Geometry of Solitons

Chuu-Lian Terng and Karen Uhlenbeck

A *solitary wave* is a traveling wave of the form $u(x,t) = f(x-ct)$ for some smooth function f that decays rapidly at infinity. It is relatively easy to find nonlinear wave equations that admit solitary wave solutions. For example,

$$u_{tt} - u_{xx} = u(2u^2 - 1)$$

has a family of solitary wave solutions

$$u(x,t) = \operatorname{sech}(x \cosh\theta + t \sinh\theta),$$

parameterized by $\theta \in \mathbb{R}$. But we do not expect that the "sum" of two such solutions will again be a solution. However, the special class of *soliton equations*, the subject of this article, does have a form of nonlinear superposition. An *n-soliton* solution is a solution that is asymptotic to a nontrivial sum of n solitary waves $\sum_{i=1}^n f_i(x - c_i t)$ as $t \to -\infty$ and to the sum of the same waves $\sum_{i=1}^n f_i(x - c_i t + r_i)$ with some nonzero phase shifts r_i as $t \to \infty$. In other words, after nonlinear interaction the individual solitary waves pass through each other, keeping their velocities and shapes but with phase shifts. Equations with multisoliton solutions are very rare (they occur nearly always in one space dimension); these equations are called *soliton equations*.

The Korteweg-de Vries equation

$$\text{(KdV)} \qquad q_t = -(q_{xxx} + 6qq_x)$$

is a well-known example of a soliton equation. If $q(x,t) = f(x-ct)$ solves KdV, then $f''' + 6ff' - cf' = 0$. The asymptotic condition on f, $\lim_{|x|\to 0} f(x) = 0$, implies that $f(x) = \frac{c}{2}\operatorname{sech}^2(\frac{\sqrt{c}}{2}x)$. So

$$q(x,t) = \frac{c}{2}\operatorname{sech}^2\left(\frac{\sqrt{c}}{2}(x - ct)\right)$$

is a family of solitary wave solutions for KdV parametrized by $c \in \mathbb{R}$.

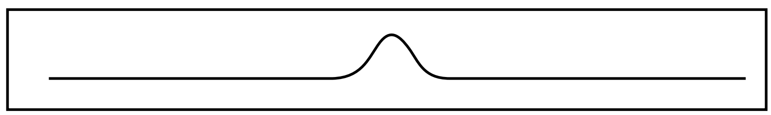


Figure 1. Wave profile for KdV 1-soliton.

KdV does have multisoliton solutions. In Figure 2 we graphically show the wave profiles of a 2-soliton solution of KdV by showing the graph of $q_i(x) = q(x, t_i)$ for a sequence of increasing times t_i. The asymptotic behavior and phase shifts can be seen in these pictures.

In classical mechanics a Hamiltonian system in $2n$-dimensions is called *integrable* if it has n independent constants of the motion whose Poisson brackets are all zero. The concept of complete integrability can be extended to infinite dimensions or partial differential equations (PDE), but is only one part of the rich structure found in the class of "completely integrable" or "soliton" equations. These equations are best described by their prototypes: KdV, nonlinear Schrödinger, and sine-Gordon equation (SGE). Phenomena that have been identified include:

- multisoliton solutions,
- Hamiltonian formulations,
- a hierarchy of commuting flows described by partial differential equations,
- formulation in terms of a Lax pair of operators,
- a scattering theory,

Chuu-Lian Terng is professor of mathematics at Northeastern University. Her e-mail address is `terng@neu.edu`. *Her research was supported in part by NSF Grant DMS 9972172.*

Karen Uhlenbeck has a Sid W. Richardson Foundation Regents Chair in mathematics at the University of Texas in Austin. Her e-mail address is `uhlen@math.utexas.edu`. *Her research was supported in part by the Sid Richardson Foundation Regents' Chair Funds of the University of Texas system.*

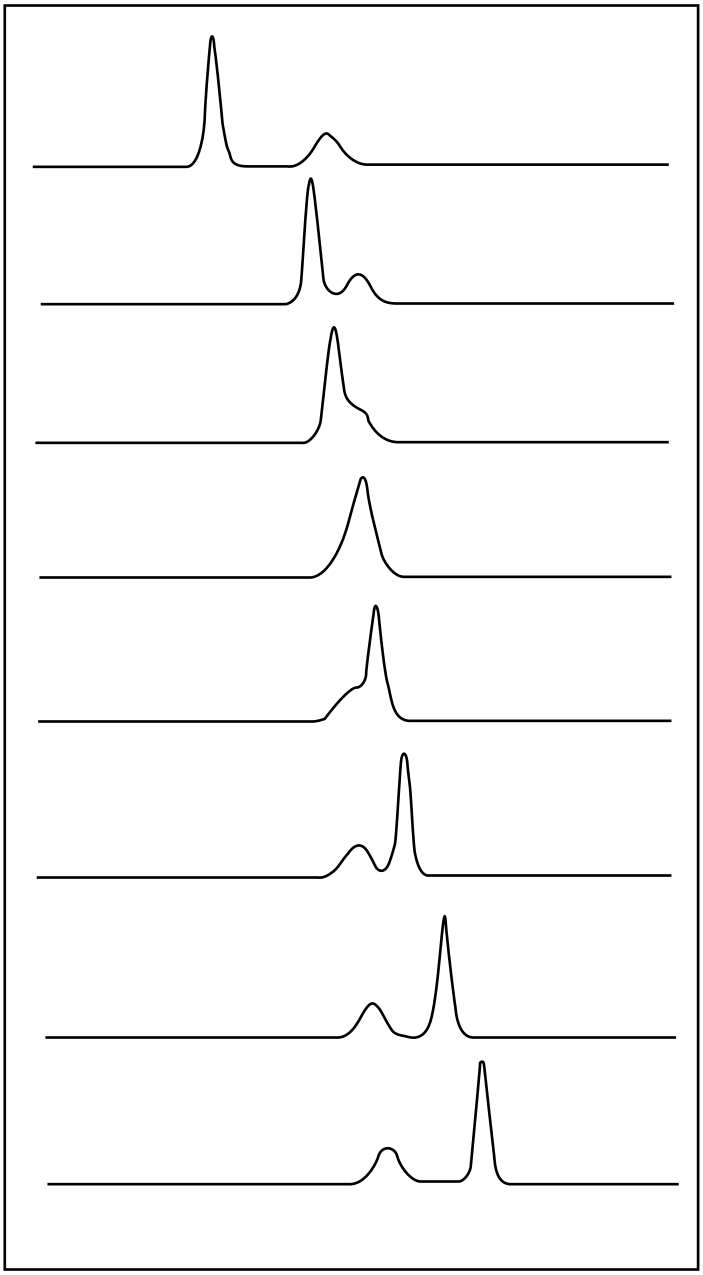


Figure 2. KdV 2-soliton.

