PURE & APPLIED MATHEMATICS
A WILEY-INTERSCIENCE SERIES OF
TEXTS, MONOGRAPHS & TRACTS

G. L. LAMB, JR.

ELEMENTS OF SOLITON THEORY

ELEMENTS OF SOLITON THEORY

G. L. LAMB, Jr.
University of Arizona

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Lamb, George L, Jr. 1931-
Elements of soliton theory.

(Pure and applied mathematics)
Bibliography: p.
Includes index.
1. Differential equations, Partial—Numerical solutions. 2. Solitons. 3. Scattering (Mathematics) I. Title.

QA377.L33 515.3'53 80-13373
ISBN 0-471-04559-4

Printed in the United States of America

10 9 8 7 6 5 4 3

To

LAUREN, LARRY, JOEY, and JOANNE

PREFACE

This book is intended to be an elementary introduction to the theory of solitons, a topic that has provided a fascinating glimpse into the inner workings of certain nonlinear processes during the past decade. The background assumed of the reader is within that usually accumulated by a senior or beginning graduate student in physics or applied mathematics. Some knowledge of integration in the complex plane is presumed, and prior exposure to eigenvalue problems, preferably in the context of quantum theory, will be found helpful but not essential. Since the subject matter is concerned with the solution of nonlinear partial differential equations, some familiarity with the rudiments of linear partial differential equations is, of course, assumed. The applications presuppose some familiarity with hydrodynamics, electromagnetic theory, and the quantum theory of a two-level atom.

The subject is thus presented at an elementary level and concentrates on the background material and introductory concepts that have played a role in setting the stage for some of the current research trends in the field. Recent reformulations involving modern differential geometry and group theory, as well as the ingenious techniques devised by R. Hirota and the results on lattice solitons pioneered by M. Toda, have not been included.

The exposition is pedagogic rather than historical and the topics chosen for consideration are those that, in the author's opinion, convey the basic ideas of the subject in the simplest and most direct way. The analytical formulations are those that present themselves naturally to a physicist raised in an applied tradition. Such workers usually find the more severe procedures of the pure mathematician to be less rather than more demonstrative.

After an introductory chapter that gives a brief indication of the connection between a nonlinear partial differential equation that exhibits soliton behavior (the Korteweg–deVries equation) and a linear eigenvalue problem (for the Schrödinger equation), the next two chapters provide an elementary account of one-dimensional scattering theory and inverse scattering methods. The Korteweg–deVries equation is then treated by inverse scattering techniques in Chapter 4. Chapter 5 provides a corresponding introduction to the other most common soliton equations. Chapters 6 and 7 present some examples of how soliton equations arise in various physical contexts. Chapter

8 introduces the subject of Bäcklund transformations and finally, in Chapter 9, the recently popular topic of soliton perturbation theory is considered.

The presentation is largely self-contained so that reference to the original literature should be unnecessary. A number of references to additional background material have been included as well as references to expositions that either extend or complement the presentation given here. They are not intended to document either priority or high points in the historical development of the subject. The reader interested in extensive bibliographies may consult the article "The soliton: a new concept in applied science" by A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, *Proc. IEEE* **63**, 1443–1483 (1973) as well as the volume *Solitons* (Springer Topics in Modern Physics Series) edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, Heidelberg, 1980).

Certain facets of soliton theory may be traced back quite directly to research in nineteenth-century mathematics. It has been this writer's experience that reference to the research of our predecessors can be especially rewarding when investigating soliton theory. The writings of A. R. Forsyth have been found to be particularly appropriate in this regard. Some aspects of soliton theory seem to provide fulfillment to the closing paragraph in the sixth volume of Forsyth's *Theory of Differential Equations*, where he writes:

> My desire has been to give a continuous exposition of those portions of the subject which . . . bear some promise of leading into paths of research that will be trodden by investigators in days yet to come.

I wish to record my appreciation to F. A. Otter, Jr., for a comment concerning the occurrence in dislocation theory of what is now known as the sine-Gordon equation. A solution technique used in this work was based upon Bäcklund transformations. The transfer of these results to coherent optical pulse propagation, where the sine-Gordon equation also arises, led to my consideration of what are now known as the solitons of coherent optics.

My thanks are due to F. A. E. Pirani for a very careful reading of many chapters of this work as well as to M. G. Forest and P. R. Schlazer, who have also read many sections. By their assistance several errors and obscurities have been eliminated, and the volume has been rendered less imperfect than it would otherwise have been.

