CURVE STRAIGHTENING AND
A MINIMAX ARGUMENT FOR CLOSED ELASTIC CURVES

by 83 35

Joel Langer and David A. Singer

Max-Planck-Institut
Für Mathematik
Bonn, Germany

Case Western Reserve University
Cleveland, Ohio
USA

SFB/MPI 83-35

CURVE STRAIGHTENING

AND A MINIMAX ARGUMENT FOR CLOSED ELASTIC CURVES

The results obtained here have to do with the following problem. Imagine the ends of a straight length of springy wire are joined together smoothly and the wire is held in some configuration described by an immersion γ of the circle into the plane or into R^3. According to the Bernoulli-Euler theory of elastic rods the bending energy of the wire is proportional to the total squared curvature of γ, which we will denote by $F(\gamma) = \int_\gamma k^2 ds$. Suppose now the wire is released and it moves so as to decrease its bending energy as efficiently as possible, i.e., following "the negative gradient of F" (so our dynamics are Aristotelian rather than Newtonian, and we are also making the physically unrealistic assumption that the wire can pass through itself freely). How does the wire evolve, and what will happen ultimately (as time goes to infinity)?

Of course, one wants to know first that one can actually define such a flow on the space of immersed circles, that it exists for all time, and that one can sensibly speak of a limiting curve γ_∞ for the trajectory through a given initial curve γ_0. It is shown here that this is indeed the case and that in fact the Palais-Smale condition holds for this flow. It is proved, moreover, that if γ_0 is a plane curve of rotation index one (e.g., if γ_0 is embedded) then the flow carries γ_0 to a circle.

Our main result, however, pertains to the non-planar case, where the situation is more complicated. In a space form it is possible to integrate the equations for an elastica, i.e., for a critical point of F, and this enables one to prove, in particular, that there is a countably infinite family of (similarit classes of) closed elastic curves in R^3 (see Theorem 0.1). Thus, not all wire loops in R^3 will flow to a circle. On the other hand, this leaves open the possibility that γ_∞ is a circle for almost any initial curve γ_0, and indeed, our concluding Theorem 3. states that the circle is the only stable closed elastica in R^3.

The proof of this theorem itself depends on the dynamical, i.e., gradient flow approach to the study of F (and avoids a detailed analysis of the Hessian of F, which is quite complicatec for non-planar elastic curves). The idea is as follows. One considers a discrete group G of rotations of R^3 and an associated pair of multiply covered circular elastic curves which are G-equi variantly regularly homotopic, and which are both local minima for the restriction of F to G-symmetric curves (though multiple circles are unstable with respect to general variations). An appeal to the minimax and symmetric criticality principles then enables one to conclude that there exists a non-circular elastica of "saddle type". Comparison with the classification theorem shows that one can account in this way for all non-circular solutions, hence all are unstable.

We remark that a similar critical structure occurs for "free" (length unconstrained) elastic curves in the standard two-sphere: it was shown in [5] (by an entirely different method) that all closed non-geodesic solutions in S^2 are unstable and can be regarded as minimax critical points arising from symmetrical regular homotopies between certain multiple coverings of a prime geodesic (though the minimax argument in [5] is made only heuristically). To the extent that a similar picture holds as well for manifolds of (non-constant) positive curvature one gains a new view of closed geodesics as the limits of almost all trajectories of $-\nabla F$.

The organization of the paper is as follows. Section 0 is a brief review of some basic facts concerning elastic curves in space forms and the classification of closed elastic curves in R^3 (details can be found in [5],[6]). Section 1 is devoted mostly to the proof of condition (C) for the curve straightening flow. We have included details and have attempted to keep the discussion as self-contained as possible. In Section 2 we derive a second variation formula which is compatible with the set-up of Section 1 and use the formula to verify the hypothesis of the splitting theorem of Gromoll and Meyer, which is required for our instability theorem. Also, we illustrate the stability problem for elastic curves by applying the formula to several concrete cases where explicit computations are possible. Section 3 combines the results of the previous three sections in the minimax argument and concludes with the instability theorem for non-circular elastic curves.

0. Closed Euclidean Elastic Curves

A classical elastica (or elastic curve) is a curve in R^2 or R^3 which is critical for the total squared curvature functional $F(\gamma) = \int_\gamma k^2 ds$ defined on regular curves of a fixed length satisfying given first order boundary data. If one removes the constraint on arclength one speaks of a free elastica. The notions of elastica and free elastica are meaningful in any Riemannian manifold, k being the geodesic curvature of γ.

