Perturbation series for the double cnoidal wave 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 1983; accepted for publication 22 June 1984)

By means of the theorems proved earlier by the author, the problem of the double cnoidal wave of the Korteweg–de Vries equation is reduced to four algebraic equations in four unknowns. Two of the unknowns are the nonlinear phase speeds c_1 and c_2. Another is a physically irrelevant integration constant. The fourth unknown is the off-diagonal element of the symmetric, 2×2 theta matrix, which in turn gives the explicit coefficients of the Riemann theta function. The double cnoidal wave $u(x,t)$ is then obtained by taking the second x-derivative of the logarithm of the theta function. Two separate forms of these four nonlinear "residual" equations are given. One is obtained from the Fourier series of the theta function and is useful for small wave amplitude. The other is based on the Gaussian series of the theta function and is highly efficient in the large amplitude regime where the double cnoidal wave is the sum of two solitary waves. Both sets of residual equations can be solved via perturbation theory and results are given to fourth order in the Fourier case and second order in the Gaussian case. The Gaussian-based perturbation series has the remarkable property that it converges more and more rapidly as the wave amplitude increases; the zeroth-order solution is the familiar double solitary wave. Numerical comparisons show that the two complementary perturbation series give accurate results in all the important regions of parameter space. (The "unimportant" regions are those in which the double cnoidal wave is an ordinary cnoidal wave subject to a very weak perturbation.) This is turn implies that even for moderate wave amplitude where the nonlinear interactions are not weak, and yet the solitary wave peaks are not well separated, at least to the eye, it is still qualitatively legitimate to describe the double cnoidal wave as either the sum of two sine waves or of two solitary waves of different heights.

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

I. INTRODUCTION

In an earlier work, the author[1] discussed the use of theta functions to study the dynamics of "polycnoidal waves," which is the term coined by the author for the spatially periodic solutions of the Korteweg–de Vries (KdV) and other soliton equations. The general theorems proved there, however, were applied only to the simplest case of the 1-polycnoidal wave, i.e., the ordinary cnoidal wave discovered by Korteweg and de Vries in 1895. This paper is the second article in a three part follow-up[2,3] which will apply the earlier results to the double cnoidal wave of the KdV equation. Throught this paper, the term "double cnoidal" will be used interchangeably with "2-polycnoidal" to denote that generalization of the cnoidal wave which is characterized by two distinct phase speeds, amplitudes, and widths.

One major theme of Ref. 1 is that by using the Gaussian series of the theta function for large amplitude waves and the complementary Fourier series for small amplitude waves, one can calculate the single cnoidal wave through perturbation series to very high accuracy for all values of the parameters. In the worst possible case, which is when the two series converge at equal rates, it was shown that the zeroth-order approximations give the phase speed to within 5% relative error while the first-order approximation is accurate to within 0.03%; the approximations for wave shape are similarly accurate. The purpose of this article is to show that one can also obtain good results for the double cnoidal wave by again deriving two complementary perturbation series based on the Gaussian and Fourier representations of the theta function, respectively.

Although the Korteweg–de Vries equation

$$u_t + u\, u_x + u_{xxx} = 0 \tag{1.1}$$

is a partial differential equation, the theta function series for the double cnoidal wave contains only four free parameters: The coefficients of the infinite series for the theta function are completely specified once these four parameters are known. Independently, Boyd[1] and Nakamura[4] were able to show that the problem of finding the double cnoidal solutions of (1.1) can be reduced to solving a system of four algebraic equations for the theta function parameters. This, together with the overlapping of the complementary large amplitude (Gaussian) and small amplitude (Fourier) expansions, makes it possible to derive efficient, accurate perturbation series that describe both the phase speeds and shape of the double cnoidal wave for all possible values of the parameters.

The next section derives these four algebraic equations, the implicit dispersion relation, for both the Fourier and Gaussian expansions. (The Fourier equations can be obtained as a special case of the Gaussian.) Section III discusses the general method of solving a set of nonlinear equations via perturbation theory. Section IV and V give the actual results for the Fourier and Gaussian expansion, respectively. Mixed Fourier–Gaussian series are described briefly in Sec. VI. The errors in these expansions are discussed in Sec. VII. The paper ends with a final section that summarizes what has gone before and discusses the possibility of extending perturbation theory to other exactly integrable soliton equations.

II. THE RESIDUAL EQUATIONS (IMPLICIT DISPERSION RELATION)

The solution $u(x,t)$ of the KdV equation is related to the theta function via the transformation

$$u = 12(\ln \theta)_{xx}. \tag{2.1}$$

The theta functions themselves satisfy a transformed version of the KdV equation which was first given by Hirota[5] and which will therefore be referred to in what follows as the "Hirota–Korteweg–de Vries" (HKdV) equation. The most compact representation of this bilinear equation is in terms of certain operators introduced by Hirota himself and defined by

$$D_x{}^n D_t{}^m (F\cdot G) \equiv \left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^n \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^m \times F(x,t)G(x',t')\right]_{\substack{x'=x\\ t'=t}} \tag{2.2}$$

where the notation indicates that x' and t' are to be replaced by x and t after the differentiation has been performed. The HKdV equation is then

$$(D_x{}^4 + D_x D_t)(\theta\cdot\theta) = 2A\theta^2 \quad [\text{HKdV}], \tag{2.3}$$

where A is a constant of integration which must be determined in the course of solution.

The theta function solutions of (2.3), dubbed "N-polycnoidal" waves in Ref. 1, are functions of the N-dimensional Riemann theta function. The double cnoidal wave, the only example considered here, is the special case $N = 2$. The classic theta function notation is discussed in part one of this three part sequence (Ref. 2).

Here it will suffice to note that the "phase" or "angle" variables are defined by

$$X = k_1(x - c_1 t) + \phi_1, \tag{2.4}$$

$$Y = k_2(x - c_2 t) + \phi_2, \tag{2.5}$$

where the constants k_i, c_i, and ϕ_i are wavenumbers, phase speeds, and phase factors, respectively. Please keep in mind that there is only a single spatial variable x; X and Y are propagating arguments with no direct physical interpretation. Reference 2 describes how to pass from X-Y space to x-t space in more detail.

The Fourier series for the theta function is

$$\theta = \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}{}_{[\text{integers}]} \exp(-\{T_{11}n_1{}^2 + 2T_{12}n_1n_2 + T_{22}n_2{}^2\}) \times \exp[2\pi i(n_1X + n_2Y)], \tag{2.6}$$

where the sums are taken over all integers including 0. The constants T_{11}, T_{12}, and T_{22} are the elements of the so-called "theta matrix." For simplicity, the notation differs slightly from the usual in that a factor of $i\pi$ has been absorbed into theta matrix elements as explained in Ref. 2.

The complementary Gaussian series is

$$\theta = \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}{}_{[\text{half-integers}]} \exp\left(-\left\{\left(\frac{R_{11}}{2}\right)(X+n_1)^2 + R_{12}(X+n_1)(Y+n_2) + \left(\frac{R_{22}}{2}\right)(Y+n_2)^2\right\}\right), \tag{2.7}$$

where the sums now range over the half-integers, i.e., $\pm\frac{1}{2}$, $\pm\frac{3}{2}$, $\pm\frac{5}{2}$, etc., instead of over the integers as in the Fourier series (2.6). The 2×2 symmetric matrix whose elements are the constants R_{11}, R_{12}, and R_{22} is loosely called the "inverse theta matrix" since it is proportional to the inverse of the matrix formed from the T_{ij}'s that appear in the Fourier series.[2]

The next step is to simply rewrite the theta series as functions of the physical variables (x,t) using the definitions of (X,Y) given above, substitute the series into the HKdV equation, and collect terms. The resulting sums depend upon how the bilinear operators of the HKdV equation affect a typical pair of terms in the series, so it is useful to define such a pair of terms as

$$F \equiv \exp(-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2) \times \exp(-[\delta_1 n_1 + \delta_2 n_2 + \delta_p]x - [\epsilon_1 n_1 + \epsilon_2 n_2 + \epsilon_p]t), \tag{2.8}$$

$$G \equiv \exp(-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2) \times \exp(-[\delta_1 n_1' + \delta_2 n_2' + \delta_p]x - [\epsilon_1 n_1' + \epsilon_2 n_2' + \epsilon_p]t), \tag{2.9}$$

where the Greek parameters $(\alpha, \beta, \gamma$, and so on) are linear functions of the theta matrix elements, wavenumbers, and phase speeds that will be given explicitly in Sec. V. The forms (2.8) and (2.9) are the natural definitions for the Gaussian series, but they can be specialized to the Fourier series, too, by setting the second-degree exponents α, β, and γ equal to zero and replacing the pseudowavenumbers δ_1 and δ_2 and pseudofrequencies ϵ_1 and ϵ_2 by $2\pi i k_1$, and so on. Thus, it is sufficient to consider the Gaussian case alone. Note that the second-degree exponents are the same for all terms in a given series; only the linear exponents are different and only through the replacement of (n_1,n_2) by (n_1',n_2').

Defining a function zeta via

$$(D_x{}^4 + D_xD_t - 2A)(F\cdot G) \equiv \zeta(n_1 - n_1', n_2 - n_2'; \alpha,\beta,\gamma,\delta_1,\delta_2,\epsilon_1,\epsilon_2,A)FG \tag{2.10}$$

one can use the theorems proved in Sec. VI of Ref. 1 to show that for the HKdV equation,

$$\zeta(m,n,\alpha,\beta,\gamma,\delta_1,\delta_2,\epsilon_1,\epsilon_2,A) = (m\delta_1 + n\delta_2)^4 + (m\delta_1 + n\delta_2)[(\epsilon_1 - 12\alpha\delta_1)m + (\epsilon_2 - 12\alpha\delta_2)n] + 12\alpha^2 - 2\beta - 2A. \tag{2.11}$$

The residual function $\rho(x,t)$, which is defined by

$$\rho(x,t) \equiv (D_x{}^4 + D_xD_t - 2A)(\theta\cdot\theta) \tag{2.12}$$

becomes, after substituting either of the theta series (2.6) or (2.7) into (2.12), invoking (2.10), and collecting terms

$$\rho = \exp(-\alpha x^2 - 2\beta xt - \gamma t^2 - 2\ \delta_p x - 2\epsilon_p t - 2\Phi_p) \text{ (times)}$$
$$\times \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \rho_{jk} \exp(-[\delta_1 j + \delta_2 k]x$$
$$\text{[integers]}$$
$$-[\epsilon_1 j + \epsilon_2 k]t - [\Phi_1 j + \Phi_2 k]), \tag{2.13}$$

where both sums are taken over the integers for either the Gaussian or Fourier theta series.

