JHEP HYPER VERSION

Received: *February 13, 1999,* Accepted: *April 3, 1999*

Barrier penetration in $1+1$-dimensional O(n) sigma models

Enrique Moreno[*]

The City University of New York,
C.C.N.Y. and Baruch College, New York
E-mail: moreno@scisun.sci.ccny.cuny.edu

Peter Orland[*†]

The Niels Bohr Institute,
Blegdamsvej 17, DK-2100, Copenhagen Ø, Denmark
and The City University of New York,
Baruch College and The Graduate School and University Center, New York, NY[‡]
E-mail: orland@alf.nbi.dk, orland@gursey.baruch.cuny.edu

Abstract: The O(n) nonlinear sigma model in $1+1$ dimensions is examined as quantum mechanics on an infinite-dimensional configuration space. Two metrics are defined in this space. One of these metrics is the same as Feynman's distance, but we show his conclusions concerning potential energy versus distance from the classical vacuum are incorrect. The potential-energy functional is found to have barriers; the configurations on these barriers are solitons of an associated sigma model with an external source. The tunneling amplitude is computed for the O(2) model and soliton condensation is shown to drive the phase transition at a critical coupling. We find the tunneling paths in the configuration space of the O(3) model and argue that these are responsible for the mass gap at $\theta = 0$. These tunneling paths have half-integer topological charge, supporting the conjecture due to Affleck and Haldane that there is a massless phase at weak coupling and $\theta = \pi$.

Keywords: Field Theories in Lower Dimensions, Sigma Models, Nonperturbative Effects, Solitons Monopoles and Instantons.

[*]Work supported by CUNY-Collaborative grant 991999.
[†]Work supported by PSC-CUNY Research Award Program grant no. 668460.
[‡]Permanent address.

JHEP04(1999)002

Contents

1. Introduction

The two-dimensional O(3) nonlinear sigma model has been regarded as interesting by elementary-particle physicists because of the features it shares with QCD, such as asymptotic freedom [1] and instantons [2]. Condensed-matter physicists have studied the model in connection with antiferromagnetic spin chains [3], the quantum hall effect [4] and (more traditionally) classical ferromagnets. The O(n) model can be studied by the $1/n$ expansion and the O(2) model can be understood a Coulomb gas of vortices [5]. The spectrum and S-matrix of the O(3) model can be determined exactly [6, 7]. Haldane argued that at $\theta = \pi$ the model should be equivalent to the half-integer XXX spin chain and is therefore gapless [3].

In spite of these successes, there are still unresolved issues. The spin-chain arguments, though extremely compelling, need to be supplemented by other evidence, since the "Haldane mapping" [3] is not strictly a mapping at all, but relies on some assumptions about local properties of the ground state. Indeed, naively one might conclude from the spin-chain arguments that any O(3) invariant, integer-spin anti-ferromagnetic chain has a gap; yet some integer-spin chains are gapless [8].

In support of Haldane's conclusion Affleck and Haldane [9] suggested that condensation of merons [10], objects with half-integer topological charge, was responsible for the gap at $\theta = 0$, while at $\theta = \pi$ and small bare coupling the merons bind in pairs, rendering the spectrum gapless. These arguments were not precise for the O(3) sigma model, but were made by a conjectured extrapolation from a model with O(3) broken to O(2). Nonetheless, numerical work of Bietenholz et. al. with the symmetry unbroken suggests that this picture is valid [11]. Other observations concerning the role of half-integer topological charge were made recently in reference [12].

In this paper we study the O(2) and O(3) nonlinear sigma models using some ideas developed for non-Abelian gauge theories [13]. Some clarification of these ideas in reference [13], as well as some preliminary work for the sigma model is presented in reference [14].

The starting point is the definition of distance functionals on the fields. We work with two such functionals. One of these is the natural distance related to the kinetic term. The other is defined on the orbit space, which consists of configurations modulo global rotations. The latter distance functional is implicit in the work of Feynman [15]. Each of these functionals is a metric. The physical utility of these metrics are discussed. The second metric is used to define a coordinate on the orbit space. We then examine the minima of the potential energy for fixed values of this coordinate. What appears is a structure of energy barriers in configuration space. On these barriers, the configurations are identical to solitons — not of the ordinary sigma model, but of a related classical field theory. These solitons are topologically stable for the O(2) model, but not for the O(3) model. They can be written down explicitly in terms of elliptic and related functions. We show that the barriers of the O(2) model can be penetrated only above a critical coupling. At this coupling a phase transition is driven by condensation of solitons. We thereby obtain an entirely quantum-mechanical explanation of the transition, complementing the standard statistical-mechanical explanation [5]. The barriers of the O(3) sigma model are considerably more complicated than those of the O(2) model. We find these and examine them in some detail. We then show that the barrier-penetration process has topological charge one half; a result which strongly supports the arguments of Affleck and Haldane [9].

Our hamiltonian analysis is quite different from the usual path-integral saddle-point approach to the sigma model. The viewpoint is one almost never considered for the model, except in the appendix in Feynman's paper [15]. As the style of his article is difficult to follow, few physicists have pursued the ideas further. We therefore see the need for a nontechnical summary of our motivation and results, which is the content of the next section.

2. Distance and potential-energy topography

Field theories are infinite-dimensional quantum systems. In the Schrödinger representation, wave functionals are mappings from a *field configuration space* to the complex plane. The elements of this space, for the time being denoted as φ, are functions from the space manifold into some target manifold. In general, there is a hamiltonian

$$H = e_0 T + \frac{1}{e_0} V[\varphi]\,,$$

where the kinetic term T is an operator which does not commute with φ and e_0 is what we will call the bare coupling constant. In fact, e_0 is not the standard coupling of some field theories, but we present the hamiltonian in this form anyway (since it will be convenient in our analysis of the sigma model). Some regularization is always at least implicit in our discussion.

To be clear as to what we mean by potential energy, the potential-energy operator $V[\varphi]$ contains gradient terms. For example, a scalar theory has the potential energy

$$V[\varphi] = \int d^D x \left\{ \frac{1}{2} [\nabla \varphi_b(x)]^2 + \mathcal{V}(\varphi_b(x)) \right\},$$

in dimension $D+1$, where the species are labeled by b and $\mathcal{V}$ is a local potential functional.

There is a natural notion of a *distance* between any pair of configurations. There is nothing vague about this idea, as we will attempt to show. Suppose the spectrum is to be calculated with a strong-coupling expansion (say on a space lattice with continuous time). Such an approximation proceeds by first diagonalizing T, then treating V as a perturbation. One way to proceed towards solving for the spectrum is through the representation of the evolution operator through the Lie-Trotter product formula:

$$\mathcal{U}(t) = e^{iHt} = \lim_{N\to\infty} \left(e^{ie_o T\epsilon} e^{iV\epsilon/e_0} \right)^N \Big|_{N\epsilon = t}\,,$$

which is then expanded in powers of V. Such an approximation is very crude and requires resummation techniques to be made of practical value; but such details, though mathematically detailed, are not conceptually sophisticated. In any case, at leading order, one expects rather generally that the matrix element of the exponential of the kinetic term between two states localized in configuration space, is of the form

$$\langle \varphi_1 | e^{ie_0 T \epsilon} | \varphi_2 \rangle \sim \left(\frac{e_0}{2\pi i\, \epsilon} \right)^{\mathcal{N}} \exp \frac{ie_0}{2\epsilon} \rho[\varphi_1, \varphi_2]^2 \,, \tag{2.1}$$

where $\mathcal{N}$ is the (ultraviolet regulated) number of degrees of freedom. The function of two configurations $\rho[\varphi_1,\, \varphi_2]$ can be identified as a distance. The expression (2.1) is difficult to evaluate exactly, and it is reasonable to expect that substitution of a function $\rho'[\varphi_1, \varphi_2]$ in place of $\rho[\varphi_1, \varphi_2]$ will make no difference as $\epsilon \to 0$, provided that this choice satisfies the criterion that as $\rho[\varphi_1, \varphi_2] \to 0$

$$\frac{\rho'[\varphi_1, \varphi_2] - \rho[\varphi_1, \varphi_2]}{\rho[\varphi_1, \varphi_2]} \to 0 \,.$$

In other words, as either distance vanishes, the two distances become indistinguishable. This indistinguishability has a geometric interpretation. It guarantees uniqueness of the infinitesimal metric, which on physical grounds should have the riemannian form:

$$d\rho^2 \equiv \rho[\varphi, \varphi + \delta\varphi]^2 = \int d^D x \int d^D y \; G_{(x,A)\,(y,B)} \; \delta\varphi^A(x)\, \delta\varphi^B(y) \,,$$

where the symbols A and B include species, Lorentz indices, etc. However, we will show that the general distance function $\rho[\cdot, \cdot]$ has its uses.

Finding the distance is easy for the example of a scalar field theory. The kinetic term is

$$T = -\frac{1}{2} \int d^D x \; \frac{\delta^2}{\delta\varphi_b(x) \delta\varphi_b(x)^*} \,.$$

A distance which satisfies (2.1) is given by the Pythagorean expression

$$\rho[\varphi_1, \varphi_2]^2 = \frac{1}{2} \int d^D x \; |\varphi_{b\,1}(x) - \varphi_{b\,2}(x)|^2 \,. \tag{2.2}$$

For a gauge theory in temporal gauge, the physical degrees of freedom are not connections $A_i(x)$ (where $i = 1, \ldots, D$ is a space-coordinate index) in some Lie algebra, but *orbits*. These are equivalence classes of connections related by gauge transformations. The set whose elements are $A_i^g = g^{-1} A_i g + i g^{-1} \partial_i g$ is an orbit α. A choice of distance between two orbits α, β is

$$\rho[\alpha, \beta]^2 = \inf_{A\in\alpha\; B\in\beta} \frac{1}{2} \int d^D x \; tr\, [A_i(x) - B_i(x)]^2 \,. \tag{2.3}$$

This expression was considered by Babelon and Viallet [16] and later (and somewhat implicitly) by Feynman [15]. It was proven to be a metric in the continuum (provided care is taken as to the mathematical details) in reference [13] and a lattice analogue was discussed in reference [14]. Since (2.3) depends only on the gauge orbits α and

β and not on the specific connections in those orbits, it is gauge invariant. Given any metric space, there is a second metric function which can be defined, called the *intrinsic metric* [17]. This latter metric is the length of the shortest path between two configurations (a consequence of the triangle inequality is that the intrinsic metric is always an upper bound on the original metric). For gauge theories, the intrinsic metric coincides with the distance (2.3) [13], as had been conjectured earlier by Babelon and Viallet [16].

Once a distance satisfying (2.1) is found, other issues can be addressed. The potential energy can be thought of as a "height function" on the configuration space. We would like to use our intuition about topography, e.g. the hills, saddles and valleys of the height function on configuration space, to understand the properties of the spectrum of the quantum theory. We give some examples of the application of this intuition below.

Our first example is a one-component scalar field with a convex potential $\mathcal{V}(\varphi)$. In such a situation, the extension of the ground-state wave functional from the origin of configuration space ($\varphi = 0$) is controlled by the potential. Even if the (regularized) theory has a vanishing bare mass, a gap will appear in the spectrum. Whether the interaction terms in the potential can survive as the cut-off is removed (and whether the renormalized theory has a gap) depends on the anomalous dimensions of these terms.

For our second example, we shall give a highly nonrigorous proof of Goldstone's theorem. Suppose, for a multi-component scalar theory there are inequivalent, continuously parametrizable, degenerate minima of the potential, characterized by an expectation value $v_b = \langle \varphi_b \rangle \neq 0$. When the volume is finite, the true vacuum is nondegenerate. Let us consider the distance between two constant low-potential-energy configurations

$$\varphi_{b\,1}(x) = v_b\,, \qquad \varphi_{b\,2}(x) = v'_b\,, \qquad |v|^2 = |v'|^2\,. \tag{2.4}$$

Then (2.2) becomes

$$\rho[\varphi_1, \varphi_2]^2 = \frac{V}{2}|v_b - v'_b|^2 \leq KV\,. \tag{2.5}$$

where V is the volume of space and K is a constant. The upper bound on quantity diverges in the thermodynamic limit. In quantum mechanics, a vanishing potential energy on a domain of diameter $\sqrt{KV}$ implies a gap between the (nondegenerate) ground state and the first excited state of order $1/\sqrt{V}$. This gap vanishes and, by Lorentz invariance, the spectrum is completely continuous as $V \to \infty$.

Actually the argument of the previous paragraph is a bit too crude. There are situations in which breaking of a continuous symmetry does not occur, despite the fact that the distance between pairs of low-potential-energy configurations is unbounded in the thermodynamic limit. In particular, when $D = 1$, a state localized near a given value of φ has low potential energy, but quantum fluctuations raise the expectation value of the kinetic energy. This latter energy is lowered through the formation of domain walls (these are point-like objects in one space dimension) between regions of different values of v_b. These domain walls cost almost no potential energy, as will be shown explicitly for the one-dimensional sigma models in sections 7 and 12. Their potential energy turns out to be of order $1/L$, where L is the one-dimensional volume. Since they also lower the expectation value of the kinetic energy, they condense in the vacuum. We caution the reader that we use the term "domain wall" somewhat loosely and only for lack of a better name (more accurate but more cumbersome is "low-energy nonlinear wave"). What we are calling domain walls are not topologically stable objects, as are the domain walls of the Ising model, though we will show that they can be thought of as solitons. When the space dimension is two, a domain wall is a one-dimensional object in space. It now has a potential energy proportional to its length l, as well as inversely proportional to the diameter of two-dimensional space L (which in turn is proportional to the square root of the volume). This energy is therefore roughly $\sim l/L$. This energy will vanish in the infinite-volume limit only if the typical length l of a domain wall is not proportional to L. If it does not vanish, spontaneous breaking of a continuous symmetry is possible.

We will now summarize the situation for our final examples, the O(2) and O(3) sigma models, for the benefit of the reader who would prefer to wade through the physical ideas before plunging into the mathematics. We study the problem of minimization of the potential energy for fixed distance from a constant configuration (the classical vacuum). More precisely, we consider this problem using two different distances, one of which is the natural distance in the sense we have already discussed, as well as a second distance (Feynman [15]) which is insensitive to global O(n) transformations. For either distance we find that the potential-energy surface has grooves or *river valleys.* The type of configuration on a river valley is specified by a parameter, or set of parameters (which are the invariants of an elliptic function [18]). We now summarize how the configurations along a river valley appear as these parameters are tuned. The configuration begins as a constant or near constant. Then a weak, long-wavelength disturbance, or spin wave appears; this is nearly sinusoidal. Next the amplitude of the spin wave increases, but has nearly vanishing derivatives, except in a small region. This configuration is the domain wall mentioned in the previous paragraph. As the configuration continues to move along the river valley, the domain wall narrows to a region of the size of the short-distance cut-off. Outside the domain wall, the configuration is nearly constant. At this stage the potential energy is enormous, and is only finite by virtue of the cut-off. Remarkably, the distance between this configuration and the constant configuration is infinitesimal, although the two configurations are separated by a high, thin potential-energy barrier. At the top of the barrier, the configuration is not a domain wall at all (the value of the field on each side of this object is nearly the same!). We will call it the *barrier configuration* (this is probably similar to what Asorey and Falceto call a "sphaeleron" in the O(3) model [12]).