The present paper is concerned mostly with closed elastic curves (where the boundary conditions specify that γ is closed to first order), though much of what we do applies to the general boundary value problem (see Remark 1 at the end of Section 1). The most obvious examples of closed free elastic curves are the closed geodesics. Almost as obvious examples of closed (non-free) elastic curves are provided by the geodesic circles in a space form M, e.g., in R^3 (actually, in the hyperbolic case, circles of radius $\frac{\sinh^{-1}(1)}{\sqrt{-G}}$ are free -- G being the sectional curvature of M). All other examples of closed elastic curves (free or not) require some work to obtain, but in the case of a space form there are some nice facts about elastic curves which make a complete classification possible.

To begin with, the Euler equations for the squared curvature $x=k^2$ and torsion τ of an elastica γ can be put in the form

$$\begin{cases} (x_s)^2 = P(x) = -x^3 + 2\mu x^2 + 4Ax - 4c^2 \\ x\tau = c \end{cases}$$

where x_s is the derivative of x with respect to the arclength parameter, μ is a Lagrange multiplier which arises when arclength is constrained, and $c>0$ and A are arbitrary constants. As P(x) is cubic polynomial in x (depending on parameters μ, A, c) it follows that the curvature and torsion of γ can be expressed in terms of elliptic functions. Furthermore, the determination of whether γ is closed comes down to an investigation of periods of associated elliptic integrals.

The latter statement, though not obvious, comes directly from the fact that the symmetry of M enables one to integrate the Frenet equations for γ. This is explained by Bryant and Griffiths in terms of the theory of exterior differential systems [1]. Independently, the present authors worked out a classification of closed free elastic curves in two dimensional space forms [5] as well as closed (non-free) elastic curves in Euclidean space [6].

In order to describe the classification in the Euclidean case it is helpful to give a brief sketch of the argument. It is proved in [5] that if γ is an elastica in a space form M^3 (it suffices to consider the three dimensional case since all higher torsions vanish) and $\{T,N,B\}$ is the Frenet frame for γ then the fields

$$\begin{cases} J_0 = (k^2-\mu)T + 2k_s N + 2k\tau B \\ J_1 = kB \end{cases}$$

extend to Killing fields on the universal cover of M^3.

We explain how this is used when $M^3 = R^3$. In this case J_0 can be shown to have constant magnitude and therefore must be a translation field, while J_1 is easily shown to be a combination of J_0 and a rotation field about an invariant line of J_1. Thus, for every elastica γ in R^3 there is naturally associated to γ a cylindrical coordinate system (r,θ,z) on R^3, the restrictions to γ of the coordinate fields $\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial z}$ being expressible in terms of k, τ, T, N, B. Setting $\gamma(s)=(r(s),\theta(s),z(s))$ one then obtains the derivatives r', θ', z' by taking dot product of T with the expressions for $\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial z}$ (and normalizing in the case of $\frac{\partial}{\partial \theta}$). Integration then yields $\gamma(s)$.

In particular, to determine whether γ closes up one considers the definite integrals over one period of the curvature function, for these are the net changes of $r(s)$, $\theta(s)$, $z(s)$ in each period. As it turns out, $\Delta r=0$, $\Delta\theta$ is an elliptic integral of the third kind and Δz is an elliptic integral of the second kind. Now γ is closed if and only if it satisfies

$$\begin{cases} \Delta z=0 \\ \dfrac{\Delta\theta}{\pi} = \text{rational}, \end{cases}$$

thus explaining the claim made earlier concerning the closedness question.

To describe the behavior of Δz, $\Delta\theta$ we note first that the general expressions for the curvature and torsion of an elastica γ can be given in terms of the maximum, α, of k^2 and two further parameters $0 \le p \le w \le 1$ which control the shape of γ:

$$\begin{cases} k^2 = x(s) = \alpha\left(1-\dfrac{p^2}{w^2}\,\mathrm{sn}^2(rs,p)\right), \quad r=\dfrac{\sqrt{\alpha}}{2w} \\ \tau(s)x(s) = c \end{cases} \tag{0.1}$$

Here $\mathrm{sn}(t,p)$ is the Jacobi elliptic sine function of modulus p, and α, p, w are of course related to the parameters μ, c, A appearing in the Euler equations. Two of these relations will be useful later and we note them here:

$$\begin{cases} 4c^2 = \dfrac{\alpha^3}{w^4}(1-w^2)(w^2-p^2) \\ \mu = \dfrac{\alpha}{2w^2}(3w^2-p^2-1). \end{cases} \tag{0.2}$$

Since we are interested only in the shape of γ it is convenient to fix $\alpha=1$ and to consider the triangular (p^2,w^2)-parameter space for all elastic curves (not necessarily closed) in R^3.