A theta series is a solution of the HKdV equation if and only if $\rho(x,t) \equiv 0$. Since the terms in (2.13) are linearly independent, this in turn implies that

$$\rho_{jk} = 0, \quad j = 0,1,2,..., \quad k = 0,1,2,..., \tag{2.14}$$

for all integers j and k. Thus, by substituting an infinite series into the differential equation, one reduces it to an infinite set of coupled algebraic equations which determine the coefficients of the series.

Since the theta function is uniquely determined by a finite number of parameters (the three theta matrix elements plus the wavenumbers and phase speeds), one seems to have a problem: infinitely more equations than unknowns! Independently, Boyd[1] and Nakamura[4] resolved this apparent paradox by proving that only four of the infinite set of "residual equations" (2.14) are independent: the rest are proportional to the chosen four, which may be conveniently taken as $j = 0,1$ and $k = 0,1$.

The goal of this paper is simply to solve these four algebraic equations via perturbation theory.

III. PERTURBATION THEORY FOR GENERAL SYSTEMS OF ALGEBRAIC EQUATIONS

Suppose one is given a system of N algebraic equations in N unknowns which depend upon a small parameter,

$$F_i(x_1,x_2,...,x_N;\epsilon) = 0, \quad i = 1,2,...,N, \tag{3.1}$$

such that a solution $\mathbf{x}^{(0)} = (x_1{}^{(0)},x_2{}^{(0)},...,x_N{}^{(0)})$ is known for $\epsilon = 0$. A regular perturbation expansion in ϵ can be calculated through the following three steps: (i) expand each F_i as a power series in the $N+1$ small variables $([x_1 - x_1{}^{(0)}], [x_2 - x_2{}^{(0)}],...,[x_N - x_N{}^{(0)}];\epsilon)$; (ii) expand each of the unknowns $(x_1,x_2,...,x_N)$ as a power series in ϵ, substitute in the series obtained in the first step and collect powers in ϵ; (iii) order-by-order in ϵ, solve the equations that result from demanding that the expansions for each F_i obtained in the second step are identically equal to 0.

If one writes

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \mathbf{x}^{(0)} + \epsilon \mathbf{x}^{(1)} + \epsilon^2 \mathbf{x}^{(2)} + \cdots, \tag{3.2}$$

then

$$\mathbf{J}\mathbf{x}^{(N)} = \mathbf{F}^{(N)}(\mathbf{x}^{(0)},\mathbf{x}^{(1)},...,\mathbf{x}^{(N-1)};\epsilon), \tag{3.3}$$

where $\mathbf{F}^{(N)}$ is the column vector whose elements are those terms at $O(\epsilon^N)$ in the power series of F_i which depend only on the lower-order coefficients in the series in the unknowns ($\mathbf{x}^{(0)}$ and so on) which have already been calculated where $\mathbf{J}$ denotes the usual $N \times N$ Jacobian matrix of the functions F_i, evaluated at $\epsilon = 0$. The elements J_{ij} of J are

$$J_{ij} \equiv \frac{\partial F_i}{\partial x_j}(\mathbf{x}^{(0)};0). \tag{3.4}$$

If the equations F_i are polynomials[6] in the unknowns, then calculating the right-hand side of (3.3) is merely a matter of rearranging power series. Since it is always possible to solve a system of linear equations like (3.3)— a matrix equation is one of the types that is always solvable—the only delicate part of the business is calculating the lowest-order solution $\mathbf{x}^{(0)}$, since this may involve solving nonlinear equations. Fortunately, in the important cases for the double cnoidal wave, solving the zeroth-order perturbation equations is easy.

Although routine, the power series rearrangements and algebra needed to repeatedly compute and then solve (3.3) quickly becomes laborious as the perturbation order increases. The perturbation series were therefore calculated using the algebraic manipulation language REDUCE 2, which can add, multiply, differentiate, and collect terms in polynomials of several variables in symbolic form without requiring the substitution of numerical values as in FORTRAN. Regular (as opposed to singular) perturbation theory is ideally suited to REDUCE 2 and vice versa: The program to compute each of the perturbation series given in the next two sections had fewer than 50 executable statements (!) and cost of the final runs was less than $10.00.

A special advantage of employing an algebraic manipulation language is that different soliton equations in the same class as the KdV differ only in the function ζ defined in Sec. II. Therefore, perturbation series for the Boussinesq water wave equation and several others can be obtained by rerunning the program after modifying only a couple of statements.

A second advantage is that the computer can substitute the perturbation series back into the original nonlinear equations to verify that the solution has indeed been calculated correctly.

The same algorithm, and very nearly the same computer program, can also be applied to higher polycnoidal waves. The major difference is that for the triple cnoidal wave, for example, which is the generalization of three sine waves (small amplitude) and three solitons (large amplitude), $N = 7$ instead of 4, and the series are more complicated because of the greater number of parameters.

IV. SMALL AMPLITUDE (FOURIER SERIES) PERTURBATION THEORY

As noted earlier, the residual equations that determine the theta function Fourier series are a special case of the corresponding more general expressions for the Gaussian series theory. Making the replacements $\delta_i \to 2\pi i k_i$, $\epsilon_i \to -2\pi_i k_i c_i$ and $\alpha = \beta = \gamma = 0$, one finds

$$\rho_{jk} = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} q_1^{n_1{}^2 + (n_1-j)^2} q_2^{n_2{}^2 + (n_2-k)^2}$$
$$\text{[integers]}$$
$$\times e^{-2T_{12}[n_1 n_2 + (n_1 - j)(n_2 - k)]}$$
$$\times \zeta(2n_1 - j, 2n_2 - k; k_1, k_2, c_1, c_2, A), \tag{4.1}$$

where

$$\zeta(m,n;k_1,k_2,c_1,c_2,A) \equiv 16\pi^4(k_1m + k_2n)^4 + 4\pi^2(k_1m + k_2n) \times (k_1c_1m + k_2c_2n) - 2A, \tag{4.2}$$

$$q_i \equiv e^{-T_{ii}} \text{ ["nomes"]}. \tag{4.3}$$

The four equations to be solved are

$$\rho_{jk} = 0, \quad j = 0,1; \quad k = 0,1. \tag{4.4}$$

The input parameters are the wavenumbers (k_1,k_2) and either the diagonal theta matrix elements (T_{11}, T_{22}) or the nomes (q_1,q_2). The perturbation series given here are power series in q_1^2 and q_2^2. The nomes (or equivalently, T_{11} and T_{22}) determine the amplitudes of the two sine waves that form the lowest-order approximation, while k_1 and k_2 are the wavenumbers of these two waves.

The four unknowns whose column vector is $\mathbf{x}$ are the two phase speeds (c_1,c_2), the constant of integration A for the HKdV equation (which has no physical significance), and the exponential of the off-diagonal theta matrix element

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ A \\ e^{-2T_{12}} \end{pmatrix}. \tag{4.5}$$

Novikov[7] has stressed that double cnoidal waves may be almost periodic in space as well as in time. In other words, the ratio of k_1/k_2 is mathematically arbitrary and may even be irrational. Therefore, it is useful to first give the general solution for symbolic k_1 and k_2 [to $O(q_1^4, q_2^4)$] and then the special $k_1 = 1$, $k_2 = 2$ solution to $O(q_1^8, q_2^8)$.

$$c_1 = -k_1^2(1 - 24q_1^2 - 72q_1^4) + 384q_2^4k_2^6/([k_1^2 - k_2^2]^2), \tag{4.6}$$

$$c_2 = -k_2^2(1 - 24q_2^2 - 72q_2^4) + 384q_1^4k_1^6/([k_1^2 - k_2^2]^2), \tag{4.7}$$

$$A = 0 + 12(k_1^4q_1^2 + k_2^4q_2^2) + 72(k_1^4q_1^4 + k_2^4q_2^4) - 768q_1^2q_2^2k_1^4k_2^4/([k_1^2 - k_2^2]^2), \tag{4.8}$$

$$\begin{aligned} e^{-T_{12}} = \{ & [(k_1 - k_2)/(k_1 + k_2)]^2\} \\ & + 32k_1k_2\{(k_1^2q_1^2 + k_2^2q_2^2)/[(k_1 + k_2)^4]\} \\ & + 32k_1k_2\{k_1^2q_1^4(-9k_1^4 + 16k_1^3k_2 \\ & - 18k_1^2k_2^2 + 3k_2^4) + k_2^2q_2^4(3k_1^4 - 18k_1^2k_2^2 \\ & + 16k_1k_2^3 - 9k_2^4) + 16k_1^2k_2^2q_1^2q_2^2(-3k_1^2 \\ & + 2k_1k_2 - 3k_2^2)\}/[(k_1 + k_2)^4(k_1^2 - k_2^2)^2]. \end{aligned} \tag{4.9}$$

