The special modular transformation for polycnoidal waves of the Korteweg–de Vries equation

John P. Boyd
Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 22 March 1984; accepted for publication 22 June 1984)

The modular transformation of the Riemann theta function is used to show that the implicit dispersion relation for the N-polycnoidal waves of the Korteweg–de Vries equation has a countable infinity of branches for $N \geqslant 2$. Although the transformation also implies that each branch or mode can be written in a countable infinity of ways, it is also shown that there is a unique "physical" representation for each mode such that the parameters of the theta function can be interpreted as wavenumbers and amplitudes in the limit of either very small or very large amplitude. Unfortunately, the small amplitude "physical" representation is different (by a modular transformation) from the large amplitude "physical" representation for a given mode, but this difference explains an apparent paradox as described in the text. The general modular transformation expresses the theta function in terms of complex wavenumbers, phase speeds, and coordinates that have no physical relevance to the Korteweg–de Vries equation, but it is shown that for $N \geqslant 2$, there is a subgroup, here dubbed the "special modular transformation," which gives a real result. This subgroup is explicitly constructed for general N and presented as a table for $N = 2$.

PACS numbers: 02.30.Jr, 02.60.Lj, 02.30.Qy

I. INTRODUCTION

This present work will focus on four themes. The first is the specialization of the general modular transformation of the theta function to that subgroup, here dubbed the "special" modular transformation, which is relevant to the Korteweg–de Vries equation. This construction is done in Secs. II, III, and IV, beginning with a description of the general transformation, then explicitly constructing the "special" transformation for general N, and finally discussing in detail the special case $N = 2$, which is the subject of the companion papers by the author.[1]

The second half of the paper will discuss in turn the three remaining issues: the multiplicity of roots of the N-polycnoidal wave dispersion relation for $N \geqslant 2$ (Sec. V); the so-called "paradox of the wavenumbers" (Sec. VI); and finally which of the infinite number of mathematically equivalent ways of writing the theta function is the "physical" representation in which the wavenumbers and phase speeds are those of the actual solitons or sine waves of the solution (Sec. VII). Before turning to the transformation itself, it is useful to describe each of these last three themes in enough detail to motivate the technical discussion of Secs. II, III, and IV.

The implicit dispersion relation for N-polycnoidal waves, derived in Refs. 1 and 2, is linear in all the unknowns for the special case $N = 1$ (the ordinary cnoidal wave discovered by Korteweg and de Vries in 1895) and thus has a unique solution. However, for $N \geqslant 2$, the dispersion relation is transcendentally nonlinear—that is, the algebraic equations we must solve are defined by infinite series in one or more unknowns—so an infinite number of roots is at least possible. The mere existence of the modular transformation raises this to a near certainty. As reviewed in Ref. 1, the N-polycnoidal wave is the second logarithmic derivative with respect to x of the N-dimensional Riemann theta function $\theta \begin{bmatrix} 0 \\ 0 \end{bmatrix} (\mathbf{T}, \zeta)$ whose N arguments are the "phase variables"

$$\zeta_j \equiv k_j (x - c_j t) + \phi_j, \tag{1.1}$$

where x is the spatial coordinate, t is time, and where the constants k_j, c_j and ϕ_j are the jth wavenumber, phase speed, and phase factor, respectively. The coefficients of the N-dimensional Fourier series in the ζ_j of the theta function are completely determined by the elements of the symmetric $N \times N$ "theta matrix" $\mathbf{T}$ as shown explicitly in Eq. (2.1) below. The special modular transformation allows the ζ_j and t_{ij} (theta matrix elements) to be simultaneously altered in a countable infinity of ways without changing the sum of the infinite series. Thus, a single root of the dispersion relation generates the solution at an infinite number of discrete points in parameter space.

A more precise picture can be obtained by looking at the limits of either very large amplitude or very small amplitude where the phase speeds c_j are known analytically,[1] and examining what we shall name the "modes" of the wave. In the small amplitude regime, the diagonal theta matrix elements $t_{ii} \gg 1$ and the N-polycnoidal wave can be approximated as the sum of N sine waves each proportional to $\cos(2\pi\zeta_j)$ for a different j. Without loss of generality,[2] one can always rescale the N-polycnoidal wave to unit period. For $N = 2$, we can thus always take $k_1 = 1$, but strict periodicity in x is preserved if $k_2 = n$, where n is any integer $\geqslant 2$. Because each different choice for k_2 gives a distinct solution, one whose graph is distinct from that for any other choice of n, we will refer to the different possibilities as "modes" and write their wavenumbers in square brackets separated by a comma, viz., $[1,n]$.

Now when the modular transformation is applied to a theta function, it alters a wavenumber by an integer. Thus, each mode $[1,n]$ can always be expressed in terms of a theta function with $k_1 = 1$ and $k_2 = 2$. Thus, the dispersion relation for the gravest [1,2] mode has roots corresponding to the modular transformations of all the other modes [1,*n*]. In Sec. V, we will tighten this argument and look at the multiple roots of the dispersion relation in some detail.

The "paradox of wavenumbers" arises because one can equally well define the "modes" of the N-polycnoidal wave in terms of the limit of large amplitude. In this regime, the peaks of the wave are very tall and narrow and essentially indistinguishable (because of their narrowness) from the solitary waves (solitons) of the spatially unbounded (as opposed to periodic) Korteweg–de Vries equation. There are solitons of N different sizes on each interval and, as explained in Ref. 1, the role of the wavenumbers k_j is quite different from that in the small amplitude limit in that the k_j specify how many solitons of the jth size appear on each spatial period. To emphasize the different role of wavenumbers, we shall denote the modes as identified in the near-soliton regime by writing the wavenumbers in braces.

The "paradox" referred to above is that $\{1,1\}$ is now the gravest $N = 2$ mode (one tall soliton and one short soliton on each unit spatial interval), whereas the simplest small amplitude mode is [1,2]. Since the linear dispersion relation gives a unique phase speed in the limit of infinitesimal waves, it is not possible to superimpose two sine waves with $k_1 = k_2 = 1$ and obtain two distinct phase speeds; such a mode would collapse into an ordinary ($N = 1$) cnoidal wave. Since the wavenumbers are fixed parameters of the implicit dispersion relation, this apparent contradiction about the identity of the gravest mode is very confusing.