JHEP04(1999)002

As we have already stated, the domain walls have nearly zero potential energy and always condense in the vacuum. Whether or not barrier configurations also condense depends on the dynamics of the particular sigma model. For the O(2) model we compute the WKB tunneling amplitude through the barrier. We find that for a sufficiently large bare coupling constant, the barriers do condense. We show that tunneling through a barrier is a vortex [5]. In this way, we obtain the first analytic demonstration of the phase transition within the hamiltonian framework of this model. The O(3) model is harder to deal with. However, we are able to find the explicit form of the river valleys. They sit together in a higher-dimensional subset of configuration space, which we call a *river delta*. Surprisingly, as the potential energy begins to grow, the river valleys in the river delta coalesce. At the top of the barrier, the configuration is similar to that of the O(2) model. It is perhaps adequate to describe the barrier as "Abelian". While we do not compute the WKB amplitude of penetrating the barrier, we do calculate the topological charge of the barrier-penetration process. It is one half.

3. The sigma-model metrics

We investigate two different choices of distance for the sigma model. Both of these are metrics, in that they satisfy the axioms for a metric space. The first metric is simpler and is easier to justify physically for the sigma model. We call it the physical metric. The second, invented by Feynman [15] has the advantage of O(n) global invariance. It is defined on pairs of points of orbit space, which will be defined below. While the second metric is no more calculationally powerful than the first and is less natural from the point of view of the sigma-model spectrum, we consider it for several reasons. First, Feynman claimed that a certain property is true of the potential-energy surface of the O(3) sigma model using this metric. In fact, this is not correct and we feel it is necessary to explain some of the subtleties as to why not (this is done in section 11). Second and more important, Feynman's metric is nice from a mathematical point of view, and it is easier to visualize the potential energy on orbit space than on the space of field configurations.

In this section and the two sections which follow we discuss some general properties of the metrics of the nonlinear sigma model in D space and one time dimension. We specify the space dimension D to be one in section 6.

Let $s(t,x)$ be a unit real n-vector-valued field on some D-dimensional space manifold (not the space-time manifold, whose dimension is $D+1$), written as

$$s(t,x) = \begin{pmatrix} s_1(t,x) \\ \vdots \\ s_n(t,x) \end{pmatrix},$$

with $s(t,x)^T s(t,x) = 1$.

The action of the O(n) sigma model is

$$S = \frac{1}{2e_0} \int dt\, d^D x\, \left(\partial_t s^T \partial_t s - \nabla_x s^T \cdot \nabla_x s \right) . \tag{3.1}$$

Wherever a regularization is not explicit in our discussion, it will usually be implicit.

The hamiltonian is built from a similar unit-vector $s(x)$, which is a c-number operator, as well as the angular-momentum operator

$$L^{k_1}(x) = -i\epsilon^{k_1 k_2 \cdots k_n} s_{k_2}(x) \cdots s_{k_{n-1}}(x) \frac{\delta}{\delta s_{k_n}(x)} .$$

Explicitly, the hamiltonian is

$$H = \int d^D x \left[\frac{e_0}{2} L^T(x) L(x) + \frac{1}{2e_0} \nabla_x s(x)^T \cdot \nabla_x s(x) \right] = e_0 T + \frac{1}{e_0} U .$$

The simplest distance one can define is the physical metric

$$r[f,s]^2 = \frac{1}{2} \int d^D x\, [f(x) - s(x)]^T [f(x) - s(x)] , \tag{3.2}$$

which is the integral of the chord between two points on a sphere. It does not actually satisfy the usual properties of a metric space, unless one is willing to use a certain amount of real analysis. For example, it is necessary to identify two field configurations which are the same except on a set of measure zero (this is similar to how Hilbert-space vectors are constructed from Schrödinger wave functions in ordinary quantum mechanics). Alternatively, one can define the lattice metric over lattice points of space x:

$$r[f,s]^2 = \frac{a^D}{2} \sum_x [f(x) - s(x)]^T [f(x) - s(x)] , \tag{3.3}$$

where a is the lattice spacing. In any case, with the appropriate definition of (3.2) or with (3.3) it is easy to show the three metric properties for field configurations s, f and g: reflexivity

$$r[f,s] = r[s,f] , \tag{3.4}$$

positivity

$$r[s,f] \geq 0 , \tag{3.5}$$

with equality holding only if $s = f$, and the triangle inequality

$$r[s,f] + r[f,g] \geq r[s,g] . \tag{3.6}$$

Just as for gauge theories, one can define an orbit space for the sigma model. This is defined as a set of equivalence classes of field configurations. Two field configurations s and f are said to be *equivalent* if $f(x) = R\ s(x)$, where $R \in$ O(n)

is a global rotation. This relation is obviously an equivalence relation. Each equivalence class is an *orbit*. If the set of unit vector fields is called $\mathcal{S}$, the *orbit space* $\mathcal{M} = \mathcal{S}/\mathrm{O}(n)$ is the set of orbits

$$\psi = \{Rs(x) : R \in \mathrm{O}(n)\}\,.$$

Two sigma-model fields s and f are equivalent if and only if there exists an orbit ψ such that $s \in \psi$ and $f \in \psi$.

The use of orbit space takes some justification. In the sigma model, unlike in gauge theories, not all states are singlets under the symmetry group. These states transform under some representation of this group (though if spontaneous symmetry breaking is absent, the ground state is a singlet, transforming under the trivial representation). However, there is an obvious reason for studying the space $\mathcal{M}$. The potential energy is invariant under $\mathrm{O}(n)$ transformations. Thus as far as the spectrum of the potential-energy operator U is concerned, there is no physical difference between different elements of ψ.

To better motivate the definition of $\mathcal{M}$, we define a modified sigma model, which has a real antisymmetric $n \times n$-matrix gauge field $\mathcal{A}(t)$ depending on time, but not on space. The euclidean path integral is

$$\begin{aligned} Z = &\int \mathcal{D}s(x,t)\, \delta(s^T s - 1) \int \mathcal{D}\mathcal{A}(t) \times \\ &\times \exp - \int d^D x \int dt \left\{ \frac{1}{2e_0} \Big| [\partial_t - \mathcal{A}(t)] s(x,t) \Big|^2 + \nabla s(x,t)^2 \right\}. \end{aligned} \tag{3.7}$$

The gauge field has only a finite number (namely $n(n-1)/2$) of degrees of freedom. The path integral (3.7) is not Lorentz invariant. Nonetheless, the Green's functions are the same as those of the usual sigma model in the thermodynamic limit. For, by making a suitable gauge transformation $\mathcal{A}(t) \to q^{-1}(t)\mathcal{A}(t)q(t) - iq^{-1}(t)q(t)$ and $s(x,t) \to q^{-1}(t)s(x,t)$, where $q(t)$ is an orthogonal matrix, $\mathcal{A}$ can be set to zero for almost every time t. Thus (3.7) can be reduced to the usual path integral except at a boundary chosen at some particular value of t.

There is a lattice version of (3.7). The continuum coordinates x, t will now be replaced by lattice coordinates. These will also be written as x, t, but are integers, equal to the corresponding continuum coordinates divided by the lattice spacing a. The lattice path integral is

$$\begin{aligned} Z = &\left[\prod_{x,t} \int d^n s(x,t) \delta\left(s(x,t)^T s(x,t) - 1 \right) \right] \left[\prod_t \int dR(t) \right] \times \\ &\times \exp - \frac{a^{D-1}}{2e_0} \sum_{x,t} \Big\{ [s(x,t+1) - R(t)s(x,t)]^T [s(x,t+1) - R(t)s(x,t)] + \\ &\qquad + \sum_{i=1}^{D} [s(x+i,t) - s(x,t)]^T [s(x+i,t) - s(x,t)] \Big\}, \end{aligned} \tag{3.8}$$

where the measure of integration over the lattice gauge field R is the $\mathrm{O}(n)$ Haar measure. Let us ask how the variable $R(t)$ behaves if no gauge fixing is imposed. It is clear that in the semiclassical approximation, the configurations with $R(t)$ chosen to minimize $\sum_x [s(x,t+1)-R(t)s(x,t)]^T[s(x,t+1)-R(t)s(x,t)]$ will dominate. Suppose then that for some fixed choice of t and $s(x,t)$, $R(t)=\bar{R}(t)$ is this minimum. Note that $\bar{R}$ is unique. This is because in the continuous-time limit $\bar{R}$ is the solution of a first-order ordinary differential equation. The fluctuations around $\bar{R}$ have a coefficient proportional to the size of physical space. Consequently, these fluctuations are suppressed in the thermodynamic limit. Therefore (3.8) can be replaced by

$$
\begin{aligned}
Z = &\left[\prod_{x,t}\int d^n s(x,t)\delta\left(s(x,t)^T s(x,t)-1\right)\right]\times \\
&\times \exp -\frac{a^{D-1}}{2e_0}\sum_{x,t}\left\{\left[s(x,t+1)-\bar{R}(t)s(x,t)\right]^T\left[s(x,t+1)-\bar{R}(t)s(x,t)\right]+\right. \\
&\left.\qquad\qquad + \sum_i [s(x+i,t)-s(x,t)]^T[s(x+i,t)-s(x,t)]\right\},
\end{aligned}
$$

Thus the path integral may be regarded as the Wick rotation of quantum mechanics of a particle with mass e_0^{-1} in a space in which the distance is $\rho[\cdot,\cdot]$, defined by

$$
\rho[\phi,\psi]^2 = \frac{a^D}{2}\inf_{R\in\mathrm{O}(n)}\sum_x [Rf(x)-s(x)]^T[Rf(x)-s(x)]\,, \tag{3.9}
$$

where $f(x)\in\phi$ and $s(x)\in\psi$. The classical kinetic energy of a time-dependent orbit is the arc-length of the curve this orbit "traces out" in orbit space.

We will now show that on the lattice $\rho[\cdot,\cdot]$ is a metric on orbit space. The orbit space is therefore a metric space. There are three properties which must hold for this to be true. Obviously for any two orbits ϕ and ψ,

$$
\rho[\phi,\psi] = \rho[\psi,\phi]\,. \tag{3.10}
$$

It is straightforward to check that for these orbits

$$
\rho[\phi,\psi]\geq 0\,, \tag{3.11}
$$

with equality holding only if $\phi=\psi$. Finally, suppose that ϕ, ψ and τ are orbits. Let $f(x)$, $s(x)$ and $t(x)$ be unit vector fields on the lattice such that $f(x)\in\phi$, $s(x)\in\psi$ and $g(x)\in\gamma$. The triangle inequality (3.6) for field configurations implies

$$
\begin{aligned}
\sum_x [Rf(x)-s(x)]^T[Rf(x)-s(x)] + \sum_x [s(x)-R'g(x)]^T[s(x)-R'g(x)] \geq \\
\geq \sum_x [Rf(x)-R'g(x)]^T[Rf(x)-R'g(x)]\,,
\end{aligned}
$$

for any R, R' in $\mathrm{O}(n)$. Taking the greatest lower bound of the sum of first two terms over R and R' gives

$$\rho[\phi,\psi] + \rho[\psi,\gamma] \geq \rho[\phi,\gamma] \,. \tag{3.12}$$

Now that the three properties (3.10), (3.11) and (3.12) are proved, it follows that $\rho[\cdot,\cdot]$ is a metric and that orbit space is a metric space.

The continuum limit of (3.9), dropping factors of the lattice spacing, is

$$\rho[\phi,\psi]^2 = \inf_{R\in\mathrm{O}(n)} \frac{1}{2}\int d^Dx\, [Rf(x) - s(x)]^T[Rf(x) - s(x)] \,. \tag{3.13}$$

This expression coincides with the "minimal distance" for the sigma model Feynman attempts to estimate in the appendix to his article [15].

Strictly speaking, in the continuum neither (3.2) nor (3.13) is actually a metric, unless we are careful about configurations differing on sets of measure zero (otherwise, the inequalities (3.5) and (3.11) can be saturated by distinct configurations). If this issue is properly dealt with, the potential-energy function is infinitely discontinuous on the space of configurations (this will be explained later in this article). However, a regularization will make this discontinuity finite or can remove it altogether.

4. Evaluation of Feynman's orbit space metric

Unlike the case of the Yang-Mills metric, it is possible to evaluate (3.2) (and (3.9)) explicitly. Let us first rewrite (3.2) as

$$\rho[\phi,\psi]^2 = \inf_{R\in\mathrm{O}(n)} \int d^Dx\, [1 - s(x)^T Rf(x)] \,. \tag{4.1}$$

Define the matrix M by

$$M_{kl} = \int d^Dx f_k(x) s_l(x) \,.$$

Then

$$\rho[\phi,\psi]^2 = V - \sup_{R\in\mathrm{O}(n)} \mathrm{tr}\, RM \,,$$

where V is the volume of space. The $\mathrm{O}(n)$-covariant variation of R is $R \to \delta\gamma R$, where $\delta\gamma$ is an infinitesimal antisymmetric matrix. The condition that $tr\, RM$ be a local extremum is therefore that RM is a real symmetric matrix, i.e. $(RM)^T = RM$. Therefore, there exists an orthogonal matrix P which diagonalizes RM:

$$RM = P^T \Lambda P \,, \qquad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} .$$

Then $\mathrm{tr}\, RM = \sum_k \lambda_k$. In fact, the absolute maximum of the trace coincides with the supremum. For this case, all the diagonal entries in Λ are nonnegative. To see that this is so, assume the contrary. If λ_k is negative for a particular k between one and n, R can be replaced by another $\mathrm{O}(n)$ matrix:

$$R \to \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} R\,,$$

where the -1 appears in the k^{th} row and k^{th} column. Such a transformation must increase the value of $tr\, RM$. Thus, if R is chosen to absolutely maximize the number $tr\, RM$, all the entries of Λ are positive definite. Now $M^T M = P^T \Lambda^2 P$, and the matrix RM has only positive eigenvalues. Therefore

$$\mathrm{tr}\, RM = \mathrm{tr}\sqrt{M^T M}\,. \tag{4.2}$$

The meaning of (4.2) should be clear; the right-hand side is the sum of the square roots of the eigenvalues of the symmetric matrix $M^T M$. Since this expression is unique, it must coincide with the supremum. Thus (4.1) reduces to

$$\rho[\phi,\psi]^2 = V - \mathrm{tr}\,\sqrt{M^T M} = V - \mathrm{tr}\sqrt{\int d^D x \int d^D y\; f(x)[s(x)^T s(y)] f(y)^T}\,. \tag{4.3}$$

Having carefully derived (4.3), let us note that in retrospect it is obvious. It is the only positive-definite $\mathrm{O}(n)$-invariant expression with the correct dimensions satisfying $\rho[\psi,\psi] = 0$.