The special solution for a wave and its second harmonic $(k_1 = 1, k_2 = 2)$ is

$$\begin{aligned} c_1 = 4\pi^2\{ & -1 + 24q_1^2 + 72q_1^4 + (\tfrac{8\,192}{3})q_2^4 \\ & + 96q_1^6 - 32(63\,488q_2^6 + 69\,632q_1^2q_2^4)/27 \\ & + 168q_1^8 + 114\,688(3\,064q_1^4q_2^4 + 16\,704q_1^2q_2^6 \\ & + 6\,025q_2^8)/243\}, \end{aligned} \tag{4.10}$$

$$\begin{aligned} c_2 = 16\pi^2\{ & -1 + 24q_2^2 + 72q_2^4 + (\tfrac{32}{3})q_1^4 \\ & + 96q_2^6 - 256(q_1^6 + 184q_1^4q_2^2)/27 \\ & + 168q_2^8 + 64(535q_1^8 + 104\,448q_1^6q_2^2 \\ & + 658\,528q_1^4q_2^4)/243\}, \end{aligned} \tag{4.11}$$

$$\begin{aligned} A = 0 + 16\pi^4\{ & 12q_1^2 + 192q_2^2 \\ & + 72q_1^4 + 1\,152q_2^4 - (\tfrac{4\,096}{3})q_1^2q_2^2 \\ & + 144q_1^6 + 2304q_2^6 \\ & + 204\,800(q_1^4q_2^2 + 13q_1^2q_2^4)/27 \\ & + 336q_1^8 + 5376q_2^8 - 16\,384(495q_1^6q_2^2 \\ & + 38\,461q_1^4q_2^4 + 89\,475q_1^2q_2^6)/243\}, \end{aligned} \tag{4.12}$$

$$\begin{aligned} e^{-T_{12}} = & (\tfrac{1}{9}) + (\tfrac{64}{81})q_1^2 + (\tfrac{256}{81})q_2^2 \\ & - (\tfrac{64}{729})[q_1^4 + 340q_2^4 + 704q_1^2q_2^2] \\ & + 256[19\,972q_2^6 + 93\,168q_1^2q_2^4 \\ & + 19\,728q_1^4q_2^2 + 37q_1^6]/6\,561 \\ & - (\tfrac{30\,994\,432}{243})q_1^4q_2^4 \\ & - (\tfrac{64}{19\,683})(2\,887q_1^8 + 8\,816\,428q_2^8) \\ & - (\tfrac{16\,384}{59\,049})(29\,029q_1^6q_2^2 + 736\,867q_1^2q_2^6). \end{aligned} \tag{4.13}$$

Several features of these expansions deserve comment. First, the numerical coefficients are rather large for high order, suggesting that the range of accuracy in the $q_1 - q_2$ plane is too small to be useful. To show that this is not true, contours of constant error for the perturbation series of various orders are given in Figs. 1 and 2 in Sec. VII.

Second, the expansions proceed in powers of $q_1^2(= \exp[-2T_{11}])$ and $q_2^2(= \exp[-2T_{22}])$ rather than q_1 and q_2 themselves even though the series for the theta functions have coefficients that are power series in the unsquared variables q_1 and q_2. This obviously improves the accuracy and usefulness of the perturbation series.

Third, the perturbation series are sparse not merely because all the odd powers vanish but because some of the expected even powers are missing, too. The series for c_1, for example, has zero coefficients for q_2^2, $q_1^2q_2^2$, $q_1^4q_2^2$, and $q_1^6q_2^2$, i.e., one missing term at each order so that the series through $O(q_1^8,q_2^8)$ contains only eleven terms. Similar sparsity exists for the other quantities.

Fourth, although the series for A has been listed for completeness (one cannot solve for the other unknowns without simultaneously obtaining A, too) A is only the constant of integration for the Hirota–Korteweg–deVries equation and has no direct physical significance. It is never necessary to evaluate A to compute the double cnoidal wave solutions of the KdV equation itself.

Fifth and most important, although $u(x,t)$, the KdV solution, is defined in terms of an infinite series, it is never necessary to explicitly tabulate the coefficients of the series. Instead the three dependent parameters T_{12}, c_1, and c_2 together with the four free parameters T_{11}, T_{22}, k_1, and k_2 completely determine all the coefficients of the theta Fourier series (2.6), which in turn determines $u(x,t)$ via (2.1)

Ordinary Stokes' expansions, obtained via the method of multiple scales as in Appendix B of Ref. 1, can be calculated for almost any wave equation, but each Fourier component—and their numbers grow as the square of the perturbation order—must be calculated through a separate expansion as complicated as that for the phase speeds c_1 and

c_2. To need only three perturbation series instead of many is thus a great simplification.

V. LARGE AMPLITUDE (GAUSSIAN SERIES) PERTURBATION THEORY

For large amplitude, the double cnoidal wave problem reduces to solving the four simultaneous nonlinear residual equations $\rho_{jk} = 0$, where the residuals are given by the Gaussian series

$$\rho_{jk} = \sum_{\substack{n_1=-\infty \\ \text{[half-integers]}}}^{\infty} \sum_{\substack{n_2=-\infty \\ \text{[half-integers]}}}^{\infty} q_1'^{\,n_1^2 + (n_1 - j)^2} q_2'^{\,n_2^2 + (n_2 - k)^2} \times \exp(-R_{12}\{n_1 n_2 + (n_1 - j)(n_2 - k)\}) \times \zeta(2n_1 - j, 2n_2 - k; \alpha, \beta, \gamma, \delta_1, \delta_2, \epsilon_1, \epsilon_2, A), \quad j = 0,1 \quad k = 0,1, \tag{5.1}$$

where

$$q_i' \equiv e^{-R_{ii}/2} \quad [\text{“complementary nomes”}]. \tag{5.2}$$

Equation (5.1) is very similar in form to its Fourier series counterpart, (4.1), but there are some noteworthy differences. As indicated in the square brackets under the summation symbols, the sums do not run over the integers but rather over the "half-integers" $\pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}, \ldots$. The "nomes" q_1 and q_2 are replaced by the "complementary nomes" q_1' and q_2' which are defined in terms of the elements of the inverse theta matrix. A factor 2 is present in (5.2) which is missing from (4.2) and the factor of 2 multiplying T_{12} in (4.1) has no counterpart in the coefficient of R_{12} in (5.1).

The major difference is in the form of ζ which is

$$\zeta(m,n;\alpha,\beta,\gamma,\delta_1,\delta_2,\epsilon_1,\epsilon_2,A) = (\delta_1 m + \delta_2 n)^4 + (\delta_1 m + \delta_2 n)[(\epsilon_1 - 12\alpha\delta_1)m + (\epsilon_2 - 12\alpha\delta_2)n] + 12\alpha^2 - 2\beta - 2A, \tag{5.3}$$

where the parameters are related to the wavenumbers and phase speeds via

$$\alpha = R_{11}k_1^2 + 2R_{12}k_1k_2 + R_{22}k_2^2, \tag{5.4}$$

$$\beta = -R_{11}k_1^2c_1 - R_{12}k_1k_2(c_1 + c_2) - R_{22}k_2^2c_2, \tag{5.5}$$

$$\gamma = R_{11}k_1^2c_1^2 + 2R_{12}k_1k_2c_1c_2 + R_{22}k_2^2c_2^2, \tag{5.6}$$

$$\delta_1 = R_{11}k_1 + R_{12}k_2, \tag{5.7}$$

$$\delta_2 = R_{12}k_1 + R_{22}k_2, \tag{5.8}$$

$$\epsilon_1 = -R_{11}k_1c_1 - R_{12}k_2c_2, \tag{5.9}$$

$$\epsilon_2 = -R_{12}k_1c_1 - R_{22}k_2c_2. \tag{5.10}$$

The parameters δ_1 and δ_2 may be named "pseudowavenumbers" because, as shown in Appendix B, they give the widths of the two solitary waves in the large amplitude, near-soliton regime in the same way that the wavenumbers k_1 and k_2 give the widths of the two sine waves in the small amplitude regime. Similarly, ϵ_1 and ϵ_2 may be labeled "pseudofrequencies" in the sense that $(-\epsilon_1/\delta_1)$ and $(-\epsilon_2/\delta_2)$ are the phase speeds of the two solitary waves for large wave amplitude.

The major complication posed by (5.4)–(5.10) is that the parameters denoted by Greek letters are functions of R_{12}, which is one of the unknowns. Thus, it is not possible to evaluate any of these parameters *a priori*; instead, one must solve for them as part of the task of solving the residual equations, which would seem to leave us facing an algebraic problem of ghastly complexity.

Fortunately, the situation is not quite as bad as it looks. The function ζ is independent of γ, which is automatically eliminated from $u(x,t)$ by taking the second logarithmic derivative with respect to x. Thus, although γ is needed to graph the theta function, it is quite irrelevant both to solving the residual equations and to evaluating the solution of the Korteweg–de Vries equation, so γ will be ignored in the rest of the discussion.

The parameter β appears in ζ only as the sum $\beta + A$. Thus, if β is artificially set equal to 0 to reduce the number of unknowns, the solution of the residual equation will be unchanged except for A, but A has no physical significance. Therefore, the calculations presented will be done with $\beta = 0$; after R_{12} and the other unknowns have been determined, one can then evaluate β and add the result to the computed A to obtain a final solution which is completely consistent with the original equations (5.1)–(5.10).

The parameter α has a slightly more complex role. One can easily show from (5.3), (5.9), (5.10), (2.1), and (2.9) that the results of a calculation in which α is artificially set equal to 0 differ from those in which α is retained via

$$A\,[\text{with }\alpha] = A(\alpha = 0) + 6\alpha^2, \tag{5.11}$$

$$c_1[\text{with }\alpha] = c_1(\alpha = 0) - 12\alpha, \tag{5.12a}$$

$$c_2[\text{with }\alpha] = c_2(\alpha = 0) - 12\alpha, \tag{5.12b}$$

$$u(x,t)[\text{with }\alpha] = -12\alpha + u(x + 12\alpha t, t), \quad (\alpha = 0). \tag{5.13}$$

Since A has no physical significance, the important role of α is to add a constant to $u(x,t)$ while simultaneously increasing all the phase speeds of the "angle" variables by the same constant. As noted in Ref. 2, the theta function solution of the Korteweg–de Vries equation is that solution which has $\langle u \rangle = 0$, where $\langle\ \rangle$ denotes an average over the periodicity interval. In the near-soliton regime, this is awkward because it implies that the solitons asymptote to $u = -12\alpha$ instead of to $u = 0$, which is the usual asymptotic solution as $|x| \to \infty$ in the spatially unbounded problem. Setting $\alpha = 0$ merely causes the solitons to asymptote to 0.[8] Thus, the parameter α is no real trouble either.

Difficulties with the "pseudofrequencies" ϵ_1 and ϵ_2 can be avoided by simply taking them as unknowns in the residual equations. After ϵ_1, ϵ_2, and R_{12} have been obtained by solving the rest of the problems, c_1 and c_2 can be obtained by solving (5.9) and (5.10) as a pair of linear equations in two unknowns. Alternatively, one could use (5.9) and (5.10) directly in ζ to replace ϵ_1 and ϵ_2 wherever they appeared by expressions in c_1, c_2, and R_{12}, which are the usual unknowns of the residual equations, but this makes the ρ_{jk} much more complicated, so it is far less work to consider ϵ_1 and ϵ_2 as the unknowns and then compute c_1 and c_2 at the end.