In Sec. VI, this paradox will be resolved with the aid of the special modular transformation. There, it is shown that if one sets $k_1 = k_2 = 1$ and begins to vary the amplitude downwards in small steps, solving the dispersion relation numerically at each step starting from the known soliton velocities, one will eventually compute the [1,2] small amplitude mode. However, the numerical phase speeds obtained with $k_1 = k_2 = 1$ will not be those of the [1,2] mode directly, but rather the modular transformation of these phase speeds. Thus, the paradox is resolved: the gravest [1,2] and $\{1,1\}$ modes are indeed the same, continuous mode, but the equivalence can be demonstrated only via the special modular transformation.

This in turn raises the third issue. Given that a particular N-polycnoidal wave can be written in a countable infinity of ways thanks to the modular transformation, what representation, if any, is best? In the limit of very large or very small amplitude, we have already answered this question: For the gravest $N = 2$ mode, for example, the [1,2] ($\{1,1\}$) representation is best for small (large) amplitude because the phase speeds can then be physically interpreted as the actual phase speeds of the two linear sine waves or of the two solitons, respectively. What is to be done for intermediate amplitude, however? In Sec. VII, we will attempt to answer this question and give several alternative ways of quantifying the (thus far) vague meaning of "small amplitude" and "large amplitude."

The final section of the paper is a summary and prospectus.

II. THE GENERAL MODULAR TRANSFORMATION

The N-polycnoidal wave is the second logarithmic derivative with respect to x of the N-dimensional Riemann theta function $\theta\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}(\mathbf{T},\boldsymbol{\zeta})$, where the N arguments ("phase variables") are $\boldsymbol{\zeta} = (\zeta_1,\zeta_2,\ldots,\zeta_n)$ with $\zeta_j = k_j(x - c_j t) + \phi_j$ as in (1.1) above and where the theta function is defined by the uniformly convergent sum

$$\theta\begin{bmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{\epsilon}'\end{bmatrix}(\mathbf{T},\boldsymbol{\zeta}) = \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}\cdots\sum_{n_N=-\infty}^{\infty}\exp\Bigg[\pi i\Bigg(\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij} \times\left(n_i+\frac{\epsilon_i}{2}\right)\left(n_j+\frac{\epsilon_j}{2}\right) + 2\sum_{i=1}^{n}\left(n_i+\frac{\epsilon_i}{2}\right)\left(\zeta_i+\frac{\epsilon_i'}{2}\right)\Bigg)\Bigg], \tag{2.1}$$

where $\boldsymbol{\epsilon} = (\epsilon_1,\epsilon_2,\ldots,\epsilon_N)$ and $\boldsymbol{\epsilon}' = (\epsilon_1', \epsilon_2',\ldots,\epsilon_N')$ are together the "characteristic" of the theta function and the t_{ij} are the elements of the $N\times N$ symmetric "theta matrix" $\mathbf{T}$. In Ref. 1 and also the works of Nakamura,[3] Hirota and Ito,[4] and their collaborators, the details of calculating the phase speeds c_j and off-diagonal theta matrix elements ($t_{ij}, i\neq j$) ["unknowns"] in terms of the wavenumbers k_j and diagonal theta matrix elements t_{ii} ["parameters"] are explained. In this work, however, we shall concentrate solely on transformations of the theta function.

The most general transformations are lucidly described in a recent book by Rauch and Farkas.[5] So as to conform with their notation and that of most other mathematics texts, this paper will use theta matrix elements t_{ij} that are imaginary in contrast to the real matrix elements T_{ij} and R_{ij} which are more convenient in the two companion papers (Ref. 1). The results given in Table I, however, are notation-independent as explained in the table caption. If N is the dimensionality of the theta function (mathematicians often use g for N because N is also the "genus" of the Riemann surface associated with the theta function), then the transformations are generated by a $(2N)\times(2N)$ dimensional matrix $\mathbf{M}$ which is a member of Sp(N,Z), the so-called "homogeneous symplectic modular group." The term "symplectic" means that if one defines an $(2N)\times(2N)$ matrix $\mathbf{J}$ via

$$\mathbf{J}\equiv\begin{pmatrix}\mathbf{0}_N & -\mathbf{I}_N\\ -\mathbf{I}_N & \mathbf{0}_N\end{pmatrix},$$

where $\mathbf{0}_N$ is the $N\times N$ matrix whose elements are all zeros and $\mathbf{I}_N$ is the $N\times N$ identity matrix, then for any matrix $\mathbf{M}$ in the general $2N\times 2N$ symplectic group,

$$\mathbf{M}\mathbf{J}\mathbf{M}^T = \mathbf{J}, \tag{2.2}$$

where $\mathbf{M}^T$ is the transpose of $\mathbf{M}$. One can show that (2.2) implies

$$\det|\mathbf{M}| = \pm 1, \tag{2.3}$$

where "det" denotes the determinant. "Modular" denotes that subgroup of the general symplectic group whose matrix elements are all integers. Rauch and Farkas[5] explain why $\mathbf{M}$ must be both symplectic and modular, but their careful and readable exposition will not be repeated here.

The individual transformations are actually expressed in terms of submatrices of **M**, so Rauch and Farkas write **M** in the block form

$$\mathbf{M} \equiv \begin{pmatrix} \mathbf{D} & \mathbf{C} \\ \mathbf{B} & \mathbf{A} \end{pmatrix}, \tag{2.4}$$

where **A**, **B**, **C** are all $N \times N$ matrices. They prove the following.