I am also indebted to W. E. Ferguson, Jr., for providing the numerical solutions of the Korteweg–deVries equation that appear in Chapter 4 and to M. O. Scully and F. A. Hopf for the numerical results indicated in Figure 7.6. The computer assistance offered to me by L. A. Appelbaum and R. C. Dillon during the preparation of the other pulse profile figures is also appreciated. Finally, I am grateful to my wife, Joan, for many hours spent in typing.

Any corrections or suggestions for improvements with which my readers may favor me will be greatly appreciated.

G. L. LAMB, JR.

Tuscon, Arizona
July 1980

CONTENTS

CHAPTER 1

Introduction

A remarkable development in our understanding of a certain class of nonlinear partial differential equations known as evolution equations has taken place in the past decade. The key to our present knowledge of these equations is the realization that they possess a special type of elementary solution. These special solutions take the form of localized disturbances, or pulses, that retain their shape even after interaction among themselves, and thus act somewhat like particles. This independence among elementary solutions is a well-known effect in processes governed by linear partial differential equations where a linear superposition principle applies but was quite unexpected when first observed in processes governed by nonlinear partial differential equations. These localized disturbances have come to be known as *solitons*.

Although the partial differential equations that govern the motion of solitons are nonlinear, they are closely related to certain linear ordinary differential equations known as Sturm–Liouville equations. A study of solitons should thus be prefaced by a summary of the relevant topics in Sturm–Liouville theory. These considerations are developed in Chapters 2 and 3.

Before taking up these preliminaries, however, we shall give a cursory introduction to certain essential elements of soliton theory in this initial chapter. First, by asking the right question regarding an ordinary differential equation of Sturm–Liouville type, we shall be led to a consideration of one of the nonlinear partial differential equations that has soliton solutions. The equation that arises is known as the Korteweg–deVries equation (Korteweg and deVries, 1895). It occurs in a number of physical problems, mostly in hydrodynamics. At our present level of understanding, there appears to be no basis for expecting so fundamental a relationship between a Sturm–Liouville equation and the Korteweg–deVries equation. Secondly, to underscore the close relation between solitons and linear ordinary differential equations, a formula that expresses the interaction between two solitons is constructed by adapting a technique devised by Bargmann (1949) for obtaining a certain class of potentials for the Schrödinger equation, which is an example of a Sturm–Liouville equation. In a somewhat loose manner of speaking, one can say the the analytical expressions that describe multisoliton interactions are merely Bargmann potentials. The particle nature of the soliton is evident when this two-soliton solution is examined. Finally, we shall consider how the

Korteweg–deVries equation arises in a simple example of nonlinear dispersive wave propagation.

1.1 A STURM–LIOUVILLE EQUATION

The differential equation

$$\frac{d^2y}{dx^2}+[\lambda-U(x)]y=0, \qquad a\leqslant x\leqslant b \tag{1.1.1}$$

plus boundary conditions imposed at point $x=a$ and b (either or both of which may be at infinity) is of frequent occurrence in applied mathematics. Such an equation is a simple example of a Sturm–Liouville equation (Ince, 1926). Equation 1.1.1 has been most thoroughly studied in the context of quantum theory, where it is known as a Schrödinger equation. This nomenclature sometimes persists even in applications of the equation to classical physics, such as wave propagation in inhomogeneous media.

For a given function $U(x)$, which would be the potential for a problem in quantum theory, imposition of the boundary conditions can lead to only certain specific values of the constant λ (the eigenvalues λ_j) for which the equation will have a nonzero solution [the eigenfunction $y_j(x)$]. The determination of the dependence of the solution y on the parameter λ and the dependence of the eigenvalues λ_j on the boundary conditions is known as a Sturm–Liouville problem.

One of the simplest examples of such an eigenvalue problem is obtained by setting $U(x)=0$ and imposing the boundary conditions $y(a)=y(b)=0$. Solution of the resulting equation $y''+\lambda y=0$ subject to the prescribed boundary conditions shows that the eigenfunctions are $y_j(x)=\sin[(\lambda_j)^{1/2}x]$, $j=1,2,3,\ldots$, with the eigenvalues $\lambda_j=[j\pi/(b-a)]^2$. As the length of the system $b-a$ increases, the λ_j become more closely spaced and in the limit $b-a\to\infty$ we obtain the continuous range of eigenvalues $0<\lambda<\infty$. When the foregoing equation arises in the study of vibrating systems, each eigenfunction $y_j(x)$ represents the shape of a normal mode of the system. An example would be a uniform string that is vibrating in free space and confined between fixed ends located at points $x=a$ and b. Since the eigenvalues λ_j are related to the resonant frequencies of vibration of the system, it is customary to refer to them as a spectrum of eigenvalues. For the case considered here, say that of a homogeneous string, the entire length of the system takes part in the vibration of each normal mode. However, it is possible to construct inhomogeneous systems, the inhomogeneity represented by the function $U(x)$, for which the vibration is confined to only a portion of the system. The vibration is then confined by the inhomogeneity rather than by the boundaries. An example would be a vibrating string that is partly embedded in elastic surroundings. This system will be discussed at length in Chapter 2.