Unfortunately, there is little one can do with the two remaining parameters, the "pseudowavenumbers" δ_1 and δ_2. The simplest set of algebraic nonlinear equations one can solve simultaneously is the set of six equations in the unknowns (A, R_{12}, ϵ_1, ϵ_2, δ_1, δ_2): the four residual equations

$\rho_{jk}=0$ plus the pair of equations which define δ_1 and δ_2, (5.7) and (5.8). However, the rather special form of these equations—(5.7) and (5.8) involve only three of the six unknowns—means that it is not necessary to solve all six equations simultaneously. Instead, one can pretend that δ_1 and δ_2 are independent free parameters and solve the four residual equations via perturbation theory exactly as for the Fourier series case in the preceding section. Adding two new parameters to the four that already exist (R_{11}, R_{22}, k_1, k_2) would seem to greatly complicate the chore of solving $\rho_{jk}=0$, but it actually does not because k_1 and k_2 do not appear explicitly in the Gaussian form of ζ (5.3). Instead, δ_1 and δ_2 appear in place of the wavenumbers in the analogous terms of ζ. Thus, this device of pretending δ_1 and δ_2 are independent parameters leads to solutions of the coupled set $\rho_{jk}=0$ which are neither more nor less complicated than the analogous solution in the Fourier case for general k_1 and k_2. Just as the Fourier solution for general k_1 and k_2 (as opposed to $k_2=2k_1$) was taken only up to and including second order $[O(q_1{}^4, q_2{}^4)]$, so also the Gaussian solution of the residual equations will only be carried to second order also.

The residual equations (implicit dispersion relation) are defined by infinite series; to calculate the solution to a given order, it is sufficient to truncate the series of ρ_{jk} after this same order. The truncated residual series in the Fourier case was omitted from the previous section because it is given (to lowest order) in Appendix A, but it is useful to give the series to second order for at least one of the two cases so that the reader can see more clearly what must be solved. The series have been simplified by exploiting the general symmetry relation (true in Fourier case also)

$$\zeta(m,n)=\zeta(-m,-n) \quad \text{for all } m,n, \tag{5.14}$$

and by dividing out common factors, which is why equal signs have been replaced by proportionality symbols. In addition

$$\chi \equiv e^{-R_{12}} \tag{5.15}$$

has been used to replace appearances of R_{12} so as to make the series rational in all parameters and unknowns.

$$\rho_{00}\propto\zeta(1,1)\chi+\zeta(1,-1)+q_1'^4[\zeta(3,1)\chi^2+\zeta(3,-1)/\chi]$$
$$+\,q_2'^4[\zeta(1,3)\chi^2+\zeta(1,-3)/\chi], \tag{5.16}$$

$$\rho_{10}\propto\zeta(0,1)+q_1'^2[\zeta(2,1)\chi+\zeta(2,-1)/\chi]+q_2'^4\zeta(0,3), \tag{5.17}$$

$$\rho_{01}\propto\zeta(1,0)+q_2'^2[\zeta(1,2)\chi+\zeta(1,-2)/\chi]+q_1'^4\zeta(3,0), \tag{5.18}$$

$$\rho_{11}\propto\zeta(0,0)+2[q_1'^2\zeta(2,0)+q_2'^2\zeta(0,2)]$$
$$+\,2q_1'^2q_2'^2[\zeta(2,2)\chi^2+\zeta(2,-2)/\chi^2]. \tag{5.19}$$

These rather innocent-looking expressions, (5.16)–(5.19), become exceedingly messy when $\zeta(m,n)$ is evaluated according to (5.3), so they were solved perturbatively using the algebraic manipulation language REDUCE 2 to perform (and check!) the algebra. Note that these series, like their solutions and the Fourier perturbation series given in Sec. IV, are "sparse": many expected terms in the series are identically equal to 0. Equation (5.16), for example, contains no first-order terms at all, and only five of the possible 12 second-order terms appear in the set.

The corresponding solutions are given below. Note that the parameters α and β have been inserted in the proper places so that the results are fully consistent with (5.3)

$$\epsilon_1=\delta_1(12\alpha-\delta_1{}^2)+24q_1'^2\delta_1{}^3$$
$$+\,24\delta_1[3q_1'^4(\delta_1{}^2\delta_2{}^4-2\delta_1{}^4\delta_2{}^2+\delta_1{}^6)$$
$$+\,16q_2'^4\delta_2{}^6]/[(\delta_2{}^2-\delta_1{}^2)^2], \tag{5.20}$$

$$\epsilon_2=\delta_2(12\alpha-\delta_2{}^2)+24q_2'^2\delta_2{}^3+24\delta_2[16q_1'^4\delta_1{}^6$$
$$+\,3q_2'^4(\delta_2{}^6-2\delta_1{}^2\delta_2{}^4+\delta_1{}^4\delta_2{}^2)]/[(\delta_2{}^2-\delta_1{}^2)^2], \tag{5.21}$$

$$A=6\alpha^2-\beta+12q_1'^2\delta_1{}^4+12q_2'^2\delta_2{}^4$$
$$+\,24[3q_1'^4(\delta_1{}^4\delta_2{}^4-2\delta_1{}^6\delta_2{}^2+\delta_1{}^8)$$
$$-\,32q_1'^2q_2'^2\delta_1{}^4\delta_2{}^4+3q_2'^4(\delta_2{}^8-2\delta_1{}^2\delta_2{}^6$$
$$+\,\delta_1{}^4\delta_2{}^4)]/[(\delta_2{}^2-\delta_1{}^2)^2], \tag{5.22}$$

$$e^{-R_{12}}=(\delta_1-\delta_2)^2/(\delta_1+\delta_2)^2$$
$$+\,32\delta_1\delta_2(q_1'^2\delta_1{}^2+q_2'^2\delta_2{}^2)/(\delta_1+\delta_2)^4$$
$$+\,32\delta_1\delta_2[(q_1'^4(3\delta_1{}^2\delta_2{}^4-18\delta_1{}^4\delta_2{}^2$$
$$+\,16\delta_1{}^5\delta_2-9\delta_1{}^6)-q_1'^2q_2'^2(48\delta_1{}^2\delta_2{}^4$$
$$-\,32\delta_1{}^3\delta_2{}^3+48\delta_1{}^4\delta_2{}^2)+q_2'^4(-9\delta_2{}^6$$
$$+\,16\delta_1\delta_2{}^5-18\delta_1{}^2\delta_2{}^4+3\delta_1{}^4\delta_2{}^2)]/$$
$$[(\delta_1+\delta_2)^6(\delta_1-\delta_2)^2]. \tag{5.23}$$

As explained before, δ_1 and δ_2 are not really independent parameters but rather are determined by R_{11} and R_{22} (or equivalently, by q_1' and q_2') through (5.7) and (5.8). It is therefore necessary to solve the triplet system of (5.23) plus (5.7) and (5.8) for the three unknowns (δ_1, δ_2, R_{12}) to obtain a completely consistent solution. However, in the large amplitude regime, δ_1 and δ_2 have a physical interpretation as giving the widths and speeds of the two solitons,

$$u(x,t)\doteq 3\delta_1{}^2\,\mathrm{sech}^2[(\delta_1/2)(x-\delta_1{}^2t+\phi_1)]$$
$$+\,3\delta_2{}^2\,\mathrm{sech}^2[(\delta_1/2)(x-\delta_2{}^2t+\phi_2)], \tag{5.24}$$

where ϕ_1 and ϕ_2 are phase constants at those times when the two solitons are well separated. In the large amplitude regime, $R_{12}\ll R_{11}, R_{22}$ so that one has approximately

$$R_{ii}=\delta_i/k_i, \quad i=1,2. \tag{5.25}$$

Under these circumstances, it may be preferable to take the pseudowavenumbers δ_1 and δ_2 as independent parameters, estimate R_{11} and R_{22} and therefore, q_1' and q_2' via (5.25) and (5.2), and then use (5.20) to (5.23) directly to estimate the importance of the corrections due to spatial periodicity to the lowest-order solution, which is just the spatially unbounded double soliton.[9]

Unfortunately, this crude estimation is all that can be done directly because the lowest-order problem for (5.7), (5.8), and (5.23), which is

$$e^{-R_{12}}=(\delta_1-\delta_2)^2/(\delta_1+\delta_2)^2, \tag{5.26}$$

$$\delta_1=k_1R_{11}+k_2R_{12}, \tag{5.7 bis}$$

$$\delta_2 = k_1 R_{12} + k_2 R_{22}, \tag{5.8 bis}$$

has no closed form solution. It is easy to solve this set numerically, however.

