{D}}^h e^{-\omega x^2}, \tag{5.5}$$
$$h_{n+1}(x) = -(1/\sqrt{\omega})e^{\omega x^2}\hat{\mathscr{D}}e^{-\omega x^2}h_n(x), \tag{5.6}$$
$$h_0(x) = 1. \tag{5.7}$$

It is easy to show by induction that $h_n(x)$ are polynomials of nth degree of the variable $\sin(\omega x/2)$. The polynomials $h_n(x)$ satisfy the following recurrence relations:

$$h_n(-x) = (-1)^h h_n(x), \tag{5.8}$$
$$\sqrt{\omega}\,\hat{\mathscr{D}}\,h_n(x) = 8e^{n\omega/4}\operatorname{sh}(n\omega/4)\cdot h_{n-1}(x), \tag{5.9}$$
$$h_{n+1}(x) - \frac{4}{\sqrt{\omega}}e^{[(n+1)/4]\omega}\sin\left(\frac{\omega x}{2}\right)h_n(x)$$
$$+\frac{8}{\omega}\operatorname{sh}\frac{n\omega}{4}e^{[(n+1)/4]\omega}h_{n-1}(x) = 0, \tag{5.10}$$
$$e^{n\omega/4}h_n(x) - \operatorname{ch}\frac{i}{2}\frac{d}{dx}h_n(x)$$
$$+ i\tan\frac{\omega x}{2}\operatorname{sh}\frac{i}{2}\frac{d}{dx}h_n(x) = 0. \tag{5.11}$$

Let us write down the explicit expressions for polynomials $h_n(x)$ for the lowest values of n:

$$h_1(x) = (4/\sqrt{\omega})e^{\omega/4}\sin(\omega x/2),$$
$$h_2(x) = \frac{8}{\omega}e^{\omega/4}\left[2e^{\omega/4}\sin^2\frac{\omega x}{2} - \operatorname{sh}\frac{\omega}{4}\right],$$
$$h_3(x) = \frac{32}{\omega^{3/2}}e^{3\omega/2}\sin\frac{\omega x}{2}\left[2\sin^2\frac{\omega x}{2}\right.$$
$$\left.+ e^{-3\omega/4}\operatorname{ch}\frac{\omega}{4} - 1\right], \tag{5.12}$$
$$h_4(x) = \frac{64}{\omega^2}e^{7\omega/4}\left[4e^{3\omega/4}\sin^4\frac{\omega x}{2}\right.$$
$$+4\left(\operatorname{ch}\frac{\omega}{4}\operatorname{ch}\frac{\omega}{2} - e^{3\omega/4}\right)\sin^2\frac{\omega x}{2}$$
$$\left.+ e^{-\omega/4}\operatorname{sh}\frac{\omega}{4}\operatorname{sh}\frac{3\omega}{4}\right].$$

The polynomials $h_n(x)$ with different n values are orthogonal with the weight $e^{-\omega x^2}\cos(\omega x/2)$. To prove this, we multiply Eq. (5.11) by $e^{-\omega x^2}\Psi_m$ $(m\neq n)$ and subtract the same expression with $m \rightleftarrows n$ and integrate over x

$$\int_{-\infty}^{\infty}\left(\Psi_m \operatorname{ch}\frac{i}{2}\frac{d}{dx}\Psi_n - \Psi_n \operatorname{ch}\frac{i}{2}\frac{d}{dx}\Psi_m\right)dx = (e^{[(2n+1)/8]\omega} - e^{[(2m+1)/8]\omega}) \times\int_{-\infty}^{\infty}\cos\left(\frac{\omega x}{2}\right)\Psi_n(x)\Psi_m(x)dx. \quad (5.13)$$

We transform the second term in the left-hand side using the identity

$$\int_{-\infty}^{\infty} f(x)\left(\operatorname{ch}\frac{i}{2}\frac{d}{dx}\varphi(x)\right)dx = \int_{-\infty}^{\infty}\left(\operatorname{ch}\frac{i}{2}\frac{d}{dx}f(x)\right)\varphi(x)dx, \quad (5.14)$$

which is valid provided that $f(x)$ and $\varphi(x)$ vanish at $x=\pm\infty$ together with all their derivatives. In our case these conditions are satisfied and we see that the left-hand side of (5.14) vanishes.

In the case $m=n$ the norm can be calculated using the recurrence relations. And finally, we arrive at the orthonormality conditions

$$\int_{-\infty}^{\infty} e^{-\omega x^2} h_n(x)h_m(x)\cos\frac{\omega x}{2}\,dx = \delta_{nm} J_n,$$
$$J_n = \sqrt{\frac{\pi}{\omega}}\left(\frac{8}{\omega}\right)^n \times\exp\left(\frac{2n^2+2n-1}{16}\omega\right)\operatorname{sh}\frac{\omega}{4}\operatorname{sh}\frac{2\omega}{4}\cdots\operatorname{sh}\frac{n\omega}{4}. \quad (5.15)$$

Now we derive the integral representation for $h_n(x)$. Let us write the identity

$$e^{-\omega x^2} = \frac{1}{\sqrt{\pi\omega}}\int_{-\infty}^{\infty} e^{-t^2/\omega}e^{2ixt}\,dt. \quad (5.16)$$

Applying the recurrence relation (5.5) we have

$$e^{-\omega x^2}h_n(x) = \frac{1}{\sqrt{\pi\omega}}\left(\frac{e^{-i\pi/2}\cdot 2}{\sqrt{\omega}}\right)^n\int_{-\infty}^{\infty} T_n(t)e^{2ixt}e^{-t^2/\omega}\,dt, \quad (5.17)$$

where $T_n(t)$ are polynomials of t, satisfying the recurrence relation

$$\operatorname{sh} t e^{-t^2/\omega}T_n(t) = \operatorname{ch}\frac{\omega}{4}\frac{d}{dt}\left(e^{-t^2/\omega}T_{n+1}(t)\right),$$
$$T_0(t)=1. \quad (5.18)$$

In the nonrelativistic limit $T_n(t)\to t$ and we come to the integral representation for the usual Hermite polynomials[16]:

$$e^{-\omega x^2}H_n(\sqrt{\omega}x) = \frac{1}{\sqrt{\pi\omega}}\left(\frac{e^{-i\pi/2}\cdot 2}{\sqrt{\omega}}\right)^n\int_{-\infty}^{\infty} t^n e^{2ixt}e^{-t^2/\omega}\,dt. \quad (5.19)$$

The $T_n(t)$ also satisfy the recurrence relations

$$\operatorname{sh}\frac{n\omega}{8}\operatorname{ch}\frac{t}{2}\operatorname{ch}\frac{\omega}{4}\frac{d}{dt}T_n(t) = \operatorname{ch}\frac{n\omega}{8}\operatorname{sh}\frac{t}{2}\operatorname{sh}\frac{\omega}{4}\frac{d}{dt}T_n(t), \quad (5.20)$$

$$e^{-t^2/\omega}T_{n+1}(t) + 2e^{[(n+1)/4]\omega}\operatorname{sh}\frac{\omega}{4}\frac{d}{dt}(e^{-t^2/\omega}T_n(t)) - 2\operatorname{sh}\frac{n\omega}{4}e^{[(n+1)/4]\omega}e^{-t^2/\omega}T_{n-1}(t)=0, \quad (5.21)$$

$$T_{n+1}(t) = \frac{e^{[(2n+3)/16]\omega}(\operatorname{ch} t - \operatorname{ch} n\omega/4)}{\operatorname{sh}(n\omega/8)\operatorname{ch}(t/2)}\operatorname{sh}\frac{\omega}{4}\frac{d}{dt}T_n(t), \quad (5.22a)$$

$$T_{n+1}(t) = \frac{e^{[(2n+3)/16]\omega}(\operatorname{ch} t - \operatorname{ch}(n\omega/4))}{\operatorname{ch}(n\omega/8)\operatorname{sh}(t/2)}\operatorname{ch}\frac{\omega}{4}\frac{d}{dt}T_n(t). \quad (5.22b)$$

For lowest values of n we have

$$T_1(t) = 2e^{3\omega/16}\operatorname{sh}(t/2),$$
$$T_2(t) = 2e^{\omega/2}(\operatorname{ch} t - \operatorname{ch}(\omega/4)), \quad (5.23)$$
$$T_3(t) = 4e^{\frac{15}{16}\omega}\operatorname{sh}(t/2)(\operatorname{ch} t - \operatorname{ch}(\omega/2)).$$

VI. COHERENT STATES

The commutator of the coordinate x and rapidity

$$\hat{\chi} = -i\frac{\hbar}{mc}\frac{d}{dx} \quad (6.1)$$

operators is equal to

$$[x,\hat{\chi}] = i(\hbar/mc). \quad (6.2)$$

Then, the uncertainty relation

$$(\Delta x)^2(\Delta\hat{\chi})^2 \geqslant \hbar^2/4m^2c^2 \quad (6.3)$$

is fulfilled.

Let us define the relativistic coherent state $|\alpha\rangle$ by

$$\cos(\omega x/2)M^-|\alpha\rangle = -i\sqrt{2/\omega}(\operatorname{sh}(i\sqrt{\omega/2}\alpha))|\alpha\rangle,$$
$$\alpha = \alpha_1 + i\alpha_2. \quad (6.4)$$

It is easily seen that $|\alpha\rangle$ has the form

$$|\alpha\rangle = \left(\frac{\omega}{\pi}\right)^{1/4}\exp\left[-\left(\sqrt{\frac{\omega}{2}}x-\alpha\right)^2 + \frac{\alpha^2}{2} - \frac{|\alpha|^2}{2}\right], \quad (6.5)$$

where

$$\alpha_1 = x_0\sqrt{\omega/2}, \quad \alpha_2 = \chi_0/\sqrt{2\omega}, \quad (6.6)$$

and x_0 and χ_0 are averages of the coordinate and rapidity, respectively. This corresponds to the fact that states (6.5) minimize the product of uncertainties, i.e., the equality sign in the relation (6.3) holds.

VII. THE GENERALIZED ALGEBRA OF THE SU(1,1) DYNAMICAL SYMMETRY

Let us introduce the operators [cf. Ref. 17],

$$L^+ = -[4(\operatorname{sh}\omega/2\operatorname{sh}\omega/4)^{1/2}]^{-1}(M^+M^-)^{1/2}M^+,$$
$$L^- = [4(\operatorname{sh}\omega/2\operatorname{sh}\omega/4)^{1/2}]^{-1}M^-(M^+M^-)^{1/2},$$
$$L^3 = (2\operatorname{sh}(\omega/2))^{-1}\hat{h}. \quad (7.1)$$

It is easy to show that the algebra of these operators is closed in the following way:

$$[L^+,L^-]_{2\omega}=L^3,$$
$$[L^+,L^3]_\omega=-L^+, \qquad (7.2)$$
$$[L^3,L^-]_\omega=-L^-,$$

where the subscript ω denotes the combination of the commutator and anticommutator according to (4.9). For the Lie algebra thus generalized we have, instead of the standard symmetry relation and the Jacobi identity, the following relations:

$$[x,y]_\omega=-[y,x]_{-\omega}, \qquad (7.3a)$$
$$[[x,y]_\omega,z]_\omega+[[y,z]_\omega,x]_\omega+[[z,x]_\omega,y]_\omega$$
$$+[[x,y]_\omega,z]_{-\omega}+[[y,z]_\omega,x]_{-\omega}$$
$$+[[z,x]_\omega,y]_{-\omega}=0. \qquad (7.3b)$$

ACKNOWLEDGMENTS

The authors would like to thank V. G. Kadyshevsky, V. I. Manko, and S. Randjibar-Daemi for useful discussions. One of the authors (R. M. M-K.) would also like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste.

[1]V. G. Kadyshevsky, R. M. Mir-Kasimov, and N. B. Skachkov, Nuovo Cimento **55**, 233 (1967).
[2]R. M. Mir-Kasimov, Sov. Phys. JETP **22**, 629 (1966).
[3]I. S. Shapiro, Sov. Phys. JETP **16**, 1219 (1963); I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Vol. 5, Integral Geometry and Representation Theory* (New York, 1966); N. Ya. Vilenkin and Ya. A. Smorodinsky, Sov. Phys. JETP **19**, 1209 (1964).
[4]A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **29**, 380 (1963).
[5]V. G. Kadyshevsky, Nucl. Phys. **6**, 125 (1968).
[6]A. Klein and T.-S. Lee H., Phys. Rev. D **10**, 4308 (1974).
[7]V. G. Kadyshevsky, R. M. Mir-Kasimov, and N. B. Skachkov, Phys. Elem. Part. Atomic Nucleus **2**, (3) 635 (1972); I. V. Amirkhanov, G. V. Grusha, and R. M. Mir-Kasimov, *ibid.* **12**, (3), 651 (1981).
[8]V. G. Kadyshevsky, R. M. Mir-Kasimov, and M. Freeman, Jadernaja Fiz. **9**, 646 (1969).
[9]V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. **165**, 59 (1985).
[10]H. Van Dam and E. P. Wigner, Phys. Rev. **138**, 1576 (1965); T. Takabayashi, Prog. Theor. Phys. Suppl. **67**, 1 (1979); D. Dominici, J. Gomis, and G. Longhi, Nuovo Cimento B **48**, 152 (1978); Nuovo Cimento A **48**, 257 (1978); I. Sogami, Prog. Theor. Phys. **50**, 1729 (1973); Ph. Droz-Vincent, Phys. Rev. D **19**, 702 (1979); W. Drechsler, J. Math. Phys. **26**, 41 (1985).
[11]P. A. M. Dirac, Proc. R. Soc. London Ser. A **322**, 435 (1971); **328**, 1 (1972); J. Math. Phys. **4**, 901 (1963).
[12]N. Mukunda, H. Van Dam, and L. C. Biedenharn, Phys. Rev. D **22**, 1938 (1980); H. Van Dam, and L. C. Biedenharn, Phys. Lett. B **81**, 313 (1979); **62**, 190 (1976); Phys. Rev. D **14**, 405 (1976); **9**, 471 (1974); L. P. Stauton and H. Van Dam, Lett. Nuovo Cimento **7**, 371 (1973).
[13]N. N. Atakishiyev, R. M. Mir-Kasimov, and Sh. M. Nagiev, Theor. Math. Phys. **44**, 47 (1980); **67**, 68 (1986); Ann. Phys. (Leipzig) **42**, 25 (1985); A. Bohm, M. Loewe, and P. Magnollay, Phys. Rev. D **31**, 2304 (1985); Y. S. Kim, and M. E. Noz, Phys. Rev. D **12**, 122, 129 (1975); **8**, 3521 (1973), Prog. Theor. Phys. **57**, 1373 (1977); **60**, 801 (1978).
[14]M. Freeman, M. D. Mateev, and R. M. Mir-Kasimov, Nucl. Phys. B **12**, 24 (1969).
[15]L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951); M. Moshinsky, *The Harmonic Oscillator in Modern Physics* (Gordon and Breach, New York, 1969); D. Basu and K. B. Wolf, J. Math. Phys. **24**, 478 (1983).
[16]*Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. 2.
[17]I. A. Malkin and V. I. Man'ko, *The Dynamical Symmetries and Coherent States of Quantum Systems* (Nauka, Moscow, 1979); A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971); A. O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965).

Symmetry operators and separation of variables for spin-wave equations in oblate spheroidal coordinates

E. G. Kalnins and G. C. Williams
Department of Mathematics and Statistics, University of Waikato, Hamilton, New Zealand

(Received 21 June 1989; accepted for publication 3 January 1990)

A family of second-order differential operators that characterize the solution of the massless spin s field equations, obtained via separation of variables in oblate spheroidal coordinates and using a null tetrad is found. The first two members of the family also characterize the separable solutions in the Kerr space-time. It is also shown that these operators are symmetry operators of the field equations in empty space-times whenever the space-time admits a second-order Killing–Yano tensor.

I. INTRODUCTION

Interest in the separation and solution of the nonscalar equations of mathematical physics in Kerr space-time began when Teukolsky[1] found that separable solutions were possible for some of the Maxwell and Weyl scalars. Chandrasekhar[2] was later able to obtain a separable solution to the Dirac equation. Separable solutions to massless spin s equations were studied by Dudley and Finlay[3] while Carter and McLenaghan[4] were able to understand Chandrasekhar's separation of Dirac's equation in terms of a differential operator that characterized the separation constant appearing in the solution. That is, the separable solutions to Dirac's equation were found to be eigensolutions of the differential operator, the eigenvalue being the separation constant appearing in the solution. Similarly, the separation constant appearing in the solution to Maxwell's equations in Kerr geometry has been characterized by Kalnins *et al.*[5] in terms of a second-order differential operator. These differential operators characterizing the separation constants are also symmetry operators of the various field equations in question. That is, they map solutions of the field equations into solutions. The essential property that allows the construction of such operators is the existence of a Killing–Yano tensor in the Kerr space-time.

The other constants associated with the separable solutions of various field equations in the Kerr space-time are the Starobinsky constants. Torres del Castillo[6–8] has shown, for various fields in type D space-times, that one can construct differential operators of order $2s$, $s = 0,\frac{1}{2},1$ that characterize these constants. Physically, Killing–Yano tensors and operators constructed from them have been associated with angular momentum by Carter and McLenaghan[4] and by Dietz and Rudiger.[9,10]

In this paper we take the Kerr metric and a Kinnersley null tetrad and subsequently place $M = 0$. We then find that the solution to the massless spin s field equations obtained via separation of variables (and with the aid of a generalized Hertz potential) are characterized by a second-order differential operator. We also show that this differential operator is a symmetry operator of the field equations.

II. PRELIMINARIES

In this paper we will use the abstract index and spinor formalisms of Penrose and Rindler.[11] For the purpose of this paper we shall also refer to those components of a symmetric spinor that are of extreme helicity as the extremal components.

The Kerr metric describes the space-time in the region exterior to a rotating black hole, its line element being

$$\begin{aligned} ds^2 = {} & \left(1 - \frac{2Mr}{\tilde\rho\tilde\rho^*}\right)dt^2 - \frac{\tilde\rho\tilde\rho^*}{\Delta}\,dr^2 \\ & - \tilde\rho\tilde\rho^*\,d\theta^2 - \left(r^2 + a^2 + \frac{2a^2Mr\sin^2\theta}{\tilde\rho\tilde\rho^*}\right) \\ & \times \sin^2\theta\,d\phi^2 + \frac{4aMr\sin^2\theta}{\tilde\rho\tilde\rho^*}\,dt\,d\phi, \end{aligned} \tag{1}$$

where

$$\tilde\rho = r + ia\cos\theta \text{ and } \Delta = r^2 - 2Mr + a^2. \tag{2}$$

We shall use the null tetrad

$$\begin{aligned} l^a &= (1/\sqrt{2}\Delta)(r^2 + a^2,\Delta,0,a), \\ n^a &= (1/\sqrt{2}\tilde\rho\tilde\rho^*)(r^2 + a^2, -\Delta,0,a), \\ m^a &= (1/\sqrt{2}\tilde\rho)(ia\sin\theta,0,1,i\csc\theta), \\ \overline{m}^a &= (1/\sqrt{2}\tilde\rho^*)(-ia\sin\theta,0,1,-i\csc\theta). \end{aligned} \tag{3}$$

In this tetrad the spin coefficients are

$$\begin{aligned} & \epsilon = 0, \quad \beta = \cot\theta/2\sqrt{2}\tilde\rho, \quad \alpha = \pi - \beta^* \\ & \gamma = \mu + \frac{(r-M)}{\sqrt{2}\tilde\rho\tilde\rho^*}, \\ & \rho = -\frac{1}{\sqrt{2}\tilde\rho^*}, \quad \tau = -\frac{ia\sin\theta}{\sqrt{2}\tilde\rho\tilde\rho^*}, \quad \pi = \frac{ia\sin\theta}{\sqrt{2}\tilde\rho^{*2}} \\ & \mu = -\frac{\Delta}{\sqrt{2}\tilde\rho\tilde\rho^{*2}}, \quad \kappa = 0, \quad \sigma = 0, \quad \lambda = 0, \quad \nu = 0, \end{aligned} \tag{4}$$

while the only nonzero component of the Weyl spinor is

$$\Psi_2 = -M/\tilde\rho^{*3}. \tag{5}$$

Following Chandrasekhar[12] we define the differential operators

$$\begin{aligned} \mathscr{D}_s &= \partial_r + iK/\Delta + 2s(r-M)/\Delta, \\ \mathscr{D}_s^\dagger &= \partial_r - iK/\Delta + 2s(r-M)/\Delta, \\ \mathscr{L}_s &= \partial_\theta + Q + s\cot\theta, \\ \mathscr{L}_s^\dagger &= \partial_\theta - Q + s\cot\theta, \end{aligned} \tag{6}$$

where

 0022-2488/90/071739-06$03.00

$$K=\sigma(r^2+a^2)+ma \text{ and } Q=\sigma a\sin\theta+m\csc\theta. \tag{7}$$

A second-order Killing–Yano tensor is an antisymmetric tensor K_{ab} that satisfies

$$\nabla_{(a}K_{b)c}=0. \tag{8}$$

Being antisymmetric, K_{ab} can be written in terms of symmetric spinors as

$$K_{ab}=K_{AA'BB'}=\tfrac{1}{2}(\epsilon_{A'B'}K_{AB}+\epsilon_{AB}\widetilde{K}_{A'B'}). \tag{9}$$

The Killing spinors K_{AB} and $\widetilde{K}_{A'B'}$ as a consequence of (8) satisfy

$$\begin{aligned}&\nabla_{(AA'}K_{BC)}=0,\\&\nabla_{A(A'}\widetilde{K}_{B'C')}=0,\\&\nabla_{BA'}K_A{}^B+\nabla_{AB'}\widetilde{K}_{A'}{}^{B'}=0.\end{aligned} \tag{10}$$

Defining the quantity $M_{AA'}$ by

$$M_{AA'}=\nabla_{BA'}K_A{}^B, \tag{11}$$

we can write the derivatives of the Killing spinors as

$$\begin{aligned}&\nabla_{AA'}K_{BC}=\tfrac{2}{3}\epsilon_{A(B}M_{C)A'},\\&\nabla_{AA'}\widetilde{K}_{B'C'}=-\tfrac{2}{3}\epsilon_{A'(B'}M_{AC')}.\end{aligned} \tag{12}$$

The derivative of $M_{AA'}$ is given by

$$\nabla_{AA'}M_{BB'}=\tfrac{1}{2}\epsilon_{A'B'}W_{AB}-\tfrac{1}{2}\epsilon_{AB}\widetilde{W}_{A'B'}, \tag{13}$$

where the symmetric spinors W_{AB} and $\widetilde{W}_{A'B'}$ are defined by

$$\begin{aligned}&W_{AB}=\tfrac{3}{2}\Psi_{ABCD}K^{CD}\\&\widetilde{W}_{A'B'}=\tfrac{3}{2}\overline{\Psi}_{A'B'C'D'}\widetilde{K}^{C'D'}.\end{aligned} \tag{14}$$

Note also from (13) that $\nabla_{AA'}M_{BB'}$ is an antisymmetric tensor, that is, we have

$$\nabla_{(a}M_{b)}=0, \tag{15}$$

which is the condition that M_a be a Killing vector. Other relations satisfied by the above quantities are

$$\begin{aligned}\Psi^E{}_{ABC}K_{DE}&=\tfrac{3}{4}\epsilon_{D(A}\Psi_{BC)EF}K^{EF}\\&=\tfrac{1}{2}\epsilon_{D(A}W_{BC)}\\\overline{\Psi}^{E'}{}_{A'B'C'}\widetilde{K}_{D'E'}&=\tfrac{3}{4}\epsilon_{D'(A'}\overline{\Psi}_{B'C')E'F'}\widetilde{K}^{E'F'}\\&=\tfrac{1}{2}\epsilon_{D'(A'}\widetilde{W}_{B'C')},\end{aligned} \tag{16}$$

and

$$\begin{aligned}&W_{AC}K_B{}^C=+\tfrac{1}{2}\epsilon_{AB}W_{CD}K^{CD},\\&\widetilde{W}_{A'C'}\widetilde{K}_{B'}{}^{C'}=+\tfrac{1}{2}\epsilon_{A'B'}\widetilde{W}_{C'D'}\widetilde{K}^{C'D'}.\end{aligned} \tag{17}$$

The antisymmetry in the free indices of the above two quantities being particularly useful. The derivatives of W_{AB} and $\widetilde{W}_{A'B'}$ are

$$\begin{aligned}&\nabla_{AA'}W_{BC}=2\Psi_{ABC}{}^DM_{DA'},\\&\nabla_{AA'}\widetilde{W}_{B'C'}=-2\overline{\Psi}_{A'B'C'}{}^{D'}M_{AD'},\end{aligned} \tag{18}$$

expressions which can be obtained by examining the consistency condition on $M_{AA'}$, that is, from

$$\begin{aligned}[\nabla_{AA'},\nabla_{BB'}]M_{CC'}=&-\epsilon_{A'B'}\Psi_{ABC}{}^DM_{DC'}\\&-\epsilon_{AB}\overline{\Psi}_{A'B'C'}{}^{D'}M_{CD'}.\end{aligned} \tag{19}$$

We define the differential operator ${}_\eta J_{AA'}$ by

$$\begin{aligned}{}_\eta J_{AA'}&=2K_{AA'}{}^{CC'}\nabla_{CC'}+(\eta/3)M_{AA'}\\&=K_A{}^C\nabla_{CA'}+\widetilde{K}_{A'}{}^{C'}\nabla_{AC'}+(\eta/3)M_{AA'}.\end{aligned} \tag{20}$$

This operator will be the essential building block for the symmetry operators we shall encounter later. The commutator of $\nabla_{BB'}$ with ${}_\eta J_{AA'}$ is

$$\begin{aligned}[\nabla_{BB'},{}_\eta J_{AA'}]=&K_A{}^C[\nabla_{BB'},\nabla_{CA'}]+\widetilde{K}_{A'}{}^{C'}[\nabla_{BB'},\nabla_{AC'}]\\&+\tfrac{2}{3}(M_{AB'}\nabla_{BA'}-M_{BA'}\nabla_{AB'})\\&+(\eta/6)(\epsilon_{B'A'}W_{BA}-\epsilon_{BA}\widetilde{W}_{B'A'}).\end{aligned} \tag{21}$$

We also define the vectors $U_{AA'}$ and $\widetilde{U}_{AA'}$ by

$$\begin{aligned}&K_A{}^BU_{BA'}=-\tfrac{1}{3}M_{AA'},\\&\widetilde{K}_{A'}{}^{B'}\widetilde{U}_{AB'}=\tfrac{1}{3}M_{AA'}.\end{aligned} \tag{22}$$

These two vectors $U_{AA'}$ and $\widetilde{U}_{AA'}$ will later be useful in choosing gauge fields that will in turn enable the separability of a decoupled equation for the extremal components of a generalized Hertz potential representing a massless spin s field. The derivative of the Weyl spinor is also related to the vector $U_{AA'}$ by

$$\nabla_{AA'}\Psi_{BCDE}=5U_{(AA'}\Psi_{BCDE)}. \tag{23}$$

In the Kerr space-time the only solution of (10) to within a common multiplicative constant, is

$$\begin{aligned}&K_{01}=-\tilde{\rho}^*, && K_{00}=K_{11}=0,\\&\widetilde{K}_{0'1'}=\tilde{\rho}, && \widetilde{K}_{0'0'}=\widetilde{K}_{1'1'}=0,\end{aligned} \tag{24}$$

whence the only nonzero components of $K_{AA'BB'}$ are

$$\begin{aligned}&K_{01'10'}=-K_{10'01'}=r,\\&K_{00'11'}=-K_{11'00'}=ia\cos\theta.\end{aligned} \tag{25}$$

The components of $M_{AA'}$ are

$$\begin{aligned}&M_{00'}=-\frac{3}{\sqrt{2}}, && M_{01'}=-\frac{3}{\sqrt{2}}\frac{ia\sin\theta}{\tilde{\rho}},\\&M_{10'}=\frac{3}{\sqrt{2}}\frac{ia\sin\theta}{\tilde{\rho}^*}, && M_{11'}=-\frac{3}{\sqrt{2}}\frac{\Delta}{\tilde{\rho}\tilde{\rho}^*},\end{aligned} \tag{26}$$

and the components of W_{AB} and $\widetilde{W}_{A'B'}$ are

$$\begin{aligned}&W_{01}=3\tilde{\rho}^*\Psi_2, && W_{00}=W_{11}=0,\\&\widetilde{W}_{0'1'}=-3\tilde{\rho}\Psi_2^*, && \widetilde{W}_{0'0'}=\widetilde{W}_{1'1'}=0,\end{aligned} \tag{27}$$

while those of the vectors $U_{AA'}$ and $\widetilde{U}_{AA'}$ are

$$\begin{aligned}&U_{00'}=\rho,\ U_{01'}=\tau,\ U_{10'}=-\pi,\ U_{11'}=-\mu,\\&\widetilde{U}_{00'}=\tilde{\rho}^*,\ \widetilde{U}_{01'}=-\pi^*,\ \widetilde{U}_{10'}=\tau^*,\ \widetilde{U}_{11'}=-\mu^*.\end{aligned} \tag{28}$$

The minimally coupled first-order equation for a massless spin s field is

$$\nabla^A{}_{A'}\phi_{AA_2\cdots A_{2s}}=0. \tag{29}$$

It is well known that in a space-time that is not conformally flat this equation is inconsistent for $s>1$. In particular when $s>1$ and for the case of an empty space-time $\phi_{A_1\ldots A_{2s}}$ must satisfy the consistency condition

$$\Psi^{BCD}{}_{(A_3}\phi_{A_4\ldots A_{2s})BCD}=0. \tag{30}$$

In the Kerr space-time and using the null tetrad (3) and defining a new set of functions Φ_k for $k=0,\ldots,2s$ by $\Phi_k=\tilde{\rho}^{*k}\phi_k$ Eqs. (29) become

$$[\mathscr{L}_{s-p} - (2s-2p-1)(ia\sin\theta/\tilde{\rho}^*)]\,\Phi_p - [\mathscr{D}_0 + (2s-2p-1)(1/\tilde{\rho}^*)]\,\Phi_{p+1} = 0,$$
$$\Delta[\mathscr{D}^\dagger_{s-p} - (2s-2p-1)(1/\tilde{\rho}^*)]\,\Phi_p + [\mathscr{L}^\dagger_{p-s+1} + (2s-2p-1) \times (ia\sin\theta/\tilde{\rho}^*)]\,\Phi_{p+1} = 0, \tag{31}$$

where $p = 0,...,2s-1$.

The following method of obtaining a solution to the massless spin s field Eqs. (29) is due to Cohen and Kegeles.[13] If the potential $\overline{P}^{A'_1\cdots A'_{2s}}$ and an associated arbitrary gauge field $G_B{}^{A'_2\cdots A'_{2s}}$ both of which are symmetric in their primed indices satisfy

$$\nabla^{B(A'_1}\nabla_{BB'}\overline{P}^{A'_2\cdots A'_{2s})B'} - \nabla^{B(A'_1}G_B{}^{A'_2\cdots A'_{2s})} - (2s-1)(s-1)\overline{\Psi}_{B'C'}{}^{(A'_1A'_2}\overline{P}^{A'_3\cdots A'_{2s})B'C'} = 0, \tag{32}$$

then a spin s field constructed from the potential and gauge fields as follows:

$$\phi_{A_1\ldots A_{2s}} = \nabla_{(A_1A'_1}\nabla_{A_2A'_2}\cdots\nabla_{A_{2s-1}A'_{2s-1}} \times\left[\nabla_{A_{2s})A'_{2s}}\overline{P}^{A'_1\ldots A'_{2s}} - G_{A_{2s})}{}^{A'_1\ldots A'_{2s-1}}\right], \tag{33}$$

will satisfy the spin s field Eqs. (29) provided those equations are consistent. When the space-time admits a second-order Killing–Yano tensor and the quantity $\widetilde{U}_{AA'}$ as defined by (22) exists, we can make the following rather special choice of the gauge field:

$$G_B{}^{A'_2\ldots A'_{2s}} = -2s\widetilde{U}_{BA'}\overline{P}^{A'A'_2\ldots A'_{2s}}. \tag{34}$$

This choice of gauge field was made by Cohen and Kegeles though not in this covariant form. With this choice of gauge and in a type D space-time Eqs. (32) decouple. In addition, in the Kerr space-time the extremal components of the potential will now satisfy separable equations. That is, if we look for solutions of the form $f(r,\theta)e^{i\sigma t + im\varphi}$ for $\overline{P}^{0'\cdots 0'}$ and $\overline{P}^{1'\cdots 1'}$ we find that

$$[\Delta\mathscr{D}^\dagger_{1-s}\mathscr{D}_0 + \mathscr{L}^\dagger_{1-s}\mathscr{L}_s + 2(2s-1)i\sigma\tilde{\rho}^*]\overline{P}^{0'\cdots 0'} = 0,$$
$$[\Delta\mathscr{D}^\dagger_{1+s}\mathscr{D}_0 + \mathscr{L}^\dagger_{1+s}\mathscr{L}_{-s} - 2(2s+1)i\sigma\tilde{\rho}^*]\tilde{\rho}^{-2s}\overline{P}^{1'\cdots 1'} = 0, \tag{35}$$

which are separable and have solutions

$$\overline{P}^{0'\cdots 0'} = R_{-s}S_{+s}e^{i\sigma t + im\varphi},$$
$$\overline{P}^{1'\cdots 1'} = \tilde{\rho}^{2s}R_{+s}S_{-s}e^{i\sigma t + im\varphi}, \tag{36}$$

where the functions $R_{\pm s}$ and $S_{\pm s}$ satisfy Teukolsky's equations, namely,

$$[\Delta\mathscr{D}^\dagger_{1-s}\mathscr{D}_0 + 2(2s-1)i\sigma r]R_{-s} = \lambda R_{-s},$$
$$[\Delta\mathscr{D}_1\mathscr{D}^\dagger_s - 2(2s-1)i\sigma r]R_{+s} = \lambda R_{+s},$$
$$[\mathscr{L}^\dagger_{1-s}\mathscr{L}_s + 2(2s-1)\sigma a\cos\theta]S_{+s} = -\lambda S_{+s},$$
$$[\mathscr{L}_{1-s}\mathscr{L}^\dagger_s - 2(2s-1)\sigma a\cos\theta]S_{-s} = -\lambda S_{-s}. \tag{37}$$

If we form $\phi_{A_1\ldots A_{2s}}$ from a potential having $\overline{P}^{0'\cdots 0'}$ as its only nonzero component, then the extremal components of the field $\phi_{A_1\ldots A_{2s}}$ are

$$\phi_0 = [1/(\sqrt{2})^{2s}]\mathscr{D}_0^{2s}R_{-s}S_{+s}e^{i\sigma t + im\varphi},$$
$$\phi_{2s} = [1/(\sqrt{2})^{2s}\tilde{\rho}^{*2s}]\mathscr{L}_{1-s} \times\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_sR_{-s}S_{+s}e^{i\sigma t + im\varphi}. \tag{38}$$

Using the Teukolsky–Starobinsky identities[14] we can write these two components, up to some constant of proportionality, in the following form:

$$\phi_0 = R_{+s}S_{+s}e^{i\sigma t + im\varphi},$$
$$\phi_{2s} = (1/\tilde{\rho}^{*2s})R_{-s}S_{-s}e^{i\sigma t + im\varphi}. \tag{39}$$

III. INTRINSIC CHARACTERIZATION OF THE TEUKOLSKY SEPARATION CONSTANT

Suppose we form a solution $\phi_{A_1\ldots A_{2s}}$ for the massless spin s field Eqs. (29) by generating it from the extremal component of a generalized Hertz potential as in (33). We will also suppose that the space-time is the Kerr space-time if $s < 1$ while if $s > 1$ we will restrict ourselves to the oblate spheroidal coordinate system and null tetrad obtained by placing $M = 0$. The extremal components of the solution may then be written in the form given by (39). The other components of the field take on more complicated forms. The separation constant λ appearing in the solution is characterized by the following operator:

$$[K_{(A_1|}{}^{B'CC'}\nabla_{CC'} - \tfrac{1}{3}M_{(A_1|}{}^{B'}]\,[K^B{}_{B'}{}^{DD'}\nabla_{DD'} + (2s/3)M^B{}_{B'}]\phi_{B|A_2\ldots A_{2s})}$$
$$= \tfrac{1}{4} - 2J_{(A_1|}{}^{B'}{}_{4s}J^B{}_{B'\phi B|A_2\ldots A_{2s})}$$
$$= \tfrac{1}{2}\lambda\phi_{A_1\cdots A_{2s}}. \tag{40}$$

For brevity we will also sometimes write the above as

$$\mathscr{I}\phi = \tfrac{1}{2}\lambda\phi. \tag{41}$$

The extremal components of this identity are relatively easy to verify using the form for the extremal components of the field given in (39). Since it is not possible to verify directly that the remaining components of this identity hold for arbitrary values of the spin parameter s we are forced to proceed by a different method. Firstly we will show that the operator and its action on the spin s field as given by (40) is a symmetry operator of the spin s field Eqs. (29) whenever those equations are consistent. That is the operator maps solutions into solutions. The following identity holds for any empty space-time which admits a second-order Killing–Yano tensor and for any spinor $\phi_{A_1\ldots A_{2s}}$. In particular we do not assume that $\phi_{A_1\ldots A_{2s}}$ satisfies a field equation of any sort. For $s > 1$ we have

$$\nabla^{CC'}{}_{\xi}J_{(C|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{A|B_2\cdots B_{2s})}$$
$$= [{}_{\xi+2}J_{(B_2|}{}^{C'}{}_{\eta-2}J^{A}{}_{A'} - {}_{\xi}J_{(B_2|A'|\eta}J^{AC'} - \tfrac{4}{9}M_{(B_2|}{}^{C'}M^{A}{}_{A'}] + (1/3s)\nabla_{(B_2|}{}^{C'}[K^{AD}M_{DA'} + \tilde{K}_{A'}{}^{D'}M^{A}{}_{D'}]]\nabla^{CA'}\phi_{AC|B_3\cdots B_{2s})}$$
$$+ [(1/6s)((\eta-4s)-(\xi+2))\nabla_{(B_2|}{}^{C'}M^{CA'}{}_{\eta}J^{A}{}_{A'} + \tfrac{1}{6}(\xi+2)[W_{(B_2|}{}^{C}{}_{\eta}J^{AC'} + \epsilon_{(B_2|}{}^{C}\tilde{W}^{A'C'}{}_{\eta}J^{A}{}_{A'}]$$
$$- \tfrac{1}{6}((\eta-4s)+(2s-2))_{\xi+4}J_{(B_2|}{}^{C'}W^{AC} + \tfrac{1}{3}(2s-2)W_{(B_2|}{}^{C}{}_{\eta}J^{AC'}]\phi_{AC|B_3\cdots B_{2s})} + (2s-2)[-(1/12s)\nabla_{(B_2|}{}^{C'}$$
$$[K^{AC}W_{|B_3|}{}^{M} + 3\tilde{K}_{A'D'}\tilde{K}^{A'D'}\Psi^{ACM}{}_{|B_3|}] + {}_{\xi}J_{(B_2|}{}^{A'}\tilde{K}_{A'}{}^{C'}\Psi^{ACM}{}_{|B_3|} + \tilde{K}^{A'C'}\Psi^{CM}{}_{(B_2B_3|\eta}J^{A}{}_{A'}]\phi_{ACM|B_4\cdots B_{2s})}. \tag{42}$$

To prove this, first we split the left hand side of (42) into two parts. Removing the index C from the symmetrization we find that

$$\nabla^{CC'}{}_{\xi}J_{(C|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{A|B_2\cdots B_{2s})}$$
$$= \nabla^{CC'}{}_{\xi}J_{(B_2|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})} + (1/4s)\nabla_{(B_2|}{}^{C'}[{}_{\xi}J^{C}{}_{A'\eta}J^{AA'} - {}_{\xi}J^{AA'}{}_{\eta}J^{C}{}_{A'}]\phi_{AC|B_3\cdots B_{2s})}. \tag{43}$$

Taking the first term on the right hand side of (43) and applying (21) twice to pass $\nabla^{CC'}$ through ${}_{\xi}J_{(B_2|}{}^{A'}{}_{\eta}J^{A}{}_{A'}$ we obtain

$$\nabla^{CC'}{}_{\xi}J_{(B_2|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= {}_{\xi}J_{(B_2|}{}^{A'}[{}_{\eta}J^{A}{}_{A'}\nabla^{CC'} + \tfrac{2}{3}[M^{AC'}\nabla^{C}{}_{A'} - M^{C}{}_{A'}\nabla^{AC'}] + (\eta/6)[\epsilon^{C'}{}_{A'}W^{CA} - \epsilon^{CA}\tilde{W}^{C'}{}_{A'}] + K^{AD}[\nabla^{CC'},\nabla_{DA'}]$$
$$+ \tilde{K}_{A'}{}^{D'}[\nabla^{CC'},\nabla^{A}{}_{D'}]]\phi_{AC|B_3\cdots B_{2s})} + [\tfrac{1}{3}[M_{(B_2|}{}^{C'}\nabla^{CA'} - M^{CA'}\nabla_{(B_2|}{}^{C'}] + (\xi/6)[\epsilon^{C'A'}W^{C}{}_{(B_2|} - \epsilon^{C}{}_{(B_2|}\tilde{W}^{C'A'}]$$
$$+ K_{(B_2|}{}^{D}[\nabla^{CC'},\nabla_{D}{}^{A'}] + \tilde{K}^{A'D'}[\nabla^{CC'},\nabla_{(B_2|D'}]]_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}. \tag{44}$$

By cycling the two contracted indices A' and the index C' we can write

$${}_{\xi}J_{(B_2|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\nabla^{CC'} = [{}_{\xi}J_{(B_2|}{}^{C'}{}_{\eta}J^{A}{}_{A'} - {}_{\xi}J_{(B_2|A'|\eta}J^{AC'}]\nabla^{CA'}. \tag{45}$$

Using (21) to pass $\nabla^{CA'}$ through ${}_{\eta}J^{A}{}_{A'}$ and applying some of the identities (10)–(18) we can also write

$$\tfrac{2}{3}M_{(B_2|}{}^{C'}\nabla^{CA'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= [\tfrac{2}{3}M_{(B_2|}{}^{C'}{}_{\eta}J^{A}{}_{A'}\nabla^{CA'} - \tfrac{8}{9}M_{(B_2|}{}^{C'}M^{A}{}_{A'}\nabla^{CA'} - \tfrac{2}{9}(\eta-4)M_{(B_2|}{}^{C'}W^{AC} + \tfrac{2}{9}(2s-2)M_{(B_2|}{}^{C'}W^{AC}]\phi_{AC|B_3\cdots B_{2s})}. \tag{46}$$

Again applying the identities (10)–(18) we obtain

$$\nabla^{CC'}{}_{\xi}J_{(B_2|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= [[{}_{\xi+2}J_{(B_2|}{}^{C'}{}_{\eta-2}J^{A}{}_{A'}\nabla^{CA'} - {}_{\xi}J_{(B_2|A'\eta}J^{AC'}\nabla^{CA'} - \tfrac{4}{9}M_{(B_2|}{}^{C'}M^{A}{}_{A'}]\nabla^{CA'} - \tfrac{1}{6}(\eta-4){}_{\xi}J_{(B_2|}{}^{C'}W^{AC} - \tfrac{2}{9}(\eta-4)$$
$$\times M_{(B_2|}{}^{C'}W^{AC} - \tfrac{2}{3}M^{CA'}\nabla_{(B_2|}{}^{C'}{}_{\eta}J^{A}{}_{A'} + (\xi/6)W^{C}{}_{(B_2|\eta}J^{AC'} + \tfrac{1}{6}(\xi+4)\epsilon_{(B_2|}{}^{C}\tilde{W}^{A'C'}{}_{\eta}J^{A}{}_{A'}$$
$$+ \tfrac{1}{6}(2s-2){}_{\xi+4}J_{(B_2|}{}^{C'}W^{AC} + \tfrac{1}{3}(2s-2)W_{(B_2|}{}^{C}{}_{\eta}J^{AC'}]\phi_{AC|B_3\cdots B_{2s})}$$
$$+ (2s-2)[{}_{\xi}J_{(B_2|}{}^{A'}\tilde{K}_{A'}{}^{C'}\Psi^{ACM}{}_{|B_3|} + \tilde{K}^{A'C'}\Psi^{CM}{}_{(B_2B_3|\eta}J^{A}{}_{A'}]\phi_{ACM|B_4\cdots B_{2s})}. \tag{47}$$

Looking at the second term of the right hand side of (43) and using the definition (20) of the differential operator ${}_{\xi}J_{AA'}$, and making use of the symmetry in the indices A and C we have

$$\nabla_{(B_2|}{}^{C'}[{}_{\xi}J^{C}{}_{A'\eta}J^{AA'} - {}_{\xi}J^{AA'}{}_{\eta}J^{C}{}_{A'}]\phi_{AC|B_3\cdots B_{2s})}$$
$$= \nabla_{(B_2|}{}^{C'}[K^{CD}K^{AE}[\nabla_{DA'},\nabla_{E}{}^{A'}] + 2K^{CD}\tilde{K}^{A'E'}[\nabla_{DA'},\nabla^{A}{}_{E'}] + \tilde{K}_{A'}{}^{D'}\tilde{K}^{A'E'}[\nabla^{C}{}_{D'},\nabla^{A}{}_{E'}] + 2K^{CD}(\nabla_{DA'}K^{AE})\nabla_{E}{}^{A'}$$
$$+ 2K^{CD}(\nabla_{DA'}\tilde{K}^{A'E'})\nabla^{A}{}_{E'} + 2\tilde{K}_{A'}{}^{D'}(\nabla^{C}{}_{D'}K^{AE})\nabla_{E}{}^{A'} + 2\tilde{K}_{A'}{}^{D'}(\nabla^{C}{}_{D'}\tilde{K}^{A'E'})\nabla^{A}{}_{E'} + (2\eta/3){}_{\xi}J^{C}{}_{A'}M^{AA'}$$
$$+ (2\xi/3)M^{C}{}_{A'0}J^{AA'}]\phi_{AC|B_3\cdots B_{2s})}. \tag{48}$$

Applying the identities (10)–(18) and noting that the field spinor and the Killing spinors are symmetric and that the quantities $K_{AC}W_{B}{}^{C}$ and $M^{C}{}_{A'}M^{AA'}$ are antisymmetric, and by cycling indices in some of the terms involving both the Killing vector $M_{AA'}$ and one or other of the Killing spinors, (48) becomes

$$\nabla_{(B_2|}{}^{C'}[{}_{\xi}J^{C}{}_{A'\eta}J^{AA'} - {}_{\xi}J^{AA'}{}_{\eta}J^{C}{}_{A'}]\phi_{AC|B_3\cdots B_{2s})}$$
$$= \tfrac{1}{3}\nabla_{(B_2|}{}^{C'}[4K^{C}{}_{D}[M^{AA'}\nabla^{D}{}_{A'} + M^{DA'}\nabla^{A}{}_{A'}]\phi_{AC|B_3\cdots B_{2s})}$$
$$+ [2\eta_{\xi}J^{C}{}_{A'}M^{AA'} + 2\xi M^{C}{}_{A'\eta}J^{AA'}]\phi_{AC|B_3\cdots B_{2s})}$$
$$- (2s-2)[K^{AC}W^{M}{}_{|B_3|} + 3\tilde{K}_{A'D'}\tilde{K}^{A'D'}\Psi^{ACM}{}_{|B_3|}]\phi_{ACM|B_4\cdots B_{2s})}]. \tag{49}$$

We now note the following three relations: first,

$$\tfrac{1}{3}M^{CA'}\nabla_{(B_2|}{}^{C'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= [\tfrac{1}{3}\nabla_{(B_2|}{}^{C'}M^{CA'}{}_{\eta}J^{A}{}_{A'} - \tfrac{1}{6}W_{(B_2|}{}^{C}{}_{\eta}J^{AC'}$$
$$+ \tfrac{1}{6}\epsilon_{(B_2|}{}^{C}\tilde{W}^{A'C'}{}_{\eta}J^{A}{}_{A'}]\phi_{AC|B_3\cdots B_{2s})}. \tag{50}$$

Second, using the definition (20) of ${}_{\eta}J^{A}{}_{A'}$ we have

$$(1/6s)\nabla_{(B_2|}{}^{C'}K^{A}{}_{D}M^{CA'}\nabla^{D}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= -(1/6s)[\nabla_{(B_2|}{}^{C'}M^{CA'}{}_{\eta}J^{A}{}_{A'}$$
$$+ \nabla_{(B_2|}{}^{C'}\tilde{K}_{A'}{}^{D'}M^{A}{}_{D'}\nabla^{CA'}]\phi_{AC|B_3\cdots B_{2s})}, \tag{51}$$

and finally

$$(\eta/12s)\nabla_{(B_2|}{}^{C'}{}_{\xi}J^{C}{}_{A'}M^{AA'}\phi_{AC|B_3\cdots B_{2s})}$$
$$= (\eta/12s)\nabla_{(B_2|}{}^{C'}M^{CA'}{}_{\eta}J^{A}{}_{A'}\phi_{AC|B_3\cdots B_{2s})}, \qquad (52)$$
since
$$_{\xi}J^{(A}{}_{A'}M^{C)A'} = 0. \qquad (53)$$

Reforming (43) from its two pieces as given by (47) and (49) and making use of the three relations (50), (51), and (52) we obtain the desired result (42). One can also prove an analogous result for the cases $s = \frac{1}{2}s^1$. It therefore follows that if $\phi_{A_1\cdots A_{2s}}$ is a solution of
$$\nabla^{A}{}_{A'}\phi_{AA_2\cdots A_{2s}} = 0, \qquad (54)$$
then the new field
$$\chi_{B_1\cdots B_{2s}} = {}_{\xi}J_{(B_1|}{}^{A'}{}_{\eta}J^{A}{}_{A'}\phi_{A|B_2\cdots B_{2s})} \qquad (55)$$
is also a solution whenever
$$\begin{aligned}&\eta - 4s = \xi + 2 \text{ and } s \geqslant \tfrac{1}{2}, \quad \text{if } \Psi_{ABCD} = 0,\\ &\eta - 4s = 0 = \xi + 2 \text{ and } s \leqslant 1, \quad \text{if } \Psi_{ABCD} \neq 0.\end{aligned} \qquad (56)$$
Thus under these conditions the differential operator (40) is a symmetry operator of the spin s field equations.