There is no essential complication in taking the analysis above to the lattice. Then (4.3) is replaced by

$$\rho[\phi,\psi]^2 = \mathcal{L} - \mathrm{tr}\sqrt{\sum_x \sum_y\; f(x)[s(x)^T s(y)] f(y)^T}\,,$$

where $\mathcal{L}$ is the total number of lattice sites. This expression is equal to (3.9).

5. The infinitesimal metrics

In this section, we discuss the riemannian form of our metrics for small separations. To do this rigorously, a lattice or alternatively Hilbert-space techniques must be used [13]. We leave it to the more mathematically oriented reader to fill in the gaps. This section is not necessary to understand the rest of the article. However, in the arguments and calculations which follow we often discuss curves and functions on

JHEP04(1999)002

configuration space. While these can be made well-defined concepts on general metric spaces [17], the reader may feel more comfortable knowing that differential-geometric intuition applies as well.

The functional riemannian metric on configuration space is directly written down from (3.2)

$$dr^2 = \frac{1}{2}\int d^Dx\ \delta s^T(x)\delta s(x)\,, \tag{5.1}$$

where δs satisfies

$$\int d^Dx\ \delta s^T(x)s(x) = 0\,. \tag{5.2}$$

There is no more work to be done for this case.

The functional riemannian metric on orbit space has the unusual feature that the metric tensor has zero eigenvalues. The corresponding zero modes are the directions along symmetry transformations. In other words, these zero modes are motions along the fibers of the fiber bundle of field configurations. The reader can find a general discussion of such metric tensors in the appendix of reference [13], though the essential idea of using a bilinear form with zero eigenvalues as an inner product was discussed much earlier [19]. Since the sigma-model orbit space is not very difficult to understand, the standard of rigor in this section is considerably less than that used for the Yang-Mills orbit space [13].

Let us find the infinitesimal form of the metric on orbits ϕ and ψ which are infinitesimally close, i.e. (4.3) is very small. Then there must exist representatives of ϕ and ψ, which we call f and s and which satisfy

$$s = f + \delta f\,.$$

Given f, we can in principle find δf by rotating s by an $\mathrm{O}(n)$ transformation minimizing $\int(f-s)^2$.

Since f and $s = f + \delta f$ are unit vectors, we can write δf in terms of an antisymmetric tensor W of rank $n-2$:

$$\delta f^i(x) = \epsilon^{ij_1j_2\cdots j_{n-1}}\ W_{j_1j_2\cdots j_{n-2}}(x)\ f^{j_{n-1}}(x)\,.$$

Global $\mathrm{O}(n)$ transformations can be parametrized as $R = e^h$, where h is a constant antisymmetric matrix, parametrized by a constant tensor b of rank $n-2$ as

$$h^{ik} = \epsilon^{ij_1j_2\cdots j_{n-2}k}\ b_{j_1j_2\cdots j_{n-2}}\,.$$

The metric evaluated on ϕ and ψ is

$$\begin{aligned} d\rho^2 = \rho[\phi,\psi]^2 &= \min_b \frac{1}{2}\int\ d^Dx \left\{\epsilon^{ij_1j_2\cdots j_{n-2}k}[W_{j_1j_2\cdots j_{n-2}}(x) - b_{j_1j_2\cdots j_{n-2}}]f(x)_k\right\}^2 \\ &= \min_b \frac{1}{2}\int\ d^Dx \left[W_{[j_1j_2\cdots j_{n-2}]}(x) - b^{[j_1j_2\cdots j_{n-2}]}\right]^2, \end{aligned} \tag{5.3}$$

where square brackets around indices denote antisymmetrization.

Carrying out the minimization with respect to b in (5.3) is straightforward. The result for the O(2) sigma model is

$$d\rho^2 = \int d^Dx \int d^Dy \; G_{\alpha(x)\alpha(y)} \; \delta\alpha(x) \; \delta\alpha(y) \,,$$

where

$$G_{\alpha(x)\alpha(y)} = \frac{1}{2}\delta^D(x-y) - \frac{1}{2V} \,, \tag{5.4}$$

and the angle $\alpha(x)$ is defined by $f_1(x) = \sin\alpha(x)$ and $f_2(x) = \cos\alpha(x)$. Notice that this functional metric tensor G has no dependence on α, and is therefore flat. The zero modes are constant rotations of $\alpha(x)$. For the O(3) model

$$d\rho^2 = \int d^Dx \int d^Dy \; G_{s^i(x)s^j(y)} \delta f^i(x) \delta f^j(y) \,,$$

where

$$G_{s^i(x)s^j(y)} = \frac{1}{2}\delta^D(x-y)\delta_{ij} - \frac{1}{2V}\epsilon_{ikl} f^k(x) (B^{-1})^{lm} \epsilon_{jmr} f^r(y) \,, \tag{5.5}$$

where the three-by-three matrix B is

$$B^{kl} = \delta^{kl} - \frac{1}{V}\int d^Dz \; f^k(z) \; f^l(z) \,.$$

It can be checked that the zero modes of the metric tensor are O(3) rotations. The metric tensor considered as a matrix in function space is idempotent, $G^2 = G$. A general treatment of how the laplacean and the curvature may be determined is given in the appendix to reference [13]. One can go further in the case of the O(3) sigma model and write the metric tensor in angular coordinates instead of unit-vector coordinates, as was done above for the O(2) case.

6. The potential-energy surface and the classical pendulum

Let us denote the "pure gauge" or "classical vacuum" orbit containing constant $s_0(x) = s_0$ by ψ_0. Consider now the following problem for $D = 1$ with $V = L$, the length of one-dimensional space. For fixed $\rho[\psi, \psi_0]$, extremize the potential energy

$$U[\psi] = \int_0^L \; dx \; \left[\frac{ds(x)}{dx}\right]^2 \,, \qquad s \in \psi \,.$$

Let us parametrize $s(x)$ using angles $\xi_1(x)$,..., $\xi_{n-1}(x)$, by

$$s(x) = \begin{pmatrix} \sin\xi_1(x) \cdots \sin\xi_{n-1}(x) \\ \sin\xi_1(x) \cdots \cos\xi_{n-1}(x) \\ \vdots \\ \cos\xi_1(x) \end{pmatrix} \,,$$

in the standard way. For O(2) we will label ξ_1 as α, and for O(3) we will label ξ_1 and ξ_2 as ξ and κ, respectively. The problem will be considered with the periodic boundary condition $s(x) = s(x + L)$, though more general boundary conditions will also be discussed for the O(3) case (in reference [14] Neumann boundary conditions were considered).

If $s(x)$ is any representative of ψ, then

$$\rho[\psi, \psi_0]^2 = L - \left| \int_0^L s(x)dx \right| , \tag{6.1}$$

from (4.3). The problem is therefore equivalent to the extremization of the functional

$$A[s] = \frac{1}{2} \int_0^L dx \, \frac{ds(x)^T}{dx} \frac{ds(x)}{dx} + \lambda \left[\left| \int_0^L s(x)dx \right| - v \right] , \tag{6.2}$$

where $v = L - \rho[\psi, \psi_0]^2$ and λ is a Lagrange multiplier. Suppose that a solution has been found. Then by rotating this solution, we can make $\int_0^L s^j(x) = 0$, for $j = 1, \ldots, n-1$, and $\int_0^L s^n(x) = v$. The solution for the extrema of $U[\psi]$ for fixed $\rho = \rho[\psi, \psi_0]$ (this defines a sphere whose center is the point ψ_0) is then described by the motion of an n-dimensional pendulum, which can be written in terms of elliptic functions and integrals.

We can choose not to work on orbit space but instead asked the same question on configuration space. In that case we need to extremize the potential energy, while holding fixed the physical metric distance from a constant configuration s_0:

$$r^2 = \frac{1}{2} \int_0^L \, [s(x) - s_0]^T [s(x) - s_0] = - \int_0^L \, s_0^T s(x) + L \, .$$

Such a constraint included with a Lagrange multiplier is once again the n-dimensional pendulum equation. We therefore see that the solutions we will obtain for either problem are the same. A solution of the problem in orbit space is a set of solutions in configuration space; the latter solutions are all the same up to global rotations.

An n-dimensional pendulum is a massive particle on the sphere S_{n-1} embedded in $\mathbb{R}^n$ under the influence of a constant gravitational force. The coordinates are s_i, $i = 1, \ldots, n$ with $\sum_i s_i^2 = 1$. For simplicity, this force can be taken to be along the n-axis.

The significance of elliptic functions [18] in the n-dimensional pendulum problem can be understood physically. There is one variable describing the "height" of the particle, namely the n^{th} coordinate, s_n. We will explain in a moment why the evolution of s_n is periodic, i.e. $s_n(t + T) = s_n(t)$, where T is the period. The behavior of other variables describing the pendulum (if $n > 2$) is not, in general, periodic. However, the evolution of these variables must satisfy certain conservation laws, namely the conservation of $(n-1)(n-2)/2$ of the $n(n-1)/2$ components of the angular-momentum tensor. The conserved components are $l_{ij} = s_i p_j - p_i s_j$ where i

and j are less than n. The other variables influence s_n only through these conserved components. In other words the equation of motion of s_n can be written so that l_{ij} can be substituted for the other degrees of freedom. Any solution of this equation is obviously periodic.

Now imagine turning the pendulum upside down; this means reversing the direction of the gravitational field. The evolution of s_n is again periodic with a new period T'. A field reversal is equivalent to the Wick rotation $t \to i\,t$. The equation of motion contains one term with a second time-derivative. Reversing the direction of the field is equivalent to changing the sign of this second time-derivative. This implies that $s_n(t)$ in the original pendulum (that is, before it was turned upside down) has an *imaginary period* iT' as well as the real period T. If this function has no singularities in the complex t-plane other than poles, it must be an elliptic function.

7. River Valleys of the O(2) model and solitons

Next we shall investigate the potential-energy surface of the O(2) nonlinear sigma model or classical XY model. More precisely, we shall examine the pendulum solutions which are the extrema of (6.2). Depending upon which is easier, sometimes we will discuss the extrema in orbit space and sometimes we will discuss the extrema in configuration space. As we pointed out in the previous section, the latter are contained in the former.

The planar pendulum displays two types of motion, namely oscillating and circulating. The lowest-energy configuration is that for which the pendulum sits at the nadir for all time. As the energy is increased, the pendulum makes small harmonic oscillations around the nadir, which can be described by trigonometric functions. Increasing the energy further leads to an increase in the period, which is an elliptic integral, eventually leading to a situation in which the pendulum will spend most of the time near the zenith, but on occasion will rapidly sweep out an angle of nearly 2π. A slight further increase in energy causes the pendulum to pass through the zenith, leading to circulating, rather than oscillating motion. The pendulum still spends most of its time near the zenith, but the motion is now consistently clockwise or counter-clockwise. As the energy continues to increase, the angular frequency of this circulating motion becomes a constant. Once more the motion can be described by trigonometric functions.

The mathematical translation of the discussion in the previous paragraph is the following. For the pendulum lagrangean

$$L = \frac{1}{2}\dot{\alpha}^2 - \mu(1 - \cos\alpha)\,,$$

the classical energy is given by

$$E_{pend} = \frac{1}{2}\dot{\alpha}^2 + \mu(1 - \cos\alpha)\,.$$

There are oscillating solutions to the equations of motion

$$\alpha(t) = 2\sin^{-1} k\, \mathrm{sn}\,(k\sqrt{\mu}t, k)\,,$$

where $\mathrm{sn}\,(u, k)$, sometimes written $\mathrm{sn}\,u$, is the Jacobi elliptic-sine function, and $E_{pend} = 2\mu k^2$. The modulus k is between zero and one. Increasing the energy means increasing k. The period is $4K$, where K is the complete elliptic integral $K(k) = \mathrm{sn}^{-1}1$. This diverges logarithmically as k approaches one (asymptotically, $K \simeq \log 4/k'$, where $(k')^2 = 1 - k^2$). At $k = 1$, these solutions are joined onto the $k = 1$ circulating solutions

$$\alpha(t) = 2\sin^{-1} \left|\mathrm{sn}(k^{-1}\sqrt{\mu}\, t, k)\right|,$$

where $E_{pend} = 2\mu/k^2$. The reason for the absolute value in this expression is that there are no turning points and the left-hand side is discontinuous with a discontinuity of 2π. As k is then decreased the energy increases further, the angular velocity eventually becoming a constant.

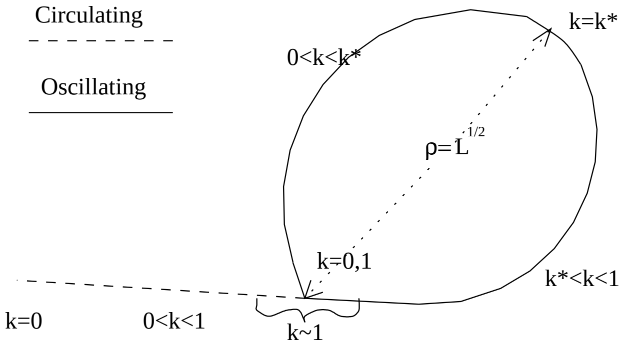

Figure 1: Two typical extremal curves for the O(2) sigma model. In the thermodynamic limit, the potential energy is vanishing along the solid (oscillating) and dashed (circulating) curves where k is not close to one. The potential energy is ultraviolet divergent in the region where k is close to one.

Now let us to turn to the extrema of (6.2), where time t is replaced by space x (up to a constant) and kinetic energy (up to another constant) is now the sigma-model potential-energy density. We take periodic boundary conditions. Up to global rotations R there are two types of solution labeled by an integer $N = 1, 2, \ldots$

$$\begin{aligned}\alpha(x) &= \pm\alpha_N^{osc}(x, k, x_0) \\ &= \pm 2\sin^{-1} k\mathrm{sn}\left[\frac{4NK}{L}(x - x_0)\right],\end{aligned} \tag{7.1}$$

which resemble the oscillating solutions of the pendulum and

$$\alpha(x) = \pm\alpha_N^{circ}(x, k, x_0) = \pm 2\sin^{-1}\left|\mathrm{sn}\left[\frac{2NK}{L}(x - x_0)\right]\right|, \tag{7.2}$$

which resemble the circulating solutions of the pendulum. These extremal curves in orbit space are nicely parametrized by the modulus k as shown in figure 1. Note: not all of these curves are the river valleys mentioned in section two.