To proceed to higher order, it is convenient to replace δ_1 and δ_2 by

$$\bar{S} \equiv \delta_1 + \delta_2, \tag{5.27}$$

$$\bar{D} \equiv \delta_1 - \delta_2. \tag{5.28}$$

There are two motives for this trick. One is the δ_1 and δ_2 often appear in (5.23) as their sum or difference rather than alone. A second reason is that for the special case of $k_1 = k_2$ (the principal branch or mode of the double cnoidal wave as explained in Ref. 2) the difference variable D is given exactly by the lowest-order solution. In terms of the new variables, the problem becomes (5.23) plus

$$\bar{S} = k_1 R_{11} + k_2 R_{22} + (k_1 + k_2) R_{12}, \tag{5.29}$$

$$\bar{D} = k_1 R_{11} - k_2 R_{22} + (k_2 - k_1) R_{12}. \tag{5.30}$$

The general solution to first order for arbitrary k_1 and k_2 is

$$\bar{S} = S - \Psi(k_1 + k_2), \tag{5.31}$$

$$\bar{D} = D - \Psi(k_2 - k_1), \tag{5.32}$$

$$e^{-R_{12}} = \chi(1 + \Psi), \tag{5.33}$$

where S, D, and χ are the solutions of the lowest-order set

$$S = k_1 R_{11} + k_2 R_{22} - (k_1 + k_2)\ln(\chi), \tag{5.34}$$

$$D = k_1 R_{11} - k_2 R_{22} + (k_1 - k_2)\ln(\chi), \tag{5.35}$$

$$\chi = D^2/S^2, \tag{5.36}$$

[which is equivalent to (5.26) plus (5.7) and (5.8)] and where

$$\Psi = 2(S+D)(S-D)[q_1'^2(S+D)^2 + q_2'^2(S-D)^2]/ \{\chi S^4 - SD^4[k_1(S+D) - k_2(S-D)]/2\}. \tag{5.37}$$

The second-order solution for the special case $k_1 = k_2 = 1$ is

$$\bar{D} = D, \tag{5.38a}$$

$$\bar{S} = S - 2\Psi + \{-32D^4\Psi(q_1'^2 + q_2'^2) + 48D^3S\Psi(q_2'^2 - q_1'^2) + 16DS^3\Psi(q_1'^2 - q_2'^2) + \Psi^2(24D^2S - \chi S^5) + 2S^5 v\}/(S^2[4D^2 - \chi S^3]), \tag{5.38b}$$

$$e^{-R_{12}} = \chi(1 + \Psi) + \{16D^4\Psi(q_1'^2 + q_2'^2) + 24D^3S\Psi(q_1'^2 - q_2'^2) + 8DS^3\Psi(q_2'^2 - q_1'^2) + 2D^2S\Psi^2(S - 6) - S^5 v\}/(S^2[4D^2 - \chi S^3]), \tag{5.39}$$

where as before S, D, and χ are the lowest-order solutions for $(\delta_1 + \delta_2)$, $(\delta_1 - \delta_2)$, and $\exp(-R_{12})$, Ψ is given by (5.37) and

$$v \equiv \{-q_1'^4[S^4 + 2S^3D + 10SD^3 + 5D^4](S+D)^3(S-D) - 8q_1'^2 q_2'^2[S^2 + 2D^2](S+D)^3(S-D)^3 - q_2'^4[S^4 - 2S^3D - 10SD^3 + 5D^4](S-D)^3(S+D)\}/(D^2S^6). \tag{5.40}$$

It goes almost without saying that the Gaussian perturbation theory is more cumbersome than its Fourier counterpart at the same; particularly annoying is the necessity of solving the lowest-order set (5.26) plus (5.7)–(5.8) numerically, even though a simple Newton's iteration initialized with $R_{12} = 0$ always seems to work. However, the Gaussian series is that oddity: a perturbation series that converges more and more rapidly as the wave amplitudes becomes larger, so it is an essential component of any complete treatment of polycnoidal waves.

VI. PERTURBED SINGLE-SOLITON REGIMES

The perturbation series derived in the previous two sections were based on the implicit assumption that both diagonal theta matrix elements T_{11} and T_{22} are either very large (Fourier series) or very small (Gaussian series). When one diagonal theta matrix element is very large and the other is very small, however, neither the Fourier series nor the Gaussian series for the theta function converges rapidly as is obvious from inspecting the form of these series.

In Ref. 1, it is shown through numerical examples that these regimes correspond to a single solitary wave slightly perturbed by a small amplitude sine wave, so these parametric neighborhoods are much less interesting than those in which the waves have amplitudes of the same order of magnitude and one or the other of the series given in the previous sections is rapidly convergent. When ($k_1 = 1$ and $k_2 = 2$)

$$T_{11} \ll \pi, \quad T_{22} \gg \pi, \tag{6.1}$$

the solitary wave is of unit period with a height and width determined solely by the magnitude of T_{11}, and the perturbation is of wavenumber 2, i.e., periodic with a period of $\frac{1}{2}$, with a small amplitude roughly equal to 4 $\exp(-T_{22})$. After the application of a modular transformation[3] to $k_1 = k_2 = 1$, this same regime is found to be characterized by either

$$R_{11} \ll 2\pi, \quad R_{22} \gg \pi, \tag{6.2}$$

or equivalently (since the wavenumbers after the modular transformation are identical) by (6.2) with the direction of the inequalities reversed.

The other perturbed-one-soliton regime occurs when

$$T_{11} \gg \pi, \quad T_{22} \ll \pi. \tag{6.3}$$

The large amplitude component is now of wavenumber 2, so that tall, narrow solitary wave is repeated with half unit period while the small amplitude perturbation is a subharmonic of period one. When the modular transformation is applied to convert to a representation with equal wavenumbers, $k_1 = k_2 = 1$, one finds that the equivalent neighborhood in terms of the inverse theta matrix elements lies around the diagonal in the $R_{11} - R_{12}$ plane,

$$R_{11} \simeq R_{12}. \tag{6.4}$$

The reason for this somewhat surprising result is that the wavenumbers are equal in the $R_{11} - R_{12}$ plane and therefore the roles of the two diagonal inverse theta matrix elements are physically interchangeable and the phase speeds, etc., must be symmetric functions of R_{11} and R_{22}. This implies that the whole of the $T_{11} - T_{22}$ plane must map into the wedge-shaped half of the $R_{11} - R_{22}$ plane which lies below the diagonal (6.4).

The Poisson summation method which was used to generate the Gaussian series from the theta Fourier series can be applied selectively to just one of the sum variables, either n_1 or n_2, in the infinite series that define the residual function ρ_{ij} (4.1). This is not the most efficient way to proceed because it causes "theta matrix-halving" as explained on p. 384 of Ref. 1, but it shows that in principle, Poisson summation can be applied to generate rapidly converging residual function series (and perturbation series derived from them) in any region of parameter space for polycnoidal waves of any genus N.

A procedure that gives more rapidly converging series is to apply partial Poisson summation directly to the multidimensional theta function and substitute the result into the Hirota–Korteweg–de Vries equation. Shirfuji[10] actually applied this idea to the double cnoidal wave of the Toda lattice problem in 1976, but the independent derivation of the residual equations by Nakamura and Boyd lay in the future, and such results as he obtained came directly from the governing equations of the Toda lattice, and not from Hirota's transformed Toda equation. The theta function can be written as

$$\theta = \theta_4(X) + e^{-T_{22}}\{\theta_4[X - (i/\pi)T_{12}]e^{2\pi i Y} + \theta_4[X + (i/\pi)T_{12}]e^{-2\pi i Y}\}, \tag{6.5}$$

when $T_{11} \ll T_{22}$, where $\theta_4(X)$ is the usual one-dimensional theta function. Representing θ_4 by its Gaussian series representation [note that (6.5) contains the lowest terms of the Fourier series in the other angle variable Y with higher terms eliminated because of the extreme smallness of $\exp(-T_{22})$] one can substitute (6.5) into the Hirota–Korteweg–de Vries equation and then use the calculus of Hirota operators developed in Ref. 1 to obtain infinite series for the residual equations.

Unfortunately, the resulting zeroth-order approximation is a quartic equation in $\exp(-T_{12})$ and cannot be solved in closed form, unlike its counterpart for the pure Fourier series representation given in Sec. IV. It follows that one is forced to resort to numerical methods even to compute the zeroth-order solution, so this kind of special treatment for the perturbed-one-soliton regimes is not very useful. In the first place, the double cnoidal wave is much more interesting when it is truly a double soliton or a double sine wave than when it is merely a perturbed ordinary cnoidal wave. In the second place, numerical solution of the "pure" Fourier or Gaussian residual equations (4.1) and (4.2) is quick and efficient unless the difference between the magnitudes of the diagonal theta matrix elements is very large, but in that case, the perturbation of the single soliton is very, very small, and therefore uninteresting.

For this reason, no further details will be given about partial Poisson summation of Shirafuji's approximation. For most purposes, the perturbation series derived in the preceding two sections are quite adequate. For the perturbed solitary wave discussed in this section, alternative perturbative methods, like that of Grimshaw,[11] might be more physical and easier than trying to work through the residual equations.

Wahlquist[12] and Kuznetsov and Mikhailov[13] report solutions obtained via Backlund transformations and the inverse scattering transform. Zagrodzinski and Jaworski[14] apply ideas similar to Shirafuji's to obtain what they dub "mixed" solutions, i.e., solitons perturbed by sine waves, for the sine-Gordon equation for general N.

VII. ACCURACY OF THE PERTURBATION SERIES

In Ref. 1, it is shown that the complementary perturbation series, one which gives the first few terms of the Fourier series of the theta function and the other which gives the Gaussian series, were very accurate for the ordinary cnoidal wave provided that each series was used in the proper regime (small wave amplitude for the Fourier series and large amplitude for the Gaussian series). In the worst case, i.e., that intermediate wave amplitude for which both series converge equally well or poorly, both gave the phase speed to within a relative error of 4.7% to zeroth order, and to within 0.027% to first order, where "zeroth" order refers to the phase speed of a linear sine wave in the Fourier case and a solitary wave on an infinite spatial interval in the Gaussian case.

For the double cnoidal wave, the overlap between the two complementary perturbation series is not quite so dramatic, but it is still good. Figures 1 and 2 compare regions in which the zeroth-order and first-order perturbation series give errors which are less than 10%. The error criterion is to take the largest of the three errors for c_1, c_2, and either T_{12} or R_{12} as appropriate using the modified relative error criterion

$$\text{Error} = (c_1^{\text{pert}} - c_1^{\text{exact}})/c, \tag{7.1}$$

where

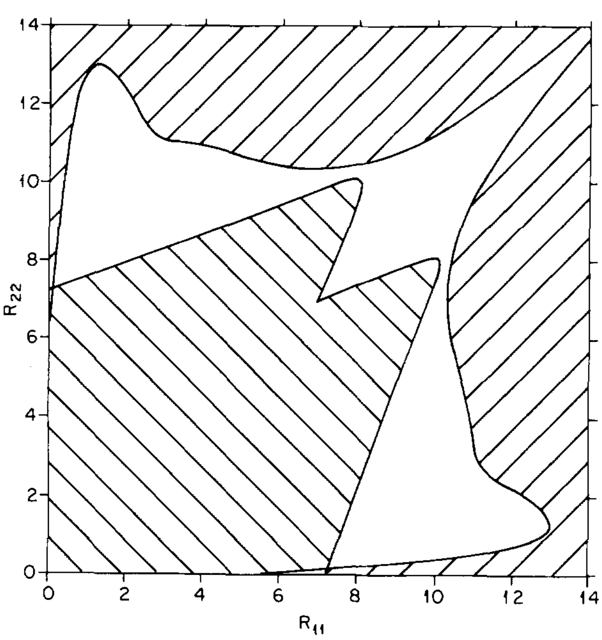

FIG. 1. The lines slanting from top right to bottom left denote that region in the $R_{11} - R_{22}$ plane where the error in all three of the quantities c_1, c_2 and R_{12}, which suffice to determine the theta function and the corresponding solution of the Korteweg–de Vries equation, is less than 10% for zeroth-order Gaussian perturbation theory, which is the double solitary wave approximation. The modified relative error is defined by Eqs. (7.1) and (7.2). The lines slanting from top left to bottom right are the 10% error region for the zeroth-order Fourier perturbation theory, which is equivalent to approximating the double cnoidal wave as the sum of two linear sine waves. The blank area is "no-man's land" where neither approximation is accurate within 10%.