Having verified that the new field $\chi_{A_1\cdots A_{2s}}$ is a solution of the massless spin s field equations we now form the following field:
$$\zeta_{A_1\cdots A_{2s}} = \chi_{A_1\cdots A_{2s}} - \tfrac{1}{2}\lambda\phi_{A_1\cdots A_{2s}}, \qquad (57)$$
where the field $\phi_{A_1\cdots A_{2s}}$ is obtained from a generalized Hertz potential which has $\bar{P}^{0'\cdots 0'}$ as its only nonzero component and where λ is the separation constant appearing in the solution for this one nonzero component. Clearly the field $\zeta_{A_1\cdots A_{2s}}$ is a solution of the spin s field equations. It is our intention to show that this field is identically zero and hence that the operator and its action as given in (40) characterizes the separation constant λ. Carter and McLenaghan[4] and Kalnins *et al.*[5] have already shown for $s \leqslant 1$ that λ is characterized by the operator (40). We can therefore restrict ourselves to a flat space-time, i.e., place $M = 0$ and consider the cases where $s > 1$. One can verify by explicit computation that $\chi_0 = \frac{1}{2}\lambda\phi_0$ and $\chi_{2s} = \frac{1}{2}\lambda\phi_{2s}$ and so the extremal components of $\zeta_{A_1\cdots A_{2s}}$ must vanish. Further from the form of the operator (40) and since t and φ are ignorable coordinates we can also conclude that the other components of $\zeta_{A_1\cdots A_{2s}}$ must have the same t and φ dependence as $\phi_{A_1\cdots A_{2s}}$. Thus we can write
$$\zeta_0 = \zeta_{2s} = 0 \qquad (58)$$
and
$$\zeta_j = f_j(r,\theta)e^{i\sigma t + im\varphi}. \qquad (59)$$

We will now compare the behavior of the left-and right-hand sides of (57) as $\theta \to 0$ and also as $\theta \to \pi$. The argument is an inductive one in that we will show that if $\zeta_{j-1} = 0$ then $\zeta_j = 0$. Note that we already have $\zeta_0 = 0$. If we write $Z_k = \tilde{\rho}^{*k}\zeta_k$ and suppose that for some $j > 0$ we have $Z_{j-1} = 0$ then from (31) we find that Z_j must satisfy
$$\begin{aligned}&[\mathscr{D}_0 + (2s - 2j + 1)(1/\tilde{\rho}^*)]Z_j = 0,\\ &[\mathscr{L}^{\dagger}_{j-s} + (2s - 2j + 1)(ia\sin\theta/\tilde{\rho}^*)]Z_j = 0.\end{aligned} \qquad (60)$$
Intergrating these equations we have
$$\zeta_j = A\tilde{\rho}^{*-(2s-j+1)}e^{-i\sigma\tilde{\rho}}\left[\frac{a - ir}{\sqrt{r^2 + a^2}}\right]^m$$
$$\times(1 + \cos\theta)^m(\sin\theta)^{s-j-m}e^{i\sigma t + im\varphi}. \qquad (61)$$
From the form of the solution it is clear that in the neighborhood of $\theta = 0$,
$$\zeta_j = \theta^{s-j-m}[b_0(r) + b_1(r)\theta + b_2(r)\theta^2 + \cdots]e^{i\sigma t + im\varphi}, \qquad (62)$$
where we have dropped the subscript j from the functions $b_i(r)$. Letting $\tilde{\theta} = \pi - \theta$ we find that in the neighborhood of $\theta = \pi$,
$$\zeta_j = \tilde{\theta}^{s-j+m}[\tilde{b}_0(r) + \tilde{b}_1(r)\tilde{\theta} + \tilde{b}_2(r)\tilde{\theta}^2 + \cdots]e^{i\sigma t + im\varphi}. \qquad (63)$$

We now consider the behavior of the right-hand side of (57) under the assumption that the field $\phi_{A_1\cdots A_{2s}}$ obtained from the generalized Hertz potential is regular at $\theta = 0$ and $\theta = \pi$. Firstly note that we can write $\phi_j = f_j(r,\theta)e^{i\sigma t + im\varphi}$ for each j and that from (31) we can generate a decoupled second order equation for ϕ_j. We find, when $M = 0$, that $\Phi_j = \tilde{\rho}^{*j}\phi_j$ satisfies the separable equation
$$[\mathscr{L}^{\dagger}_{j-s+1}\mathscr{L}_{s-j} + \Delta\mathscr{D}_1\mathscr{D}^{\dagger}_{s-j}$$
$$- 2(2s - 2j - 1)i\sigma\tilde{\rho}]\Phi_j = 0. \qquad (64)$$
It therefore follows that we can write ϕ_j as a sum over λ of terms of the form
$$(1/\tilde{\rho}^{*j})R(r;\lambda)S(\theta;\lambda). \qquad (65)$$
Now since to first order in θ the function $\tilde{\rho}^* = r - ia\cos\theta$ is independent of θ as $\theta \to 0$ and also as $\theta \to \pi$ it follows that to first order the θ dependence of ϕ_j in the neighborhood of $\theta = 0$ and $\theta = \pi$ will be determined by the behavior of the function S in these regions. From (64) we find that S satisfies
$$[\mathscr{L}^{\dagger}_{j-s+1}\mathscr{L}_{s-j} + 2(2s - 2j - 1)\sigma a\cos\theta]S(\theta;\lambda)$$
$$= -\lambda S(\theta;\lambda). \qquad (66)$$
From an examination of this equation we find that the regular solution for S behaves in the neighborhood of $\theta = 0$ as
$$S = \theta^{|s-j+m|}(c_0 + c_1\theta + c_2\theta^2 + \cdots), \qquad (67)$$
while in the neighborhood of $\theta = \pi$ we find
$$S = \tilde{\theta}^{|s-j-m|}(\tilde{c}_0 + \tilde{c}_1\tilde{\theta} + \tilde{c}_2\tilde{\theta}^2 + \cdots). \qquad (68)$$
Given that ϕ is nonsingular we must have in the neighborhood of $\theta = 0$ that
$$\tfrac{1}{2}\lambda\phi_j = \theta^{|s-j+m|}[g_0(r) + g_1(r)\theta$$
$$+ g_2(r)\theta^2 + \cdots]e^{i\sigma t + im\varphi}. \qquad (69)$$
Now $(\mathscr{T}\phi)_j$ will in general be formed from second-order derivatives of ϕ_j and first-order derivatives of both ϕ_{j-1} and ϕ_{j+1}. Thus in the neighborhood of $\theta = 0$ we will have
$$(\mathscr{T}\phi)_j = \theta^{|s-j+m|}[h_{-2}(r)\theta^{-2} + h_{-1}(r)\theta^{-1}$$
$$+ h_0(r) + h_1(r)\theta + \cdots]e^{i\sigma t + im\varphi}. \qquad (70)$$
Subtracting (69) from (70) we find that we must have
$$\zeta_j = \theta^{|s-j+m|-2}[k_0(r) + k_1(r)\theta$$
$$+ k_2(r)\theta^2 + \cdots]e^{i\sigma t + im\varphi} \qquad (71)$$

in the neighborhood of $\theta = 0$. Similarly in the neighborhood of $\theta = \pi$ we have

$$\zeta_j = \tilde{\theta}^{|s-j-m|-2}[\tilde{k}_0(r) + \tilde{k}_1(r)\theta + \tilde{k}_2(r)\theta^2 + \cdots]e^{i\sigma t + im\varphi}. \tag{72}$$

Now if $\zeta_{j-1} = 0$ and we assume that $\zeta_j \neq 0$ then it is also required that ζ_j have the behavior specified in (62) and (63). Thus if the two differing behaviors of ζ_j are to be consistent we must have

$$s - j - m \geqslant |s - j + m| - 2,$$
$$s - j + m \geqslant |s - j - m| - 2. \tag{73}$$

The above inequalities have a solution only when

$$-1 \leqslant m \leqslant 1 \text{ and } j \leqslant s + 1. \tag{74}$$

Thus for $|m| > 1$ we must have $\zeta_j = 0$ and so by induction all the components of $\zeta_{A_1 \cdots A_{2s}}$ will vanish and so for $|m| > 1$ we obtain

$$\mathscr{T}\phi = \tfrac{1}{2}\lambda\phi, \tag{75}$$

as desired.

To deal with the case where $|m| \leqslant 1$ we note that since we obtained $\phi_{A_1 \cdots A_{2s}}$ from a generalized Hertz potential we can write ϕ_j as

$$\phi_j = \mathscr{H}_j R_{-s} S_{+s}, \tag{76}$$

where $\mathscr{H}_j$ is a differential operator of order $2s$. We can therefore write (75) as

$$[\tilde{L}_{j-1}\mathscr{H}_{j-1} + L_j\mathscr{H}_j + \tilde{\tilde{L}}_{j+1}\mathscr{H}_{j+1}]R_{-s}S_{+s} = \tfrac{1}{2}\lambda\mathscr{H}_j R_{-s}S_{+s}, \tag{77}$$

where the differential operators L_j are of second order while the operators $\tilde{L}_j$ and $\tilde{\tilde{L}}_j$ are of first order. The only relations existing on the functions R_{-s} and S_{+s} by which this identity could hold are the Teukolsky equations for the functions R_{-s} and S_{+s}. Accordingly for some given j we must be able to write

$$[\tilde{L}_{j-1}\mathscr{H}_{j-1} + L_j\mathscr{H}_j + \tilde{\tilde{L}}_{j+1}\mathscr{H}_{j+1}] - \tfrac{1}{2}\lambda\mathscr{H}_j = \mathscr{G}_r\mathscr{T}_r + \mathscr{G}_\theta\mathscr{T}_\theta, \tag{78}$$

where $\mathscr{G}_r$ and $\mathscr{G}_\theta$ are in general differential operators of order $2s$ and $\mathscr{T}_r$ and $\mathscr{T}_\theta$ are the "Teukolsky operators", that is

$$\mathscr{T}_r \equiv \Delta\mathscr{D}^\dagger_{1-s}\mathscr{D}_0 + 2(2s-1)i\sigma r - \lambda,$$
$$\mathscr{T}_\theta \equiv \mathscr{L}^\dagger_{1-s}\mathscr{L}_s + 2(2s-1)\sigma a\cos\theta + \lambda, \tag{79}$$

for which $\mathscr{T}_r R_{-s} = 0$ and $\mathscr{T}_\theta S_{+s} = 0$. Now we note that Eq. (78) may be split into two parts, namely those terms which are independent of λ and those terms which are linear in λ. We find that

$$\mathscr{G}_r - \mathscr{G}_\theta = \tfrac{1}{2}\mathscr{H}_j, \tag{80}$$

and hence that

$$[\tilde{L}_{j-1}\mathscr{H}_{j-1} + L_j\mathscr{H}_j + \tilde{\tilde{L}}_{j+1}\mathscr{H}_{j+1}] - \tfrac{1}{2}\lambda\mathscr{H}_j - \tfrac{1}{2}\mathscr{H}_j\mathscr{T}_\theta = \mathscr{G}_\theta(\mathscr{T}_r + \mathscr{T}_\theta), \tag{81}$$

thus $\mathscr{G}_r$ and $\mathscr{G}_\theta$ will be uniquely determined.

We have established that the above identities amongst the various differential operators must hold for $|m| > 1$. Further, the m dependence of the various terms in any given identity is described by a polynomial in m. Since any given identity holds for an infinite number of values of m it must also hold when $m \leqslant 1$. We therefore have

$$\mathscr{T}\phi = \tfrac{1}{2}\lambda\phi, \tag{82}$$

for all values of m.

[1] S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).
[2] S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[3] A. L. Dudley and J. D. Finlay III, J. Math. Phy. **20**, 311 (1979).
[4] B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[5] E. G. Kalnins, W. Miller, Jr., and G. C. Williams, J. Math. Phys. **30**, 2360 (1989).
[6] G. F. Torres del Castillo, Proc. R. Soc. London Ser. A **400**, 119 (1985).
[7] G. F. Torres del Castillo, J. Math. Phys. **27**, 1583 (1986).
[8] G. F. Torres del Castillo, J. Math. Phys. **29**, 971 (1988).
[9] W. Dietz and R. Rüdiger, Proc. R. Soc. London Ser. A **375**, 361 (1981).
[10] W. Dietz and R. Rüdiger, Proc. R. Soc. London Ser. A **381**, 315 (1982).
[11] R. Penrose and W. Rindler, *Spinors and Spacetime Vol. 1. Two Spinor Calculus and Relativistic Fields* (Cambridge U.P., London, 1984).
[12] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., Oxford 1984).
[13] J. M. Cohen and L. S. Kegeles, Phys. Rev. D **19**, 1641 (1979).
[14] E. G. Kalnins, W. Miller Jr., and G. C. Williams, J. Math. Phys. **30**, 2925 (1989).

Lobachevskian Dirac fields

G. A. Kerimov
Institute of Physics Academy of Sciences of the Azerbaijan SSR, Baku 143, USSR

(Received 15 March 1989; accepted for publication 6 December 1989)

Free Dirac fields in a Lobachevskian space-time (the Euclideanization of the anti-de Sitter space-time) are studied. These fields, which are related in the flat-space-time limit to the Euclidean Dirac fields, act in a space $\mathscr{H}$, which again has Fock space structure. The two-point Green's function for Lobachevskian Dirac fields is obtained in terms of intrinsic geometric objects.

I. INTRODUCTION

In this article we continue the study of the quantum field theory in a Lobachevskian space-time (the Euclideanization of the anti-de Sitter space-time) by group-theoretical methods that we began in Ref. 1. Having constructed a Lobachevskian Bose field, our aim is to find a Lobachevskian Dirac field by a similar procedure.

The following is a short outline of the material to be presented here. We will construct Lobachevskian Dirac fields by starting with creation and annihilation operators satisfying the usual canonical anticommutation relations and then defining the fields in terms of these operators. This Fock space construction is carried out in Sec. IV and goes quite parallel to ordinary Euclidean Fock space construction. For this purpose in Sec. II we recall the definition of a Euclidean free Dirac field, while in Sec. III, we discuss the interrelation between P-induced and K-induced pictures of the principal series representation of $SO_0(4,1)$. Some mathematical results necessary for Sec. IV are given in Appendices A–C.

II. FREE EUCLIDEAN DIRAC FIELDS

In this section we summarize the theory of free Euclidean Dirac field of mass $m_f > 0$.

Euclidean spin 1/2 fields have been constructed several years ago by Osterwalder *et al.*[2,3] An important conclusion reached by these authors was that the construction of the interacting field models requires a doubling of the fermionic degrees of freedom. They also showed that the Hermiticity of the action has to be abandoned in favor of Osterwalder–Schrader positivity.

We find it expedient to use, throughout this paper, a formalism suggested by Osterwalder and Schrader.[2]

The Euclidean Fock space $\mathscr{E}$ for spin 1/2 fermions is given by

$$\mathscr{E} = \bigoplus_{n=0}^{\infty} \mathscr{E}^{(n)},$$

and

$$\mathscr{E}^{(n)} = \bigwedge^{n}_{n} \mathscr{E}^{(1)},$$

where $\wedge$ means the n-fold completely antisymmetric tensor product. The no-particle (vacuum) $\mathscr{E}^{(0)}$ consists of elements that are complex numbers and

$$\mathscr{E}^{(1)} = \mathscr{E}^{(1)}_{+} \oplus \mathscr{E}^{(1)}_{-},$$

where each $\mathscr{E}^{(1)}_{\pm}$ is isomorphic to $C^4 \otimes L^2(R^4)$. A state in $\mathscr{E}^{(1)}_{\pm}$ is described by a four-component function of the momentum $\mathbf{p}$ ($\mathbf{p}\in R^4$) labeled by an index j taking on the values 1,...,4. The vacuum is denoted by $|0\rangle$. In the standard fashion we introduce fermion annihilation and creation operators $b_j(\mathbf{p})$ and $b_j^*(\mathbf{p})$ ($\mathbf{p}\in R^4, j=1,...,4$) and antifermion annihilation and creation operators $d_j(\mathbf{p})$ and $d_j^*(\mathbf{p})$ ($\mathbf{p}\in R^4, j=1,...,4$), satisfying the anticommutation relations

$$\{b_j(\mathbf{p}),b_{j'}^*(\mathbf{p}')\} = \{d_j(\mathbf{p}),d_{j'}^*(\mathbf{p}')\} = \delta_{jj'}\delta^4(\mathbf{p}-\mathbf{p}'), \tag{2.1}$$

with all other anticommutators vanishing. The $b_j^*(\mathbf{p})$ is the creation operator related to the one-particle space $\mathscr{E}^{(1)}_{+}$ and $d_j^*(\mathbf{p})$ is the creation operator related to the one-particle space $\mathscr{E}^{(1)}_{-}$.

For the purposes of analyzing the flat space-time limit of the Lobachevskian quantum field theory, however, it is more convenient to introduce polar coordinates in R^4: $\mathbf{p} = p\mathbf{n}$, where $p>0$ and $\mathbf{n}$ is a unit vector in R^4. The anticommutation relations (2.1) can be rewritten as

$$\{b_j(p;\mathbf{n}),b_{j'}^*(p';\mathbf{n}')\} = \{d_j(p;\mathbf{n}),d_{j'}^*(p';\mathbf{n}')\} = \delta_{jj'}p^{-3}\delta(p-p')\delta^3(\mathbf{n}-\mathbf{n}'), \tag{2.2}$$

where $\delta^3(\mathbf{n}-\mathbf{n}')$ is the Dirac distribution on the unit sphere S^3 in R^4.

Now we will give a transformation law of the creation and annihilation operators under the inhomogeneous rotation group $ISO(4) = R^4 \ominus SO(4)$ (Euclidean group) which is the semi-direct product of R^4, considered as an additive group, and $SO(4)$

$$(\mathbf{a},k)\in ISO(4), \quad \mathbf{a}\in R, \quad k\in SO(4).$$

First we note that an arbitrary element $k\in SO(4)$ can be uniquely decomposed as follows;

$$k = \varkappa m, \tag{2.3}$$

with

 0022-2488/90/071745-10$03.00

$$\varkappa \equiv \varkappa(\mathbf{s}) = \begin{pmatrix} 1-\dfrac{s_1^2}{1+s_4} & -\dfrac{s_1s_2}{1+s_4} & -\dfrac{s_1s_3}{1+s_4} & s_1 \\ -\dfrac{s_1s_2}{1+s_4} & 1-\dfrac{s_2^2}{1+s_4} & -\dfrac{s_2s_3}{1+s_4} & s_2 \\ -\dfrac{s_1s_3}{1+s_4} & -\dfrac{s_2s_3}{1+s_4} & 1-\dfrac{s_3^2}{1+s_4} & s_3 \\ -s_1 & -s_2 & -s_3 & s_4 \end{pmatrix}, \tag{2.4}$$

and

$$m \equiv m(\mathbf{e}) = \begin{pmatrix} e_1^2+e_4^2-e_2^2-e_3^2 & 2e_1e_2+2e_3e_4 & 2e_1e_3-2e_2e_4 & 0 \\ 2e_1e_2-2e_3e_4 & e_2^2+e_4^2-e_1^2-e_3^2 & 2e_2e_3+2e_1e_4 & 0 \\ 2e_1e_3+2e_2e_4 & 2e_2e_3-2e_1e_4 & e_3^2+e_4^2-e_1^2-e_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.5}$$

where $\mathbf{s},\mathbf{e}$ are the unit vectors in R^4. The set of all matrices m given by Eq. (2.5) is a subgroup of SO(4) being isomorpic to SO(3). [As is easily seen, (2.5) is nothing but the Cayley's parametrization of SO(3).] On the Fock space $\mathscr{F}$ there is a unitary representation $\mathrm{U}(\mathbf{a},k)$ of ISO(4) [more precisely: of the covering group of ISO(4)], determined by

$$\mathrm{U}(\mathbf{a},k)|0\rangle = |0\rangle,$$

$$\mathrm{U}(\mathbf{a},k)b_j^*(p;\mathbf{n})\mathrm{U}^{-1}(\mathbf{a},k) = e^{-ip(k\mathbf{n},\mathbf{a})}\sum_{j'=1}^{4} b_{j'}^*(p;k\mathbf{n})S_{j'j}(r),$$

$$\mathrm{U}(\mathbf{a},k)d_j^*(p;\mathbf{n})\mathrm{U}^{-1}(\mathbf{a},k) = e^{-ip(k\mathbf{n},\mathbf{a})}\sum_{j'=1}^{4} S_{jj'}^{-1}(r)d_{j'}^*(p;k\mathbf{n}), \tag{2.6}$$

[here and in the following we use the scalar product $(\mathbf{p},\mathbf{x}) = \Sigma_{i=1}^4 p_i x_i$] where $r \in \mathrm{SO}(3)$ is a so-called "Wigner rotation" that is determined by

$$r = \varkappa^{-1}(k\mathbf{n})k\varkappa(\mathbf{n}). \tag{2.7}$$

Furthermore,

$$m \to S(m) \tag{2.8}$$

is a four-dimensional unitary (reducible) representation of SO(3)

$$S(m) = \begin{pmatrix} i\mathbf{e}\boldsymbol{\sigma} + e_4\sigma_4 & 0 \\ 0 & i\mathbf{e}\boldsymbol{\sigma} + e_4\sigma_4 \end{pmatrix}, \tag{2.9}$$

where m is determined by (2.5) and σ_i, $i = 1,2,3$ are Pauli matrices, $\sigma_4 = 1$. Note that the representation (2.8) may be extended to a representation of SO(4), such that

$$S(\varkappa) = \left(\frac{1+s_4}{2}\right)^{1/2} \begin{pmatrix} 1 & \dfrac{i\mathbf{s}\boldsymbol{\sigma}}{1+s_4} \\ \dfrac{i\mathbf{s}\boldsymbol{\sigma}}{1+s_4} & 1 \end{pmatrix}, \tag{2.10}$$

where $\varkappa \in \mathrm{SO}(4)$ is given by (2.4). We have the important relations

$$S^*(\varkappa) = S^{-1}(\varkappa) = S(\varkappa^{-1}), \tag{2.11}$$

$$\gamma_4 S(\varkappa)\gamma_4 = S^{-1}(\varkappa), \tag{2.12}$$

$$S(\varkappa)\gamma_4 S^{-1}(\varkappa) = \not{s}, \tag{2.13}$$

where

$$\gamma_j = \begin{pmatrix} 0 & -i\sigma_j \\ i\sigma_j & 0 \end{pmatrix}, \quad (j = 1,2,3), \quad \gamma_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are Euclidean γ matrices and $\not{s} = \Sigma_{i=1}^4 s_i\gamma_i$.

Having extended the definition of the four-dimensional representations S (2.8), we can split the rotation matrix appearing in (2.6) into three factors

$$S(r) = S(\varkappa^{-1}(k\mathbf{n}))S(k)S(\varkappa(\mathbf{n})), \tag{2.14}$$

where r is defined by (2.7). We are now in position to define two independent Euclidean fields $\psi_\alpha^{(1)}$ and $\psi_\alpha^{(2)}$, $\alpha = 1,\dots,4$. We set

$$\psi_\alpha^{(1)}(\mathbf{x}) = (2\pi)^{-2}\sum_{j=1}^{4}\int_{R^+}\int_{S^3}\frac{1}{(p^2+m_f^2)^{1/2}} \times\{b_j(p;\mathbf{n})u_\alpha^j(p;\mathbf{n})e^{ip(\mathbf{n},\mathbf{x})} + d_j^*(p;\mathbf{n})v_\alpha^j(p;\mathbf{n})e^{-ip(\mathbf{n},\mathbf{x})}\}p^3\,dp\,d\mathbf{n},$$

$$\psi_\alpha^{(2)}(\mathbf{x}) = (2\pi)^{-2}\sum_{j=1}^{4}\int_{R^+}\int_{S^3}\frac{1}{(p^2+m_f^2)^{1/2}} \times\{b_j^*(p;\mathbf{n})\hat{u}_\alpha^j(p;\mathbf{n})e^{-ip(\mathbf{n},\mathbf{x})} + d_j(p;\mathbf{n})\hat{v}_\alpha^j(p;\mathbf{n})e^{ip(\mathbf{n},\mathbf{x})}\}p^3\,dp\,d\mathbf{n}, \tag{2.15}$$

where $R^+ = \{p \in R : p \geqslant 0\}$ and $d\mathbf{n}$ is the Euclidean measure on the S^3. The $u^j(p;\mathbf{n}), v^j(p;\mathbf{n})$, etc. are the Euclidean spinors, which can be written in the form

$$u^j(p;\mathbf{n}) = S(\varkappa(\mathbf{n}))w_+^j(p),$$
$$v^j(p;\mathbf{n}) = S(\varkappa(\mathbf{n}))w_-^j(p),$$
$$\hat{u}^j(p;\mathbf{n}) = w_-^{jT}(p)S^{-1}(\varkappa(\mathbf{n})),$$
$$\hat{v}^j(p;\mathbf{n}) = w_-^{jT}(p)S^{-1}(\varkappa(\mathbf{n})), \tag{2.16}$$

i.e.,

$$u_\alpha^j(p;\mathbf{n}) = \sum_{\beta=1}^{4} S_{\alpha\beta}(\varkappa(\mathbf{n}))w_{+\beta}^j(p),$$

etc., where $w_\pm^j(p)$ $(p > 0, j = 1,\dots,4)$ are constant spinors defined by

$$w_{\pm}^{1}(p)=\begin{pmatrix}(ip\pm m_f)^{1/2}\\0\\0\\0\end{pmatrix},\quad w_{\pm}^{2}(p)=\begin{pmatrix}0\\(ip\pm m_f)^{1/2}\\0\\0\end{pmatrix},$$

$$w_{\pm}^{3}(p)=\begin{pmatrix}0\\0\\(-ip\pm m_f)^{1/2}\\0\end{pmatrix},\quad w_{\pm}^{4}(p)=\begin{pmatrix}0\\0\\0\\(-ip\pm m_f)^{1/2}\end{pmatrix}. \tag{2.17}$$

The spinors (2.16) satisfy the following relations:

$$\sum_{j=1}^{4} u_{\alpha}^{j}(p;\mathbf{n})\hat{u}_{\beta}^{j}(p;\mathbf{n})=(ip\not{n}+m_f)_{\alpha\beta},$$

$$\sum_{j=1} v_{\alpha}^{j}(p;\mathbf{n})\hat{v}_{\beta}^{j}(p;\mathbf{n})=-(ip\not{n}+m_f)_{\alpha\beta}, \tag{2.18}$$

which is a consequence of (2.13) and

$$\sum_{j=1}^{4} w_{\pm\alpha}^{j}(p)w_{\pm\beta}^{j}(p)=(ip\gamma_4\pm m_f)_{\alpha\beta}. \tag{2.19}$$

Observe that the Euclidean Dirac fields $\psi^{(1)}$ and $\psi^{(2)}$ anticommute as they should do, i.e.,

$$\{\psi_{\alpha}^{(i)}(\mathbf{x}),\psi_{\beta}^{(j)}(\mathbf{x}')\}=0,\quad \text{for all } \mathbf{x},\mathbf{x}'\in R^4. \tag{2.20}$$

(Nonvanishing anticommutators such as $\{\psi^{(1)},\psi^{(1)*}\}$ are irrelevant objects that play no further role.) These fields transform according to

$$\mathbf{U}(\mathbf{a},k)\psi_{\alpha}^{(1)}(\mathbf{x})\mathbf{U}^{-1}(\mathbf{a},k)=\sum_{\beta=1}^{4} S_{\alpha\beta}^{-1}(k)\psi_{\beta}^{(1)}(k\mathbf{x}+\mathbf{a}),$$

$$\mathbf{U}(\mathbf{a},k)\psi_{\alpha}^{(2)}(\mathbf{x})\mathbf{U}^{-1}(\mathbf{a},k)=\sum_{\beta=1}^{4} \psi_{\beta}^{(2)}(k\mathbf{x}+\mathbf{a})S_{\beta\alpha}(k). \tag{2.21}$$

The verification of the covariance properties (2.21) of the fields are based on the relation (2.14) and the fact that

$$\sum_{j=1}^{4} S_{jj'}^{-1}(m)w_{\pm\beta}^{j}(p)=\sum_{\alpha=1}^{4} S_{\beta\alpha}^{-1}(m)w_{\pm\alpha}^{j'}(p),$$

$$\sum_{j=1}^{4} S_{j'j}(m)w_{\pm\beta}^{j}(p)=\sum_{\alpha=1}^{4} w_{\pm\alpha}^{j'}(p)S_{\alpha\beta}(m), \tag{2.22}$$

where $m\in SO(3)$ [see Eqs. (2.9) and (2.17)].

The two-point Euclidean Green's function $G_{\alpha\beta}^{E}(\mathbf{x},\mathbf{x}')$ is given by

$$\begin{aligned}G_{\alpha\beta}(\mathbf{x},\mathbf{x}') &\equiv \langle 0|\psi_{\alpha}^{(1)}(\mathbf{x})\psi_{\beta}^{(2)}(\mathbf{x}')|0\rangle\\ &=(2\pi)^{-4}\int_{R^+}\int_{S^3}\frac{(ip\not{n}+m_f)_{\alpha\beta}}{p^2+m_f^2}\\ &\quad\times e^{ip(\mathbf{n},\mathbf{x}-\mathbf{x}')}p^3\,dp\,d\mathbf{n}\\ &=\left(\frac{m_f}{2\pi}\right)^2\frac{1}{|\mathbf{x}-\mathbf{x}'|}\Big\{K_1(m_f|\mathbf{x}-\mathbf{x}'|)\\ &\quad-\frac{\not{x}-\not{x}'}{|\mathbf{x}-\mathbf{x}'|}K_2(m_f|\mathbf{x}-\mathbf{x}'|)\Big\},\end{aligned} \tag{2.23}$$

where K_ν is the modified Bessel function of the third kind.

In concluding this section, we remark that each subspace carries a unitary (highly reducible) representation of ISO(4) which is equivalent to the induced representation of ISO(4) induced, in the sense of Mackey,[4] by the four-dimensional representation S of SO(4) defined above.

III. INTERTWINING OPERATORS FOR $SO_0(4,1)$

It is apparent from (2.6) and (2.21) that the formula (2.15) is essentially a relation between representations of ISO(4) which are induced by $R^4 \ominus SO(3)$ and SO(4) subgroups of ISO(4). Therefore, if we want to construct a Lobachevskian field possessing the property that it goes over into a Euclidean field in the flat-space limit, it is necessary to obtain a relation between representations of $SO_0(4,1)$ that are induced by minimal parabolic P (see, below) and maximal compact $K\cong SO(4)$ subgroups of $SO_0(4,1)$. [We mention that the minimal parabolic subgroup of $SO_0(4,1)$ reduce to $R^4\ominus SO(3)$ subgroup of ISO(4) in the Inönü–Wigner contraction[5] of $SO_0(4,1)$ with respect SO(4).] For this purpose we shall write down here (without proofs) the explicit form of intertwining integrals between the P- and K-induced realizations of the principal series representation of $SO_0(4,1)$.

Let us start the discussion with the fact that a four-dimensional Lobachevskian space Λ^4 can be realized as a four-dimensional hypersurface

$$\xi_0^2-\xi_1^2-\cdots-\xi_4^2=1,\quad \xi_0>0, \tag{3.1}$$

in a five-dimensional pseudo-Euclidean space $R^{4,1}$ with the bilinear form

$$[\xi,\eta]\equiv\xi_0\eta_0-\xi_1\eta_1-\cdots-\xi_4\eta_4.$$

In what follows we shall denote the boundary of Λ^4 by B, i.e., $B=\{\zeta\in\Lambda^4:[\zeta,\zeta]=0\}$. It is clear from (2.1) that the group of motions for Lobachevskian space Λ^4 is $SO_0(4,1)$ which acts transitively in Λ^4.

A convenient way to parametrize any noncompact semisimple Lie group is given by means of the Iwasawa decomposition. This tells us that, if in a noncompact semisimple Lie group G we pick a maximal compact subgroup K and a suitable Abelian subgroup A then there is a nilpotent subgroup N normalized by A, such that any group element $g\in G$ can be written uniquely as

$$g=kan, \tag{3.2}$$

with $k\in K, a\in A, n\in N$. For $G=SO_0(4,1)$ the elements of maximal compact subgroup $K\cong SO(4)$ are

$$k=\begin{pmatrix}1&0&0\\0&\check{k}&\\0&&\end{pmatrix},\quad \check{k}\in SO(4), \tag{3.3}$$

and we may take as the subgroup A the set of all elements in $SO_0(4,1)$ of the form

$$a \equiv a(t) = \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & 1 & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix}, \quad t \in R. \tag{3.4}$$

Then the elements of nilpotent subgroup N (here is actually Abelian) are[6,7]

$$n = \begin{pmatrix} 1 + y^2/2 & y^T & -y^2/2 \\ y & 1 & -y \\ y^2/2 & y^T & 1 - y^2/2 \end{pmatrix}, \quad y \in R^3, \tag{3.5}$$

where y is the column vector (y_1, y_2, y_3), y^T its transpose, and $y^2 = y_1^2 + y_2^2 + y_3^2$. Due to (2.3), the Iwasawa decomposition (3,2) can be carried further and one has

$$g = \varkappa man, \tag{3.6}$$

with

$$\varkappa \equiv \varkappa(\zeta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \breve{\varkappa}(\mathbf{s}) & \\ 0 & & \end{pmatrix}, \quad \zeta \equiv (1, \mathbf{s}), \quad \mathbf{s} \in S^3, \tag{3.7}$$

$$m = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \breve{m} & \\ 0 & & \end{pmatrix}, \tag{3.8}$$

where the matrices $\breve{\varkappa}$ and $\breve{m}$ are determined by Eqs. (2.4) and (2.5), respectively.

Another important parametrization of any element of $SO_0(4,1)$ is due to Cartan, i.e., the element of $SO_0(4,1)$ can be written as

$$g = \varkappa a k, \quad \varkappa, k \in K, \quad a \in A, \tag{3.9}$$

where the matrices $\varkappa$, a, and k are given by Eqs. (3.7), (3.4), and (3.3), respectively.

Obviously, the matrices m given by (3.8) form a subgroup of $SO_0(4,1)$ which we denote by M. The $M \cong SO(3)$ is at the same time a centralizer of A in K and a normalizer of N in K. Thus $P = MAN$ is also a subgroup of $SO_0(4,1)$ which is called the minimal parabolic subgroup of $SO_0(4,1)$. A finite-dimensional irreducible unitary representation of P has the form[6,7]

$$man \to \chi^\rho(a) D^j(m), \tag{3.10}$$

where $\chi^\rho(a) = e^{-i\rho t}$, $\rho > 0$, is a unitary character of A and D^j is an irreducible unitary representation of M. The number j is an integer for single-valued irreducible unitary representation of SO(3) and half-odd integer for the double-valued irreducible unitary representation [i.e., for the faithful irreducible unitary representation of the two-fold covering group SU(2) of SO(3)].

The principal series of irreducible unitary representations of $SO_0(4,1)$ is parametrized by (ρ, j) and is obtained by inducing the representations (3.10) of P to $SO_0(4,1)$. Denoting the Hilbert space for the finite dimensional representations D^j of SO(3) by H^j, the principal series of representations of $SO_0(4,1)$ may be realized (in a standard way) in the Hilbert space $L^2(X, H^j)$ of square-integrable functions over the coset space $X = SO_0(4,1)/P \cong K/M \cong S^3$ with values in H^j, and this Hilbert space may be identified with $L^2(S^3, H^j)$. The representations $T^{(\rho,j)}$ of $SO_0(4,1)$ are then given by the formula[6,7]

$$T^{(\rho,j)}(g) f(\mathbf{s}) = \chi^{\rho + 3i/2}(a^{-1}) D^j(m^{-1}) f(\mathbf{s}'), \tag{3.11}$$

where $g \in SO_0(4,1)$, $\mathbf{s} \in S^3$, $f \in L^2(S^3, H^j)$, and the quantities $\mathbf{s}'$ are determined from

$$g^{-1}\varkappa(\zeta) = \varkappa(\zeta')man, \quad \zeta \equiv (1, \mathbf{s}), \quad \zeta' \equiv (1, \mathbf{s}'). \tag{3.12}$$

To get a more explicit form of the transformation law (3.11) we use the Cartan decomposition of g (3.9); the factors $a \in A$ and $k \in K$ yield:

(a) $$T^{(\rho,j)}(a) f_\mu(\mathbf{s}) = [(a\zeta)_0]^{-3/2 + i\rho} f_\mu(\mathbf{s}'), \quad \mu = -j, \ldots, j, \tag{3.13}$$

with $\zeta' = a\zeta/(a\zeta)_0$,

(b) $$T^{(\rho,j)}(k) f_\mu(\mathbf{s}) = \sum_{\nu=-j}^{j} D^j_{\mu\nu}(\varkappa^{-1}(\zeta) k \varkappa(k^{-1}\zeta)) f_\nu(\mathbf{s}'), \tag{3.14}$$

with $\zeta' = k\zeta$.

In what follows we will give alternative realizations of $T^{(\rho,j)}$ on certain function spaces over $SO_0(4,1)/K \cong AN \cong \Lambda^4$.

Let $f \in L^2(S^3, H^j)$ and D^j be unitary irreducible representations of $K \cong SO(4)$ which remain irreducible when restricted to SO(3); define

$$\varphi(\xi) = \int_{S^3} [\xi, \zeta]^{-3/2 + i\rho} D^j(\varkappa(\zeta_\xi)) f(\mathbf{s}) d\mathbf{s}, \tag{3.15}$$

where $\xi \in \Lambda^4$, $\zeta \equiv (1, \mathbf{s}) \in B$, and $\zeta_\xi \equiv \alpha^{-1}(\xi)\zeta/(\alpha^{-1}(\xi)\zeta)_0$. Here $\alpha(\xi)$ is a coset representative for $\xi \in \Lambda^4 \cong AN$,

$$\alpha(\xi) = \begin{pmatrix} \xi_0 & \dfrac{1}{\xi_0 - \xi_4}\boldsymbol{\xi}^T & \dfrac{1}{\xi_0 - \xi_4} - \xi_0 \\ \boldsymbol{\xi} & 1 & \boldsymbol{\xi} \\ \xi_4 & \dfrac{1}{\xi_0 - \xi_4}\boldsymbol{\xi}^T & \dfrac{1}{\xi_0 - \xi_4} - \xi_4 \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}. \tag{3.16}$$

Then, $\varphi(\xi)$ are infinitely differentiable functions on Λ^4 with values in H^j and the representations $T^{(\rho,j)}$ of $SO_0(4,1)$ can be realized in a space of such functions. In this realization the representations $T^{(\rho,j)}$ of $SO_0(4,1)$ are given by

$$T^{(\rho,j)}(g)\varphi(\xi) = D^j(\alpha^{-1}(\xi) g \alpha(g^{-1}\xi))\varphi(g^{-1}\xi). \tag{3.17}$$

Furthermore, $\varphi(\xi)$ are eigenfunctions of the second-order Casimir operator of $SO_0(4,1)$.

Proof of these statements for $SO_0(2,1)$ can be found in Ref. 8.

In fact, the formula (3.15) gives the interrelation between the P-induced [see Eq. (3.11)] and K-induced [see Eq. (3.17)] pictures of the principal series.

We note that the verification of Eq. (3.17) is based on the relations

$$\varkappa\left(\frac{\alpha^{-1}(\xi)k\zeta}{(\alpha^{-1}(\xi)k\zeta)_0}\right)\varkappa^{-1}(k\zeta)k\varkappa(\zeta) = \alpha^{-1}(\xi)k\alpha(k^{-1}\xi)\varkappa\left(\frac{\alpha^{-1}(k^{-1}\xi)\zeta}{(\alpha^{-1}(k^{-1}\xi)\zeta)_0}\right), \tag{3.18}$$

which are straightforward exercises in matrix multiplication, and the fact that

$$a^{-1}\alpha(\xi)=\alpha(a^{-1}\xi), \quad a\in A. \tag{3.19}$$

IV. FREE LOBACHEVSKIAN DIRAC FIELDS

In this section we introduce free Lobachevskian Dirac fields, which will be related in the flat-space limit to the Euclidean fields constructed in Sec. II.

In analogy to the Euclidean case, we choose the one-particle Hilbert space for fermions $\mathscr{H}^{(1)}$ to be

$$\mathscr{H}^{(1)}=\mathscr{H}_{+}^{(1)}\oplus\mathscr{H}_{-}^{(1)},$$

where each $\mathscr{H}_{\pm}^{(1)}$ is isomorphic to $C^4\otimes h$ with h denoting the Hilbert space of all functions $F(p;\mathbf{n})$ on $R^+\times S^3$ for which

$$\|F\|\equiv\int_{R^+}\int_{S^3}|F(p;\mathbf{n})|^2\omega_{1/2}(p/\lambda)dp\,d\mathbf{n}<\infty,$$

where $\omega_{1/2}(p)=|\Gamma(ip+2)/\Gamma(ip+1/2)|^2$ is the Plancherel weight for the principal series representations $T^{(p,1/2)}$.[6] [We note that, each subspace $\mathscr{H}_{\pm}^{(1)}$ carries a unitary (highly reducible) representation of $SO_0(4,1)$ induced by the representation S (2.8) of SO(4) (see below).] Then the Lobachevskian Fermi Fock space $\mathscr{H}$ is defined as usual to be the Hilbert space completion of the alternating tensor algebra over $\mathscr{H}^{(1)}$

$$\mathscr{H}=\bigoplus_{n=0}^{\infty}\bigwedge^{n}\mathscr{H}^{(1)}.$$

The vacuum will be denoted by $|0\rangle$. We introduce (in a standard fashion) fermion annihilation and creation operators $B_j(p;\mathbf{n})$ and $B_j^*(p;\mathbf{n})$ ($p\in R^+,\mathbf{n}\in S^3, j=1,...,4$) and antifermion annihilation and creation operators $D_j(p;\mathbf{n})$ and $D_j^*(p;\mathbf{n})$ ($p\in R^+,\mathbf{n}\in S^3, j=1,...,4$) satisfying the anticommutation relation

$$\begin{aligned}\{B_j(p;\mathbf{n}),B_{j'}^*(p';\mathbf{n}')\}&\\ =\{D_j(p;\mathbf{n}),D_{j'}^*(p';\mathbf{n}')\}&\\ =\delta_{jj'}\lambda^{-3}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(\frac{1}{2}+ip/\lambda)}\right|^2&\delta(p-p')\delta^3(\mathbf{n}-\mathbf{n}'),\end{aligned} \tag{4.1}$$

all other anticommutators vanishing. The $B_j^*(p;\mathbf{n})$ is the creation operator related to the one-particle space $\mathscr{H}_{+}^{(1)}$ and $D_j^*(p;\mathbf{n})$ is the creation operator related to the one-particle space $\mathscr{H}_{-}^{(1)}$.

The Fock space $\mathscr{H}$ carry a unitary representation $\mathbb{U}(g)$ of $SO_0(4,1)$ defined by

$$\begin{aligned}&\mathbb{U}(g)|0\rangle=|0\rangle,\\ &\mathbb{U}(g)B_j^*(p;\mathbf{n})\mathbb{U}^{-1}(g)\\ &\quad=[(g\zeta)_0]^{-3/2+ip/\lambda}\sum_{j'=1}^{4}B_{j'}^*(p;\mathbf{n}')S_{jj'}(m),\\ &\mathbb{U}(g)D_j^*(p;\mathbf{n})\mathbb{U}^{-1}(g)\\ &\quad=[(g\zeta)_0]^{-3/2+ip/\lambda}\sum_{j'=1}^{4}S_{jj'}^{-1}(m)D_{j'}^*(p;\mathbf{n}'),\end{aligned} \tag{4.2}$$

where

$$g\varkappa(\zeta)=\varkappa(\zeta')man, \quad \zeta\equiv(1,\mathbf{n}), \quad \zeta'\equiv(1,\mathbf{n}'),$$

with notation as in Sec. III.

We are now prepared to define two independent Lobachevskian Dirac fields $\Psi_\mu^{(1)}$ and $\Psi_\mu^{(2)}$, $\mu=1,...,4$. Equations (2.15) and (3.15) tell us to do this by setting

$$\begin{aligned}\Psi_\mu^{(1)}(\xi)=(2\pi)^{-2}\sum_{j=1}^{4}\int_{R^+}\int_{S^3}&\frac{1}{\sqrt{p^2+(m_f+\lambda/2)^2}}\{B_j(p;\mathbf{n})U_\mu^j(p;\mathbf{n}_\xi)[\zeta,\xi]^{-3/2-ip/\lambda}+D_j^*(p;\mathbf{n})V_\mu^j(p;\mathbf{n}_\xi)\\ &\times[\zeta,\xi]^{-3/2+ip/\lambda}\}|\Gamma(2+ip/\lambda)/\Gamma(1/2+ip/\lambda)|^2dp\,d\mathbf{n},\\ \Psi_\mu^{(2)}(\xi)=(2\pi)^{-2}\sum_{j=1}^{4}\int_{R^+}\int_{S^3}&\frac{1}{\sqrt{p^2+(m_f+\lambda/2)^2}}\{B_j^*(p;\mathbf{n})\hat{U}_\mu^j(p;\mathbf{n}_\xi)[\zeta,\xi]^{-3/2+ip/\lambda}+D_j(p;\mathbf{n})\hat{V}_\mu^j(p;\mathbf{n}_\xi)\\ &\times[\zeta,\xi]^{-3/2-ip/\lambda}\}|\Gamma(2+ip/\lambda)/\Gamma(1/2+ip/\lambda)|^2dp\,d\mathbf{n},\end{aligned} \tag{4.3}$$

with

$$\begin{aligned}U^j(p;\mathbf{n}_\xi)&=S(\varkappa(\zeta_\xi))W^j_+(p),\\ V^j(p;\mathbf{n}_\xi)&=S(\varkappa(\zeta_\xi))W^j_-(p),\\ \hat{U}^j(p;\mathbf{n}_\xi)&=W^{jT}_+(p)S^{-1}(\varkappa(\zeta_\xi)),\\ \hat{V}^j(p;\mathbf{n}_\xi)&=W^{jT}_-(p)S^{-1}(\varkappa(\zeta_\xi)),\end{aligned}$$

i.e.,

$$U_\mu^j(p;\mathbf{n}_\xi)=\sum_{\nu=1}^{4}S_{\mu\nu}(\varkappa(\zeta_\xi))W^j_{+\nu}(p),$$

etc., where

$$\zeta_\xi=\alpha^{-1}(\xi)\zeta/(\alpha^{-1}(\xi)\zeta)_0, \quad \zeta\equiv(1,\mathbf{n}), \quad \zeta_\xi\equiv(1,\mathbf{n}_\xi)$$

and

$$W^1_\pm(p)=\begin{pmatrix}[ip\pm(m_f+\lambda/2)]^{1/2}\\0\\0\\0\end{pmatrix},$$

$$W^2_\pm(p)=\begin{pmatrix}0\\ [ip\pm(m_f+\lambda/2)]^{1/2}\\0\\0\end{pmatrix},$$

$$W^3_\pm(p)=\begin{pmatrix}0\\0\\ [-ip\pm(m_f+\lambda/2)]^{1/2}\\0\end{pmatrix},$$

$$W^4_{\pm}(p)=\begin{pmatrix}0\\0\\0\\[-ip\pm(m_f+\lambda/2)]^{1/2}\end{pmatrix}.$$

It follows from (4.2) that Lobachevskian Dirac fields transform as [cf. Eq. (3.17)]

$$\mathbb{U}(g)\Psi^{(1)}_\mu(\xi)\mathbb{U}^{-1}(g)=\sum_{\nu=1}^4 S^{-1}_{\mu\nu}(\alpha^{-1}(g\xi)g\alpha(\xi))\Psi^{(1)}_\nu(g\xi),$$

$$\mathbb{U}(g)\Psi^{(2)}_\mu(\xi)\mathbb{U}^{-1}(g)=\sum_{\nu=1}^4 \Psi^2_\nu(g\xi)S_{\nu\mu}(\alpha^{-1}(g\xi)g\alpha(\xi)).$$

In Appendix C we show that the Lobachevskian Dirac fields are actually anticommutative, i.e.,

$$\{\Psi^{(i)}_\mu(\xi),\Psi^{(j)}_\nu(\xi')\}=0,\quad \text{for all } \xi,\xi'\in\Lambda^4. \tag{4.4}$$

The two-point Lobachevskian Green's function $G^L(\xi,\xi')$ is defined by

$$\begin{aligned}G^L_{\mu\nu}(\xi,\xi')&\equiv\langle 0|\Psi^{(1)}_\mu(\xi)\Psi^{(2)}_\nu(\xi')|0\rangle\\
&=\frac{\lambda^3}{(2\pi)^4}\sum_{\tau,\beta=1}^4\int_{R^+}\int_{S^3}\frac{1}{p^2+(m_f+\lambda/2)^2}S_{\mu\tau}(\varkappa(\zeta_\xi))(ip\gamma_4+m_f+\lambda/2)_{\tau\beta}\\
&\quad\times S^{-1}_{\beta\nu}(\varkappa(\zeta_{\xi'}))[\xi,\xi]^{-3/2-ip/\lambda}[\xi,\xi']^{-3/2+ip/\lambda}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2dp\,d\mathbf{n}\\
&=\frac{\lambda^3}{(2\pi)^4}\sum_{\tau,\beta=1}^4\int_{R^+}\int_{S^3}\frac{1}{p^2+(m_f+\lambda/2)^2}S_{\mu\tau}(\varkappa(\zeta))(ip\gamma_4+m_f+\lambda/2)_{\tau\beta}\\
&\quad\times S^{-1}_{\beta\nu}(\varkappa(\zeta_\Xi))[\zeta,\Xi]^{-3/2+ip/\lambda}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2dp\,d\mathbf{n},\end{aligned}\tag{4.5}$$

where

$$\zeta_\Xi=\alpha^{-1}(\Xi)\zeta/(\alpha^{-1}(\Xi)\zeta)_0,\quad \alpha^{-1}(\Xi)\equiv\alpha^{-1}(\xi')\alpha(\xi),$$
$$\Xi=\alpha(\Xi)\mathring{\Xi},\quad \mathring{\Xi}\equiv(1,0,0,0,0).$$

The last equality in (4.5) is based on quasi-invariance of the measure $d\mathbf{n}$.

Let us now calculate the explicit form of $G^L(\xi,\xi')$. For this purpose, we introduce polar coordinates in Λ^4:

$$\Lambda^4\ni\Xi=(\cosh\lambda r,\mathbf{s}\sinh\lambda r),\quad \mathbf{s}\in S^3,\quad 0\leqslant r<\infty.$$

Then

$$\alpha(\Xi)=\varkappa(\eta)a\varkappa^{-1}(a^{-1}\eta/(a^{-1}\eta)_0), \tag{4.6}$$

where

$$\eta\equiv(1,\mathbf{s}),\quad a\equiv a(\lambda r)\in A.$$

[The Eq. (4.6) is due to the Cartan decomposition.] We insert (4.6) and make a change in the variable of integration over. This gives

$$\begin{aligned}G^L_{\mu\nu}(\xi,\xi')=&\frac{\lambda^3}{(2\pi)^4}\sum_{\tau,\beta=1}^4\int_{R^+}\int_{S^3}\frac{1}{p^2+(m_f+\lambda/2)^2}\\
&\times S_{\mu\tau}(\varkappa(\zeta'))(ip\gamma_4+m_f+\lambda/2)_{\tau\beta}\\
&\times S^{-1}_{\beta\nu}\left(\varkappa\left(\frac{\alpha^{-1}(\Xi)\zeta'}{(\alpha^{-1}(\Xi)\zeta')_0}\right)\right)[\zeta,\Xi]^{-3/2+ip/\lambda}\\
&\times\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2dp\,d\mathbf{n},\end{aligned}\tag{4.7}$$

where

$$\zeta'\equiv\varkappa(\eta)\zeta,\quad \zeta\equiv(1,\mathbf{n}).$$

We now use the relations

$$\varkappa(\varkappa(\eta)\zeta)\varkappa^{-1}\left(\frac{\alpha^{-1}(\Xi)\varkappa(\eta)\zeta}{(\alpha^{-1}(\Xi)\varkappa(\eta)\zeta)_0}\right)=\varkappa(\eta)\varkappa(\zeta)\varkappa^{-1}\left(\frac{a^{-1}\zeta}{(a^{-1}\zeta)}\right)\varkappa^{-1}\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right), \tag{4.8}$$

which is a consequence of (3.18) in the special case $k=\varkappa(\eta)$ and

$$\varkappa^{-1}(\varkappa(\eta)\zeta)\varkappa^{-1}\left(\frac{\alpha^{-1}(\Xi)\varkappa(\eta)\zeta}{(\alpha^{-1}(\Xi)\varkappa(\eta)\zeta)_0}\right)=\varkappa^{-1}(\eta)\varkappa^{-1}(\zeta)\varkappa^{-1}\left(\frac{a^{-1}\zeta}{(a^{-1}\zeta)_0}\right)\varkappa^{-1}\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right). \tag{4.9}$$

Relation (4.9) follows from the relation (4.8) and the fact

$$\varkappa^2(\varkappa(\eta)\zeta)=\varkappa(\eta)\varkappa^2(\zeta)\varkappa(\eta). \tag{4.10}$$

Thus we obtain

$$G^L(\xi,\xi')=\frac{\lambda^3}{(2\pi)^4}\int_{R^+}\int_{S^3}\frac{1}{p^2+(m_f+\lambda/2)^2}\left\{ip\gamma_4 S^{-1}(\varkappa(\eta))S^{-1}(\varkappa(\zeta))S^{-1}\left(\varkappa\left(\frac{a^{-1}\zeta}{(a^{-1}\zeta)_0}\right)\right)S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right.$$
$$\left.+(m_f+\lambda/2)S(\varkappa(\eta))S(\varkappa(\zeta))S^{-1}\left(\varkappa\left(\frac{a^{-1}\zeta}{(a^{-1}\zeta)_0}\right)\right)S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right\}[\zeta,a\mathring{\Xi}]^{-3/2+ip/\lambda}$$
$$\times\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2 dp\,d\mathbf{n}. \tag{4.11}$$

In deriving Eq. (4.11) we have used the relation

$$S(\varkappa)\gamma_4=\gamma_4 S^{-1}(\varkappa),$$

[see Eq. (2.12)] and the fact that $S(\cdot)$ is a representation of SO(4).

We now introduce the polar coordinates on S^3 given by

$$\mathbf{n}=(\sin\beta\sin\theta\sin\varphi,\ \sin\beta\sin\theta\cos\varphi,$$
$$\sin\beta\cos\theta,\cos\beta),$$
$$0<\beta,\theta<\pi,\quad 0<\varphi<2\pi.$$

Then the Euclidean measure on S^3 has the form

$$d\mathbf{n}=\sin^2\beta\sin\theta\,d\beta\,d\theta\,d\varphi.$$

After integration over the angles θ and φ we find that

$$G^L(\xi,\xi')=\frac{2\lambda^3}{(2\pi)^3}\int_0^\infty\frac{1}{p^2+(m_f+\lambda/2)^2}\times\left\{ip\gamma_4 S^{-1}(\varkappa(\eta))S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\mathscr{I}_1+(m_f+\lambda/2)S(\varkappa(\eta))\times S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\mathscr{I}_2\right\}, \tag{4.12}$$

where

$$\mathscr{I}_1\equiv\int_0^\pi\left(-\sinh\frac{\lambda r}{2}+\cosh\frac{\lambda r}{2}\cos\beta\right)\times(\cosh\lambda r-\cos\beta\sinh\lambda r)^{-2+ip/\lambda}\sin^2\beta\,d\beta, \tag{4.13}$$

$$\mathscr{I}_2\equiv\int_0^\pi\left(\cosh\frac{\lambda r}{2}-\sinh\frac{\lambda r}{2}\cos\beta\right)\times(\cosh\lambda r-\cos\beta\sinh\lambda r)^{-2+ip/\lambda}\sin^2\beta\,d\beta. \tag{4.14}$$

In Appendix A we explicitly evaluate the integrals $\mathscr{I}_1$ and $\mathscr{I}_2$.