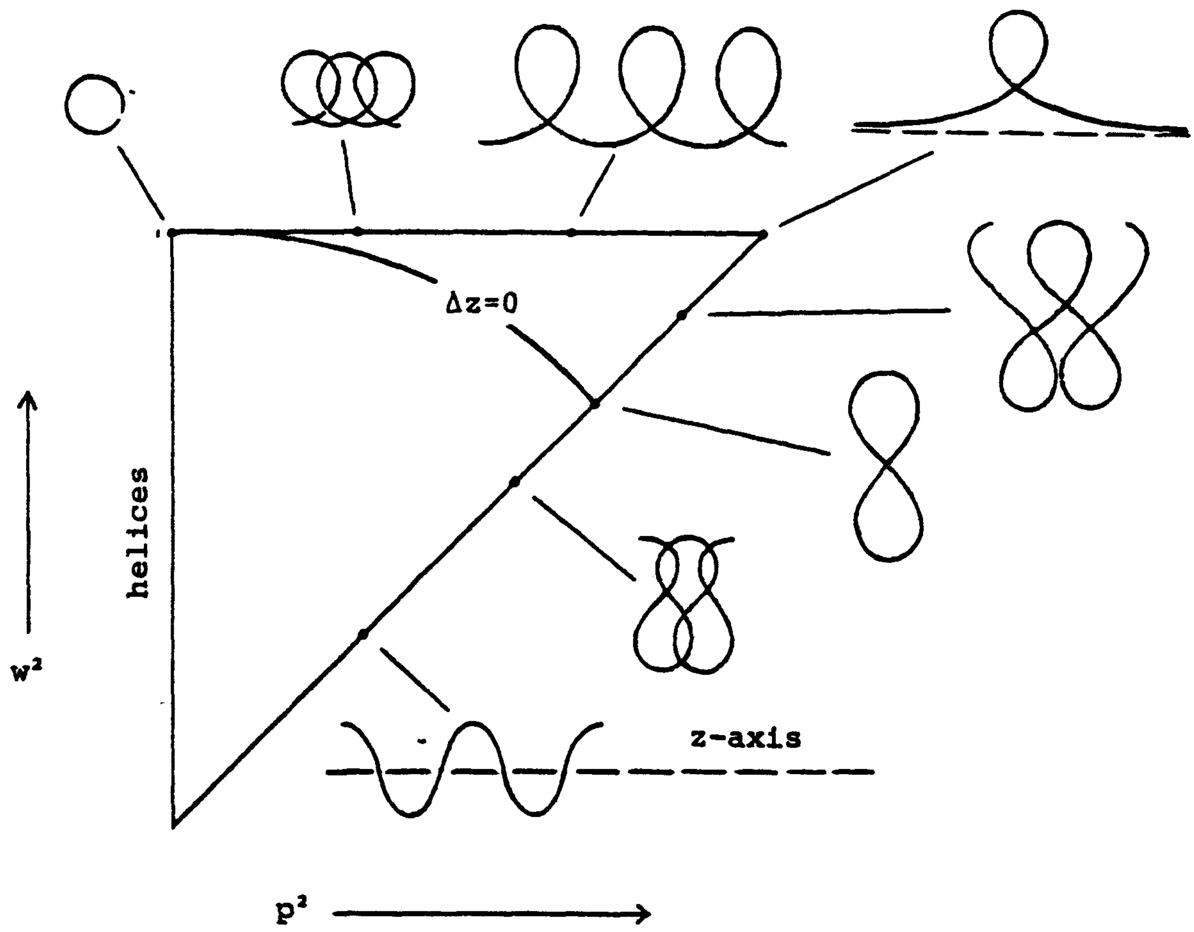

Figure 1

From (0.1) one sees that the planar elastic curves are represented in Figure 1 by points on the upper and diagonal boundaries, and the non-planar elastic curves (except for the helices) correspond to points in the interior (so all curves drawn in Figure 1 are planar).

Analysis of Δz shows that the curve $\Delta z=0$ behaves qualitatively as pictured in Figure 1. Note that the circle and the figure eight are the two points of intersection of $\Delta z=0$ with the boundary. Analysis of $\Delta\theta$ is much more difficult but the computations in [6] show that $\Delta\theta$ decreases monotonically from 0 to $-\pi$ as the curve $\Delta z=0$ is traversed from left to right.