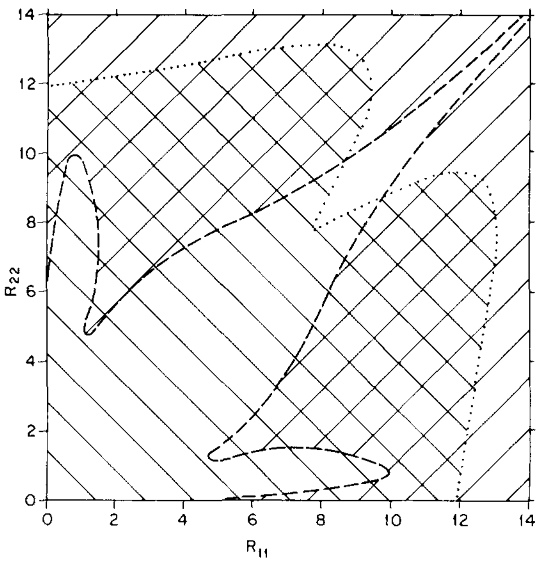

FIG. 2. Same as Fig. 1 except for the first-order Fourier and Gaussian approximations, which incorporate the first correction to the double solitary wave and double sine wave. The first-order theories overlap very well.

$$c = \text{larger of} \begin{cases} c_1 \\ c_1^{\text{linear sine wave}} = -39.4, \end{cases} \tag{7.2}$$

and similarly for c_2 and the off-diagonal theta matrix element. The reason for the modification, i.e., the replacement of the exact variable by its value in the linear limit, is that both phase speeds vanish along certain curves in the two-dimensional parameter space spanned by R_{11} and R_{22}, which implies infinite unmodified relative errors in the neighborhood of these curves even though the absolute errors may be very small.

Figures 1 and 2 show the principal branch of the double cnoidal wave with $k_1 = k_2 = 1$. Because the wavenumbers are identical, the graph is symmetric about the diagonal $R_{11} = R_{22}$. The neighborhood of this diagonal corresponds to a perturbed ordinary cnoidal wave of unit period. As explained in the preceding section, neither perturbation series can be expected to work well in these neighborhoods because both that derived in Sec. IV and the Gaussian series of Sec. V implicitly assume that the amplitudes of both waves are either very small or very large. However, the graphs show that the near-diagonal and near-axis regions where the Gaussian and Fourier perturbation series fail are quite narrow—almost invisible on the scale of the graph. This is a strong pragmatic justification for omitting a detailed treatment of the mixed Gaussian–Fourier perturbation series which, as noted in Sec. VI, can be calculated, but which would hardly ever be of any practical value.

Even outside these narrow perturbed-single-soliton areas, the zeroth-order perturbation curves do not quite overlap; there is a small region of moderate R_{11} and R_{22} where neither approximation gives all three dependent variables to within 10%. However including the first-order corrections to the sum of the two noninteracting linear sine waves and to the double solitary wave reduces the error to less than 10% everywhere except very close to the diagonals and the axes.

The physical implication is clear: The double cnoidal wave of the Korteweg–de Vries equation can always be considered both qualitatively and quantitatively to be either (i) the sum of two noninteracting sine waves; (ii) a pair of solitary waves of different heights, repeated with unit period over all x; or (iii) a single soliton plus a weak sinusoidal perturbation. When one wants to obtain numerical values for the double cnoidal wave, the perturbation series derived earlier will usually be adequate. If high accuracy is needed, it is straightforward to solve the residual equations numerically using the peturbation series to initialize the iteration.

The one serious complication is that the Fourier and Gaussian perturbation series involve different parameters—the Fourier expansion uses T_{11} and T_{22} while Gaussian employs R_{11} and R_{22}—and it is not possible to transform from one pair of parameters to the other unless one knows either T_{12} and R_{12}. In practical terms, this means that if one wants to make a contour plot of the phase speed c_1 as a function of T_{11} and T_{22} including such small values of these diagonal theta matrix elements that one passes into the double-soliton regime, one must use an iteration instead of a direct evaluation. One must guess T_{12} (in the large amplitude, double-soliton regime, one cannot calculate it from the Fourier perturbation series), perform a modular transformation as in Ref. 3 to obtain the three inverse theta matrix elements, apply the Gaussian perturbation series, determine the difference between the R_{12} obtained by the modular transformation and that calculated by the Gaussian perturbation series, transform back to T_{11}-T_{22} space, and guess a new value for T_{12} and so on. The fact that the off-diagonal theta matrix elements are unknowns rather than independent parameters is a considerable practical difficulty.

Fortunately, it is one that arises only when one is attempting to simultaneously explore the dynamics of double cnoidal waves in both the large and small amplitude regimes. If one is content instead to examine the double cnoidal wave strictly as the sum of two solitary waves, then one can stick to the inverse theta matrix elements R_{11} and R_{22} as parameters and use the Gaussian perturbation series alone. If one wants to investigate polycnoidal waves as a sum of quasilinear waves, the Fourier perturbation series is more than adequate.

VIII. SUMMARY

Following the plan outlined in Ref. 1, the problem of the double cnoidal wave for the Korteweg–de Vries equation has been reduced to four algebraic equations in four unknowns. Because the four functions of this set are defined only via infinite series, it is extremely advantageous to express these four residual functions in two quite different ways: one obtained by using the ordinary Fourier series of the theta function and then applying the theorems of Ref. 1, and a second representation derived via the alternative Gaussian series. These representations are mutually complementary in the sense that the Fourier representation, obtained independently by Nakamura,[4] is very efficient for small amplitude double cnoidal waves while the Gaussian representation, obtained here for the first time, is highly effective for large amplitude, i.e., when the double cnoidal wave is approximately equal to two solitary waves of unequal heights repeated periodically over all space.

It is also straightforward to solve the residual equations using perturbation theory. Comparisons with numerical solutions show that even the zeroth-order perturbation series have good overlap while the two first-order series cover almost all of parameter space with errors of 10% or less. By using the algebraic manipulation language REDUCE 2, it is trivial to extend the series to fairly high order for the principal branch of the double cnoidal wave (fourth order for the Fourier case and second order for the Gaussian series) so as to cover all the physically interesting regimes in parameter space.

The methods employed here, which explicitly use the properties of the Riemann theta function, are only applicable to partial differential equations which are exactly integrable by the periodic analog of the inverse scattering method, which is known as the "Hill's spectrum" procedure. Within this class, however, the ideas developed here extend very readily to other equations. For the Boussinesq water wave equation, for example, the calculations presented here can be repeated merely by altering the residual equations (and the appropriate line in the REDUCE 2 computer program) to use a new function $\zeta(m,n)$, where $\zeta(m,n)$ is defined (for the Korteweg–de Vries equation) by (2.11).

The Gaussian perturbation series is remarkable in that it converges most rapidly when the wave amplitude is large rather than small, which makes it well suited for exploring the effects of spatial periodicity on solitary waves. The Fourier perturbation series is useful, too, because its form is simpler and easier to evaluate than the Gaussian series and it remains accurate even for moderately large waves. Both series share the common property that it is not necessary to write down separate series for each of an infinite number of Fourier coefficients or the like: one need only have series for three parameters, and these determine the whole infinite series for the theta function, and thus for the double cnoidal wave itself.

Future work to calculate perturbation series for other partial differential equations integrable via the "Hill's spectrum" method is now in progress. It is hoped that the results will be useful whenever equations of soliton type are applied with spatially periodic boundary conditions, or wherever there is a high density of solitons so that soliton–soliton overlap is important.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants OCE8108530 and OCE8305648.

APPENDIX A: PERTURBATION THEORY IN AN UNPHYSICAL REPRESENTATION

The companion paper (Ref. 3) has shown that via the "special" modular transformation, a given theta function can be expressed in a denumerably infinite number of ways. Each of these allowed representations involves theta functions of two "angle" variables, $X = k_1(x - c_1 t)$ and $Y = k_2(x - c_2 t)$, but in general the phase speeds c_1 and c_2 have no actual physical interpretation unless the representation is that unique one defined to be the "physical" representation in Ref. 3. Fortunately, the perturbation series given earlier automatically calculate in this "physical" representation so that c_1 and c_2 are the actual speeds at which individual peaks of the polycnoidal wave are moving.

Nonetheless, it is still of interest to see how perturbation theory can cope with the problem of calculating in an "unphysical" representation because this both provides an additional demonstration of the existence of an infinite number of alternative representations of the theta function and also illuminates the assumptions and details of the perturbation method. For simplicity, attention will be limited to the lowest-order Fourier case for a polycnoidal wave consisting of a sine wave of unit period and its second harmonic (plus very small high harmonics created by their interaction which will not be explicitly calculated).