They are

$$\mathscr{I}_1=\frac{\pi}{\sinh\lambda r}\frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)}P_{3/2,-1/2}^{-1/2+ip/\lambda}(\cosh\lambda r),$$

$$\mathscr{I}_2=\frac{\pi}{\sinh\lambda r}\frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)}P_{3/2,1/2}^{-1/2+ip/\lambda}(\cosh\lambda r),$$

where $P^l_{mn}(z)$ is the generalized Legendre function of the first kind (see Appendix A).

Finally, the remaining p integral can be evaluated in terms of generalized Legendre function of the second kind $Q^l_{mn}(z)$ (see Appendix B):

$$G^L_{\mu\nu}(\xi,\xi')=\frac{2\lambda^3}{(2\pi)^3\sinh\lambda r}\frac{\Gamma(m_f/\lambda+3/2)}{\Gamma(m_f/\lambda-1/2)}\left\{\left[S(\varkappa(\eta))\gamma_4\times S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right]_{\mu\nu}Q^{m_f/\lambda}_{3/2,-1/2}(\cosh\lambda r)+\left[S(\varkappa(\eta))S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right]_{\mu\nu}\times Q^{m_f/\lambda}_{3/2,1/2}(\cosh\lambda r)\right\}. \tag{4.15}$$

Furthermore we check that the Euclidean Green's function indeed a limiting case of the Lobachevskian Green's function, i.e., that

$$G^L_{\mu\nu}(\xi,\xi')\underset{(\lambda\to 0)}{\to}G^E_{\mu\nu}(\mathbf{x},\mathbf{x}'),$$

with $|\mathbf{x}-\mathbf{x}'|=r$, since

$$\lim_{\lambda\to 0}Q^{m_f/\lambda}_{\mu\nu}(\cosh\lambda r)=K_{\mu-\nu}(m_f r)$$

[see (B9)].

After all, it remains to be shown that the Lobachevskian Dirac fields (4.3) goes over in the flat-space limit into the Euclidean Dirac fields (2.15). To do so, it is convenient to introduce an horospherical coordinate x_μ, $\mu=1,\ldots,4$ on Λ^4

$$\xi_0=\cosh\lambda x_4+(\lambda^2/2)x^2e^{\lambda x_4},$$
$$\xi_i=\lambda x_i,\quad i=1,2,3,$$
$$\xi_4=\sinh\lambda x_4+(\lambda^2/2)x^2e^{\lambda x_4},$$

where $-\infty<x_\mu<\infty$ and $x^2=x_1^2+x_2^2+x_3^2$.

First we denote that the discussion of the Lobachevskian plane wave factor $[\xi,\zeta]^{-3/2+ip/\lambda}$ is analogous here to the one given for the scalar case leading, in the limit $\lambda\to 0$, to the plane wave factor $\exp[ip(\mathbf{n},\mathbf{x})]$, i.e.,

$$\lim_{\lambda\to 0}[\xi,\zeta]^{-3/2+ip/\lambda}=\exp[ip(\mathbf{n},\mathbf{x})],$$

(see Ref. 1). Hence using the relations

$$\lim_{\lambda\to 0}\alpha(\xi)=1$$

and

$$\lim_{\lambda\to 0}\lambda^3\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2=p^3,$$

it follows that the Lobachevskian Dirac fields (4.3) reduce in the flat-space limit $\lambda\to 0$ to the Euclidean Dirac fields (2.15).

APPENDIX A: EVALUATION OF THE INTEGRALS (4.13) AND (4.14)

In this appendix we shall evaluate the following integrals:

$$\mathcal{I}_1 \equiv \int_0^\pi \left(- \sinh \frac{\lambda r}{2} + \cosh \frac{\lambda r}{2} \cos \beta \right) (\cosh \lambda r - \cos \beta \sinh \lambda r)^{-2 + ip/\lambda} \sin^2 \beta \, d\beta,$$

and

$$\mathcal{I}_2 \equiv \int_0^\pi \left(\cosh \frac{\lambda r}{2} - \sinh \frac{\lambda r}{2} \cos \beta \right) (\cosh \lambda r - \cos \beta \sinh \lambda r)^{-2 + ip/\lambda} \sin^2 \beta \, d\beta.$$

The calculation is based on the formula [see Eq. 10.3.7(1) of Ref. 9]

$$\int_0^\pi (\cosh \omega - \cos \beta \sinh \omega)^\sigma C_k^1(\cos \beta) \sin^2 \beta \, d\beta = \frac{(-1)^k 2\sqrt{\pi}\Gamma(3/2)\Gamma(\sigma+1)}{k!\Gamma(\sigma-k+1)} \times \sinh^{-1} \omega P_{\sigma+1}^{-1-k}(\cosh \omega),$$

where $C_k^m(z)$ and $P_\sigma^k(z)$ are the Gegenbauer polynomials and the associated Legendre functions of the first kind, respectively.[10] By making use of the relations

$$C_0^1(\cos \beta) = 1, \quad C_1^1(\cos \beta) = 2 \cos \beta,$$

it is now straightforward to eavluate the integrals $\mathcal{I}_1$ and $\mathcal{I}_2$; we obtain

$$\mathcal{I}_1 = \pi \sinh^{-1} \lambda r \left\{ - \sinh \frac{\lambda r}{2} P_{-1+ip/\lambda}^{-1}(\cosh \lambda r) + \cosh \frac{\lambda r}{2} \frac{\Gamma(-1+ip/\lambda)}{\Gamma(-2+ip/\lambda)} P_{-1+ip/\lambda}^{-2}(\cosh \lambda r) \right\} \tag{A1}$$

and

$$\mathcal{I}_2 = \pi \sinh^{-1} \lambda r \left\{ \cosh \frac{\lambda r}{2} P_{-1+ip/\lambda}^{-1}(\cosh \lambda r) - \sinh \frac{\lambda r}{2} \frac{\Gamma(-1+ip/\lambda)}{\Gamma(-2+ip/\lambda)} P_{-1+ip/\lambda}^{-2}(\cosh \lambda r) \right\}. \tag{A2}$$

However, the integrals $\mathcal{I}_1$ and $\mathcal{I}_2$ can also be expressed in terms of the functions $P_{mn}^l(z)$.[9] (The explicit definitions of these functions will be given below.) Indeed, using the relation [see Eq. 6.3.5(5) of Ref. 9]

$$P_l^m(\cosh \omega) = \frac{\Gamma(l+m+1)}{\Gamma(l+1)} P_{m0}^l(\cosh \omega),$$

and the recurrence formulas [see Eqs. 6.5.6(7) and 6.5.6(8) of Ref. 9]

$$P_{m+1/2,1/2}^{l+1/2}(\cosh \omega) = \cosh \frac{\omega}{2} P_{m0}^l(\cosh \omega) + \sinh \frac{\omega}{2} P_{m+1,0}^l(\cosh \omega)$$

and

$$P_{m+1/2,-1/2}^{l+1/2}(\cosh \omega) = \sinh \frac{\omega}{2} P_{m0}^l(\cosh \omega) + \cosh \frac{\omega}{2} P_{m+1,0}^l(\cosh \omega),$$

with l a complex number, and m integer, we find

$$\mathcal{I}_1 = \frac{\pi}{\sinh \lambda r} \frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)} P_{3/2,-1/2}^{-1/2+ip/\lambda}(\cosh \lambda r) \tag{A3}$$

and

$$\mathcal{I}_2 = \frac{\pi}{\sinh \lambda r} \frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)} P_{3/2,1/2}^{-1/2+ip/\lambda}(\cosh \lambda r). \tag{A4}$$

Here we have used the symmetry relation [see Eq. 6.3.6(1) of Ref. 9]

$$P_{mn}^l(z) = P_{-m,-n}^l(z). \tag{A5}$$

The functions $P_{mn}^l(z)$ (with $z = \cosh \omega$, l a complex number, and m and n which are simultaneously either integers or half odd integers) have been discussed extensively in the book by Vilenkin referred to previously [see Chap. 6 of Ref. 9]. They can be expressed in terms of hypergeometric functions ${}_2F_1$:

$$P_{mn}^l(z) = \frac{\Gamma(l+1-n)}{\Gamma(l+1-m)} \frac{2^{-m}}{\Gamma(1+m-n)} \times (z-1)^{(m-n)/2}(z+1)^{(m+n)/2} \times {}_2F_1(m-l, m+l+1; 1+m-n; (1-z)/2) \quad (m \geqslant n) \tag{A6}$$

and

$$P_{mn}^l(z) = \frac{\Gamma(l+1+n)}{\Gamma(l+1+m)} \frac{2^{-n}}{\Gamma(1+n-m)} \times (z-1)^{(n-m)/2}(z+1)^{(n+m)/2} \times {}_2F_1(n-l, n+l+1; 1+n-m; (1-z)/2) \quad (n \geqslant m). \tag{A7}$$

We shall call the function $P_{mn}^l(z)$ a generalized Legendre function of the first kind, since

$$P_l(z) = P_{00}^l(z), \tag{A8}$$

where $P_l(z)$ is the Legendre function of the first kind.[10]

In concluding this appendix we give some of the properties of the $P_{mn}^l(z)$ which we need later:

(a) $$\int_0^\infty P_{mn}^{-1/2+i\rho}(z) P_{mn}^{-1/2-i\rho}(t) \rho \tanh(\rho + i\varepsilon) d\rho = \delta(z-t), \tag{A9}$$

if $M = 0$ or $1/2$, where

$$M = \begin{cases} \min(|m|,|n|), & \text{for } mn \geqslant 0, \\ 0, & \text{for } mn < 0, \end{cases}$$

and

$\varepsilon = 0$, if m and n are integers,

$\varepsilon = 1/2$, if m and n are half odd integers.

(b) $P_{mn}^{-l-1}(z) = (-1)^{m-n}\dfrac{\Gamma(l+1-m)\Gamma(l+1+m)}{\Gamma(l+1-n)\Gamma(l+1+n)} \times P^{l}_{mn}(z).$ (A10)

APPENDIX B: INTEGRALS REPRESENTING GENERALIZED LEGENDRE FUNCTIONS OF THE SECOND KIND

In this appendix we want to evaluate the integrals

$$\mathscr{F}_1 \equiv \int_0^\infty \frac{ip}{p^2+(m_f+\lambda/2)^2}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2 \times \frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)} P_{3/2,-1/2}^{-1/2+ip/\lambda}(\cosh\lambda r)dp, \tag{B1}$$

and

$$\mathscr{F}_2 \equiv \int_0^\infty \frac{1}{p^2+(m_f+\lambda/2)^2}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2 \times \frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)} P_{3/2,1/2}^{-1/2+ip/\lambda}(\cosh\lambda r)dp. \tag{B2}$$

It will be useful to introduce the function $Q^{l}_{mn}(z)$ by

$$Q^{l}_{mn}(z) = \frac{2^l\Gamma(l+1-m)\Gamma(l+1+m)}{\Gamma(2l+2)} \times \left(\frac{z+1}{z-1}\right)^{(m+n)/2}(z-1)^{-l-1} \times {}_2F_1(l+1+m,l+1+n;2l+2;2/(1-z)) \tag{B3}$$

which we shall call the generalized Legendre function of the second kind. The integrals (B1) and (B2) can be evaluated by means of

$$\int_0^\infty \frac{1}{(l+1/2)^2+\rho^2}\rho\tanh(\rho+i\varepsilon)\frac{\Gamma(1/2+i\rho+m)}{\Gamma(1/2+i\rho+n)} \times P_{mn}^{-1/2+i\rho}(\cosh\omega)d\rho = (-1)^{n-m}\frac{\Gamma(l+1-n)}{\Gamma(l+1-m)}Q^{l}_{mn}(\cosh\omega), \tag{B4}$$

valid for $\mathrm{Re}\, l > -1/2$, $m \pm n > 0$, $M = 0$ or $1/2$, where

$$M = \begin{cases}\min(|m|,|n|), & \text{for } mn>0,\\ 0, & \text{for } mn<0,\end{cases}$$

and $\varepsilon = 0$ or $1/2$ according as m and n are integers or half odd integers. This formula can be derived from the integral [see Eq. (7.12) of Ref. 11]

$$\int_1^\infty P_{mn}^{-1/2+i\rho}(z)Q^{l}_{mn}(z)dz = \frac{\Gamma(l-m+1)\Gamma(1/2+i\rho-n)}{\Gamma(l-n+1)\Gamma(1/2+i\rho-m)}\frac{1}{\rho^2+(l+1/2)^2}, \tag{B5}$$

($\mathrm{Re}\, l > -1/2, m \pm n > 0$) with the aid of Eq. (A9). By making use of Γ-function identities, it is now straightforward to evaluate the integrals $\mathscr{F}_1$ and $\mathscr{F}_2$; we obtain

$$\mathscr{F}_1 = -\frac{\Gamma(m_f/\lambda+3/2)}{\Gamma(m_f/\lambda-1/2)}Q_{3/2,-1/2}^{m_f/\lambda}(\cosh\lambda r), \tag{B6}$$

and

$$\mathscr{F}_2 = \frac{1}{\lambda}\frac{\Gamma(m_f/\lambda+1/2)}{\Gamma(m_f/\lambda-1/2)}Q_{3/2,1/2}^{m_f/\lambda}(\cosh\lambda r). \tag{B7}$$

Let us now determine the limit $\lambda\to 0$ of $Q_{\mu\nu}^{m/\lambda}(\cosh\lambda r)$. In order to do this it is convenient to perform an analytic continuation of the hypergeometric function in Eq. (B3). Using a formula 2.10(2) of Ref. 10 we obtain from Eq. (B3)

$$Q^{l}_{\mu\nu}(z) = \frac{\Gamma(l+1+\mu)\Gamma(\nu-\mu)}{2^{1+\mu}\Gamma(l+1-\nu)}(z+1)^{(\mu+\nu)/2}(z-1)^{(\mu-\nu)/2}{}_2F_1\left(l+1+\mu,-l+\mu;1-\nu+\mu;\frac{1-z}{2}\right)$$
$$+\frac{\Gamma(l+1-\mu)\Gamma(-\nu+\mu)}{2^{1+\nu}\Gamma(l-\nu+1)}(z+1)^{(\mu+\nu)/2}(z-1)^{(\nu-\mu)/2}{}_2F_1\left(l+1+\nu,-l+\nu;1+\nu-\mu;\frac{1-z}{2}\right).$$

Now it is quite straightforward to establish the following relation for $Q^{l}_{\mu\nu}(z)$:

$$\lim_{\lambda\to 0} Q_{\mu\nu}^{m/\lambda}(\cosh\lambda r) = \frac{1}{2}\left\{\Gamma(\nu-\mu)\left(\frac{r}{2}\right)^{\mu-\nu}{}_0F_1\left(\mu-\nu;\frac{r^2}{4}\right) + \Gamma(\mu-\nu)\left(\frac{r}{2}\right)^{\nu-\mu}{}_0F_1\left(\nu-\mu;\frac{r^2}{4}\right)\right\} \tag{B8}$$

$$= K_{\mu-\nu}(mr). \tag{B9}$$

In deriving the last equality we used the definition of the modified Bessel functions [see Eqs. 7.2(12) and 7.2(13) of Ref. 12]

$$I_\nu(z) = [(z/2)^\nu/\Gamma(\nu+1)]{}_0F_1(\nu+1;z^2/4),$$

and

$$K_\nu(z) = (\pi/2\sin\nu\pi)[I_{-\nu}(z) - I_\nu(z)].$$

For the case $\mu=0$ or $\nu=0$ Eq. (B9) reduces to the formula 7.8(4) of Ref. 12, since

$$Q^{\mu}_{l}(z) = e^{i\mu\pi}\frac{\Gamma(l+1)}{\Gamma(l+1-\mu)}Q^{l}_{\mu 0}(z) = e^{i\mu\pi}\frac{\Gamma(1+l+\mu)}{\Gamma(1+l)}Q^{l}_{0\mu}(z),$$

which follows from (B3) and from the explicit expression for the associated Legendre functions of the second kind [see Eq. 3.2(37) of Ref. 10].

APPENDIX C: PROOF OF ANTICOMMUTATIVITY OF THE LOBACHEVSKIAN DIRAC FIELDS

In this appendix we show that the Lobachevskian Dirac fields anticommute as they should do.

It is not difficult to see that

$$\{\Psi^{(1)}_\mu(\xi),\Psi^{(2)}_\nu(\xi')\} = \frac{\lambda^3}{(2\pi)^4}\int_{R^+}\int_{S^3}\frac{1}{p^2+(m_f+\lambda/2)^2}\{[S(\varkappa(\zeta_\xi))(ip\gamma_4+m_f+\lambda/2)S^{-1}(\varkappa(\zeta_{\xi'}))]_{\mu\nu}$$
$$\times[\zeta,\xi]^{-3/2-ip/\lambda}[\zeta,\xi']^{-3/2+ip/\lambda}+[S(\varkappa(\zeta_\xi))(ip\gamma_4-m_f-\lambda/2)S^{-1}(\varkappa(\zeta_{\xi'}))]_{\mu\nu}$$
$$\times[\zeta,\xi]^{-3/2+ip/\lambda}[\zeta,\xi']^{-3/2-ip/\lambda}\}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2 dp\,d\mathbf{n}. \tag{C1}$$

By arguments very similar to those used in arriving at (4.12) we can show that anticommutator (C1) may be written as

$$\{\Psi^{(1)}_\mu(\xi),\Psi^{(2)}_\nu(\xi')\} = \frac{2\lambda^3}{(2\pi)^3}\frac{1}{\sinh\lambda r}\int_0^\infty\frac{1}{p^2+(m_f+\lambda/2)^2}\left|\frac{\Gamma(2+ip/\lambda)}{\Gamma(1/2+ip/\lambda)}\right|^2\left\{ip\left[S(\varkappa(\eta))\gamma_4 S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right]_{\mu\nu}\right.$$
$$\times\frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)}P^{-1/2+ip/\lambda}_{3/2,-1/2}(\cosh\lambda r)+(m_f+\lambda/2)\left[S(\varkappa(\eta))S^{-1}\left(\varkappa\left(\frac{a^{-1}\eta}{(a^{-1}\eta)_0}\right)\right)\right]_{\mu\nu}$$
$$\left.\times\frac{\Gamma(-1+ip/\lambda)}{\Gamma(ip/\lambda)}P^{-1/2+ip/\lambda}_{3/2,1/2}(\cosh\lambda r)-(p\leftrightarrow -p)\right\}dp, \tag{C2}$$

with notation as in Sec. IV. This expression vanishes because of symmetry relation (A10). Thus

$$\{\Psi^{(i)}_\mu(\xi),\Psi^{(j)}_\nu(\xi')\}=0\quad(i,j=1,2),\quad\text{for all }\xi,\xi'\in\Lambda^4.$$

(We note that this condition is trivially fulfilled for $i=j$.)

[1]G. A. Kerimov, J. Math. Phys. **27**, 1358 (1986).
[2]K. Osterwalder and R. Schrader, Helv. Phys. Acta **46**, 277 (1973).
[3]J. Frohlich and K. Osterwalder, Helv. Phys. Acta **47**, 781 (1974).
[4]G. W. Mackey, Ann. Math. **55**, 101 (1952); **58**, 193 (1953).
[5]E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[6]R. Takahasi, Bull. Soc. Math. Fr. **91**, 289 (1963).
[7]V. K. Dobrev, G. Mack, V. B. Petkova, S. G. Petrova, and I. T. Todorov, in *Lecture Notes in Physics*, Vol. 63 (Springer, New York, 1977).
[8]I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskii-Shapiro, *Representation Theory and Authormorphic Functions* (Nauka, Moscow, 1966) (in Russian).
[9]N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968).
[10]*Higher Transcendental Functions*, Vol. 1 of Bateman Manuscript Project, edited by A. Erdelyi (McGraw-Hill, New York, 1953).
[11]M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964).
[12]*Higher Transcendental Functions*, Vol. 2 of Bateman Manuscript Project, edited by A. Erdelyi (McGraw-Hill, New York, 1953).

On the violation of causality in relativistic quantum mechanics

Piotr Kosiński
Institute of Physics, University of Lodz, Nowotki 149, 90-236 Lodz, Poland
Paweł Maślanka
Institute of Mathematics, Banacha 22, 90-238 Lodz, Poland

(Received 28 June 1989; accepted for publication 7 March 1990)

The estimates on causality violating part of the wave function are given. The problem of acausal behavior of charge and current densities is discussed.

I. INTRODUCTION

It has been observed that there are some difficulties concerning the notion of causality in relativistic quantum mechanics. Hegerfeld[1] proved a theorem that a particle at $t = 0$ is localized with probability 1 in a finite volume of space immediately develops infinite "tails" irrespective of the particular notion of localization (cf. also Ref. 2). Later[3] he extended his theorem to the case of states described by the wave functions with exponentially bounded tails at $t = 0$. Rujisenaars[4] proved that the amount of causality violation tends to zero as $t \to \infty$. Moreover, he has shown that it would be extremely difficult to verify experimentally the violation of causality. While we are not completely convinced that the arguments of this kind provide a real solution of the problem, we want to concentrate here only on the mathematical aspects of the problem. Namely, we shall give very simple proofs of two theorems that characterize more precisely the acausal part of the wave function. For simplicity we restrict our attention to the one-dimensional case; the discussion of the generalization to three dimensions is given in the last section.

To make the paper self-contained let us recall the essence of the Newton–Wigner approach[5] to the problem of particle localization. The particles of spin zero are described by the representation of the Poincarè group corresponding to the $J = 0$ representation of little group. The invariant scalar product reads

$$(\psi,\phi) = \int \frac{d^3\mathbf{p}}{p^0}\, \bar{\psi}(\mathbf{p})\phi(\mathbf{p}), \tag{1.1}$$

with supports of ψ and ϕ concentrated on the hyperboloid $p^2 = m^2$. By invoking some physically plausible assumptions, Wigner and Newton showed that the state vector describing the particle localized at the point $\mathbf{x}$ reads

$$\psi_{\mathbf{x}}(\mathbf{p}) = (2\pi)^{-3/2}\sqrt{p^0}\, e^{i\mathbf{p}\cdot\mathbf{x}}. \tag{1.2}$$

One can introduce the wave function corresponding to a given state vector ϕ by the standard formula

$$\tilde{\phi}(\mathbf{x}) \equiv \langle \psi_{\mathbf{x}} \mid \phi\rangle = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3\mathbf{p}}{p^0} \sqrt{p^0}\, \phi(\mathbf{p}) e^{i\mathbf{p}\cdot\mathbf{x}}. \tag{1.3}$$

The scalar product can be now written as

$$(\phi_1,\phi_2) = \int d^3\mathbf{x}\, \bar{\tilde{\phi}}_1(\mathbf{x})\tilde{\phi}_2(\mathbf{x}). \tag{1.4}$$

Now, the probability that in the state ϕ the particle is localized within the volume V (or, more precisely, that it is in the localized state described by $\mathbf{x} \in V$) reads

$$P(V) = \int_V d^3\mathbf{x} \mid \tilde{\phi}(\mathbf{x})|^2. \tag{1.5}$$

Let us note that the notion of localized state is not covariant.

Noting that P_0 is the generator of time translations one can introduce the time-dependent wave function

$$\tilde{\phi}_t(\mathbf{x}) = \langle \psi_{\mathbf{x}} \mid \phi_t\rangle = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3\mathbf{p}}{p^0} \sqrt{p^0}\phi(\mathbf{p}) e^{i\mathbf{p}\cdot\mathbf{x} - ip^0x^0}, \tag{1.6}$$

which is the solution to the Klein–Gordon equation.

The wave function of the localized state is

$$\tilde{\phi}_{\mathbf{x}}(\mathbf{y}) = \langle \psi_{\mathbf{y}} | \psi_{\mathbf{x}}\rangle = \delta^{(3)}(\mathbf{x} - \mathbf{y}), \tag{1.7}$$

as it should be.

Finally, let us note that one can consider the Fourier transform of the localized state

$$\int \frac{d^3\mathbf{p}}{p^0} \sqrt{p^0} e^{i\mathbf{p}(\mathbf{x}-\mathbf{y})} = \left(\frac{m}{r}\right)^{5/4} H^{(1)}_{5/4}(imr), \quad r \equiv |\mathbf{x} - \mathbf{y}|. \tag{1.8}$$

Its shape reflects the standard wisdom that the relativistic particle can be localized only within a distance $\sim 1/m$. However, the above "wave function" is not to be interpreted as probability density.

Now, let us fix some $a > 0$ and assume that at $t = 0$ the particle of mass m is localized within the interval $\langle -a,a\rangle$. Therefore its state is described by a wave function $f \in L^2(R)$ such that suppf $\subset \langle -a,a\rangle$. We shall trace the time development of this state:

$$f_t(x) = (U(t)f)(x). \tag{1.9}$$

Here $U(t) \equiv \exp(-i\sqrt{-\Delta + m^2}\cdot t)$ is the pseudodifferential operator of relativistic time evolution. The causal region at any $t > 0$ is defined to be the interval $\langle -a - t, a + t\rangle$. The natural measure of the amount of causality violation is given by the formula

$$\mathrm{I}(t) \equiv \int_{|x| > a+t} dx\, |f_t(x)|^2. \tag{1.10}$$

We introduce the following notation:

$$F_t^{(+)}(\lambda) \equiv f_t(x), \quad \lambda \equiv x - a - t > 0,$$
$$F_t^{(-)}(\lambda) \equiv f_t(x), \quad \lambda \equiv -x - a - t > 0. \tag{1.11}$$

Now we can formulate our theorems.

Theorem 1: (i) The functions $F_t^{(\pm)}(\lambda)$ are infinitely differentiable for $\lambda > 0$ (modulo the set of vanishing measure).

 0022-2488/90/071755-06$03.00

(ii) For any $\lambda_0 > 0$ there exists a constant $c(\lambda_0, f, t)$ such that for all $\lambda \geqslant \lambda_0$,

$$|F_t^{(\pm)}(\lambda)| \leqslant C(\lambda_0, f, t) e^{-m\lambda}, \tag{1.12}$$

and for any natural k,

$$\lim_{t\to\infty} t^k C(\lambda_0, f, t) = 0. \tag{1.13}$$

(iii) If $f \in C_0^N(R)$, $N \geqslant 3$, then there exists a constant $C(f,t)$ such that for $\lambda > 0$,

$$|F_t^{(\pm)}(\lambda)| \leqslant C(f,t) e^{-m\lambda}, \tag{1.14}$$

and

$$\lim_{t\to\infty} t^{N-2} C(f,t) = 0. \tag{1.15}$$

(iv) If $f \in L^q(R)$, $q \geqslant 2$, then for $\lambda > 0$,

$$|F_t^{(\pm)}(\lambda)| \leqslant C(f,q) \lambda^{-1/q}, \quad 2 \leqslant q < \infty,$$
$$|F_t^{(\pm)}(\lambda)| \leqslant C(f) |\ln \lambda|, \quad q = \infty. \tag{1.16}$$

The estimates given by Eq. (1.16) are the best possible for $q = \infty$ and the best possible among the power estimates for $2 \leqslant q < \infty$.

Theorem 2: (i) If $f \in C_0^N(R)$, $N \geqslant 1$, then

$$\lim_{t\to\infty} t^{2N-2} I(t) = 0. \tag{1.17}$$

(ii) If $f \in L^q(R)$, $2 < q \leqslant \infty$, then for any $\varepsilon > 0$,

$$\lim_{t\to\infty} t^{1-2/q-\varepsilon} I(t) = 0, \tag{1.18}$$

and this estimate is the best possible among the power estimates.

(iii) If $f \in L^2(R)$, then

$$\lim_{t\to\infty} I(t) = 0, \tag{1.19}$$

but no f-independent estimate exists.

(iv) $I(t) < \frac{1}{4}$ and for any $\varepsilon > 0$ there exist such $m > 0$, $a > 0$, $t > 0$, and $f \in L^2(R)$, supp$f \subset \langle -a, a\rangle$ that $I(t) > \frac{1}{4} - \varepsilon$.

II. PROOFS OF THE THEOREMS

If $f \in L^2(R)$, supp$f \subset \langle -a, a\rangle$, then $f \in L^1(R)$. Therefore we can write[6]

$$f_t(x) = \text{l.i.m.} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dp \tilde{f}(p) e^{ipx - i\omega(p)t}, \tag{2.1}$$

where

$$\tilde{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx f(x) e^{-ipx}, \tag{2.2}$$

and $\omega(p) = \sqrt{p^2 + m^2}$.

Moreover, $\tilde{f}(p)$ is entire function and the following estimate holds:

$$|\tilde{f}(p)| < Ce^{a|Imp|}. \tag{2.3a}$$

For $f \in C_0^N(R)$ the estimate may be improved to yield

$$|\tilde{f}(p)| < C_n e^{a|Imp|}/(1 + |p|)^N. \tag{2.3b}$$

To prove both theorems we look for the more convenient representation of the wave function f_t in the region $|x| > a + t$. To this end let us note that $\omega(p)$ is analytic on the complex plane with two cuts along the imaginary axis starting from $\pm im$. Therefore, using the inequality (2.3b) we can for any $f \in C_0^\infty(R)$ deform the contour in Eq. (2.1) for $|x| > a + t$ as it is shown on Fig. 1. In this way we arrive at the following expressions for the wave function f_t valid outside the causality region:

$$f_t(x) = \begin{cases} \dfrac{2i}{\sqrt{2\pi}} \displaystyle\int_m^{\infty} dp \tilde{f}(ip) e^{-px} \operatorname{sh}(\sqrt{p^2 - m^2}\, t), & x > a + t, \quad (2.4a) \\ \dfrac{2i}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{-m} dp \tilde{f}(ip) e^{-px} \operatorname{sh}(\sqrt{p^2 - m^2}\, t), & x < -a - t. \quad (2.4b) \end{cases}$$

The above representation is valid for any $f \in L^2(R)$, supp$f \subset \langle -a, a\rangle$. This is easily seen by taking the sequence $f_n \xrightarrow{L^2} f$, $f_n \in C_0^\infty(R)$, using the inequality (2.3a) and unitarity of $U(t)$.

The corresponding expressions for $F_t^{(\pm)}(\lambda)$ are

$$F_t^{(+)}(\lambda) = \frac{2i}{\sqrt{2\pi}} \int_m^{\infty} dp [\tilde{f}(ip) e^{-pa}] \times [\operatorname{sh}(\sqrt{p^2 - m^2}\, t) e^{-pt}] e^{-p\lambda}, \tag{2.5a}$$

$$F_t^{(-)}(\lambda) = \frac{2i}{\sqrt{2\pi}} \int_{-\infty}^{-m} dp [\tilde{f}(ip) e^{pa}] \times [\operatorname{sh}(\sqrt{p^2 - m^2}\, t) e^{pt}] e^{p\lambda}. \tag{2.5b}$$

Now we can prove our first theorem. (i) The statement follows at once from Eqs. (2.5) and the Lebesque theorem because the functions in square brackets are bounded.

(ii) From Eq. (2.5a) we get for $\lambda \geqslant \lambda_0 > 0$

$$|F_t^{(+)}(\lambda)| \leqslant \frac{2}{\sqrt{2\pi}} e^{-m(\lambda - \lambda_0)} \int_m^{\infty} dp [|\tilde{f}(ip)| e^{-ap}] \times [\operatorname{sh}(\sqrt{p^2 - m^2}\, t) e^{-pt}] e^{-p\lambda_0},$$

and (1.12) holds with

$$C(\lambda_0, f, t) \equiv \frac{2}{\sqrt{2\pi}} e^{m\lambda_0} \int_m^{\infty} dp [|\tilde{f}(ip)| e^{-ap}] \times [\operatorname{sh}(\sqrt{p^2 - m^2}\, t) e^{-pt}] e^{-p\lambda_0}.$$

Note that for any $\alpha > 0$

$$\lim_{t\to\infty} [t^\alpha \operatorname{sh}(\sqrt{p^2 - m^2 t}) e^{-pt}] = 0,$$

and

$$\sup_{t>0}\,[t^{\alpha}\,\mathrm{sh}(\sqrt{p^2-m^2}t)e^{-pt}]\leqslant\text{const.}\times p^{\alpha}. \tag{2.6}$$

Therefore (1.13) follows from the Lebesque convergence theorem. The above arguments apply also to $F_t^{(-)}(\lambda)$.

(iii) Using the inequality (2.3b) we obtain (1.14) with [for $F_t^{(+)}(\lambda)$],

$$C(f,t)=\frac{2C_N}{\sqrt{2\pi}}\int_m^{\infty}dp\,\frac{\mathrm{sh}(\sqrt{p^2-m^2}t)e^{-pt}}{(1+|p|)^N},$$

and one has only to apply again (2.6) and the Lebesque theorem.

(iv) If $f\in L^q(R)$, $q\geqslant 1$, $\operatorname{supp} f\subset\langle -a,a\rangle$, then by Hölder inequality,

$$|\tilde f(p)|=\frac{1}{\sqrt{2\pi}}\left|\int_{-a}^{a}dx\,f(x)e^{-ipx}\right|$$

$$\leqslant C(q)\cdot\|f\|_q\cdot\frac{e^{a|Imp|}}{|Imp|^{1-1/q}}, \tag{2.7}$$

where $C(q)$ is some constant depending on q only. Using (2.7) and (2.5a) we get

$$|F_t^{(+)}(\lambda)|\leqslant\frac{C(q)\cdot\|f\|_q}{\sqrt{2\pi}}\int_m^{\infty}dp\,\frac{e^{-m^2t/2p-p\lambda}}{p^{1-1/q}}.$$

Estimating the integral on the right-hand side we get (1.16).

To show that the estimates are the best possible we consider the function

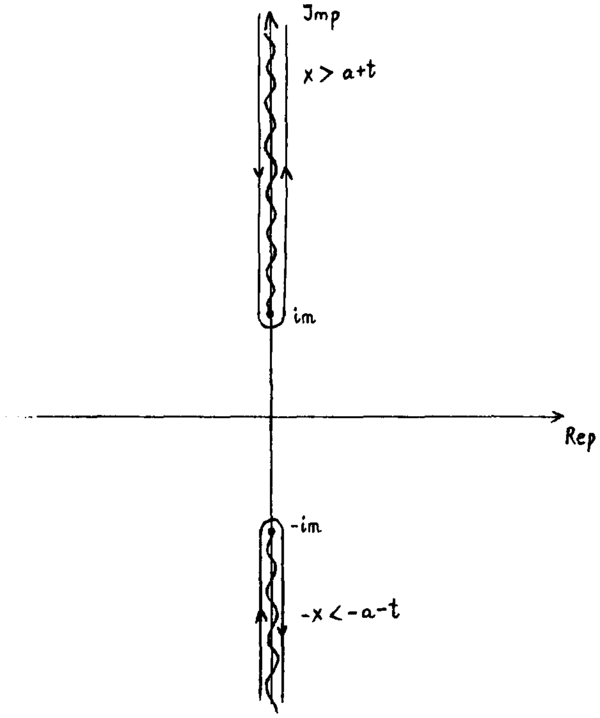


FIG. 1. The deformed contour.

$$f_q(x)=\begin{cases}((q-2)/q)^{1/2}(2a)^{(2-q)/2q}(a-x)^{-1/q}, & \text{for } |x|<a,\\ 0 & \text{for } |x|>a.\end{cases} \tag{2.8}$$

Then for $q>2, f_q\in L^{q-\varepsilon}(R)$ for any $\varepsilon>0$ and $f_q\in L^2(R)$, $\|f_q\|_2=1$. It is easy to check that for the corresponding functions $F_t^{(+)}(\lambda)$ we have

$$|F_t^{(\pm)}(\lambda)|\geqslant\text{const}\begin{cases}\lambda^{-1/q}, & 2<q<\infty,\\ |\ln\lambda|, & q=\infty.\end{cases}$$

This concludes the proof of the first theorem.

To prove the second one let us note that we can write

$$I(t)=I_+(t)+I_-(t),$$

where

$$I_+(t)\equiv\int_0^{\infty}d\lambda\,|F_t^{(+)}(\lambda)|^2$$

$$=\int_0^{\infty}d\lambda\int_m^{\infty}dp\int_m^{\infty}dp'\,e^{-(p+p')\cdot\lambda}\cdot\tilde f(ip)$$

$$\times\overline{\tilde f(ip')}\,\mathrm{sh}(\sqrt{p^2-m^2}\cdot t)\mathrm{sh}(\sqrt{p'^2-m^2}\cdot t)$$

$$\times e^{-(p+p')(a+t)}, \tag{2.9}$$

and a similar formula is valid for $I_-(t)$.

Let us assume that $f\in L^q(R)$, $q>2$. Using Eq. (2.7) one easily checks that the order of integration may be changed. After integrating over λ we arrive at the following expression:

$$I_+(t)=\int_m^{\infty}\int_m^{\infty}\frac{dp\,dp'}{(p+p')}\tilde f(ip)\;\overline{\tilde f(ip')}\,\mathrm{sh}(\sqrt{p^2-m^2}\cdot t)$$

$$\times\mathrm{sh}(\sqrt{p'^2-m^2}\cdot t)e^{-(p+p')(a+t)}. \tag{2.10}$$

But $p+p'\geqslant 2\sqrt{pp'}$ so we have to estimate the integral

$$\int_m^{\infty}\frac{dp}{\sqrt{p}}\,|\tilde f(ip)|e^{-(a+t)\cdot p}\cdot\mathrm{sh}(\sqrt{p^2-m^2}\cdot t)$$

$$\leqslant\text{const}\int_m^{\infty}dp\,\frac{e^{-m^2t/2p}}{p^{3/2-1/q}}.$$

Using again the Lebesque convergence theorem and (2.6) we get for any $\varepsilon>0$,

$$\lim_{t\to\infty}t^{1/2-1/q-\varepsilon}\int_m^{\infty}\frac{dp}{\sqrt{p}}\,|\tilde f(ip)|\mathrm{sh}(\sqrt{p^2-m^2}\cdot t)$$

$$\times e^{-(a+t)p}=0.$$

This gives Eq. (1.18) for $I_+(t)$, the same reasoning holds for $I_-(t)$. To obtain (1.17) we repeat the above arguments taking into account the inequality (2.3b). To prove that the above estimate is the best possible we use the function given by Eq. (2.8) for which

$$I(t)\geqslant\text{const}\;t^{-(1-2/q)}.$$

(iii) Equation (1.19) follows from (1.18) and the unitarity of $U(t)$. We prove that there exists no estimate of the form

$$I(t) \leqslant C(f)\rho(t),$$

where ρ is the function independent of the choice of f and $\lim_{t\to\infty} \rho(t) = 0$. To this end we choose any such function $\rho_{(t)}$. Then there exists a sequence $\{t_n\}$, $t_n \underset{n\to\infty}{\to} \infty$ and the function $f \in L^2(R)$, $\operatorname{supp} f \subset \langle -a,a\rangle$ such that

$$\lim_{n\to\infty} I(t_n)/\rho(t_n) = \infty. \tag{2.11}$$

The proof goes as follows. We put $q_n \equiv 2n/(n-1)$ and denote by f_n the corresponding function given by Eq. (2.8). If $I_n(t)$ is the quantity $I(t)$ calculated with f_n as the initial wave function then it is straightforward to check that

$$I_n(t) \geqslant C(a,m)t^{-1/n}, \tag{2.12}$$

where $C(a,m)$ is a constant not depending on n. For any sequence $\{b_n\}$ such that $b_n \geqslant 0$, $\Sigma_n b_n < \infty$ the function $f_b \equiv \Sigma_n b_n f_n$ belongs to $L^2(R)$ and $\operatorname{supp} f \subset \langle -a,a\rangle$. Moreover

$$\|f_b\|_2 \leqslant \sum_n b_n,$$

and because of the positivity of all f_n;

$$I_b(t) \geqslant \frac{C(a,m)}{(\Sigma_n b_n)^2} \sum_n b_n^2 t^{-1/n}. \tag{2.13}$$

Our conclusion follows now from Eq. (2.13) and the following lemma.

Lemma: For any positive function ρ such that $\rho(t) \underset{t\to\infty}{\to} 0$ there exist the sequences $\{b_n\}$ and $\{t_k\}$ such that

(i) $b_n \geqslant 0, \sum_n b_n < \infty$,

(ii) $t_k \underset{k\to\infty}{\to} \infty$,

(iii) $\displaystyle \lim_{k\to\infty} \frac{\Sigma_n b_n^2 t_k^{-1/n}}{\rho(t_k)} = \infty.$

Proof of the lemma: Let $\{t_k\}$, $t_k \underset{k\to\infty}{\to} \infty$ be a sequence such that $\rho(t_k) \leqslant 1/k^6$; for any natural k we can choose a natural n_k such that $t_k^{1/2n_k} \leqslant 2$. Therefore, if we define

$$b_n = \begin{cases} 0, & \text{for } n \neq n_k, \\ (1/k^2)t_k^{1/2n_k}, & \text{for } n = n_k, \end{cases}$$

then (i) $b_n \geqslant 0$, (ii) $\Sigma_n b_n = \Sigma_k (1/k^2) t_k^{1/2n_k} \leqslant 2\,\Sigma_k (1/k^2) < \infty$, and

$$\frac{\Sigma_n b_n^2 t_k^{1/n_k}}{\rho(t_k)} \geqslant \frac{b_{n_k}^2 t_k^{-1/n_k}}{1/k^b} = k^2.$$

(iv) To prove that $I(t) < \frac{1}{4}$ we assume first that $f \in C_0^\infty(R)$ and $f(x) \geqslant 0$. It follows then from (2.2) that also $\tilde f(ip) \geqslant 0$. From Eqs. (2.5) we get

$$\begin{aligned} |F_t^{(+)}(\lambda)| &\leqslant \frac{1}{\sqrt{2\pi}} \int_0^\infty dp\, \tilde f(ip) e^{-p(a+\lambda)}, \\ |F_t^{(-)}(\lambda)| &\leqslant \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 dp\, \tilde f(ip) e^{p(a+\lambda)}. \end{aligned} \tag{2.14}$$

We can now rotate the integration contours back to real axis to obtain

$$\begin{aligned} |F_t^{(+)}(\lambda)| &\leqslant \frac{-i}{\sqrt{2\pi}} \int_0^\infty dp\, \tilde f(p) e^{ip(\lambda+a)} \\ &\equiv -if_+(\lambda+a), \\ |F_t^{(-)}(\lambda)| &\leqslant \frac{-i}{\sqrt{2\pi}} \int_{-\infty}^0 dp\, \tilde f(p) e^{-ip(\lambda+a)} \\ &\equiv -if_-(\lambda+a). \end{aligned} \tag{2.15}$$

Obviously

$$f(x) = f_+(x) + f_-(-x). \tag{2.16}$$

From the Plancherel theorem we have

$$\begin{aligned} &\int_{-\infty}^\infty dx |f_+(x)|^2 + \int_{-\infty}^\infty dx |f_-(-x)|^2 \\ &= \int_0^\infty dp |\tilde f(p)|^2 + \int_{-\infty}^0 dp |\tilde f(p)|^2 = 1. \end{aligned} \tag{2.17}$$

But from Eq. (2.16)

$$\begin{aligned} &\int_{-\infty}^\infty dx |f_+(x)|^2 + \int_{-\infty}^\infty dx |f_-(-x)|^2 \\ &+ \int_{-\infty}^\infty dx [\bar f_+(x) f_-(-x) + \bar f_-(-x) f_+(x)] = 1, \end{aligned}$$

and therefore

$$\int_{-\infty}^\infty dx [\bar f_+(x) f_-(-x) + \bar f_-(-x) f_+(x)] = 0.$$

Consequently,

$$\begin{aligned} &\operatorname{Re} \int_{-\infty}^{-a} dx\, \bar f_+(x) f_-(-x) + \operatorname{Re} \int_a^\infty dx\, \bar f_+(x) f_-(-x) \\ &+ \operatorname{Re} \int_{-a}^a dx\, \bar f_+(x) f_-(-x) = 0. \end{aligned} \tag{2.18}$$

Using again (2.16) we conclude that $f_+(x) = -f_-(-x)$ for $|x| > a$; Eq. (2.18) then takes the form

$$\begin{aligned} &\int_a^\infty dx |f_+(x)|^2 + \int_{-\infty}^{-a} dx |f_-(-x)|^2 \\ &= \operatorname{Re} \int_{-a}^a dx\, \bar f_+(x) f_-(-x). \end{aligned} \tag{2.19}$$

Our result follows now from Eqs. (2.15), (2.19), and the relations

$$\begin{aligned} 1 &= \int_{-a}^a dx |f(x)|^2 \\ &= \int_{-a}^a dx |f_+(x)|^2 + \int_{-a}^a dx |f_-(-x)|^2 \\ &\quad + 2\operatorname{Re} \int_{-a}^a dx\, \bar f_+(x) f_-(-x), \end{aligned} \tag{2.20}$$

$$0 \leqslant \int_{-a}^{a} dx |f_+(x) - f_-(-x)|^2$$
$$= \int_{-a}^{a} dx |f_+(x)|^2 + \int_{-a}^{a} dx |f_-(-x)|^2$$
$$- 2\,\mathrm{Re} \int_{-a}^{a} dx \bar{f}_+(x) f_-(-x). \qquad (2.21)$$

Making use of the fact that $U(t)$ is unitary and $C_0^\infty(R)$ is dense in $L^2(R)$ we extend our result to any positive $f \in L^2(R)$. If f is an arbitrary real element of $L^2(R)$ then

$$|F_t^{(+)}(\lambda)| \leqslant \frac{1}{\sqrt{2\pi}} \int_m^\infty dp\, \tilde{g}(ip) e^{-(a+t+\lambda)p}$$
$$\times \mathrm{sh}(\sqrt{p^2 - m^2} \cdot t),$$

with $g(x) = |f(x)|$ and the same inequality holds for $F_t^{(-)}(\lambda)$. Finally, for any complex f the result follows from the fact that real (imaginary) part of $\tilde{f}(ip)$ corresponds to real (imaginary) part of $f(x)$.

It remains only to prove that this estimate is the best possible. To this end we consider again the functions f_q given by Eq. (2.8). It is straightforward to check that if $am \to 0$, $mt \to \infty$, $(am)(mt) \to 0$ (with a fixed, for simplicity) then $|F_t^{(+)}(\lambda)| \nearrow |f_{q+}(a+\lambda)|$ pointwise; here f_{q+} is the function defined in Eq. (2.15) calculated for $f = f_q$. Therefore, from the Lebesque theorem it follows that

$$I_+(t) \to \int_0^\infty d\lambda\, |f_{q_+}(\lambda + a)|^2. \qquad (2.22)$$

But it is easy to check by explicit calculations that

$$\lim_{q \to 2^+} \int_0^\infty d\lambda\, |f_{q^+}(\lambda + a)|^2 = \frac{1}{4}, \qquad (2.23)$$

(see the Appendix). Our result follows now from Eqs. (2.22) and (2.23).

III. CONCLUSIONS

We discussed above in some detail the properties of the acausal tail for the propagation of positive-frequency part of the solution to the Klein–Gordon equation. We restricted our considerations to the case of one space dimension but the generalization to three (and more) dimensions is quite straightforward. For fixed point **x** we rotate the axes in momentum space to make the (say) first axis parallel to **x**. The considerations of Sec. II can then be repeated with some minor changes. Therefore we could prove similar theorems for R^3, for example if $f \in C_0^\infty(R)$ then $I(t)$ tends to zero faster than t^{-N} for any natural N (this result is implicit in the paper of Ruijsenaars[4]). What is more important, the estimate $I(t) < \frac{1}{4}$ holds true too. One can ask whether the result depends on the choice of the notion of localization. The definition of coordinate operator given by Wigner and Newton seems to be the most reasonable one. However, the results of previous sections rely in fact only on the assumption that the momentum is the generator of space translations and the energy is positive. On the other hand one can skip the notion of position operator and consider the behavior of charge and current densities:

$$\rho(x,t) = \frac{ie}{2m} \left(\bar{\psi} \frac{\partial \psi}{\partial t} - \psi \frac{\partial \bar{\psi}}{\partial t} \right),$$
$$j(x,t) = \frac{e}{2mi} \left(\bar{\psi} \frac{\partial \psi}{\partial x} - \psi \frac{\partial \bar{\psi}}{\partial x} \right),$$

where $\psi(x,t)$ is the solution to the Klein–Gordon equation containing positive frequency part only.

The formulas (2.4) are also valid in this case. Let us assume that $\rho(x,t)$ and $j(x,t)$ the causality principle is not violated, i.e., $\rho(x,t) = 0 = j(x,t)$ for $|x| > a + t$. Then it follows from the above equations that $\overline{\psi(x,t)} = C\psi(x,t)$ for $|x| > a + t$, $|C| = 1$. Therefore after a suitable redefinition we may assume that $\psi(x,t)$ is purely imaginary for $|x| > a + t$. Using Eq. (2.4) we conclude that $\overline{\tilde{\psi}(ip)} = \tilde{\psi}(ip)$ for $|p| > m$. Applying the Schwarz reflection principle to the function $g(z) \equiv \tilde{\psi}(iz)$ and using the fact that $\tilde{\psi}(z)$ is an entire function we obtain

$$\overline{\tilde{\psi}(p)} = \tilde{\psi}(-p), \quad p \in \mathbb{R}. \qquad (3.1)$$

Therefore $\psi(x,0)$ is a real function. Summarizing, the necessary and sufficient condition for the wave function to describe the causal behavior of charge and current density is the following constraint for the initial wave function

$$\overline{\psi(x,0)} = C\psi(x,0), \quad |C| = 1.$$

It is easy to understand this result. Using Eq. (3.1) we can write

$$\mathrm{Re}\, \psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dp\, \tilde{\psi}(p) \cos[\omega(p)t] e^{ipx}.$$

The function $\cos[\omega(p) \cdot t]$ is an entire function and $\cos[\omega(p)t] \leqslant \mathrm{const} \times \exp\{t \cdot |\mathrm{Im} p|\}$. It follows then that $\mathrm{Re}\, \psi(x,t)$ behaves causally. Therefore

$$\rho \sim \mathrm{Im}\, \psi\, \mathrm{Re}\, \dot{\psi} - \mathrm{Re}\, \psi\, \mathrm{Im}\, \dot{\psi},$$

also behaves causally.

It should be stressed that there is nothing mysterious in the noncausality discussed here (from the mathematical point of view). Usually we pick up the particular solution to the Klein–Gordon equation by imposing the initial value conditions on $f(x,0)$ and $\partial f(x,t)/\partial t\,|_{t=0}$; here the second condition is replaced by demanding that only positive frequency part contribute. This is highly nonlocal condition and as a result $\partial f(x,t)/\partial t\,|_{t=0}$ does not have a compact support [its Fourier transform is $-i\omega(p)\tilde{f}(p)$, which is not an entire function].

ACKNOWLEDGMENTS

We thank Professor B. Nowak and Dr. S. Giler for very helpful discussions and careful reading of the manuscript.

This work is supported by CPBP 01.03.

APPENDIX: THE VERIFICATION OF EQ. (2.23)

We verify here Eq. (2.23). First, using Eqs. (2.2) and (2.8) we get

$$e^{-pa}\tilde{f}(ip)=\left(\frac{q-2}{2\pi q}\right)^{1/2}(2a)^{(2-q)/2q}\times\int_{-a}^{a}dx(a-x)^{-1/q}e^{p(x-a)}$$

$$=e^{i\pi(1/q-1)}\left(\frac{q-2}{2\pi q}\right)^{1/2}(2a)^{(2-q)/2q}p^{1/q-1}\times\gamma(1-1/q,2ap).$$

The corresponding function $f_{q+}(a+\lambda)$ reads

$$f_{q^+}(a+\lambda)=\frac{i}{\sqrt{2\pi}}\int_0^\infty dp\tilde{f}(ip)e^{-p(a+\lambda)}=\frac{i}{2\pi}e^{i\pi(1/q-1)}\left(\frac{q-2}{q}\right)^{1/2}(2a)^{(2-q)/2q}\int_0^\infty dp\,p^{1/q-1}\gamma\left(1-\frac{1}{q},2ap\right)e^{-p\lambda}$$

$$=\frac{i}{2\pi}e^{i\pi(1/q-1)}\left(\frac{q-2}{q}\right)^{1/2}(2a)^{1/2}\frac{q}{(q-1)(2a+\lambda)}\cdot{}_2F_1\left(1,1;2-\frac{1}{q};\frac{2a}{2a+\lambda}\right).$$

Therefore

$$\int_0^\infty d\lambda\,|f_{q^+}(a+\lambda)|^2=\frac{1}{4\pi^2}\left(\frac{q-2}{q}\right)\frac{q^2}{(q-1)^2}\times\int_0^1 du\left|{}_2F_1\left(1,1;2-\frac{1}{q};u\right)\right|^2. \quad \text{(A1)}$$

We need only the divergent part of the last integral for $q\to 2$. It comes from the divergence of the integrand for $u\approx 1$. But

$${}_2F_1\left(1,1;2-\frac{1}{q};u\right)\underset{u\to 1^-}{\approx}\frac{\Gamma(2-1/q)}{\Gamma(1-1/q)}B\left(1-\frac{1}{q},\frac{1}{q}\right)(1-u)^{-1/q},$$

and inserting this expression into (A1) we get Eq. (2.23).