Together with a few simple observations about the qualitative behavior of $r(s)$ and $z(s)$ this proves the classification

Theorem 0.1 a) Let γ be a closed planar elastica. Then γ is a circle or (up to similarity) the unique figure eight elastica which closes up in one period of its curvature $k(s)=cn(rs,p)$ or a multiple cover of one of these two.

b) For each pair of integers $0<2n<m$ there is (up to similarity) a unique non-planar elastica $\gamma_{m,n}$ which closes up in m periods of its curvature (k given by (0.1)) as it makes n windings around the z-axis (its axis of symmetry). All of these curves $\gamma_{m,n}$ are embedded and lie on embedded tori of revolution. The knot types represented by elastic curves in R^3 are precisely the (m,n)-torus knots satisfying $m>2n$. Any closed Euclidean elastic curve is one of the above.

1. Condition (C) for Total Squared Curvature

In this section we define a gradient flow for total squared curvature and show that this flow satisfies the appropriate compactness condition for applying the minimax principle in the infinite dimensional situation. We begin by recalling the Palais-Smale condition on a C^1 map $f:M \to R$, M a C^1 Riemannian Hilbert manifold:

(C) Given any sequence $\{s_n\}$ in M on which f is bounded but on which $\|\nabla f\|$ is not bounded away from zero there exists a convergent subsequence $\{s_{n_j}\}$.

Since the total squared curvature of a regular curve is unchanged under arbitrary reparametrizations it is apparently impossible to obtain condition (C) without restricting to a manifold of specially parametrized curves, e.g., arclength-parametrized curves. In fact it is convenient to consider the total squared curvature functional as being defined directly on the manifold of unit tangent fields, i.e., on curves in the unit sphere $S^2=\{x \in R^3: |x|=1\}$.

Observe that if γ is an arclength-parametrized curve in R^3 its total squared curvature is just twice the energy of its derivative; that is, if $\omega=\gamma':[0,1] \to R^3$ lies on the unit sphere then $F(\gamma) = \int_0^1 k^2 ds = \int_0^1 \langle \omega', \omega' \rangle dt = 2J(\omega)$. Furthermore, γ is a regular closed curve precisely when ω is a "balanced" closed curve, i.e., ω is a closed curve satisfying $\int_0^1 \omega dt=0$. Thus, the study of F on unit speed closed curves in R^3 is equivalent to the study of J on balanced closed curves in S^2. Finally, since J is invariant under rotations of S^2, it will suffice to fix a point $P \in S^2$ and study J on balanced curves in S^2 which begin and end at P.

Now it is well known that if M is a Riemannian manifold and $P,Q\in M$ then one can obtain condition (C) for $J:\Omega(P,Q)\to R$ where $\Omega(P,Q)$ is a certain Riemannian Hilbert manifold of curves in M which start at P and end at Q. Thus, setting $M=S^2$ and $P=Q$, we can view our goal in this section as that of showing that condition (C) persists upon restriction of J to the submanifold Ω_B of balanced curves in $\Omega=\Omega(P,P)$. This conclusion is by no means automatic even though the restriction is to a smooth submanifold of finite codimension (consider, e.g., the functional $f(x)=\langle x,x\rangle$ on an infinite dimensional Hilbert space, the unit sphere as a submanifold, and $\{s_n\}$ an orthonormal sequence).

The outline of the proof is as follows. We first show that $J:\Omega\to R$ satisfies a slightly stronger version of condition (C), condition $(\bar{C})$, which is the same except that the phrase "$\|\nabla f\|$ is not bounded away from zero" is replaced by "$\nabla f(s_n)$ has a Cauchy subsequence". We will be considering Ω as a submanifold of a Hilbert space H so the term "Cauchy" here refers simply to the norm on H (though one could also make sense of $(\bar{C})$ in an intrinsic way).

We then consider the restriction $J_B=J|_{\Omega_B}$ and a sequence $\{\omega_n\}$ in Ω_B for which $J_B(\omega_n)$ is bounded but $\|\nabla J_B(\omega_n)\|$ is not bounded away from zero, the objective being to satisfy the hypothesis of $(\bar{C})$ -- that $J(\omega_n)$ has a Cauchy subsequence. On achieving this we conclude that the subsequence converges to a critical point of J_B by virtue of Ω_B being closed and ∇J_B being continuous.