- inverse scattering transforms,
- the Painlevé property,
- an algebraic geometric description of solutions periodic in space variable,
- formal algebraic solutions in terms of loop-group data.

In this article, after giving a little historical background on how soliton equations arose in classical differential geometry, we reinterpret the classical Bäcklund transformations in terms of actions of loop groups, or dressing transformations; we outline how different categories of solutions to the equations come about due to different choices of scattering data; and we finish with two of the most challenging open problems in the area.

A detailed list of references can be found in the *Journal of Differential Geometry Survey,* volume 4, on integrable systems [9] and in an expository article on solitons by R. Palais [7].

Solitons and the Classical Differential Geometry of Surfaces in $\mathbb{R}^3$

Most expositions of soliton theory outline the history of the Korteveg de Vries (KdV) equation, beginning with the physical observation of S. Russell of a bow wave in a canal in 1834. The equations were first written down by Boussinesq in 1871 and in 1895 by Korteweg and de Vries. In addition to describing water waves, the KdV equation also arises as a universal limit of lattice vibrations as the spacing goes to zero. The surprising numerical experiments of Fermi, Pasta, and Ulam in 1955 on an anharmonic lattice and the ingenious explanation by Zabusky and Kruskal in 1965 in terms of *solitons* of the KdV equation were quickly followed by a ground-breaking paper of Gardner, Greene, Kruskal, and Miura [3], which introduced the method of solving KdV using the inverse scattering transform for the Hill's operator. This brings us into the modern era. However, there is a separate circle of ideas that originates in geometry.

A central theme in the nineteenth century geometry was the local theory of surfaces in $\mathbb{R}^3$, which we might regard as the prehistory of modern constructions in soliton theory. The SGE arose first through the theory of surfaces of constant Gauss curvature -1 in $\mathbb{R}^3$, and the reduced 3-wave equation can be found in Darboux's work on triply orthogonal systems of $\mathbb{R}^3$. In 1906 a student of Levi-Civita, da Rios, wrote a master's thesis in which he modeled the movement of a thin vortex by the motion of a curve propagating in $\mathbb{R}^3$ along its binormal. It was much later, in 1971, that Hasimoto showed the equivalence of this system with the nonlinear Schrödinger equation $q_t = \frac{i}{2}(q_{xx} + 2\,|\,q\,|^2 q)$. These equations were rediscovered independently of their geometric history. The main contribution of the classical geometers lies in their methods for constructing explicit solutions of these equations rather than in their discovery of the equations themselves.

A few of the basic ideas from classical geometry of surfaces are needed to describe the geometric problems. Let M be a surface in $\mathbb{R}^3$. The *first fundamental form* I is the induced metric on tangent planes of M, i.e.,

$$I(v, w) = \langle v, w \rangle,$$

the dot product of v, w in $\mathbb{R}^3$. The negative of the unit normal e_3 is a map from M to the unit sphere S^2 of $\mathbb{R}^3$. Its differential A at a point maps the tangent space at that point to itself and is given by a symmetric map relative to I. A is called the *shape operator*. The *second fundamental form* is the canonical bilinear form defined by A, i.e.,

$II(v,w) = A(v)\cdot w$. If v is a unit tangent vector at p, then $II(v,v)$ is the curvature of the plane curve σ at p, where σ is the intersection of M and the plane spanned by v and the normal line at p. The *principal curvatures* of M are the eigenvalues of II with respect to I, which is the same thing as the eigenvalues of the shape operator A. The *mean curvature* is the arithmetic mean of the principal curvatures. The *Gaussian curvature* is the intrinsic curvature of the surface and is given by the product of the principal curvatures.

In regions where the principal curvatures are never equal, it is possible to choose *line of curvature coordinates*, which are coordinates along the eigendirections of A, or the directions of principal curvature. If the Gauss curvature is negative, it is possible to choose *asymptotic coordinates*, which are coordinates along the directions v in which $II(v,v) = 0$. Calculations are much easier in these special coordinate systems, and it is the choice of such special coordinate systems that links special geometry to interesting partial differential equations.

The Fundamental Theorem of Surfaces is based on a compatibility condition for ordinary differential equations in two independent variables, which is the same *zero-curvature condition* we will revisit when we discuss Lax pairs. If A and B are $n\times n$ matrices depending on two variables x and y, then the pair of equations

$$V_x = VA, \qquad V_y = VB \tag{1}$$

can be solved for $n\times n$ matrix valued maps V for all initial values exactly when

$$A_y - B_x = [A,B]. \tag{2}$$

Here $[A,B] = AB - BA$. This condition comes from requiring that $V_{xy} = V_{yx}$. We will call (2) the *compatibility condition* for (1).

We obtain such a system if we choose a parametrization $X(x,y)$ for a surface in $\mathbb{R}^3$ and $E = (e_1, e_2, e_3)$ a local orthonormal frame such that $e_3(x,y)$ is normal to M at $X(x,y)$. Note that $E(x,y)$ lies in the group $O(3)$ of 3×3 orthogonal matrices. Then we can write

$$X_x = E\begin{pmatrix} p \\ 0\end{pmatrix}, \quad X_y = E\begin{pmatrix} q \\ 0\end{pmatrix},$$
$$E_x = EP, \quad E_y = EQ,$$

where p and q are 2×1-matrix valued functions and P and Q are $so(3)$-valued functions (here $so(3)$ is the space of real 3×3 skew-symmetric matrices). Moreover, p, q, P, and Q can be expressed in terms of coefficients of I and II. Letting $V = \begin{pmatrix} E & X \\ 0 & 1\end{pmatrix}$, we obtain

$$V_x = VA, \quad V_y = VB,$$

where $A = \begin{pmatrix} p & P \\ 0 & 0\end{pmatrix}$, $B = \begin{pmatrix} q & Q \\ 0 & 0\end{pmatrix}$ are elements of the Lie algebra of the group of rigid motions of $\mathbb{R}^3$.

The compatibility conditions for the above first-order system for X and E are called the *Gauss and Codazzi equations*. This is a system of second-order partial differential equations involving six functions (the coefficients of I and II). As we will see later, the number of functions involved in the Gauss and Codazzi equations can be reduced by special coordinate choices. The Fundamental Theorem of Surfaces says that two fundamental forms that satisfy the Gauss and Codazzi equations determine a surface in $\mathbb{R}^3$ up to rigid motion.