Theorem: (Modular Transformation) If **M** is a member of the $(2N)\times(2N)$ homogeneous symplectic modular group and if **A**, **B**, **C**, and **D** are its $(N \times N)$ submatrices as defined by (2.4) above, then if

$$\hat{\boldsymbol{\zeta}} = [(\mathbf{CT} + \mathbf{D})^T]^{-1}\boldsymbol{\zeta}, \tag{2.5}$$

$$\hat{\mathbf{T}} = (\mathbf{AT} + \mathbf{B})\,(\mathbf{CT} + \mathbf{D})^{-1}, \tag{2.6}$$

where **T** is symmetric and positive definite imaginary and

$$\hat{\boldsymbol{\epsilon}} = \mathbf{D}\boldsymbol{\epsilon} - \mathbf{C}\boldsymbol{\epsilon}' + \operatorname{diag}(\mathbf{CD}^T), \tag{2.7}$$

$$\hat{\boldsymbol{\epsilon}}' = -\mathbf{B}\boldsymbol{\epsilon} + \mathbf{A}\boldsymbol{\epsilon}' + \operatorname{diag}(\mathbf{AB}^T), \tag{2.8}$$

where diag (**R**) is the vector-valued function of an arbitrary square matrix that returns the diagonal matrix elements of its argument as its result, i.e.,

$$\operatorname{diag}(\mathbf{R}) \equiv \begin{pmatrix} r_{11} \\ r_{22} \\ \vdots \\ r_{NN} \end{pmatrix}, \tag{2.9}$$

then

$$\theta\begin{bmatrix} \hat{\boldsymbol{\epsilon}} \\ \hat{\boldsymbol{\epsilon}}' \end{bmatrix}(\hat{\boldsymbol{\zeta}},\hat{\mathbf{T}}) = K \exp\left[\pi i\left(\sum_{k=1}^{N}\sum_{l=1}^{N} P_{kl}\zeta_k\zeta_l\right)\right]\theta\begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}' \end{bmatrix}(\boldsymbol{\zeta},\mathbf{T}), \tag{2.10}$$

where the P_{kl} are the elements of the $N \times N$ matrix **P**, where

$$\mathbf{P} = (\mathbf{CT} + \mathbf{D})^{-1}\mathbf{C} \tag{2.11}$$

and where K is a constant dependent on **M**, $\boldsymbol{\epsilon}$, and $\boldsymbol{\epsilon}'$. Furthermore, $\hat{\mathbf{T}}$ is a theta matrix, i.e., symmetric and positive definite imaginary.

This theorem is far too general for the theory of the Korteweg–de Vries equation. First, the Korteweg–de Vries solution $u(x,t)$ is proportional to the second logarithmic derivative with respect to x of the theta function, so the constant K in (2.10), which is explicitly computed by Rauch and Farkas, is irrelevant to the theory of the Korteweg–de Vries equation and shall be ignored here.

Second, since the theta matrix **T** is positive definite imaginary, i.e., it must be complex, most of the transformations described by the theorem will yield complex phase variables $\hat{\zeta}_j$ even if the untransformed variables are all real. Since complex coordinates are physically meaningless for the waves of the Korteweg–de Vries equation,[6] it follows that one loses nothing by concentrating only on that subgroup of transformations which yields real coordinates.

Inspecting (2.5) reveals two possibilities for such a subgroup: (i) $\mathbf{D} = 0$, in which case $\hat{\zeta}$ is pure imaginary and all the factors of $i = \sqrt{-1}$ in (2.5) and the theta series (1.2) cancel to give a series involving real x and t only or (ii) $\mathbf{C} = 0$ so that $\hat{\zeta}$ is real if $\boldsymbol{\zeta}$ is real. The first possibility is equivalent to applying the Poisson summation method to rewrite the Fourier series of the theta function as a series of Gaussian functions as explained in an earlier paper by the author.[2] Strictly speaking, the Poisson sum is merely the special case $\mathbf{C} = -\mathbf{I}_N, \mathbf{B} = \mathbf{I}_N$. One can easily show, however, that using the most general **C**, **B** allowed by the symplectic condition (2.2) is equivalent to possibility (ii) for some **A**, **B** followed by Poisson summation. Consequently, (i) adds nothing to (ii) except the possibility of Poisson summation which was already thoroughly explored in earlier work. Therefore, the rest of this article will focus on the second case $\mathbf{C} = 0$. Keeping **T** pure imaginary then requires that $\mathbf{B} = 0$, so the general sympletic modular transformation has been reduced to those for which **A** and **D**, the diagonal submatrices, are the only nonzero blocks.

By substituting such a block diagonal **M** into the symplectic condition (2.2) and noting

$$\mathbf{M}^T = \begin{pmatrix} \mathbf{D}^T & 0 \\ 0 & \mathbf{A}^T \end{pmatrix} \tag{2.12}$$

one can show that the symplectic condition is satisfied if and only if

$$\mathbf{A} = [\mathbf{D}^{-1}]^T. \tag{2.13}$$

Rauch and Farkas[5] give a complete set of that finite number of generators whose products and inverse give the most general symplectic matrix which is also modular, i.e., has integral elements. For the special case considered here, one can discuss these $(2N)\times(2N)$ generating matrices in terms of a single $N \times N$ block (say the **A** block) because the rest of **M** is specified uniquely by $\mathbf{B} = \mathbf{0}$, $\mathbf{C} = \mathbf{0}$, and $\mathbf{D} = [\mathbf{A}^T]^{-1}$ according to (2.13). In the next section, these generators will be explicitly constructed.

III. THE SPECIAL MODULAR TRANSFORM

The modular transformation with $\mathbf{B} = \mathbf{C} = \mathbf{0}$ so the **M** is block diagonal will be referred to as the "special" modular transformation. As shown in the previous section, the "special" transformation is, excluding Poisson summation, the most general transformation of theta functions which is physically relevant to polycnoidal waves of the Korteweg–de Vries equation. It is useful to restate the theorems of the previous section for this special case.

Theorem: (special modular transformation) Let

$$\hat{\boldsymbol{\zeta}} = \mathbf{A}\boldsymbol{\zeta}, \tag{3.1}$$

$$\hat{\mathbf{T}} = \mathbf{ATA}^T, \tag{3.2}$$

$$\hat{\boldsymbol{\epsilon}} = [\mathbf{A}^{-1}]^T\boldsymbol{\epsilon}, \tag{3.3}$$

$$\hat{\boldsymbol{\epsilon}}' = \mathbf{A}\boldsymbol{\epsilon}', \tag{3.4}$$

where **A** is the lower right-hand block of a block-diagonal matrix **M** of the symplectic modular group. (The general form of **A** is constructed below.) Then

$$\theta\begin{bmatrix} \boldsymbol{\epsilon} \\ \hat{\boldsymbol{\epsilon}}' \end{bmatrix}(\hat{\boldsymbol{\zeta}};\hat{\mathbf{T}}) = K\theta\begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}' \end{bmatrix}(\boldsymbol{\zeta};\mathbf{T}), \tag{3.5}$$

where K is a constant.

Note that the Gaussian factor has disappeared because the matrix **P** of the general theorem is now identically zero; the special transformation takes a theta Fourier series directly into another theta Fourier series.