We will next investigate the properties of each branch of solutions. First let us consider the oscillating branch (7.2). We denote the particular value of k where $E = 2K$ by k^*, where

$$E = E(k) = \int_0^1 \mathrm{dn}^2\, u\, du \tag{7.3}$$

is another standard elliptic integral, not to be confused with the energy E_{pend}. The number k^* has the numerical value $k^* \approx 0.82$. For $0 \le k \le k^*$ the distance from ψ_0 is independent of N and has the form

$$\rho^{osc}(k)^2 = L - \left| \int_0^L \cos \alpha_N^{osc} dx \right| = 2L \left(1 - \frac{E}{K} \right),$$

while for $k^* \le k \le 1$

$$\rho^{osc}(k)^2 = L - \left| \int_0^L \cos \alpha_N^{osc} dx \right| = 2L \frac{E}{K}. \tag{7.4}$$

The function $\rho(k)$ rises smoothly from 0 to L as k goes from 0 to k^*, then falls off to zero again as $k \to 1$. Notice that at $k = 1$ the orbit is identical to that at $k = 0$. An orbit along an extremal curve is maximally far from the origin at $k = k^*$. The potential-energy functional $U = \frac{1}{2} \int_0^L (\frac{d\alpha}{dx})^2 dx$ for the oscillating branch (7.2) is

$$U^{osc}(k) = \frac{32 N^2 K}{L} [E - k'^2 K],$$

where $k'^2 = 1 - k^2$. For fixed volume L, $U^{osc}(k)$ diverges at $k = 1$, but, as mentioned earlier, this divergence is regularized by a lattice (or some other ultraviolet cut-off).

Physically the oscillating branch of solutions (7.2) is a spin wave of wavelength L/N. As $k \to 0$, the spin wave dies off in amplitude, approaching ψ_0. As $k \to 1$, the spin wave begins to *wind*. If the angle α is represented on a circle, then the configuration (up to global rotations) is a curve on a cylinder. This is shown in figure 2*abc*. However the $k = 1$ limit is not a simple kink of winding number N. For not only is the curve in figure 2*c* beginning to wind, but the amplitude is nearly a constant for all x. In other words, the spin wave has become a domain wall of the type discussed in section 2. The potential energy of the domain walls, like those of the spin waves is very small. The winding takes place in narrow regions of space. As k approaches one, the width of these regions collapses to zero. Thus the $k = 1$ solution coincides with the $k = 0$ solution. Since the potential energy diverges as k approaches one, we can see that there is an *infinite discontinuity* in the potential-energy function on orbit space. This discontinuity can be removed or made finite with an ultraviolet cut-off.

For the circulating branch the distance from ψ_0 is given by

$$\rho^{circ}(k)^2 = L - \left| \int_0^L \cos \alpha_N^{circ} dx \right| = \frac{2L}{k^2} \left(\frac{E}{K} - k'^2 \right),$$

while the potential energy is

$$U^{circ}(k) = \frac{8 N^2 K E}{L}.$$

Again there is an ultraviolet divergence at $k = 1$.

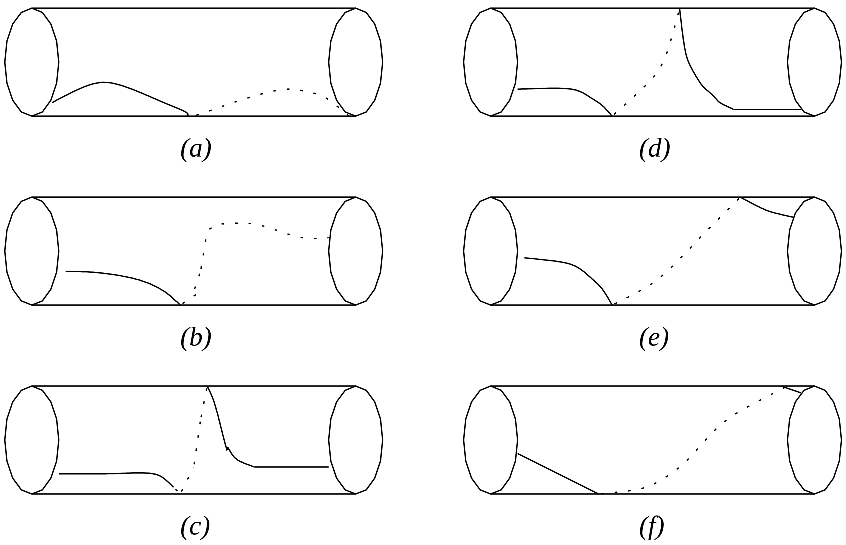

Figure 2: How extremal-curve configurations depend on k. Here we show a path in configuration space, starting from ψ_0, moving away as an oscillating configuration, reaching ψ_0 again, then moving away as a circulating configuration. In each case, the horizontal axis of the cylinder is x, while the angle is α. For the oscillating solutions, small k implies a small amplitude spin wave (a), and as k increases the potential energy becomes more localized into a domain wall (b). When k is close to one, short kinks begin to form (c), while away from these kinks, the configuration is nearly constant. For the circulating solutions with k close to one (d), the kinks are fully formed and the winding number is N (the cylinder in this figure is too short to show the remaining kinks). Decreasing k does not change the number of windings, but the kinks begin to spread out into domain walls (e). As k decreases further, the derivatives become of order $1/L$ for all x. If $k \approx 1$ ((c) and (d)) the potential energy becomes divergent; otherwise it vanishes as $L \to \infty$.

The circulating branch of solutions wind around the cylinder (as shown in figure $2def$) N times. In other words, these are configurations with N kinks. As $k \to 1$, the potential energy diverges, but the orbit approaches ψ_0. This is because the regions where the kinks occur will narrow until they disappear. As k decreases, the potential energy decreases as well, but $\rho^{circ}(k)$ becomes of order $\sqrt{L}$. What happens is that the derivative of α becomes of order N/L in this limit.

Not all of the extremal configurations we have constructed are minima of the potential energy on a sphere in orbit space. This issue is examined in detail in Appendix A. We find that all the circulating solutions (7.2) are indeed local minima on the orbit-space sphere, for any value of k. Consequently, these are river valleys (note: in reference [14] the term "river valley" was used for all the extremal curves). We also show that the $N = 1$ oscillating solution (7.2) is a local minimum on the orbit-space sphere for sufficiently small k.

We will now interpret our results thus far. We have found that the potential energy on orbit space $U[\psi]$ has one-dimensional valleys. In the thermodynamic limit

$L \to \infty$, the bottoms of these valleys are flat (the potential energy vanishes there) except for the special region where k approaches unity. The $k = 1$ point is actually ψ_0. The energy of a point in a river valley diverges as ψ_0 is approached. Therefore the potential energy at ψ_0 is discontinuous. If a regularization is introduced, the energy vanishes at ψ_0, but for $k = 1 - \epsilon$, where ϵ is determined by the ultraviolet cut-off, the river-valley potential energy rapidly rises to a large number. Even in this region, the energy must be a local minimum in the direction perpendicular to the river valley.

For large volume, the potential energy is almost constant nearly everywhere in a river valley. The one-dimensional domain where this is so has length $\mathrm{O}(\sqrt{L})$. If we naively view k as a collective variable, and ignore fluctuations in other degrees of freedom, the gap is of order $\mathrm{O}(1/L)$.

We note that the extremal curves are not straight lines in configuration space. Their tangent vectors at k for the oscillating solution (7.2) are

$$\beta^{osc}(x,k) = \frac{\partial \alpha_1^{osc}(x,k)}{\partial k} = \frac{2\mathrm{sn}\,u\;\mathrm{dn}\,u - Z(u)\mathrm{cn}\,u}{1-k^2}\,, \tag{7.5}$$

where $u = 4K(x - x_0)$ and $Z(u)$ is the Jacobi zeta function. The inner product of β_N and its derivative with respect to k is not zero, which means that the extremal curves have curvature. One can define the unit tangent vector $\hat{\beta}(x,k) = \beta(x,k)/\sqrt{\int_0^L \beta(y,k)^2\,dy}$.

All extremal-curve configurations, including those of river valleys, are related to solitons in a finite volume. Consider adding to the O(2) sigma model action (3.1) an external source h (from the viewpoint of lattice spin systems, this is an external magnetic field), modifying it to

$$S = \frac{1}{2e_0}\int dt\,dx\,(\partial_t s^T \partial_t s - \partial_x s^T \partial_x s + h^T s)\,. \tag{7.6}$$

This is the sine-Gordon action, with the different classical vacuua $\alpha(x) = 0, \pm 2\pi, \ldots$ identified (in this respect it is closer to the textbook example of the twisting band than the usual sine-Gordon action). Without loss of generality, the direction of h can be chosen so that (7.6) reduces to

$$S = \frac{1}{2e_0}\int dt\,dx\,[(\partial_t\alpha)^2 - (\partial_x\alpha)^2 + h\,\cos\alpha]\,,$$

and the equation of motion is the sine-Gordon equation, in the compact field α:

$$(\partial_t^2 - \partial_x^2)\alpha = -h\sin\alpha\,. \tag{7.7}$$

Substituting $\alpha(t,x) = \alpha(u)$, where $u = t \pm x/v$ into (7.7) yields

$$\partial_u^2\alpha(u) = -\frac{v^2 h}{(v^2-1)}\sin\alpha\,,$$

which is the equation of motion of the planar pendulum. The solutions to this equation are a train of evenly-spaced solitons, of the form

$$\alpha = 2\sin^{-1} k\,\mathrm{sn}\left(\frac{4Ku}{l}, k\right), \tag{7.8}$$

or

$$\alpha = 2\sin^{-1} \left|\mathrm{sn}\left(\frac{2Ku}{l}, k\right)\right|, \tag{7.9}$$

where the modulus k is a free parameter between zero and one and the spacing between solitons is given by

$$l = \sqrt{\frac{32K^2(v^2-1)}{v^2 h}}\,.$$

The solitons are located at the nodes $u = 0, \pm l, \pm 2l, \ldots$ The $k \to 1$ limit of either (7.8) or (7.9) reduces to the standard one-soliton solution.

Unlike the solutions (7.8) and (7.9), the extremal-curve configurations (7.2) and (7.2) have no explicit time dependence. However, they do contain the translation parameter x_0. Thus the soliton solutions of (7.6) really can be thought of as extremal-curve configurations with moving x_0. Conversely, a river-valley configuration is just a "snapshot" of a soliton configuration (7.9) at a given time. The solutions (7.2) contain $2N$ solitons (actually they are solitons alternating with anti-solitons) and (7.2) contain N solitons.

The reason we emphasize the mathematical resemblance between extremal-curve configurations and solitons is to give a physical picture of quantum barrier penetration, which is discussed in the next section. The river-valley approaches the trivial configuration ψ_0 as k tends to one. Thus a soliton configuration (7.9), with diverging potential energy can be made arbitrarily close to ψ_0 (meaning that the metric separation in configuration space between the orbit containing the soliton configuration and ψ_0 is small).

8. Barrier penetration and vortices

We have gone to a lot of trouble to investigate the extremal curves, and if the reader has been patient enough to follow our discussion thus far, we will now explain their physical significance.

An orbit in an extremal curve has very small potential energy, unless it approaches ψ_0 from particular directions, corresponding to k close to one. For small k *superpositions* of extremal-curve configurations,

$$\alpha = a_1\alpha_1 + a_2\alpha_2 + \cdots\,, \tag{8.1}$$

where the real constants $a_1, a_2, \ldots$ are not too large, are also configurations of small potential energy. There is nothing surprising in this fact. For small k, Jacobi elliptic functions become trigonometric functions and (8.1) is a general spin-wave configuration. For small amplitude waves, the sigma model is approximately a free field theory. However, when k is large, the extremal-curve configurations cannot be superposed as in (8.1), because their nonlinear character is important.

We now make the following

Assertion In the Schrödinger representation of the XY model, the vacuum wave functional can be significantly different from zero for only two kinds of field configurations. These are:

1. Spin waves, i.e. configurations of the form (8.1).

2. River-valley configurations with k very close to one, as well as slight deformations of such configurations.

By "slight deformations" we mean two things:

- First, small-amplitude spin waves can be added to large k river-valley configurations without significantly changing the potential energy.

- Second, river-valley configurations are very different from zero only in small regions of space where kinks begin to appear (we have shown in the last section that these kinks can be thought of as solitons). There is no significant gain in either the potential energy or in the distance from ψ_0 if the position of these regions is changed.

To see why the assertion must be true, let us first examine each case 1. and 2. individually. Clearly configurations of type 1. which have small potential energy must exist as fluctuations around ψ_0. There is nothing to suppress such fluctuations. Why should configurations of type 2. be important? They should be highly suppressed by virtue of their huge potential energy. However, they may not be completely suppressed. The reason is just that the orbits containing these configurations are close to ψ_0 in orbit space. Even if quantum fluctuations from ψ_0 to one of these orbits may be small, there is the possibility that they do not disappear entirely. Only the configurations 1. and 2. are either of small potential energy or close to ψ_0. Therefore the vacuum wave functional should be vanishly small when evaluated for any configuration other than 1. and 2. .

Whether or not configurations of type 2. indeed appear through quantum fluctuations depends upon the barrier-penetration amplitude. It is the computation of this amplitude we examine in the next section. However, first we will discuss further the physical interpretation of the barrier penetration. The energy barriers are more than just the $k \approx 1$ region of river valleys (7.2). Consider the point of intersection ψ

of a sphere whose center is ψ_0 and a river valley. Let G be a small neighborhood of ψ in the sphere (see figure 3). If this neighborhood G is sufficiently small, the point of intersection ψ is actually a *minimum* of potential energy in G. The $k \approx 1$ part of a river valley is a path of least resistance through the barrier.