Pseudospherical Surfaces and Bäcklund Transformations

An important contribution of the classical geometers is the study of pseudospherical surfaces, which led to SGE. A *pseudospherical surface* is a surface of Gaussian curvature -1 in $\mathbb{R}^3$. Such a surface has a special set of asymptotic coordinates in which the two fundamental forms are

$$I = dx^2 + \cos q\,(dx\,dt + dt\,dx) + dt^2,$$
$$II = \sin q\,(dx\,dt + dt\,dx),$$

where q is the angle between the x and t-curves. With this special choice of coordinates, the Gauss and Codazzi equations boil down to a single equation in q

$$q_{xt} = \sin q. \tag{SGE}$$

The Fundamental Theorem of Surfaces gives us a local correspondence between solutions of SGE and surfaces of constant Gaussian curvature -1 in $\mathbb{R}^3$ up to rigid motions. Although SGE has many global solutions defined on $\mathbb{R}^2$, the corresponding surfaces always have singularities. In fact, Hilbert proved that there is no complete immersed surface in $\mathbb{R}^3$ with sectional curvature -1.

The idea of Bäcklund transformations comes from a construction on pseudospherical surfaces called a line congruence. A *line congruence* in $\mathbb{R}^3$ is a two-parameter family of lines

$$L(u,v): \quad x(u,v) + \tau\xi(u,v), \quad -\infty < \tau < \infty.$$

A surface M given by $Y(u,v) = x(u,v) + t(u,v)\xi(u,v)$ for some smooth function t is called a *focal surface* of the line congruence if the line $L(u,v)$ is tangent to M at $Y(u,v)$ for all (u,v). Hence $\xi(u,v)$ lies in the tangent plane of M at $Y(u,v)$, which is spanned by $x_u + t_u\xi + t\xi_u$ and $x_v + t_v\xi + t\xi_v$. This implies that t satisfies the following quadratic equation: $\det(\xi, x_u + t\xi_u, x_v + t\xi_v) = 0$. In general, this quadratic equation has two distinct solutions for t. Hence generically each line congruence has two focal surfaces, M and M^*. This results in a diffeomorphism $\ell : M \to M^*$ such that the line joining p and $p^* = \ell(p)$ is tangent to

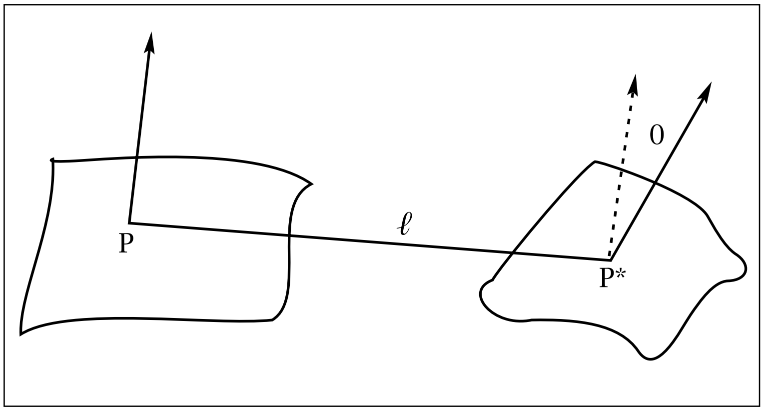


Figure 3. Pseudospherical line congruence.

both M and M^*. We will call ℓ also a *line congruence*.

A line congruence $\ell : M \to M^*$ is called *pseudospherical* with constant θ if the angle between the normal of M at p and the normal of M^* at $p^* = \ell(p)$ is θ and the distance between p and p^* is $\sin\theta$ for all $p \in M$ (see Figure 3).

In 1883 A. Bäcklund showed that if ℓ is a pseudospherical line congruence, then both M and M^* are pseudospherical and ℓ maps asymptotic lines to asymptotic lines. However, the transformations come about from showing that this construction can always be realized. For any pseudospherical surface M, constant θ, and unit vector $v_0 \in TM_{p_0}$ not a principal direction, there exist a unique surface M^* and a pseudospherical congruence $\ell : M \to M^*$ with constant θ such that $\overrightarrow{p_0 p_0^*} = (\sin\theta)v_0$. Analytically this is equivalent to the statement that if q is a solution of SGE, then the following overdetermined system of ordinary differential equations is solvable for q^*:

$$(3) \qquad \begin{cases} q^*_x = q_x + 4s\sin(\frac{q^*+q}{2}), \\ q^*_t = -q_t + \frac{2}{s}\sin(\frac{q^*-q}{2}), \end{cases}$$

where $s = \tan\frac{\theta}{2}$. Moreover, a solution q^* is again a solution of SGE. We will call both ℓ and the transform from q to q^* a *Bäcklund transformation*. This description of Bäcklund transformations gives us an algorithm for generating families of solutions of the PDE by solving a pair of ordinary differential equations. The procedure can be repeated, but the miracle is that after the first step, the procedure can be carried out algebraically. This is the Bianchi Permutability Theorem. Let $\ell_i : M_0 \to M_i$ be two pseudospherical congruences with angles θ_i respectively and with $\sin\theta_1^2 \neq \sin\theta_2^2$. Then there exist an algebraic construction of a unique surface M_3, and pseudospherical congruences $\tilde\ell_1 : M_2 \to M_3$ and $\tilde\ell_2 : M_1 \to M_3$ with angles θ_1 and θ_2 respectively such that $\tilde\ell_2\ell_1 = \tilde\ell_1\ell_2$. The analytic reformulation of this theorem is the following: Suppose q is a solution of the SGE and q_1, q_2 are two solutions of the above system (3) with angles $s_1 = \tan\frac{\theta_1}{2}$ and $s_2 = \tan\frac{\theta_2}{2}$ respectively. The Bianchi Permutability Theorem gives a third local solution q_3 to the SGE

$$\tan\frac{q_3 - q}{4} = \frac{s_1 + s_2}{s_1 - s_2}\tan\frac{q_1 - q_2}{4}.$$

To see how the scheme works, we start with the trivial solution $q = 0$ of SGE. Then (3) can be solved explicitly to get $q^*(x,t) = 4\tan^{-1}(e^{sx+\frac{t}{s}})$, which is the 1-soliton solution of SGE. Application of the Bianchi Permutability Theorem gives the 2-soliton solutions

$$q(x,t) = 4\tan^{-1}\left(\frac{s_1+s_2}{s_1-s_2}\,\frac{e^{s_1x+\frac{1}{s_1}t} - e^{s_2x+\frac{1}{s_2}t}}{1+e^{(s_1+s_2)x+(\frac{1}{s_1}+\frac{1}{s_2})t}}\right).$$

Repeated applications of the theorem give complicated but explicit n-soliton solutions. Note that the parameters s_1, s_2 in the above formula for 2-solitons are real. But for $s_1 = e^{i\theta}$ and $s_2 = -e^{-i\theta}$, although q_1, q_2 are not real valued,

$$q_3(x,t) = 4\tan^{-1}\left(\frac{\sin\theta\sin(T\cos\theta)}{\cos\theta\cosh(X\sin\theta)}\right)$$

is real and a solution of the SGE, where $X = x - t$ and $T = x + t$ are the laboratory coordinates. This solution is periodic in T and is called a *breather*.