Assuming that

$$T_{11}, T_{22} \gg 1 \tag{A1}$$

the four residual equations are, to lowest order with common factors omitted,

$$\rho_{00} = \zeta(0,0) \tag{A2}$$

$$\rho_{10} = \zeta(1,0), \tag{A3}$$

$$\rho_{01} = \zeta(0,1) + e^{-2T_{11} - 2T_{12}} \zeta(2,1), \tag{A4}$$

$$\rho_{11} = e^{-2T_{12}} \zeta(1,1) + \zeta(1,-1), \tag{A5}$$

where the zeta function for the Korteweg–de Vries equation is defined by

$$\zeta(m,n) \equiv 16\pi^4 (k_1 m + k_2 n)^4 + 4\pi^2 (k_1 m + k_2 n)(k_1 c_1 m + k_2 c_2 n) - 2A. \tag{A6}$$

[The zeta function satisfies the general symmetry relation $\zeta(m,n) = \zeta(-m,-n)$ as evident in (A6), and this has been used to simplify (A3) through (A5).]

In the physical representation for which $k_1 = 1$ and $k_2 = 2$, (A2) and (A3) may be solved to give

$$c_1 = -4\pi^2 + O(e^{-2T_{11}}, e^{-2T_{22}}), \tag{A7}$$

$$A = 0 + O(e^{-2T_{11}}, e^{-2T_{22}}), \tag{A8}$$

If one assumes

$$|T_{12}| \ll T_{11}, T_{22} \tag{A9}$$

as done implicitly in earlier sections, then the second term in ρ_{01} must be neglected to give

$$c_2 = -16\pi^2 + O(e^{-T_{11}}, e^{-2T_{22}}). \tag{A10}$$

Equation (A9) is the key assumption that ensures that we calculate in the physical representation. The phase speed c_2 is indeed that of a second harmonic of the linearized Korteweg–de Vries equation. The residual $\rho_{11} = 0$ gives

$$T_{12} = \log(3) = 1.0986. \tag{A11}$$

It is, however, equally possible to calculate in the unphysical representation $k_1 = k_2 = 1$. As stressed in the author's companion paper on the modular transformation,[3] the linear dispersion relation gives a unique phase speed for each wavenumber, so it is quite absurd to suppose that the two waves of different phase speeds which are the dominant terms in the Fourier series of a small amplitude double cnoidal wave can both have identical wavenumbers. (If wavenumbers and phase speeds are the same, then the two waves are identical and we have an ordinary cnoidal wave which depends on but a single "angle" variable.) Nonetheless, it is still possible to represent the solution using a theta function with the unphysical wavenumber $k_2 = 1$ if (A9) is replaced by

$$T_{12} = -T_{11} + \Delta. \tag{A12}$$

The first two residuals are unaffected (to lowest order!) by the change in k_2 and by (A12), so the phase speed c_1 and constant of integration are still given by (A7) and (A8). The invariance of c_1 is in fact true to all orders in perturbation theory because a modular transformation that alters k_2 and Y does not affect X and c_1 at all as may be seen in Table I of Ref. 3. The invariance of A is also exact because the special modular transformation leaves the theta function unchanged, which means that after the angle variables have been converted to x and t, the theta function has the same dependence on space and time as before. The theta function must therefore satisfy the Hirota–Korteweg–de Vries equation with the same constant of integration A.

The other two residual equations, however, are quite drastically changed. When the wavenumbers are identical and $A = 0$,

$$\zeta(1,1) = 0 \tag{A13}$$

so that

$$\rho_{11} = e^{-2T_{12}}\zeta(1,1). \tag{A14}$$

The only way that $\rho_{11} = 0$ is if either (i) $T_{12} = \infty$, which is impossible since the theta series would diverge or (ii)

$$\zeta(1,1) = 0 \tag{A15}$$

which demands

$$c_2 = -28\pi^2. \tag{A16}$$

This is not the phase of any linear wave of the Korteweg–de Vries equation with an integer wavenumber.

Because of the large magnitude ot T_{12}, it is no longer legitimate to neglect the second term in ρ_{01}, which becomes [using (A12)]

$$\rho_{01} = \zeta(0,1) + e^{-2\Delta}\zeta(2,1), \tag{A17}$$

which gives

$$T_{12} = -T_{11} + 1.0986. \tag{A18}$$

These alterations in c_2 and T_{12} [from the values given in (A10) and (A11)] are exactly as listed in Table I of the companion paper by Boyd for a transformation by the modular generator $\mathbf{A}_2^{-1}$. Equation (A16) is the limit of the numerically calculated values of c_2 as given in Table II of the same paper, while (A10) gives the limit of what is called c_2^{mod} in the same table. Thus, there is a gratifying consistency between the numerical solutions of the residual equation, the perturbation theory, and the special modular transformation.

The lowest three terms of the theta function itself can be written in either representation,

$$\theta\begin{bmatrix}0\\0\end{bmatrix} = 1 + e^{-T_{11}}\cos(2\pi X) + e^{-T_{22}}\cos(2\pi Y) + e^{-T_{11}-T_{22}-2T_{12}}\cos(2\pi[X+Y]). \tag{A19}$$

In the "physical" representation [1,2], i.e., $k_1 = 1$ but $k_2 = 2$, the first two cosine terms are dominant in the limit of T_{11}, $T_{22} \gg 1$. In the "unphysical" representation [1,1], i.e., $k_1 = k_2 = 1$, (A18) implies

$$e^{-T_{11}-T_{22}-2T_{12}} = e^{-T_{22}}e^{T_{11}-2.2972} \gg e^{-T_{22}} \tag{A20}$$

for large T_{11}, so that the $\cos(2\pi Y)$ term is exponentially small in comparison to the "mixed" terms $\cos(2\pi[X+Y])$. This is as it should be because $\cos(2\pi[X+Y])$ is of wavenumber 2 in x and is in fact identical with the term which is written $\cos(2\pi Y)$ in the other representation.

APPENDIX B: THE RELATIONSHIP BETWEEN THE "TETRA-GAUSSIAN" DOUBLE SOLITON AND HIROTA'S DOUBLE SOLITON

As noted in the main body of this paper and the author's previous work,[1] the sum of the four Gaussians with peaks at the corners of the unit square, named the "tetra-Gaussian" and labeled by an upper case Greek Θ, can be given two interpretations. First, it is the lowest-order approximation to the full theta function series $\theta(X,Y)$. Second, it is an exact solution of the Hirota–Korteweg–de Vries equation for the spatially unbounded problem, representing two solitary waves of unequal height. This second interpretation is important because it justifies interpreting the double cnoidal wave as a double soliton when the wave amplitudes are large enough so that the tetra-Gaussian is an accurate approximation.

It therefore, is useful to explain how the tetra-Gaussian, which seemingly is very different, is physically equivalent to Hirota's own solution to the HKdV equation, which is

$$H(x,t) = 1 + e^{-\bar\delta_1 x + \bar\delta_1^3 t - \phi_1} + e^{-\bar\delta_2 x + \bar\delta_2^3 t - \phi_2} + \left(\frac{\bar\delta_1 - \bar\delta_2}{\bar\delta_1 + \delta_2}\right)^2 \times \exp(-(\bar\delta_1 + \bar\delta_2)x + (\bar\delta_1^3 + \bar\delta_2^3)t - \phi_1 - \phi_2). \tag{B1}$$

The tetra-Gaussian in contrast is

$$\Theta(x,t) = \sum_{n_1 = \pm 1/2}\sum_{n_2 = \pm 1/2} \exp(-\{(R_{11}/2)(X+n_1)^2 + R_{12}(X+n_1)(Y+n_2) + (R_{22}/2)(Y+n_2)^2\}) \tag{B2}$$

or written in terms of x and t

$$\Theta(x,t) = \exp(-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2 - (R_{11}/8) - (R_{22}/8)) \times \sum_{n_1 = \pm 1/2}\sum_{n_2 = \pm 1/2} \exp(-\{(\delta_1 n_1 + \delta_2 n_2 + \delta_p)x + (\epsilon_1 n_1 + \epsilon_2 n_2 + \epsilon_p)t\}) \times \exp(-\{R_{12}n_1 n_2 + \Phi_1 n_1 + \Phi_2 n_2 + \Phi_p\}) \tag{B3}$$

where the Greek letters are related to the theta matrix elements and k_1, k_2, etc., via (2.8) through (2.19). Thus, Hirota's solution is an unsymmetrical sum of four exponentials of linear arguments and is spatially unbounded, whereas the tetra-Gaussian is a symmetrical sum of four exponentials of quadratic arguments. Because of (i) the form of the function $\zeta(i,j)$ which appears in the residual equation (2.22), (2.23), and (2.26); and (ii) the second logarithmic derivative transformation, these differences are almost entirely cosmetic if one matches the pseudowavenumbers, i.e.,

$$\bar{\delta}_1 = \delta_1, \quad \bar{\delta}_2 = \delta_2. \tag{B4}$$

It was stressed in the author's previous work[1] that the reason one can prove that all but four of the residual equations $\rho_{jk} = 0$ are redundant is because $\zeta(i,j)$ depends only on differences in the exponentials of a pair of terms in the theta function whose interaction in the bilinear HKdV equation is described by $\zeta(i,j)$. This implies that if $H(x,t)$ is a solution of the HKdV equation, then $\exp[\nu x + \omega t + \Xi]H(x,t)$ is also a solution for arbitrary constants ν, ω, and Ξ. This theorem was widely used by Hirota himself a decade ago to manipulate his solutions into convenient form. Here, recalling that the explicit, exact solution of the residual equations for the tetra-Gaussian (which is also the lowest-order approximate solution for the full θ-series) implies that $\epsilon_1 = \delta_1^3 - 12\alpha\delta_1$, $\epsilon_2 = \delta_2^3 - 12\alpha\delta_2$, and $\exp[-R_{12}] = (\delta_1 - \delta_2)^2/(\delta_1 + \delta_2)^2$, one can verify through routine multiplication that $\exp[\nu x + \omega t + \Xi]H(x,t)$ matches $\Theta(x,t)$ except for the dependence of the latter on α, β, and γ provided that

$$\nu = \tfrac{1}{2}(\delta_1 + \delta_2), \tag{B5}$$

$$\omega = \tfrac{1}{2}(\delta_1^3 + \delta_2^3), \tag{B6}$$

$$\Xi = -R_{11}/8 - R_{22}/8 + \Phi_p - \phi_1/2 - \phi_2/2 - R_{12}/4, \tag{B7}$$

and that one adjusts the phase factors ϕ_1 and ϕ_2 in the angle variables X and Y, which determine Φ_1 and Φ_2 in (B2) via $\Phi_1 = R_{11}\phi_1 + R_{12}\phi_2$ and $\Phi_2 = R_{12}\phi_1 + R_{22}\phi_2$, so that

$$\Phi_1 = \phi_1 + R_{12}/2, \tag{B8}$$

$$\Phi_2 = \phi_2 + R_{12}/2. \tag{B9}$$

The two phase factors in X and Y are neither more nor less than what is needed to match the two phase factors in $H(x,t)$ and vice versa.