[1]G. C. Hegerfeldt, Phys. Rev. D **10**, 3320 (1974).
[2]C. N. Fleming, Phys. Rev. **139**, B963 (1965).
[3]G. C. Hegerfeldt, Phys. Rev. Lett. **54**, 2395 (1985).
[4]S. N. M. Ruijsenaars, Ann. Phys. **137**, 33 (1981).
[5]T. Newton and E. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.

On the covariant quantization of a first-class superparticle

U. Lindström

Institute of Theoretical Physics, University of Stockholm, Vanadisvägen 9, S-11346 Stockholm, Sweden

M. Roček, W. Siegel, and P. van Nieuwenhuizen

Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794-3840

A. E. van de Ven

Department of Physics, New York University, New York, New York 10003

(Received 2 November 1989; accepted for publication 3 January 1990)

The Batalin–Vilkovisky formalism is used to quantize covariantly a formulation of the superparticle with only first-class constraints. Towers of ghosts are found that split into ghosts-for-ghosts and ghost gauge fields. Subsolutions to the master equation with spinor and bispinor ghosts contain terms cubic in antifields and lead to free gauge-fixed BRST-invariant actions. The final formulation with multispinor ghosts is nilpotent only on the classical shell.

I. INTRODUCTION

Recently we proposed a supercovariant quantization of the superparticle[1] that led to a free gauge-fixed action.[2] The methods we used were subsequently applied (with varying degrees of completeness) to the heterotic string[3–7] and type II strings,[8,9] as well as to further aspects of the superparticle.[10,11] All these papers applied the Batalin–Vilkovisky (BV) quantization procedure[12] to a system with an infinite tower of ghosts-for-ghosts,[13] with mixed first- and second-class constraints,[14] used a novel gauge condition that was consistent only on shell,[2] and resulted in a gauge-fixed first-quantized action that was free (up to interactions with a background[10]). However, all of these papers suffer from a technical defect[15,16] explained below, which has a rather serious consequence: The final gauge-fixed action is *not* BRST invariant[15]! (In Refs. 3 and 4 we found that the methods of Ref. 2 gave the correct conformal anomaly for the superstring. It is tempting to conjecture that new methods can be found that will lead to the *same* action with some appropriately modified BRST laws.)

In this paper we attempt to solve this problem by using the alternate formulation of the superparticle that is free of second-class constraints.[17,18] This formulation involves new symmetries, new gauge fields, and new infinite towers of ghosts-for-ghosts.[19] We find that the whole ghost-for-ghost sector splits into two parts: half the fields are ordinary ghost-for-ghosts while the other half act as gauge fields for the former. We call the latter "ghost gauge fields." We consider a nested set of intermediate solutions to the master equation $(S,S) = 0$, with ever-increasing numbers of ghosts. Each of these intermediate solutions can be gauge-fixed by fixing all (physical and ghost) gauge fields. For the first time in the BV formalism, we have found the need for an S_3 (S_k denotes the part of S with k antifields), satisfying $(S_1,S_2) + (S_0,S_3) = 0$, but no higher S_k are needed. However, all non-gauge-fixed actions S contain additional gauge symmetries, and the most straightforward approach would seem to be to apply the BV formalism to all these local symmetries, leading to further ghost-for-ghosts and ghost gauge fields. We believe that all these intermediate theories lead to the wrong BRST cohomology (this point is under study). This has led us to consider the final model of this paper, in which all local symmetries are treated with the BV formalism. We find towers of ghosts-for-ghosts c_i and $\eta_{i_1\cdots i_n}$ and ghost gauge fields ω_i and $\lambda_{i_1\cdots i_n}$ (where both i and $i_n \geqslant 0$ and $n \geqslant 2$). The action for this model satisfies $(S_1,S_1) = 0$ modulo the S_0 field equations, but we have not been able to integrate this result by constructing an S_2 satisfying $(S_1,S_1) + 2(S_0,S_2) = 0$. It may be that further fields and/or fewer ghosts, not found from the most straightforward application of the BV approach, are needed to obtain an S satisfying $(S,S) = 0$.

We begin by reviewing the results of Ref. 2 and explaining the defect.[15,16] The starting point in Ref. 2 was the supersymmetric action

$$S_0 = \dot{x}p - i\dot{\theta}\not{p}\theta - \tfrac{1}{2}gp^2, \tag{1.1}$$

which is invariant under diffeomorphisms (ξ) and κ symmetry:

$$\delta\theta = \not{p}\kappa, \quad \delta g = \dot{\xi} - 4i\dot{\theta}\kappa,$$
$$\delta x = \xi p - i\theta\gamma\,\delta\theta, \quad \delta p = 0. \tag{1.2}$$

(We use the conventions of Ref. 10, where all fields have their spinor indices either up or down and no charge-conjugation matrix is ever needed.)

The BV procedure[12] involves three steps: First, one finds the "minimal" solution $S_{\min}\,[\phi,\phi^*]$ to the master equation

$$\frac{1}{2}(S,S) = \frac{\partial_r S}{\partial\phi^A}\frac{\partial_l S}{\partial\phi_A{}^*} = 0, \tag{1.3}$$

where the ϕ^A are all physical fields, ghosts, and ghosts-for-ghosts, and the $\phi_A{}^*$ are the corresponding antifields. In Ref. 2 this minimal solution was

$$\begin{aligned} S_{\min} = {} & (\dot{x} - i\dot{\theta}\gamma\theta)p - \frac{1}{2}gp^2 + x^*(p\chi + i\theta\gamma\not{p}c_1) \\ & + g^*(\dot{\chi} + 4i\dot{\theta}c_1) + 2i\chi^*(c_1\not{p}c_1) \\ & - \sum_{i=0}^{\infty} c_i{}^*\not{p}c_{i+1} + 2g^*\Big[ix^*(c_1\gamma c_1 - \theta\gamma c_2) \\ & + \sum_{i=0}^{\infty} c_i{}^*c_{i+2} - 4i\chi^*c_1c_2\Big], \end{aligned} \tag{1.4}$$

0022-2488/90/071761-09$03.00

where $c_0{}^* = \theta^*$, and the κ ghost is $-c_1$ [we redefined $c_i = (-)^i c_i$ as compared to Ref. 2]. Next, one adds "nonminimal" terms that trivially satisfy the master equation and contain the Nakanishi–Lautrup auxiliary fields as well as the (antifields of the) antighosts. In Ref. 2 this nonminimal term was

$$S_{\text{nonmin}} = \hat{\chi}^* \pi_\xi + \sum_{k=1}^{\infty} \sum_{l=1}^{k} c_k{}^{l*} \pi_k{}^l. \tag{1.5}$$

Finally, one introduces the gauge-fixing fermion Ψ, and replaces all antifields $\phi_A{}^*$, nonminimal as well as minimal, by $\partial\Psi/\partial\phi^A$. In Ref. 2 the gauge-fixing fermion was

$$\begin{aligned}\Psi &= \hat{\chi}(g-1) - i \sum_{k=0}^{\infty} \sum_{l=0}^{k} c_{k+1}{}^{k-l+1} \dot{c}_k{}^l \\ &= \hat{\chi}(g-1) - i \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} c_{k+2l+1}{}^{k+l+1} \\ &\quad \times (\dot{c}_{k+2l}{}^l - \dot{c}_{k+2l+2}{}^{l+1}).\end{aligned} \tag{1.6}$$

At this point, there is already a clear problem which is essentially the problem discussed in Ref. 16: Ψ is not a good gauge-fixing fermion, as it has a gauge invariance

$$\delta c_k{}^0 = 0, \quad \delta c_k{}^l = \epsilon_{k-2l}, \quad \epsilon_m = 0, \quad \text{for } m<0, \tag{1.7}$$

which S_Ψ inherits after the antifields are eliminated. (This is the origin of the residual gauge invariance discovered in Ref. 16.) In Ref. 2 this problem was "avoided" by a singular field redefinition that effectively fixed this gauge invariance at the expense of BRST invariance of the final action.[15] One may attempt to modify Ψ (and/or S_{nonmin}) to avoid this gauge invariance. For example, one could consider

$$\Psi = \hat{\chi}(g-1) - i \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} c_{k+2l+1}{}^{k+l+1} \dot{c}_{k+2l}{}^l, \tag{1.8}$$

which has no gauge invariance of its own. However, this does not solve the problem: Basically, after gauge fixing we find terms of the form $\pi\dot{c}$ for all ghosts *except* c_i^i. For the superparticle this leads to a strange new gauge invariance of the effective action, whereas for the superstring it becomes impossible to remove the nonlinear terms in the action by various redefinitions. In both cases it appears that we would need to introduce a term

$$c_i{}^* \pi_i{}^0 \tag{1.9}$$

into S_{nonmin}. Indeed, in Ref. 7 such a term appears. However, c_i is a *minimal* field, and hence this clearly violates the master equation (and thus ultimately BRST invariance of the final action). Nevertheless, this is a suggestive idea because, as we shall see, the first-class superparticle has an analogous term in the *minimal* sector. (An alternate proposal, based on a "second round" of BV quantization to fix the residual symmetries, is discussed in Ref. 20. The results seem closely related.)

The first-class superparticle[17,18] is a modification of the original that is free of second-class constraints. The classical action is

$$S_0 = S_0{}^1 + S_0{}^2, \tag{1.10}$$

$$S_0{}^1 = \dot{x}p - i\dot{\theta}\not{p}\theta - \tfrac{1}{2}gp^2 + i\dot{\theta}d - i\psi\not{p}d, \tag{1.11}$$

$$S_0{}^2 = \tfrac{1}{16} d\lambda d, \quad \lambda^T = -\lambda, \tag{1.12}$$

and leads to the field equations

$$\begin{aligned}&gp = \dot{x} - i\dot{\theta}\gamma\theta - i\psi\gamma d,\\ &\dot{p} = 0, \quad p^2 = 0, \quad \not{p}d = 0,\\ &\dot{d} = 2\not{p}\dot{\theta}, \quad \dot{\theta} = \not{p}\psi - i\tfrac{1}{8}\lambda d, \quad d \otimes d = 0.\end{aligned} \tag{1.13}$$

The terms additional to (1.1) are manifestly supersymmetric, due in particular to the invariance of d (upon quantization, the usual supersymmetry covariant spinor derivative). (Note that d plays the role of $\pi_0{}^0$, as discussed above.) The superparticle with the action $S_0{}^1$ (see Ref. 21) is free of second-class constraints; the extra constraint imposed by $S_0{}^2$ is needed only to give an irreducible multiplet of states for $D>4$ (or $N>1$). In general, the arbitrary antisymmetric bispinor λ always contains a three-form piece $\gamma^{abc}\lambda_{abc}$, which constrains the multiplet to have superspin 0. (For $N=1$ and $D=3$ or 10, this is all of λ.) When applicable, it also includes a term $\sigma_{\text{INT}}\gamma^a\lambda_{\text{INT},a}$ constraining the multiplet to superisospin 0. Sometimes λ has further terms beyond λ_{abc} and $\lambda_{\text{INT},a}$, whose constraints are redundant, as follows from the analysis of Ref. 22. We include these redundant constraints here, and work with the bispinor $\lambda^{\alpha\beta}$, to simplify the analysis and allow treatment of all D and N simultaneously.

This action has a large number of symmetries: $S_0{}^1$ by itself is invariant under the usual diffeomorphisms

$$\delta_\xi x = \xi p, \quad \delta_\xi g = \dot{\xi}, \quad \delta_\xi(\text{rest}) = 0, \tag{1.14}$$

and κ symmetry,

$$\delta_\kappa \theta = \not{p}\kappa, \quad \delta_\kappa g = 4i\kappa\dot{\theta}, \quad \delta_\kappa \psi = \dot{\kappa}, \quad \delta_\kappa(\text{rest}) = 0, \tag{1.15}$$

as well as two new symmetries:

$$\delta_\omega \psi = -\not{p}\omega, \quad \delta_\omega g = 2i\omega d, \quad \delta_\omega(\text{rest}) = 0, \tag{1.16}$$

$$\delta_a \psi = ad, \quad \delta_a(\text{rest}) = 0. \tag{1.17}$$

The a symmetry is redundant in the sense of Ref. 11 (it can be generated by a nonlocal ω symmetry). The ω symmetry, however, plays a crucial role. Adding S_0^2, we find that the total S_0 has additional symmetries. The symmetry generated by the constraint to which S_0^2 leads is

$$\begin{aligned}&\delta_\eta\theta = -i\tfrac{1}{8}\eta d, \quad \delta_\eta x = -\tfrac{1}{8}d\eta\gamma\theta, \quad \delta_\eta d = i\tfrac{1}{8}\not{p}\eta d,\\ &\delta_\eta g = -\tfrac{1}{2}\psi\eta d, \quad \delta_\eta\lambda = \dot{\eta} + i\tfrac{1}{4}(\lambda\not{p}\eta - \eta\not{p}\lambda),\\ &\delta_\eta(\text{rest}) = 0,\end{aligned} \tag{1.18}$$

where also $\eta^T = -\eta$.

There are two further symmetries for $S_0{}^1 + S_0{}^2$:

$$\begin{aligned}&\delta_A\psi = Ad, \quad \delta_A\lambda = 8i(A^T\not{p} - \not{p}A), \quad \delta_A(\text{rest}) = 0,\\ &\delta_X\lambda^{\alpha\beta} = -X^{\alpha\beta\gamma}d_\gamma, \quad \delta(\text{rest}) = 0,\end{aligned} \tag{1.19}$$

where $A_\alpha{}^\beta$ is a bispinor with no symmetry, and $X^{\alpha\beta\gamma}$ is antisymmetric in $\alpha\beta$ while its cyclic part in $\alpha\beta\gamma$ vanishes. The A symmetry includes the a symmetry as a special case (the identity matrix). We shall return to these symmetries in Sec. IV and consider for now only the ξ, κ, and η symmetries, which close among themselves.

II. SOLUTIONS WITHOUT BISPINORS OR MULTISPINORS

The algebra of the κ and η symmetries is

$$[\delta_{\kappa_1},\delta_{\kappa_2}] = \delta_\xi(-4i\kappa_1 \not{p}\kappa_2), \quad [\delta_\kappa,\delta_\eta] = \delta_\xi(-\tfrac{1}{2}\kappa\eta d), \quad [\delta_{\eta_1},\delta_{\eta_2}] = \delta_\eta(i\tfrac{1}{4}\eta_{[1}\not{p}\eta_{2]}). \tag{2.1}$$

(We define $[ab] = ab - ba$.) The ghosts for ξ, κ, and η symmetries are denoted by χ, $-c_1$, and η, respectively.

Applying the BV procedure, we find the action that takes care of the ξ, κ, and η symmetries by solving $(S_1^0,S_1^0) + 2(S_0,S_2^0) = 0$:

$$\begin{aligned}S_1{}^0 &= x^*[p\chi + i(d\gamma + \theta\gamma\not{p})c_1 + \tfrac{1}{8}d\eta\gamma\theta] + \theta^*[-\not{p}c_1 + i\tfrac{1}{8}\eta d] + g^*[\dot\chi + 4i\dot\theta c_1 + \tfrac{1}{2}\psi\eta d] + \psi^*[-\dot c_1] + d^*[i\tfrac{1}{4}\not{p}\eta d]\\ &\quad + \tfrac{1}{2}\lambda^*[\dot\eta + i\tfrac{1}{4}(\lambda\not{p}\eta - \eta\not{p}\lambda)] + \chi^*[2ic_1\not{p}c_1 + \tfrac{1}{2}d\eta c_1] + i\tfrac{1}{8}\eta^*[\eta\not{p}\eta], \qquad (2.2)\\ S_2{}^0 &= 2ig^*x^*c_1\gamma c_1 - i\tfrac{1}{8}x^*\lambda^*(\eta\gamma\eta),\end{aligned}$$

where we normalize $(\lambda^{\alpha\beta},\lambda_{\gamma\delta}{}^*) = \delta_{[\gamma}{}^\alpha\delta_{\delta]}{}^\beta$ and $\lambda^*\dot\eta$ stands for $\lambda_{\alpha\beta}{}^*\dot\eta^{\alpha\beta}$ with unrestricted summation over α,β. We gauge-fix this system by

$$\begin{aligned}S_{\text{nonmin}} &= \hat\chi^*\pi_\xi + \hat c_1{}^*\pi_1 + \hat\eta^*\pi_\eta,\\ \Psi &= \hat\chi(g-1) - \hat c_1\psi + \tfrac{1}{2}\hat\eta\lambda.\end{aligned} \tag{2.3}$$

Note that (in contrast to Ref. 2) we have no need for a "pyramid" of ghosts. We find, for the antifields,

$$\hat\chi^* = g - 1, \quad g^* = \hat\chi, \quad \hat c_1{}^* = -\psi, \quad \psi^* = -\hat c_1, \tag{2.4}$$

$$\hat\eta^* = \lambda, \quad \lambda^* = \hat\eta, \quad \text{rest} = 0, \tag{2.5}$$

which gives the gauge-fixed action

$$\begin{aligned}S_\Psi &= \dot x p - i\dot\theta\not{p}\theta - \tfrac{1}{2}gp^2 + i\dot\theta d - i\psi\not{p}d + \tfrac{1}{16}d\lambda d\\ &\quad + \hat\chi(\dot\chi + 4i\dot\theta c_1 + \tfrac{1}{2}\psi\eta d) + \hat c_1\dot c_1\\ &\quad + \tfrac{1}{2}\hat\eta[\dot\eta + i\tfrac{1}{4}(\lambda\not{p}\eta - \eta\not{p}\lambda)]\\ &\quad + (g-1)\pi_\xi - \psi\pi_1 + \lambda\pi_\eta.\end{aligned} \tag{2.6}$$

Introducing the shifted variables

$$\begin{aligned}\tilde\pi_\xi &= \pi_\xi - \tfrac{1}{2}p^2, \quad \tilde\pi_1 = \pi_1 + i\not{p}d + \tfrac{1}{2}\hat\chi\eta d,\\ \tilde\pi_\eta &= \pi_\eta + \tfrac{1}{16}d\otimes d - i\tfrac{1}{8}(\not{p}\eta\hat\eta + \hat\eta\eta\not{p}),\\ \tilde d &= d - 4\hat\chi c_1 - \not{p}\theta,\end{aligned} \tag{2.7}$$

and dropping the tildes, we find

$$\begin{aligned}S_\Psi &= \dot x p - \tfrac{1}{2}p^2 + i\dot\theta d + \hat\chi\dot\chi + \hat c_1\dot c_1 + \tfrac{1}{2}\hat\eta\dot\eta\\ &\quad + (g-1)\pi_\xi + \lambda\pi_\eta - \psi\pi_1.\end{aligned} \tag{2.8}$$

This is clearly a fully gauge-fixed action. However, we believe that its cohomology is wrong and therefore we proceed.

The next simplest case takes the ghost ω_0 into account. In S_1 there are two new terms coming from (1.16),

$$S_1{}^1 = -g^*2id\omega_0 - \psi^*\not{p}\omega_0, \tag{2.9}$$

while there are no new terms in S_2. Hence $S_1 = S_1{}^0 + S_1{}^1$, $S_2 = S_2{}^0$, and $(S,S) = 0$. Adding a term $\hat c_2{}^*\pi_2$ to the nonminimal action, and a term $\hat c_2\omega_0$ to the gauge fermion, the final quantum action reads

$$S_\Psi = S_\Psi\ [\text{in (2.6)}] + \omega_0\pi_2 - \hat\chi 2id\omega_0 + \hat c_1\not{p}\omega_0, \tag{2.10}$$

and, shifting ω_0, we recover the free gauge-fixed BRST-invariant action $S_\Psi = S_\Psi$ [in (2.8)] $+ \tilde\omega_0\pi_2$. It is, however, independent of the antighost $\hat c_2$. This is a signal that this system has ghosts-for-ghosts. *On shell*, the transformations (1.14)–(1.16) and (1.18) are inert under

$$\delta\kappa = \not{p}\kappa_2, \quad \delta\omega = \dot\kappa_2 + \not{p}\omega_1, \quad \delta\xi = -2i\kappa_2 d, \quad \delta\eta = 0. \tag{2.11}$$

We see the beginning of *two* towers of ghosts emerging: the usual κ-symmetry tower c_i, and a second tower of ghost gauge fields ω_i. From (2.11) we see that we should add

$$S_1{}^2 = -c_1{}^*\not{p}c_2 + \omega_0{}^*(\dot c_2 + \not{p}\omega_1) - 2i\chi^*c_2 d. \tag{2.12}$$

This also leads to new terms in S_2. Proceeding, we find the final answer for this case to be

$$\begin{aligned}S_1 &= x^*[p\chi + i(d\gamma + \theta\gamma\not{p})c_1 + \tfrac{1}{8}d\eta\gamma\theta] + i\tfrac{1}{8}\theta^*\eta d + g^*[\dot\chi + 4i\dot\theta c_1 + \tfrac{1}{2}\psi\eta d - 2id\omega_0] + d^*[i\tfrac{1}{4}\not{p}\eta d]\\ &\quad + \tfrac{1}{2}\lambda^*[\dot\eta + i\tfrac{1}{4}(\lambda\not{p}\eta - \eta\not{p}\lambda)] + \chi^*[2ic_1\not{p}c_1 + \tfrac{1}{2}d\eta c_1 + 2idc_2] + i\tfrac{1}{8}\eta^*[\eta\not{p}\eta] - c_i{}^*\not{p}c_{i+1} + \omega_{i-1}{}^*[\dot c_{i+1} + \not{p}\omega_i],\\ S_2 &= 2g^*[ix^*(c_1\gamma c_1 - \theta\gamma c_2) - 4i\chi^*c_1c_2 + c_i{}^*c_{i+2} + \omega_{i-1}{}^*\omega_{i+1}] \qquad (2.13)\\ &\quad - x^*[\omega_{i-1}{}^*\gamma c_{i+2} + i\tfrac{1}{8}\lambda^*(\eta\gamma\eta)] - 2\chi^*\omega_{i-1}{}^*c_{i+3},\end{aligned}$$

where $c_0{}^* = \theta^*$, $\omega_{-1}{}^* = -\psi^*$, and all sums over i run from 0 to infinity. [The last term in S_2 is needed to ensure $(S_1,S_2) = (S_2,S_2) = 0$.]

The gauge fixing is very similar to that of the first model, except that we now also have to fix the c_i and ω_i symmetries. The nonminimal action contains an additional term $\Sigma_{i=2}^\infty \hat c_i{}^*\pi_i$ and the gauge fermion contains the extra term $\Sigma_{i=0}^\infty \hat c_{i+2}\omega_i$. The antifields $\omega_i{}^*$ become $\hat c_{i+2}$, and $\hat c_i{}^*$ becomes ω_{i-2}, for $i \geqslant 2$. The gauge-fixed action is given by

$$S_\Psi = S_\Psi \ [\text{in } (2.6)] - 2i\hat{\chi} d\omega_0 + \sum_{i=0}^{\infty} \hat{c}_{i+1} \not{p}\omega_i$$

$$+ \sum_{i=2}^{\infty} \hat{c}_i \dot{c}_i + 2\hat{\chi} \sum_{i=1}^{\infty} \hat{c}_i \omega_i + \sum_{i=0}^{\infty} \omega_i \pi_{i+2}. \qquad (2.14)$$

Introducing the same shifts as in (2.7), and further

$$\tilde{\pi}_2 = \pi_2 - 2i\hat{\chi}(d + \not{p}\theta) + \not{p}\hat{c}_1,$$

$$\tilde{\pi}_{i+1} = \pi_{i+1} - 2\hat{c}_{i-1}\hat{\chi} - (-)^i \not{p}\hat{c}_i \quad (i > 1), \qquad (2.15)$$

and dropping tildes, we finally obtain

$$S_\Psi = \dot{x}p - \tfrac{1}{2}p^2 + i\dot{\theta} d + \hat{\chi}\dot{\chi} + \sum_{i=1}^{\infty} \hat{c}_i \dot{c}_i + \frac{1}{2}\hat{\eta}\dot{\eta}$$

$$+ (g-1)\pi_\xi + \sum_{i=0}^{\infty} \omega_{i-1}\pi_{i+1} + \lambda\pi_\eta. \qquad (2.16)$$

This is clearly again a fully gauge-fixed action, but once more we believe that its cohomology is wrong.

III. THE SUPERPARTICLE WITH TOWERS OF BISPINOR GHOSTS

In Sec. II we considered two simple solutions that are building blocks for our final formulation and that both satisfy $(S,S) = 0$: (i) solution I without any towers of ghosts, which takes only the ghosts c_1, χ, and η for the classical symmetries κ, ξ, and η into account; and (ii) solution II with a tower of spinorial ghosts-for-ghosts c_i ($c_0 \equiv \theta$) and a corresponding tower of ghost gauge fields ω_i ($\omega_{-1} \equiv -\psi$), which takes the classical $\omega \equiv \omega_0$ symmetry and its descendants ω_i into account, but not yet the classical A and X symmetries. Solution II clearly contains solution I.

In this section we will extend solution II and take the A symmetry and its descendants into account. We obtain two new towers of ghost fields which will turn out to be bispinors. In Sec. IV we take the X symmetry into account and obtain further towers of multispinors. The model with bispinor ghosts is completely solved in this section. We present an action S satisfying $(S,S) = 0$ and construct the corresponding quantum action by fixing all gauge and ghost gauge fields. Despite the fact that all fields in the quantum action have propagators, this solution cannot be considered a correct quantization of the superparticle since we did not take the X symmetry into account; presumably, the cohomology of this model is wrong. As a building block for our final formulation, however, it is very useful. A new aspect of this solution is the occurrence of terms in S with three antifields, denoted by S_3. We first deduce by general arguments the structure of the terms in S_1, and then fix the few free constants by evaluating (S_1,S_1) in a few easy sectors. These general arguments are based partly on the symmetry structure of the bispinors and partly on analogies between the spinor ghosts and the bispinor ghosts. Having fixed S_1, we then determine S_2 from $(S_1,S_2) + 2(S_0,S_2) = 0$. Terms without any classical antifields in S_2 are not found by this procedure, but follow from the requirement that $(S_1,S_2) = 0$ on the S_0 shell. However, (S_1,S_2) does not vanish by itself, but contains terms proportional to the S_0 field equations, and in this way we find S_3.

Since the solutions in Sec. II were rather simple, we did not elaborate on their formal structure. However, for the bispinor ghosts complications arise concerning the (anti) symmetry of their indices, and hence we first go back to the ω_i and c_i towers and discuss their index structure. Defining θ to have a contravariant spinor index, $\theta = \theta^\alpha$, it follows from the terms in the classical action S_0 and the transformation rules $\delta\psi \sim \dot{\kappa} - \not{p}\omega_0$, $\delta\omega_i \sim \dot{c}_{i+2} + \not{p}\omega_{i+1} + \cdots$, and $\delta c_i \sim -\not{p}c_{i+1} + \cdots$ that the spinor index structure is as follows:

$$c_{2i}^\alpha; \quad c_{2i+1,\alpha}; \quad d_\alpha; \quad \omega_{21}^\alpha; \quad \omega_{2i+1,\alpha}; \quad \lambda^{\alpha\beta}; \quad \eta^{\alpha\beta}. \qquad (3.1)$$

It follows that the classical gauge parameter A has mixed spinor indices, for example, $\delta\psi_\alpha = A_\alpha{}^\beta d_\beta$. At this point, one could decide to introduce only the ghost corresponding to $A_\alpha{}^\beta$, denoted again by $A_\alpha{}^\beta$. Its transformation rule follows from the classical gauge algebra

$$[\delta(\eta),\delta(A)] = \delta(\hat{A} = -i\tfrac{1}{4}A\eta), \quad [\delta(A),\delta(\text{rest})] = 0. \qquad (3.2)$$

It is then straightforward to verify that the only new terms to be added to the action of model II in (1.28) are as follows:

$$\Delta S_1 = -\psi^* A d - 8i\lambda^*(\not{p}A) + i\tfrac{1}{4}A^*(A\not{p}\eta),$$

$$\Delta S_2 = -4g^*\lambda^*(\eta A). \qquad (3.3)$$

[We recall that expressions like $A^*(A\not{p}\eta)$ stand for $\Sigma_{\alpha,\beta} A_\alpha{}^{\beta *}(A\not{p}\eta)_\beta{}^\alpha$ with unrestricted summation over α,β. Since λ^* is antisymmetric while A^* has no symmetry, the corresponding transformation law for λ in (1.19) contains an extra factor 2 w.r.t. δS_1.]

The resulting model satisfies $(S,S) = 0$. Choosing the gauge fermion and nonminimal action as an extension of (2.3) and the addition below (2.13),

$$S_{\text{nonmin}} = \hat{\chi}^* \pi_\xi + \sum_{i=1}^{\infty} \hat{c}_i{}^* \pi_i + \hat{\eta}^* \pi_\eta + \hat{\eta}_A{}^* \pi_A,$$

$$\Psi = \hat{\chi}(g-1) + \sum_{i=0}^{\infty} \hat{c}_{i+1}\omega_{i-1} + \frac{1}{2}\hat{\eta}\lambda + \hat{\eta}_A A, \qquad (3.4)$$

we find that

$$S_\Psi = S_\Psi [\text{in } (2.14)] + A\text{-dependent terms}$$

$$\text{from } \Delta S_1 \text{ and } \Delta S_2 + A\pi_A. \qquad (3.5)$$

Clearly, a shift of π_A leads to $S_\Psi = S_\Psi \ [\text{in } (2.14)] + A\tilde{\pi}_A$, which is independent of $\hat{\eta}_A$, and hence, as for (2.10), we need ghosts-for-ghosts.

Comparing the BRST transformation laws of solution II,

$$\delta\lambda \sim (-A^T\not{p} + \not{p}A) + \dot{\eta} + \cdots, \quad \delta\eta \sim \cdots,$$

$$\delta\psi \sim \not{p}\omega_0 + \dot{c}_1, \quad \delta c_i \sim \not{p}c_{i+1}, \qquad (3.6)$$

one is led to expect that λ and A actually form the beginning of an infinite tower of gauge ghost fields, and that η is the tip of a corresponding tower of ghosts-for-ghosts. The simplest case would be the following generic set of BRST rules:

$$\delta c_i \sim \not{p}c_{i+1}, \quad \delta\omega_i \sim \not{p}\omega_{i+1} + \dot{c}_{i+1},$$

$$\delta\eta_i \sim \not{p}\eta_{i+1}, \quad \delta\lambda_i \sim \not{p}\lambda_{i+1} + \dot{\eta}_i. \qquad (3.7)$$

Indeed, these are "symmetries of symmetries." However, since $\lambda \equiv \lambda_0$ is antisymmetric, it transforms not as $\delta\lambda_0 \sim \not{p}\lambda_1$ but rather as $\delta\lambda_0 \sim \not{p}\lambda_1 - \lambda_1^T\not{p}$. This raises the possibility that, in general, one has $\delta\lambda_i \sim \not{p}\lambda_{i+1} + \mu_{i+1}\not{p}$. Since the λ_i with mixed spinor indices cannot have symmetry properties, the corresponding λ_{i+1} and μ_{i+1} could be independent fields, and thus lead to further symmetries. Since $\not p$ converts upper (lower) indices into lower (upper) indices, this implies that in the λ_{i+1} series the first index is alternatingly up or down as i increases, whereas for μ_{i+1} this holds for the second index. However, both $\not p\lambda_{i+1}$ and $\mu_{i+1}\not p$ appear in the variation of the same field λ_i, and if we want to keep the correlation between the even-or-odd i indices and the up-or-down spinor indices found in the c_i and ω_i series, there is no alternative but to introduce pairs of ij indices. To be clear, we give a few examples:

$$\lambda \equiv (\lambda_{00})^{\alpha\beta}, (\lambda_{01})^{\alpha}{}_{\beta}, (\lambda_{10})_{\alpha}{}^{\beta}, (\lambda_{11})_{\alpha\beta},$$
$$\eta \equiv (\eta_{00})^{\alpha\beta}, (\eta_{01})^{\alpha}{}_{\beta}, \text{etc.} \tag{3.8}$$

The index structure of λ_{ij} and η_{ij} leads us to consider the following set of generic transformation rules:

$$\delta\lambda_{ij} \sim \dot\eta_{ij} + \lambda_{i,j+1}\not p + \not p\lambda_{i+1,j} + \lambda_{i0}\not p\eta_{0j} + \eta_{i0}\not p\lambda_{0j},$$
$$\delta\eta_{ij} \sim \eta_{i,j+1}\not p + \not p\eta_{i+1,j} + \eta_{i0}\not p\eta_{0j}. \tag{3.9}$$

The last two terms in $\delta\lambda_{ij}$ and the last term in $\delta\eta_{ij}$ are new and are needed to satisfy $(S,S) = 0$. We can understand why they arise as follows. For $\delta\lambda_{00}$ these terms are required by classical gauge invariance of the action. This leads us to consider general terms like $\lambda_{mn}\not p\eta_{pq}$ (since $\not p$ has indices as in $\not p^{\alpha\beta}$ or $\not p_{\alpha\beta}$, it must be contracted with indices on λ and η that are both up or both down). However, conservation of ghost number and symmetries of the λ and η fields (to be discussed) suggest that we restrict ourselves to terms with $n = p = 0$.

As mentioned at the beginning of this section, the classical A symmetry is given by $\delta\psi \sim Ad$, $\delta\lambda \sim \not pA - A^T\not p$, and the A ghost transforms as $\delta A \sim A\not p\eta$. Identifying $\not pA$ with $\not p\lambda_{i0}$, we extend these terms to the ω and λ towers. The term $\delta\theta \sim \eta d$ leads similarly to new ηd terms in the c tower. (The field d is not the beginning of another tower; rather, d may be identified with $\omega_{-2}{}^*$.) In this way, we arrive at the following set of generic transformation rules:

$$\delta c_i \sim \not p c_{i+1} + \eta_{i0} d,$$
$$\delta\omega_i \sim \dot c_{i+2} + \not p\omega_{i+1} + \lambda_{i+2,0} d,$$
$$\delta\eta_{ij} \sim \eta_{i,j+1}\not p + \not p\eta_{i+i,j} + \eta_{i0}\not p\eta_{oj},$$
$$\delta\lambda_{ij} \sim \dot\eta_{ij} + \lambda_{i,j+1}\not p + \not p\lambda_{i+1,j} + \lambda_{i0}\not p\eta_{0j} + \eta_{i0}\not p\lambda_{oj}. \tag{3.10}$$

The coefficients of these terms must now be determined; they may be ij dependent and should follow from $(S_1,S_1) = 0$ on the S_0 shell. Before tackling this problem we must deal with another important issue, the symmetries of λ_{ij} and η_{ij}. To illustrate the general procedure to determine the symmetries of η_{ij} and λ_{ij}, we perform a minicalculation in a simplified example. Assume that, to linear order in fields,

$$\delta\eta_{00} = \eta_{01}\not p + \not p\eta_{10} + \cdots,$$
$$\delta\eta_{01} = \eta_{02}\not p - \not p\eta_{11} + \cdots, \tag{3.11}$$
$$\delta\eta_{,0} = \alpha_1\eta_{11}\not p + \not p\eta_{20} + \cdots.$$

All coefficients in this simplified example have been made equal to ± 1 by scaling of η_{01}, η_{10}, and η_{11} to reflect the actual case, except the coefficient α_1, which cannot be freely chosen, since it stands in front of a field (η_{11}) that did appear before. Since η_{00} is antisymmetric, $\eta_{10}^T = -\eta_{01}$, where the transformation symbol T indicates transposition in spinor space. Therefore, also $\delta\eta_{10}^T = -\delta\eta_{01}$, and one finds

$$\alpha_1\eta_{11}^T = \eta_{11}, \quad \eta_{20}^T = -\eta_{02}. \tag{3.12}$$

Requiring that $(S,S) = 0$ implies for η_{00} that $\delta(\delta\eta_{00}) = 0$ modulo S_0 field equations. Since

$$\delta(\delta\eta_{00}) = \eta_{02}p^2 - \not p\eta_{11}\not p + \alpha_1\not p\eta_{11}\not p + p^2\eta_{20}, \tag{3.13}$$

it follows that $\alpha_1 = +1$. Hence η_{11} is symmetric. At the next level we have

$$\delta\eta_{20} = \eta_{21}\not p + \not p\eta_{30} + \cdots,$$
$$\delta\eta_{02} = \eta_{03}\not p + \not p\eta_{12} + \cdots, \tag{3.14}$$
$$\delta\eta_{11} = \alpha_2\eta_{12}\not p + \alpha_3\not p\eta_{21} + \cdots.$$

We find, from nilpotency at the previous level,

$$\delta\delta\eta_{01} = \eta_{03}p^2 + \not p\eta_{12}\not p - \alpha_2\not p\eta_{12}\not p - \alpha_3 p^2\eta_{21},$$
$$\delta\delta\eta_{10} = (\alpha_2\eta_{12}p^2 + \alpha_3\not p\eta_{21}\not p) + (\not p\eta_{21}\not p + p^2\eta_{30}). \tag{3.15}$$

Clearly, $\alpha_2 = +1$ and $\alpha_3 = -1$.

From the symmetry properties at this level, $\eta_{20}^T = -\eta_{02}$ and $\eta_{11}^T = \eta_{11}$, we then find the symmetry properties at the next level.

$$\eta_{21}^T = -\eta_{12}, \quad \eta_{30}^T = -\eta_{03}, \quad \eta_{12}^T = -\eta_{21}. \tag{3.16}$$

Continuing in this way we find the following symmetry properties of η_{ij}:

$$\eta_{ij}^T = -(-)^{ij}\eta_{ji}. \tag{3.17}$$

Since $\delta\lambda_{ij}$ contains a term $\dot\eta_{ij}$, λ_{ij} also has this symmetry:

$$\lambda_{ij}^T = -(-)^{ij}\lambda_{ji}. \tag{3.18}$$

Using these symmetry properties, we can now correlate various coefficients in the transformation rules. For example, from

$$\delta\eta_{ij} = \beta^1_{ij}\eta_{i,j+1}\not p + \beta^2_{ij}\not p\eta_{i+1,j} + \cdots, \tag{3.19}$$

we obtain

$$\beta^2_{ij} = \beta^1_{ji}(-)^j, \quad \beta^1_{ij} = \beta^2_{ji}(-)^i. \tag{3.20}$$

In this way we arrive at the following transformation rules for η and λ:

$$\delta\eta_{ij} = -(\eta_{i,j+1}\not p + (-)^j\not p\eta_{i+1,j}) + \beta^3_{ij}\eta_{i0}\not p\eta_{oj},$$
$$\delta\lambda_{ij} = \beta^4_{ij}\dot\eta_{ij} + (\lambda_{i,j+1}\not p + (-)^j\not p\lambda_{i+1,j}) \tag{3.21}$$
$$+ \beta^5_{ij}(\lambda_{i0}\not p\eta_{oj} - (-)^j\eta_{i0}\not p\lambda_{0j}).$$

We shall now fix the coefficients β^3_{ij}, β^4_{ij}, and β^5_{ij} by requiring that the $\eta_{ij}{}^*$ and $\lambda_{ij}{}^*$ terms in (S_1,S_1) vanish modulo the S_0 field equations. Since the variation of, for example, λ_{i0} in $\lambda_{i0}\not p\eta_{0j}$ is obtained by first moving λ_{i0} to the right, then replacing it by $\delta\lambda_{i0} = (\lambda_{i0}{}^*, S_1)$, which has opposite statistics from λ_{i0}, and then moving this $\delta\lambda_{i0}$ back to its original position, one picks up in this process the fermion number of η_{0j}. These fermion numbers F can be deduced from the transformation laws, and are given by

$$F(c_i) = (-)^{i+1}, \quad F(\omega_i) = (-)^i,$$
$$F(\eta_{ij}) = (-)^{i+j+1}, \quad F(\lambda_{ij}) = (-)^{i+j}. \tag{3.22}$$

We are now ready to evaluate $\delta\delta\lambda_{ij}$ and $\delta\delta\eta_{ij}$. From the $\dot{\eta}\not{p}$ terms in $\delta\delta\lambda_{ij}$ we learn that

$$\beta^4_{i,j+1} = \beta^4_{ij} = \beta^4_{i+1,j} = \beta^4 = \text{const.} \tag{3.23}$$

(More precisely, only the symmetric part of β^4_{ij} appears in S_1, and this part is ij independent.) Next we evaluate all $\eta_{ij}{}^*$ terms in (S_1,S_1), and find

$$\beta^3_{ij} = (-)^j\beta^3, \quad \beta^3 = \text{const.} \tag{3.24}$$

[Again only $\hat{\beta}^3_{ij} \equiv \beta^3_{ij} + (-)^{i+j}\beta^3_{ji}$ appears in the action, and one finds that $\hat{\beta}^3_{ij} = \beta^3_i(-)^j$. Using the freedom to redefine η_{i0}, we arrive then at $\hat{\beta}^3_{ij} = (-)^j\beta^3$.] The $\lambda_{ij}{}^*$ terms in (S_1,S_1) vanish modulo the S_0 field equations provided the following conditions hold. From the $\lambda^*\eta\dot{\eta}$ terms we learn that

$$\beta^5_{ij} = (-)^j\beta^3_{ij}, \tag{3.25}$$

and these two relations are compatible. We can always redefine $\eta_{i,j+1}$ w.r.t. $\eta_{i,j}$ such that β^1_{ij} is an ij-independent constant, and find then that BRST nilpotency (which requires that $\beta^1_{ij}\beta^1_{j,i+1}$ be symmetric in ij) is satisfied. For later use we note that the scale of η_{i0} is still left free at this point. Similar steps can be performed for $\delta_{\lambda ij}$. The rest of the λ^* terms vanish provided

$$\beta^5_{ij} = \beta^5 = \text{const}, \quad \beta^3_{ij} = (-)^j\beta^5. \tag{3.26}$$

Hence at this point all $\delta\lambda_{ij}$ and $\delta\eta_{ij}$ terms are fixed up to overall constants β^4 and β^5.

Next we analyze the $\delta\omega_i$ and δc_i laws. Using the freedom one has in redefining fields at higher levels, we may start from

$$\delta c_i = -\not{p}c_{i+1} + \beta^6_i\eta_{i0}d,$$
$$\delta\omega_i = \not{p}\omega_{i+1} + \beta^7_i\dot{c}_{i+2} + \beta^8_i\lambda_{i+2,0}d. \tag{3.27}$$

Since the d transformation rule is known from solution II,

$$\delta d = i\tfrac{1}{4}\not{p}\eta_{00}d, \tag{3.28}$$

we expect to find definite values for (some of) the constants. From $\delta\delta c_i = 0$ we find

$$\beta^6_i = \beta^6 = \text{const}, \quad \beta^3 = i/4. \tag{3.29}$$

From the $\omega_i{}^*$ sector we then find that $\beta^5 = i\beta^4/4$; $\beta^5 = i/4$, $\beta^7_i = \beta^7 = \text{const}$, $\beta^8_i = \beta^8 = \text{const}$, $\beta^6\beta^7 = \beta^8\beta^4$. Hence we have at this moment only three free constants left: β^6, β^7, and β^8 in the δc_i and $\delta\omega_i$ laws.

Finally, we fix the remaining constants by requiring nilpotency also in other sectors. Using the g^* terms in S_1,

$$g^*[\dot{\chi} + 4i\dot{\theta}c_1 + \tfrac{1}{2}\psi\eta_{00}d - 2id\omega_0], \tag{3.30}$$

we find $\beta^7 = 1$ from the $\dot{c}_2$ terms and $\beta^6 = i/8$ from δc_1 in the x^* sector. Only the product $\beta^8\beta^4$ is fixed, because changing the scale of λ_{ij} only affects the $\dot{\eta}$ terms in $\delta\lambda$ and the λ term in $\delta\omega$. We choose $\beta^4 = 1$.

The final result for S_1 reads

$$\begin{aligned} S_1 = {} & x^*[\not{p}\chi + i(d\gamma + \theta\gamma\not{p})c_1 + \tfrac{1}{8}d\eta_{00}\gamma\theta] + d^*[i\tfrac{1}{4}\not{p}\eta_{00}d] + g^*[\dot{\chi} + 4i\dot{\theta}c_1 + \tfrac{1}{2}\psi\eta_{00}d - 2id\omega_0] \\ & + \chi^*[2ic_1\not{p}c_1 + \tfrac{1}{2}d\eta_{00}c_1 + 2idc_2] + \omega_{i-1}{}^*[\dot{c}_{i+1} + \not{p}\omega_i + i\tfrac{1}{8}\lambda_{i+1,0}d] + c_i{}^*[-\not{p}c_{i+1} + i\tfrac{1}{8}\eta_{i0}d] \\ & + \tfrac{1}{2}\lambda_{ij}{}^*[\dot{\eta}_{ij} + (\lambda_{i,j+1}\not{p} + \not{p}\lambda_{i+1,j}(-)^j) + i\tfrac{1}{4}(\lambda_{i0}\not{p}\eta_{0j} - \eta_{i0}\not{p}\lambda_{0j}(-)^j)] \\ & + \tfrac{1}{2}\eta_{ij}{}^*[-(\eta_{i,j+1}\not{p} + \not{p}\eta_{i+1,j}(-)^j) + i\tfrac{1}{4}\eta_{i0}\not{p}\eta_{0j}(-)^j], \end{aligned} \tag{3.31}$$

where all sums over i and j run from 0 to infinity. The factors $\tfrac{1}{2}$ in front of the $\lambda_{ij}{}^*$ and $\eta_{ij}{}^*$ terms account for the fact that in expressions like $\lambda_{ij}{}^*\dot{\eta}_{ij} \equiv \lambda^{\alpha\beta *}_{ij}\dot{\eta}_{ij,\alpha\beta}$ a given field $\lambda^{\alpha\beta *}_{ij}$ appears twice because of its symmetry $\lambda^{\alpha\beta}_{ij} = (-)^{i+j}\lambda^{\beta\alpha}_{ji}$. (The same factor of $\tfrac{1}{2}$ is needed when not both indices are up.) In a nonredundant but noncovariant notation, these factors $\tfrac{1}{2}$ would be absent, but we prefer to work with unrestricted summations in order to maintain covariance.

Having found an expression for S_1 satisfying $\tfrac{1}{2}(S_1,S_1) = 0$ modulo the S_0 field equations, we go back to the calculations of (S_1,S_1) and collect the terms in the matrix M^{AB} in

$$\frac{1}{2}(S_1,S_1) = \phi_A{}^*M^{AB}\frac{\partial_l S_0}{\partial\phi_B{}^*}.$$

This leads then to $S^0_2 = k\phi_A{}^*M^{AB}\phi_B{}^*$, where $k = 1$ in all cases except when both $\phi_A{}^*$ and $\phi_B{}^*$ correspond to classical antifields, in which case one needs $k = \tfrac{1}{2}$ to avoid overcounting in $\tfrac{1}{2}(S_1,S_1) + (S_2,S_0) = 0$. [Since such terms should occur twice in (S_1,S_1) this constitutes a useful check on the algebra.] Terms without any classical antifieds are not found in this way, as already noted. At the lowest level we find variations of the form

$$(c_i{}^*\eta_{i0} + \omega_{i-1}{}^*\lambda_{i+1,0})\not{p}d, \tag{3.32}$$

and we extend them to the η and λ towers as follows:

$$S^{(1)}_2 = (-)^j\tfrac{1}{8}c_i{}^*\eta_{i,j+1}\omega_{j-1}{}^* - \tfrac{1}{16}\omega_{i-1}{}^*\lambda_{i+1,j+1}\omega_{j-1}{}^*. \tag{3.33}$$

Note the extra factor $\tfrac{1}{2}$ in the second term. The correctness of this extension must follow from the (S_1,S_2) analysis. Similarly we extend the $\chi^*\eta\psi^*$ terms and the $\lambda_{00}{}^*\chi^*\eta$ and $\lambda_{00}{}^*\chi^*\eta\eta$ terms found in solution II. The final result for S_2 found in this way and satisfying $(S_1,S_2) = 0$ modulo the S_0 field equations reads

$$\begin{aligned} S_2 = {} & 2g^*\{ix^*(c_1\gamma c_1 - \theta\gamma c_2) - 4i\chi^*c_1c_2 + c_i{}^*c_{i+2} + \omega_{i-1}{}^*(\omega_{i+1} + i\tfrac{1}{4}\eta_{i+1,0}\psi) \\ & + i\tfrac{1}{4}d^*\eta_{10}d + \lambda_{ij}{}^*[\lambda_{i,j+2} - \tfrac{1}{4}(-)^j(\lambda_{i1}\eta_{0j} + \lambda_{i0}\eta_{1j})] + \eta_{ij}{}^*(\eta_{i,j+2} + i\tfrac{1}{4}\eta_{i1}\eta_{0j})\} \\ & - x^*[\omega_{i-1}{}^*\gamma c_{i+2} - i\tfrac{1}{8}\omega_{i-1}{}^*\eta_{i+1,0}\gamma\theta + \lambda_{ij}{}^*(-)^j(\gamma\eta_{i+1,j} + i\tfrac{1}{8}\eta_{i0}\gamma\eta_{0j})] \\ & - 2\chi^*[\omega_{i-1}{}^*c_{i+3} + i\tfrac{1}{4}(-)^i c_1\eta_{0,i+1}\omega_{i-1}{}^* + \lambda_{ij}{}^*(\eta_{i,j+2} - \tfrac{1}{4}\eta_{i0}\eta_{1j})] \\ & - \tfrac{1}{4}\omega_{i-1}{}^*\eta_{i+1,0}\not{p}d^* + \tfrac{1}{8}(-)^j c_i{}^*\eta_{i,j+1}\omega_{j-1}{}^* - \tfrac{1}{16}\omega_{i-1}{}^*\lambda_{i+1,j+1}\omega_{j-1}{}^*. \end{aligned} \tag{3.34}$$

Having obtained S_2 satisfying $(S_1,S_2)=0$ modulo the S_0 field equations, we next collect all terms in (S_1,S_2) that are proportional to the S_0 field equations. We do find such terms, but only in the g^*d^*, $g^*\omega_i{}^*$, $d^*\omega_i$ sectors. The classical field equations with a time derivative involve only $\dot p$, $\dot x$, $\dot\theta$, and $\dot d$ and can only come from terms in S_1 with a $\partial/\partial\tau$. In

$$(S_1,S_2)=\frac{\partial_r S_1}{\partial\phi^A}\frac{\partial_l S_2}{\partial\phi_A{}^*}+\frac{\partial_r S_2}{\partial\psi^A}\frac{\partial_l S_1}{\partial\psi_A{}^*}, \tag{3.35}$$

we need then only consider

$$\phi^A=\{\chi,c_i,\eta_{ij}\}\quad\text{and}\quad\psi_A{}^*=\{g^*,\omega_i{}^*,\lambda_{ij}{}^*\}. \tag{3.36}$$

However, g^* may be dropped as S_2 is g independent, and the only time derivatives of classical fields that remain are $\dot\theta$ in the $g^*\omega_i{}^*$ sector. This yields a $g^*d^*\omega_i{}^*$ term in S_3.

The S_0 field equations without time derivative are p^2, $\not p d$, and dd. Since none of these expressions is present in the field equations with a time derivative, we may directly collect all terms in (S_1,S_2) bilinear in p, and d's, and try to integrate them to S_3. Of course, since we already found a $g^*d^*\omega_i{}^*$ term in S_3 we should find at this point a p^2 term in the $d^*\omega_i{}^*$ sector of (S_1,S_2) and also a $\not p d$ term in the g^*d^* sector. We indeed find these terms. The S_3 we find is given by

$$S_3=\tfrac{1}{2}(-)^i g^*\omega_{i-1}{}^*\eta_{i+1,1}d^*. \tag{3.37}$$

We now complete the determination of S_3. First of all, $(S_2,S_3)=0$ and $(S_3,S_3)=0$. The latter is obvious, but also (S_2,S_3) vanishes since one always produces two g^*'s. Next, we consider the classical field equations in $(S_1,S_3)+\frac{1}{2}(S_2,S_2)$. The field equations with time derivatives are absent for the following reasons. We may omit the term (S_2,S_2) from consideration, as it contains no $\partial/\partial\tau$. Repeating the discussion below (3.35) but now with S_2 replaced by S_3, we see that $\phi^A=\{\chi,c_i,\eta_{ij}\}$ does not contribute, since S_3 does not contain their antifields. Analogously, the $\psi_A{}^*=\{g^*,\omega_i{}^*,\lambda_{ij}{}^*\}$ do not contribute since the corresponding fields are absent from S_3. Finally we consider the classical field equations without time derivatives, p^2, $\not p d$, and dd. The term (S_2,S_2) cannot produce such terms, as the part of S_2 linear in p and d is given by

$$S_2=i\tfrac{1}{8}g^*d^*\eta_{10}d-\tfrac{1}{4}\omega_{i-1}{}^*\eta_{i+1,0}\not p d^*+\cdots. \tag{3.38}$$

Clearly, (S_2,S_2) contains no p^2, $\not p d$, or dd terms. In S_3 we find no p or d; hence we need p^2 or $\not p d$ or dd terms in S_1. The only such term in S_1 is $S_1=\frac{1}{4}d^*\not p\eta_{00}d$ and since S_3 contains no d fields, also (S_1,S_3) cannot produce these field equations. Thus (3.37) is our final S_3.