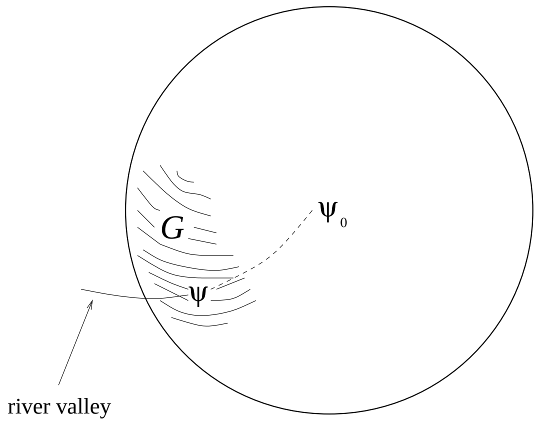

Figure 3: River valleys for $k \approx 1$ are paths of lowest potential energy through the ultraviolet divergent energy barriers. The potential at ψ is a minimum in the neighborhood G in the sphere of constant metric distance from ψ_0.

A typical tunneling event can be visualized with the aid of either figures $2abc$ or $2def$. A constant configuration makes a transition to the configuration $2c$ ($2d$), which contains short localized regions of large potential energy. These stretch out in physical space, meaning that the configuration evolves first to $2b$ ($2e$), then $2a$ ($2f$) and finally becomes nearly asymptotically constant (winding with minimal derivatives with respect to x).

The amount of time required to make the initial transition to 2c (2d) depends on the coupling e_0 and the nature of the cut-off. In a sensible regularization, the potential energy in the small region $1 > k > 1^*$ can be set to zero, where 1^* is a number very close to one but strictly less than one. The energy barrier is thereby rendered finite. If the initial state is a delta function localized in ψ_0, i.e. $k = 1$, the orbit will lie in a spreading wave packet, which will eventually reach 1^*. Now e_0^{-1} plays the role of mass in this quantum-mechanical system. The form of this wave packet is therefore, for short times,

$$\Psi(k,t) = \sqrt{\frac{1}{2\pi i\, t e_0^{-1}}}\, \exp \frac{i\,\rho(k)^2}{2t\, e_0^{-1}} \,. \qquad (8.2)$$

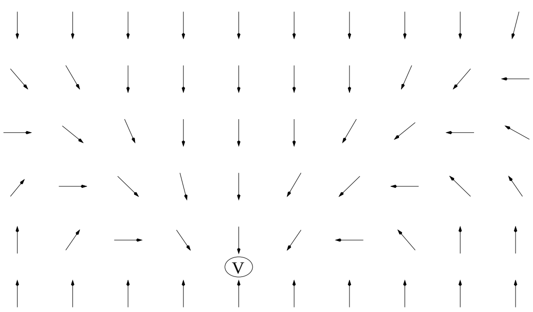

Figure 4: The tunneling process at weak coupling. Time is represented by the vertical direction, and a lattice is used for ease of visualization. The initial configuration is constant. The vortex (located at the letter V) appears suddenly. Then the configuration slowly decays back to a constant. As the coupling increases, the number of lattice spacings between the initial constant configuration and the vortex also increases.

The separation of these two points in orbit space is $\rho(1^*)$. Therefore, the typical time an orbit takes to travel from $k = 1$ to $k = 1^*$ is proportional to $e_0 \rho(1^*)^2$. For small e_0 the transition is very sudden, and the tunneling process (shown on a lattice) resembles that of figure 4. At stronger coupling, this transition time is large, but still finite.

The tunneling process is a *vortex* [5]. The closed line integral of the gradient of α enclosing the "core" of the vortex $\int \nabla\theta \cdot d\mathbf{l}$, where the topology of the one-dimensional configuration changes, is 2π. From fig. 4, the reader can see that our vortex is asymmetrical, unlike the Berezinskii-Kosterlitz-Thouless vortices, which are nearly rotational invariant (figure 5). In the tunneling process corresponding to the latter, the topology change does not happen suddenly, even for small e_0. The two types of vortices do not resemble each other very much, except at infinitely-strong coupling. Actually, euclidean field configurations which dominate the lattice partition function are not isolated vortices, but a superposition of vortices along with spin waves. Most of these resemble neither our asymmetrical vortex of figure 4 nor the more standard variety of figure 5.

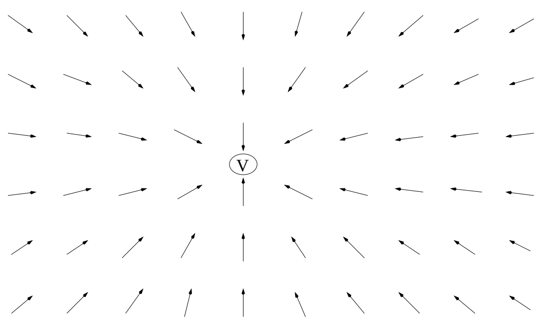

Figure 5: The Berezinskii-Kosterlitz-Thouless vortex. The picture of the tunneling process is similar, except that it is nearly rotationally invariant. The vortex appears halfway between constant configurations during a long time evolution.

9. The tunneling amplitude in the O(2) model

The ultraviolet divergence in the barrier height is due to the fact that the derivative $\partial\alpha(x,k)/\partial x$ diverges as $k \to 1$. Any sensible regularization imposes the restriction

$$\left|\frac{\partial\alpha(x,k)}{\partial x}\right| \leq \frac{2Q}{a}\,, \tag{9.1}$$

where a is the short-distance cut-off and Q is a constant depending on the details of the regularization. For example, on a lattice, a is the lattice spacing and the difference in α at adjacent lattice sites is at most π, so that $Q = \pi/2$.

For the circulating river-valley solution with one soliton, that is (7.2) with $N = 1$, the condition (9.1) implies that for each x

$$\frac{2K}{L}\left|\,\mathrm{dn}\left[\frac{2K}{L}(x - x_0)\right]\right| \leq \frac{2Q}{a}\,,$$

or

$$K(k) \leq \frac{QL}{a} \equiv K(1^*)\,.$$

The path in orbit space of the tunneling is along the river valley from $k = 1^*$ to some value of $k = \tilde{k}$ less than 1^*, for fixed x_0. Recall that the role of the mass is played by e_0^{-1}. The WKB formula for the tunneling amplitude for fixed x_0 is

$$\mathcal{T} = \exp -W\,, \qquad W = \int_{\tilde{k}}^{1^*} dk \sqrt{\int_0^L \beta(x,k)^2 dx} \sqrt{\frac{2}{e_0}\left[\frac{U(k)}{e_0} - \frac{U(\tilde{k})}{e_0}\right]}\,, \tag{9.2}$$

where β is the tangent vector to the river valley defined in (7.5). The square root of the length of the tangent vector is included in the integrand so that the integration measure is the orbit-space line element.

A useful approximation is the replacement

$$dk\sqrt{\int_0^L \beta(x,k)^2 dx} \quad \longrightarrow \quad d\rho(k) = \frac{d\rho(k)}{dk} dk$$

in (9.2). This approximation is valid because what remains in the integrand is small unless $\rho(k)$ is small. In this regime, $\rho(k)$ is given by (7.4).

It is convenient to use the formulas for the derivatives of the complete elliptic integrals of the first and second kind

$$dE = \frac{E-K}{2k^2} d(k^2)\,, \qquad dK = \frac{E - k'^{\,2} K}{2k^2 k'^{\,2}} d(k^2)\,,$$

the approximations valid when $k \approx 1$ that

$$E \approx 1\,, \qquad K \approx \log \frac{4}{k'}\,,$$

and to make a change of variable from k to K. We also assume that in the dominant part of the range of integration $U(\tilde{k})$ can be neglected. Then (9.2) becomes

$$W \approx \int_{K(\tilde{k})}^{\frac{QL}{a}} \frac{2\sqrt{2}}{e_0 K} dK \approx \frac{2\sqrt{2}}{e_0} \log \frac{QL}{a}\,. \tag{9.3}$$

Notice that (9.3) implies that the WKB tunneling amplitude at any given point vanishes in the thermodynamic limit. The alert reader may have noticed that W is similar to the usual expression for the vortex action [5].

10. The phase transition in the O(2) model

We will now explain the Kosterlitz-Thouless transition using our hamiltonian quantum-mechanical methods.

Thus far we have considered a change in the winding around the cylinder at a specific location. To determine the total transition amplitude, it is necessary to sum over all the possibilities of this location. On a lattice, the number of such

JHEP04(1999)002

possibilities is L/a. With a different regularization, this number should be L divided by the short-distance cut-off and multiplied by some (non-universal) constant Y. In any case

$$\frac{L}{a}\mathcal{T} = Y\, Q^{-\frac{2\sqrt{2}}{e_0}} \left(\frac{L}{a}\right)^{1-\frac{2\sqrt{2}}{e_0}} . \tag{10.1}$$

In the thermodynamic limit $L \to \infty$ this amplitude vanishes unless e_0 is greater than its critical value $2\sqrt{2}$ (as pointed out earlier, this number is not universal). For small e_0, tunnelings are rare and the vacuum wave functional is gaussian. In this weak-coupling phase, winding modes or solitons are stable particles. For sufficiently large e_0 the solitons condense in the vacuum, disordering correlation functions. The quantity in the exponential of (10.1) is really not very different from Kosterlitz and Thouless' famous mean-field free-energy estimate. Having said this, we started from a different point of view, which will later prove useful for the O(3) sigma model.

Many years ago, soliton condensation in the Schrödinger picture was argued to be responsible for the phase transition of the hamiltonian XY model [20]. Our calculation shows that this is the case. Our solitons (7.8), (7.9) are *not* classical solutions of the XY model, but the associated model (7.6). We will later argue that soliton condensation takes place in the O(3) sigma model for any coupling (at topological angle $\theta = 0$). Once again, these are not solitons of the classical O(3) sigma model, but are solutions of a related classical field theory.

The most obvious discrepancy between our result (9.3) and the energy-entropy argument of Kosterlitz and Thouless is that we find a critical coupling of $e_0 = 2\sqrt{2} = 2.83$ instead of $e_0 = \pi/2 = 1.57$. This need not concern us, for this number is not universal. A recent Monte-Carlo computation [21] for the square-lattice Villain model indicates a critical coupling of $e_0 = 1.33$. The numerical result certainly disagrees with both analytical estimates.

11. River deltas and solitons in the O(3) sigma-model

Next let us turn to the O(3) case. The extrema of (6.2) are related to solutions of the spherical pendulum (see for example Whittaker [22] and appendix B in this article).

The main new ingredient in the spherical pendulum is the presence of a new conserved quantity. This quantity is the angular momentum about the vertical axis. It is not possible for the pendulum to reach the zenith or nadir on the sphere unless the angular momentum is zero. This is easy to understand physically. If the pendulum is placed at either point, the angular momentum about the vertical axis is zero unless the velocity diverges; but if the velocity does diverge, the total energy would also be divergent. Therefore there is a minimum distance from the zenith and another from the nadir to the vertical axis. We have just proved that the motion always lies between two circles on the sphere, each circle perpendicular to the vertical axis. Since we already know that the vertical component of velocity is periodic in time (in fact an elliptic function) it follows that the pendulum actually touches the circles alternately and that the time to go from one circle to the other is half the period. It can be proved that the azimuthal angle always advances each half-period by $\Delta\kappa$, which satisfies the Halphen inequality $|\Delta\kappa| \leq \pi$ and the Puiseux inequality $|\Delta\kappa| \geq \pi/2$ (both of these inequalities can be proved using elementary complex analysis [23]). An illustration of the pendulum motion is given in figure 6.

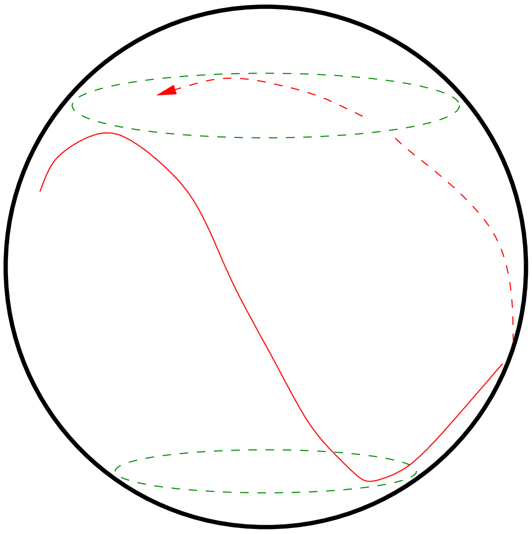

Figure 6: The path of the spherical pendulum (shown in red). The motion is restricted between two circles (shown in green). The advance in azimuthal angle as the pendulum moves from one circle to the other is between $\pi/2$ and π. This also depicts an extremum of (6.2) for the O(3) sigma model, where the role of time is played by the position coordinate x.

The configuration which extremizes the functional (6.2) can be written immediately once the spherical pendulum solution is known. It is most compactly written in terms of Weierstrass functions (our conventions are those of Whittaker and Watson [18]) $\mathcal{P}(z) = \mathcal{P}(z; g_2, g_3)$ and their integrals. The period $2\omega_1$ is purely real, while the period $2\omega_3$ is purely imaginary. Explicitly, the analysis of appendix B gives

$$\xi(x) = \cos^{-1}\left[-\frac{\mathcal{E}}{3} - \mathcal{P}(y+\omega_3)\right], \tag{11.1}$$

$$\begin{aligned} e^{i\,\kappa(x)} &= \exp\left\{(\eta_1+\eta_2)(\nu_+ - \nu_-) - [\zeta(\nu_+) - \zeta(\nu_-)]y\right\} \times \\ &\quad \times \sqrt{\frac{\sigma(y+\omega_3+\nu_+)\,\sigma(y+\omega_3-\nu_-)}{\sigma(y+\omega_3-\nu_+)\,\sigma(y+\omega_3+\nu_-)}}\,, \end{aligned} \tag{11.2}$$

where

$$y = \frac{2N\omega_1}{L}(x-x_0)\,.$$

There are two independent real parameters in this solution, because the constants in (11.1), (11.2) are related by

$$\begin{aligned} \frac{g_2}{4} &= 1 + \frac{\mathcal{E}^2}{3}\,, \qquad \frac{g_3}{4} = \frac{2\mathcal{E}^3}{27} - \frac{2\mathcal{E}}{3} + b\,, \\ \nu_+ &= \omega_1 + i\beta\,, \qquad \nu_- = i\gamma\,, \\ \mathcal{P}(\nu_\pm) &= -\frac{\mathcal{E}}{3} \pm 1\,, \qquad \mathcal{P}'(\nu_\pm) = -2i\sqrt{b}\,, \end{aligned} \tag{11.3}$$

g_2 and g_3 are related to ω_1 and ω_3 in the standard way and β and γ are real positive numbers, which are less than $-2i\omega_3$. This is the most general solution, but it usually does not satisfy the periodic boundary condition. The expression (11.1) is periodic in ξ, that is $\xi(L) = \xi(0)$, but in general (11.2) is not periodic in κ.