The surface corresponding to $q = 0$ degenerates to a straight line. In Figure 4 we show the wave profile $\frac{\partial q}{\partial X}(\cdot, T)$ of a 1-soliton of SGE and the corresponding surfaces for a sequence of different s_i. We show in Figures 5 and 6 wave profiles $\frac{\partial q}{\partial X}(\cdot, T_i)$ of a 2-soliton and a breather of SGE for a sequence of increasing times T_i. Alongside each is the corresponding pseudospherical surface. We use $s_1 = 1$ and $s_2 = 1/\sqrt{3}$ for the 2-soliton, and we use $s_1 = \frac{1}{5}(4+3i)$ and $s_2 = -\bar{s}_1$ for the breather. The time sequence T_i for the breather covers a half period.

Lie Transformations and Lax Pairs

A key development in the understanding of soliton theory occurred in 1968 when P. Lax observed that many properties of KdV are explained from an associated linear isospectral problem [4]. This associated linear system has come to be called a "Lax pair".

If a PDE is given with q as dependent variable, a linear system depending on a parameter

$$E_x = EA, \quad E_t = EB$$

in which A and B are $n \times n$ matrix-valued functions of q and x-derivatives of q is called a *Lax pair* of the PDE if the compatibility condition $A_t - B_x - [A,B] = 0$ is the PDE in question. Fix t, and let $L(t)$ denote the operator $\frac{\partial}{\partial x} + A$. The compatibility condition can be rewritten as a Lax equation

$$\frac{\partial L}{\partial t} = [L, B].$$

This implies that $L(t)$ is conjugate to $L(0)$ for all t. In other words, $L(t)$ is isospectral.

A Lax pair for SGE was constructed by M. Ablowitz, D. Kaup, A. Newell, and H. Segur in 1973. We will explain how this Lax pair can be obtained from the classical Lie transformations. Sophus Lie observed that the SGE is invariant under Lorentz transformations. In asymptotic coordinates, which correspond to light cone coordinates, a Lorentz transformation is $(x,t) \mapsto (\frac{1}{\lambda}x, \lambda t)$. If q is a solution of the SGE, then so is $q^\lambda(x,t) = q(\frac{1}{\lambda}x, \lambda t)$ for any nonzero real constant λ. Let M_λ denote the pseudospherical surface corresponding to q^λ. There is a relation between Lie transformations and the modern notion of a Lax pair. Namely, choose for each λ a local orthonormal frame $e_1^\lambda, e_2^\lambda, e_3^\lambda$ on M^λ such that e_1^λ is $\frac{\partial}{\partial x}$ and e_3^λ is normal to M^λ. Let E^λ be the $O(3)$-valued map such that $e_1^\lambda, e_2^\lambda, e_3^\lambda$ are the columns, and $w^\lambda = (E^\lambda)^{-1}dE^\lambda$ the $so(3)$-valued 1-form. Substitute $(\frac{1}{2\lambda}x, 2\lambda t)$ for (x,t) in ω^λ to get a one-parameter family of flat $so(3)$-valued 1-forms. Here a $n \times n$-valued 1-form w is *flat* if $dw + w \wedge w = 0$. To get the Lax pair of SGE, we need to identify the Lie algebra $so(3)$ with the Lie algebra $su(2)$ of skew Hermitian 2×2 matrices of trace 0. This allows us to rewrite the family of the flat $so(3)$-connections as a family of flat $su(2)$- valued 1-forms:

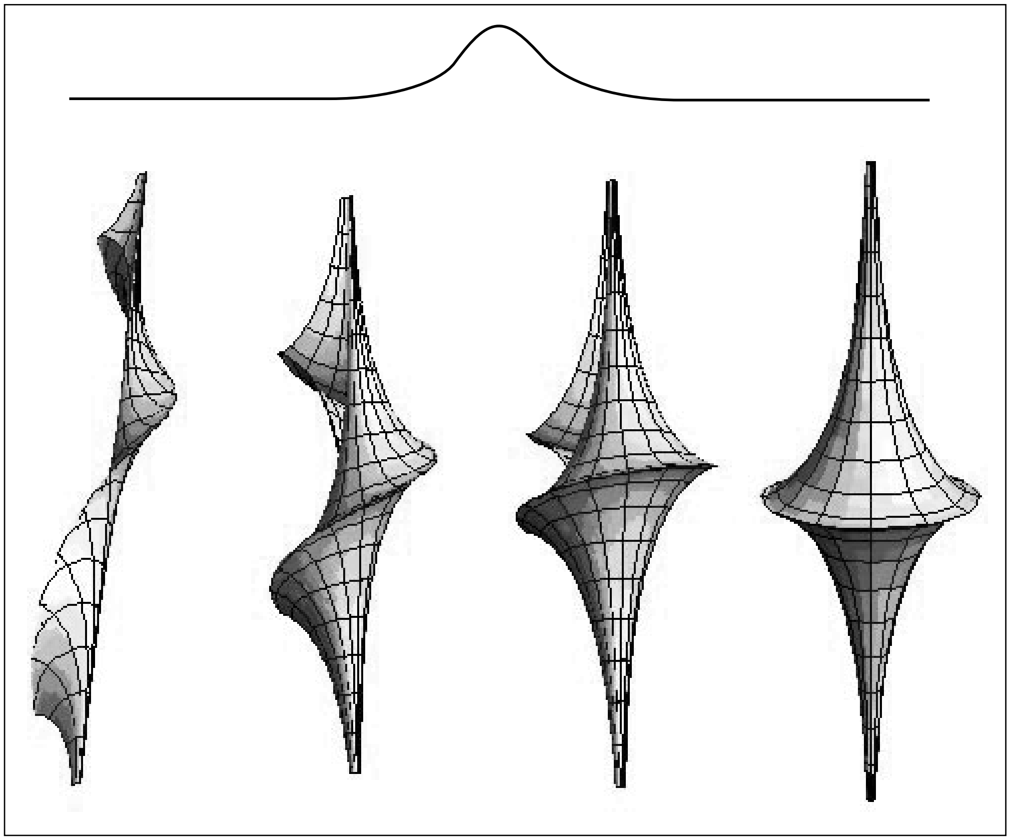


Figure 4. SGE 1-soliton wave and corresponding surfaces.

$$\begin{pmatrix} -i\lambda & -\frac{q_x}{2} \\ \frac{q_x}{2} & i\lambda \end{pmatrix} dx + \frac{i}{4\lambda}\begin{pmatrix} \cos q & -\sin q \\ -\sin q & -\cos q \end{pmatrix} dt.$$

Flatness is equivalent to the solvability of the following first-order system for all λ:

$$(4) \qquad \begin{cases} E_x = E\begin{pmatrix} -i\lambda & -\frac{q_x}{2} \\ \frac{q_x}{2} & i\lambda \end{pmatrix}, \\ E_t = \frac{i}{4\lambda}E\begin{pmatrix} \cos q & -\sin q \\ -\sin q & -\cos q \end{pmatrix}. \end{cases}$$

The compatibility condition for this system is exactly that $q_{xt} = \sin q$. System (4) is the Lax pair for the SGE.