As a final check on our results for the bispinor model, we make a dimensional analysis. If we assign dimensions as

$$\begin{aligned}
&\dim(x,p,\theta,d)=0,\quad \dim(x^*,p^*,\theta^*,d^*)=1,\\
&\dim(g,\omega_i,\lambda_{ij})=1,\quad \dim(g^*,\omega_i^*,\lambda_{ij}^*)=0,\\
&\dim(\chi,c_i,\eta_{ij})=0,\quad \dim(\chi^*,c_i{}^*,\eta_{ij}{}^*)=1,\\
&\dim\left(\frac{\partial}{\partial\tau}\right)=1,
\end{aligned} \tag{3.39}$$

then one could in principle have any S_k. This is different from the Green–Schwarz superparticle, where we found that only g^* had dimension zero, which explained the absence of S_k, for $k\geqslant 3$. However, ghost number conservation and spinor contractions restrict the possibilities further, and we have found no S_k with $k\geqslant 4$.

To obtain the gauge-fixed action we choose

$$\begin{aligned}
S_{\text{nonmin}}&=\hat\chi^*\pi_\xi+\sum_{i=1}^\infty\hat c_i{}^*\pi_i+\sum_{i,j=0}^\infty\hat\eta_{ij}{}^*\pi_{ij},\\
\Psi&=\hat\chi(g-1)+\sum_{i=0}^\infty\hat c_{i+1}\omega_{i-1}+\frac{1}{2}\sum_{i,j=0}^\infty\hat\eta_{ij}\lambda_{ij}.
\end{aligned} \tag{3.40}$$

The antifields are then eliminated by

$$\begin{aligned}
&\hat\chi^*=g-1,\quad \hat c_{i+1}{}^*=\omega_{i-1},\quad \hat\eta_{ij}{}^*=\lambda_{ij},\\
&g^*=\hat\chi,\quad \omega_{i-1}{}^*=\hat c_{i+1},\quad \lambda_{ij}{}^*=\hat\eta_{ij}.
\end{aligned} \tag{3.41}$$

After shifting fields, the gauge-fixed action becomes

$$\begin{aligned}
S_\Psi={}&\dot x p-\frac{1}{2}p^2+i\dot\theta d+\hat\chi\dot\chi+\sum_{i=1}^\infty\hat c_i\dot c_i\\
&+\frac{1}{2}\sum_{i,j=0}^\infty\hat\eta_{ij}\dot\eta_{ij}+(g-1)\pi_\xi\\
&+\sum_{i=0}^\infty\omega_{i-1}\pi_{i+1}+\sum_{i,j=0}^\infty\lambda_{ij}\pi_{ij}.
\end{aligned} \tag{3.42}$$

IV. THE FORMULATION WITH MULTISPINOR GHOSTS

As our final model, we take the X symmetry of (1.19), as well as its descendants, into account. Including these symmetries amounts to treating *all* local symmetries with the BV formalism, as we advocated in the Introduction, and leads to further towers of ghosts-for-ghosts $\eta_{i_1\ldots i_n}$ and ghost gauge fields $\lambda_{i_1\ldots i_n}$, with $n\geqslant 3$. Adding the X transformations of (1.19), the classical gauge algebra contains the following new commutators:

$$[\delta(\eta),\delta(X)]=\delta(\hat X),\quad [\delta(X),\delta(\text{rest})]=0, \tag{4.1}$$

where

$$\hat X^{\alpha\beta\gamma}=i\tfrac{1}{4}[(\eta\not p)^\alpha{}_\delta X^{\delta\beta\gamma}+(\eta\not p)^\beta{}_\delta X^{\alpha\delta\gamma}+(\eta\not p)^\gamma{}_\delta X^{\alpha\beta\delta}]. \tag{4.2}$$

Observe that $\hat X$ is again $\alpha\beta$ antisymmetric and its $\alpha\beta\gamma$-cyclic part vanishes. This leads us to introduce the following new terms in S_1:

$$\begin{aligned}
\Delta S_1={}&\tfrac{1}{2}\lambda_{00,\alpha\beta}{}^*\lambda_{000}^{\alpha\beta\gamma}d_\gamma-\tfrac{1}{3}\lambda_{000,\alpha\beta\gamma}{}^*i\tfrac{1}{4}[(\eta_{00}\not p)^\alpha{}_\delta\lambda_{000}^{\delta\beta\gamma}\\
&+(\eta_{00}\not p)^\beta{}_\delta\lambda_{000}^{\alpha\delta\gamma}+(\eta_{00}\not p)^\gamma{}_\delta\lambda_{000}^{\alpha\beta\delta}],
\end{aligned} \tag{4.3}$$

where $\lambda_{000}^{\alpha\beta\gamma}$ is the commuting three-spinor ghost for the X symmetry. Requiring the $\lambda_{00}{}^*$ terms in $(S_1,\Delta S_1)+\frac{1}{2}(\Delta S_1,\Delta S_1)$ to vanish on the S_0 shell allows the extra terms in the BRST transformations, for λ_{10}, λ_{01}, and η_{00},

$$\begin{aligned}
&\delta'\lambda_{10,\alpha}{}^\beta=\lambda_{100,\alpha}{}^{\beta\gamma}d_\gamma,\quad \delta'\lambda_{01,}{}^\alpha{}_\beta=\lambda_{010,}{}^\alpha{}_\beta{}^\gamma d_\gamma,\\
&\delta'\eta_{00}^{\alpha\beta}=\eta_{000}^{\alpha\beta\gamma}d_\gamma,
\end{aligned} \tag{4.4}$$

and, for λ_{000},

$$\begin{aligned}
\delta'\lambda_{000}^{\alpha\beta\gamma}={}&\dot\eta_{000}^{\alpha\beta\gamma}+\not p^{\alpha\delta}\lambda_{100,\delta}{}^{\beta\gamma}+\not p^{\beta\delta}\lambda_{010,}{}^\alpha{}_\delta{}^\gamma+\not p^{\gamma\delta}\lambda_{001,}{}^{\alpha\beta}{}_\delta\\
&-i\tfrac{1}{4}[\eta_{000}^{\alpha\beta\delta}(\not p\lambda_{00})_\delta{}^\gamma+\eta_{000}^{\alpha\delta\gamma}(\not p\lambda_{00})_\delta{}^\beta\\
&+\eta_{000}^{\delta\beta\gamma}(\not p\lambda_{00})_\delta{}^\alpha],
\end{aligned} \tag{4.5}$$

where λ_{100}, λ_{010}, and λ_{001} are new (anticommuting) gauge ghost fields, and η_{000} is the beginning of a new tower of ghosts-for-ghosts. One may verify that $\delta'\delta\lambda_{00}^{\alpha\beta}$ indeed vanishes on the classical shell.

Similarly, requiring the $\eta_{00}{}^*$ terms to vanish (up to classical field equations) allows for the additions to the BRST transformations of η_{10} and η_{01},

$$\delta'\eta_{10,\alpha}{}^{\beta} = \eta_{100,\alpha}{}^{\beta\gamma}\mathrm{d}_\gamma, \quad \delta'\eta_{01,}{}^{\alpha}{}_{\beta} = \eta_{010,}{}^{\alpha}{}_{\beta}{}^{\gamma}\mathrm{d}_\gamma, \tag{4.6}$$

and yields the BRST transformation for η_{000},

$$\begin{aligned}\delta\eta_{000}^{\alpha\beta\gamma} = &-(\not{p}^{\alpha\delta}\eta_{100,\delta}{}^{\beta\gamma} + \not{p}^{\beta\delta}\eta_{010,}{}^{\alpha}{}_{\delta}{}^{\gamma} + \not{p}^{\gamma\delta}\eta_{001,}{}^{\alpha\beta}{}_{\delta})\\ &+ i\tfrac{1}{4}[\eta_{000}^{\alpha\beta\delta}(\not{p}\eta_{00})_\delta{}^\gamma + \eta_{000}^{\alpha\delta\gamma}(\not{p}\eta_{00})_\delta{}^\beta\\ &+ (\eta_{00}\not{p})^\alpha{}_\delta\eta_{000}^{\delta\beta\gamma}],\end{aligned} \tag{4.7}$$

where η_{100}, η_{010}, and η_{001} are (commuting) ghosts-for-ghosts.

Generalizing (4.4) and (4.6) to the case of arbitrary λ_{ij} and η_{ij}, we have

$$\begin{aligned}\delta\lambda_{ij} &= \dot{\eta}_{ij} + \lambda_{i,j+1}\not{p} + \not{p}\lambda_{i+1,j}(-)^j + i\tfrac{1}{4}(\lambda_{i0}\not{p}\eta_{0j}\\ &\quad - \eta_{i0}\not{p}\lambda_{0j}(-)^j) + \lambda_{ij0}d,\\ \delta\eta_{ij} &= -\eta_{i,j+1}\not{p} - \not{p}\eta_{i+1,j}(-)^j\\ &\quad + i\tfrac{1}{4}\eta_{i0}\not{p}\eta_{0j} + \eta_{ij0}d,\end{aligned} \tag{4.8}$$

where the last terms on the right-hand side are new as compared to (3.25). Doing the same for (4.5) and (4.7), we obtain

$$\begin{aligned}\delta\lambda_{ijk} = &\,\dot{\eta}_{ijk} + (\lambda_{ij,k+1} + \lambda_{i,j+1,k}(-)^k + \lambda_{i+1,jk}(-)^{j+k})\not{p} + i\tfrac{1}{4}(\lambda_{ij0}\not{p}\eta_{0k} + \lambda_{i0k}\not{p}\eta_{0j}(-)^{jk} - \eta_{i0}\not{p}\lambda_{0jk}(-)^{j+k})\\ &- i\tfrac{1}{4}(\eta_{ij0}\not{p}\lambda_{0k}(-)^k + \eta_{i0k}\not{p}\lambda_{0j}(-)^{j(k+1)} - \lambda_{i0}\not{p}\eta_{0jk}),\\ \delta\eta_{ijk} = &-(\eta_{ij,k+1} + \eta_{i,j+1,k}(-)^k + \eta_{i+1,jk}(-)^{j+k})\not{p}\\ &+ i\tfrac{1}{4}(\eta_{ij0}\not{p}\eta_{0k}(-)^k + \eta_{i0k}\not{p}\eta_{0j}(-)^{j(k+1)} - \eta_{i0}\not{p}\eta_{0jk}(-)^{j+k}),\end{aligned} \tag{4.9}$$

where we determined the minus sign factors by requiring (S_1,S_1) to vanish modulo the S_0 field equations in the $\lambda_{ij}{}^*$ and $\eta_{ij}{}^*$ sectors. Here the λ_{ijk} satisfy

$$\lambda_{ijk} = -(-)^{ij}\lambda_{jik} \tag{4.10}$$

and

$$(-)^{ki}\lambda_{ijk} + (-)^{ij}\lambda_{jki} + (-)^{jk}\lambda_{kij} = 0, \tag{4.11}$$

i.e., λ_{ijk} is *graded* antisymmetric under ij interchange and its *graded* cyclic part vanishes (we suppressed all spinor indices as they merely "travel along" with the level indices). Furthermore, we have fermion number $F(\lambda_{ijk}) = i+j+k$ (mod 2) and ghost number $G(\lambda_{ijk}) = i+j+k+1$. The ghost-for-ghosts η_{ijk} have the same symmetry properties as λ_{ijk}, but $F(\eta_{ijk}) = i+j+k+1$ (mod 2) and $G(\eta_{ijk}) = i+j+k+2$.

Next, we should verify whether (S_1,S_1) also vanishes on shell in the $\lambda_{ijk}{}^*$ and $\eta_{ijk}{}^*$ sectors. However, it is here that we must take account of the fact that *on shell* the X transformation (1.19) is inert under $(X_3 \equiv X)$

$$\delta X_3^{\alpha\beta\gamma} = X_4^{\alpha\beta\gamma\delta}d_\delta, \tag{4.12}$$

and this in turn is left inert by

$$\delta X_4^{\alpha\beta\gamma\delta} = X_5^{\beta\gamma\delta\epsilon}d_\epsilon, \tag{4.13}$$

and so on, where each X_n is $\alpha\beta$ antisymmetric and its $\alpha\beta\gamma$-cyclic part vanishes, but otherwise has no symmetry. Indeed, from the BRST transformation for λ_{ijk} in (4.9), we have so far, for the $\lambda_{ijk}{}^*$ terms in S_1,

$$\begin{aligned}\tfrac{1}{3}\lambda_{ijk}{}^*[\dot{\eta}_{ijk} &+ (\lambda_{ij,k+1} + \lambda_{i,j+1,k}(-)^k + \lambda_{i+1,jk}(-)^{j+k})\not{p}\\ &+ i\tfrac{1}{4}(\lambda_{ij0}\not{p}\eta_{0k} + \lambda_{i0k}\not{p}\eta_{0j}(-)^{jk} - \eta_{i0}\not{p}\lambda_{0jk}(-)^{j+k})\\ &- i\tfrac{1}{4}(\eta_{ij0}\not{p}\lambda_{0k}(-)^k + \eta_{i0k}\not{p}\lambda_{0j}(-)^{j(k+1)} - \lambda_{i0}\not{p}\eta_{0jk}),\end{aligned} \tag{4.14}$$

and, upon calculating the $\lambda_{ijk}{}^*$ terms in $\tfrac{1}{2}(S_1,S_1)$, we obtain, dropping field equations,

$$\begin{aligned}\tfrac{1}{3}\lambda_{ijk}{}^*[i\tfrac{1}{4}(\lambda_{ij0}\not{p}\eta_{0k0} &+ \lambda_{i0k}\not{p}\eta_{0j0}(-)^{jk}\\ &+ \lambda_{0jk}\not{p}\eta_{0i0}(-)^{i(j+k)}) - i\tfrac{1}{4}(\eta_{ij0}\not{p}\lambda_{0k0}(-)^k\\ &+ \eta_{i0k}\not{p}\lambda_{0j0}(-)^{j(k+1)}\\ &+ \eta_{0jk}\not{p}\lambda_{0i0}(-)^{i(j+k+1)})]d.\end{aligned} \tag{4.15}$$

This can be canceled by adding further terms,

$$\begin{aligned}\Delta S_1 = \tfrac{1}{3}\lambda_{ijk}{}^*\lambda_{ijk0}d &+ \tfrac{1}{3}\lambda_{ijk0}{}^*[i\tfrac{1}{4}(\lambda_{ij0}\not{p}\eta_{0k0} + \lambda_{i0k}\not{p}\eta_{0j0}(-)^{jk}\\ &+ \lambda_{0jk}\not{p}\eta_{0i0}(-)^{i(j+k)}) - i\tfrac{1}{4}(\eta_{ij0}\not{p}\lambda_{0k0}(-)^k\\ &+ \eta_{i0k}\not{p}\lambda_{0j0}(-)^{j(k+1)} + \eta_{0jk}\not{p}\lambda_{0i0}(-)^{i(j+k+1)})],\end{aligned} \tag{4.16}$$

where the first term in the case $i=j=k=0$ is as expected from (4.12). However, it is not difficult to see that we now have an integrability problem in obtaining S_2. Namely, $\tfrac{1}{2}(S_1,S_1)$ contains a term

$$\tfrac{1}{2}\lambda_{ij}{}^*\lambda_{ij00}dd \tag{4.17}$$

and we therefore expect a term $\lambda_{ij}{}^*\lambda_{ijkl}\lambda_{kl}{}^*$ in S_2, but then (S_0,S_2) not only produces the desired term in (4.17), but also a term $\lambda_{00kl}\lambda_{kl}{}^*$.

We may try to solve this integrability problem by generalizing (4.16) to

$$\begin{aligned}\Delta S_1 = \tfrac{1}{3}\lambda_{ijk}{}^*[a_0\lambda_{ijk0}d &+ \tfrac{2}{3}a_1(\lambda_{ij0k} - \tfrac{1}{2}\lambda_{jk0i}(-)^{i(j+k)}\\ &- \tfrac{1}{2}\lambda_{ki0j}(-)^{k(i+j)})d\\ &+ \tfrac{2}{3}a_2(\lambda_{k0ij}(-)^{k(i+j)} - \tfrac{1}{2}\lambda_{i0jk}\\ &- \tfrac{1}{2}\lambda_{j0ki}(-)^{i(j+k)})d] + \lambda_{ijkl}{}^*\text{-terms},\end{aligned} \tag{4.18}$$

where we allow for the spinor index of d to contract in all possible ways. It is easy to verify that integrability of terms of the form (4.17) requires $a_0 - a_1 + a_2 = 0$. However, upon recalculating the $\lambda_{ijk}{}^*$ terms in (S_1,S_1) we find that, after all, $a_0 = 1$, $a_1 = a_2 = 0$.

The extension to arbitrarily many spinor indices proceeds in a similar way. However, since the aforementioned integrability problem precludes a solution satisfying $(S,S) = 0$, this analysis comes to a dead end and other approaches to the quantization of the superparticle should be followed.

ACKNOWLEDGMENTS

We are grateful to Ingemar Bengtsson for drawing our attention to Refs. 17 and 18. M.R. and U.L. are happy to thank the ITP's at Stockholm and Stony Brook for hospitality.

The work of M.R., N.S., and P.v.N. was supported in part by National Science Foundation Grant No. PHY 89-08495. The work of A.E.v.d.V. was supported in part by National Science Foundation Grant No. PHY 87-15995.

[1]L. Brink and J. H. Schwarz, Phys. Lett **100**, B, 310 (1981).

[2]U. Lindström, M. Roček, W. Siegel, P. van Nieuwenhuizen, and A. E. van de Ven, Phys. Lett. B **224**, 285 (1989).

[3]W. Siegel, "Lorentz-covariant gauges for Green–Schwarz superstrings," Stony Brook preprint ITP-SB-89-19, March 1989; talk presented at Strings '89, College Station, TX 13–18 March 1989.

[4]S. J. Gates, Jr., M. T. Grisaru, U. Lindström, M. Roček, W. Siegel, P. van Nieuwenhuizen, and A. E. van de Ven, Phys. Lett B **225**, 44 (1989).

[5]U. Lindström, M. Rŏcek, W. Siegel, P. van Nieuwenhuizen, and A. E. van de Ven, "Construction of the covariantly quantized heterotic superstring," Stony Brook preprint ITP-SB-89-38, May 1989.

[6]M. B. Green and C. H. Hull, Phys. Lett. B **225**, 57 (1989).

[7]R. Kallosh, Phys. Lett. B **224**, 273 (1989).

[8]D. A. Depireux, S. J. Gates, Jr., P. Majumdar, B. Radak, and S. Vashakidze, "Towards covariantly quantized $D = 10$ type II Green–Schwarz σ-models," Maryland preprint UMDEPP 89–169, May 1989.

[9]R. E. Kallosh, Phys. Lett. B **225**, 49 (1989).

[10]M. Roček, W. Siegel, P. van Nieuwenhuizen, and A. E. van de Ven, Phys. Lett. B **227**, 87 (1989).

[11]U. Lindström, M. Roček, W. Siegel, P. van Nieuwenhuizen, and A. E. van de Ven, Phys. Lett B **228**, 53 (1989).

[12]I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **102**, 27 (1981); **120**, 166 (1983); Phys. Rev. D **28**, 2567 (1983); **30**, 508 (1984); Nucl. Phys. B **234**, 106 (1984); J. Math. Phys. **26**, 172 (1985).

[13]T. Curtright (private communication); J. Thierry-Mieg, "BRS structure of antisymmetric tensor gauge fields," Harvard preprint HUTMP-79/B85, 1980; W. Siegel, Phys. Lett. B **93**, 170 (1980).

[14]M. B. Green and J. H. Schwarz, Nucl. Phys. B **243**, 285 (1984); I. Bengtsson and M. Cederwall, "Covariant superstrings do not admit covariant gauge fixing," Göteborg preprint 84–21, June 1984; T. Hori and K. Kamimura, Prog. Theora. Phys. **73**, 476 (1985).

[15]F. Bastianelli, G. W. Delius, and E. Laenen (private communication).

[16]J. M. L. Fisch and M. Henneaux, "A note on the covariant BRST quantization of the superparticle," Université Libre de Bruxelles preprint ULB TH2/89-04, 1989.

[17]W. Siegel, Nucl. Phys. B **263**, 93 (1986).

[18]W. Siegel, Phys. Lett. B **203**, 79 (1988); I Bengtsson, Phys. Rev. D **39**, 1158 (1989).

[19]I. A. Batalin, R. E. Kallosh, and A. Van Proeyen, "Symmetries of superparticle and superstring action," in *Quantum Gravity*, edited by M. A. Markov, V. A. Berezin, and V. P. Frolov (World Scientific, Singapore, 1988); R. Kallosh, W. Troost and A. Van Proeyen, Phys. Lett. B **212**, 428 (1988).

[20]F. Bastianelli, G. W. Delius, and E. Laenen, "Free quantum action and BRST-charge for the superparticle," Stony Brook preprint ITP-SB-89-51 July 1989.

[21]W. Siegel, Class. Quant. Grav. **2**, L95 (1985).

[22]A. R. Miković and W. Siegel, Phys. Lett. B **209**, Lett. 47 (1988).

Some symmetries of quantum dimensions

Jürgen Fuchs
NIKHEF-H, Kruislaan 409, NL-1098 SJ Amsterdam, The Netherlands

Peter van Driel
Institute for Theoretical Physics, University of Amsterdam, Valckenierstraat 65, NL-1018 XE Amsterdam, The Netherlands

(Received 19 September 1989; accepted for publication 28 February 1990)

Several useful properties of the quantum dimensions of representations of $\mathscr{U}_q(\mathrm{sl}(n))$ for q a root of unity are established. The results can be used to calculate truncated Kronecker products of these representations and hence to investigate the isomorphism between these truncated tensor products and the fusion rules of the WZW theories associated with $\mathscr{U}_q(\mathrm{sl}(n))$.

I. INTRODUCTION

It has recently become more and more clear that the relevance of quantum groups to physics is not restricted to the theory of integrable systems, but that they play a major role in two-dimensional conformal field theory, too.

By a quantum group (sometimes, a more general definition of the term quantum group is used), we mean a quasitriangular Yang–Baxter algebra, i.e., (Refs. 1–5) a Hopf algebra $\mathscr{A}$ (i.e., an associative algebra with unit endowed with the operations of comultiplication, antipode, and counit) with the additional property (called quasitriangularity[3,5]) that the comultiplication Δ and the composition $\pi\circ\Delta$ of Δ with the permutation π in $\mathscr{A}\otimes\mathscr{A}$ are related by conjugation by a matrix $\mathscr{R}\in\mathscr{A}\otimes\mathscr{A}$ obeying certain restrictions (in particular it follows that $\mathscr{R}$ satisfies the Yang–Baxter equation). It is expected that to any rational (and perhaps also to nonrational) two-dimensional conformal field theory, one can associate a quantum group, and that most, if not all, of the properties of the conformal field theory are consequences of the underlying quantum group structure. The most direct connection would certainly be to express the generators of the quantum group in terms of the fields belonging to the chiral algebra of the conformal field theory (and vice versa), but so far no such construction is known. However, more indirect connections that are immediate consequences of such a construction have already been uncovered. First, the matrices describing the braiding and fusing of the chiral block functions (which are the building blocks of the partition function and correlation functions of the conformal field theory) have been identified[6–9] with quantum group analogs of Wigner–Racah coefficients. Second, arguments have been given[8,10,11] that the fusion rules of the conformal field theory are isomorphic to appropriately defined tensor products of quantum group representations; in particular, for the quantum group underlying a rational conformal field theory, it should be possible to identify such a tensor product realized on a finite set of "good" representations. (In contrast, so far there is no hint on how to interprete in terms of the quantum group the normalization factors that are needed[12,9] to compute the structure constants of the full operator algebra and hence to solve the conformal field theory completely.)

In the present paper, we provide additional support for the connection between fusion rules and tensor products of quantum group representations. We consider the case where the conformal field theory is the Wess–Zumino–Witten theory[13] on a simply connected Lie group manifold G; the underlying quantum group $\mathscr{U}_q(g)$ is described in Sec. II (g is the Lie algebra of G, and q is related via $q^{k+h}=1$ to the dual Coxeter number h of g and to the level k of the affine Kac–Moody algebra $g^{(1)}$, which generates the chiral algebra of the WZW theory). In Sec. III, we discuss the Kronecker tensor products of $\mathscr{U}_q(g)$ for q a root of unity; in general, representations that are not fully reducible appear; it is however possible to define a truncated version of the Kronecker product that acts on a restricted set of irreducible representations. We then derive several useful symmetry properties of the so-called quantum dimensions of the representation of $\mathscr{U}_q(\mathrm{sl}(n))$ [Sec. IV, Eqs. (8), (10), (12), and (14)]. The corresponding properties of WZW fusion rules are described in Sec. V. In the course of proving these symmetry relations, we encountered a peculiar correspondence between representations of $\mathscr{U}_q(\mathrm{sl}(n))$ and $\mathscr{U}_q(\mathrm{sl}(k))$ for q an $(n+k)$th root of unity; this correspondence is explored further in Sec. VI.

II. QUANTUM DIMENSIONS

To any simple or affine Lie algebra g, one can associate a quantum group as the q-deformation $\mathscr{U}_q(g)$ of its universal enveloping algebra by defining generators and relations as follows.[1,2,4] For each simple root $\alpha^{(i)}=1,\dots,r$, of g one has three generators H_i, E_i^+, E_i^-; they satisfy the commutation relations

$$[H_i,H_j]=0,$$
$$[H_i,E_j^\pm]=\pm A^{ij}E_j^\pm,$$
$$[E_i^+,E_j^-]=\delta_{ij}[H_i],$$

as well as

$$\sum_{m=0}^{1-A^{ij}}(-1)^m\begin{bmatrix}1-A^{ij}\\ m\end{bmatrix}_i (E_i^\pm)^m E_j^\pm (E_i^\pm)^{1-A_{ij}-m}=0,\qquad \text{for } i\neq j.$$

Here, A is the Cartan matrix of g, and we have introduced the notations

 0022-2488/90/071770-06$03.00

$$[x] \equiv [x]_q = (q^{x/2} - q^{-x/2})/(q^{1/2} - q^{-1/2}),$$
$$\begin{bmatrix} n \\ m \end{bmatrix} = \frac{[n]!}{[m]![n-m]!},$$
$$[n]! = \prod_{m=1}^{n} [m],$$

and also $[x]_i \equiv [x]_{q_i}$ with $q_i = q^{(\alpha^{(i)},\alpha^{(i)})/2}$. (In the "classical limit" $q \to 1$, the above relations just become the commutation and Serre relations for a Chevalley basis of g.) For generic values of q, the representation theory of $\mathscr{U}_q(g)$ is completely parallel to that of g (Refs. 2, 5, 14–17). Irreducible finite-dimensional representations are highest-weight representations with dominant integral highest weight $\Lambda = \Sigma_{i=1}^{r} \lambda^i \Lambda_{(i)}, \lambda^i \in \mathbf{Z}_{\geq 0}$ (the $\Lambda_{(i)}$ denote the fundamental weights of g). Also, the usual Kronecker tensor product $R \times R'$ of two irreducible representations is fully reducible, and (due to the quasitriangularity of $\mathscr{A}$) $R \times R'$ and $R' \times R$ are isomorphic.

As in the case of g, one can define, for any representation R of $\mathscr{U}_q(g)$ and any nonnegative integer N, a trace on the centralizer of $\mathscr{U}_q(g)$ in the N-fold tensor product $R^{\otimes N}$, and from this a Markov trace on the braid group B_N (Refs. 4, 5, 18) [for $g = \mathrm{sl}(n)$ and R the defining representation, the centralizer algebra is the Hecke algebra (of type A_{N-1}) $H_N(q)$, and the trace is the Ocneanu[19] trace.[4,20]]. This can be extended to the formal union of the centralizer algebras for arbitrary N, and hence to $\cup_N B_N$. Evaluating this trace on the identity element of the braid group (and leaving out a normalization factor that is a certain power of $[n]$) one gets for any irreducible representation of $\mathscr{U}_q(g)$ a number called the *statistical dimension*[18] or *quantum dimension* of R. From the properties of the Markov trace, it follows[18] that these quantum dimensions obey the usual sum rule for tensor products: If the tensor product of two irreducible representations R and R' is written as a sum of irreducible representations as

$$R \times R' = \bigoplus_i R_i,$$

then the quantum dimensions $\mathscr{D}$ obey

$$\mathscr{D}(R) \cdot \mathscr{D}(R') = \sum_i \mathscr{D}(R_i). \tag{1}$$

From now on, we specialize to the case $g = \mathrm{sl}(n)$ unless noted otherwise. Then the quantum dimension of an irreducible finite-dimensional representation with highest weight Λ is given by the formula[4]

$$\mathscr{D}^{(n)}(\Lambda) = \prod_{(i,j) \in Y_\Lambda} \frac{[n-i+j]}{[h_{ij}]}. \tag{2}$$

Here, Y_Λ is the Young diagram of the representation; (i,j) labels the boxes of Y_Λ, with i counting rows from top to bottom and j counting columns from left to right; finally, h_{ij} is the length of the hook with corner (i,j) in Y_Λ. Explicitly, we have, for example,

$$\mathscr{D}^{(n)}(\Lambda_{(l)}) = \begin{bmatrix} n \\ l \end{bmatrix} \quad \text{for } l = 1,\ldots,r,$$

and

$$\mathscr{D}^{(n)}(l \cdot \Lambda^{(1)}) = \begin{bmatrix} n+l-1 \\ l \end{bmatrix} \quad \text{for } l \in \mathbf{Z}_{>0}$$

for the fundamental representations and for the symmetric tensor representations, respectively. Also, in the classical limit $q \to 1$, the quantum dimension just becomes the ordinary dimension.

III. THE TRUNCATED TENSOR PRODUCT

So far, q just played the role of a formal expansion parameter. When q is interpreted as a complex number, it turns out that for q a root of unity the representation theory is markedly different from the generic case.[21–23] The basic observation[20,24] is that, while for generic q the centralizer of $\mathscr{U}_q(g)$ in $R^{\otimes N}$ is semisimple, this ceases to be true for q a root of unity. Now the representations of the centralizer algebra determine the decomposition of the Kronecker product $R \times R'$; as a consequence the Kronecker product of irreducible representations is fully reducible for generic q, but in general not fully reducible for q a root of unity. (Consider, e.g., the case of the defining representation of $\mathscr{U}_q(\mathrm{sl}(n))$ where the centralizer is the Hecke algebra $H_N(q)$; for $q = 1$ $H_N(q)$ is isomorphic to the semisimple group algebra of the symmetric group S_N; when q is deformed away from 1, the representation theory remains isomorphic to that of S_N [and hence the Kronecker products of $\mathscr{U}_q(\mathrm{sl}(n))$ are isomorphic to those of $\mathrm{sl}(n)$] as long as q is not a root of unity[20,24].)

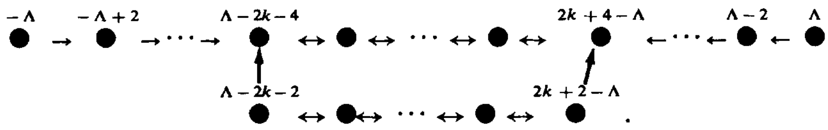
Suppose that R is a non-fully reducible representation (not containing any irreducible part) for $q = q_0$ a root of unity. If q is deformed away from q_0, R decomposes into the direct sum of irreducible representations. In fact, inspection shows that exactly two irreducible representations are paired up for $q = q_0$; at least one of them obeys $(\Lambda,\theta) > k$, where θ is the highest root of g and k is the minimal positive integer such that $q_0^{k+h} = 1$ for h the dual Coxeter number of g. In terms of the generators of $\mathscr{U}_q(g)$, this mixing of representations occurs because at $q = q_0$ the step operators $E_i^{\pm}$ are nilpotent, $(E_i^{\pm})^{k+h} = 0$, so that the representation contains states that cannot be reached from the highest weight state, and, in addition, states that are new highest (or lowest) weight states; the latter states have zero norm, i.e., are null states (the norm is defined as for the corresponding irreducible representations of g; also, q must be treated as a formal parameter, i.e., is not changed under complex conjugation).

As an example consider the case $g = \mathrm{sl}(2)$. For generic q, the irreducible finite-dimensional representations are characterized by a single nonnegative integer Λ (twice the spin). The dimension of such a representation R_Λ is $\Lambda + 1$, and the states in R_Λ are connected by the step operators $E^{\pm}$ schematically as

$$\overset{-\Lambda}{\bullet} \leftrightarrow \overset{-\Lambda+2}{\bullet} \leftrightarrow \cdots\cdots \leftrightarrow \overset{\Lambda-2}{\bullet} \leftrightarrow \overset{\Lambda}{\bullet}.$$

More precisely, application of $E^{\pm}$ to a state of weight μ gives the state with weight $\mu \pm 2$, but only up to some factors $[m_i], m_i \in \mathbf{Z}$. Now for $q^{k+2} = 1$, due to $[k+2] = 0$, $E^{\pm}$

annihilates some states that would not be annihilated for generic q. As a consequence, representations with highest weight $k+1<\Lambda<2k+2$ that appear in a tensor product of representations with highest weight $\leqslant k$ contain $\Lambda-k-1$ highest weight null states and pair up with would-be irreducible representations of highest weight $2k+2-\Lambda$ to form a reducible, but not fully reducible representation according to the scheme



(For $\Lambda>2k+2$, Λ not a multiple of $k+1$, the situation is similar; for Λ a multiple of $k+1$, the representation is still irreducible.)

The definition of quantum dimension can be extended to the case of reducible, but not fully reducible representations by analytic continuation. Namely, if R is not fully reducible, but is the deformation for q tending to a root q_0 of unity of the direct sum of two irreducible representations $R_{\Lambda_1}, R_{\Lambda_2}$, then we define

$$\mathscr{D}(R)=(\mathscr{D}(\Lambda_1)+\mathscr{D}(\Lambda_2))|_{q=q_0}.$$

It turns out (see also the next section) that for this type of representation the quantum dimensions of R_{Λ_1} and R_{Λ_2} become in fact equal up to a sign at $q=q_0$, and hence $\mathscr{D}(R)=0$. In addition, there exist also irreducible representations for which $\mathscr{D}(R)=0$; among these, there are in particular all representations with highest weight Λ such that $(\Lambda,\theta)=k+1$.

By analytic continuation, the dimension sum rule (1) holds also for non-fully reducible representations. The fact that the non-fully reducible representations all have zero quantum dimension suggests that—together with the irreducible representations that also possess zero quantum dimension—they form an ideal in the category of $\mathscr{U}_q(g)$ representations; according to Ref. 25 this is indeed the case. This observation implies that the following *truncated tensor product* "$\star$" is well defined:

$$R\star R'=(R\times R')|_{\text{positive part}}, \tag{3}$$

where the right-hand side denotes those irreducible representations in $R\times R'$ that have positive quantum dimension.

In terms of the centralizer algebras, this truncation corresponds to quotienting the centralizer by its Abelian ideals.[20,24] In short, (3) defines a well-behaved tensor product on the set of irreducible representations with positive quantum dimension; these representations are precisely the highest weight representations R_Λ with $(\Lambda,\theta)\leqslant k$.

IV. IDENTITIES FOR QUANTUM DIMENSIONS

Returning to the case $g=\mathrm{sl}(n)$, from now on we assume that $q^{k+n}=1$. Then, we have the identities

$$[x]=[x+2(k+n)]=[k+n-x]. \tag{4}$$

Using the fact that the length of the ith row of the Young diagram Y_Λ equals $\Sigma_{j=i}^{n-1}\lambda^j$ (for $\Lambda=\Sigma_{i=1}^{n-1}\lambda^i\Lambda_{(i)}$), it then follows from the formula (2) that the quantum dimension of a (highest weight or would-be highest weight) representation is a positive real number as long as

$$\lambda^0\equiv k-\sum_{i=1}^{n-1}\lambda^i \tag{5}$$

is non-negative while for negative λ^0 the quantum dimension can also be zero or negative. Moreover, in accordance with the discussion in the previous section, it turns out that many distinct representations lead to the same (absolute) value of the quantum dimension. The purpose of this section is to analyze this situation in more detail.

First, consider the case $n=2$. Then the quantum dimension of the (irreducible or would-be irreducible) representation R_Λ is

$$\mathscr{D}^{(2,k)}(\Lambda)=[\Lambda+1].$$

Employing (4) for $n=2$, it is easy to see that

$$\mathscr{D}^{(2,k)}(k-\Lambda)=\mathscr{D}^{(2,k)}(\Lambda) \tag{6}$$

and

$$\mathscr{D}^{(2,k)}(2k+2-\Lambda)=-\mathscr{D}^{(2,k)}(\Lambda) \tag{7}$$

[In particular it follows that $\mathscr{D}^{(2,k)}(k+2+\Lambda)=-\mathscr{D}^{(2,k)}(\Lambda)$ and $\mathscr{D}^{(2,k)}(k+1)=0$.] (These equations hold for arbitrary real Λ, but of course only for Λ a non-negative integer [and $\Lambda\leqslant k$ in the case of (6)] they actually relate quantum dimensions of finite-dimensional representation of $\mathscr{U}_q(\mathrm{sl}(2))$.)

It is not too difficult to derive analogs of (6) and (7) for arbitrary n. We find that (6) generalizes to

$$\mathscr{D}^{(n,k)}(\sigma(\Lambda))=\mathscr{D}^{(n,k)}(\Lambda), \tag{8}$$

where the highest weight $\sigma(\Lambda)$ is obtained from Λ as

$$\sigma(\Lambda)=\sum_{i=1}^{n-1}\lambda^{i-1}\Lambda_{(i)}, \tag{9}$$

with λ^i, $i=1,\dots,n-1$, the Dynkin labels of Λ and λ^0 as in (5). [In order that $\sigma(\Lambda)$ is dominant, we have to restrict here to representations with $\lambda^0\geqslant 0$.] Similarly, (7) gets replaced by

$$\mathscr{D}^{(n,k)}(\Lambda+(\lambda^0+1)\theta)=-\mathscr{D}^{(n,k)}(\Lambda), \tag{10}$$

where $\theta\equiv\Lambda_{(1)}+\Lambda_{(n-1)}$ is the highest root of $\mathrm{sl}(n)$.

As an example for (8), if we take R to be the singlet representation, the j-fold application of σ gives

$$\mathscr{D}^{(n,k)}(k\cdot\Lambda_{(j)})=1,$$

for $j=1,\dots,n-1$. An example for (10) is

$$\mathscr{D}^{(n,k)}(k\cdot\theta) = -\mathscr{D}^{(n,k)}(\theta) = -[k-1][k+1].$$

It also follows from (10) that $\mathscr{D}^{(n,k)}(\Lambda) = 0$ if $\lambda^0 = -1$, i.e., if the Young diagram Y_Λ has exactly $k+1$ columns.

To prove the relations (8) and (10) [and also (12) and (14) below], one simply takes the dimension formula (2) and rearranges the various factors using in particular the identities (4). This is a rather lengthy, but straightforward exercise. In the case of (8), the calculation can be simplified by observing that the quantum dimension does not change if the Young diagram under consideration is reflected along the diagonal (so that the meaning of rows and columns is exchanged) and if at the same time n and k are interchanged, i.e., that

$$\mathscr{D}^{(n,k)}(\Lambda) = \mathscr{D}^{(k,n)}(\delta(\Lambda)), \tag{11}$$

where δ denotes the transformation resulting in the reflection of Y_Λ. The identity (11) holds because $[n-i+j] = [k-j+i]$ and because the hook lengths h_{ij} are invariant under the reflection δ. Applying (11) to both sides of (8) (and interpreting σ diagrammatically, see below), we arrive at the assertion that the quantum dimension does not change if a column of length k is added to a Young diagram for $\mathrm{sl}(k)$, a statement that follows rather easily from (2).

Actually, there is also a second possibility to generalize (6), namely,

$$\mathscr{D}^{(n,k)}(\omega(\Lambda)) = \mathscr{D}^{(n,k)}(\Lambda), \tag{12}$$

with

$$\omega(\Lambda) = \sum_{i=1}^{n-1} \lambda^{n-i-1}\Lambda_{(i)} \tag{13}$$

(here, we assume again $\lambda^0 \geqslant 0$). However, this in fact does not provide much new information because $\omega\circ\sigma(\Lambda) = \Sigma_{i=1}^{n-1}\lambda^{n-i}\Lambda_{(i)} = \overline{\Lambda}$ is the weight conjugate to Λ, and it is no surprise that conjugate representations possess identical quantum dimensions.

The operations σ and ω can be understood diagrammatically as follows. In terms of Young diagrams, σ correspond to adding a row of length k to the diagram (due to $\lambda_0 \geqslant 0$, this becomes the first row of the new diagram $Y_{\sigma(\Lambda)}$) and deleting columns of length n (if present), while the Young diagrams Y_Λ and $Y_{\omega(\Lambda)}$ (with one of them turned upside down) add up to a rectangular diagram with k columns of length $n-1$. There is also an interpretation in terms of Dynkin diagrams: σ and ω correspond to automorphisms of the extended Dynkin diagram of $\mathrm{sl}(n)$, namely, to the primitive rotation (that maps the ith node in the diagram to the $i+1$th node for $i = 0,...,n-1$ defined mod n), and to the reflection of the diagram which exchanges the 0th and $n-1$th nodes, respectively. (The automorphism group of the extended Dynkin diagram is the group generated by these two transformations, subject to the constraints $\sigma^n = \omega^2 = \omega\sigma\omega\sigma = 1$.)

The interpretation in terms of the extended Dynkin diagram of $\mathrm{sl}(n)$ makes it clear that (8) and (12) are the only independent generalizations of (6) to general values of n. It is tempting to search for a similar symmetry principle lying behind the possible extensions of (7). In fact, while (10) is the most natural way to generalize (7), there are many other ways to do so. For example, one has

$$\mathscr{D}^{(n,k)}(\sigma_i(\Lambda)) = -\mathscr{D}^{(n,k)}(\Lambda), \tag{14}$$

where $\sigma_i(\Lambda) = \Sigma_j \lambda^j_{\sigma_i}\Lambda_{(j)}$ has Dynkin labels

$$\lambda^j_{\sigma_i} = \begin{cases} 2(k+n-1) - \lambda^j, & \text{for } j = i, \\ \lambda^i + \lambda^j - k - n + 1, & \text{for } |j-i| = 1, \\ \lambda^j, & \text{for } |j-i| \geqslant 2. \end{cases}$$

The transformations (10) and (14) are special cases of modified Weyl reflections. The most general transformation of this kind has been found in Ref. 23, using the q-analog of the Weyl dimension formula in order to define the quantum dimension, i.e.,

$$\mathscr{D}^{(n,k)}(\Lambda) = \prod_{\alpha>0} \frac{[(\Lambda+\rho,\alpha)]}{[(\rho,\alpha)]}, \tag{15}$$

with $\alpha > 0$ denoting the positive roots, and $2\rho = \Sigma_{\alpha>0}\alpha$. [It is, of course, highly plausible that (15) is equivalent to (2) for $g = \mathrm{sl}(n)$, but to the best of our knowledge this has not yet been proven rigorously.] Namely,

$$\mathscr{D}^{(n,k)}(\sigma_{\alpha,\beta}(\Lambda)) = \epsilon(\alpha)\mathscr{D}^{(n,k)}(\Lambda), \tag{16}$$

for

$$\sigma_{\alpha,\beta}(\Lambda) = \bar\sigma_\alpha(\Lambda+\rho) - \rho + (k+n)\beta.$$

Here, $\bar\sigma_\alpha$ is the Weyl reflection with respect to an arbitrary root α, β is an arbitrary vector in the coroot lattice, and ϵ denotes the homomorphism from the Weyl group to $\{1, -1\}$, i.e., $\epsilon(\alpha) = (-1)^{l(\alpha)}$, where $l(\alpha)$ is the length of $\bar\sigma_\alpha$. Equations (10) and (14) follow from (16) by setting $\alpha = \beta = \alpha^{(i)}$ and $\alpha = \beta = \theta$, respectively.

To prove (16), observe that any Weyl reflection permutes the positive roots up to possibly sign factors, with the number of minus signs given by $l(\alpha)$ (Ref. 26); thus (15) changes only by a factor of $\epsilon(\alpha)$ if $\Lambda+\rho$ is replaced by $\bar\sigma_\alpha(\Lambda+\rho)$. Moreover, because of (4) the addition of $(k+n)\beta$ to Λ changes $[(\Lambda+\rho,\alpha)]$ only by a sign $(-1)^{(\beta,\alpha)}$, and hence (15) by $\Pi_{\alpha>0}(-1)^{(\beta,\alpha)} = (-1)^{2(\beta,\rho)} = 1$ for any β in the coroot lattice. Note that this proof works for arbitrary roots α and coroot lattice elements β, but that for generic choices of α, β the formula (15) is actually not a relation between quantum dimensions because for Λ dominant integral, $\sigma_{\alpha,\beta}(\Lambda)$ typically is not a dominant weight.

The explicit formula for $\sigma_{\alpha,\beta}$ also shows that the relation to the Weyl group is rather indirect. Therefore, even though we do not know any counterexample, it is not clear to us whether (16) [together with (8), (10), and (12)] is sufficient to obtain *all* representations possessing the same (up to sign) quantum dimension as some given representation.

V. WZW FUSION RULES

As already mentioned in the Introduction, it is widely expected that the truncated tensor products of $\mathscr{U}_q(g)$ are isomorphic to the fusion rules of the corresponding WZW theories. The WZW fusion rules can in principle be obtained from the Kronecker products of g with the help of the so-called depth rule,[13] although in practice it can be rather tedious to apply this rule. However, some general results can be obtained by using directly the Kac–Moody null vector equation which is the source[13] of the depth rule. Namely,[27,28] for any automorphism η of the extended Dynkin diagram of g an action on the set of primary WZW fields ϕ can be defined under which the fusion rules

$$\phi * \phi' = \bigoplus_i \phi_i$$

transform covariantly. More precisely, if ϕ carries the representation R_Λ of g, then $\eta(\phi)$ is the primary field carrying the representation $R_{\eta(\Lambda)}$, and the action on the fusion rules is determined by $\eta(\phi \oplus \phi') = \eta(\phi) + \eta(\phi')$ together with

$$\eta(\phi * \phi') = \begin{cases} \eta(\phi) * \eta(\phi'), & \text{if } \eta \in \Gamma, \\ \eta(\phi) * \phi' = \phi * \eta(\phi'), & \text{if } \eta \in \hat{\Gamma}/\Gamma, \end{cases}$$

where Γ and $\hat{\Gamma}$ are the groups of automorphisms of the Dynkin diagram and extended Dynkin diagram of g, respectively.

A nontrivial necessary condition for the isomorphism between WZW fusion rules and truncated tensor products is therefore that any automorphism $\eta \in \hat{\Gamma}$ acts analogously on the quantum dimensions, i.e., that, for $R_\Lambda * R_{\Lambda'} = \oplus_i R_{\Lambda_i}$, one has

$$\mathscr{D}\Big(\bigoplus_i R_{\eta(\Lambda_i)}\Big) = \begin{cases} \mathscr{D}(R_{\eta(\Lambda)}) \cdot \mathscr{D}(R_{\eta(\Lambda')}), & \text{if } \eta \in \Gamma, \\ \mathscr{D}(R_\Lambda) \cdot \mathscr{D}(R_{\eta(\Lambda')}) = \mathscr{D}(R_{\eta(\Lambda)}) \cdot \mathscr{D}(R_{\Lambda'}), & \text{if } \eta \in \hat{\Gamma}/\Gamma. \end{cases}$$

For $g = \mathrm{sl}(n)$, Γ is generated by $\omega \circ \sigma$ while $\hat{\Gamma}/\Gamma$ is generated by σ. Now, as we have shown in the previous section, $\mathscr{D}(\eta(\Lambda)) = \mathscr{D}(\Lambda)$ for all $\eta \in \hat{\Gamma}$. Moreover, since the truncation of the tensor products only removes representations of zero quantum dimension, the q-dimension sum rule (1) also holds for truncated tensor products, and $\mathscr{D}(\Lambda * \Lambda') = \mathscr{D}(\Lambda) \cdot \mathscr{D}(\Lambda')$. Together, it then follows that the condition is indeed met. Thus the symmetry properties (8) and (12) of quantum dimensions strongly support the conjecture that WZW fusion rules and truncated tensor products are isomorphic.

Note that the primitive rotation σ(9) changes the conjugacy class $\mathscr{C}(\Lambda) \in \mathbf{Z}_n$ of a highest weight representation by k mod n. As a consequence, by applying $\eta = \sigma^m$ with appropriate power m to $R * R'$, one can achieve that $\eta(R * R')$ belongs to some fixed (say, the trivial) conjugacy class. Thus the covariance of the fusion rules under η can be used to reduce the number of fusion rules of the SU(n) WZW theory to a set of independent ones that is smaller by a factor of n. [For k a multiple of n, σ does not change the conjugacy class; nevertheless inspection shows that also in this case the number of independent fusion rules is smaller by almost a factor of n than the total number of fusion rules. Also note that as a consequence of the $k-n$ duality (11), it is no loss of generality to take $k \leqslant n$ so that only the case $k = n$ is exceptional.]

VI. $k-n$ DUALITY

It is already apparent from the relation $q^{n+k} = 1$ that in some sense the roles played by k and n should be interchangeable. This is made precise by formula (11). Via (11), we can relate restricted Kronecker products of $\mathscr{U}_q(\mathrm{sl}(n))$ and $\mathscr{U}_q(\mathrm{sl}(k))$ at the same value of q [and hence, assuming the validity of the conjecture that truncated tensor products and WZW fusion rules are isomorphic, relate the fusion rules of the SU(n) level k WZW theory to the SU(k) level n theory]. It is clear that this does not provide an isomorphism of the respective algebras of truncated tensor products, and inspection shows that it is not even a homomorphism. Nevertheless, interesting observations can be made. In particular, the relation can be employed to simplify the explicit calculation of truncated tensor products.

Of particular interest is the case $k = 2$ because the fusion rules of the SU(2) WZW theory are already known[13] for arbitrary level. Under the action of δ, a representation of $\mathscr{U}_q(\mathrm{sl}(n))$ with highest weight $\Lambda = \Lambda_{(i)} + \Lambda_{(j)}$ (with $0 \leqslant i \leqslant j \leqslant n-1$ and $\Lambda_{(0)} := 0$) is mapped to the representation of $\mathscr{U}_q(\mathrm{sl}(2))$ having highest weight $j - i$; thus we learn that

$$\mathscr{D}^{(n,2)}(\Lambda_{(i)} + \Lambda_{(j)}) = [j - i + 1]. \tag{17}$$

Moreover, in terms of quantum dimensions the truncated tensor products of $\mathscr{U}_q(\mathrm{sl}(2))$ just read

$$[l] \cdot [m] = [|l-m| + 1] + [|l-m| + 3] + \cdots + [l + m - 1] \tag{18}$$

(provided that $l, m \leqslant (k+1)/2$; this can always be arranged by making use of $[k + 2 - x] = [x]$). But this means that we can rather easily write down all restricted tensor products of $\mathscr{U}_q(\mathrm{sl}(n))$ for $q^{n+2} = 1$.