We note that the invariants g_2 and g_3 are not completely arbitrary, because of the inequalities

$$e_1 > 1 - \frac{\mathcal{E}}{3}, \qquad -1 - \frac{\mathcal{E}}{3} \leq e_3 < e_2 \leq 1 - \frac{\mathcal{E}}{3}. \tag{11.4}$$

From (11.2) it is found that the advance in κ during a half-period is

$$\Delta\kappa = \kappa(L) - \kappa(0) = 2i\omega_1[\zeta(\nu_+) - \zeta(\nu_-)] - 2i\eta_1(\nu_+ - \nu_-).$$

Imposing periodic boundary conditions forces $\Delta\kappa$ to be a rational multiple of π. The possibilities include $\Delta\kappa = \pm\pi$. Then the spherical pendulum solution reduces to that of the planar pendulum with the identification $\sin\alpha = \sin\xi \sin\kappa$ and $\cos\alpha = \cos\xi$. In this case, $U[\psi]$ coincides with the O(2) expression as does $\rho[\psi, \psi_0]$ (this follows immediately from (4.3)).

The results of the previous paragraph show the conclusion of Feynman [15], that orbits containing configurations of nearly-constant derivatives are a small metric distance from ψ_0, is false. He considered the distance between orbits containing circulating configurations such as (7.2) and ψ_0 (his configurations were not explicit mathematically, but he correctly argued that circulating configurations could be made to have arbitrarily small potential energy). The distance $\rho^{circ}(k)$ we have already calculated shows that these are far (of order $\sqrt{L}$) from ψ_0 for lowest potential energy. Feynman claimed that a path in orbit space between these two orbits could be made short by what he called "slipping the loops about". He was referring to the fact that the first homotopy group of the two-sphere is trivial. However, this statement is false because the triangle inequality (3.12) implies any such path has a length greater than $\rho^{circ}(k)$ [17]. Nonetheless, there is some merit in Feynman's notion of making paths by "slipping the loops about". We will show later in this section that most of the important tunneling paths are of this type.

The configurations (11.1), (11.2) for the O(3) sigma model are related to the soliton configurations of a closely related model. Let us consider adding an external source to the O(3) sigma model action (3.1) as we did for the O(2) model in (7.6):

$$S = \frac{1}{2e_0} \int dt\, dx\, (\, \partial_t s^T \partial_t s - \partial_x s^T \partial_x s + h^T s\,).$$

This is no longer the periodic sine-Gordon action, as there are two independent field components, instead of one. The direction of h can be chosen so that the action becomes

$$S = \frac{1}{2e_0} \int dt\, dx \left\{ (\partial_t\xi)^2 - (\partial_x\xi)^2 + \sin^2\xi\, [(\partial_t\kappa)^2 - (\partial_x\kappa)^2] + h\, \cos\xi \right\},$$

and the equation of motion is

$$(\partial_t^2 - \partial_x^2)\xi + (\partial_t \sin^2\xi\, \partial_t - \partial_x \sin^2\xi\, \partial_x)\kappa = -h\sin\xi\,. \tag{11.5}$$

Now substituting $\xi(t,x) = \xi(u)$, $\kappa(t,x) = \kappa(u)$ where $u = t \pm x/v$, as before, into (11.5) yields the equations of motion of the spherical pendulum:

$$\partial_u^2\xi(u) + \partial_u(\sin^2\xi\, \partial_u\kappa) = -\frac{v^2 h}{(v^2-1)}\sin\xi\,,$$

The solutions to this equation are a train of solitons, just as in the O(2) case, with spacing

$$l = \sqrt{\frac{32K^2(v^2-1)}{v^2 h}}\,. \tag{11.6}$$

The difference is that this train is not, in general, a periodic configuration. The solitons change from one to the next, because of the angular shift $2\Delta\kappa$. Periodic soliton solutions do exist when $\Delta\kappa = \pi$, but these are a small subclass. While (7.7) is a completely integrable partial differential equation, we do not expect that (11.5) is integrable. Furthermore, there is no notion of topological stability for the solutions of the O(3) equation (11.5).

From the discussion above and Appendix A, it is clear that the O(2) circulating solutions (7.2) are a local minimum of the potential energy, as is the $N = 1$ oscillating solution for small k, when periodic boundary conditions are imposed.

It is not sufficient to examine only the O(2) solutions (7.2) and (7.2) embedded in the O(3) model. The reason is that (11.1) and (11.2) contain a much larger class of highly nonlinear minimal-energy configurations, which should be just as important in the thermodynamic limit. How do we impose the boundary condition? We will answer this question after examining first some general properties of (11.1) and (11.2).

For the time being we assume no special boundary condition, but a very large system, where $\cos\xi$ has period $l = L/M$, which is large compared to the short-distance cut-off. Then we can see that integrating over the long distance L should give a nonvanishing result for $\int_0^L s_3 dx/N$, but zero for $\int_0^L s_1 dx/N$ and $\int_0^L s_2 dx/N$. This is an ergodicity argument, and the result seems obvious, but is probably difficult to prove. Then the distance from ψ_0 is given by

$$\begin{aligned}\rho(N; g_2, g_3)^2 &= L - \left|\int_0^L \cos\xi\, dx\right| \qquad (11.7)\\ &= \left(1 + \frac{\mathcal{E}}{3}\right)L - \frac{L}{2N\omega_1}[\zeta(2N\omega_1 + \omega_3) - \zeta(\omega_3)] = L\left(1 + \frac{\mathcal{E}}{3} - \frac{\eta_1}{\omega_1}\right).\end{aligned}$$

The potential energy can be written down as an integral:

$$U(N; g_2, g_3) = \frac{8N^2\omega_1^2}{L^2}\int_0^L \left\{(1+\mathcal{E}) - \left[1 + \frac{\mathcal{E}}{3} + \mathcal{P}\left(\frac{2N\omega_1 x}{L} + \omega_3\right)\right]\right\} dx\,.$$

This integral can be evaluated and we find

$$U(N;g_2,g_3)=\frac{8N^2\omega_1^2}{L^2}\left(\frac{2}{3}\mathcal{E}+\frac{\eta_1}{\omega_1}\right). \tag{11.8}$$

We know that for some choice of invariants g_2 and g_3, the right-hand side of (11.8) will diverge in the ultraviolet, since it must reduce to the O(2) result when the angular-momentum parameter b vanishes. Examining the integral expression for ω_1:

$$\omega_1=\int_{e_1}^{\infty}[4(z-e_1)(z-e_2)(z-e_3)]^{-\frac{1}{2}}\,dz\,,$$

it is clear that it diverges only when $e_1-e_2\to 0$. On the other hand, η_1/ω_1 does not diverge anywhere except as $e_1\to\infty$, in which case (11.8) vanishes. This can be seen by writing this expression in term of Jacobi elliptic integrals as

$$\frac{\eta_1}{\omega_1}=-e_1+(e_1-e_3)\frac{E(k)}{K(k)}\,, \tag{11.9}$$

where

$$k^2=\frac{e_2-e_3}{e_1-e_3}\,. \tag{11.10}$$

In fact, we can also do this for ω_1:

$$\omega_1=\frac{K(k)}{\sqrt{e_1-e_3}}\,. \tag{11.11}$$

The relation (11.11) can be obtained from the identity

$$\mathcal{P}(z)=e_3+\frac{e_1-e_3}{\mathrm{sn}^2((e_1-e_3)^{1/2}z,k)}\,, \tag{11.12}$$

and identifying the periods, where k is given by (11.10). Integration of (11.12) and substitution of $\zeta(\omega_1)=\eta_1$ yields (11.9). The potential energy diverges logarithmically when $e_1-e_2\to 0$, or $k\to 1$. It is otherwise finite, vanishing in the thermodynamic limit.

To understand better where $U(N;g_2,g_3)$ diverges, we write

$$e_1=W+\delta\,,\qquad e_2=W-\delta\,,\qquad e_3=-2W\,.$$

If $\delta<1/4$, the inequalities (11.4) become

$$\frac{2+\delta}{3}>W>\frac{2-\delta}{3}\,. \tag{11.13}$$

These new inequalities are completely consistent with the relations (11.3). The divergence occurs as $\delta\to 0$. We will compute the angular-momentum parameter b and the azimuthal-angle advance $\Delta\kappa$ in this limit.

As δ vanishes, e_1, e_2 and e_3 approach $2/3$, $2/3$ and $-4/3$, respectively, by (11.13). We also find that $\mathcal{E} \to 1$. Then $\mathcal{P}(\nu_\pm) = -\mathcal{E}/3 \pm 1$ implies that $\mathcal{P}(\nu_+) = 2/3$ and $\mathcal{P}(\nu_-) = -4/3$. Hence $\nu_+ = \omega_1$ and $\nu_- = \omega_3$. But then $\mathcal{P}'(\nu_\pm)^2 = 0$, so that $b = 0$. It must therefore be true that a configuration in the orbit becomes an O(2) extremal-curve configuration with $k \approx 1$. This can be checked by seeing whether

- $|\Delta\kappa|$ is π, and
- $\rho(N; g_2, g_3)$ approaches zero as $e_1 - e_2 \to 0$.

Explicitly

$$\Delta\kappa \to 2i\omega_1[\zeta(\omega_1) - \zeta(\omega_3)] - 2i\eta_1(\omega_1 - \omega_3) = 2i(\eta_1\omega_2 - \eta_2\omega_1) = \pi\,.$$

From (11.8), (11.9) and (11.13),

$$\rho(N; g_2, g_3)^2 \to L\left(1 + \frac{1}{3} - \frac{2}{3}\right) = 0\,.$$

This shows that as the potential energy diverges, the distance between the configurations (11.1), (11.2) and ψ_0 vanishes. This property was also shown to be true for the O(2) model. What is also significant is that in this limit the configuration becomes periodic! There are potential-energy barriers just as in the O(2) model. In fact, the orbits close to ψ_0 become what is essentially the O(2) configuration for $k \approx 1$. The situation is shown schematically in figure 7.

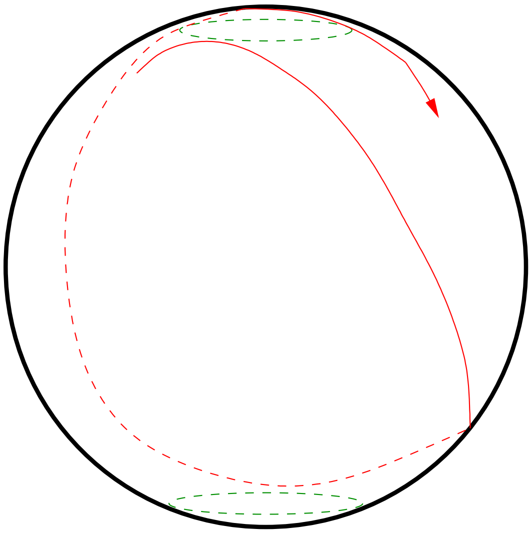

Figure 7: The configurations (11.1) and (11.2) with large potential energy are spherical pendulum solutions for which the circles begin to close around the vertical axis. As the circles close completely, the advance in azimuthal angle approaches π and the dynamics of the spherical pendulum becomes the dynamics of the planar pendulum.

Imagine moving along the barrier from ψ_0. As ρ increases and the energy decreases the parameter e_1 becomes more important, as it is no longer forced to be $2/3$. There is now an *extremal fan* consisting of many adjacent extremal curves (with $e_1 - e_2$ small), converging at the point of orbit space ψ_0. Consider a neighborhood G in the sphere around a $b = 0$ solution ψ of the form (11.1), (11.2) (this is identical to an O(2) extremal-curve solution). There is a one-dimensional curve containing ψ where the fan intersects the sphere.

The endpoints of this curve are fixed by (11.13). For any N, the fan contains a circulating O(2) river valley (7.2). We call the fan a river delta. The potential energy on this river delta is less than on the rest of G. The amplitude for barrier penetration is greatest along the delta. The situation is depicted in figure 8.

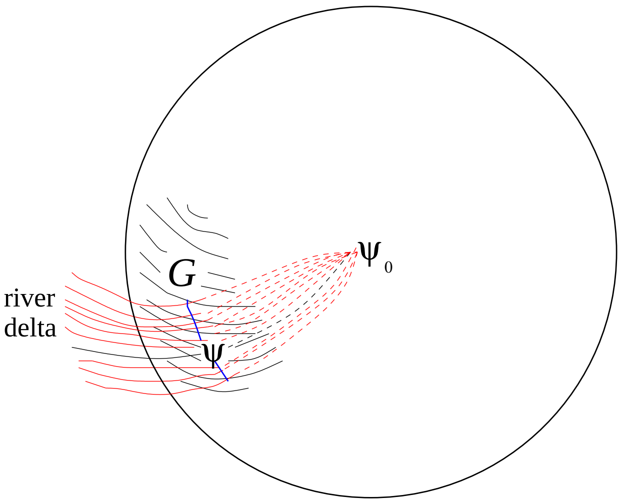

Figure 8: A two-dimensional river delta where tunneling can occur most readily though ultraviolet-divergent energy barriers. The O(2) river valley passing through the point ψ in the orbit-space sphere around ψ_0 is drawn in black, while the remainder of the delta is drawn in red. The curve of intersection of the delta and the sphere is drawn in blue.

The paths along the river delta are essentially Feynman's "slipping the loops about" paths. The O(2) river valley orbits inside the O(3) river delta contain very symmetrical configurations, which are preserved, up to a sign, under reflection $x \to L - x$. This is not so for the general river-delta configurations. The first homotopy group of O(3) is trivial and it is not hard to visualize these paths as the decay of a non-topologically-stable soliton.

Since the extremal-fan configurations with significant potential energy are periodic, we make the approximation that *all* river-delta configurations can be replaced with modified configurations which satisfy periodic boundary conditions. We add an extra linear term to $\kappa(x)$, resulting in these boundary conditions. These are not strict minima of the potential energy for a given ρ. However their potential energy is the same as that of the corresponding O(2) extremal-curve configurations within order $1/L$. The modification we make is to replace (11.2) by

$$\begin{aligned} e^{i\,\kappa(x)} = \exp\Big\{&(\eta_1+\eta_2)(\nu_+-\nu_-) - [\zeta(\nu_+)-\zeta(\nu_-)]y + \\ &+ \frac{\pi}{2\omega_1}y + 2[\zeta(\nu_+)-\zeta(\nu_-)]y - 2\eta_1(\nu_+-\nu_-)y\Big\} \times \\ &\times \sqrt{\frac{\sigma(y+\omega_3+\nu_+)\,\sigma(y+\omega_3-\nu_-)}{\sigma(y+\omega_3-\nu_+)\,\sigma(y+\omega_3+\nu_-)}}\,. \end{aligned} \tag{11.14}$$

It is now permissible to consider such a configuration with any choice of N. The river deltas in the model with periodic boundary conditions are paths through configurations of the form (11.1), (11.14).