To this end we introduce auxilliary functionals $Y:\Omega\to R$, $Y(\omega)=\int_0^1\langle y,\omega\rangle dt$, one for each unit vector $y\in R^3$ and note that ω

is balanced precisely when $Y(\omega)=0$ for all y. In the spirit of Lagrange multipliers we then write

$$\nabla J(\omega_n) = \nabla J_B(\omega_n) + \lambda_n \nabla Y_n(\omega_n) \quad ,$$

the numbers $\lambda_n>0$ and unit vectors y_n being uniquely determined by the curves ω_n (we are assuming here that $\nabla J(\omega_n) \neq \nabla J_B(\omega_n)$). As will be seen below the sequence $\{\omega_n\}$ is bounded in $\Omega \subset H$ and a subsequence of $\{\lambda_n\}$ is bounded in R. Therefore, the proof will require just one additional ingredient, that the mapping $(y,\omega) \longmapsto \nabla Y(\omega) : S^2 \times \Omega \to H$ is compact.

We now begin to fill in the details. Let $H_0=L^2(I,R^3)$ be the Hilbert space of square integrable maps of the unit interval $I=[0,1]$ into R^3 with inner product $\langle u,v\rangle_0 = \int_0^1 \langle u(t),v(t)\rangle dt$, and let $H=H_1=H_1(I,R^3)$ be the Hilbert space of absolutely continuous maps $\omega: I \to R^3$ having square integrable first derivative, the inner product on H given by $\langle \omega,\eta\rangle_1 = \langle \omega,\eta\rangle_0 + \langle \omega',\eta'\rangle_0$.

Now consider the subsets $\Omega_B \subset \Omega \subset H$ defined by
$\Omega=\{\omega \in H: |\omega(t)| = 1 \text{ for all } t \text{ and } \omega(0)=\omega(1)=P\}$, $\Omega_B=\{\omega \in \Omega: \int_0^1 \omega(t)\,dt=0\}$.
According to the standard theory (see [7]), Ω is a closed C^∞ submanifold of H and by the implicit function theorem Ω_B is in turn a closed C^∞ submanifold of Ω of codimension three. Note that all the functionals $Y:\Omega \to R$ are C^∞ since they are restrictions of continuous linear maps on H. Note also that the C^∞ map $J:\Omega \to R$ is given simply by $2J(\omega)=\langle \omega,\omega\rangle_1 - 1$ and therefore a subset of Ω is bounded in H precisely when it is J-bounded (a fact we alluded to earlier).

We now wish to equip Ω with a natural Riemannian structure and consider the gradient of J relative to that metric. The metric we will use has been considered before (e.g., by Tromba [11]) for studying the geodesic problem.

Since we are considering S^2 as a submanifold of R^3 we can take advantage of the usual identification of the tangent spaces $S^2_{\omega(t)}$ with subspaces of R^3 and define spaces of "tangential H_k vectorfields w(t) along ω", k=0,1, by
$H_0(\omega^*TS^2)=\{w\in H_0(I,R^3): w(t)\in S^2_{\omega(t)}$ for almost all t} and
$H_1(\omega^*TS^2)=\{w\in H_1(I,R^3): w(t)\in S^2_{\omega(t)}$ for all t}.

Given $\omega\in\Omega$ and $w\in H_1(\omega^*TS^2)$ we let $\frac{Dw}{\partial t}$ denote the (almost everywhere defined) covariant derivative of w along ω. Here the covariant derivative is that of the standard metric on S^2 and can be written $\frac{Dw}{\partial t}=\Pi(\omega(t))w'(t)=w'(t)-\langle w'(t),\omega(t)\rangle\omega(t)$ with w' the ordinary derivative of w as a map into R^3 and $\Pi(\omega(t$ the orthogonal projection of S^3 onto $S^2_{\omega(t)}$. As is well known the tangent space at ω to the Hilbert manifold Ω can be identified with the space $\Omega_\omega=\{w\in H_1(\omega^*TS^2): w(0)=w(1)=0\}$. So we can now define our Riemannian metric $\langle\ ,\ \rangle_\omega$ on Ω_ω by $\langle w,v\rangle_\omega=\left\langle\frac{Dw}{\partial t},\frac{Dv}{\partial t}\right\rangle_0$.

Now set $\alpha=\nabla J(\omega)$, the gradient of J at ω (relative to the metric $\langle\ ,\ \rangle_\omega$). Then by definition $DJ(\omega)w=\langle\alpha,w\rangle_\omega=\left\langle\frac{D\alpha}{\partial t},\frac{Dw}{\partial t}\right\rangle_0$. On the other hand the formula for the differential of J is $DJ(\omega)w=\langle\omega',w'\rangle_0=\left\langle\omega',\frac{Dw}{\partial t}\right\rangle_0$. Thus, $\left\langle\frac{D\alpha}{\partial t}-\omega',\frac{Dw}{\partial t}\right\rangle_0=0$ for all $w\in\Omega_\omega$. The significance of this is apparent from the "Du Bois-Raymond lemma" (the version we give here is from Klingenberg [4], and applies to any Riemannian manifold M).