The generators are of two classes. For each, one begins with the $N \times N$ identity matrix and modifies it according to a prescription given by Rauch and Farkas. For the first class, ${}^{+}A_{ij}$ in their notation with i,j restricted so that $i \neq j$, add -1 to the (j,i) element of the identity matrix. For the second class D_i change the sign of the (i,i) element of the identity matrix. Rauch and Farkas also include the inverses of the ${}^{+}A_{ij}$ in their generator set as the ${}^{-}A_{ij}$; the D_i are their own inverses. (The ${}^{-}A_{ij}$ are obtained by adding $+1$ to the (j,i) element of the identity matrix.) One thus obtains a complete (but not necessarily minimal) generating set with $2N^2 - N$ members. The statement that this set is the complete generator of the special transformations means that the most general $N \times N$ matrix A which appears in (3.1)–(3.3) is the product of an arbitrary number of the generating matrices each raised to an arbitrary, non-negative power, i.e., if one adopts the revised notation of labeling the generators A_i, $i = 1,2,\ldots,2N^2 - N$, then the most general transformation is

$$A = A_{i_1}^{n_1} A_{i_2}^{n_2} A_{i_3}^{n_3} \cdots A_{i_m}^{n_m}, \tag{3.6}$$

where $n_j \geqslant 0$, $1 \leqslant i_j \leqslant 2N^2 - N$ but are otherwise arbitrary; m, the number of factors, is arbitrary also. The generators do not commute even for the special transformation, so (3.6) usually cannot be simplified.

TABLE I. The transformations produced by the generators of the special modular group $\mathbf{A}_1$ and $\mathbf{A}_2$, and their inverses for $N = 2$ (double cnoidal wave). The plus signs correspond to $\mathbf{A}_1$; the minus signs to $\mathbf{A}_1^{-1}$ and $\mathbf{A}_2^{-1}$. The (^) quantities are the new variables created by the transformation. The matrices whose elements are T_{ij} and R_{ij} are explained in Boyd[1]; the T_{ij} transform exactly as those of the usual imaginary theta matrix elements t_{ij} used by mathematicians. The k_i and c_i are the wavenumbers and phase speeds that appear in the "angle" variables X and Y when the special modular transformation is applied to the double cnoidal wave of the Korteweg–de Vries equation.

$\mathbf{A}_1$ (and $\mathbf{A}_1^{-1}$):

$$\begin{pmatrix}\hat{X}\\ \hat{Y}\end{pmatrix} = \begin{pmatrix}X \pm Y\\ Y\end{pmatrix},$$

$$\begin{pmatrix}\hat{T}_{11} & \hat{T}_{12}\\ \hat{T}_{12} & \hat{T}_{22}\end{pmatrix} = \begin{pmatrix}[T_{11} \pm 2T_{12} + T_{22}] & [T_{12} \pm T_{22}]\\ [T_{12} \pm T_{22}] & [T_{22}]\end{pmatrix},$$

$$\begin{pmatrix}\hat{R}_{11} & \hat{R}_{12}\\ \hat{R}_{12} & \hat{R}_{22}\end{pmatrix} = \begin{pmatrix}[R_{11}] & [R_{12} \mp R_{11}]\\ [R_{12} \mp R_{11}] & [R_{11} \mp 2R_{12} + R_{22}]\end{pmatrix},$$

$$\begin{pmatrix}\hat{k}_1\\ \hat{k}_2\end{pmatrix} = \begin{pmatrix}k_1 \pm k_2\\ k_2\end{pmatrix},$$

$$\begin{pmatrix}\hat{c}_1\\ \hat{c}_2\end{pmatrix} = \begin{pmatrix}[k_1c_1 \pm k_2c_2]/[k_1 \pm k_2]\\ k_2\end{pmatrix},$$

$\mathbf{A}_2$ (and $\mathbf{A}_2^{-1}$):

$$\begin{pmatrix}\hat{X}\\ \hat{Y}\end{pmatrix} = \begin{pmatrix}X\\ Y \pm X\end{pmatrix},$$

$$\begin{pmatrix}\hat{T}_{11} & \hat{T}_{12}\\ \hat{T}_{12} & \hat{T}_{22}\end{pmatrix} = \begin{pmatrix}[T_{11}] & [T_{12} \pm T_{11}]\\ [T_{12} \pm T_{11}] & [T_{11} \pm 2T_{12} + T_{22}]\end{pmatrix},$$

$$\begin{pmatrix}\hat{R}_{11} & \hat{R}_{12}\\ \hat{R}_{12} & \hat{R}_{22}\end{pmatrix} = \begin{pmatrix}[R_{11} \mp 2R_{12} + R_{22}] & [R_{12} \mp R_{22}]\\ [R_{12} \mp R_{22}] & [R_{22}]\end{pmatrix},$$

$$\begin{pmatrix}\hat{k}_1\\ \hat{k}_2\end{pmatrix} = \begin{pmatrix}k_1\\ k_2 \pm k_1\end{pmatrix},$$

$$\begin{pmatrix}\hat{c}_1\\ \hat{c}_2\end{pmatrix} = \begin{pmatrix}c_1\\ [k_2c_2 \pm k_1c_1]/[k_2 \pm k_1]\end{pmatrix}.$$

The D_i generators merely replace $\zeta_i \rightarrow -\zeta_i$. This is not a very interesting transformation since one can always take $\theta\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ or $\theta\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}$, whichever is convenient, as the theta function by adjusting the phase factors ϕ_j in (1.1). These theta functions are always even in each of $\zeta_1,\zeta_2,\ldots,\zeta_N$, so the transformation described by the N D_i invariably leaves the theta function unaltered and is physically irrelevant. Consequently, the only interesting nontrivial transformations are those generated by the smaller set of the $2N^2 - 2N$ matrices that Rauch and Farkas label ${}^{+}A_{ij}$ and ${}^{-}A_{ij}$.

Rauch and Farkas[5] prove that $\mathbf{T}$ is a theta matrix, i.e., symmetric and positive definite imaginary, for the general modular transformation of any N, so the transformed theta function series is always convergent. Although the determinant of $\mathbf{T}$ is invariant under transformation (proof: $\det \mathbf{A}_i = 1$ for all i and the determinant of the product of two arbitrary matrices is the product of their determinants), the trace of $\mathbf{T}$, i.e., the sum of the diagonal elements, generally is altered by the transformation as evident in Table I. Since the trace of the theta matrix is the sum of the eigenvalues, it follows that the eigenvalues, and therefore the rate of convergence of the series are normally changed by the special modular transformation even though the fact of convergence (does it converge or diverge?) is never altered.