As an example, consider the product of the representation R_Λ of highest weight $\Lambda = \Lambda_{(1)} + \Lambda_{(n-2)}$ and of its conjugate representation. The ordinary sl(n) Kronecker product reads (for $n \geqslant 6$)

$$\begin{aligned} R_\Lambda \times R_{\bar{\Lambda}} = {} & \mathbf{1} + 2R_\theta + 2R_{\Lambda_{(2)} + \Lambda_{(n-2)}} + R_{\Lambda_{(3)} + \Lambda_{(n-3)}} \\ & + (R_{2\Lambda_{(1)} + \Lambda_{(n-2)}} + \text{c.r.}) \\ & + (R_{\Lambda_{(1)} + \Lambda_{(2)} + \Lambda_{(n-3)}} + \text{c.r.}) \\ & + R_{2\theta} + R_{\Lambda + \bar{\Lambda}}, \end{aligned}$$

where c.r. stands for the conjugate representation. Removing representations with $\lambda^0 < 0$ from this product, one is left with the representations in the first line; for the quantum dimension of these representations, (17) gives

$$\mathscr{D}^{(n,2)}(\Lambda) = [4], \quad \mathscr{D}^{(n,2)}(\theta) = [3],$$
$$\mathscr{D}^{(n,2)}(\Lambda_{(2)} + \Lambda_{(n-2)}) = [5],$$
$$\mathscr{D}^{(n,2)}(\Lambda_{(3)} + \Lambda_{(n-3)}) = [7].$$

Now, from (18) we learn that the quantum dimension sum rule for the correct truncated tensor product should read

$[4]\cdot[4] = [1] + [3] + [5] + [7]$; there is only one possibility to achieve this, namely, to also remove one of the adjoint representations and one of the representations with highest weight $\Lambda_{(2)} + \Lambda_{(n-2)}$, i.e., the truncated tensor product reads

$$R_\Lambda \star R_{\bar{\Lambda}} = \mathbf{1} + R_\theta + R_{\Lambda_{(2)} + \Lambda_{(n-2)}} + R_{\Lambda_{(3)} + \Lambda_{(n-3)}}.$$

In general, one would not expect that the restriction (18) (or similar restrictions obtained by considering other values of k) on the quantum dimensions is already sufficient to determine uniquely which of the irreducible representations with $\lambda^0 \geqslant 0$ have to be removed from the $\mathrm{sl}(n)$ Kronecker product in order to arrive at the truncated tensor product. However, the recipe does give unique results for the tensor products of representations possessing Young diagrams with sufficiently low number of boxes (and, in fact, for all cases we checked).

[1]V. G. Drinfeld, Sov. Math. Dokl. **32**, 254 (1985).

[2]M. Jimbo, Lett. Math. Phys. **10**, 254 (1985); **11**, 247 (1986).

[3]V. G. Drinfeld, in *Proceedings of the International Congress of Mathematicians* (American Mathematical Society, Berkeley, CA, 1986), p. 798.

[4]N. Yu. Reshetikhin, preprints LOMI E-4-87 and E-17-87.

[5]A. N. Kirillov and N. Yu. Reshetikhin, preprint LOMI E-9-88.

[6]G. Moore and G. Seiberg, Commun. Math. Phys. **123**, 177 (1989).

[7]L. Alvarez-Gaumé, C. Gomez, and G. Sierra, Nucl. Phys. B. **319**, 155 (1989).

[8]L. Alvarez-Gaumé, C. Gomez, and G. Sierra, Phys. Lett. B **220**, 142 (1989).

[9]G. Felder, J. Fröhlich, and G. Keller, IAS Princeton/Zürich preprint (January, 1989).

[10]G. Moore and N. Yu. Reshetikhin, Nucl. Phys. B. **328**, 557 (1990).

[11]L. Alvarez-Gaumé, C. Gomez, and G. Sierra, Nucl. Phys. B **330**, 347 (1990).

[12]Vl. S. Dotsenko and V. A. Fateev, Phys. Lett. B **154**, 291 (1985); J. Fuchs and A. Klemm, Ann. Phys. **194**, 303 (1989).

[13]D. Gepner and E. Witten, Nucl. Phys. B **278**, 493 (1986).

[14]I. V. Cherednik, Sov. Math. Dokl. **33**, 507 (1986).

[15]M. Rosso, Commun. Math. Phys. **117**, 581 (1988).

[16]G. Lusztig, Adv. Math. **70**, 237 (1988).

[17]D.-J. Smit, Commun. Math. Phys. **128**, 1 (1990).

[18]K. Fredenhagen, K. H. Rehren, and B. Schroer, Commun. Math. Phys. **125**, 201 (1989).

[19]P. Freyd, D. Yetter, W. Lickorish, K. Millet, A. Ocneanu, and J. Hoste, Bull. Am. Math. Soc. **12**, 239 (1985).

[20]V. F. R. Jones, Ann. Math. **126**, 335 (1987).

[21]G. Lusztig, Contemp. Math. **82**, 59 (1989).

[22]P. Roche and D. Arnaudon, Lett. Math. Phys. **17**, 295 (1989).

[23]V. Pasquier and H. Saleur, Nucl. Phys. B **330**, 523 (1990).

[24]H. Wenzl, Invent. Math. **92**, 349 (1988).

[25]I. Frenkel, Talk at the conference on "Infinite-dimensional Lie Algebras and Infinite-dimensional Lie Groups," Paris (1989).

[26]J. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972), p. 51.

[27]J. Fuchs and D. Gepner, Nucl. Phys. B **294**, 30 (1987).

[28]J. Fuchs, Nucl. Phys. B (Proc. Suppl.) **6**, 157 (1989).

On an exceptional nonassociative superspace

M. Günaydin[a)]
CERN, CH-1211 Geneva 23, Switzerland and Department of Physics, Penn State University, University Park, Pennsylvania 16802[b)]

(Received 20 November 1989; accepted for publication 21 February 1990)

The Jordan superalgebra JF(6/4) is the unique exceptional Jordan superalgebra that has no realization in terms of Z_2-graded associative supermatrices. It is proposed to be the basis of an exceptional superspace that is non-Clifford algebraic. The JF(6/4) is constructed in a basis that renders itself to such an interpretation and its derivation, reduced structure, and Möbius superalgebras are studied. These algebras are simply the Lie superalgebras of generalized rotation, the Lorentz group, and conformal supergroup of the Jordan superalgebra JF(6/4). We also comment on the implications of the exceptionality of JF(6/4).

I. INTRODUCTION

The twistor formalism in four-dimensional space-time $(d=4)$ leads naturally to the representation of four-vectors in terms of 2×2 Hermitian matrices over the field of complex numbers C: In particular, the coordinate four-vectors can be represented as such. In this form the actions of four-dimensional space-time symmetry groups on the Minkowski space take on particularly elegant forms. For example, the action of the conformal group on the Minkowski coordinates can be realized as a group of linear fractional transformations of the corresponding 2×2 matrices.[1] Since Hermitian matrices close under anticommutation one can consider them as elements of a Jordan algebra with the symmetric Jordan product.[2] Then the rotation, Lorentz, and conformal groups in $d=4$ can be interpreted as the automorphism, reduced structure, and Möbius groups of the Jordan algebra of 2×2 complex Hermitian matrices.[3] Conversely, this interpretation allows one to define the concepts of rotation, Lorentz groups, and conformal groups for any Jordan algebra:[3,4] In the mathematics literature they have been studied under the names automorphism, reduced structure, and superstructure groups ("super" in this case having nothing to do with supersymmetry).[5] Denoting as J_n^A the Jordan algebra of $n\times n$ Hermitian matrices over the division algebra A, one finds the symmetry groups shown in Table I.

The symbols R, C, H, O in Table I represent the four division algebras and RG, LG, and CG are the rotation (automorphism), Lorentz (reduced structure), and conformal (Möbius) groups, respectively. In addition to the Jordan algebras in Table I there is another infinite family of simple Jordan algebras,[6] namely the Jordan algebras $\Gamma(d)$ of Dirac gamma matrices in d dimensions:

$$\{\Gamma_\mu,\Gamma_\nu\}=\delta_{\mu\nu}1,\quad \mu,\nu,...=1,2,...,d. \tag{1.1}$$

The automorphism, reduced structure, and Möbius groups of the Jordan algebras $\Gamma(d)$ are simply the rotation, Lorentz, and conformal groups of $(d+1)$-dimensional Minkowski space-times. One finds the following special isomorphisms between the Jordan algebras of 2×2 Hermitian matrices over the four division algebras and the Jordan algebras of gamma matrices:

$$J_2^R\approx\Gamma(2),\quad J_2^C\approx\Gamma(3),\quad J_2^H\approx\Gamma(5),\quad J_2^O\approx\Gamma(9). \tag{1.2}$$

As has been observed by many authors, the dimensions of the Minkowski space-times to which the above gamma matrices correspond are precisely the critical dimensions for the existence of super Yang–Mills theories, as well as of the classical Green–Schwarz superstrings. These Jordan algebras are quadratic and their norm forms are precisely the quadratic invariants constructed using the Minkowski metric. The reduced structure group is defined as the invariance group of the norm form and the superstructure (conformal) group is simply the invariance group of the zero-norm condition: Clearly, for the quadratic norms in question these definitions agree with the usual ones.

Jordan superalgebras are Z_2-graded algebras with a supersymmetric product and have been classified by Kac.[7] One can similarly define their rotation, Lorentz groups, and conformal supergroups as their automorphism, reduced structure, and Möbius supergroups. A complete list of these supergroups was given in Ref. 3. In Table II we list the simple Jordan superalgebras and their symmetry supergroups using a modified version of the Kac notation for labeling Jordan superalgebras. For example, the Jordan superalgebra of type X with m even elements and n odd elements is denoted as JX(m/n).

In Table II we denoted the super rotation, Lorentz groups, and conformal groups of Jordan superalgebras as SRG, SLG, and SCG, respectively: They appear from left to right, in that order, on the line immediately below the corresponding Jordan superalgebra JX. The term $U(1)_{(F)}$ denotes the "fermionic" $U(1)$ factor generated by a single odd

TABLE I. The symmetry groups RG, LG, and CG of the Jordan algebra J.

J	RG	LG	CG
J_n^R	SO(n)	SL(n,R)	Sp$(2n,R)$
J_n^C	SU(n)	SL(n,C)	SU(n,n)
J_n^H	USp$(2n)$	SU*$(2n)$	SO*$(4n)$
J_3^O	F_4	$E_{6(-26)}$	$E_{7(-25)}$

[a)] Bitnet: GXT@PSUVM.
[b)] Permanent address.

TABLE II. Simple Jordan superalgebras and their symmetry supergroups using a modified version of the Kac[7] notation for labeling Jordan superalgebras.

SRG	SLG	SCG
$JA(m^2+n^2/2mn)$: $SU(m/n)$	$SU(m/n)\times SU(m/n)$	$SU(2m/2n)$
$JBC(\frac{1}{2}m(m+1)+n(2n-1)/2mn)$: $OSp(m/2n)$	$SU(m/2n)$	$OSp(4n/2m)$
$JD(m/2n)$: $OSp(m-1/2n)$	$OSp(m/2n)$	$OSp(m+2/2n)$
$JP(n^2/n^2)$: $P(n-1)$	$SU(n/n)$	$P(2n-1)$
$JQ(n^2/n^2)$: $Q(n-1)\times U(1)_F$	$Q(n-1)\times Q(n-1)\times U(1)_F$	$Q(2n-1)$
$JD(2/2)_\alpha$: $OSp(1/2)$	$SU(1/2)$	$D(2,1,;\alpha)$
$JF(6/4)$: $OSp(1/2)\times OSp(1/2)$	$OSp(2/4)$	$F(4)$
$JK(1/2)$: $OSp(1/2)$	$SU(1/2)$	$SU(2/2)$

generator.[3,8]

Some of the supergroups appearing in Table II correspond to space-time symmetry groups in various dimensions: In these cases one may consider the underlying Jordan superalgebra as the basis of a superspace by identifying the even subspace with the space-time coordinates and the odd subspace with the Grassmann coordinates. By multiplying the odd elements with anticommuting Grassmann coefficients one can work with the symmetric Jordan product in both sectors.[3,8]

In the list of simple Jordan superalgebras in Table II one is truly unique, namely the exceptional Jordan superalgebra JF(6/4). It is the only simple Jordan superalgebra that has no realization in terms of Z_2-graded associative supermatrices. That it is exceptional has been proved rather recently.[9] Our aim in this article is to study JF(6/4) and its symmetry supergroups. We shall propose it as the basis of an exceptional superspace. Its exceptionality implies that the corresponding supersymmetry is non-Clifford algebraic. For a certain real form of JF(6/4) the Möbius group F(4) is simply the $N=2$ superconformal group in five-dimensional Minkowski space (or the $N=2$ anti-de Sitter supergroup in $d=6$). Another real form of JF(6/4) leads to that real form of F(4) that corresponds to an exceptional $N=8$ conformal supergroup in one dimension.

The plan of our paper is as follows. In Sec. II we review the Tits–Koecher–Kantor (TKK) construction of Lie algebras from Jordan algebras[10–12] and its generalization to Jordan superalgebras.[3,8] In Sec. III we formulate the Lie superalgebra of F(4) in a split basis with a three-graded (Jordan) structure. In Sec. IV we construct the exceptional Jordan superalgebra in a basis that is suitable for interpretation as the basis of a superspace. We then study the Lie superalgebra of the automorphism (rotation), reduced structure (Lorentz) and Möbius (conformal) supergroups of JF(6/4). In Sec. V we discuss some of the implications of the exceptionality of JF(6/4).

II. TKK CONSTRUCTION OF LIE ALGEBRAS FROM JORDAN ALGEBRAS AND ITS SUPERSYMMETRIC GENERALIZATION

A Jordan algebra J is a nonassociative algebra with a symmetric product $a\cdot b=b\cdot a$ satisfying the Jordan identity

$$a\cdot(b\cdot a^2)=(a\cdot b)\cdot a^2,\quad \forall a,b\in J. \tag{2.1}$$

A derivation D of J is an endomorphism of J such that

$$D(a\cdot b)=(Da)\cdot b+a\cdot(Db). \tag{2.2}$$

Derivations form a Lie algebra under commutation, which is referred to as the derivation algebra (Der J) of J. If we denote the left and right multiplication by an element a as L_a and R_a, respectively, we have $L_a=R_a$ since

$$L_a b\equiv a\cdot b=b\cdot a\equiv R_a b. \tag{2.3}$$

We can rewrite the derivation condition (2.2) as

$$[D,L_a]=L_{Da}. \tag{2.4}$$

Linearization of the Jordan identity leads to the identity

$$(a\cdot b)\cdot(c\cdot d)+(a\cdot c)\cdot(b\cdot d)+(a\cdot d)\cdot(b\cdot c)$$
$$=a\cdot(b\cdot(c\cdot d))+c\cdot(b\cdot(a\cdot d))+d\cdot(b\cdot(a\cdot c)), \tag{2.5}$$

which is equivalent to the identities

$$[L_{a\cdot b},L_c]+[L_{c\cdot a},L_b]+[L_{b\cdot c},L_a]=0 \tag{2.6}$$

and

$$[[L_a,L_b],L_c]=L_{[L_a,L_b]c}. \tag{2.7}$$

Identity (2.7) shows that $[L_a,L_b]$ is a derivation of the Jordan algebra J. In fact, one can prove that all inner derivations of a Jordan algebra can be represented in this form. Therefore, we can formally write

$$[L_J,L_J]=\mathrm{Der}\,J. \tag{2.8}$$

Exponentiating the derivations one obtains the inner automorphisms of J:

$$e^D(a\cdot b) = (e^D a)\cdot(e^D b). \tag{2.9}$$

Therefore, the derivation algebra of J is simply the Lie algebra of its automorphism (rotation) group. It is obvious from above that the multiplication operators close under commutation into derivations of J. The Lie algebra generated by multiplications with elements of J and its derivations is referred to as the structure algebra $\mathrm{St}(J)$. Clearly, the multiplication by the identity element of J commutes with the other elements of $\mathrm{St}(J)$. Excluding this identity operator one is left with what is referred to as the reduced structure algebra $\mathrm{St}(J)_0$ of J. The reduced structure algebra is simply the Lie algebra of the reduced structure group or what we called the Lorentz group of J in Sec. I.

Another way to describe the $\mathrm{St}(J)$ is via the Jordan triple product, which is defined as

$$\{abc\}\equiv a\cdot(b\cdot c) + (a\cdot b)\cdot c - b\cdot(a\cdot c). \tag{2.10}$$

Using the triple product (2.10) we can define a bilinear transformation S_{ab} on J as

$$S_{ab}c\equiv\{abc\}, \quad \forall a,b,c\in J. \tag{2.11}$$

The transformations S_{ab} generate $\mathrm{St}(J)$ under commutation:

$$[S_{ab},S_{cd}] = -S_{\{cda\},b} + S_{a,\{dcb\}}, \tag{2.12}$$

which is readily seen from the fact that

$$S_{ab} = [L_a,L_b] + L_{a\cdot b} = D_{a,b} + L_{a\cdot b}. \tag{2.13}$$

The TKK construction defines a three-graded Lie algebra g over a given Jordan algebra J:

$$g = g^{-1}\oplus g^0\oplus g^{+1}. \tag{2.14}$$

The elements of grade $+1$ and grade -1 spaces are labeled by the elements of J (hence we are actually using two copies of J, which are referred to as a Jordan pair[5]):

$$\left.\begin{array}{l} U_a\in g^{+1} \\ \widetilde{U}_b\in g^{-1} \\ S_{ab}\in g^0 \end{array}\right\}, \forall a,b\in J. \tag{2.15}$$

Then one defines the Lie product among these elements as

$$\begin{aligned} &[U_a,\widetilde{U}_b]\equiv S_{ab}, \quad [S_{ab},U_c]\equiv U_{\{abc\}}, \\ &[S_{ab},\widetilde{U}_c] = -\widetilde{U}_{\{bac\}}, \\ &[S_{ab},S_{cd}] = S_{\{abc\},d} - S_{c,\{bad\}}, \end{aligned} \tag{2.16}$$

with the other products vanishing. The Jacobi identities follow from the Jordan identity. The resulting Lie algebra g is referred to as the superstructure algebra and is isomorphic to the Lie algebra of the superstructure (Möbius) group of J, which we called the conformal group in Sec. I. There is a nonlinear realization of the conformal group on J as a group of linear fractional transformations.[13]

The TKK construction has been generalized to the construction of Lie superalgebras from Jordan superalgebras.[3,8] A Jordan superalgebra is a Z_2-graded algebra $J = J^0\oplus J^1$, with a supercommutative product

$$a\cdot b = (-1)^{\alpha\beta} b\cdot a, \quad a\in J^\alpha, \quad b\in J^\beta, \quad \alpha,\beta = 0,1, \tag{2.17}$$

which satisfies the identity

$$\begin{aligned} &(-1)^{\alpha\gamma}[L_{a\cdot b},L_c\} + (-1)^{\beta\alpha}[L_{b\cdot c},L_a\} \\ &\quad + (-1)^{\gamma\beta}[L_{c\cdot a},L_b\} = 0, \end{aligned} \tag{2.18}$$

where the mixed bracket $[,\}$ denotes the usual Lie superbracket. The defining conditions (2.17) and (2.18) imply the identity

$$[[L_a,L_b\},L_c\} = (-1)^{\beta\gamma} L_{a\cdot(c\cdot b) - (a\cdot c)\cdot b}. \tag{2.19}$$

By multiplying the odd elements (grade 1) with anticommuting Grassmann parameters and the even elements (grade zero) with ordinary complex parameters, the supercommutative product can be replaced with a commutative one and the TKK construction can be carried over to the super case in a straightforward manner.[3,8]

One can invert the TKK construction and define Jordan (super) algebras starting from Lie (super) algebras with the appropriate three-graded structure.[7,14] In fact, this is how Kac defined and classified Jordan superalgebras.[7] We shall also use the inverse TKK functor to write the multiplication table of the exceptional Jordan superalgebra JF(6/4). We shall choose a basis that renders itself readily to a space-time interpretation, in contrast to the basis given in Ref. 7. Our starting point will be the exceptional Lie superalgebra F(4), to which we turn next.

III. THE EXCEPTIONAL LIE SUPERALGEBRA F(4)

The exceptional Lie superalgebra F(4) has $\mathrm{Spin}(7)\times\mathrm{SU}(2)$ as its even subalgebra. The odd generators transform in the spinor representation (8,2) of $\mathrm{Spin}(7)\times\mathrm{SU}(2)$. Explicit supercommutation relations of the generators of F(4) were given in Refs. 15 and 16. Denoting the generators of Spin(7) and SU(2) as J_{ij} and T_m, respectively, we have

$$\begin{aligned} &[J_{ij},J_{kl}] = -\delta_{ik}J_{jl} + \delta_{jk}J_{il} - \delta_{jl}J_{ik} + \delta_{il}J_{jk}, \\ &J_{ij} = -J_{ji}, \quad i,j,... = 1,2,...,7, \\ &[T_m,T_n] = i\epsilon_{mnp}T_p, \quad m,n,... = 1,2,3. \end{aligned} \tag{3.1}$$

The odd generators V_{xa} carry the spinor indices x and a of SU(2) and Spin(7), respectively. The remaining nonvanishing supercommutation relations of F(4) are[15]

$$\begin{aligned} &[T_m,V_{xa}] = \tfrac{1}{2}\tau^m_{yx}V_{ya}, \quad [J_{ij},V_{xa}] = \tfrac{1}{2}(\Gamma_{ij})_{ba}V_{xb}, \\ &\{V_{xa},V_{yb}\} = 2\sigma\widetilde{C}_{ab}(\widehat{C}\tau^m)_{xy}T_m \\ &\qquad + (\sigma/3)\widehat{C}_{xy}(\widetilde{C}\Gamma_{ij})_{ab}J_{ij}, \end{aligned} \tag{3.2}$$

where $\widehat{C}$ and $\widetilde{C}$ are the charge conjugation matrices in dimensions $d=2$ and $d=7$, respectively; Γ_{ij} are the spinor representation matrices of J_{ij}; τ^m are the Pauli matrices; and σ is an arbitrary real parameter. We shall take $\sigma = \frac{3}{2}$ for later convenience and for $\widehat{C}$ we shall choose $\widehat{C} = i\tau^2$.

The matrix $\widetilde{C}$ is defined by the condition

$$\widetilde{C}\Gamma_i = -\Gamma_i^T\widetilde{C}, \quad \widetilde{C}^T = \widetilde{C}, \tag{3.3}$$

where Γ_i are the gamma matrices in $d=7$ satisfying

$$\begin{aligned} &\{\Gamma_i,\Gamma_j\} = 2\delta_{ij}, \quad i,j = 1,2,...,7, \\ &\Gamma_{ij}\equiv\tfrac{1}{2}[\Gamma_i,\Gamma_j]. \end{aligned} \tag{3.4}$$

We shall choose Γ_i to be antisymmetric and $\widetilde{C}$ to be the identity matrix.[12]

For applying the inverse TKK functor to construct the exceptional Jordan superalgebra JF(6/4) we need to go to a split basis for F(4), with a three-graded decomposition (Jordan structure) $F(4) = g^{-1} \oplus g^0 \oplus g^{+1}$, where the three graded subspaces contain the following generators:

$$
\begin{aligned}
g^{-1} &= \{J_I^-, T^-, V_a^-\},\\
g^0 &= \{J_{IJ}, N, T^3, W_a, Z_a\},\\
g^{+1} &= \{J_I^+, T^+, V_a^+\}.
\end{aligned} \tag{3.5}
$$

Various new quantities appearing in (3.5) are defined as

$$
\begin{aligned}
&J_I^\mp = J_{I6} \mp iJ_{I7}, \quad I,J,... = 1,2,...,5,\\
&T^\mp = T^1 \mp iT^2, \quad V_a^+ = (\Gamma_6 + i\Gamma_7)_{ab} V_{1b},\\
&V_a^- = (\Gamma_6 - i\Gamma_7)_{ab} V_{2b}, \quad N = -iJ_{67},\\
&W_a = (\Gamma_6 + i\Gamma_7)_{ab} V_{2b}, \quad Z_a = (\Gamma_6 - i\Gamma_7)_{ab} V_{1b}.
\end{aligned} \tag{3.6}
$$

Note, also, that the grading is defined by the generator $Q = N + T^3$ and only four of the eight components of V_a^+, V_a^-, W_a, and Z_a are linearly independent.

In this split basis the generators of F(4) satisfy the following supercommutation relations:

$$
\begin{aligned}
&[J_{IJ}, J_K^\pm] = -\delta_{IK} J_J^\pm + \delta_{JK} J_I^\pm,\\
&[J_I^+, J_J^-] = -2J_{IJ} - 2\delta_{IJ} N,\\
&[N, J_I^\mp] = \mp J_I^\mp, \quad [N, V_a^\mp] = \mp\tfrac{1}{2} V_a^\mp,\\
&[T_3, V_a^\mp] = \mp\tfrac{1}{2} V_a^\mp, \quad [N, W_a] = \tfrac{1}{2} W_a = -[T_3, W_a],\\
&[N, Z_a] = -\tfrac{1}{2} Z_a = -[T_3, Z_a], \quad [T^+, T^-] = 2T_3.
\end{aligned} \tag{3.7}
$$

IV. THE EXCEPTIONAL JORDAN SUPERALGEBRA JF (6/4) AND ITS SYMMETRY SUPERGROUPS

As mentioned earlier, one may invert the TKK construction and construct Jordan algebras[12] and Jordan superalgebras[7] from certain three-graded Lie (super) algebras. We shall now apply this inverse TKK functor to construct the exceptional Jordan superalgebra starting from F(4). Consider now the subspace g^{+1} of F(4) as given in Sec. III. Among any two elements A and B of g^{+1} we define the product

$$A \cdot B \equiv [\{A,P\}, B\}, \quad \forall A,B \in g^{+1}, \tag{4.1}$$

where P is a fixed element of the g^{-1} subspace, which we take to be

$$P = \tfrac{1}{2}(T^- - J_5^-). \tag{4.2}$$

From the Jacobi identity and the three-grading it then follows that

$$(-1)^{\alpha\beta} [\{A,P\}, B\} + [\{P,B\}, A\} = 0. \tag{4.3}$$

This shows that the product is supercommutative and closure is a trivial consequence of the three-grading,

$$A \cdot B = (-1)^{\alpha\beta} B \cdot A. \tag{4.4}$$

Under this product (4.4) we then find

$$
\begin{aligned}
&J_I^+ \cdot J_J^+ = -\delta_{IJ} J_5^+ + \delta_{5J} J_I^+ + \delta_{I5} J_J^+,\\
&J_I^+ \cdot T^+ = 0, \quad T^+ \cdot T^+ = T^+,\\
&\Theta_a \cdot \Theta_b = (2\sigma/3)[\Gamma_I(\Gamma_6 + i\Gamma_7)]_{ab} J_I^+\\
&\qquad - 2\sigma[\Gamma_5(\Gamma_6 + i\Gamma_7)]_{ab} T^+,\\
&T^+ \cdot \Theta_a = \tfrac{1}{2}\Theta_a, \quad J_I^+ \cdot \Theta_a = \tfrac{1}{2}\delta_{I5}\Theta_a - \tfrac{1}{2}(\Gamma_{I5})_{ab}\Theta_b,\\
&J_5^+ \cdot J_5^+ = J_5^+,
\end{aligned} \tag{4.5}
$$

where we let $\Theta_a = V_a^+$. Keeping in mind that not all Θ_a are linearly independent, let us study the derivations of JF(6/4). The superderivation algebra of JF(6/4) is generated by the following operators:

$$
\begin{aligned}
&D_{I,J} = [L_{J_I^+}, L_{J_J^+}] = -D_{J,I},\\
&S_{a,b} = \{L_{\Theta_a}, L_{\Theta_b}\} = S_{b,a}, \quad F_{I,a} = [L_{J_I^+}, L_{\Theta_a}].
\end{aligned} \tag{4.6}
$$

The operators $D_{I,J}$ and $S_{a,b}$ are the even generators and the $F_{I,a}$ are the odd generators. The even generators $D_{I,J}$ act on the elements of JF(6/4) as

$$
\begin{aligned}
&D_{I,5} J_K^+ = 0 = D_{I,J} T^+,\\
&D_{\mu,\nu} J_5^+ = 0, \quad \mu,\nu = 1,2,3,4,\\
&D_{\mu,\nu} J_\lambda^+ = \delta_{\mu\lambda} J_\nu^+ - \delta_{\nu\lambda} J_\mu^+,\\
&D_{\mu,\nu} \Theta_a = \tfrac{1}{4}[\Gamma_\mu, \Gamma_\nu]_{ab} \Theta_b,\\
&D_{I,5} \Theta_a = 0.
\end{aligned} \tag{4.7}
$$

The above commutation relations imply that $D_{I,5} = 0$ and the generators $D_{\mu,\nu}$ close under commutation and form the Lie algebra of SO(4):

$$[D_{\mu,\nu}, D_{\lambda,\rho}] = \delta_{\nu\lambda} D_{\mu,\rho} + \delta_{\mu\rho} D_{\nu,\lambda} - \delta_{\mu\lambda} D_{\nu\rho} - \delta_{\nu\rho} D_{\mu,\lambda}. \tag{4.8}$$

The derivations $S_{a,b}$ turn out not to be independent of $D_{\mu,\nu}$. One finds that

$$S_{a,b} = -(\sigma/3) D_{\mu,\nu} [\Gamma_{\mu\nu} \Gamma_5 (\Gamma_6 + i\Gamma_7)]_{ab}. \tag{4.9}$$

Thus far we have been using the same symbols for denoting the elements of JF(6/4) and those of the subspace g^{+1} considered as elements of the Lie superalgebra F(4). Let us now introduce different symbols for denoting the elements of JF(6/4) and work only with the linearly independent components of Θ_a. We choose the convention of Ref. 16 for Γ_i, i.e.,

$$
\begin{aligned}
&\Gamma_1 = 1 \otimes \sigma_3 \otimes \sigma_2, \quad \Gamma_2 = 1 \otimes \sigma_1 \otimes \sigma_2,\\
&\Gamma_3 = \sigma_2 \otimes 1 \otimes \sigma_3, \quad \Gamma_4 = \sigma_2 \otimes 1 \otimes \sigma_1,\\
&\Gamma_5 = \sigma_3 \otimes \sigma_2 \otimes 1, \quad \Gamma_6 = \sigma_1 \otimes \sigma_2 \otimes 1,\\
&\Gamma_7 = \sigma_2 \otimes \sigma_2 \otimes \sigma_2,
\end{aligned} \tag{4.10}
$$

where σ_1, σ_2, and σ_3 are the Pauli matrices. Then we find the following relations among the components of Θ:

$$
\begin{aligned}
&\Theta_2 = i\Theta_1, \quad \Theta_4 = i\Theta_3, \quad \Theta_6 = -i\Theta_5,\\
&\Theta_8 = -i\Theta_7.
\end{aligned} \tag{4.11}
$$

We set $\sigma = \tfrac{3}{2}$ and define

$$
\begin{aligned}
&B_0 \equiv J_5^+, \quad S \equiv T^+, \quad B_\mu \equiv J_\mu^+, \quad \mu,\nu = 1,...,4,\\
&(Q_\alpha) = \begin{pmatrix} \Theta_1 \\ \Theta_3 \\ \Theta_5 \\ \Theta_7 \end{pmatrix}, \quad \alpha,\beta,... = 1,2,3,4.
\end{aligned} \tag{4.12}
$$

The multiplication table of JF(6/4) now takes on a more transparent form:

$$\begin{aligned}
&B_\mu\cdot B_\nu = -\delta_{\mu\nu}B_0, \quad B_0\cdot B_\mu = B_\mu, \quad B_0\cdot B_0 = B_0,\\
&B_0\cdot S = 0 = B_\mu\cdot S, \quad S\cdot S = S,\\
&Q_\alpha\cdot Q_\beta = (i\gamma_5\gamma_\mu C)_{\alpha\beta}B^\mu + (\gamma_5 C)_{\alpha\beta}(B_0 - 3S),\\
&B_\mu\cdot Q_\alpha = (i/2)(\gamma_\mu)_{\alpha\beta}Q_\beta, \quad B_0\cdot Q_\alpha = \tfrac{1}{2}Q_\alpha,\\
&S\cdot Q_\alpha = \tfrac{1}{2}Q_\alpha,\\
&\mu,\nu,\ldots = 1,\ldots,4, \quad \alpha,\beta,\ldots = 1,\ldots,4.
\end{aligned} \tag{4.13}$$

The B_0 and S are the two idempotents and $I = B_0 + S$ is the identity element of JF(6/4). The matrices γ_μ are the four-dimensional Dirac gamma matrices and C is the charge conjugation matrix:

$$\{\gamma_\mu,\gamma_\nu\} = 2\delta_{\mu\nu}, \quad \gamma_\mu C = -C\gamma_\mu^T. \tag{4.14}$$

The indices μ, ν,... and α, β,... are the vector and spinor indices of SO(4), respectively, generated by $D_{\mu,\nu}$. The above convention for the Γ_i leads to the following expressions for γ_μ and C:

$$\begin{aligned}
&\gamma_1 = 1\otimes\sigma_1, \quad \gamma_2 = -1\otimes\sigma_3, \quad \gamma_3 = -\sigma_1\otimes\sigma_2,\\
&\gamma_4 = -\sigma_2\otimes\sigma_3, \quad \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4, \quad C = \gamma_3.
\end{aligned} \tag{4.15}$$

In the new basis the elements of the derivation algebra Der JF(6/4) are

$$\begin{aligned}
&D_{\mu,\nu} = [L_{B_\mu},L_{B_\nu}] = -D_{\nu,\mu},\\
&G_{\alpha,\beta} = \{L_{Q_\alpha},L_{Q_\beta}\} = G_{\beta,\alpha},\\
&F_{0,\alpha} = [L_{B_0},L_{Q_\alpha}], \quad F_{\mu,\alpha} = [L_{B_\mu},L_{Q_\alpha}],\\
&E_{S,\alpha} = [L_S,L_{Q_\alpha}].
\end{aligned} \tag{4.16}$$

However, they are not all independent. We find that

$$\begin{aligned}
&F_{\mu,\alpha} = i(\gamma_\mu)_{\alpha\beta}F_{0,\beta}, \quad E_{S,\alpha} = -F_{0,\alpha},\\
&G_{\alpha,\beta} = -\tfrac{1}{4}(\gamma_{\mu\nu}\gamma_5 C)_{\beta\alpha}D_{\mu,\nu}.
\end{aligned} \tag{4.17}$$

The independent generators satisfy the supercommutation relations

$$\begin{aligned}
&[D_{\mu,\nu},D_{\lambda,\rho}] = \delta_{\nu\lambda}D_{\mu,\rho} + \delta_{\mu\rho}D_{\nu,\lambda} - \delta_{\mu\lambda}D_{\nu,\rho} - \delta_{\nu\rho}D_{\mu,\lambda},\\
&[D_{\mu,\nu},F_{0,\alpha}] = \tfrac{1}{2}(\gamma_{\mu\nu})_{\alpha\beta}F_{0,\beta},\\
&\{F_{0,\alpha},F_{0,\beta}\} = -\tfrac{1}{8}(\gamma_5\gamma_{\mu\nu})_{\beta\alpha}D_{\mu,\nu}.
\end{aligned} \tag{4.18}$$

Clearly, the even subalgebra is the Lie algebra of SO(4). It is straightforward to verify that the full derivation algebra is the Lie superalgebra of OSp(1/2)$\times$OSp(1/2). This is best done by decomposing the generators of SO(4) into its two commuting SU(2) subalgebras. Therefore, we have

$$\text{Der JF}(6/4) = \text{OSp}(1/2)\otimes\text{OSp}(1/2). \tag{4.19}$$

The structure algebra of JF(6/4) is generated by the derivations and the multiplications by the elements of JF(6/4): They satisfy the following supercommutation relations:

$$\begin{aligned}
&[D_{\mu,\nu},L_{B_\lambda}] = \delta_{\mu\lambda}L_{B_\nu} - \delta_{\nu\lambda}L_{B_\mu},\\
&[D_{\mu,\nu},L_{Q_\alpha}] = \tfrac{1}{4}[\gamma_\mu,\gamma_\nu]_{\alpha\beta}L_{Q_\beta},\\
&[F_{0,\alpha},L_S] = \tfrac{1}{4}L_{Q_\alpha},\\
&[F_{0,\alpha},L_{B_0}] = -\tfrac{1}{4}L_{Q_\alpha},\\
&[F_{0,\alpha},L_{B_\mu}] = -(i/4)(\gamma_\mu)_{\alpha\beta}L_{Q_\beta},\\
&[F_{0,\alpha},L_{Q_\beta}] = \tfrac{1}{2}(i\gamma_5\gamma_\mu C)_{\alpha\beta}L_{B_\mu}\\
&\qquad + \tfrac{1}{2}(\gamma_5 C)_{\alpha\beta}(L_{B_0} + 3L_S).
\end{aligned} \tag{4.20}$$

In addition, there are those supercommutation relations that define the derivations in terms of multiplication operators of JF(6/4). The structure algebra with 11 even and 8 odd generators is simply the Lie superalgebra of OSp(2/4)$\times U(1)$. The $U(1)$ generator is $L_I = L_{B_0} + L_S$. The even subgroup of the reduced structure group OSp(2/4) is SO(2)$\times$Sp(4).

At this point we should stress that different real forms of JF(6/4) will have different real forms of the derivation and structure algebras. For example, there exist other real forms of JF(6/4) for which the even subgroup SO(4) of the automorphism group goes over to SO(3,1) or SO(2,2). These different real forms of JF(6/4) can be obtained via the application of the inverse TKK functor to different real forms of F(4).

Finally, one can now rewrite F(4) as the Möbius (conformal) superalgebra of JF(6/4) using the supergeneralization of Eqs. (2.16) as given in Refs. 3 and 8. Perhaps, for physical applications the nonlinear realization of F(4) as the supergroup of linear fractional transformations of JF(6/4) in the sense of Ref. 13 is more relevant: It simply corresponds to the superconformal group action on the corresponding superspace.

V. ON THE EXCEPTIONALITY OF JF(6/4)

A Jordan algebra is said to be special if it can be represented in terms of associative matrices, with the Jordan product defined as one-half the anticommutator: Otherwise, it is said to be exceptional. In their classic work, Jordan *et al.* proved that with but one possible exception all Jordan algebras are special.[17] This possible exception is the Jordan algebra J_3^O of 3$\times$3 Hermitian octonionic matrices, whose exceptionality was proved by Albert.[18] In the 1980's Zelmanov extended these results to the infinite-dimensional case and proved that there are no infinite-dimensional exceptional Jordan algebras.[19]

In the Jordan formulation of quantum mechanics the observables and density matrices representing a physical system are elements of a Jordan algebra. If the underlying Jordan algebra is special, then the Jordan formulation is equivalent to the Dirac formulation of quantum mechanics over a Hilbert space (or a vector space in the finite-dimensional case). The existence of an exceptional Jordan algebra raised the question as to whether or not one can formulate quantum mechanics over the exceptional Jordan algebra which satisfy the axioms of Von Neumann: This was answered in the affirmative relatively recently.[20] This so-called octonionic quantum mechanics has no Hilbert space formulation. Somewhat later, the Jordan formulation was generalized to the quadratic Jordan formulation of quantum mechanics, which is applicable to the exceptional octonionic case as well.[3] The axioms of quantum mechanics as formulated by Von Neumann are equivalent to the axioms of projective geometry.[21] The projective geometry corresponding to octonionic quantum mechanics is nondesarguian,[20] which implies that it cannot be embedded in a higher dimensional projective geometry. This is consistent with the exceptionality of J_3^O.

In the 1980's the exceptional Jordan algebra J_3^O has made its appearance within the framework of supergravity theories through the work of Günaydin, Sierra, and Townsend (GST).[22] In their construction and classification of $N=2$ Maxwell–Einstein supergravity theories, GST showed that there exist four remarkable theories of this type which are uniquely determined by simple Jordan algebras of degree three. These are the Jordan algebras J_3^R, J_3^C, J_3^H, and the exceptional Jordan algebra J_3^O: Their symmetry groups in five, four, and three space-time dimensions give the famous magic square of Freudenthal, Rozenfeld, and Tits.[23] The exceptional theory defined by J_3^O shares the remarkable features of the maximally extended $N=8$ supergravity theory in the respective dimensions. More recently, several authors have speculated on the possible role J_3^O may play in the framework of string theories.[24]

In going over to the Jordan superalgebras most of the concepts and definitions for ordinary Jordan algebras carry over in a natural way. For example, consider a Z_2-graded associative algebra $R = R^0 \oplus R^1$. The elements of R define a Jordan superalgebra under the superanticommutator:

$$A \cdot B \equiv \tfrac{1}{2}[AB + (-1)^{\alpha\beta} BA], \quad \forall A,B \in R. \tag{5.1}$$

In most applications the algebra R is the algebra of operators acting on a Z_2-graded vector space $V = V^0 \oplus V^1$: Hence one can represent them as supermatrices.[25] For space-time supersymmetry the corresponding vector space V is referred to as the superspace. We shall use this term in a more general sense.

A Jordan superalgebra is then special if it can be represented in terms of associative Z_2-graded supermatrices, with the product being one-half the superanticommutator. All Jordan superalgebras are special except for JF(6/4). (Here we are only considering Jordan superalgebras whose odd subspaces J^1 are not empty.) The exceptionality of JF(6/4) has been shown by Shtern rather recently.[9] The existence of an exceptional Jordan superalgebra raises many interesting questions. For example, if we give it a space-time interpretation as the basis of an exceptional superspace, then this superspace has no realization as an ordinary Z_2-graded vector space on which the supersymmetry is represented Clifford algebraically. For special Jordan superalgebras this can always be done. In the language of ordinary quantum mechanics this means that for special Jordan algebras one may choose to work either with density matrices or vectors in a Hilbert space: These two formulations are equivalent. One major advantage of the Hilbert or vector space formulation is the fact that one can define tensor products of vector spaces in a straightforward manner. For ordinary superspace this means that one can develop a tensor calculus. On the other hand, defining tensor products of Jordan algebras is a notoriously difficult problem, e.g., the tensor product of two Jordan algebras is in general not a Jordan algebra. Therefore, if we were to consider JF(6/4) as the basis of an exceptional superspace, then it is desirable to formulate it in the language of ordinary Z_2-graded vector spaces. Assume that there exists such a "generalized vector space" and let us refer to it as the exceptional supermodule MF. The action of the supersymmetry operators, as well as the tensor product operation over MF will then have to be nonassociative. In other words, there will be a nonassociative tensor calculus corresponding to this exceptional superspace. Interestingly enough, the exceptional $N=2$ supergravity theories discovered by GST (Ref. 22) cannot be written in the language of ordinary conformal tensor calculus.[26] It could well be that the underlying tensor calculus is a nonassociative one. In fact, the exceptional $N=2$ theories originate in five space-time dimensions and the real form of F(4) with the even subgroups $SO(5,2) \times SU(2)$ is simply the $N=2$ conformal superalgebra in that dimension. Another way to approach this problem is to look at the unitary representations of F(4) and try to interpret them in the language of a 10-dimensional superspace corresponding to JF(6/4). A pure gauged anti-de Sitter supergravity in $d=6$ with F(4) symmetry was constructed in Ref. 27. Matter couplings of that theory, as well as the conformal supergravity in $d=5$ with F(4) symmetry are yet to be constructed. We hope to address these problems and related issues in the future.

We should also point out that the quadratic Jordan formulation of quantum mechanics[3] can be formally extended to Jordan superalgebras, as well as Jordan algebras defined over finite fields. In the super-Jordan case one takes the odd subspace over Grassmann numbers. Since the trace form is no longer a real number, one loses the usual probability interpretation. This might be remedied by interpreting the odd sector as unobservable. In any case, the formal quantum mechanics defined over JF(6/4) would again not have a usual vector space formulation.

Finally, the algebra JF(6/4) considered as the basis of a superspace corresponds to six bosonic and four "fermionic" (Grassmann) coordinates. Because of its potential relevance to the superstring one may want to know if there exists a larger algebraic structure, with 10 bosonic elements and 16 fermionic ones, which can be considered as the basis of a superspace. From Kac's[7] classification we know that this algebraic structure cannot be a simple Jordan superalgebra, although it may contain JF(6/4) as a subalgebra. If such an algebraic structure exists, then it may give us the clue as to how to generalize conformal supersymmetry to 10 space-time dimensions.

ACKNOWLEDGMENT

This work is dedicated to Murray Gell-Mann on the occasion of his 60th birthday.

[1] F. Gürsey, Nuovo Cimento **3**, 988 (1956).

[2] P. Jordan, Nach. Ges. Wiss. Göttingen **XX**, 209 (1932); Z. Physik **80**, 285 (1933).

[3]M. Günaydin, Ann. Israel Phys. Soc. **3**, 279 (1980).
[4]M. Günaydin, Nuovo Cimento A **29**, 467 (1975).
[5]See the excellent review by K. McCrimmon, Bull. Am. Math. Soc. **84**, 612 (1978) and the references therein.
[6]R. D. Schafer, *An Introduction to Non-associative Algebras* (Academic, New York, 1966).
[7]V. Kac, Commun. Algebra **5**, 1375 (1977).
[8]I. Bars and M. Günaydin, J. Math. Phys. **20**, 1977 (1979).
[9]A. S. Shtern, Funkt. Anal. Ego Prilozheniya **21**, 93 (1987).
[10]J. Tits, Nederl. Akad. van Wetens. **65**, 530 (1962).
[11]M. Koecher, Am. J. Math. **89**, 787 (1967).
[12]I. L. Kantor, Sov. Math. Dokl. **5**, 1404 (1964).
[13]M. Koecher, Inv. Math. **3**, 136 (1967).
[14]M. Günaydin, J. Math. Phys. **30**, 937 (1989).
[15]W. Nahm, V. Rittenberg, and M. Scheunert, J. Math. Phys. **9**, 1640 (1976).
[16]B. S. De Witt and P. van Nieuwenhuizen, J. Math. Phys. **23**, 1953 (1982).
[17]P. Jordan, J. von Neumann, and E. Wigner, Ann. Math. **36**, 29 (1934).
[18]A. A. Albert, Ann. Math. **36**, 65 (1934).
[19]E. I. Zelmanov, Sib. Math. J. 24, 89 (1983).
[20]M. Günaydin, C. Piron, and H. Ruegg, Commun. Math. Phys. **61**, 69 (1978).
[21]C. Piron, *Foundations of Quantum Physics* (Benjamin, Massachusetts, 1976).
[22]M. Günaydin, G. Sierra, and P. K. Townsend, Phys. Lett. B **133**, 72 (1983); Nucl. Phys. B **242**, 244 (1984); Nucl. Phys. B **253**, 573 (1985).
[23]H. Freudenthal, Nederl. Akad. Wetensch. Proc. A **62**, 447 (1959); B. A. Rozenfeld, Dokl. Akad. Nauk. SSSR **106**, 600 (1956); J. Tits, Mem. Acad. R. Belg. Sci. **29**, fasc. 3 (1955).
[24]D. B. Fairlie and C. Manouge, Phys. Rev. D **34**, 1832 (1986); G. Sierra, Class. Quant. Grav. **4**, 385 (1987); P. Goddard, W. Nahm, D. Olive, H. Ruegg, and A. Schwimmer, Commun. Math. Phys. **112**, 385 (1987); F. Gürsey, Mod. Phys. Lett. A **2**, 967 (1987); G. Chapline and M. Günaydin, UCRL preprint 95290 (1986), unpublished; E. Corrigan and T. J. Hollowood, Phys. Lett. B **203**, 47 (1988); M. Günaydin and S. J. Hyun, Phys. Lett. B **209**, 498 (1988).
[25]M. Scheunert, *The Theory of Lie Superalgebras*, Lecture Notes in Math. Vol. 716 (Springer, New York, 1979).
[26]B. de Wit and A. van Proeyen (private communication).
[27]L. J. Romans, Nucl. Phys. B **269**, 691 (1986).

Unified models from gauged supergroups. I. Dimensional reduction formalism. Unitary and orthosymplectic series

P. D. Jarvis[a)]
Department of Physics, University of Tasmania, Box 252C, GPO, Hobart Tas 7001, Australia

(Received 26 October 1989; accepted for publication 31 January 1990)

Principles are elaborated for constructing consistent grand unified models (GUT's) plus Higgs sectors from dimensional reduction of gauged supergroups over graded coset spaces. A search of possible GUT groups is made for unitary and orthosymplectic supergroups over a space parametrized by a single pair of Grassmann coordinates.

I. INTRODUCTION AND MAIN RESULTS

It is a commonly acknowledged failing of so-called grand unified models (GUT's),[1] which introduce a simple gauge group larger than the $SU(3)\times SU(2)\times U(1)$ of the standard electroweak model of elementary particle physics, that in many respects the procedure for arriving at physically acceptable models is too loose, while at the same time leaving many other basic questions unanswered.

This paper, and a sequel in preparation, concern a hitherto unexplored avenue for constructing constrained classes of unified models from *gauged supergroups* via coset space dimensional reduction (CSDR). An introduction to this method, and the main results of the present study, are given below. For completeness, the difficulties besetting the GUT's will first be outlined,[2] and the supergroup CSDR approach compared with other generalizations of GUT's.

Conventional GUT's (renormalizable, spontaneously broken non-Abelian gauge theories) involve choices of (i) a local gauge group (a compact Lie group); (ii) scalar multiplet(s) (Higgs sector) to implement spontaneous symmetry breaking; and (iii) fermion multiplets (quark and lepton matter fields). There is considerable flexibility[3] in all of these choices, even after satisfying physical requirements such as, for example, for (i) that the group should have complex representations (to avoid fermion masses at tree level); for (ii) and (iii) that the theory should be asymptotically free; and for (iii) that the model should be anomaly-free (to ensure renormalizability), and perhaps extra conditions such as only color triplets, antitriplets, and singlets. Systematic studies[4] taking these requirements into account still leave a virtually unlimited choice of gauge groups and scalar and matter multiplets. Even so, there are questions left unanswered centering around the coupling constant ratios (scalar self-couplings to implement the desired symmetry-breaking patterns and Higgs masses; Yukawa couplings to produce acceptable fermion mass ratios and weak boson masses [e.g., $M_W(L) \ll M_W(R)$], etc.). Often these couplings must be artificially "fine-tuned" order-by-order in perturbation theory. Above all, there is no prediction of the number of fermion generations. Around 20 parameters (masses, couplings, mixing angles, etc.) must be fitted to data, and even then a consensus is emerging[5] that the proton lifetime tends to be predicted seriously lower than the current experimental limits.

Three related avenues have been proposed for improving on conventional GUT model building as outlined above: (i) technicolor[6] and/or subconstituent models;[7] (ii) supersymmetric unified models[8] (possibly deriving from supergravity;[9] and (iii) realistic Kaluza–Klein models.[10] Option (i) replaces some of the questions about choices of scalar and matter multiplets and couplings, with dynamical issues concerning composite states (e.g., Higgs). Options (ii) alleviate the standard GUT difficulties with fine-tuning of parameters and the generation of mass hierarchies,[11] at the expense of an abundance of "superpartners" of the usual matter fields, none of which are as yet experimentally confirmed. Realistic gauge groups are hard to extract directly from extended supergravities. The latter, together with (iii), go beyond the bounds of renormalizable local quantum field theories, which form the main framework for the present discussion (for remarks on string theories, see Sec. IV).

The CSDR formalism as applied to pure gauge symmetries[12] is related to work on invariant connections on principal fiber bundles with applications to the topology of the solutions of the Yang–Mills equations for various types of field configuration. In the context of ordinary (nongraded) gauge groups, CSDR has been extensively studied[13–16] as a means of supplying a consistent rationale for starting with a higher-dimensional model and constructing constrained GUT models with predetermined scalar multiplets, scalar potential, and hence the unbroken gauge group and the pattern of massive and massless fields. While the CSDR approach to GUT's is perhaps more modest than the more sweeping alternatives listed above, nevertheless a phenomenologically compelling model has yet to emerge. [It should be stressed that these remarks apply to pure gauge theories and CSDR. There is a large literature on homogeneous spaces and symmetric potentials in the Kaluza–Klein context (see Refs. 10 and 17, and references therein). For remarks on string theory see Sec. IV.] Furthermore there remains the question of whether the formalism should be regarded as having dynamical meaning, or merely as a recipe for model building.

The variant of obtaining GUT models via CSDR with gauged supergroups as proposed here has its genesis in the work of Ne'eman and others on supergroup classification schemes for the electroweak model and extensions.[18] It was

[a)] Alexander von Humboldt Fellow.

 0022-2488/90/071783-08$03.00

claimed[19] that the supergroup models could make sense if formulated over a graded space-time manifold $(x^\mu,\theta,\bar{\theta})$ such that the parts of the potential gauging the odd generators have components only in the fermionic directions. These coordinates arise naturally in a superspace approach to BRST symmetry,[20] as the vehicle whereby the ghost fields are linked with the gauge potential (i.e., for the even generators an ordinary Lie algebra) in a superfield expansion in covariant quantization. These ingredients were married via CSDR[21] and a constrained O (n) model based on OSp$(n/2)$ was subsequently proposed.[22]

The aim of the present work, and a sequel in preparation, is to set out the general principles for obtaining consistent, constrained GUT models arising as solutions to the CSDR constraints (see Table I).

Briefly, the CSDR ansatz[23] is that only potentials A are admitted which are invariant under a transformation group S, in the sense that if $A \to {}^sA$, then ${}^sA$ is gauge equivalent to A. Closure under the group law then imposes constraints on such A, and there is a relation between the isotropy group R and the gauge group G, from which the structure of the scalar sector can be determined (see especially Schwarz and Tyupkin,[12] and Kubyshin *et al.*[16] for discussions). In the present case where R, S, and/or G are supergroups, additional consistency requirements must be satisfied for a solution to be physically acceptable. The formalism is outlined in Sec. II; in particular, the emergence of the consistency requirements, and the necessity of gauging noncompact real forms of the supergroups, is explained in detail.