<u>Proposition 1.1</u> Let ω be an H_1 curve in M and let $v\in H_0(\omega^*TM)$. Suppose that $\left\langle v,\frac{Dw}{\partial t}\right\rangle_0=0$ for all $w\in H_1(\omega^*TM)$ satisfying w(0)=w(1)=0. Then v lies in $H_1(\omega^*TM)$ and is parallel, i.e $\frac{Dv}{\partial t}=0$ a.e.

<u>Proof</u>: Let $z\in H_1(\omega^*TM)$ be defined by $\frac{Dz}{\partial t}=v$, z(0)=0. In other words, in terms of local coordinates, $z=(z^1,\dots,z^n)$ is the

unique solution to the linear system with H_1 coefficients

$$\frac{d}{dt} z^k + \sum_{i,j=1}^{n} (\frac{d}{dt} \omega^i) \Gamma_{ij}^k z^j = v^k, \quad z^k(0)=0, \quad k=1,2,\ldots,n.$$

Let $x \in H_1(\omega^*TM)$ be defined by $\frac{Dx}{\partial t} = 0$, $x(1)=z(1)$. Then $w(t) = z(t) - tx(t)$ lies in $H_1(\omega^*TM)$ and satisfies $w(0)=w(1)=0$. Thus we have $\langle v-x, v-x\rangle_0 = \langle v, v-x\rangle_0 - \langle x, v-x\rangle_0 = \langle v, \frac{Dw}{\partial t}\rangle_0 - \int_0^1 (\frac{d}{dt}\langle x,w\rangle)dt = 0$. So $v=x$ a.e. ///

We can now conclude that there exists $x \in H_1(\omega^*TS^2)$ such that

$$\frac{D\alpha}{\partial t} = \omega' + x, \qquad \frac{Dx}{\partial t} = 0 \qquad (1.1)$$

Recall our goal is to get the curves ω_n to converge in $\Omega \subset H$. For this we will need the x_n to converge in H_0. Here an elementary comparison between the covariant derivative and the ordinary derivative is called for:

<u>Proposition 1.2</u> If A is an H_1 bounded subset of Ω then there exists a constant C such that, for any $\omega \in A$ and any $w \in H_1(\omega^*TS^2)$,

$$\|w\|_1 \leq C(\|\frac{Dw}{\partial t}\|_0 + \|w\|_0).$$

<u>Proof</u>: As noted earlier w' can be resolved into components tangential and normal to S^2, $w' = \frac{Dw}{\partial t} + \langle w', \omega\rangle\omega$. Since $\frac{d}{dt}\langle \omega, w\rangle = 0$ we can rewrite this as $w' = \frac{Dw}{\partial t} - \langle w, \omega'\rangle\omega$. On A we have an H_0 bound on ω' and a C^0 bound on ω (since H_1 is continuously embedded in C^0), so $\|w'\|_0 \leq C_1(\|\frac{Dw}{\partial t}\|_0 + \|w\|_0)$ for some constant C_1. Therefore, for a new constant C,

$$\|w\|_1 = (\|w'\|_0^2 + \|w\|_0^2)^{\frac{1}{2}} \leq C(\|\frac{Dw}{\partial t}\|_0 + \|w\|_0). \qquad ///$$

It is now a simple matter to complete the first step in the outline:

<u>Proposition 1.3</u> $J:\Omega \to R$ satisfies condition $(\bar{C})$.

<u>Proof</u>: Let $\{\omega_n\}$ be a J bounded hence H_1 bounded sequence for which $\alpha_n = \nabla J(\alpha_n)$ converges in H_1. Then $\omega_n' = \frac{D\alpha_n}{\partial t} - x_n$ for certain parallel fields x_n in H_1. Since $\{\alpha_n'\}$ is H_0 bounded

and since $|\frac{D\alpha}{\partial t}| \le |\alpha'|$ for all t we have an H_0 bound on $|\frac{D\alpha_n}{\partial t}|$. As we also have an H_0 bound on $\{\omega_n'\}$ this gives an H_0 bound on $\{x_n\}$. Proposition 1.2 now implies that $\{x_n\}$ is in fact H_1 bounded. Since H_1 is compactly embedded in H_0 we can therefore assume that $\{x_n\}$ converges in H_0.