These matrices ${}^{\pm}A_{ij}$ exist only for $N \geqslant 2$, i.e., for the double cnoidal wave or higher. All modular transformations for the ordinary cnoidal wave ($N = 1$) yield a result in which the spatial coordinate has both real and imaginary parts except the Poisson summation discussed in the author's earlier paper.[1] In the next section, the simplest nontrivial case $N = 2$ will be described and its generators will be given explicitly.

IV. THE SPECIAL MODULAR TRANSFORMATION FOR $N = 6$: DOUBLE CNOIDAL WAVE

In Rauch and Farkas[5] terminology, there are four generators for $N = 2$, but since half of these are inverses of the other half, there are only two generators[7] in the usual terminology of group theory where inverses are not counted, i.e.,

$$A_1 = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \tag{4.1}$$

$$A_2 = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}. \tag{4.2}$$

[The inverses are obtained by merely changing the sign of the off-diagonal element for both (4.1) and (4.2).] From Table I, it is apparent that the transformations wrought by A_1 can be obtained from those created by A_2 by interchanging the X and Y coordinates and the subscripts on the theta matrix elements: there is effectively only a single transformation (and its inverse) which can, however, be applied to replace either X or Y by the sum $(X \pm Y)$.

In the companion papers by Boyd,[1] it was convenient to work in terms of modified theta matrices whose elements are denoted by T_{ij} and R_{ij}. Since,

$$T_{ij} \equiv \pi \operatorname{Im}(t_{ij}) \tag{4.3}$$

the T_{ij} transform exactly as the complex elements t_{ij} used by mathematicians and vice versa.

The matrix elements R_{ij}, which give the coefficients of the Gaussian series of the theta function, are those of a matrix $\mathbf{R}$ which is proportional to the inverse of $\mathbf{T}$. When $\mathbf{T} \rightarrow \hat{\mathbf{T}} = \mathbf{A}\mathbf{T}\mathbf{A}^T$, the corresponding matrix $\mathbf{R} = (2\pi^2)\,\mathbf{T}^{-1}$ transforms as

$$\hat{\mathbf{R}} = [\mathbf{A}^T]^{-1}\,\mathbf{R}\mathbf{A}^{-1}. \tag{4.4}$$

Equation (4.4) applies for the special modular transformation of any N; the transformed R for the special case $N = 2$ are also given in Table I.

Recalling that $X = k_1\ (x - c_1 t) + \phi_1$ and $Y = k_2\ (x - c\,t) + \phi_2$, it follows that the phase speeds and wavenumbers are also altered by the transformation. The changes made by the generators and their inverses are given in Table I.

V. THE BRANCHING OF THE DISPERSION RELATION

As shown in the Introduction, the existence of the modular transformation implies that each mode of the 2-polycnoidal wave can be transformed into a theta function with $k_1 = 1$, $k_2 = 2$. This suggests that the dispersion relation for any given set of parameters is infinitely multibranched. At a minimum, it has been shown that if we attempt to make a contour plot of c_1, c_2, and t_{12} for a single mode, say the gravest, then all the higher modes, which are countably infinite in number, provide extraneous roots of the dispersion relation for at least some values of the parameters.

Two issues remain. First, when $k_1 = 1$ and $k_2 = 2$, for example, does the [1,3] mode give a solution of the implicit dispersion relation for all values of (t_{11}, t_{22}) or only for some limited region in parameter space? It is reasonable to conjecture that the correct answer is "all" since the modular transformation is not subject to any parametric restrictions: The transform of a theta function is always a uniformly convergent series of proper theta function form. However, the transformed values of the diagonal theta matrix elements depend on the off-diagonal elements which unfortunately are part of the solution of the dispersion relation, so this conjecture cannot be proved without deeper analysis than that done here.

The second issue is whether all solutions of the implicit dispersion relation are "regular," that is to say, are modes which have well-defined limits for small and large amplitude. Here again, one is tempted to speculate that the answer is that all solutions are regular. The N-polycnoidal waves have an analytic structure which is extraordinarily tidy and simple in other respects. Furthermore, as shown by Boyd,[1] the implicit dispersion can be solved by perturbation theory for both small and large amplitude with good overlap between the two regimes; the overlap implies it is difficult to insert irregular modes that exist only for intermediate parameter values. However, these are not proofs but plausibility arguments. It is known that at least one physically important type of solitary wave, the "modon" of Stern, Flierl, McWilliams, Larichev, and Reznik,[9] does not have a small amplitude limit, but exists only when the amplitude is above some threshold.

Thus, a full resolution of these two issues must remain for future research. It is certain, however, that the implicit dispersion relation for N-polycnoidal waves has multiple solutions for $N \geqslant 2$.

VI. THE WAVENUMBER PARADOX AND THE CONTINUATION METHOD

A good numerical procedure for tracing the structure of a mode is the so-called "continuation" method. The basic idea is to vary one or more of the parameters in small steps. At each step, the algebraic equations are solved for the unknowns via Newton's method. The first guess which is needed to initialize the Newton iteration is obtained by using the results from the previous point in parameter space (or by linear extrapolation from the results at the two previous points); this will always give convergence if the steps in the parameter are sufficiently small. To initialize the parameter march, one needs to know the approximate solution at some point in parameter space. The perturbation theory of Boyd[1] provides such approximate solutions for both very large and very small amplitudes i.e., for very large and small values of t_{ii}, so the continuation method can be easily applied to all the regular solutions of the implicit dispersion relation.

The "paradox of the wavenumbers" is that the gravest mode in the near-soliton limit is $\{1,1\}$, i.e., has $k_1 = k_2 = 1$ while the lowest mode in the small amplitude regime is denoted [1,2] because $k_2 = 2k_1 = 2$. When the continuation method is applied to the $\{1,1\}$ mode, beginning in the large amplitude, near-soliton regime, and both diagonal theta matrix elements are simultaneously increased, what does the algorithm give when $|t_{ii}| \gg 1$? This limit is the small amplitude regime where a mode that is [1,1], i.e., the sum of two sine waves of different phase speeds but identical wavenumbers, cannot exist. Table II answers this question: The continuation method, applied with $k_1 = k_2 = 1$, gives values for c_2 which are the modular transform of those for the gravest small amplitude mode [1,2].