Bäcklund Transformations and Dressing Actions

Many beautiful transformations constructed by classical geometers play an important role in constructing solutions to various soliton equations. In addition to Bäcklund and Lie transformations, there are Ribaucour, Bianchi, and Darboux transformations. They were natural to the classical geometers, but seem to come from nowhere from the point of view of partial differential equations.

The modern explanation begins with "dressing transformations", which appeared first in an article by Zakharov and Shabat [12]. Bäcklund transformations will be the simplest kind of dressing transformations. Let $E(x,t,\lambda)$ be the solution of (4) satisfying the initial condition $E(0,0,\lambda) = I$. We will call this E the *frame* for the solution q (it is also called a *wave function*). Since the coefficients of system (4) depend holomorphically on the parameter $\lambda \in \mathbb{C} \setminus \{0\}$, the solution $E(x,t,\lambda)$ depends holomorphically on $\lambda \in \mathbb{C} \setminus 0$. Let G denote the group of meromorphic maps $f : \mathbb{C} \to SL(2,\mathbb{C})$ that are regular at $\lambda = 0$ and ∞ and satisfy

$$(*) \qquad f(\bar\lambda)^* f(\lambda) = I, \quad \overline{f(-\bar\lambda)} = f(\lambda).$$

We can factor the product $f(\lambda)E(x,t,\lambda)$ as the product $\tilde E(x,t,\lambda)\tilde f(x,t,\lambda)$ so that $\tilde E(x,t,\lambda)$ is

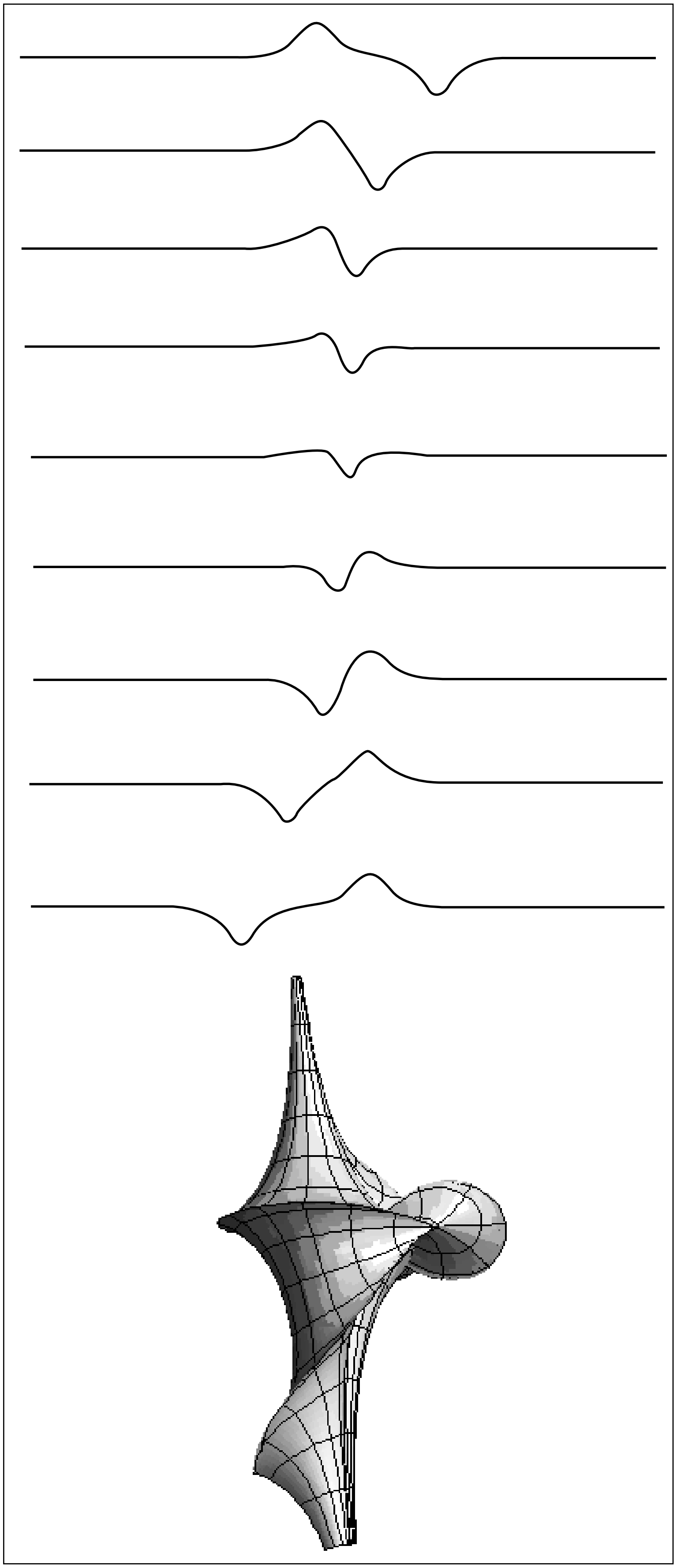


Figure 5. SGE 2-soliton wave and corresponding surface.

holomorphic for $\lambda \in \mathbb{C} \setminus 0$ and $\tilde{f}(x,t,\cdot)$ is in G. (Note we reverse the order.) A direct computation shows that $\tilde{E}$ satisfies an equation of the form (4) for some $\tilde{q}$. Hence $\tilde{E}$ is the frame for a new solution $\tilde{q}$ of SGE. The transformation

$$q \mapsto f \sharp q := \tilde{q}$$

defines an action of G on the space of solutions of SGE and is an example of a *dressing transformation*.