By hypothesis we have $\{\alpha_n'\}$ hence $\{\frac{D\alpha_n}{\partial t}\}$ converging in H_0. Combining this with (1.1) and the previous paragraph we conclude $\{\omega_n'\}$ converges in H_0. Since $\{\omega_n\}$ is H_1 bounded we can also assume $\{\omega_n\}$ converges in H_0. So $\{\omega_n\}$ converges in H_1. ///

We now carry out a similar analysis of the gradient of Y. Set

$$\beta = \nabla Y(\omega), \quad \hat{y}(t) = \Pi(\omega(t))\omega = \omega - \langle y, \omega(t)\rangle \omega(t) \qquad (1.2)$$

On the one hand $\langle \frac{D\beta}{\partial t}, \frac{Dw}{\partial t}\rangle_0 = DY(\omega)w$ for all w in Ω_ω and on the other hand $DY(\omega)w = Y(w) = \int_0^1 \langle y, w\rangle dt = \int_0^1 \langle \hat{y}, w\rangle dt$. Define $z \in H_1(\omega^* TS^2)$ b

$$\frac{Dz}{\partial t} = -\hat{y}, \quad z(0) = 0 \qquad (1.3)$$

Since any $w \in \Omega_\omega$ satisfies $w(1)=0$ integration by parts now gives $DY(\omega)w = \langle z, \frac{Dw}{\partial t}\rangle_0$, hence $\langle \frac{D\beta}{\partial t} - z, \frac{Dw}{\partial t}\rangle_0 = 0$ for all w in Ω_ω. Therefore, by Proposition 1.1, there exists $\underset{\sim}{x} \in H_1(\omega^* TS^2)$ such that

$$\frac{D\beta}{\partial t} = z + \underset{\sim}{x}, \quad \frac{D\underset{\sim}{x}}{\partial t} = 0 \qquad (1.4)$$

Using formulas (1.1) - (1.4) we can easily prove smoothness of critical points of J_B as well as the compactness statement mentioned above:

<u>Proposition 1.4</u> Let $\omega \in \Omega_B$ be a critical point of J_B, i.e., $0 = \nabla J_B(\omega) = \nabla J(\omega) - \lambda \nabla Y(\omega)$ for some λ, y. Then ω is C^∞.

<u>Proof</u>: Using the above notation we have $\alpha = \lambda\beta$. Taking $\frac{D}{\partial t}$ gives $\omega' = (\lambda(z + \underset{\sim}{x}) - x)$, so ω' is actually in H_1. But the H_0 function ω'' can now be written $\omega'' = \frac{D\omega'}{\partial t} - \langle \omega', \omega'\rangle \omega = -\lambda\hat{y} - \langle \omega', \omega'\rangle \omega$, so in fact ω'' is itself in H_1. The result follows now from the formula for ω''. ///

Proposition 1.5 Let $\{y_n\}$ be a sequence in S^2 and let $\{\omega_n\}$ be a bounded sequence in Ω. Then $\beta_n = \nabla Y_n(\omega_n)$ has a subsequence which converges in H_1. I.e., the map $(y,\omega)\mapsto \nabla Y(\omega): S^2\times\Omega\to H$ is compact.

Proof: By (1.3), (1.4) we have an H_0 bound on second covariant derivatives: $\|\frac{D^2\beta_n}{\partial t^2}\|_0 = \|\hat{y}_n\|_0 \le 1$. Meanwhile, we can obtain an H_0 bound on the first covariant derivative as follows. Since

$$\|\beta\|_0^2 = \int_0^1\langle\beta,\beta\rangle du = \int_0^1\int_0^u\left(\frac{d}{dt}\langle\beta,\beta\rangle\right)dtdu = 2\int_0^1\int_0^u\langle\beta,\frac{D\beta}{\partial t}\rangle dtdu \le 2\|\beta\|_0\|\frac{D\beta}{\partial t}\|_0$$

we have $\|\beta\|_0 \le 2\|\frac{D\beta}{\partial t}\|_0$. Therefore, $\langle\frac{D\beta}{\partial t},\frac{D\beta}{\partial t}\rangle_0 = DY(\omega)\beta = \int_0^1\langle\hat{y},\beta\rangle dt$ $\|\beta\|_0 \le 2\|\frac{D\beta}{\partial t}\|_0$. So $\|\frac{D\beta_n}{\partial t}\|_0 \le 2$.

Proposition 1.2 now implies that $\frac{D\beta_n}{\partial t}$ is H_1 bounded. So we can assume $\frac{D\beta_n}{\partial t}$ converges in H_0. Further use of Proposition 1.2 now implies the result. ///

It remains only to prove

Proposition 1.6 Suppose $\{\omega_n\}$ is a sequence of curves in Ω_B such that $J_B(\omega_n)$ and $\|\nabla J_B(\omega_n)\|_1$ are both bounded sequences. Then (a subsequence of) the sequence $\{\lambda_n\}$ defined by the equation $\nabla J(\omega_n) = \nabla J_B(\omega_n) + \lambda_n\nabla Y_n(\omega_n)$ is also bounded.