12. Tunneling and the topological charge

Adding the term

$$\begin{aligned} S_\theta = \theta\, q = \frac{\theta}{4\pi} \int dt\, dx\; \epsilon^{a\,b\,c}\, s_a\, \partial_t s_b\, \partial_x s_c \\ = \frac{1}{4\pi} \int dt\, dx\, \left[\partial_t (\cos\xi\, \partial_x \kappa) - \partial_x (\cos\xi\, \partial_t \kappa)\right]. \end{aligned} \tag{12.1}$$

to the action has a significant effect on tunneling through the energy barriers. If the spacetime is the two-sphere and the action of the spacetime configuration $s(t,x)$ is finite, then the number q is an integer, called the topological charge, and is the degree of the mapping from the two-sphere to itself. If the action is not finite, q is not quantized and the term "topological charge" is a misnomer. We will show in this section that the space-time field configuration generated by moving from ψ_0 through the $N=1$ river delta has a half-integer value of q.

As in the O(2) model, an ultraviolet cut-off must be introduced. The tunneling for $N=1$ proceeds as follows:

1. First a wave packet of the form (8.2) expands freely from a constant configuration in ψ_0 to an $N=1$ extremal-fan configuration whose derivatives are limited by the cut-off.
2. Next this river-delta configuration begins to stretch out continuously in space (its derivatives with respect to x becoming smaller) as ρ increases. Eventually the configuration becomes a constant (in ψ_0) once again.

Since boundary conditions in x are periodic, Stokes' theorem implies

$$q = \lim_{e_1 - e_2 \to 0} \frac{1}{4\pi} \int_0^L \cos\xi\, \partial_x \kappa\,. \tag{12.2}$$

Instead of doing a detailed calculation of the degree using our explicit configurations (11.1), (11.14), we can see what (12.2) must be from an intuitive argument. The azimuthal-angle advance in the initial river-delta configuration must approach $\Delta\kappa = \pi$ as $e_1 - e_2$ approaches zero, as we have already shown. This advance takes place over extremely short intervals in x. Since $N=1$, the advance happens once on the pendulum down-swing $x = x_0 + L/2$, and again on the up-swing at $x = x_0$. Since κ is a compact variable, we must subtract 2π from the advance at the up-swing. As $e_1 - e_2 \to 0$

$$\kappa \to \pi \left[H(x - x_0) - H\left(x - x_0 - \frac{L}{2}\right)\right], \tag{12.3}$$

where H is the step function, namely $H(x) = 0$ for $x < 0$, $H(x) = 1$ for $x \geq 0$. Thus (12.2) is

$$\begin{aligned} q &= \frac{1}{4\pi} \int_0^L \cos\xi(x) \cdot \pi\, \left[\delta(x - x_0) - \delta\left(x - x_0 - L/2\right)\right]\, dx \\ &= \frac{1}{4}\left[\cos\xi(x_0) - \cos\xi\left(x_0 + L/2\right)\right]. \end{aligned}$$

Since $b \to 0$, the pendulum almost swings through the zenith and the nadir, so that $\xi(x_0) \to 0$ and $\xi(x_0 + L/2) \to \pi$. The final result is

$$q = \frac{1}{4}[1 - (-1)] = \frac{1}{2}\,.$$

As an orbit moves along the path through the barrier joining ψ_0 to itself, the wave functional must pick up a phase $\exp \pm i\theta/2$. At $\theta = \pi$, passage through the barriers contribute factors of oscillating sign to the wave functional. This means that the wave functional will vanish somewhere along the barrier. Barrier configurations with $N = 1$ are suppressed. However, multi-soliton barrier configurations may survive.

13. Discussion

Our hamiltonian methods, though semiclassical in nature, are very different from the saddle-point techniques most field theorists are accustomed to. They are conceptually more complicated than saddle-point methods, but are at least partially successful for the O(3) sigma model. We do not know yet whether they can be developed to obtain quantitative results or if they might suggest a more powerful version of the saddle-point methods.

We have not evaluated the barrier-penetration amplitude in the O(3) sigma model. This is a formidable problem for several reasons. The regions of barrier penetration in orbit space are two-dimensional, unlike the case of the XY model, for which they are one-dimensional. More significantly, the task of understanding the role of nonlinear spin-wave fluctuations near the barriers will probably not be easy. In the XY model, further spin-wave corrections are not important because the spin waves are effectively gaussian.

Though there is still much we do not understand about the O(3) model, our analysis has been revealing. We have found the explicit form of the barrier configurations. These truly have half-integer topological charge, just as predicted by Affleck and Haldane [9]. However, the one-parameter family of barrier configurations are more general and more complicated than both the modified meron in Affleck's paper and the original meron of Gross [10].

When $\theta = 0$, the Schrödinger wave functional on configuration space can be significantly different from zero only where the potential energy is small or near a river-delta configuration. This is strong evidence that barrier penetration produces the mass gap. The fact that the topological charge is half-integer means that at $\theta = \pi$, there can be a massless phase. A phase transition may separate this from a massive strong-coupling phase, driven by pairs of barrier-penetration events (of vanishing or integer topological charge).

Feynman's [15] point of view as to how the gap arises was rather different from ours. He argued that the mass gap should arise because the vacuum wave functional should vanish everywhere on the orbit space, except on a region of finite diameter. This may be true, though it is a subtle matter in an asymptotically-free theory where all length scales are important. Quantum barrier penetration will occur for configurations other than the classical vacuum ψ_0; one effect of such barrier penetration may be to suppress configurations which are a distance $\rho \sim \sqrt{L}$ from ψ_0.

We have left many questions unanswered and see three further directions for this research.

While we find the phase transition in the hamiltonian XY model, our technique is not yet powerful enough to fully understand even this model. Our methods are as good as vortex-mean-field theory [5], but we are not yet able to do a renormalization group analysis, as Kosterlitz first did for the Coulomb gas (or its equivalent, the sine-Gordon model). It may be that a hamiltonian renormalization group of the barrier configurations is possible.

We would like to understand better the condensation of barrier configurations in the O(3) model. This will require a proper incorporation of spin-wave effects. In any case, our results indicate that the instanton gas of integer topological charges [24] is not sufficient to understand mass generation in the model. The one-parameter configurations we have found appear to be more general than merons [10]. We expect that if a calculation could be done by a saddle-point method in euclidean space, it would also reveal this to be the case.

Finally, the extremal problem for the sigma model (6.2) has an analogue in gauge theories. Some progress has been made towards its solution [25], building on some of the work here and in reference [13].

Acknowledgments

P.O. thanks the Niels Bohr Insitute staff for their hospitality. We would also like to thank W. Bietenholz for a careful reading of the manuscript.

A. Which extremal curves are minimal curves?

In this appendix we determine which of our extremal curves for the O(2) model are what we call river valleys. These must be minimal curves. We need to know whether a point (actually a field configuration or an orbit) on such a curve, constrained to be on the sphere (of constant r from s_0 or of constant ρ from ψ_0) is in stable classical equilibrium.

Once the situation is understood for the O(2) model, it is possible to decide the issue of river deltas in the O(3) model, discussed in section 11.

Consider a solution of (6.2) for the O(2) case. Suppose that the angle $\alpha(x)$ is an extremal curve configurations such as (7.2) or (7.2). If we vary this angle by $\delta\alpha(x)$, the potential energy changes to second order in $\delta\alpha$ by

$$\begin{aligned} \delta U &= -\int_0^L \left(\delta\alpha\frac{d^2}{dx^2}\alpha + \frac{1}{2}\delta\alpha\frac{d^2}{dx^2}\delta\alpha\right)\,dx \\ &= -\frac{1}{2}\int_0^L \delta\alpha\left(\frac{d^2}{dx^2}\delta\alpha - 2\lambda\sin\alpha\right)\,dx\,, \end{aligned} \tag{A.1}$$

where λ is the Lagrange multiplier in (6.2) and we have used the pendulum equation of motion. We want to know whether δU is positive for any acceptable $\delta\alpha$, other than the zero mode of translation invariance. We say "acceptable" because $\delta\alpha$ is not arbitrary. There is a constraint in (6.2) on $\delta\alpha$, namely that the variation must not change the distance in orbit space to ψ_0.

The constraint means that $\delta\alpha$ must satisfy

$$\begin{aligned} v^2 &= \left[\int_0^L \cos(\alpha+\delta\alpha)\,dx\right]^2 + \left[\int_0^L \sin(\alpha+\delta\alpha)\,dx\right]^2 \\ &= \left(\int_0^L \cos\alpha\,dx\right)^2 + \left(\int_0^L \sin\alpha\,dx\right)^2 . \end{aligned}$$

Expanding this up to second order in $\delta\alpha$ gives

$$\begin{aligned} 0 = &-v_1\int_0^L \left[\delta\alpha\,\sin\alpha + \frac{1}{2}(\delta\alpha)^2\,\cos\alpha\right]\,dx + \left(\int_0^L \delta\alpha\,\sin\alpha\,dx\right)^2 \\ &+ v_2\int_0^L \left[\delta\alpha\,\cos\alpha - \frac{1}{2}(\delta\alpha)^2\,\sin\alpha\right]\,dx + \left(\int_0^L \delta\alpha\,\cos\alpha\,dx\right)^2 , \end{aligned} \tag{A.2}$$

where

$$v_1 = \int_0^L \cos\alpha\,dx\,, \qquad v_2 = \int_0^L \sin\alpha\,dx\,.$$

The second-order form of the constraint (A.2) can be simplified somewhat.

Let us rotate α so that $v_1 = v$ and $v_2 = 0$, as is the case for (7.2), (7.2). The same conditions can be imposed on $\alpha + \delta\alpha$. The expansion of these conditions to second order yields

$$0 = \int_0^L \left[\delta\alpha\,\sin\alpha + \frac{1}{2}(\delta\alpha)^2\,\cos\alpha\right]\,dx \tag{A.3}$$

$$0 = \int_0^L \left[\delta\alpha\,\cos\alpha - \frac{1}{2}(\delta\alpha)^2\,\sin\alpha\right]\,dx\,. \tag{A.4}$$

Conditions (A.3) and (A.4) are equivalent to (A.2). The variation of the potential energy (A.1) upon substitution of (A.3) is

$$\delta U = -\frac{1}{2}\int_0^L \delta\alpha\left(\frac{d^2}{dx^2} + \lambda\cos\alpha\right)\delta\alpha\,dx\,, \tag{A.5}$$

Substituting (7.2) and (7.2), equation (A.5) becomes

$$\delta U = \frac{1}{2}\int_0^L \delta\alpha \left[-\frac{d^2}{dx^2} + \frac{16N^2K^2}{L^2}\left(-1 + 2k^2\,\mathrm{sn}^2\frac{4NK(x-x_0)}{L}\right)\right]\delta\alpha\, dx \quad \text{(A.6)}$$

and

$$\delta U = \frac{1}{2}\int_0^L \delta\alpha \left[-\frac{d^2}{dx^2} + \frac{4N^2K^2k^2}{L^2}\left(-1 + 2\,\mathrm{sn}^2\frac{2NK(x-x_0)}{L}\right)\right]\delta\alpha\, dx\,, \quad \text{(A.7)}$$

respectively.

The eigenvalue equation for the operator in square brackets in each of (A.6) and (A.7) is a Hill equation, specifically a Lamé equation [18] of order one. It is still necessary to impose the constraints (A.3) and (A.4).

Our problem has become the following question: under what circumstances are the quadratic forms in (A.6) and (A.7) positive on variations satisfying (A.3) and (A.4)? After an appropriate rescaling, the Lamé operators are

$$\mathcal{H}_{osc} = \mathcal{H} - 1\,, \quad \text{(A.8)}$$

on periodic functions of period $4NK$, for the oscillating extremal curves (7.2) and

$$\mathcal{H}_{circ} = \mathcal{H} - k^2\,, \quad \text{(A.9)}$$

on periodic functions of period $2NK$, for the circulating extremal curves (7.2), where in each case

$$\mathcal{H} = -\frac{d^2}{du^2} + 2k^2\,\mathrm{sn}^2\, u\,.$$

To answer this question it is first necessary to see whether the spectrum of (A.8) or (A.9) is positive-definite with a single zero mode (corresponding to translation invariance of the extremal curve configuration (7.2) or (7.2), respectively). If this is the case for a particular Lamé operator, any variation raises the potential energy. If not, meaning that there is a negative eigenvalue in the spectrum, it must then be checked on a case-by-case basis whether variations, other than the zero mode of translation invariance, satisfying (A.3) and (A.4) can give a nonpositive δU.

The zero mode corresponding to translations for (A.8) is $\mathrm{cn}\, u$ and for (A.9) is $\mathrm{dn}\, u$, which are both doubly-periodic functions. These modes are obtained by simply differentiating (7.2) and (7.2) with respect to x_0. Recall that the ground-state eigenfunction of a quantum-mechanical hamiltonian must be a unique (i.e. nondegenerate) real function which vanishes nowhere in the physical region of u. Every other real eigenfunction must have at least one node. In fact if a given eigenfunction possesses no zeros in the physical region it must be the ground-state eigenfunction. This eigenfunction is $\mathrm{dn}\, u$ (which satisfies the boundary conditions for (A.8) and (A.9)).

Let us first consider the case of the oscillating extremal curves, (A.8). We can see readily that there is a negative eigenvalue in the spectrum. We have shown that the ground-state eigenfunction for either the oscillating and the circulating case, is $\mathrm{dn}\, u$ which has period $2K$ and no nodes. Its eigenvalue of $\mathcal{H}_{osc}$ is $-k'^2$. We will not do the analysis here to determine the effect of the constraints (A.3), (A.4) on δU (though it can probably be done, as the spectrum of the order-one Lamé equation with our boundary conditions can be completely determined). We will only note that the $N = 1$ oscillating solution (7.2) for sufficiently small k *must* be a local minimum on the sphere of radius ρ_0. For there are only three kinds of extremal configurations on the sphere (see figure 1):

- the $k < k^*$ oscillating solutions (7.2).
- the $k > k^*$ oscillating solutions.
- the circulating solutions (7.2).

If ρ_0 is sufficiently small, a configuration of the third type is a barrier configuration. In that case, among the three possibilities, the $N = 1$ oscillating solution with $k < k^*$ has the smallest potential energy. Since the potential energy is bounded below, it must have a minimum value on the sphere.

For the circulating extremal curves (A.9), the ground state $\mathrm{dn}\, u$ is also the zero mode, so the remainder of the spectrum is positive. Hence all the circulating extremal curves are river valleys.