In the subsequent development we shall always be concerned with systems of infinite extent. Hence, any localized solutions will always be due to inhomogeneities. Such localized solutions have been perhaps most thoroughly studied in quantum theory, where they are used to describe the discrete energy levels of atomic systems.

There are relatively few functions $U(x)$ for which the corresponding ordinary differential equation (1.1.1) may be solved in terms of the standard transcendental functions. As an example (which will be considered in detail in Chapter 2) the choice $U(x)=-2\ \mathrm{sech}^2 x$ plus the boundary conditions $y(\pm\infty)=0$ leads to the single eigenvalue $\lambda_1=-1$ with the associated eigenfunction $y_1=\mathrm{sech}\ x$. That is, $y_1=\mathrm{sech}\ x$ is the solution of the equation $y_1''+(-1+2\,\mathrm{sech}^2 x)y_1=0$ that vanishes as $x\to\pm\infty$. In quantum theory, the interpretation of this result is that a particle is confined by a potential well having a shape proportional to $\mathrm{sech}^2 x$ while the single value of λ is proportional to the energy that the particle confined by this well can possess. As a classical interpretation of the same equation, we may consider the channeling of a wave in a medium having the depth-dependent refractive index $n^2=1+2\ \mathrm{sech}^2 x$, where depth is measured from the location of the maximum value of $n(x)$. The inhomogeneity establishes a waveguide in the medium. A wave can be confined to the depth about which the refractive index takes on its maximum value. When the sign of the potential is reversed so that $U(x)=2\ \mathrm{sech}^2 x$, the potential is repulsive and no bound state occurs. Similarly, when the refractive index is given by $n^2=1-2\ \mathrm{sech}^2 x$, waves tend to emanate away from the region of refractive index variation and no channeling effect takes place. This example will be developed further in Chapter 2.

In addition to the discrete negative values λ_j, with their associated localized wave functions $y_j(x)$, equations such as (1.1.1) can also possess a continuous range of solutions for positive values of λ when $b-a$ becomes infinite. In the quantum case the physical interpretation of such solutions is that of the scattering of an incident particle with energy proportional to λ by some obstacle that is characterized by the potential $U(x)$. In the one-dimensional problems being considered here, the presence of the scatterer usually manifests itself in terms of reflection and transmission of the incident wave. [The fact that certain choices for the potential $U(x)$ can result in perfect transmission with no reflected wave will play a central role in later considerations.] Particles with any positive energy may be incident upon the scattering center, of course, and hence we expect the continuous range of positive eigenvalues $0<\lambda<\infty$. In the example of the inhomogeneous medium mentioned above, the situation corresponding to $\lambda>0$ is the reflection and transmission of a wave of arbitrarily high frequency that is incident upon the inhomogeneous layer from the outside. In Chapter 2 we will show that, for the potential function $U(x)=-2\ \mathrm{sech}^2 x$ given above, the scattering solutions are made up of linear combinations of the functions $y_\pm=e^{\pm i\sqrt{\lambda}\,x}(i\sqrt{\lambda}\mp\tanh x)$. Determination of the solution of a Schrödinger equation when the potential function $U(x)$ is specified is frequently referred to as solving a scattering problem.