The top line of the table shows the near-soliton regime; the $|t_{ii}|$ are fairly small and the "Gaussian," large amplitude perturbation theory[1] is quite accurate. By the time the bottom is reached, we are in the small amplitude regime and the Gaussian perturbation theory is inaccurate while perturbation theory beginning with two Fourier components as the lowest approximation gives excellent results. However, the continuation method gives $c_2 = -28\pi^2$ in the limit whereas $\cos(2\pi[x - c_1 t])$ and $\cos(4\pi[x - c_2 t])$ have small-amplitude phase speeds of $-4\pi^2$ and $-16\pi^2$, respectively.

TABLE II. Numerical solutions of the residual equations, obtained through the "continuation"method, are compared with zeroth-order Fourier perturbation theory (linear sine waves) and Gaussian perturbation theory (two solitary waves). The percentages are the relative errors. The wavenumbers k_1 and k_2 for the numerical and solitary wave calculations are both equal to 1; $k_2 = 2$ for the Fourier series approximation. The variable c_2^{mod} ($= [c_1 + c_2]/2$) is the second phase speed after a modular transformation via the generator matrix A_2. This transformation leaves c_1 and k_1 unchanged, but it alters k_2 from 1 to 2.

T_{11}	T_{22}	c_1	c_2	c_2^{mod}	c_1 error	c_2 error
0.8	1.6	313.16	− 402.08	− 44.46		
	(Gaussian)	313.15	− 402.06		9.E-6	6.E-5
	(Fourier)	− 39.48		− 157.92	893.23%	71.84%
1.2	2.4	73.30	− 342.29	− 134.49		
	(Gaussian)	73.31	− 341.68		0.02%	0.18%
	(Fourier)	− 39.48		− 157.91	205.67%	14.83%
1.6	3.2	4.17	− 308.46	− 152.14		
	(Gaussian)	4.27	− 305.72		2.42%	0.89%
	(Fourier)	− 39.48		− 157.91	89.44%	3.65%
2.0	4.0	− 21.14	− 291.40	− 156.27		
	(Gaussian)	− 20.61	− 285.52		2.53%	2.02%
	(Fourier)	− 39.48		− 157.91	46.45%	1.04%
2.4	4.8	− 31.49	− 283.25	− 157.37		
	(Gaussian)	− 29.61	− 274.83		5.97%	2.97%
	(Fourier)	− 39.48		− 157.91	20.25%	0.35%
2.8	5.6	− 35.94	− 279.48	− 157.71		
	(Gaussian)	− 31.11	− 270.65		13.44%	3.16%
	(Fourier)	− 39.48		− 157.91	8.97%	0.13%
3.2	6.4	− 37.90	− 277.76	− 157.83		
	(Gaussian)	− 27.75	− 272.01		26.76%	2.07%
	(Fourier)	− 39.48		− 157.91	4.01%	0.05%

Agreement comes only after making a modular transformation that converts the theta function representation from one with $k_2 = 1$ to $k_2 = 2$, which is the actual wavenumber of one of the two dominant components of this mode in this limit of $|t_{ii}| \to \infty$.

It is useful to see explicitly how a mode can thus diguise itself. The lowest four terms of the Fourier series of the theta function are (in any representation)

$$\theta = 1 + e^{-T_{11}} \cos(2\pi X) + e^{-T_{22}} \cos(2\pi Y) + e^{-T_{11} - T_{22} - 2T_{12}} \cos(2\pi[X + Y]), \quad (6.1)$$

where

$$X = k_1(x - c_1 t), \quad Y = k_2(x - c_2 t). \quad (6.2)$$

The author's companion paper on perturbation theory[1] shows how to evaluate the phase speeds and T_{12} in the "physical" representation $[1,2]^P$; the results are compared against c_1 and c_2^{mod} in Table II. It is also possible, although one would never want to do it except to make a point, to calculate perturbatively in the "unphysical" [1,1] representation as done in Appendix A of that same paper where it is shown that

$$c_1 = -4\pi^2, \quad (6.3)$$

$$c_2 = -28\pi^2, \quad (6.4)$$

$$T_{12} = -T_{11} + \log(3), \quad (6.5)$$

for $T_{11}, T_{22} \gg 1$. Note tht c_2 in (6.4) is the limit of the numerical calculations in Table II for the column labeled "c_2."

In the physical representation $[1,2]^P$, $T_{12} = \log(3)$ and the Fourier series (6.1) is well approximated by the sum of the first three terms. In the "unphysical" [1,1] representation, however, (6.5) shows that the fourth term in (6.1), proportional to $\cos(2\pi[X + Y])$ is larger than the third by a factor $O(e^{T_{11}})$. Discarding $\cos(2\pi Y)$ and taking the double logarithmic derivative gives, using (6.1)–(6.5), the Korteweg–de Vries solution

$$u(x,t) \doteq -48\pi^2(e^{-T_{11}} \cos[2\pi\{x - (-4\pi^2)t\}] + \tfrac{4}{9} e^{-T_{22} + T_{11}} \cos[2\pi\{2x - (-4\pi^2 - 28\pi^2)t\}]). \quad (6.6)$$

The second term in (6.6) travels at a phase velocity of $-16\pi^2$; it is just the expected second harmonic with $k_2 = 2$. In the [1,1] representation, this term appears in disguise as $\cos(2\pi[X + Y])$ [as opposed to $\cos(2\pi Y)$ in the physical $[1,2]^P$ representation], but this disguise cannot change its physical nature.

In the limit of small amplitude, the $N = 2$ Fourier series for $u(x,t)$ is always dominated by just two terms, but the terms wear different disguises in different representations. The second harmonic is $\cos(2\pi[X + Y])$ in the [1,1] representation, $\cos(2\pi Y)$ in the $[1,2]^P$, $\cos(2\pi[-X + Y])$ in the [1,3], and so on.