Since $\zeta(i,j)$ is independent of γ, it follows that $\exp[-(\gamma/2)t^2]H(x,t)$ is a solution if $H(x,t)$ is. The function $\zeta(i,j)$ does depend on β, but only in the combination of $\beta - A$. Thus, if $H(x,t)$ solves the HKdV equation with $A = 0$, then $\exp[-\beta xt]H(x,t)$ is a solution of the HKdV equation with the new constant of integration $A = \beta$. This same reasoning explains why Fourier series numerical integration of the HKdV equation instead of the KdV equation, which is otherwise tempting because the Fourier series of the theta function converges much more rapidly than that for the mermorphic function which is the corresponding solution of the KdV equation, will not work unless A is known in advance: There is only a single value of A which the HKdV equation has a periodic solution. Arbitrary choices of A will yield solutions that are the products of a periodic function with $\exp[-(\text{const})xt]$. This is strictly a numerical difficulty, however; neither β nor A has any effect on the solution of the Korteweg–de Vries equation because the $\exp[-\beta xt]$ factor is automatically eliminated when the second logarithmic derivative is taken.

The factor of $\exp[-(\alpha/2)x^2]$ does alter $u(x,t)$, but only by the addition of a constant and simultaneously a shift in all the phase speeds by the same constant. Stated formally, one can easily show from the form of $\zeta(i,j)$ (or from the theorem given in Sec. VII of the author's previous work) that if $H(x,t)$ is a solution of the HKdV equation with $u(x,t) = 12(\ln H)_{xx}$ as the corresponding KdV solution, then $\exp[-(\alpha/2)x^2]H(x + 12\alpha t, t)$ is also an HKdV solution with the new constant of integration $A' = A - 6\alpha^2$ with

$$v(x,t) = -12\alpha + u(x + 12\alpha t, t) \tag{B10}$$

as the corresponding solution of the Korteweg–de Vries equation. Thus, aside from the α dependence in (B10), the tetra-Gaussian is physically equivalent to Hirota's double-soliton HKdV solution, even though their mathematical form is rather different.

APPENDIX C: THE "TETRA-GAUSSIAN" AND THE GEOMETRY OF THE X-Y PLANE

For large R_{11}, R_{22}, one can accurately approximate the full theta function series by a tetra-Gaussian and deduce a number of simple facts that have been exploited here and in Ref. 2. First, note that, using Θ to denote the tetra-Gaussian as in Appendix B,

$$\ln\Theta \equiv \ln\Bigg[\sum_{n_1=-1/2}^{1/2}\sum_{n_2=-1/2}^{1/2}\exp\bigg(-\bigg\{\Big(\frac{R_{11}}{2}\Big)(X+n_1)^2 + R_{12}(X+n_1)(Y+n_2) + \Big(\frac{R_{22}}{2}\Big)(Y+n_2)^2\bigg\}\bigg)\Bigg], \tag{C1}$$

which by extracting the common factor is

$$= -(R_{11}/2)X^2 + R_{12}XY + (R_{22}/2)Y^2 + \ln\Bigg[\sum_{n_1=-1/2}^{1/2}\sum_{n=-1/2}^{1/2}\exp(-R_{11}Xn_1 - R_{22}Yn_2 - R_{12}Xn_2 - R_{12}Yn_1)\Bigg]. \tag{C2}$$

When we take the second derivative of $\ln\Theta$, the quadratic terms in X^2, XY, and Y^2 are converted to a constant (-12α), so the shape of the double soliton is determined entirely by the remaining logarithm in (C2), i.e.,

$$L \equiv \ln\Bigg[\sum_{n_1=-1/2}^{1/2}\sum_{n_2=-1/2}^{1/2}\exp(-R_{11}Xn_1 - R_{22}Yn_2 - R_{12}Xn_2 - R_{12}Yn_1)\Bigg]. \tag{C3}$$

As done in Appendix B, one can then show that the sum of the four exponentials with linear arguments in (C2) is equivalent to Hirota's sum of four linear exponentials that generate the double solitary wave in the spatially unbounded problem.

Here, a different strategy will be adopted. When R_{11} and R_{22} are very large, the "tetra-Gaussian" has four narrow peaks at each of four corners of the unit square $X = \pm\frac{1}{2}$, $Y = \pm\frac{1}{2}$. Over most of the square, Θ is dominated by a single term. The logarithm of a single exponential of linear argument in (C2) can be evaluated explicitly to give a result linear in X and Y, which is then eliminated by taking two derivatives. Thus, solitons occur only where at least two peaks of the tetra-Gaussian are of comparable magnitude.

One such region is the neighborhood of the positive Y axis where the important peaks are $n_1 = \pm\frac{1}{2}$, $n_2 = -\frac{1}{2}$ and

$$L \doteq \ln[2\cosh(R_{11}X/2 + R_{12}Y/2 - R_{12}/4)] + R_{22}Y/2 + R_{12}X/2. \quad \text{(C4)}$$

The valley in the graph of Θ [which corresponds to a ridge of the function $U(X,Y)$ graphed in Figs. 7, 8, and 10 of Ref. 2] occurs along the line where the argument of the hyperbolic cosine is 0, i.e.,

$$R_{12}Y = -R_{11}X + R_{12}/2. \quad \text{(C5)}$$

Repeating the argument along the negative Y axis gives (C5) again except for a sign change for the Y-intercept, $R_{12}/2$. Thus, one finds, as quoted in Sec. VI of Ref. 2, that the slopes of the soliton valleys are $(-R_{11}/R_{12})$ and by similar reasoning, $(-R_{12}/R_{22})$.

Using the definitions $\delta_1 = R_{11}k_1 + R_{12}k_2$, $X \equiv k_1(x - c_1t)$, etc., as in (5.7) and (2.4) above, one can write

$$L = \ln(\cosh[(\delta_1/2)X + (\epsilon_1/2)t - R_{12}/4]) + [*], \quad \text{(C6)}$$

where the $[*]$ denotes terms that will be eliminated by differentiation. Then

$$u(x,t) \equiv 12\frac{d^2}{dx^2}L = 3\delta_1{}^2\,\mathrm{sech}^2[(\delta_1/2)\{x + (\epsilon_1/\delta_1)t - R_{12}/(2\delta_1)\}]. \quad \text{(C7)}$$

Thus, the soliton whose width and amplitude are determined by the diagonal inverse theta matrix element R_{11} corresponds to a trough in the graph of $\theta(X,Y)$ which runs roughly parallel to the Y axis. Repeating the analysis for negative Y gives (C7) again except for a change of sign in R_{12}. Now, the region around the origin is where the two soliton troughs turn and merge. The jump represented by the sign change in R_{12} is therefore the collisional phase shift, which is then

$$\text{phase shift} = R_{12}/\delta_1. \quad \text{(C8)}$$

The tilting of the soliton troughs so that they only approximately parallel the X and Y axes is intimately related to this collisional phase shift. Since the full theta series is dominated within the unit square entirely by the four peaks of the tetra-Gaussian, it follows that the soliton valleys must intercept the edges of the unit square at the same value of X (Y) for the trough paralleling the $Y(X)$ axis, to within $O(\exp[-R_{11}/2], \exp[-R_{22}/2])$, or the theta function will not be periodic. Were it not for the phase shift, the troughs could preserve periodicity simply by running parallel to the axes. As it is, the tilt insures that the troughs, whose equations are

$$R_{12}Y = -R_{11}X \pm R_{12}/2, \quad (+)Y>0,\ (-)Y<0, \quad \text{(C9)}$$

both intersect the edges of the unit square, $Y = \pm\frac{1}{2}$, at $X = 0$. As explained in Sec. VI of Ref. 2, this tilting of trough lines also implies that the phase velocities are not the speeds of the "free" solitary waves, i.e., the rate at which the solitons travel when not enmeshed in a collision; c_1 and c_2 are rather the time-averaged velocities of the peaks of $u(x,t)$.

[1] J. P. Boyd, J. Math. Phys. **23**, 375 (1982).
[2] J. P. Boyd, J. Math. Phys. **25**, 3390 (1984).
[3] J. P. Boyd, J. Math. Phys. **25**, 3415 (1984).
[4] A. Nakamura, J. Phys. Soc. Jpn. **47**, 1701 (1979).
[5] R. Hirota, Phys. Rev. Lett. **27**, 1192 (1971).

[6] Strictly speaking, the residual equations involve negative powers of the exponential of the "off-diagonal" theta matrix element, but these can be easily converted into polynomials through the binomial theorem. Only the higher-order correction to this variable need be expanded, so the expansion is consistent with the perturbation series.

[7] S. P. Novikov, in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, New York, 1980), pp. 325–338.

[8] The freedom to replace α in (5.13) by an arbitrary constant gives polycnoidal wave solutions of the Korteweg–de Vries equation which are not the second logarithmic derivative of a theta function. Since the shapes and the relative locations of the peaks and troughs are unaffected by this single non-theta degree of freedom, however, it is still reasonable to state the theta functions generate all possible polycnoidal solutions of the Korteweg–de Vries equation.

[9] G. B. Whitham, *Nonlinear Waves* (Wiley–Interscience, New York, 1974), pp. 580–585.
[10] T. Shirafuji, Suppl. Prog. Theor. Phys. **59**, 126 (1976).
[11] R. Grimshaw, Proc. R. Soc. London Ser. A **368**, 359, 377 (1979).
[12] H. D. Wahlquist, in *Backlund Transformations*, edited by R. M. Miura (Springer-Verlag, New York, 1976), p. 162.
[13] E. A. Kuznetsov and A. V. Mikhailov, Z. Eksp. Teor. Fiz. **67**, 1717 (1974).
[14] J. Zagrodzinski and M. Jaworski, Z. Phys. B. **49**, 77 (1982).