1.2 THE KORTEWEG–deVRIES EQUATION

If the function $U(x)$ in (1.1.1) should contain a parameter, say α, so that $U = U(x, \alpha)$, then variation of the shape of the potential by variation of α could be expected to lead to some corresponding variation in the eigenvalues λ_j; that is, we would expect the values of the λ_j to depend upon α. It is perhaps not too unnatural to ask whether or not there are potential functions $U(x, \alpha)$ for which the λ_j remain unchanged as the parameter α is varied. One rather trivial example suggests itself immediately. Replacement of any $U(x)$ by $U(x+\alpha)$ merely translates the potential or refractive index inhomogeneity along the x axis. This merely changes the location of the confined particle or depth of the sound channel and has no effect upon the bound-state energy or the frequency of the confined wave. Such a variation of α thus has no effect upon the eigenvalues λ_j. For comparison with future results, it should be noted that functions $U(x+\alpha)$ satisfy the linear partial differential equation $U_\alpha - U_x = 0$. We shall find that there are other more interesting possibilities that lead to *nonlinear* partial differential equations. In particular, functions $U(x, \alpha)$ that satisfy the nonlinear partial differential equation

$$U_\alpha + UU_x + U_{xxx} = 0 \tag{1.2.1}$$

will also be shown to leave the eigenvalues invariant. When the parameter α is interpreted as time (the associated Sturm–Liouville equation is, of course, not the time-dependent Schrödinger equation), then (1.2.1) is the Korteweg–deVries equation.

Finding solutions to the Korteweg–deVries equation can thus be related to the determination of parameter-dependent potentials in a Sturm–Liouville equation, and vice versa. The scattering problem mentioned in Section 1.1 was concerned with the determination of a wave function y when the potential U was specified. In the present situation we are concerned with determining the potential when certain information about the wave function is specified (in a manner that will be considered in detail in Chapter 4). Determination of a potential from information about the wave function is appropriately referred to as an *inverse* scattering problem.

By using a method devised by P. Lax (1968), we may see quite easily that the Korteweg–deVries equation is one of an infinite number of equations that govern the variation in the potential of a Schrödinger equation in such a way that the eigenvalues remain constant. To see this, it is convenient to write the Schrödinger or Sturm–Liouville equation in the form

$$Ly = \lambda y \tag{1.2.2}$$

where $L = D^2 - u(x,t)$ and $D = d/dx$. A time derivative of this equation yields

$$(Ly)_t = Ly_t + L_t y = \lambda_t y + \lambda y_t \tag{1.2.3}$$

Since $(Ly)_t = y_{xxt} - uy_t - u_t y = Ly_t - u_t y$, we see that $L_t = -u_t$. We are interested in imposing a time variation on u and hence y, such that $\lambda_t = 0$. Let us consider the possibility that the time dependence of y may be expressed in the form $y_t = By$ where B is some linear differential operator (not necessarily unique) that must be determined. The spatial variation of y is, of course, given by (1.2.2). Equation 1.2.3 may now be written

$$(-u_t + [L,B])y = \lambda_t y \tag{1.2.4}$$

where $[L,B] \equiv LB - BL$. We note that λ will be a constant so that $\lambda_t = 0$ provided that B is chosen to satisfy the equation $-u_t + [L,B] = 0$. In general, this would be an operator equation. However as we shall presently see, certain restrictions on the form of B can yield an expression for $[L,B]$ that is devoid of differential operators and contains merely u and its spatial derivatives. In such cases we shall have constructed a partial differential equation for $u(x,t)$ which, when satisfied, will imply $\lambda_t = 0$; that is, the eigenvalues remain constant in time.

As a first example of a differential operator B that can lead to constant eigenvalues, let us consider $B_1 = aD$, where a is initially allowed to be a function of u and its spatial derivatives. Then, from the definition of L,

$$\begin{aligned}[L,B_1]y &= (LB_1 - B_1 L)y \\ &= (D^2 - u)(ay_x) - aD(y_{xx} - u_y) \\ &= 2a_x D^2 y + a_{xx} Dy + au_x y \end{aligned} \tag{1.2.5}$$

If the coefficients of $D^2 y$ and Dy in this last expression vanish, that is, if a is a constant, then $[L,B_1] = au_x$ and (1.2.4) becomes

$$(u_t - au_x)y = -\lambda_t y \tag{1.2.6}$$

Therefore, λ_t will be zero and λ thus constant in time provided that u satisfies the partial differential equation $u_t - au_x = 0$. Since the solution of this equation is any function of $x + at$, we see that any potential of the form $u(x+at)$ will leave λ unchanged in time. This somewhat uninteresting example in which the parameter at merely translates the potential along the x axis at a velocity $-a$ has already been alluded to in the preceding discussion.

To obtain a more interesting example we might try $B_2 = aD^2 + fD + g$, where f and g are, in general, functions of u and its spatial derivatives and a is again a constant, as in the previous example. However, a simple calculation shows that no extension of the previous result is obtained. We are merely led to the same linear partial differential equation for u.