A meromorphic map f with a single simple pole is called a *simple factor*, and the dressing transformation given by a simple factor can be obtained algebraically. Let s be a nonzero real number, V_0 a one-dimensional linear subspace of $\mathbb{R}^2$, π_0 the orthogonal projection of $\mathbb{R}^2$ onto V_0, and $\pi_0^\perp = I - \pi_0$. Then

$$f_{is,\pi_0}(\lambda) = \pi_0 + \frac{\lambda - is}{\lambda + is}\,\pi_0^\perp$$

is a simple factor. Note that $f_{is,\pi_0}(\lambda)^{-1} = f_{-is,\pi_0}(\lambda)$. By residue calculus, we see that if we can factor $f_{is,\pi_0}E$ as $\tilde{E}\tilde{f}$, then $\tilde{f}(x,t,\lambda)$ must be of the form $f_{is,\pi(x,t)}$ for some projection $\pi(x,t)$. So we need to find $\pi(x,t)$ such that

$$\tilde{E}(x,t,\lambda) = \left(\pi_0 + \frac{\lambda - is}{\lambda + is}\,\pi_0^\perp\right) \times \; E(x,t,\lambda)\left(\pi(x,t) + \frac{\lambda + is}{\lambda - is}\,\pi^\perp(x,t)\right)$$

is holomorphic for $\lambda \in \mathbb{C} \setminus 0$. Hence the residue of the right-hand side at $\pm is$ should be zero. This implies that $\pi(x,t)$ has to be the projection of $\mathbb{R}^2$ onto the subspace

$$V(x,t) = \overline{E(x,t,is)}^{\,t}(V_0).$$

Moreover, $\tilde{E}$ turns out to be the frame for a new solution $\tilde{q}$ of the SGE. Set $A = E^{-1}dE$ and $\tilde{A} = \tilde{E}^{-1}d\tilde{E}$. The formula for $\tilde{E}$ implies that

$$\tilde{f}A - d\tilde{f} = \tilde{A}\tilde{f}.$$

We obtain the pair of ordinary differential equations (3) relating q and $\tilde{q}$ by comparing the residues at $\lambda = \pm is$.

The Bianchi Permutability Theorem also follows naturally from this point of view. Given nonzero real numbers s_1, s_2 such that $s_1^2 \neq s_2^2$ and two projections π_1, π_2 of $\mathbb{R}^2$, we can find two projections ξ_1, ξ_2 and g such that

$$g = f_{is_1,\xi_1} f_{is_2,\pi_2} = f_{is_2,\xi_2} f_{is_1,\pi_1}.$$

Moreover, ξ_1, ξ_2 can be written algebraically in terms of s_1, s_2, π_1, and π_2 and hence g. By the same reasoning, if $gE = \tilde{E}\tilde{g}$, then $\tilde{g}$ is algebraic in $\tilde{f}_1(x,t,\lambda)$ and $\tilde{f}_2(x,t,\lambda)$. This reinterprets the classical permutability formulas as being the consequence of the noncommutativity of simple factors. Moreover, this procedure works for any equations with Lax pairs. Hence it gives a general scheme for constructing Bäcklund transformations for any soliton equations. A detailed description of this can be found in [10].

Types of Solutions

Given a rapidly decaying function u on the line, consider the first equation of (4) with asymptotic and boundedness conditions:

$$\begin{cases} \psi_x = \psi(a\lambda + u), \\ \lim_{x\to-\infty} e^{-a\lambda x}\psi(x,\lambda) = I, \\ e^{-a\lambda x}\psi(x,\lambda) \text{ bounded in } x. \end{cases} \tag{5}$$

Beals and Coifman showed that this system has a unique solution $\psi^u(x,\lambda)$ and

$$m^u(x,\lambda) = e^{-a\lambda x}\psi^u(x,\lambda)$$

is meromorphic for $\lambda \in \mathbb{C} \setminus \mathbb{R}$, has asymptotic expansion at $\lambda = \infty$, and may have a discontinuity along the real line in the λ-plane. The singularity data S^u of m^u is the scattering data for u, and the map from u to S^u is called the *scattering transform*. The inverse of this map is the *inverse scattering transform*. Let $f(\lambda) = m^u(0,\lambda)$, and define

(6)
$$E(x,\lambda) = f(\lambda)^{-1}e^{a\lambda x}m^u(x,\lambda) = f(\lambda)^{-1}\psi^u(x,t,\lambda).$$

Since ψ^u is a solution of $\psi^{-1}\psi_x = a\lambda + u$ and since $E(x,\lambda)$ and $\psi^u(x,\lambda)$ differ by a multiplicative constant matrix $f(\lambda)$,

$$f(\lambda)^{-1}e^{a\lambda x} = E(x,\lambda)m^u(x,\lambda)^{-1}. \tag{7}$$

This means that we have replaced the inverse scattering transform by a factorization problem. Namely, to obtain $u(x)$ from $f(\lambda)$, we first factor $f(\lambda)^{-1}e^{a\lambda x}$ as a product $E(x,\lambda)m^{-1}(x,\lambda)$ so that $E(x,\lambda)$ is holomorphic for $\lambda \in \mathbb{C} \setminus 0$ and $m(x,\lambda)$ has the same type of singularities as $f(\lambda)$. Then $u = E^{-1}E_x - a\lambda$. Moreover, if $u(x,t)$ is a solution of SGE, then use of the Lax pair shows that the scattering data of $u(\cdot,t)$ can be expressed explicitly in terms of that of $u(\cdot,0)$. Therefore, the inverse scattering transform can be used to solve Cauchy problems with rapidly decaying inital data. This motivates us to introduce a more general notion of scattering data to include many different classes of solutions other than the rapidly decaying class.

In the language of this article, the goal of scattering theory is to identify from a given solution q an f such that $q = f^{-1}\sharp 0$. In other words, for the frame E of q, we wish to solve

$$f(\lambda)^{-1}e^{a(\lambda x - \frac{1}{4\lambda}t)} = E(x,t,\lambda)\tilde{f}(x,t,\lambda)^{-1}$$

for some f and $\tilde{f}$ so that f and $\tilde{f}$ have power series expansion at 0 and ∞. Here $a = \mathrm{diag}(-i,i)$, and $e^{a(\lambda x - \frac{1}{4\lambda}t)}$ is the frame for the trivial solution