In Sec. III the formalism is illustrated for the case of the real forms of the classical unitary and orthosymplectic superalgebras. [A sequel in preparation will consider the exceptional simple superalgebras $D(2,1\alpha)$, $G(3)$, and $F(4)$: the remaining simple superalgebras $p(n)$ and $q(n)$ lack an invariant bilinear form and cannot be gauged in the conventional way.] To make the analysis manageable, R is restricted mainly to the case $\mathrm{Sp}(2,R) \simeq \mathrm{SU}(1,1)$ (i.e., the covariance group of the extended BRST symmetry) or its U(1) subgroup, and S to the corresponding inhomogeneous extension including anticommuting supertranslations. More sophisticated possibilities for S/R are occasionally mentioned, but a general analysis is beyond the scope of this paper.

The results of this partial search for constrained GUT models via CSDR from gauged supergroups are set out in Table I. The original $\mathrm{O}(n)$ models[22] now appear as a particular case of a one- (integer) parameter family with $\mathrm{O}(n)$ symmetry and Higgs field in the fundamental representation based on $\mathrm{OSp}(n(j+1),nj/m(2j+1),R)$; in addition there are several new classes based on real forms of $B(0,n)$ and $A(m,n)$. For discussions of these solutions, their scalar sectors and possible spontaneous symmetry breaking, see Secs. III and IV. Other real forms than those of Table I (for which no solutions were found with the present choices of R and S) are also discussed in Sec. III.

The paper concludes in Sec. IV with some discussion of further extensions of the models, and likely generalizations to other choices of coset space, the inclusion of supersymmetry, and remarks on string theories. The analysis of Sec. III depends on some details of real forms and embeddings of $\mathrm{Sp}(2,R) \simeq \mathrm{SU}(1,1)$ in Lie algebras and superalgebras, and the necessary material is included as Appendices A and B.

TABLE I. Classes of unified models derived from gauged supergroups via CSDR, showing supergroup G, residual gauge group H, isotropy group H of the graded homogeneous space, and Higgs scalar multiplets λ_H (identified by their dimension as representations of H). See Sec. III for further comments and more complicated (reductive) possibilities for H not entered here.

R	G	H	λ_H
SU(1,1)	$\mathrm{SU}(j+1,j/n)$ $(2j+1>n)$	$\mathrm{SU}(n)(\times\mathrm{U}(1))$	(none)
SU(1,1)	SU(1,1/2)/(center)	SU(2)	$(2+\bar{2})$
U(1)	$\mathrm{OSp}(1/2n,R)$	$\mathrm{SU}(n)\times\mathrm{U}(1)$	$(n+\bar{n})$
$\mathrm{Sp}(2,R)$	$\mathrm{OSp}(nj+n,nj/2mj+2m,R)$ $(m=1$ or $n=1)$	$\mathrm{O}(n)$ or $\mathrm{O}(m)$	(n) or (m)

II. CSDR FORMALISM AND CONSISTENCY REQUIREMENTS FOR SUPERGROUPS

In this section we review[23] the CSDR formalism[12–16,21] in order to establish notation and to illustrate the special considerations necessary for dealing with gauged supergroups.[22]

Consider a gauge theory over a (super)manifold $\mathscr{M}\times\mathscr{N}$ with $\mathscr{M}$ a D-dimensional space-time manifold, $\mathscr{N}$ a (super)coset space, with points $X=(x,\theta)$, and gauge potential

$$A = A^aT_a = dX^M A_M{}^aT_a \tag{1}$$

referred to local coordinates $X^M=(x^\mu,\theta^m)$, $1\leqslant M\leqslant D+\widetilde{D}$, $1\leqslant\mu\leqslant D$, $1\leqslant m\leqslant\widetilde{D}$, in $\mathscr{M}\times\mathscr{N}$ and a basis $\{T_a\}$ for the Lie (super)algebra G. Then the (gauge plus Higgs) sector of a constrained GUT model arises with the higher-dimensional action

$$S \propto \int dx\, d\theta (F^{MN},F_{MN}) \tag{2}$$

and expanding in the fields that are solutions to the CSDR constraints. Here F_{MN} is the field strength

$$F_{MN} = \partial_M A_N - [MN]\partial_N A_M + [A_M,A_N]$$

the local coordinates. (For grading sign factors, see Appendix A.) Index contractions are with respect to an appropriate invariant metric tensor

$$G_{MN} = \begin{bmatrix} g_{\mu\nu} & 0 \\ 0 & g_{mn} \end{bmatrix}; \tag{3}$$

and (,) is an invariant bilinear form on the (super)algebra G.

Let the space-time (super)symmetry transformations act infinitesimally by

$$X^M \to X^M + \epsilon^A\xi_A{}^M(X). \tag{4}$$

(For a discussion in terms of finite transformations see Schwarz and Tyupkin.[12] For supermanifolds, see, for example, Ref. 24.) Here the invariant vector fields satisfy

$$\xi_A{}^K \partial_K \xi_B{}^L - [AB]\xi_B{}^K \partial_K \xi_A{}^L = -C_{AB}{}^C \xi_C{}^L,$$

corresponding to the (super)algebra $[J_A, J_B] = C_{AB}{}^C J_C$. The action on superspace induces an action on superfields, $\delta\phi = \epsilon^A \delta_A \phi$. This for the gauge potential

$$\delta_A A_K = -\xi_A{}^L \partial_L A_K - [AK](\partial_K \xi_A{}^L) A_L; \tag{5}$$

and this will be invariant if compensated by an infinitesimal gauge transformation with parameter $\epsilon^A W_A = \epsilon^A W_A{}^a T_a$:

$$[AK]\delta_A A_K = \partial_K W_A + [A_K, W_A]. \tag{6}$$

The composition of two such transformations will be of the same form provided the $W_A(X)$ satisfy the constraints

$$\xi_A{}^K \partial_K W_B - [AB]\xi_B{}^K \partial_K W_A - [W_A, W_B] + C_{AB}{}^C W_C = 0, \tag{7}$$

where the brackets are in the Lie (super)algebra of the gauge group.

The consequences of (7) are easily seen in a basis of S such that $\{J_a,\ 1 \leqslant a \leqslant \dim R\}$ generate R, and $\{J_p,\ \dim R + 1 \leqslant p \leqslant \dim S\}$ span S/R. First note that, for isometries $J_a \in R$, $\xi_a{}^K \equiv 0$. Hence, from (7),

$$[W_a, W_b] = C_{ab}{}^c W_c \tag{8}$$

and thus the $\{W_a\}$ generate a (super)algebra $\widetilde{R}$ homomorphic to R in G. Assuming this is X independent, then since also $\partial_\mu \xi_a{}^K = 0$ is always true, from (5) and (6) follows

$$[A_\mu, W_a] = 0, \tag{9}$$

that is, the unbroken gauge group corresponds to the commutant H of $\widetilde{R}$ in G. Finally if $\theta^m \to \exp(\theta^m \delta_m{}^p J_p)$ is the explicit parametrization for cosets S/R we have infinitesimally

$$\exp \epsilon^a J_a \cdot \exp\ \theta^m \delta_m{}^p J_p = \exp(\theta^m - \epsilon^a \theta^r f_{ra}{}^m)\delta_m{}^p J_p \cdot \exp \epsilon^a J_a,$$

so that

$$\partial_m \xi_a{}^k = -f_{ma}{}^k.$$

Substituting in (5) and (6) yields, at $\theta = 0$,

$$f_{ma}{}^k A_k = [A_m, W_a], \tag{10}$$

and, since $A_m \equiv A_m{}^a T_a$, the $\widetilde{D} \times \dim G$ matrix $A_m{}^a$ intertwines the representations $J_a \to \mathrm{ad}\, J_a|_{S/R}$ of R and $W_a \to \mathrm{ad}\ W_a|_G$ of $\widetilde{R} \simeq R$. [The adjoint action $A_m \to f_{ma}{}^k A_k$ is orthosymplectic so R can be regarded as being imbedded in $\mathrm{OSp}(D/\widetilde{D})$, for example. See also Salam and Strathdee.[10]] Applying Schur's lemma, if

$$S \supset R:\ \mathrm{ad}\, S|_{S/R} \downarrow \oplus_\rho \rho_R,$$

$$G \supset \widetilde{R} \times H:\ \mathrm{ad}\, G \downarrow (\mathrm{ad}\, R \otimes 1_H) \oplus \oplus_\sigma (\sigma_R \otimes \lambda_H), \tag{11}$$

then for each $\sigma_R \otimes \lambda_H$ for which there is an equivalent $\rho_R \simeq \sigma_R$ in the decomposition of $\mathrm{ad}\, S|_{S/R}$, there will be a scalar field in the representation λ_H of H. An analysis similar to (11) can be given for matter fields under CSDR, but will not be required here[23] (see also below).

We now formulate some general guiding principles necessary when implementing CSDR with gauged supergroups. First we discuss Hermiticity assignments, and then turn to the CSDR process itself.

(i) Hermiticity: The adjoint operator on superalgebra generators (and hence gauge fields) is determined by the chosen real form. As will emerge below, given an arbitrary complex superalgebra (usually from the list of simple Lie superalgebras[25]), we choose a certain real form[25,26] G (possibly noncompact). If this has basis $\{t_a\}$ consider a unitary representation $t_a \to T_a$. In the associated Cartan–Weyl basis there will be root vectors T_α such that $(T_\alpha)^2 = \epsilon_\alpha T_{-\alpha}$, with $\epsilon_\alpha = +1$ or -1 specified for the different α by the real form in question. Correspondingly, for the associated gauge potential (1), definite Hermiticity will be conferred on the A^α:

$$(A^\alpha T_\alpha)^\dagger = (T_\alpha)^\dagger (A^\alpha)^\dagger = \epsilon_\alpha [\alpha](A^\alpha)^\dagger T_{-\alpha}; \tag{12}$$

i.e.,

$$(A^\alpha)^\dagger = \epsilon_\alpha [\alpha] A^{-\alpha},$$

where $[\alpha] = (-1)^{(\alpha)}$ and (α) is 0 or 1 according as α is even or odd. Reality of the component (super)fields $A_M{}^\alpha$ can then be decided by applying (12) to invariant combinations like $\Phi^M A_M{}^\alpha$ for some fields Φ^M transforming like X^M.

(ii) Residual gauge group H: The residual (unbroken) gauge group H ($G \supset \widetilde{R} \times H$) must be a compact Lie group. Moreover the Killing form $(\, , \,)|_H$ must be positive definite. This is obviously necessary if the final model is to be of conventional Yang–Mills type in D-dimensional space-time. Specifically, the Lagrangian density will contain a term $(F_{\mu\nu}, F^{\mu\nu})_H$, where $F_{\mu\nu}$ is derived from the H-valued part of the potential A_μ, and this must be positive definite. (For noncompact real forms, infinite-dimensional representations are required, for consistency with the adjoint operator. We simply assume that the usual invariant bilinear form is still well-defined for such irreps.) This is not automatic for superalgebras in general: for example, the root spaces[25] are usually described in terms of unit vectors (ϵ, δ) such that $\epsilon \cdot \epsilon = 1$ and $\delta \cdot \delta = -1$. In fact, in the case SU(2/1), proposed as an electroweak unification group,[18,19] the Killing form has a Minkowskian signature, and SU(2/1) $\subset \mathrm{OSp}(3,1/2,R)$. [This difficulty has been overcome[27] by a nonlinear realization of SU(2/1), at the expense of a loss of predictive power (e.g., the weak angle becomes an arbitrary parameter).]

(iii) Higgs sector: In the reductions (11) the gradings of the matching pairs $\rho_R, \sigma_R \otimes \lambda_H$, where $\rho_R \simeq \sigma_R$, must be equal. This is nothing but the requirement that the spin-statistics theorem be obeyed, so that the Lorentz-scalar fields in representations λ_H are true (bosonic) Higgs fields. As to their contribution to the Lagrangian density, there is, of course, the requirement that their kinetic and potential energy occur with the conventional signs. However, this can usually be arranged via the choice of the relative sign of g_{mn} in the S-invariant metric (3). Furthermore, for case (iv) to be studied below, the grading of ρ_R and $\sigma_R \otimes \lambda_H$ is odd, and for most real forms of simple superalgebras,[25,26] the choices $T_{\delta^2} = T_{-\delta}$ or $T_{\delta^2} = -T_{-\delta}$ are both allowed.

(iv) Ghost sector and BRST[28] supersymmetry: The isometries of S/R should include supertranslations on at least one a-number coordinate θ. This requirement is not strictly necessary for the implementation of CSDR, but as part of the

motivation for considering supergroups is to incorporate the ghost fields, it is a logical accompaniment of (i)–(iii) above. Indeed, in the cases to be studied in Sec. III, a slight generalization will dictate the minimal choice of homogeneous space S/R. To see how the ghost fields arise, consider (5)–(7), where there is a Grassmannian coordinate undergoing supertranslations $\theta \to \theta + \epsilon$, i.e., $\xi_\epsilon{}^\theta \equiv 1$. Then

$$\partial_\theta A\mu = D_\mu W_\theta$$

and

$$2\,\partial_\theta W_\theta - \{W_\theta, W_\theta\} = 0,$$

using the fact that the supertranslations anticommute. The explicit solution of the "zero curvature" condition for W_θ is $W_\theta = -U\partial_\theta U^{-1}$, leading to the superfield expansions

$$U = \exp(\theta\omega + \cdots), \quad W_\theta = \omega + \tfrac{1}{2}\theta\{\omega,\omega\} + \cdots,$$

$$A_\mu = a_\mu + \theta D_\mu \omega + \cdots,$$

i.e., the usual BRST transformations appear as transformations among the component fields under supertranslation.[21] Finally, given that the ghost fields are incorporated, a slight generalization of (2) can be made to allow for gauge fixing terms, usually in the form[20,21]

$$S_{\rm gf} \propto \int dx\, d\theta (A^M, A_M). \tag{13}$$

III. EXAMPLES: UNITARY AND ORTHOSYMPLECTIC SUPERGROUPS WITH MINIMAL S/R

It is not practicable to make a systematic study of CSDR for supergroups even with the conditions (i)–(iv) of Sec. II in place, given the large number of choices of starting group G, graded manifold $\mathscr{N}$, and isotropy group embedding. One restricted possibility, taken up here, is that the requirements of BRST symmetry and gauge fixing should dictate the choice of homogeneous space $\mathscr{N} \simeq S/R$. As mentioned above, the ghost fields are associated with superfield expansions with respect to a-number coordinates of the manifold. With $2k$ such, there is a natural symplectic symmetry $\mathrm{Sp}(2k,R)$. Here we take simple $k=1$, and spaces S/R whose isometries include supertranslations on the two Grassmann coordinates (to implement extended BRST transformations) and $\mathrm{Sp}(2,R) \simeq \mathrm{SU}(1,1)$ rotations (so that there is maximum symmetry between ghosts and antighosts[20]). The simplest choice[21] is just the inhomogeneous extension $\mathrm{ISp}(2,R)$ or "Grassmann Euclidean" group[29]; occasionally we take only the Abelian U(1) or "ghost number" part of $\mathrm{Sp}(2,R)$ as the isotropy subgroup. (A larger symmetry[30] carried as "conformal transformations" on θ, $\bar\theta$ is $\mathrm{OSp}(1,1/2,R)$ [with little group $R \simeq \mathrm{O}(1,1) \times \mathrm{ISp}(2,R)$]. The group manifold of the similar U(1) extension is the "supercircle" $\phi, \theta, \bar\theta$, $0 \leqslant \phi \leqslant 2\pi$. For further comments, see the concluding remarks in Sec. IV.)

Let us consider case by case the various real forms[25,26] of the simple superalgebras from the unitary and orthosymplectic series, and their Killing forms (Appendix A), and embeddings of $\mathrm{Sp}(2,R) \simeq \mathrm{SU}(1,1)$ [or simply U(1)] (Appendix B) consistent with (i)–(iv) of Sec. II. The resultant residual gauge groups and Higgs scalar multiplets are summarized in Table I. The real forms are in each case characterized by the real form of the underlying Lie algebra.[26]

For $A(m,n)$, there are the following cases.

$sl(m,R) \oplus sl(n,R) \oplus R$: The real Abelian generator corresponds to noncompact scale transformations, which we do not consider in the present study. Therefore we require $A(n,n) \simeq \mathrm{sl}(n/n)/\text{center}$ with Lie part $\mathrm{sl}(n,R) \oplus \mathrm{sl}(n,R)$. From Appendix A, the SU(1,1) embedding $f = f_1 \oplus f_2$ must be such that either f_1 or f_2 has zero centralizer, e.g., of type (ii) of Appendix B. The other part, f_1 say, must have a compact centralizer. Let the n-dimensional representation of A_{n-1} decompose with respect to the A_1 subalgebra as

$$\{1\} \downarrow \bigoplus_j g_j \{j\}, \quad \sum_j g_j (2j+1) = n, \tag{14}$$

with half-odd integer or integer j. Then there exist subalgebras

$$\begin{aligned}\mathrm{sl}(n) &\supset gR \oplus \bigoplus_j \mathrm{sl}(g_j(2j+1))\\ &\supset gR \oplus \Big[\bigoplus_j (\mathrm{sl}(g_j j) \oplus \mathrm{sl}(2j+1))\Big],\end{aligned} \tag{15}$$

where the R degeneracy g is one *less* than the number of nonzero g_j. Thus the centralizer $gR \oplus [\oplus_j \mathrm{sl}(g_j)]$ is always either noncompact [which is unacceptable by condition (ii) above], or it vanishes altogether.

$SU^*(2m) \oplus SU^*(2n) \oplus R$: As above we require $m=n$. The principal SU(1,1) embeddings in Appendix B do not include $\mathrm{SU}^*(2n)$, although as noted there special cases may occur. However, we expect the same difficulties to occur here as in the $\mathrm{sl}(m,R) \oplus \mathrm{sl}(n,R)$ case above, and so we do not consider this case further.

$SU(p,m-p) \oplus SU(q,n-q) \oplus iR$ (with $m \geqslant n$): The Abelian generator corresponds to compact U(1) transformations. From Appendix A the SU(1,1) [or U(1) embedding in $\mathrm{SU}(p,m-p)$] must have zero centralizer (if $m=n$ then the embedding in either summand is maximal, as above). Suppose a fundamental k-dimensional representation of an A_{k-1} algebra [in its real form $\mathrm{SU}(r,k-r)$] decomposes with respect to an A_1 subalgebra as

$$\{1\} \downarrow \bigoplus_j g_j \{j\}, \quad 2g_{1/2} + \sum_j g_j (2j+1) = k \tag{16}$$

(where $j = \frac{1}{2}$ or integer). Then, from Appendix B, there is an embedding

$$\begin{aligned}\mathrm{SU}(r,k-r) &\supset g\mathrm{U}(1) \oplus [\mathrm{SU}(g_{1/2}, g_{1/2})\\ &\quad \oplus \bigoplus_j (\mathrm{SU}(a_j j, a_j(j+1)) \oplus \mathrm{SU}(b_j(j+1), b_j j))]\\ &\supset g\mathrm{U}(1) \oplus [\mathrm{SU}(g_{1/2}) \oplus \mathrm{SU}(1,1)]\\ &\quad \oplus \bigoplus_j [\mathrm{SU}(a_j) \oplus \mathrm{SU}(j,j+1) \oplus \mathrm{SU}(b_j)\\ &\quad \oplus \mathrm{SU}(j+1,j)],\end{aligned} \tag{17}$$

where

$$g_j = a_j + b_j, \quad g_{1/2} + \sum_j (j a_j + (j+1) b_j) = r,$$

$$g_{1/2} + \sum_j ((j+1) a_j + j b_j) = k - r,$$

and the U(1) degeneracy $\wp$ is one less than the number of nonzero $\wp_j$. The embedding therefore has centralizer

$$H = \wp U(1) \oplus \bigoplus_j [SU(a_j) \oplus SU(b_j)].$$

Returning to the case in hand, clearly the embedding in the SU$(p,m-p)$ must be of the maximal type (ii) of Appendix B (i.e., $m-p=p+1$ or $p-1$), and the residual gauge group [from the embedding of SU(1,1) in the other factor SU$(q,n-q)$] can be of the above reductive form [including an additional U(1) summand if $m>n$]. The simplest possibility is, of course, if $\wp_0 = 1$ and $\wp_{1/2} = \wp_{j>1} \equiv 0$, in which case H is SU$(n) \oplus$ U(1); it is this case which is entered in Table I.

Higgs multiplets for SU(p,m−p) ⊕ SU(q,n−q) ⊕ iR: According to the discussion in Sec. II [c.f. (11)], the Higgs fields come from doublets of SU(1,1) in the decomposition of the adjoint of the real form derived from SU$(p,m-p)$ $\oplus$ SU$(q,n-q) \oplus$ U(1). In the case of an embedding of A_1 in A_{n-1} of the form (16) for the fundamental representation,

$$\{\bar{1};1\} \downarrow [\oplus_{j,k} \wp_j \wp_k \{j\} \otimes \{k\}] - \{0\},$$

for the adjoint representation. If $\wp_{1/2} \neq 0$, then doublets (which would correspond to wrong-statistics fields) must be forbidden by ensuring $\wp_0 = \wp_1 \equiv 0$. If $\wp_{1/2} = 0$, then only integer spins appear. For the odd generators (transforming as $\{1\} \otimes \{\bar{1}\} \oplus \{\bar{1}\} \otimes \{1\}$ under $A_{m-1} \oplus A_{n-1}$) the A_1 content is

$$2\{p\} \otimes \left[\bigoplus_j \wp_j \{j\}\right].$$

Thus doublets appear provided $p=0$ or 1 and $\wp_{1/2} \neq 0$, or $p=\frac{1}{2}$ and $\wp_0 = 0$ or $\wp_1 \neq 0$ (and $\wp_{1/2} = 0$). In the first case, $p=0$, $\wp_{1/2} \neq 0$ contradicts $m>n$, while, for $p=1$, $\wp_{1/2}=1$, SU(2,1/1,1) gives simply a U(1) gauge group with complex Higgs. This case is again reproduced by $p=\frac{1}{2}$, $\wp_0 = 1$, while, for $p=\frac{1}{2}$, $\wp_0 = 2$, i.e., SU(1,1)/center, we have an SU(2) group with complex Higgs doublet [if the central extension is included then this becomes SU(2) $\oplus$ U(1)]. For the simple case SU$(j+1,j/n)$ listed in Table I $(2j+1>n>2)$, the gauge group is SU$(n) \oplus$ U(1) but no SU(1,1) doublets, and thus no Higgs scalar fields, appear. [More exotic embeddings, for example, SU(2,2) $\simeq$ O(4,2) $\supset$ O(3,2) $\supset$ SU(1,1), such that $\{1\} \downarrow \{\frac{3}{2}\}$ (see Appendix B), or real forms of SU$(mn) \supset$ SU$(m) \oplus$ SU(n), have not been exhaustively studied.] This is also the case for more involved embeddings such as (17): an examination of low-dimensional cases (taking into account the restrictions on $\wp_{1/2}$), and an inductive proof on $m \geqslant n$, show that the SU(1,1) content will always have spin at least 1 or $\frac{3}{2}$. (Part of the root cause of this disappointing result is the restrictive choice of homogeneous space S/R. See Sec. IV for comments.)

sl(n,C): This is a special case of $A(n-1,n-1)$ where one A_{n-1} factor provides the noncompact generators. Let the fundamental n-dimensional representation of A_{n-1} (as a complex algebra) decompose as before as

$$\{1\} \downarrow \bigoplus_j \wp_j \{j\}, \quad j = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots,$$

under the A_1 subalgebra. Then from (ii) of Appendix B there exists the chain [c.f. (15) and (17) above]

$$\begin{aligned} \mathrm{sl}(n,C) &\supset \wp C \oplus \left[\bigoplus_j \mathrm{sl}(\wp_j(2j+1),C)\right] \\ &\supset \wp C \oplus \left[\bigoplus_j \mathrm{sl}(\wp_j,C) \oplus \mathrm{sl}(2,C)\right] \\ &\downarrow \wp C \oplus \left[\bigoplus_j [\mathrm{sl}(\wp_j,C)] \oplus \mathrm{sl}(2,R)\right], \end{aligned} \tag{18}$$

with the C degeneracy being one *less* than the number of nonzero $\wp_j$. Thus the centralizer is either noncompact, or it vanishes altogether.

For $C(n)$ and $B(0,n)$, there are the following cases.

OSp(2/2n,R): For the present discussion this is a special case of OSp$(m/2n)$ below.

OSp(1/2n,R): See also OSp$(m/2n)$ below. However, in this case the Killing form is positive on the *whole* of the even subalgebra, and we can use suitable embeddings of the smaller isotropy group U(1), for example, corresponding to

$$\langle 1 \rangle \downarrow \{1\} \otimes \{+1\} \oplus \{\bar{1}\} \otimes \{-1\}.$$

In this case H is SU$(n) \oplus$ U(1) and the Higgs fields are a complex $\{1\}$. [It may also be possible to break SU(n) further, viz., SU$(p) \oplus$ SU$(n-p) \oplus$ U(1) $\oplus$ U(1).]

For $B(m,n)$, $m>0$ and $D(m,n)$, $m>1$, there are the following cases.

O(2m) ⊕ Sp(2p,2n−2p):* Positivity of the Killing form (from Appendix A) requires a maximal SU(1,1) embedding in O* $(2m)$ or in Sp$(2p,2n-2p)$. Examination of low-dimensional cases, e.g.,

$$\mathrm{O}^*(4) \simeq \mathrm{SO}(2,1) \oplus \mathrm{SO}(2,1),$$
$$\mathrm{O}^*(6) \simeq \mathrm{SU}(3,1), \quad \mathrm{O}^*(8) \simeq \mathrm{O}(6,2),$$
$$\mathrm{Sp}(2,2) \simeq \mathrm{SO}(4,1),$$

suggests that a maximal embedding of the type given in Appendix B is ruled out for this class of real forms (for further comments, see Appendix B). [The same applies to the real form of $C(n)$ corresponding to O(2) $\oplus$ Sp$(2p,2n-2p)$.] As mentioned above, exceptions may occur for higher-rank cases, especially in connection with spinor representations.

O(p,m−p) ⊕ Sp(2n,R): Using the embedding types (i) and (ii) of Appendix B, there is now plenty of scope for maximal and nonmaximal embeddings. Rather than elaborate in detail [c.f. (14)–(18) above], Table I contains a simple case where the resulting gauge group is O(k) and the Higgs fields are in the fundamental, k-dimensional representation [1] of O(k). [Use must be made of the embeddings[31]

$$\mathrm{O}(mn) \supset \mathrm{O}(m) \oplus \mathrm{O}(n), \quad \mathrm{Sp}(2mn) \supset \mathrm{O}(m) \oplus \mathrm{Sp}(2n).]$$

IV. CONCLUSIONS

In the foregoing (in Sec. II) we have set out general guiding principles for constructing consistent unified models from gauged supergroups via CSDR, and (in Sec. III) have attempted a comprehensive search of models for one specific homogeneous space (parametrized by a single pair of Grassmann coordinates). The results in Table I generalize earlier work[22] and indicate, first, that the CSDR formalism (see Sec. II) does admit consistent solutions, and, second, that there is a variety of admissible GUT groups. The latter may be nonsimple (see the more elaborated solutions in Sec. III not summarized in Table I), and the advantage of CSDR is that there is only one gauge coupling (if a simple superalgebra is used as starting material). Furthermore, the Higgs sector, and the (classical) mass spectrum consequent on the spontaneous symmetry breaking thereof, are, in principle, predetermined.

In the present study no detailed computations of such mass and coupling relations are given. It is sufficient to note that such constraints exist in principle for the models listed in Table I (further details are given for the exceptional superalgebras studied in the sequel to this paper[32]). The general situation is similar to the particular case studied in earlier work[22]: as noted there, there is no natural mass parameter in the θ, $\bar{\theta}$ a-number manifold, and hence no tree level spontaneous symmetry breaking. In the absence of a calculable dynamical scheme, concrete predictions are therefore difficult. In any case, all CSDR schemes suffer the defect that higher-loop corrections tend to destroy the classical predictions.[33] Only the presence of conventional Fermi–Bose supersymmetry is likely to ameliorate this difficulty (just as it can do for hierarchy problems in conventional GUT's[11]); however, such additional structure is somewhat anathema to the CSDR approach.

Table I shows that for the $SU(n) \oplus U(1)$-type models (see, also, Sec. III) derived from $SU(j, j+1/n)$, no Higgs multiplets occur. Although this may be seen as a failing, it may also turn out to be an advantage if elementary scalars are to be avoided, or if they are expected to arise from supersymmetry. However, even in this case, the scheme does not lack some predictive power: matter fields [*a priori* in multiplets of $SU(j, j+1/n)$] are constrained by CSDR to occur in specific $SU(n) \oplus U(1)$ representations (leading perhaps to exotic particles, or certain generation structures).

It should be emphasized that some of the limitations of the present models may be a reflection of the minimal choice (Sec. III) of homogeneous space. In particular, graded spaces with extra *c-number* coordinates (as well as *a-number* coordinates of the type used here) [e.g., $OSp(2/2)/OSp(1/2)$ with coordinates $(\phi;\xi,\bar{\xi})$, $0\leqslant\phi\leqslant 2\pi$; $OSp(3/2)/OSp(2/2)$ with coordinates $(\theta,\phi;\xi,\bar{\xi})$, $0\leqslant\theta\leqslant\pi$, $0\leqslant\phi\leqslant 2\pi$] contain natural length scales and may provide tree level mass spectra. An extension with bosonic ghosts was given in Ref. 34.

Aside from extensions to Poincaré supersymmetry, it is clearly important to generalize the present work to include general coordinate invariance, and utilize the body of existing work on symmetry aspects of (conventional) Kaluza–Klein theories and invariant metrics.[10,17] In studies of a-number Kaluza–Klein models to date, a natural invariant metric ansatz has been used that, with appropriate torsion constraints, provides an alternative route to the standard model[35]; meanwhile the CSDR formalism [for the $ISp(2,R)/Sp(2,R)$ space] has been extended to vielbein gauge fields.[36]

Finally it is of interest to consider the present work in a superstring or supermembrane context. Coset spaces have been considered as alternatives to Calabi–Yau compactifications,[37] although the presence of continuous isometries is generally unwanted. The philosophy of the present approach, with the central role played by BRST symmetry, suggests that the appropriate starting point is the ubiquitous affine exceptional group structures in superstring spectra,[38] where BRST symmetry dictates the appropriate extended Lorentzian lattice for the (bosonized) ghost field content.

ACKNOWLEDGMENTS

It is a pleasure to thank Bob Delbourgo for his guidance and insights over several years in helping to nurture the ideas of the present paper. Present and past collaborators to whom I am also indebted include Peter Dondi, George Thompson, Richard Farmer, Jenny Henderson, Simon Twisk, and Ruibin Zhang.

This work was supported by an Australian Research Council chief investigator grant. Ongoing and previous grants from the ARC, the Alexander von Humboldt Foundation, and contracts from the University of Tasmania are gratefully acknowledged.

APPENDIX A: KILLING FORMS AND UNITARY AND ORTHOSYMPLECTIC SUPERALGEBRAS

The classical (special) unitary and orthosymplectic superalgebras can be conveniently characterized by giving their generators and (anti)commutation relations in matrix form. In each case we give the adjoint operation pertaining to the compact real form, and determine the signature of the Killing form by examining the second-order Casimir operator in this basis.

(i) SU(m/n): The generators are $E^A{}_B = (E^B{}_A)^\dagger$, $\Sigma_A E^A{}_A = 0$, $A,B = 1,\ldots,m+n$. The $SU(m/n)$ superalgebra in this basis is

$$[E^A{}_B, E^C{}_D] = \delta^C{}_B E^A{}_D - \left[{}^{AC}_{BD}\right]\delta^A{}_D E^C{}_B,$$

where $[XY] = -1$ if the quantities X, Y are both odd, and otherwise $+1$; this convention is here applied to $A+B$ and $C+D$. We take $[A] = +1$ for $A = 1,\ldots,m$, and -1 for $A = m+1,\ldots,m+n$.

The $SU(m) \oplus SU(n) \oplus U(1)$ generators can be identified as (repeated indices summed)

$$SU(m): \; A^a{}_b = E^a{}_b - \delta^a{}_b E^c{}_c/m,$$
$$a,b,c = 1,\ldots,m;$$
$$SU(n): \; A^\alpha{}_\beta = E^\alpha{}_\beta - \delta^\alpha{}_\beta E^\gamma{}_\gamma/n,$$
$$\alpha,\beta,\gamma = m+1,\ldots,m+n;$$
$$U(1): \; Z = \sum_c E^c{}_c = -\sum_\gamma E^\gamma{}_\gamma.$$

The quadratic Casimir invariant is[39]

$$C_2 = E^A{}_B [B] E^B{}_A = A^a{}_b A^b{}_a - A^\alpha{}_\beta A^\beta{}_\alpha - (m-n)/mnZ^2 + (\text{odd}),$$

showing that in all cases the Killing form is indefinite, with the sign of the compact $U(1)$ part the same as the $SU(n)$ $(m>n)$, $SU(m)$ $(m<n)$, contribution. If $n=1$, then the $A^\alpha{}_\beta$ terms are absent, but then the Z^2 term is always negative (if $m>1$). If $m=n$, the above analysis still applies but the Z generator is dropped from the superalgebra in the $A(n,n)$

$\simeq SU(n/n)$/center case.

(ii) OSp(m/n): The generators are $J_{AB} = -[AB]J_{BA}$, $A,B = 1,...,m+n$, n even. The OSp(m/n) superalgebra in this basis is[39]

$$[J_{AB},J_{CD}] = \eta_{BC}J_{AB} - [ab]\eta_{AC}J_{BD} - [CD]\eta_{BD}J_{AC} + [AB][CD]\eta_{AD}J_{BC},$$

where $\eta_{AB} = [AB]\eta_{BA}$ is even and graded symmetric. The $O(m)\oplus Sp(n)$ generators are simply the subsets $J_{ab} = -J_{ba}$, $a,b = 1,...,m$, and $J_{\alpha\beta} = J_{\beta\alpha}$, $\alpha,\beta = m+1,..., m+n$, for which we impose $(J_{ab})^\dagger = J^{ba}$, $(J_{\alpha\beta})^\dagger = -J^{\beta\alpha}$ for the compact real forms (this follows from an embedding of the type

$$J_{AB} = \eta_{AC}E^C{}_B - [AB]\eta_{BC}E^C{}_A$$

in $U(m/n)$ and assuming $\eta_{ab} = \eta^{ab}$, $\eta_{\alpha\beta} = -\eta^{\alpha\beta}$ for the metric).

The quadratic Casimir invariant is simply[39]

$$J^{AB}J_{BA} = \eta^{AC}\eta^{BD}J_{CD}J_{BA} = J^{ab}J_{ba} + J^{\alpha\beta}J_{\beta\alpha} + (\text{odd})$$

and so the Killing form is again of indefinite sign (except in the case $m=1$). Taking noncompact real forms will not change the signature for the compact generators, but will change the sign of the noncompact ones. However, as detailed in Secs. II and III, the latter are not gauged as it is arranged that the residual gauge group H is compact.

APPENDIX B: SU(1,1) EMBEDDINGS

We take SU(1,1) generators e_+, e_-, h with commutation relations

$$[e_+,e_-] = 2h, \quad [e_\pm,e_\pm] = 0, \quad [h,e_\pm] = \pm e_\pm,$$

and adjoint operation $h^\dagger = h$, $(e_+)^\dagger = -e_-$. By an *embedding* of SU(1,1) in a Lie algebra G over R is meant an algebra homomorphism

$$f\colon SU(1,1)\to G,\ e_\pm \to E_\pm,\ h\to H,$$

$$[H,E_\pm] = \pm E_\pm, \text{ etc.},$$

and such that the adjoint operation (associated with the real form of G) is compatible: $H_2 = H$, and $(E_+)_2 = -E_-$. The principal SU(1,1) embeddings[31] used in Sec. III arise by considering the matrix elements of the $e\pm$, h in finite-dimensional representations, namely, for $j = 0, \frac{1}{2}, 1,...,$

$$E_\pm |j,m\rangle = [j\mp m]^{1/2}[j\pm m+1]^{1/2}|j,m\pm 1\rangle,$$

$$H|j,m\rangle = m|j,m\rangle.$$

(i) $SU(1,1)\subset SO(j+1,j)\subset SU(j+1,j)$, $j=1,2,...$: Consider the above matrix elements in the basis

$$\{|j,-j\rangle,|j,-j+2\rangle,...,|j,j\rangle;\ |j,-j+1\rangle,...,|j,j-1\rangle\}.$$

Then it is easily seen that as real matrices of dimension $(2j+1)$ they fulfill $X^TG + GX = 0$, where G in block form is $\mathrm{diag}(\tilde{1}_{j+1}, -\tilde{1}_j)$ and $\tilde{1}_n$ is the $n\times n$ *antidiagonal* unit matrix. This means that $E_\pm$, H are linear combinations of generators of $SO(j+1,j)\subset SU(j+1,j)$.

(ii) $SU(1,1)\subset Sp(2j+1,R)\subset SL(2j+1,R)$, $j=\frac{1}{2},\frac{3}{2}$ ... : Consider the SU(1,1) matrix elements in the basis

$$\{|j,-1j\rangle,|j,-j+2\rangle,...,|j,j-1\rangle;\ |j,j\rangle,|j,j-2\rangle,...,|j,-j+1\rangle\}.$$

Then it is easily seen that as real matrices of dimension $(2j+1)$ they fulfill $X^TG + GX = 0$, where G in block form is antidiag $(-1_{j+1/2}, 1_{j+1/2})$ with 1_n the $n\times n$ unit matrix. This means that $E_\pm$ are linear combinations of generators of $Sp(2j+1,R)\subset SL(2j+1,R)$.

Finally, cases where the principal embeddings (i) and (ii) are not maximal require additional study. For example, the possibility $SU(1,1)\subset SO^*(2n)\subset Sp(2^{n-1},R)$, which may arise for appropriate symplectic spinor representations of $SO^*(2n)$, would allow different real forms of $D(m,n)$ to be used than those considered in Sec. III. Similarly, intervening $SU^*(2n)$ algebras may enlarge the choice of $A(m,n)$-type models.

[1] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **28**, 1491 (1972); J. Pati and A. Salam, *ibid.* **31**, 661 (1973); H. Fritzsch and P. Minkowski, Ann. Phys. (NY) **93**, 193 (1975).

[2] For a comprehensive introduction to grand unified models see, for example, C. Kounnas, A. Masiero, D. V. Nanopoulos, and K. A. Olive, *Grand Unification With and Without Supersymmetry and Cosmological Implications* (World Scientific, Singapore, 1984).

[3] For ground rules for GUT models, see L. O'Raifeartaigh, *Group Structure of Gauge Theories* (Cambridge U.P., Cambridge, 1986).

[4] M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978); R. Slansky, Phys. Rep. **79**, 1 (1981); R. C. King and D. Grech, Fortschr. Phys. **33**, 269 (1985).

[5] See, for example, *Sixth Workshop on Grand Unification*, edited by S. Rudasz and T. Walsh (World Scientific, Singapore, 1986).

[6] R. K. Kaul, Rev. Mod. Phys. **55**, 449 (1983).

[7] R. R. Volkas and G. C. Joshi, Phys. Rep. **159**, 303 (1988).

[8] P. Fayet, Phys. Rep. **105**, 21 (1984).

[9] See, for example, D. V. Nanopoulos, "The physics of supersymmetry and supergravity," in Ref. 2, p. 283.

[10] E. Witten, Nucl. Phys. B **186**, 412 (1981); A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982); W. Mecklenburg, Fortschr. Phys. **32**, 207 (1984); M. J. Duff, B. E. W. Nilsson, and C. N. Pope, Phys. Rep. **130**, 1 (1986).

[11] G. G. Ross, "Realistic models of supersymmetry," in *Supersummetry—a Decade of Development*, edited by P. West (Hilger, London, 1986), p. 374.

[12] P. Forgacs and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980); A. S. Schwarz, *ibid.* **56**, 79 (1977); E. Witten, Phys. Rev. Lett. **38**, 121 (1977); A. S. Schwarz and Yu. S. Tyupkin, Nucl. Phys. B **187**, 321 (1981); J. Harnad, S. Shnider, and J. Tafel, Lett. Math. Phys. **4**, 41 (1980); L. B. Hudson and R. Kantowski, J. Math. Phys. **25**, 3094 (1984); G. Rudolph and I. Volobujev, Theor. Math. Phys. **62**, 261 (1985) [Teor. Mat. Fiz. **62**, 388 (1985)].

[13] N. S. Manton, Nucl. Phys. B **158**, 141 (1979); A. S. Schwarz and Yu. S. Tyupkin, Ref. 12; G. Chapline and N. S. Manton, Nucl. Phys. B **184**, 391 (1981); N. S. Manton, *ibid.* **193**, 502 (1981); G. Chapline and R. Slansky, *ibid.* **209**, 461 (1982); G. Chapline and B. Grossmann, Phys. Lett. B **143**, 161 (1984).

[14] D. I. Olive and P. C. West, Nucl. Phys. B **217**, 248 (1983); L. Palla, Z. Phys. C **24**, 195 (1984); P. Forgacs and G. Zoupanos, Phys. Lett. B **148**, 99 (1984); D. Lüst and G. Zoupanos, *ibid.* **165**, 309 (1985); K. Farakos, G. Koutsoumbas, M. Surridge, and G. Zoupanos, Nucl. Phys. B **291**, 128 (1987); G. Zoupanos, Phys. Lett. B **201**, 322 (1988); K. Farakos, D. Kapetanakis, G. Koutsoumbas, and G. Zoupanos, *ibid.* **211**, 301 (1988).

[15] F. A. Bais, K. J. Barnes, P. Forgacs, and G. Zoupanos, Nucl. Phys. B **263**, 557 (1986); K. J. Barnes and M. Surridge, Z. Phys. C **33**, 89 (1986); K. J. Barnes, R. C. King, and M. Surridge, Nucl. Phys. B **281**, 253 (1987); K. J. Barnes, P. Forgacs, M. Surridge, and G. Zoupanos, Z. Phys. C **33**, 437 (1987).

[16] I. P. Volobujev and Yu. A. Kubyshin, Theor. Math. Phys. **68**, 778 & 885 (1986) [Teor. Mat. Fiz. **68**, 225 & 368 (1986)]; Yu. A. Kubyshin and I.

P. Volobujev, "Weinberg–Salam model in the dimensional reduction scheme," in *Quarks '86*, edited by A. N. Tavkhelidze *et al.* (VNU Science, Utrecht 1987); N. G. Kozimirov and I. I. Tkachev, Z. Phys. C **36**, 83 (1987); Yu. A. Kubyshin, J. M. Mourão, and I. P. Volobujev, Phys. Lett. B **203**, 349 (1988).
[17]F. Müller-Hoissen and R. Stückl, Class. Quant. Grav. **5**, 27 (1988); F. Müller-Hoissen, *ibid.* **4**, L 143 (1987).
[18]S. A. Wouthuysen, in *Current Trends in the Theory of Fields*, edited by J. E. Lannutti and P. K. Williams (AIP, New York, 1978), p. 153; Y. Ne'eman, Phys. Lett. **818**, 190 (1979); D. B. Fairlie, *ibid.* **828**, 97 (1979); J. G. Taylor, Phys. Lett. B **83**, 331 (1979); J. G. Taylor, Phys. Rev. Lett. **43**, 826 (1979); Y. Ne'eman and S. Sternberg, Proc. Natl. Acad. Sci. USA **77**, 3127 (1980).
[19]P. H. Dondi and P. D. Jarvis, Phys. Lett. B **84**, 75 (1979); Z. Phys. C **4**, 201 (1980).
[20]L. Bonora and M. Tonin, Phys. Lett. B. **98**, 48 (1981); L. Bonora, P. Pasti, and M. Tonin, Nuovo Cimento A **64**, 307 (1981); R. Delbourgo and P. D. Jarvis, J. Phys. A **15**, 611 (1982); R. Delbourgo, P. D. Jarvis, and G. Thompson, Phys. Rev. D **26**, 775 (1982).
[21]R. Delbourgo, R. J. Farmer, and P. D. Jarvis, Phys. Lett. B **123**, 319 (1983).
[22]R. Delbourgo and P. D. Jarvis, Phys. Lett. D **28**, 2122 (1983).
[23]N. S. Manton, Ann. Phys. (NY) **167**, 328 (1986).
[24]A. Rogers, J. Math. Phys. **21**, 1352 (1980); Commun. Math. Phys. **105**, 375 (1986); J. M. Rabin and L. Crane, *ibid.* **100**, 141 (1985); **102**, 123 (1985); B. S. de Witt, *Supermanifolds* (Cambridge, U.P. Cambridge, 1984).
[25]V. G. Kac, Adv. Math. **26**, 8 (1977).
[26]M. Parker, J. Math. Phys. **21**, 689 (1980).
[27]S. Deguchi, J. Phys. A **22**, 227 (1989).
[28]C. Becchi, A. Rouet, and R. Stora, Commun. Math. Phys. **42**, 127 (1975); I. V. Tyutin, FIAN report 39, 1975.
[29]R. Delbourgo and J. A. Henderson, J. Phys. A **20**, 529 (1987).
[30]W. Siegel and B. Zwieback, Nucl. Phys. B **282**, 125 (1987).
[31]M. Lorente and B. Gruber, J. Math. Phys. **13**, 1639 (1972); E. Dynkin, Am. Math. Soc. Transl. Ser. 2 **6**, 111 (1957); A. I. Mal'cev, Am. Math. Soc. Transl. **33**, 154 169 (1950); *ibid.* **9**, 172 (1962).
[32]P. D. Jarvis, "Unified models from gauged supergroups. II. Exceptional series $D(2,1;\alpha)$, $G(3)$, $F(4)$" (in preparation).
[33]J. A. Henderson and P. D. Jarvis, Nucl. Phys. B **297**, 539 (1988).
[34]P. D. Jarvis, Int. J. Mod. Phys. A **4**, 401 (1989).
[35]R. Delbourgo and R. B. Zhang, Phys. Lett. B **202**, 296 (1988); Phys. Rev. D **38**, 2470 (1988); R. Delbourgo, R. B. Zhang, and S. E. Twisk, Mod. Phys. Lett. A **3**, 1073 (1988).
[36]P. D. Jarvis, Class. Quant. Grav. **5**, 619 (1988).
[37]T. R. Govindarajan, A. S. Joshipura, S. D. Rindani, and U. Sarkar, Phys. Rev. Lett. **57**, 2489 (1986); D. Lüst, Nucl. Phys. B **276**, 220 (1986).
[38]See, for example, W. Lerche, A. N. Schellekens, and N. P. Warner, "Lattices and strings," Phys. Rep. (to appear), and references therein.
[39]P. D. Jarvis and H. S. Green, J. Math. Phys. **20**, 2115 (1979).

Trapped surfaces in monopole-like Cauchy data of Einstein–Yang–Mills–Higgs equations

Edward Malec
ICTP, P.O.B. 586, Trieste, Italy and Institute of Physics, Jagellonian University, Reymonta 4, 30-064 Cracow, Poland
Piotr Koc
Institute of Physics, Jagellonian University, Reymonta 4, 30-064 Cracow, Poland

(Received 29 August 1989; accepted for publication 13 December 1989)

The non-Abelian monopole solution of Bogomolny, Prasad, and Sommerfield is chosen as a part of the Cauchy data for the evolution of Einstein–Yang–Mills–Higgs equations. Momentarily static spherically symmetric data for gravitational fields are obtained numerically via the Lichnerowicz equation. In the case of generic scaling of fields initial data with trapped surfaces were found.

I. INTRODUCTION

Recently, several authors attempted to find field-theoretic configurations that collapse to a black hole.[1–3] To solve this problem completely for a given model one should be able to study the global in-time evolution of a system; this is not feasible in general, unfortunately.

An alternative would be to use some qualitative criteria, e.g., the hoop conjecture[4] that is a generalization of the condition $m > 2r$ (in Schwarzschild coordinates) for the existence of black holes in spherical systems. That was the strategy used in Ref. 2 and to some extent also in Ref. 3. The point is, however, that for noncompact distributions of matter the precise form of the hoop conjecture is not known while the standard version fails.[5] Another qualitative condition has been proven recently[6] for particular geometries with coefficients that are precisely defined; in contrast to the hoop conjecture that is expressed in terms of the ADM mass and the largest circumference of a body, this new criterion uses the rest mass and the proper radius. It is necessary to point out that both criteria need information that can be obtained only via integration of a part of the Einstein-matter equations; for instance, it is sufficient to solve the constraint equations at a particular instant of time.

Yet another possibility would be to solve the initial data for gravitational and matter fields and to look for trapped surfaces on the Cauchy surface. Trapped surfaces prove the existence of black holes, modulo certain reservations related to the validity of the cosmic censorship hypothesis.[7] If we restrict ourselves to data that are momentarily static, then the problem of finding trapped surfaces of a particular shape is not more laborious than application of the above aforementioned criteria while the information we get is more conclusive.

Certainly the assumption of momentarily static initial data is quite restrictive and most black holes would never meet that condition in their history. But, on the other hand, the scenario with symmetric time-initial data strongly appeals to physical intuition. Intuitively it is almost obvious that if one takes a static configuration and adds more and more matter inside a fixed volume then at a certain stage an apparent horizon should surround a center of the region; the system ceases to be static at later times and collapses to a black hole.

Our aim is to apply the above approach to the study of a magnetic monopole configuration of Einstein–Yang–Mills–Higgs theory. Magnetic monopole is a specific field theory configuration with substantial energy concentrated in a small volume and it is natural to ask whether it can collapse to a black hole. This possibility was raised by Hiscock.[8] Such a small black hole with a magnetic charge can be stable against evaporation via the Hawking mechanism.[8]

To be specific, we have studied the Bogomolny–Prasad–Sommerfield[9] (BPS) monopole coupled to gravitation. An important part of this work is a definition of a gravitational analog of the BPS sholution. We have employed the conformal approach to initial data of Einstein equations.[10]

There are many ways to define the gravitational BPS monopole as a result of an arbitrariness in scalings of spherically symmetric fields. That point is explained in Sec. II. We have discussed two most obvious cases to get quite different results; while Sec. III contains a description of initial data that are proven to contain trapped surfaces, Sec. IV describes initial data with no apparent horizons. Theorems of Sec. IV were proven elsewhere,[11] but here the absence of trapped surfaces is proven in a different and simpler way. As an aside question let us remark that we have found numerically the Wheeler's "bag of gold configuration" (reminded recently in Ref. 12) in each of the two scalings. We do not discuss this point as it has been explained in detail elsewhere.[11]

II. CONFORMAL APPROACH TO MOMENT IN TIME SYMMETRY INITIAL DATA OF EINSTEIN–YANG–MILLS–HIGGS EQUATIONS

We will study a system of Einstein–Yang–Mills–Higgs equations

$$(D_\mu F^{\mu\nu})^a = 0, \tag{1a}$$

$$(D_\mu D^\mu \phi)^a = 0, \tag{1b}$$

$$|\phi|_\infty = M,$$

$$R^{\nu\mu} - g^{\nu\mu} R/2 = 8\pi T^{\nu\mu}. \tag{1c}$$

 0022-2488/90/071791-05$03.00

Here D_μ is the double (gauge and gravity) covariant derivative, the non-Abelian strength field tensor is given by

$$F^{a\mu\nu} = \partial^\mu A^{a\nu} - \partial^\nu A^{a\mu} + \epsilon^{abc} A^{b\mu} A^{c\nu}, \tag{2}$$

where $A^{a\mu}$ denotes a gauge potential and ϵ^{abc} is the completely antisymmetric Levi-Civita tensor. Here, $T^{\mu\nu}$ is the energy-momentum tensor of the Yang–Mills–Higgs field,

$$T^{\mu\nu} = F^{a\mu\gamma} F^{a\nu}{}_{\gamma} - g^{\mu\nu} F^{a\alpha\beta} F^{a}{}_{\alpha\beta}/4 + (D^\mu\phi)^a (D^\nu\phi)^a - g^{\mu\nu}(D_\gamma\phi)^a (D^\gamma\phi)^a/2, \tag{3}$$

$R^{\mu\nu}$ is the Ricci tensor, ϕ denotes a scalar field in the adjoint representation of the SU(2) gauge group, and M is its (positive) length at spatial infinity. The spatial part of the metric $(g^{\mu\nu})$ is assumed to be positively definite.