If we proceed a step further and consider $B_3 = aD^3 + fD + g$, we find that

$$[L, B_3]y = (2f_x + 3au_x)D^2y + (f_{xx} + 2g_x + 3au_{xx})Dy$$
$$+ (g_{xx} + au_{xxx} + fu_x)y \tag{1.2.7}$$

A new partial differential equation for u now results when we again require that the coefficients of D^2y and Dy vanish. The vanishing of these coefficients yields simple differential relations that are readily integrated. We find that $f = -\frac{3}{2}au + c_1$ and $g = -\frac{3}{4}au_x + c_2$, where c_1 and c_2 are arbitrary functions of time that arise from integration. Then from (1.2.7),

$$[L, B_3]y = \left[\tfrac{1}{4}a(u_{xxx} - 6uu_x) + c_1 u_x\right]y \tag{1.2.8}$$

The partial differential equation satisfied by u again follows from the relation $-u_t + [L, B_3] = 0$. The constant a may be set equal to -4 to simplify the coefficients in the resulting equation. Also, the function $c_1(t)$ may be set equal to zero since it may be eliminated in the final equation for u by merely transforming to new independent variables given by $dx' = dx + c_1(t)\,dt$ and $dt' = dt$. The new equation for u is thus found to be

$$u_t - 6uu_x + u_{xxx} = 0 \tag{1.2.9}$$

If u is governed by this equation, the left-hand side of (1.2.4) will vanish and hence we again obtain $\lambda_t = 0$. Except for the factor of -6 (which could be eliminated by setting $u = -\frac{1}{6}U$) this is the nonlinear partial differential equation that was given in (1.2.1). It is one of the standard forms of the Korteweg–deVries equation. Thus, if the potential in a Schrödinger equation evolves according to the Korteweg–deVries equation, the eigenvalue parameter λ remains constant.

Finally, since the functions f and g in the operator B_3 are now known, the time dependence of the solution y is also known. It is given by

$$y_t = B_3 y = (-4D^3 + 6uD + 3u_x)y \tag{1.2.10}$$

The function $c_2(t)$ has also been set equal to zero since it may be eliminated by introducing a new dependent variable $\bar{y} = y \exp(\int c_2\,dt)$. It should be noted that both the spatial variation of y,

$$y_{xx} - uy = \lambda y \tag{1.2.11}$$

and the temporal variation given by (1.2.10) are expressed in terms of linear differential equations.

As might be expected at this point, an infinite sequence of higher-order equations, characterized by the odd linear operators $B_5, B_7, \ldots,$ may be constructed (Lax, 1968; Gardner et al., 1974). However, these higher-order evolution equations do not seem to arise in physical applications at present and will not be considered here. Instead, we shall examine two simple solutions of the Korteweg–deVries equation (1.2.9).

Single-Soliton Solution of the Korteweg–deVries Equation

The simplest solution of the Korteweg–deVries equation is the steady-state solution which is obtained by looking for a solution in the form $u(x-ct)$. The solution thus represents a disturbance that moves in the positive x direction at a constant velocity c. It will be shown in Chapter 4 that a steady-state pulse solution of the Korteweg–deVries equation (1.2.9) is

$$u = -\frac{c}{2}\,\mathrm{sech}^2\left[\frac{\sqrt{c}}{2}(x-ct)\right] \tag{1.2.12}$$

This solution exhibits a common feature of nonlinear waves in that the amplitude and velocity of the pulse are related. Larger-amplitude pulses move more rapidly and also are narrower in width. A simple integration shows that the width and amplitude of the pulse are related in such a way that

$$\int_{-\infty}^{\infty} dx\sqrt{|u|} = \pi \tag{1.2.13}$$

The solution given in (1.2.12), which represents a localized disturbance that is symmetric about its midpoint, is the single-soliton solution of the Korteweg–deVries equation. However, the true soliton nature of this expression is not yet evident. The essential element of the soliton is that the analytical form above is preserved, except for a phase shift, after the interaction of two or more such pulses. To see this preservation of form upon interaction, we must consider a more complicated solution than the mere steady-state result quoted above. Procedures for generating multisoliton solutions based upon inverse scattering techniques will be described in later chapters. To give a preview of these more general results, we shall here obtain the two-soliton solution by a simple method that predates the more sophisticated inverse scattering techniques. The method is one devised by Bargmann (1949) for application to the radial Schrödinger equation but which may be applied equally well in the present instance where the range of the independent variable is the entire x axis. The close connection between potentials and multisoliton solutions is brought out quite clearly and simply by this method.