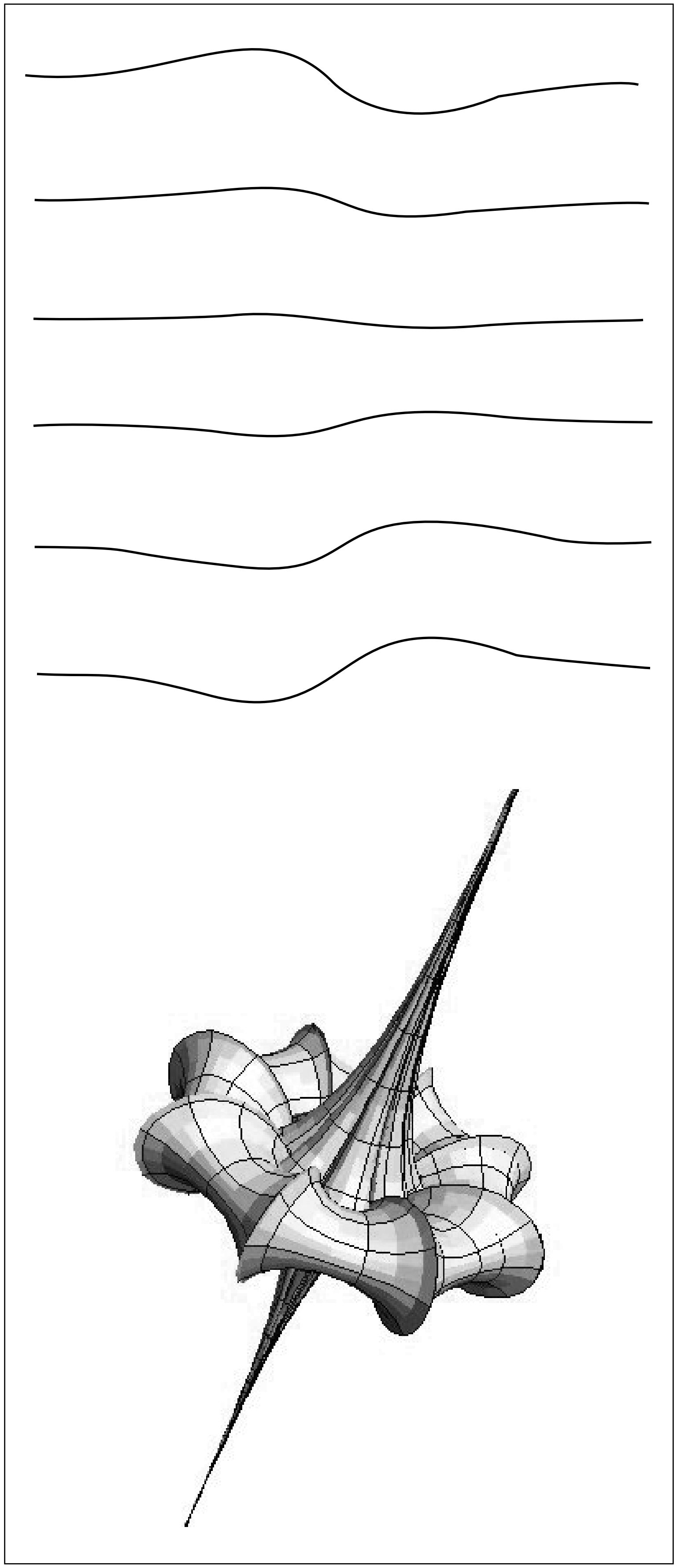


Figure 6. SGE breather wave and corresponding surface.

$q = 0$ of SGE. We will call f *scattering data* for the solution q.

Inverse scattering theory concerns reconstructing solutions q from f. There are classical methods in complex variables that in principle accomplish this. For example, Birkhoff factorizations and Riemann-Hilbert problems are two of the techniques. The text by Pressley and Segal [8] and the paper by Beals and Coifman [1] are good basic references.

To make the theory rigorous, it is necessary to identify exactly the class of solutions one wishes to construct. Here are some examples. In the following, we assume f satisfies condition $(*)$ and is regular at 0 and ∞.

• Soliton solutions. Choose f as a rational map. This is the case of applying Bäcklund transformations repeatedly, starting with the trivial solution.

• Local analytic solutions. Choose f as a local holomorphic map on a neighborhood of 0 and ∞. Solutions obtained from such f are locally defined analytic functions.

• Algebraic geometry solutions. Choose f as for the local analytic solutions but require that $f^{-1}(\lambda)af(\lambda)$ is a finite Laurent polynomial. The factorization can be done by solving a system of ordinary differential equations. This class of solutions includes the so-called "finite-gap" periodic solutions obtained by theta function theory on Riemann surfaces. McKean's article [5] gives an exposition of this type of solution.

• Formal algebraic solutions. Choose f as a pair of formal power series based at 0 and ∞. The factorization can be done formally, and hence many formal solutions can be constructed. Solutions from this class appear in conformal field theory and string theory. Van Moerbeke has a nice set of expository lectures on this [6].

• Scaling invariant solutions. Each soliton equation has an action of the multiplicative group $\mathbb{R}^*$ of nonzero real numbers on the space of solutions. For example, the $\mathbb{R}^*$-action for SGE is the Lie transformation, i.e., $r \cdot q(x,t) = q(rx, \frac{1}{r}t)$. A *scaling invariant* solution is a solution that is invariant under the $\mathbb{R}^*$-action. These solutions can be understood by isomonodromy methods. Beals and Sattinger give a good description of this in [2]. Solutions of this type appear in quantum cohomology.

• Schwartz class (rapidly decaying) solutions. Here f is meromorphic with finitely many poles on $\mathbb{C} \setminus \mathbb{R}$, f has an asymptotic expansion at $\lambda = 0$ and ∞, and the limits $\lim_{s \to 0^+} f(r \pm is)$ are smooth functions. Soliton solutions are a subclass of this class.

We have not explained in detail the role of condition $f(\bar{\lambda})^* f(\lambda) = I$, which is a reality condition for both SGE and nonlinear Schrödinger equations. It is a condition associated to the compact group $SU(2)$. This condition can be dropped or replaced by conditions describing a real form of any complex simple Lie group or one of the symmetric spaces. In general, inverse scattering transforms, $f \mapsto f\sharp 0$, have singularities unless the scattering data f are chosen carefully. It is one of our contributions to observe that if the reality conditions contain those associated to a compact Lie group, then the inverse scattering transforms yield solitons and Schwartz class solutions.

Some Intriguing Open Problems

The recent interest of geometers in integrable systems was initiated by S. S. Chern in the 1970s. Chern and his coworkers were interested in using submanifold geometry to find new examples of soliton equations. Both E. Calabi and Chern had used geometric methods to describe minimal surfaces in spheres. In the early 1980s physicists discovered that 2-dimensional Einstein equations, harmonic maps, and self-dual Yang-Mills all had formulations in terms of Lax pairs. Wente's construction of a counterexample to the Hopf conjecture, an immersed torus in $\mathbb{R}^3$ with constant mean curvature, was later observed to be constructible by methods from soliton theory. This has led to several decades of work by both mathematicians and physicists on a new class of problems that can be static (elliptic) or evolution equations, depending on the signature of space-time. We describe two different open problems at this interface.