Cauchy data for the evolution of the system (1) consist of A^{a}_{i}, F^{a0k}, ϕ^a, $d\phi/dt$, which are prescribed on a surface Σ with a three geometry defined by a metric tensor g_{ik} and with an immersion of Σ into a four-dimensional space-time given by an external curvature K^{ij}. This whole set of functions is not arbitrary, since they have to satisfy the initial value equations

$$R^{0\mu} - g^{0\mu} R/2 = 8\pi T^{0\mu}, \tag{4a}$$

$$(D_i F^{i0})^a = \epsilon^{abc} \phi^b (D^0\phi)^c. \tag{4b}$$

These equations are hightly nonlinear, so it is a nontrivial task to find a configuration admitted by Eqs. (4a) and (4b).

For time-symmetric data, which we are interested in, $F^{ai0} = 0$, $K^{ij} = 0$, and the initial Einstein constraints reduce to the single equation

$$R^{00} - g^{00} R/2 = 8\pi T^{00}. \tag{5}$$

We will employ the well-known conformal approach[10] to choose initial data that satisfy Eq. (5). According to this, we assume

$$g_{ij} = f^4 \eta_{ij}, \tag{6}$$

where f solves the following equation:

$$\Delta f = -2\pi(\rho_h f^\alpha + \rho_{\mathrm{YM}} f^\beta), \tag{7}$$

(Δ is the flat Laplacian) with the boundary condition at spatial infinity $f = 1$.

In (7) ρ_h, ρ_{YM} are energy densities of instantly at rest Higgs and Yang–Mills configurations on Σ. We assume them to be given by the BPS solution in Minkowski space, i.e.,

$$A^{a0} = 0,$$
$$A^{a}_{j} = \epsilon^{aij} x_i [-1 + Mr/\mathrm{sh}(Mr)]/r^2,$$
$$\phi^a = M x^a [\coth(Mr) - 1/(Mr)]/r \tag{8}$$

and

$$\rho_h = (D_i\phi)^a (D^i\phi)^a/2,$$
$$\rho_{\mathrm{YM}} = F^{a}_{ij} F^{aij}/4,$$

where

$$\rho_h = \rho_{\mathrm{YM}} = T_{00} = r^{-4} + M^4 \frac{2\,\mathrm{ch}^2(Mr)+1}{\mathrm{sh}^4(Mr)} - M^3 \frac{4\,\mathrm{ch}(Mr)}{r\,\mathrm{sh}^3(Mr)}. \tag{9}$$

If f is a solution of Eq. (7) then the data

$$K^{ij} = 0, \quad g_{ij} = f^4 \eta_{ij}, \quad F^{ai0} = 0,$$
$$\mathscr{F}^{aij} = f^{(\alpha+\beta)/2} F^{aij},$$
$$(D_i\Phi)^a = f^{(\alpha-1)/2} (D_i\phi)^a, \tag{10}$$

will satisfy Eq. (5), that is also the full system of initial constraints (4a), because of time symmetry. The Yang-Mills initial constraint (4b) for rescaled fields is satisfied trivially, since $F^{ai0} = 0$ and $\epsilon^{abc}\phi^b(D^0\phi)^c = 0$ for the flat BPS solution. There is no initial value constraints for scalar fields. Therefore, the above data supplemented by $d\phi/dt = 0$ give rise to the time evolution of Einstein–Yang–Mills–Higgs equations. It is an obvious remark that at spatial infinity as well as in the limit of vanishing gravitation, the Yang–Mills–Higgs counterpart of the solution coincides with the BPS monopole.

There is freedom in scaling initial data, according to various values of conformal factors α,β in Eq. (10). Two possibilities, however, appear to be most attractive. Let us recall after Ref. 13 that the option with $\beta = -3$ is the unique one. Now we can assume $\alpha = \beta$ ($= -3$) to have the simplest possible Eq. (7); that case was studied in Ref. 11. The second possibility is still to retain $\beta = -3$ but to take $\alpha = 1$, which is the most plausible option for generic nonspherical fields.[13] In this paper we will study just the second case, with $\beta = -3$, $\alpha = 1$, although the configuration is spherically symmetric.

Our intention is to find a solution f of Eq. (7) that tends to 1 at spatial infinity. The energy densities behave asymptotically like r^{-4}, thus very far from the center of the configuration a solution of Eq. (7) would be well approximated by the Reissner–Nordström solution

$$f = (1 + A/r + B/r^2)^{1/2}.$$

Inside the region of high density, however, no solution is explicitly available; to get a solution we have to use numerical methods. It will be convenient to introduce new variables,

$$x = Mr \tag{11a}$$

and

$$u = xf/M^{1/2}. \tag{11b}$$

Equation (7) now reads (assuming $\alpha = 1, \beta = -3$)

$$u'' = -\pi x^4 \rho u^{-3} [1 + (u/x)^4 M^2], \tag{12}$$

where

$$\rho = 1/x^4 + (2\,\mathrm{ch}^2(x) + 1)/\mathrm{sh}^4(x) - 4\,\mathrm{ch}(x)/(x\,\mathrm{sh}^3(x)).$$

In (12) prime denotes differentiation with respect to the variable x.

A qualitative investigation shows that near the origin $u = x$ const $+ o(x)$. Thus the initial date $u(0) = 0$, $u'(0) \neq 0$ are acceptable, in accordance with our knowledge about a solution of the elliptic equation (7), which has to achieve a global maximum at $r = 0$: $(d/dr)f = 0, f > 1$.

III. GRAVITATING BPS MONOPOLES AND TRAPPED SURFACES

We will treat Eq. (12) as a dynamical system with initial data $u(0) = 0$, $u'(0) \neq 0$. Our final aim is to find a smooth solution satisfying the asymptotic condition $\lim_{x\to\infty} u(x)M^{1/2}/x = 1$. To achieve that we need a property expressed in the following theorem.

Theorem 1: Assume that solutions u of Eq. (12) exist in the semiaxis $(0,\infty)$ and that $\lim_{x\to\infty} u/x > 0$. If $M^2 < 0.61$ then scattering data $u'(\infty)$ depend continuously on initial values, i.e.,

$$|u'(\infty) - v'(\infty)| \leqslant C|u'(0) - v'(0)|,$$

where u,v are two solutions generated by $u'(0),v'(0)$, respectively.

Before going to the proof let us prove a few propositions.

Proposition 1: A smooth positive solution u of Eq. (12) satisfies the inequality

$$\frac{d}{dx}\left(\frac{u}{x}\right) \leqslant 0.$$

Proof: u/x is proportional to f, a solution of the elliptic equation (7) with nonnegative densities $\rho_h,\rho_{\rm YM}$. The inequality follows from the mini-max principle.

Proposition 2: Let $M^2 < 0.61$ and u,v be solutions of Eq. (12) which exist in $(0,\infty)$. If initially $u(0) = v(0) = 0$ and

$$\epsilon \equiv u'(0) - v'(0) > 0.$$

then the difference $u'(x) - v'(x)$ decreases as long as the value of u/x is greater than $(3)^{1/4}/M^{1/2}$ and increases below this point, but always $u'(x) > v'(x)$.

Proof: Since $u'(0) > v'(0)$, there is an interval $(0,x_0)$ such that $u(x) > v(x)$. Equation (12) can be written in the integral form,

$$u(x) = xu'(0) - \pi\int_0^x dt \int_0^t d\tau\, \tau\rho(\tau) \times [1/(u/\tau)^3 + M^2 u/\tau].$$

The difference between the two solutions reads

$$\frac{u}{x} - \frac{v}{x} = u'(0) - v'(0) - \frac{\pi}{x}\int_0^x dt\int_0^t d\tau\,\tau\rho(\tau) \times\left[\frac{1+(u/\tau)^4M^2}{(u/\tau)^3} - \frac{1+(v/\tau)^4M^2}{(v/\tau)^3}\right] \equiv \epsilon + \delta.$$

Here the quantity δ defined by the last equality is negative in $(0,x_0)$, since the integrand is positive for $u/x > v/x > (3)^{1/4}/M^{1/2}$. Now it is easy to check that a difference between derivatives of u and v is positive in $(0,x_0)$:

$$\begin{aligned}
u'(x) - v'(x) &= u'(0) - v'(0) - \pi\int_0^x dt\, t\rho \times\left[\frac{1+(u/t)^4M^2}{(u/t)^3} - \frac{1+(v/t)^4M^2}{(v/t)^3}\right] \\
&= \epsilon - \pi\int_0^x dt\, t\rho\,\frac{(u/t)^4M^2 - 3}{(u/t)^4}\left(\frac{u}{t} - \frac{v}{t}\right) + o(\epsilon) \\
&= \epsilon - \pi\int_0^x dt\, t\rho\,\frac{(u/t)^4 - 3}{(u/t)^4}(\epsilon + \delta) + o(\epsilon) \\
&\geqslant \epsilon - \pi M^2\epsilon\int_0^x dt\, t\rho\left(\epsilon + \frac{\delta}{\epsilon}\right) + o(\epsilon) \\
&\geqslant \epsilon\left(1 - \pi M^2\int_0^\infty dt\, t\rho\right) + o(\epsilon) \\
&= \epsilon(1 - \pi M^2 0.495) + o(\epsilon) > 0,
\end{aligned} \tag{13}$$

if ϵ is sufficiently small. It means that as long as $u(x)$ is greater than $v(x)$, the derivative of u is always greater than the derivative of v, although the difference between them is decreasing for $x < x_n$, where $x_n \equiv \{x : u(x)/x = (3)^{1/4}/M^{1/2}\}$. (The existence of a unique x_n follows from Proposition 1.) For $x > x_n$ the derivative

$$\frac{d}{dx}(u' - v') = -\pi\rho x\left[\frac{1+M^2(u/x)^4}{(u/x)^3} - \frac{1+M^2(v/x)^4}{(v/x)^3}\right]$$

becomes positive, hence the difference $u' - v'$ starts to grow up. Thus under the conditions stated in Proposition 2, always $u' > v'$ and $u > v$, provided that the initial values are sufficiently close. Now, to finish the proof one should take a set $u_1'(0),...,u_n'(0)$ such that the difference between $u_i'(0) - u_{i+1}'(0)$ is small while the end points are far apart.

Now we are ready to prove Theorem 1. Let ϵ be small (we keep the notation from Proposition 2). We know that for each value of the argument x differences $u'(x) - v'(x)$, $u - v$ are positive. Take very large $x > x_n$; from general theory of ordinary differential equations one infers that $u(x_n)$, $u'(x_n)$ depend continuously on initial datum $u'(0)$ (one can also show this by a direct calculation, analogous to that of Proposition 2). Hence it is sufficient to consider only the interval (x,∞).

Let us compute the difference

$$u'(\infty) - v'(\infty) = u'(x) - v'(x) - \pi\int_x^\infty dt\, t\rho \times\left[\frac{1+M^2(u/t)^4}{(u/t)^3} - \frac{1+M^2(v/t)^4}{(v/t)^3}\right]. \tag{14}$$

We have to estimate the difference $u/x - v/x$:

$$\begin{aligned}
\frac{u(t) - v(t)}{t} &= \frac{1}{t}\left[\int_x^t d\tau(u'(\tau) - v'(\tau)) + u(x) - v(x)\right] \\
&\leqslant \frac{1}{t}[[u'(t) - v'(t)](t - x) + u(x) - v(x)] \\
&\leqslant u'(\infty) - v'(\infty) + (u(x) - v(x))/x \equiv \beta.
\end{aligned} \tag{15}$$

In (15) β is positive, by Proposition 2, which yields also the first inequality. From (14) and (15) we obtain, after some algebra

$$\begin{aligned}
u'(\infty) - v'(\infty) &\leqslant u'(x) - v'(x) + \pi\int_x^\infty dt\, t\rho[(v/t)^{-3} + (v/t + \beta)^{-3} + \beta M^2] \\
&\leqslant C_1\epsilon + C_2\beta \\
&\leqslant \epsilon(C_1 + C_3) + [u'(\infty) - v'(\infty)]C_2,
\end{aligned} \tag{16}$$

where $C_2 = \pi\int_x^\infty dt\, t\rho[3(v/t)^{-2} + M^2]$ and C_3

$= C_2 \sup\{1,c_1\}$. The constant C_1 is that one which appears in the estimation $u'(x) - v'(x) \leqslant C_1\epsilon$, and is finite (that is to be shown by explicit calculation or by invoking standard results of the theory of ordinary differential equations).

By assumption C_2 is finite [notice that we assumed that $(\lim_{x\Rightarrow\infty} u/x)^{-1}$ is finite; this condition, stated in Theorem 1 is not necessary, but in cases of interest it is always satisfied], and even more, it can be arbitrarily small, since ρ vanishes like r^{-4} at large distances and x can be very large. Thus

$$0 < u'(\infty) - v'(\infty) \leqslant \epsilon(C_1 + C_9)/(1 - C_2),$$

which accomplishes the proof of Theorem 1.

As a trivial implication of Proposition 1 we have $0 \leqslant \lim_{x\Rightarrow\infty} u/x < u'(0)$ for any positive solution of Eq. (12). If to denote $\lim_{x\Rightarrow\infty} u/x \equiv a$, then asymptotically a positive solution u of Eq. (12) solves the equation

$$u'' + -\pi u^{-3}(1 + M^2a^4). \tag{17}$$

Its solution reads $u(x) = \sqrt{Ax^2 + Bx + C}$ with $B^2 - 4AC = 4\pi(1 + M^2a^4)$. Now, if to take into account the asymptotic condition $uM^{1/2}/x \Rightarrow 1$ we may infer that $M^{1/2}A^{1/2} = 1$, that is $M^{1/2}a = 1$.

Thus very far from the center of configuration a solution is approximated by $u = \sqrt{x^2/M + Bx + C}$ with $B^2 - 4C/M = 8\pi$. The parameter B happens to be equal to the ADM mass of the whole system.

These observations justify our numerical strategy, which consists of the following points:

(i) Fix $u(0) = 0$, $u'(0) \neq 0$, and do calculations for a fixed value of M;

(ii) keep M and change $u'(0)$ so many times as to get for large values of x the situation in which there are two sets of data, say $u_1'(0)$ and $u_2'(0)$ such that $u_1'(x)M^{1/2} < 1$ and $M^{1/2}u_2'(x) > 1$. Here u_1, u_2 develop from $u_1'(0)$, $u_2'(0)$, respectively.

Now we infer from Theorem 1 that there should exist an intermediate value $u_*'(0)$, $u_1'(0) < u_*'(0) < u_2'(0)$ such that $u_*'M^{1/2}$ approaches 1 as x tends to infinity.

In terms of the variables x, u, and a parameter M the following expression is of particular interest:

$$H(u) \equiv M^{1/2}(2u_M'(x) - u_M/x).$$

The subscript attached to u specifies a value of M connected with the solution. In the original variables $H = (d/dr)(f^2r)/f$; hence for H nonpositive somewhere, spherical trapped surfaces do occur.[7]

We have done several numerical calculations to show the existence of trapped surfaces. The investigation was made easier thanks to a technical result of the following proposition:

Proposition 3: A positive solution of Eq. (12) with a horizon (i.e., with $H(u)(x) < 0$ somewhere) can exist only for $M \in (0.37, 1.6)$. The proof is done in the Appendix.

The existence of trapped surfaces was established in configurations corresponding to several pairs of data, e.g.,

(i) $u'(0) = 5.68$, $M \cong 0.80$,
(ii) $u'(0) = 15.7$, $M \cong 0.90$,
(iii) $u'(0) = 100$, $M \cong 0.94$.

Remark: The above data show that in fact we need a stronger result than stated in Theorem 1, to cover at least the case with $M^2 = 0.64$. We have restricted to the previous number just to avoid complications which would obscure the main idea of the proof. It is essentially easy to improve the upper bound by careful inspection of coefficients in Proposition 2 above; this is the unique step of the proof of Theorem 1 that depends on the estimate $M^2 < 0.61$.

We conclude this section by pointing out that the initial data (10) [with $\alpha = 1$, $\beta = -3$ and fields given by formulas (8)] of Einstein–Yang–Mills–Higgs equations with $f = uM^{1/2}/x$ given by the above pairs (i)–(iii), contain trapped surfaces in the Cauchy surface.

IV. THE CASE WITH NO APPARENT HORIZONS

In this section we will study initial data (10) with $\alpha = \beta = -3$ and $f = uM^{1/2}/x$ generated by a solution u of the equation

$$u'' = -2\pi\rho u^{-3}. \tag{18}$$

That case was studied in Ref. 11, where trapped surfaces were shown to be absent. The key step of the proof consisted in the application of a necessary result for the existence of trapped surfaces proven earlier.[6] Here we will give a different and much simpler proof.

We recall a few facts related to Eq. (18).[11] There is a number $c_* \cong 1.624$, such that for $u'(0) < c_*$ all solutions have to vanish somewhere; this exemplifies Wheeler's bag of gold configuration[12] and was described in detail in Ref. 11. For $u'(0) > c_*$ all solutions of Eq. (18) have correct asymptotic behavior. The numerical investigation has shown the absence of apparent horizons for a finite number of examples. The problem is to show that this is a generic property, i.e., that for $u'(0) > c_*$ trapped surfaces are absent. In Ref. 11 was employed the above mentioned qualitative result. Here we solve this problem in a different way.

Theorem 2: Let u be a solution of Eq. (18) generated by $u'(0) > c_*$ and let $H(u_1)/M^{1/2} = 2u_1' - u_1/x$ be positive everywhere, i.e., trapped surfaces are absent. Then for all Cauchy data (10) with $\alpha = \beta = -3$ corresponding to solutions u of Eq. (18) generated by $u'(0) > u_1'(0)$, trapped surfaces are absent.

Proof: We will show that under the above conditions

$$H(u)/M^{1/2} \geqslant H(u_1)/M_1^{1/2}, \tag{19}$$

which proves the thesis, since $H(u_1)$ is known to be positive. In fact, at $x = 0$, $H(u) = u'(0) > u_1'(0) = H(u_1)$. Notice that $u''(x) \geqslant u_1''(x)$, $u'(x) \geqslant u_1'(x)$, and $u(x) \geqslant u_1(x)$; these inequalities follow directly from Eq. (16) and $u'(0) > u_1'(0)$. Now it is easy to see that

$$\frac{d}{dx}\frac{xH(u)}{M^{1/2}} = 2u''x + u'x \geqslant 2u_1''x + u_1'x$$

$$= \frac{d}{dx}\frac{xH(u)}{M^{1/2}}$$

at every point $x \geqslant 0$. This proves (19).

Thus, it suffices to take u_* as close as possible to c_* and to check that its corresponding configuration does not contain trapped surfaces; that program has been done in Ref. 11. The above theorem allows us to conclude that trapped surfaces are absent in all configurations with the scaling $\alpha=\beta=-3$.

ACKNOWLEDGMENTS

Part of this work has been done at the International Centre for Theoretical Physics in Trieste. One of the authors (EM) would like to thank the organizers of Summer School on High Energy Physic and Cosmology for the invitation. We are grateful to Dr. Tadeusz Chmaj for his help in the numerical analysis. We thank Professor Niall O'Murchadha for an inspiring discussion.

This work was supported in part by Polish Government Grant CPBP 01.03.

APPENDIX: PROOF OF PROPOSITION 3 FROM SEC. III

Proposition 3: A solution of Eq. (12) with a horizon and such that $u/x>0$ ($x\to\infty$), can exist only for $M\in(0.37,1.6)$.

Proof: In the first part we prove the estimation from above. Let w_α be a solution of the equation

$$w_\alpha''=-\pi C_\alpha M^2 w_\alpha, \tag{A1}$$

where $C_\alpha\equiv\inf(\rho{:}x\leqslant\alpha)$ and ρ is given below the formula (12). If u is a solution of Eq. (12) such that $u(0)=w_\alpha(0)=0$, $u'(0)=w_\alpha'(0)$, then obviously $u\leqslant w_\alpha$.

Equation (A1) possesses bound states for $C_\alpha M^2\alpha^2\pi>\pi^2/4$; thus the necessary condition for w_α (and u) to have the required asymptotic behavior reads

$$M<(\pi\alpha^{-2}/(4C_\alpha))^{1/2}. \tag{A2}$$

Equation (A2) yields the above estimation if to take $\alpha=2.0$ and observe that then $C_\alpha\cong 0.074$.

In the second part we derive the estimate from below. This is done in two steps.

Lemma: If $u'(0)>c_*[1/2+(u'(0))^4M^2/2]^{1/4}$, where c_* is specified in Sec. IV, then a solution $u(x)$ of Eq. (12) does not possess a horizon.

Proof of Lemma: Notice (see Sec. III) that $u/x\leqslant u'(0)$ and therefore $u\geqslant h$, where $h'(0)=u'(0)$, $h(0)=u'(0)=0$, and h is a solution of

$$h''=-\pi\rho x^4 h^{-3}[1+(u'(0))^4M^2]. \tag{A3}$$

Now define

$$v=h[1/2+(u'(0))^4M^2/2]^{-1/4}, \tag{A4}$$

v solves the equation

$$v''=-2\pi\rho x^4 v^{-3} \tag{A5}$$

and $v'(0)>c_*$. Hence $H(v)=2v'-v/x$ is positive, according to results of Sec. IV. Methods analogous to those from Sec. IV show that $H(u)\geqslant H(v)$; that is, under the conditions stated in the lemma, a horizon is absent.

To finish the proof of the second part of the proposition, let us note, that if we want to allow the existence of trapped surfaces, we must have solutions with initial data satisfying the converse of the inequality stated in the lemma, i.e.,

$$u'(0)\leqslant c_*[1/2+(u'(0))^4M^2/2]^{1/4}. \tag{A6}$$

Equation (A6) and the obvious estimation $u'(0)M>1$ imply the desired estimate $M>0.37$.

That kind of reasoning can be repeated to get even better estimates from below.

[1]T. D. Lee, Phys. Rev. D **35**, 3637 (1987); R. Friedberg, T. D. Lee, and Y. Pang, Phys. Rev. D **35**, 3658 (1987); T. D. Lee and Y. Pang, *ibid.* **35**, 3678 (1987).

[2]J. A. Frieman and Ch. T. Hill, SLAC preprint 4283 (1987).

[3]N. K. Glendenning, T. Kodama, and F. R. Klinkhamer, Phys. Rev. D **38**, 3226 (1987).

[4]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 868.

[5]W. B. Bonnor, Phys. Lett. A **99**, 424 (1983).

[6]P. Bizon, E. Malec, and N. O'Murchadha, Phys. Rev. Lett. **56**, 1147 (1988); P. Bizon, E. Malec, and N. O'Murchadha, Class. Quant. Grav. **6**, 961 (1989).

[7]S. W. Hawking and G. F. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).

[8]W. A. Hiscock, Phys. Rev. Lett. **50**, 1734 (1982).

[9]E. B. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976); M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).

[10]Y. Choquet-Bruhat and J. W. York, Jr., *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 1, pp. 99–172.

[11]T. Chmaj and E. Malec, Class. Quantum Grav. **6**, 1687 (1989).

[12]N. O'Murchadha, Class. Quantum Grav. **4**, 1609 (1987).

[13]J. Isenberg, N. O'Murchadha, and J. W. York, Phys. Rev. D **13**, 1532 (1976).

Wave functions of bound states of a fermion and a Dirac dyon and matrix elements in an external electromagnetic field

Jian-zu Zhang[a)]
CERN, Geneva, Switzerland and Center of Theoretical Physics, Chinese Center of Advanced Science & Technology (World Laboratory), Beijing, People's Republic of China
Yong-chang Qi
Department of Physics, Shanxi University, Taiyuan, Shanxi 030006, People's Republic of China

(Received 22 November 1988; accepted for publication 7 January 1990)

Wave functions of bound states for a fermion and a Dirac dyon with charge $Z_d < Z_d^c$ and for $j \geqslant |q| + \frac{1}{2}$ are obtained. Matrix elements of this system in the external electromagnetic field are calculated and the corresponding selection rules are shown.

I. INTRODUCTION

In recent years, the problem of bound states of a fermion in a fixed Dirac monopole[1] or in a 't Hooft–Polyakov monopole[2] has been extensively discussed.[3–20] In this paper the bound system of a fermion and a fixed Dirac dyon will be discussed further.[9,17] As is well known for this system of a fermion and a Dirac monopole, there is the Lipkin–Weisberger–Peshkin (LWP) difficulty[21] in the angular momentum states $j = |q| - \frac{1}{2}$, which shows up in the fermion's radial wave functions at the origin. The radial wave functions of the fermion in the angular momentum state $j = |q| - \frac{1}{2}$ do not vanish at the origin. This means that the fermion in these states goes through the monopole; thus, the Hamiltonian of the system is ill defined at the origin. To avoid this difficulty, an infinitesimal extra magnetic moment is endowed to the fermion by Kazama and Yang.[5,6] In Ref. 9 we showed that for the system of a fermion and a Dirac dyon there is also the LWP difficulty in the angular momentum states $j \geqslant |q| + \frac{1}{2}$ when the dyon charge Z_d exceeds some critical value Z_d^c. In order to avoid the LWP difficulty, besides the Kazama–Yang term $-(Kq/2Mr^3)\beta \boldsymbol{\Sigma}\cdot\mathbf{r}$, the term $i(KZZ_d e^2/2Mr^3)\boldsymbol{\gamma}\cdot\mathbf{r}$ should be considered also. But in the case $Z_d < Z_d^c$, the Hamiltonian of the system is well defined at the origin, so we can solve the bound-state energy for $j \geqslant |q| + \frac{1}{2}$ without the Kazama–Yang term[5] and the term $(KZZ_d e^2/2Mr^3)\boldsymbol{\gamma}\cdot\mathbf{r}$ (Ref. 9). The results show that the bound-state energy is hydrogenlike.[17]

In this paper, wave functions of bound states for a fermion and a Dirac dyon with charge $Z_d < Z_d^c$ and for $j \geqslant |q| + \frac{1}{2}$ are obtained. Using these wave functions, matrix elements of this system in the external electromagnetic field are calculated and the corresponding selection rules are shown.

II. THE BASIC EQUATION AND ITS SOLUTION

In this section we first review the basic equation to fix our conventions. The Hamiltonian of this system is[17]

$$H = \boldsymbol{\alpha}\cdot(-i\nabla - Ze\mathbf{A}) + \beta M - \lambda/r, \tag{2.1}$$

where $\mathbf{A}$ is the vector potential of the dyon. In order to remove the string of singularities, $\mathbf{A}$ is defined in terms of two or more functions in a corresponding number of overlapping regions.[3] Note that $\lambda = ZZ_d e^2$, where Z is the electric charge of the fermion that is an integer and Z_d is the electric charge of the dyon that need not be an integer. For the states $j \geqslant |q| + \frac{1}{2}$ there are two types of simultaneous eigensections of $\mathbf{J}^2$, J_Z, and H (Ref. 4):

$$\text{Type A}\quad \psi_{jm}^{(1)} = \frac{1}{r}\begin{pmatrix} h_1(r)\xi_{jm}^{(1)} \\ -ih_2(r)\xi_{jm}^{(2)} \end{pmatrix} \quad (j \geqslant |q| + \tfrac{1}{2}), \tag{2.2}$$

$$\text{Type B}\quad \psi_{jm}^{(2)} = \frac{1}{r}\begin{pmatrix} h_3(r)\xi_{jm}^{(2)} \\ -ih_4(r)\xi_{jm}^{(1)} \end{pmatrix} \quad (j \geqslant |q| + \tfrac{1}{2}), \tag{2.3}$$

where

$$\xi_{jm}^{(1)} = c\phi_{jm}^{(1)} - s\phi_{jm}^{(2)}, \tag{2.4}$$

$$\xi_{jm}^{(2)} = s\phi_{jm}^{(1)} + c\phi_{jm}^{(2)}, \tag{2.5}$$

$$c = q[(2j+1+2q)^{1/2} + (2j+1-2q)^{1/2}]/2|q|(2j+1)^{1/2}, \tag{2.6}$$

$$s = q[(2j+1+2q)^{1/2} - (2j+1-2q)^{1/2}]/2|q|(2j+1)^{1/2}, \tag{2.7}$$

$$\phi_{jm}^{(1)} = \begin{pmatrix} ((j+m)/2j)^{1/2} Y_{q,j-1/2,m-1/2} \\ ((j-m)/2j)^{1/2} Y_{q,j-1/2,m+1/2} \end{pmatrix}, \tag{2.8}$$

$$\phi_{jm}^{(2)} = \begin{pmatrix} -((j-m+1)/(2j+2))^{1/2} Y_{q,j+1/2,m-1/2} \\ ((j+m+1)/(2j+2))^{1/2} Y_{q,j+1/2,m+1/2} \end{pmatrix}, \tag{2.9}$$

where $Y_{q,L,M}$ is the monopole harmonic.[3,22–24]

In (2.2) and (2.3), $h_i(r)$ ($i = 1,2,3.4$) are defined in a rather different way than in Ref. 4; thus, the system of equations satisfied by $h_i(r)$ is obtained in the compact form, which is easily treated. According to Lemma I of Ref. 4, from (2.1)–(2.3), we obtain, for type A:

$$(M - E - \lambda/r)h_1(r) + (\partial_r + \mu/r)h_2(r) = 0,$$
$$(\partial_r - \mu/r)h_1(r) + (M + E + \lambda/r)h_2(r) = 0, \tag{2.10}$$

and for type B:

$$(M - E - \lambda/r)h_3(r) + (\partial_r - \mu/r)h_4(r) = 0,$$
$$(\partial_r + \mu/r)h_3(r) + (M + E + \lambda/r)h_4(r) = 0, \tag{2.11}$$

where

$$\mu = [(j+\tfrac{1}{2})^2 - q^2]^{1/2} > 0, \tag{2.12}$$

[a)] On leave of absence from: Department of Physics, Shanxi University, Taiyuan, Shanxi 030006, People's Republic of China.

 0022-2488/90/071796-04$03.00

$q = Zeg$, g is the strength of the magnetic monopole; the Dirac quantization condition is $eg = n/2$ $(n = 0, \pm 1, \pm 2, \ldots)$ (Ref. 1).

We can solve (2.10) and (2.11) according to the standard treatment in quantum mechanics textbooks.[25] When $r \to 0$, (2.10) is reduced to

$$(\lambda/r)h_1(r) - (\partial_r + \mu/r)h_2(r) = 0,$$
$$(\partial_r - \mu/r)h_1(r) + (\lambda/r)h_2(r) = 0. \quad (2.13)$$

Setting $h_1(r) = ar^\nu$, $h_2(r) = br^\nu$, where a and b are constants, from (2.13) we obtain

$$\lambda a - (\nu + \mu)b = 0,$$
$$(\nu - \mu)a + \lambda b = 0.$$

From conditions of nonzero a and b, the finiteness of $h_1(r)$ and $h_2(r)$ when $r \to 0$, we have

$$\nu = (\mu^2 - \lambda^2)^{1/2} = [(j+\tfrac{1}{2})^2 - q^2 - (zz_d e^2)^2]^{1/2} > 0. \quad (2.14)$$

Setting

$$\rho = 2(M^2 - E^2)^{1/2} r = 2pr$$
$$(p = (M^2 - E^2)^{1/2}), \quad (2.15)$$
$$h_1(\rho) = 2p(M+E)^{1/2} e^{-\rho/2}\rho^\nu (Q_1(\rho) + Q_2(\rho)),$$
$$h_2(\rho) = 2p(M-E)^{1/2} e^{-\rho/2}\rho^\nu (Q_1(\rho) - Q_2(\rho)), \quad (2.16)$$

we have

$$\rho Q_1'(\rho) + (\nu - \lambda E/p)Q_1(\rho) - (\mu + \lambda M/p)Q_2(\rho) = 0,$$
$$\rho Q_2'(\rho) + (\nu - \rho + \lambda E/p)Q_2(\rho) - (\mu - \lambda M/p)Q_1(\rho) = 0, \quad (2.17)$$

and

$$\rho Q_1''(\rho) + (2\nu + 1 - \rho)Q_1'(\rho) - (\nu - \lambda E/p)Q_1(\rho) = 0, \quad (2.18)$$
$$\rho Q_2''(\rho) + (2\nu + 1 - \rho)Q_2'(\rho) - (\nu + 1 - \lambda E/p)Q_2(\rho) = 0. \quad (2.19)$$

If we set

$$h_3(\rho) = 2p(M+E)^{1/2} e^{-\rho/2}\rho^\nu (Q_3(\rho) - Q_4(\rho)),$$
$$h_4(\rho) = 2p(M-E)^{1/2} e^{-\rho/2}\rho^\nu (Q_3(\rho) + Q_4(\rho)), \quad (2.20)$$

then $Q_3(\rho)$ satisfies (2.18), and $Q_4(\rho)$ satisfies (2.19).

Equations (2.18) and (2.19) are standard confluent hypergeometric equations. Their finite solutions at the origin are the confluent hypergeometric function $F(a,b,\rho)$:

$$Q_{1,3}(\rho) = A_{1,3} F(\nu - \lambda E/p, 2\nu + 1, \rho),$$
$$Q_{2,4}(\rho) = A_{2,4} F(\nu + 1 - \lambda E/p, 2\nu + 1, \rho). \quad (2.21)$$

From (2.2), (2.16), and (2.21), the radial wave functions are

$$R_{\substack{1\\2}}(\rho) = 2ph_{\substack{1\\2}}(\rho)/\rho = 4p^2(M \pm E)^{1/2} e^{-\rho/2}\rho^{\nu-1} \times [A_1 F(\nu - \lambda E/p, 2\nu + 1, \rho) \pm A_2 F(\nu + 1 - \lambda E/p, 2\nu + 1, \rho)]. \quad (2.22)$$

When $\rho \to 0$, $F(a,b,\rho) \to 0$, from (2.17), we have

$$A_2 = \frac{\nu - \lambda E/p}{\mu + \lambda M/p} A_1. \quad (2.23)$$

Similarly, for A_3 and A_4, we have

$$A_4 = \frac{\nu - \lambda E/p}{\mu - \lambda M/p} A_3. \quad (2.24)$$

When $\rho \to \infty$, $F(a,b,\rho) \to e^\rho$, so $R_i(\rho)$ is divergent. In order to avoid the divergence, we must set

$$\nu - \lambda E/p = -n \quad (n = 0,1,2,\ldots). \quad (2.25)$$

When $n = 0$, $F(\nu - \lambda E/p, 2\nu + 1, p) = 0$; though $F(\nu + 1 - \lambda E/p, 2\nu + 1, p)$ is still divergent, $A_2 F(\nu + 1 - \lambda E/p, 2\nu + 1, \rho) = 0$. Thus we obtain[17]

$$E_{n,q,j} = \pm M \left[1 + \frac{1}{(n+\nu)^2/\lambda^2}\right]^{-1/2} = \pm M \left(1 + \left[\frac{ZZ_d e^2}{n + [(j+\tfrac{1}{2})^2 - q^2 - (zz_d e^2)^2]^{1/2}}\right]^2\right)^{-1/2}, \quad (2.26)$$

where $n = 0,1,2,3,\ldots; j \geqslant |q| + \tfrac{1}{2}$; $q = Zeg \neq 0$; $eg = \pm\tfrac{1}{2}, \pm 1, \pm\tfrac{3}{2}, \ldots$, $\mu = [(j+\tfrac{1}{2})^2 - q^2]^{1/2} > 0$, $\lambda = ZZ_d e^2$. Notice that the total angular momentum j, which is defined in Ref. 4, is different from the total angular momentum in ordinary quantum mechanics. Here, j can take integer as well as half-integer values. The spectrum (2.26) is hydrogenlike, but is different from the atomic or the molecular spectrum. If dyons exist in nature, (2.26) leads to the possibility to look for dyonic bound states, for example, from astronomical observations.

By the relation

$$L_n^\nu(x) = \frac{\Gamma(\nu + 1 + n)}{n!\Gamma(\nu + 1)} F(-n, \nu + 1, x),$$

and using (2.23) and (2.25), (2.22) is reduced to

$$R_{\substack{1\\2}}^{nqj}(\rho) = 4p_{nqj}^2 (M \pm E_{nqj})^{1/2} A_1^{nqj} e^{-\rho/2}\rho^{\nu-1} \times \left\{\frac{n!\Gamma(2\nu+1)}{\Gamma(2\nu+1+n)} L_n^{2\nu}(\rho) \mp \frac{n}{\mu + (\mu^2 + n^2 + 2n\nu)^{1/2}} \times \frac{(n-1)!\Gamma(2\nu+1)}{\Gamma(2\nu+n)} L_{n-1}^{2\nu}(\rho)\right\}, \quad (2.27)$$

Similarly, we have

$$R_{\substack{3\\4}}^{nqj}(\rho) = 4p_{nqj}^2 (M \pm E_{nqj})^{1/2} A_3^{nqj} e^{-\rho/2}\rho^{\nu-1} \times \left\{\frac{n!\Gamma(2\nu+1)}{\Gamma(2\nu+1+n)} L_n^{2\nu}(\rho) \pm \frac{n}{\mu - (\mu^2 + n^2 + 2n\nu)^{1/2}} \times \frac{(n-1)!\Gamma(2\nu+1)}{\Gamma(2\nu+n)} L_{n-1}^{2\nu}(\rho)\right\}. \quad (2.28)$$

Because $\xi_{jm}^{(1)}$ and $\xi_{jm}^{(2)}$ are normalized, the radial wave functions $R_i(\rho)$ satisfy the following normalization condition:

$$\int_0^\infty \sum_{i=1(3)}^{2(4)} |R_i(2p_{nqj} r)|^2 r^2 dr = 1.$$

Using the normalization condition of $L_n^{\nu}(x)$,

$$\int_0^{\infty} dx\, x^{\nu} e^{-x} L_n^{\nu}(x) L_{n'}^{\nu}(x) = \frac{\Gamma(\nu+n+1)}{n!}\,\delta nn',$$

we have

$$A_{\substack{1\\3}}^{nqj} = \frac{1}{2\Gamma(2\nu+1)}\left(Mp_{nqj}\,\frac{n!}{\Gamma(2\nu+n)}\left[\frac{1}{2\nu+n} + \frac{n}{[\mu \pm (\mu^2+n^2+2\nu n)^{1/2}]^2}\right]\right)^{-1/2}. \qquad (2.29)$$

III. TRANSITION MATRIX ELEMENTS IN THE EXTERNAL ELECTROMAGNETIC FIELD

In this section we show examples of the calculation of matrix elements by using wave functions of bound states of a fermion with a Dirac dyon.

If the external electromagnetic field is described by the vector potential

$$\mathbf{A}(\mathbf{x},t) = \mathbf{A}_0 \exp[i(\mathbf{K}\cdot\mathbf{x} - \omega t)], \qquad (3.1)$$

then the interaction Hamiltonian of the fermion–Dirac dyon system in the external electromagnetic field is

$$H_i = -Ze\boldsymbol{\alpha}\cdot\mathbf{A}(\mathbf{x},t). \qquad (3.2)$$

Treat H_i as the perturbation. Suppose that within the scale of the bound system of a fermion and a Dirac dyon, $\mathbf{K}\cdot\mathbf{X} \ll 1$, the $\exp(i\mathbf{K}\cdot\mathbf{X}) \approx 1 + i\mathbf{K}\cdot\mathbf{X} + \cdots$. In order to show the general method of the calculation of matrix elements, we discuss not only the first term, but also the $\mathbf{K}\cdot\mathbf{X}$ term in the above expansion. Take $\mathbf{A}_0$ along the X axis, $\mathbf{K}$ along the Z axis, and work in the Coulomb gauge. Let N represent the quantum number set (n,q,j,m). We have

$$H'_{N',N} = H'^{(0)}_{N',N} + H'^{(1)}_{N',N}, \qquad (3.3)$$

where H' represents the part of H_i separated from the time-dependent factor $\exp(-i\omega t)$. In (3.3),

$$H'^{(0)}_{N',N} = -ZeA_0\left\langle \psi_{N'}^{(1)}\left|\begin{pmatrix}0 & \sigma_1\\ \sigma_1 & 0\end{pmatrix}\right|\psi_N^{(1)}\right\rangle = -ZeA_0\left\langle\begin{pmatrix}0 & \sigma_1\\ \sigma_1 & 0\end{pmatrix}\right\rangle_{N',N}, \qquad (3.4)$$

$$H'^{(1)}_{N',N} = -iZeA_0K\sqrt{\frac{4\pi}{3}}\left\langle \psi_{N'}^{(1)}\left| rY_{010}\begin{pmatrix}0 & \sigma_1\\ \sigma_1 & 0\end{pmatrix}\right|\psi_N^{(1)}\right\rangle = -iZeA_0K\sqrt{\frac{4\pi}{3}}\left\langle rY_{010}\begin{pmatrix}0 & \sigma_1\\ \sigma_1 & 0\end{pmatrix}\right\rangle_{N',N}. \qquad (3.5)$$

For type A (2.2),

$$H'^{(0)a}_{N',N} = -iZeA_0\left[\Omega_{21}^{(2)qn'j'nj}(\sigma_1)_{j'm'jm}^{21} - \Omega_{12}^{(2)qn'j'nj}(\sigma_1)_{j'm'jm}^{12}\right], \qquad (3.6)$$

$$H'^{(1)a}_{N',N} = ZeA_0K\sqrt{\frac{4\pi}{3}}\left[\Omega_{21}^{(3)qn'j'nj}(Y_{010}\sigma_1)_{j'm'jm}^{21} - \Omega_{12}^{(3)qn'j'nj}(Y_{010}\sigma_1)_{j'm'jm}^{12}\right], \qquad (3.7)$$

where

$$\Omega_{st}^{(2)qn'j'nj} = \int_0^{\infty} r^2\, dr\, R_s^{n'qj'*}(2p_{n'qj'}r) R_t^{nqj}(2p_{nqj}r) \quad (s,t=1,2), \qquad (3.8)$$

$$\Omega_{st}^{(3)qn'j'nj} = \int_0^{\infty} r^3\, dr\, R_s^{n'qj'*}(2p_{n'qj'}r) R_t^{nqj}(2p_{nqj}r) \quad (s,t=1,2), \qquad (3.9)$$

$$(\sigma_1)_{j'm'jm}^{st} = \int d\Omega\, \xi_{j'm'}^{(s)\dagger}\sigma_1\xi_{jm}^{(t)} \quad (s,t=1,2), \qquad (3.10)$$

$$(Y_{010}\sigma_1)_{j'm'jm}^{st} = \int d\Omega\, \xi_{j'm'}^{(s)\dagger}Y_{010}\sigma_1\xi_{jm}^{(t)} \quad (s,t=1,2). \qquad (3.11)$$

From the calculation of matrix elements we obtain the selection rules of $H'^{(0)a}_{N',N}$:

$$\Delta j = 0, \pm 1, \quad \Delta m = \pm 1. \qquad (3.12)$$

In the general coordinates, they are $\Delta j = 0, \pm 1$, $\Delta m = 0, \pm 1$.

The selection rules of $H'^{(1)a}_{N',N}$ are

$$\Delta j = 0, \pm 1, \pm 2, \quad \Delta m = \pm 1. \qquad (3.13)$$

In the general coordinates, they are $\Delta j = 0, \pm 1, \pm 2$, $\Delta m = 0, \pm 1, \pm 2$.

For type B (2.3), we have similar results.

Now we show the calculations of the matrix elements $Q_{st}^{(\eta)}$ $(\eta = 2,3)$.

We notice that $F(a,b,Z)$ is reduced to a polynomial when a is equal to 0 or a negative integer:

$$F(-n,b,\rho) = \sum_{l=0}^{n}\frac{(-1)^l n(n-1)\cdots(n-l+1)}{l!\,b(b+1)\cdots(b+l-1)}\rho^l \quad (n = 0,1,2,\cdots).$$

From (3.8) we have

$$\Omega_{st}^{(2)qn'j'nj} = G^{(1)}_{n'qj'(\mp)}G^{(1)}_{nqj(\pm)}2^{\nu+\nu'-2}p_{n'qj'}^{\nu'-1}p_{nqj}^{\nu-1}\left[\sum_{l=0}^{n}\sum_{l'=0}^{n'}C^n_{2\nu+1,l}C^{n'}_{2\nu+1,l'} \mp K_{nj}^{(+)}\sum_{l=0}^{n-1}\sum_{l'=0}^{n'}C^{n-1}_{2\nu+1,l}C^{n'}_{2\nu+1,l'} \pm K_{n'j'}^{(+)}\sum_{l=0}^{n}\sum_{l'=0}^{n'-1}C^{n}_{2\nu+1,l}C^{n'-1}_{2\nu+1,l'} \mp K_{nj}^{(+)}K_{n'j'}^{(+)}\sum_{l=0}^{n-1}\sum_{l'=0}^{n'-1}C^{n-1}_{2\nu+1,l}C^{n'-1}_{2\nu+1,l'}\right]\cdot I_{ll'}^{\nu\nu'}, \qquad (3.14)$$

where

$$G^{(1)}_{nqj(\pm)} = 4p^2_{nqj}(M \pm E_{nqj})^{1/2} A^{nqj}_1,$$

$$K^{(+)}_{nj} = \frac{n}{\mu + (\mu^2 + n^2 + 2n\nu)^{1/2}},$$

$$C^{(n)}_{\nu l} = \frac{(-1)^l n(n-1)\cdots(n-l+1)}{l!(2\nu+1)(2\nu+2)\cdots(2\nu+l)},$$

$$C^{(n)}_{\nu 0} \equiv 1, \quad C^{0}_{\nu 0} \equiv 1,$$

$$I^{\nu\nu'}_{ll'} = \int_0^\infty dr\, e^{-(p_{nqj}+p_{n'qj'})r} r^{\nu+\nu'+l+l'}$$

$$= (p_{nqj} + p_{n'qj'})^{-(\nu+\nu'+l+l'+1)}$$

$$\times \Gamma(\nu+\nu'+l+l'+1).$$

In (3.14), the upper (lower) sign is taken for $st = 12(21)$.

Similarly, we can calculate $Q^{(3)qn'j'nj}_{st}$.

Using properties of monopole harmonics[3,22–24] and Wigner 3-j symbols,[26] we have obtained $\Omega^{(\eta)}_{st}$ ($\eta = 1,2$) and $(Y_{010}\sigma_1)^{st}_{j'm'jm}$ in detail also. The results, which are not necessary to list here, are quite tedious.

IV. DISCUSSIONS

(i) The discussion of the singular case of large dyon charge ($Z_d > Z^c_d$) and the ground state $j = |q| - \frac{1}{2}$ is very important. But in these two cases, in order to avoid LWP difficulty, we must introduce the Kazama–Yang term, $-(kq/2Mr^3)\beta\,\mathbf{\Sigma}\cdot\mathbf{r}$, and the term, $-(kZZ_de^2/2Mr^3)\boldsymbol{\gamma}\cdot\mathbf{r}$ (Ref. 9), which are very difficult to solve. Besides the zero energy bound state,[5] we have not obtained the explicit analytical radial wave functions yet, so we cannot calculate the radial parts of transition matrix elements in these two cases. The transition to the ground state would presumably be crucial in identifying emissions from such a physical system if it exists. For a given q the selection rules of electric dipole transition from type A or type B to ground state type C,

$$\psi^{(3)}_{jm} = \frac{1}{r}\begin{bmatrix} f(r)\xi^{(2)}_{jm} \\ g(r)\xi^{(2)}_{jm} \end{bmatrix}, \quad j = |q| - \tfrac{1}{2},$$

are

$$\Delta j = -1; \quad \Delta m = 0, \pm 1. \tag{4.1}$$

Notice that in (4.1) the $\Delta j = 0$ parity violation transition and the $\Delta j = 1$ transitions are absent.

(ii) In the fermion–dyon bound system, the spatial parity is violated by the magnetic charge of the dyon[7] that leads to a modification of the selection rules for electromagnetic transitions. Particularly , the $\Delta j = 0$ dipole transitions shown in (3.12) are allowed., These $\Delta j = 0$ parity violation electric dipole transitions are the most outstanding feature of the fermion–dyon bound system, which distinguishes it from the hydrogenlike atom.

(iii) When $\lambda \ll 1$, (2.26) is reduced to

$$\frac{M-E}{M} = \frac{\lambda^2}{2(n+\mu)^2}\left[1 + \frac{\lambda^2}{\mu(n+\mu)}\right]. \tag{4.2}$$

Reference 8 presents analytic approximate results for dyon–fermion binding energies and the corresponding bound-state wave functions for angular momentum $j \geqslant |q| + \frac{1}{2}$ with the Kazama–Yang term. But their results are valid only in the limit of weak binding, $M - E \ll M$, and for small charges $ZZ_de^2 \ll 1$. It is easy to check from the last paper of Ref. 8, taking the limit $\kappa \to 0$, that their result coincides with our Eq. (2.26) to the leading order in λ [just the first term of Eq. (4.2)], whereas our formula is valid without a restriction to $\lambda \ll 1$.

Note added in proof: In the general coordinates, i.e., when $\mathbf{A}_0$ is not taken along the X axis, the $\Delta j = 0$, $\Delta m = 0$ transition exists. The corresponding transition matrix elements are

$$(H'_I)_{qn'jm,njm} = 16iZeA_{03}p^2_{n'qj}p^2_{nqj} \cdot \frac{qm}{j(j+1)}\left(I^{qn'j,nj}_1 - I^{qn'j,nj}_2\right),$$

where $p^2_{nqj} = M^2 - E^2_{nqj}$, A^{nqj}_1 is the normalization constant given by (2.29), and $I^{qn'j,nj}_{1,2}$ are the radial integrals.

ACKNOWLEDGMENTS

One of us (Z.J.Z.) would like to thank the Theoretical Physics Division of CERN for its hospitality. He also would like to thank John Ellis and T. T. Wu for helpful discussions. He is indebted to the referee for calling his attention to the discussion of the transitions to the ground state.

[1]P. A. M. Dirac, Proc. R. Soc. London, Ser. A **133**, 60 (1931); Phys. Rev. **74**, 817 (1948).

[2]G. 't Hooft, Nucl. Phys. B **79**, 276, (1974); A. M. Polyakov, Pis'ma ZETF **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)].

[3]T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976).

[4]Y. Kazama, C. N. Yang, and A. S. Goldhaber, Phys. Rev. D **15**, 2287 (1977).

[5]Y. Kazama and C. N. Yang, Phys. Rev. D **15**, 2300 (1977).

[6]C. N. Yang, *Proceedings of the Monopole Meeting*, Trieste, Italy, Dec. 1981, edited by N. S. Craigie, P. Goddard, and W. Nahm (World Scientific, Singapore, 1982), p. 237.

[7]A. S. Goldhaber, Phys. Rev. D **16**, 1815 (1977); Y. Kazama, *ibid.* **16**, 3078 (1977).

[8]P. Osland and T. T. Wu, Nucl. Phys. B **247**, 421, 450 (1984); **256**, 13, 32, 4491 (1985); **261**, 687 (1985).

[9]Xin-zhou Li, Ke-lin Wang, and Jian-zu Zhang, Phys. Lett. B **148**, 89 (1984).

[10]R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976); R. Jackiw and J. R. Schrieffer, Nucl. Phys. B **190**, [FS3], 253 (1981); J. Goldstone and F. Wilczek, Phys. Rev. Lett. **47**, 988 (1981).

[11]R. Jackiw and C. Rebbi, Phys. Rev. Lett. **36**, 1122 (1976); P. Hasenfratz and G. 't Hooft, *ibid.* **36**, 1119 (1976); A. Goldhaber, *ibid.* **36**, 1112 (1976).

[12]J. Preskill, Phys. Rev. Lett. **43**, 1365 (1979).

[13]C. Callan, Phys. Rev. D **25**, 2141 (1982); **26**, 2058 (1982); Nucl. Phys. B **212**, 391 (1983); V. Rubakov, Pis'ma ZETF **33**, 658 (1981) [JETP Lett. **33**, 644 (1981)]; Nucl. Phys. B **203**, 311 (1982); F. Wilczek, Phys. Rev. Lett. **48**, 1144 (1982).

[14]Xin-zhou Li, Ke-lin Wang, and Jian-zu Zhang, Nuovo Cimento A **75**, 87 (1983); **80**, 311 (1983).

[15]Xin-zhou Li, Ke-lin Wang, and Jian-zu Zhang, Phys. Lett. B **140**, 209 (1984).

[16]Ke-lin Wang and Jian-zu Zhang, Phys. Energi. Fort. Phys. Nuclear **9**, 161 (1985).

[17]Xin-zhou Li and Jian-zu Zhang, Phys. Rev. D **33**, 562 (1986).

[18]Xin-zhou Li, Feng Yu, and Jian-zu Zhang, Phys. Rev. D **34**, 1124 (1986).

[19]Yong-chang Qi, Phys. Lett. B **176**, 115 (1986).

[20]Yong-chang Qi, Phys. Energi. Fort. Phys. Nuclear. **10**, 545 (1986).

[21]H. J. Lipkin, W. I. Weisberger, and M. Peshkin, Ann. Phys. (N.Y.) **53**, 203 (1969).

[22]T. T. Wu and C. N. Yang, Phys. Rev. D **16**, 1018 (1977).

[23]T. Dray, J. Math. Phys. **26**, 1030 (1985).

[24]T. Dray, J. Math. Phys. **27**, 781 (1986).

[25]V. B. Berestetskii, E. M. Lifshitz, and L. P.Pitaevskii, *Relativistic Quantum Theory, Part 1* (Pergamon, New York, 1971).

[26]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960).