B. Appendix B: The spherical pendulum

The lagrangean of the spherical pendulum is

$$L = \frac{1}{2}\dot{\xi}^2 + \frac{1}{2}\sin^2\xi\, \dot{\kappa}^2 - \mu(1 - \cos\xi)\,.$$

The equation of motion for κ is the statement that the vertical component of the angular-momentum vector is conserved, that is

$$\frac{dl_z}{dt} = 0\,, \qquad l_z = \sin^2\xi\, \dot{\kappa}\,. \tag{B.1}$$

The conserved energy is

$$E_{pend} = \frac{1}{2}\dot{\xi}^2 + \frac{l_z^2}{2\sin^2\xi} + \mu(1 - \cos\xi)\,. \tag{B.2}$$

Define the new time coordinate $\tau = \sqrt{\mu/2}t$ and the new conserved quantities $b = l_z^2/2\mu$ and $\mathcal{E} = E_{pend}/\mu - 1$. The energy-conservation relation, (B.2) becomes

$$\left(\frac{d\xi}{d\tau}\right)^2 = 4(\mathcal{E} + \cos\xi) - \frac{4b}{\sin^2\xi}\,.$$

Let $\mathcal{Z} = -\cos\xi$. The energy-conservation equation is then

$$\left(\frac{d\mathcal{Z}}{d\tau}\right)^2 = 4\mathcal{Z}^3 - 4\mathcal{E}\mathcal{Z}^2 - 4\mathcal{Z} + 4(\mathcal{E} - b) \equiv M(\mathcal{Z})\,. \tag{B.3}$$

The physical region is, by definition, $\mathcal{Z} \in [-1, 1]$. Since $(d\mathcal{Z}/d\tau)^2$ must be positive in this region, the cubic polynomial $M(\mathcal{Z})$ on the left-hand side of (B.3) must have some positive values in this region. If $\mathcal{Z} = \pm 1$, then $M(\mathcal{Z})$ has the value $-4b$. Hence the polynomial has at least two real roots between -1 and 1. The motion of the pendulum is such that the value of $\mathcal{Z}$ is between these two roots. Furthermore, if (B.3) is continued outside of the physical region, in the limit as $\mathcal{Z} \to \infty$, the behavior of $M(\mathcal{Z})$ is $M(\mathcal{Z}) \to \infty$. Therefore there is one (unphysical) root of $M(\mathcal{Z})$ in the open interval $\mathcal{Z} \in (1, \infty)$.

After an appropriate shift of $\mathcal{Z}$ by $-\mathcal{E}/3$, eliminating the quadratic term in the polynomial $M(\mathcal{Z})$, it is straightforward to solve (B.3) in terms of the Weierstrass elliptic function $\mathcal{P}(z; g_2, g_3) = \mathcal{P}(z)$. The solution is

$$\mathcal{Z} = \mathcal{P}(\tau - \tau_0) + \frac{\mathcal{E}}{3}$$

where

$$g_2 = 4\left(1 + \frac{\mathcal{E}^2}{3}\right), \qquad g_3 = 4\left(\frac{2\mathcal{E}^3}{27} - \frac{2\mathcal{E}}{3} + b\right).$$

We can parametrize the Weierstrass function by e_1, e_2 and e_3 instead of g_2 and g_3. These numbers are just the roots of the polynomial on the right-hand side of (B.3) shifted by $-\mathcal{E}/3$. They are real, add up to zero and satisfy

$$e_1 > 1 - \frac{\mathcal{E}}{3}, \qquad -1 - \frac{\mathcal{E}}{3} \le e_3 < e_2 \le 1 - \frac{\mathcal{E}}{3}\,.$$

Since e_1 is positive, e_3 is negative, while e_2 can be of either sign. The motion of the pendulum always lies between two circles on the sphere. The pendulum reaches the bottom circle at time τ such that $\mathcal{P}(\tau - \tau_0) = e_3$ and the top circle at time τ such that $\mathcal{P}(\tau - \tau_0) = e_2$. If the initial condition at $\tau = 0$ is chosen so that $\mathcal{Z}$ lies on the bottom circle, then $-\tau_0$ is fixed to be ω_3.

Notice that

$$\omega_1 = \int_{e_1}^{\infty} (4t^3 - g_2 t - g_3)^{-1/2} dt$$

is purely real, while

$$\omega_3 = -i \int_{-\infty}^{e_3} (g_3 + g_2 t - 4t^3)^{-1/2} dt$$

is purely imaginary.

Equation (B.1) can now be written

$$\frac{d\kappa}{d\tau} = \frac{2\sqrt{b}}{1-[\mathcal{P}(\tau+\omega_3)+\frac{\mathcal{E}}{3}]^2}\,. \tag{B.4}$$

To integrate (B.4) requires two steps. The first is to reduce the denominator, which is quadratic in $\mathcal{P}(\tau+\omega_3)$ to the sum of terms with denominators linear in this function, that is

$$\frac{1}{1-(\mathcal{P}+\frac{\mathcal{E}}{3})^2} = \frac{1}{2}\frac{1}{\mathcal{P}+\frac{\mathcal{E}}{3}+1} - \frac{1}{2}\frac{1}{\mathcal{P}+\frac{\mathcal{E}}{3}-1}\,.$$

To integrate such an expression, we prove the identity

$$\frac{\mathcal{P}'(\nu)}{\mathcal{P}(z)-\mathcal{P}(\nu)} = 2\zeta(\nu)+\zeta(z-\nu)-\zeta(z+\nu)\,. \tag{B.5}$$

We will show that each of the two sides of (B.5) has the same zeros and poles (and multiplicities thereof) with the same residues of the poles. Equation (B.5) then follows. Since $\zeta(z+2\omega_i)=\zeta(z)+2\eta_i$, the right-hand side is an elliptic function. The zeros of the right-hand side of (B.5) are points congruent to zero, since the Weierstrass zeta function is odd. Furthermore, the order of each of these zeros is two, for as $z\to 0$,

$$2\zeta(\nu)+\zeta(z-\nu)-\zeta(z+\nu) \to -z[\mathcal{P}(-\nu)-\mathcal{P}(\nu)]+\mathrm{O}(z^2)=\mathrm{O}(z^2)\,.$$

The poles of the right-hand side are points congruent to $z=\pm\nu$, since $\zeta(z)$ has poles congruent to zero. These are simple poles. We therefore have in each period cell one zero of order two and two poles of order one. The poles have residue ± 1 at $\pm\nu$. Now the left-hand side of (B.5) has double zeros at the double poles of $\mathcal{P}(z)$; these are the points congruent to zero. Furthermore, it has simple poles at the points congruent to $\pm\nu$, with residue

$$\lim_{z\to\pm\nu}\frac{\mathcal{P}'(\nu)}{\mathcal{P}(z)-\mathcal{P}(\nu)}\,(z\pm\nu)=\pm 1\,.$$

Therefore (B.5) is correct.

The right-hand side of (B.5) is easy to integrate, giving the result

$$\int\frac{dz}{\mathcal{P}(z)-\mathcal{P}(\nu)} = \frac{1}{\mathcal{P}'(\nu)}\left[2z\zeta(\nu)+\log\frac{\sigma(z-\nu)}{\sigma(z+\nu)}\right].$$

Define two complex numbers ν_+ and ν_- by

$$\mathcal{P}(\nu_\pm) = -\frac{\mathcal{E}}{3}\pm 1\,. \tag{B.6}$$

These numbers are determined up to $2m\omega_1 + 2n\omega_3$ as well as an overall sign. Now

$$\mathcal{P}'(\mathcal{P}^{-1}(z)) = (4z^3 - g_2 z - g_3)^{1/2} \,,$$

and substituting (B.6) gives

$$\mathcal{P}'(\nu_\pm)^2 = -4b \,.$$

Since $\mathcal{P}'$ is an odd function, we have the option of choosing

$$\mathcal{P}'(\nu_\pm) = -2i\sqrt{b} \,.$$

This choice is satisfied by the expressions for $\nu_\pm$ evaluated below.

Inverting (B.6) gives

$$\nu_\pm = \omega_3 + \int_{-\frac{\mathcal{E}}{3}\pm 1}^{e_3} \frac{dz}{\sqrt{4\,(z-e_1)\,(z-e_2)\,(z-e_3)}} \,. \tag{B.7}$$

We can immediately see from (B.7) and $-\frac{\mathcal{E}}{3} - 1 < e_3 < -\frac{\mathcal{E}}{3} + 1$ that ν_- is purely imaginary. On the other hand, ν_+ has a nonzero real part. Explicitly,

$$\begin{aligned}
\nu_+ &= \omega_3 - \int_{e_3}^{-\frac{\mathcal{E}}{3}\pm 1} \frac{dz}{\sqrt{4\,(z-e_1)\,(z-e_2)\,(z-e_3)}} \\
&= \omega_3 + \int_{e_2}^{-\frac{\mathcal{E}}{3}+1} \frac{dz}{\sqrt{-4\,(e_1-z)\,(z-e_2)\,(z-e_3)}} - \\
&\quad - \int_{e_3}^{e_2} \frac{dz}{\sqrt{4\,(e_1-z)\,(e_2-z)\,(z-e_3)}} \,.
\end{aligned} \tag{B.8}$$

The first term of (B.8) is purely imaginary. So is the second; this follows from $e_1 > -\frac{\mathcal{E}}{3} + 1$. The last term is $\omega_2 - \omega_3 = \omega_1 + 2\omega_2$. We therefore have (adding $-2\omega_2$ to ν_+ which changes nothing)

$$\nu_+ = \omega_1 + i\beta \,, \qquad \nu_- = i\gamma \,,$$

where β and γ are real constants between zero and $-2i\omega_3$.

The solution of (B.4) can now be written down. It is

$$\begin{aligned}
e^{i\,\kappa(x)} &= \exp\Big\{(\eta_1+\eta_2)(\nu_+ - \nu_-) - [\zeta(\nu_+) - \zeta(\nu_-)]\tau\Big\} \times \\
&\quad \times \sqrt{\frac{\sigma(\tau+\omega_3+\nu_+)\,\sigma(\tau+\omega_3-\nu_-)}{\sigma(\tau+\omega_3-\nu_+)\,\sigma(\tau+\omega_3+\nu_-)}} \,.
\end{aligned} \tag{B.9}$$

References

[1] A.M. Polyakov, *Phys. Lett.* **B 59** (1975) 79.

[2] A.A. Belavin and A.M. Polyakov, *Sov. Phys. JETP Lett.* **22** (1975) 245.

[3] F.D.M. Haldane, *Phys. Lett.* **A 93** (1983) 464; *Phys. Rev. Lett.* **50** (1983) 1153; *J. Appl. Phys.* **57** (1985) 3359.

[4] H. Levine, S. Libby and A. Pruisken, *Phys. Rev. Lett.* **51** (1983) 1915, erratum *ibid.* **52** (1984) 1254.

[5] V.J. Berezinskii, *Sov. Phys. JETP* **32** (1971) 493;
J.M. Kosterlitz and D.V. Thouless, *J. Phys.* **C6** (1973) 1181;
J.M. Kosterlitz, *J. Phys.* **C7** (1974) 1046.

[6] A.B. Zamolodchikov and Al.B. Zamolodchikov, *Ann. Phys. (NY)* **120** (1979) 253.

[7] P.B. Wiegmann, *Phys. Lett.* **B 152** (1985) 209;
A.M. Polyakov and P.B. Wiegmann, *Phys. Lett.* **B 131** (1983) 121.

[8] V.O. Tarasov, L.A. Takhtadjan and L.D. Fadeev, *Theor. Math. Fiz.* **57** (1983) 163.

[9] I. Affleck, *Phys. Rev. Lett.* **56** (1986) 408; *J. Phys. Condens. Matter* **1** (1989) 3047;
F.D.M. Haldane (unpublished).

[10] D.J. Gross, *Nucl. Phys.* **B 132** (1978) 439.

[11] W. Bietenholz, A. Pochinski and U.-J. Wiese, *Phys. Rev. Lett.* **75** (1995) 4524 [`hep-lat/9505019`].

[12] M. Asorey and F. Falceto, *Phys. Rev. Lett.* **80** (1998) 234 [`hep-th/9712126`].

[13] P. Orland, NBI preprint, NBI-HE-96-35 (1996), *The metric on the space of Yang-Mills configurations*, `hep-th/9607134`.

[14] M. Kudinov, E. Moreno and P. Orland, in *Proc. of NATO Adv. Research Workshop*, Zakopane, *New Developments in Quantum Field Theory*, ed. P.H. Damgaard and J. Jurkiewicz, 1998, 315.

[15] R.P. Feynman, *Nucl. Phys.* **B 188** (1981) 479.

[16] O. Babelon and C.M. Viallet, *Comm. Math. Phys.* **81** (1981) 515; *Phys. Lett.* **B 103** (1981) 45.

[17] A.D. Alexandrov and V.A. Zalgaller, *Intrinsic Geometry of Surfaces*, *AMS Translations of Mathematical Monographs*, **15**, ed. J. M. Danskin, 1967.

[18] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, Cambridge University Press.

[19] I.M. Singer, *Physica Scripta* **24** (1980) 817;
M.F. Atiyah, N.J. Hitchin and I.M. Singer, *Proc. R. Soc. Lond.* **A362** (1978) 425;
P.K. Mitter and C.M. Viallet, *Comm. Math. Phys.* **79** (1981) 457;
O. Babelon and C.M. Viallet, *Phys. Lett.* **B 85** (1979) 246;
P.K. Mitter, in *Proc. of NATO Adv. Study Inst.*, Cargèse, Aug. 1979, Plenum Press, New York, 1980;
M.S. Narasimhan and T.R. Ramadas, *Comm. Math. Phys.* **67** (1979) 121;
M. Asorey and P.K. Mitter, *Comm. Math. Phys.* **80** (1981) 43.

[20] E. Fradkin and L. Susskind, *Phys. Rev.* **D 17** (1978) 2637.

[21] M. Hasenbusch and K. Pinn, *J. Phys.* **A30** (1997) 63.

[22] E.T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, 1937. Note: there is an error in Whittaker's discussion of the spherical pendulum in that he asserts that there is a choice for ν_+ (called λ in his notation) which is purely imaginary.

[23] A. Weinstein, *Amer. Math. Monthly* **49** (1942) 521.

[24] V.A. Fateev, I.V. Frolov and A.S. Schwarz, *Nucl. Phys.* **B 154** (1979) 1;
B. Berg and M. Lüscher, *Comm. Math. Phys.* **69** (1979) 57.

[25] P. Orland, to appear.