A map $s : \mathbb{C} \to SU(n)$ is *harmonic* if it is a critical point for the energy functional

$$\mathcal{E}(s) = \int_{\mathbb{C}} \mathrm{tr}((s^{-1}s_x)^2 + (s^{-1}s_y)^2)\, dx\, dy.$$

The Euler-Langrange equation written in terms of $P := \frac{1}{2}s^{-1}\frac{\partial s}{\partial \bar{z}}$ and $Q := \frac{1}{2}s^{-1}\frac{\partial s}{\partial z} = -P^*$ is

$$\frac{\partial P}{\partial \bar{z}} = \frac{\partial Q}{\partial z} = -[P, Q].$$

K. Pohlmeyer was the first to find the Lax pair for this equation:

$$E^{-1}E_z = (1 - \lambda)P, \qquad E^{-1}E_{\bar{z}} = (1 - \lambda^{-1})Q.$$

The first definite progress in understanding this equation was due to the physicists A. Din, V. Glaser, R. Stora, and W. Zakrzewski. Their work was publicized among mathematicians by Eells and his coworkers and led to a good understanding of harmonic maps from S^2 to $SU(n)$. Perhaps it is not so surprising that information on the solutions on T^2 is recovered via the algebraic geometry methods that were developed for periodic solutions for soliton equations. The Gauss map of a constant mean curvature surface in $\mathbb{R}^3$ is harmonic, and these methods have been applied to constant mean curvature tori in $\mathbb{R}^3$. The recent survey article by N. Hitchin discusses the state of knowledge about these geometric problems in [9], as well as the Euclidean monopoles we mention below.

On the other hand, one does not understand the type of scattering data $f(\lambda)$ that will give a harmonic map from a surface of genus greater than one to $SU(n)$. To characterize these scattering data, one needs a more complete understanding of periodic solutions than is currently available.

The anti-self-dual Yang-Mills equations in $\mathbb{R}^4 = \mathbb{C}^2$ can be described rather simply in terms of their Lax pairs. The equations are for an $SU(2)$-connection, which we write in complex coordinates in terms of two 2×2 traceless complex matrices $A_{\bar z}, A_{\bar w}$ and their Hermitian adjoints $A_z = -A_{\bar z}^*$, $A_w = -A_{\bar w}^*$. The commutator of two complex operators

$$D_\lambda = \frac{\partial}{\partial \bar w} + A_{\bar w} + \lambda \frac{\partial}{\partial z} + A_z,$$

$$D'_\lambda = \frac{\partial}{\partial w} + A_w \mp \lambda^{-1} \frac{\partial}{\partial \bar z} + A_{\bar z}$$

encode the equation in its Lax pair

$$[D_\lambda,\ D'_\lambda] = 0.$$

The choice of minus sign in D'_λ gives the elliptic equation on $\mathbb{R}^4$ familiar to mathematicians, while the choice of plus sign gives an equation on $\mathbb{R}^{2,2}$.

The monopole equations are obtained by letting $w = t + iu$ and assuming all the fields are independent of u. If the minus sign is used, a monopole equation on $\mathbb{R}^3$ is obtained, which has been analyzed in a beautiful sequence of papers by Nahm, Donaldson, and Hitchin. If the plus sign is used, a wave equation on $\mathbb{R}^{2,1}$ is obtained, and multisoliton solutions have been found by R. Ward and his coworkers. The cover image shows some features of a 3-soliton solution. This is an interesting and natural equation, which is gauge invariant and has both solitons and solutions of the linear wave equation as special solutions. Further investigation of scattering and inverse scattering theory for this equation may lead to some new ideas about integrable systems.

Acknowledgments

The authors would like to thank A. Knapp for many excellent suggestions on exposition, as well as D. Freed, T. Karcher, T. Kimura, X. Liu, R. Palais, and C. Taubes, who read the first draft of our paper and gave us many useful comments. Illustrations for this article were produced using R. Palais's 3D-Filmstrip program.

Bibliography

[1] R. Beals and R. R. Coifman, Scattering and inverse scattering for first order systems, *Comm. Pure Appl. Math.* **37** (1984), 39–90.

[2] R. Beals and D. H. Sattinger, Integrable systems and isomonodromy deformations, *Physica D* **65** (1993), 17–47.

[3] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Method for solving the Korteweg-de Vries equation, *Physics Rev. Lett.* **19** (1967), 1095–1097.

[4] P. D. Lax, Integrals of nonlinear equations of evolution and solitary waves, *Comm. Pure. Appl. Math.* **31** (1968), 467–490.

[5] H. P. McKean, Integrable systems and algebraic curves, *Global Analysis* (Proc. Biennial Sem. Canad. Math. Congr., Univ. Calgary, Calgary, 1978), Lecture Notes in Math., vol. 755, Springer, Berlin, 1979, pp. 467-490.

[6] P. van Moerbeke, Integrable foundations of string theory, *Lectures on Integrable Systems* (Sophia-Antipolis, 1991), World Sci. Publ., NJ, 1994, pp. 163-267.

[7] R. S. Palais, The symmetries of solitons, *Bull. Amer. Math. Soc. (N.S.)* **34** (1997), 339–403.

[8] A. Pressley and G. B. Segal, *Loop Groups*, Oxford Science Publ., Clarendon Press, Oxford, 1986.

[9] C. L. Terng and K. Uhlenbeck, *Integrable Systems*, J. Differential Geometry Survey, vol. 4, International Press, Cambridge, MA, 1999.

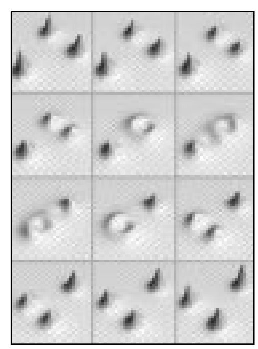
[10] ____, Bäcklund transformations and loop group actions, to appear in *Comm. Pure. Appl. Math.*

[11] R. S. Ward, Soliton solutions in an integrable chiral model in $2 + 1$ dimensions, *J. Math. Phys.* **29** (1988), 386–389.

[12] V. E. Zakharov and A. B. Shabat, Integration of nonlinear equations of mathematical physics by the inverse scattering method, II, *Funct. Anal. Appl.* **13** (1979), 166–174.

About the Cover

Solitons are usually associated with certain so-called integrable nonlinear wave equations in one space dimension (in particular the Korteweg-de Vries, Sine-Gordon, and nonlinear Schrödinger equations), and for a long time it was not clear whether soliton behavior was possible in higher space dimensions. Indeed, a simple scaling argument ("Derrick's Theorem") showed that in more dimensions, wave equations that arise from classical field theories could not have soliton solutions. But Richard Ward's Modified Chiral Model is a dimension reduction of the self-dual Yang-Mills gauge theory model, and Derrick's Theorem does not apply. In fact, Ward was able to use twistor methods to prove the existence of soliton solutions of his model, and Christopher Anand devised an algorithm for writing down explicit n-soliton solutions. The cover picture shows twelve frames of a flipbook animation of an Anand-Ward 3-soliton interaction, and was produced by R. Palais's 3D-Filmstrip visualization program using Pascal routines provided by Anand. In the first four frames, the two solitons on the right approach each other along a line. They merge in frame five and in frame six this combined object fissions into two solitons that move apart at right angles to the original direction of approach. One of these fission product solitons then coalesces with the third of the original solitons in frame seven, and in frames eight to twelve there is another fission of this second combined object, with the two solitons that result moving away from each other in a line parallel to the line along which the original pair approached.



— R. Palais