The moral of the story is that while one can legitimately solve the dispersion relation using any of the infinite number of disguises for a mode which are allowed by the special modular transformation, there is in general only one representation for which the c_1 and c_2 are the actual rates at which components of the 2-polycnoidal wave are traveling. Identifying this "natural" or "physical" representation is clearly an important issue and is therefore the theme of the next section. Table II also shows that the {1,1} and [1,2] modes are indeed the same, but the equivalence of these two different disguises of the gravest mode of the 2-polycnoidal wave is obvious only through the modular transformation.

VII. THE CANONICAL OR "PHYSICAL" REPRESENTATION

Because of the freedom provided by the special modular transform, each N-polycnoidal wave for $N \geqslant 2$ can wear a countable infinity of diguises. It follows that a major issue is to identify what representation, i.e., what set of wavenumbers k_1 and k_2, give the "physical" representation in which the wavenumbers and phase speeds directly describe the wave.

We will assert, and then demonstrate below, that the following descriptions of the "physical" or canonical representation are equivalent.

(i) It is the representation in which the phase speeds c_1 and c_2 give the actual average rates of travel of the solitons or sine waves.

(ii) It is the representation which the perturbation series of Boyd[1] calculate in, i.e., the perturbation methods automatically give phase speeds which are the true average rates at which the wave crests move.

(iii) It is the representation in which (small amplitude) the off-diagonal matrix elements t_{ij} $(i \neq j)$ are very small in absolute value in comparison to the diagonal elements t_{ii} or equivalently (large amplitude) the off-diagonal elements R_{ij} are small in comparison to the diagonal elements of the inverse theta matrix R_{ii}.

The first description is simply a definition of what we mean by a "physical" representation. For sufficiently large or small amplitude, the N-polycnoidal wave reduces to the usual N-soliton solution or to a sum of N sine waves, so this definition of a canonical representation is always unambiguous if we are sufficiently close to one or the other of these limiting cases.

The second description is an obvious consequence of the first because the zeroth-order solutions of the perturbation theory are the limits of infinitely large or small amplitude. Thus, the wavenumbers that appear in the zeroth-order solution always count the number of solitons on the interval or are the actual wavenumbers of the sine waves, and this is not changed by adding the higher-order corrections.

The third description is consequence of the following theorem proved in Rauch and Farkas.[5]

Theorem: When the theta matrix is diagonal, i.e., $t_{ij} = 0$ if $i \neq j$, then the N-dimensional theta function may be written as the product of N one-dimensional theta functions

$$\theta(\zeta_1,\zeta_2 \ldots,\zeta_N;\mathbf{T}) = \prod_{i=i}^{N} \theta(\zeta_1,t_{ii}). \tag{7.1}$$

The significance of the theorem is that since $u(x,t)$ is proportional to the logarithm of the theta function, each term in the product in (7.1) will contribute additively to the solution of the Korteweg–de Vries equation:

$$u(x,t) = 12 \sum_{i=1}^{N} \frac{d^2}{dx^2} \ln \theta(\zeta_i,t_{ii}). \tag{7.2}$$

This is precisely the situation which occurs in the limits of very large or very small amplitude: the N-polycnoidal wave reduces to a sum of N solitons or sine waves each with its own wavenumber, phase speed, and amplitude. Equation (7.2) and definition (iii) of the "physical" representation are also consistent with the perturbation theory of Boyd[1]: in the limit that the diagonal theta matrix elements (or inverse theta matrix elements) become very large, the off-diagonal elements, which have finite zeroth-order values, necessarily become small relative to the diagonal elements. Thus, both the limiting behavior of the N-polycnoidal wave together with (7.2) and the perturbation theory show that (iii) is true in the physical representation at least for sufficiently large or small diagonal theta matrix elements.

Strictly speaking, of course, no theta matrix for a Korteweg–de Vries solution is ever exactly diagonal; as shown in Appendix B of Ref. 1, the off-diagonal theta matrix elements are responsible for the phase shifts that occur whenever solitons collide. Still, the basic argument is correct, and it can be reversed to justify definition (ii) from (iii). The implicit assumption of the perturbation series of Boyd[1] is that the order of magnitude of different terms in the series can be determined solely from the diagonal theta matrix elements, which is sensible only if the off-diagonal theta matrix elements are small in comparison as required by (iii). Appendix A of the companion paper[1] on perturbation theory is able to calculate perturbation series in an "unphysical" representation only by assuming the diagonal and off-diagonal matrix elements are of the same magnitude. Although no rigorous proof will be given, the fact that the elements of the special modular transformation are always integers strongly suggest that such a transformation will invariably destroy the smallness of the off-diagonal theta matrix elements relative to the diagonal elements at it does in (6.5) so that this smallness is a unique property of the physical representation.

The only flaw with these three equivalent descriptions of the "physical" representation is that they are all in some way tied to the limiting cases of extremes of amplitude or equivalently, of diagonal theta matrix element size. What does one do for intermediate amplitude?

The mathematical response is to use analytic continuation in the parameters: If a given intermediate amplitude solution is the smooth continuation as the parameters are slowly varied of an infinitesimally small amplitude solution whose physical representation is $[1,2]^P$, then this same description is the physical representation of the intermediate amplitude solution, too. Since there is no ambiguity in the limit, there is no ambiguity in this extended definition either as long as the solution branches are continuous with either infinitesimal or infinite amplitude. Note that we use a superscript "P" to denote that the physical representation of a mode is meant, and not one of the infinite number of other representations allowed by the special modular transformation.

There is a remaining physical ambiguity in that Table II shows that the $[1,2]^P$ and $\{1,1\}^P$ representations both describe the same continuous mode: For intermediate amplitude, is it better to describe the polycnoidal wave as a sum of sine waves or of solitons? The answer is given in Ref. 2: For intermediate values of amplitude or of theta matrix elements, both descriptions, as solitons and as sine waves, are qualitatively and even quantitatively correct, and which is better is a matter of individual preference.

VIII. SUMMARY

Mathematicians have known for at least half a century that the theta functions could be expressed in an infinite number of ways via the so-called "modular transformation." The general modular transformation, however, usually gives complex results even though only real values of the space and time variables x and t are relevant to the theory of the Korteweg–de Vries equation. For the ordinary cnoidal wave, which was discovered eighty years ago, the only noncomplex modular transformation is that single transformation which can alternatively be obtained by taking the Poisson sum of the Fourier series of the theta function. Boyd[2] has already discussed the usefulness of this Poisson sum.

It is shown in this paper, however, that for the N-polycnoidal wave with $N\geqslant 2$, where N is the number of arguments of the theta function, there exists a subgroup of the general transformation which does yield nontrivial real results. This subgroup is labeled the "special" modular transformation and is defined to specifically exclude the Poisson sum, which also gives real-valued results but multiplies the theta function by a Gaussian factor. Only the "special" modular transformation and the Poisson sum are useful in the physics of the Korteweg–de Vries equation.

By specializing the rules for the general transformation given in Rauch and Farkas,[5] the generators of the "special" modular transformation are explicitly constructed for arbitrary N. The two generators and their inverses for $N = 2$ are given in Table I above, which also shows the effects of the transformation on the phase speeds and wavenumbers which appear in the "phase" variable that are the arguments of the theta functions.

Since the "special" modular transformation allows each polycnoidal wave to be expressed in an infinite number of ways, a "physical" representation is defined to be that in which the wavenumbers and phase speeds of the theta function match those of the peaks and troughs of the actual wave. Since different polycnoidal waves are obtained for different (physical) wavenumbers, it is helpful to introduce the notation of writing the wavenumbers in [ ] (when θ is represented by a Fourier series) or { } (Gaussian series) and adding a superscript P when the physical representation is meant.

The importance of the special modular transformation in physical applications of the Korteweg–de Vries equations and its cousins is twofold.

First, it shows that the implicit dispersion relation of Boyd[1] for the phase speeds of the polycnoidal wave has an infinite number of solution branches for a given set of parameters (including a fixed set of wavenumbers) even though there is only a single branch for which the wavenumbers are those of the wave's physical representation. Perturbation theory and the "continuation" method are offered as useful ways of computing the physical branch rather than one of the infinite number of other real solutions permitted by the mathematics.

Second, the branch which is the sum of a simple linear sine wave and its second harmonic, $[1,2]^P$ in the notation introduced here, is the sum of one tall solitary wave and one shorter solitary wave on each periodicity interval in the opposite limit of large wave amplitude. This solitary wave limit is written $\{1,1\}^P$; a large amplitude polycnoidal wave for which the wavenumbers $k_1 = 1$ and $k_2 = 2$ are the physical ones is a solution with three solitary waves on each periodicity interval—two of one height and one of a different size. Thus,

$$[1,2]^P = \{1,1\}^P \tag{8.1}$$

in the sense that this single branch must be expressed using a different set of wavenumbers in different amplitude limits if the phase speeds that appear in its theta function are to match those of the actual troughs and crests of the wave.

The special modular transformation is thus essential to understanding the polycnoidal wave because it allows us to change wavenumbers and phase speeds at will so that for any amplitude, we can make the mathematics reflect the physics. Numerically solving the residual equations for fixed wavenumbers, for example, $k_1 = 1$ and $k_2 = 2$, will always give us phase speeds to insert into the theta functions. When we have passed from small wave amplitude to large, however, the phase speeds of the theta functions have only mathematical significance, and differ radically from the actual rates at which the two solitary waves of the branch indicated in (8.1) are traveling, unless we use the special modular transformation to alter the second wavenumber to $k_2 = 1$.

For the Korteweg–de Vries equation and many other soliton equations which are real valued, the special modular transformation (and the Poisson sum discussed in Refs. 1 and 2) are the whole story. Other soliton equations like the cubic Schrödinger equation, however, are intrinsically complex. It is no longer obvious that we should reject the complex-valued transformations which belong to the general modular group but not to the special subgroup defined and constructed here. Future work should explore the physical significance of the general modular transformation for the cubic Schrödinger equation and its complex-valued cousins.

Note added in proof: H. Segur and A. Finkel (unpublished preprint) have applied two-dimensional theta functions and the modular transformation to the Kadomtsev–Petviashvili equation (two space dimensions but only a single phase speed). Their concept of a "basic" theta matrix is an attempt to remove the ambiguity allowed by the modular transformation; in the limits of large and small amplitude, at least, their "basic" matrix is that of the "physical" representation defined here. An earlier work on this same equation (with H. Philander) is "Nonlinear Phenomena" in *Lecture Notes in Physics*, No. 189, edited by K. B. Wolf (Springer-Verlag, Berlin, 1983).

ACKNOWLEDGMENTS

I wish to thank Professor Stanley Jacobs for helpful discussions.

This work was supported by the National Science Foundation through grants OCE 8108530 and OCE 8305648.

[1]J. P. Boyd, J. Math Phys. **25**, 3390(1984); **25**, 3402(1984).
[2]J. P. Boyd, J. Math. Phys. **23**, 375 (1982).
[3]A. Nakamura, J. Phys. Soc. Jpn. **47**, 1701 (1979).
[4]R. Hirota and M. Ito, J. Phys. Soc. Jpn. **50**, 338 (1981).
[5]H. E. Rauch and H. M. Farkas, *Theta Functions with Applications to Riemann Surfaces* (Williams and Wilkins, Baltimore, MD, 1974).
[6]For soliton equations with complex coefficients and also those whose equivalent in Hirota bilinear operators is a coupled set rather than a single equation (the nonlinear Schrödinger equation simultaneously is both) the general modular transformation may have physical significance. Future work will resolve this question.
[7]Interestingly, the matrices (4.1) and (4.2), here used to construct a special subgroup of the 4×4 matrices of the $N=2$ modular group, are themselves the generators of the general symplectic modular group for $N=1$, $\mathrm{Sp}(1,Z)$. This group of 2×2 matrices therefore has a double role in theta function theory: (i) for $N=1$, to give the general transformation of the real and imaginary parts (X_r, X_{im}) of the single complex variable $X=X_r+iX_{im}$ and (ii) for $N=2$, to construct the special modular transformaton of the real parts only of two complex variables (X,Y). [Only application (ii) is physically interesting.] Books on ordinary elliptic function theory such as Rauch and Lebowitz[8] usually take $\mathbf{T}=\mathbf{A}_2$ and $\mathbf{S}=\mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{A}_1$ as the generators, but the same group is generated by (4.1) and (4.2).
[8]H. E. Rauch and A. Lebowitz, *Elliptic Functions, Theta Functions and Riemann Surfaces* (Williams and Wilkins, Baltimore, MD, 1973).
[9]G. R. Flierl, V. D. Larichev, J. C. McWilliams, and G. M. Reznik, Dyn. Atmos. Oceans **5**, 1–41 (1980).