Proof: The sequence $\|\alpha_n\|_{\omega_n} = \|\nabla J(\omega_n)\|_\omega$ is bounded since

$\langle\alpha_n,\alpha_n\rangle_{\omega_n} = DJ(\omega_n)\alpha_n = \int_0^1\langle\omega_n',\frac{D\alpha_n}{\partial t}\rangle dt \le \|\omega_n\|_0\|\alpha_n\|_{\omega_n}$. Therefore, $\|\lambda_n\beta_n\|_{\omega_n} = \|\lambda_n\nabla Y_n(\omega_n)\|_{\omega_n}$ is also bounded.

But $|\lambda_n| = \dfrac{\|\lambda_n\beta_n\|_{\omega_n}}{\|\beta_n\|_{\omega_n}}$, so it will suffice to bound $\|\beta_n\|_{\omega_n}$ away from zero. The strategy for doing this is to compare β_n with $\hat{y}_n$ (which also tends to push ω_n in the direction of increasing Y), and to bound $\|\hat{y}_n\|_0$ away from zero.

Let us first see how to bound (a subsequence of) $\|\hat{y}_n\|_0$ away from zero. Since ω_n is H_1 bounded we can assume it converges in C^0 to some balanced curve ω. For each $y \in S^2$, $\hat{y}(t) = \Pi(\omega(t))y$ vanishes at no more than two points $\omega(t)$, and since ω is balanced and continuous it follows that $\|\hat{y}\|_0 > 0$. By compactness, $\|\hat{y}_n\|_0$ must therefore be uniformly bounded away from zero as y varies over S^2. Finally, continuity of the functionals Y on C^0 allows us to conclude that a positive lower bound exists for $\|\hat{y}_n\|_0$.

To compare β_n with $\hat{y}_n$ we first subtract from the $\hat{y}_n$ certain fields $R_n(t)$ obtained by restricting to ω_n a rotation field on S^2, so chosen as to put $\hat{y}_n - R_n$ in Ω_ω, i.e., so that $\hat{y}_n(0) = R_n(0)$ (hence also $\hat{y}_n(1) = R_n(1)$). Clearly this can be done so that R_n has maximum length $\hat{y}_n(0) \leq 1$.

Now balanced curves have the property that rotations of S^2 do not affect any of the functionals Y. Therefore, we can write $\langle \beta_n, \hat{y}_n - R_n \rangle_{\omega_n} = DY_n(\omega_n)[\hat{y}_n - R_n] = DY_n(\omega_n)\hat{y}_n = \langle \hat{y}_n, \hat{y}_n \rangle_0$ and hence

$$\|\beta_n\|_{\omega_n} \geq \frac{\langle \beta_n, \hat{y}_n - R_n \rangle_{\omega_n}}{\|y_n - R_n\|_{\omega_n}} = \frac{\|\hat{y}_n\|_0^2}{\|\hat{y}_n - R_n\|_{\omega_n}} .$$

It remains to obtain upper bounds for $\|\hat{y}_n\|_{\omega_n}$ and $\|R_n\|_{\omega_n}$. One easily sees that $\frac{D\hat{y}}{\partial t} = -\langle y, \omega \rangle \omega'$ and hence that $\langle \hat{y}, \hat{y} \rangle_\omega = \int_0^1 \langle \frac{D\hat{y}}{\partial t}, \frac{D\hat{y}}{\partial t} \rangle dt = \int_0^1 \langle y, \omega \rangle^2 \langle \omega', \omega' \rangle \, dt \leq 2J(\omega)$. Finally, since $R_n(t)$ is the restriction of a rotation field $\bar{R}: R^3 \to R^3$ with maximum length on S^2 less than one, the chain rule gives $\|R_n\|_{\omega_n} = \|\frac{DR_n}{\partial t}\|_0 \leq \|R_n'\|_0 \leq \|D\bar{R}_n \cdot \omega_n'\|_0 \leq \|\omega_n'\|_0 = 2J(\omega_n)$. ///

Thus we have proved

<u>Theorem 1.7</u> $J_B : \Omega_B \to R$ satisfies the Palais-Smale condition (C)

By virtue of the correspondence described at the beginning of this section and Proposition 1.4 we have the immediate

<u>Corollary 1.8</u> The total squared curvature F achieves its infimum in each regular homotopy class of unit length immersions $\gamma : S^1 \to R^2$. Thus, for each integer p there exists a closed C^∞ elastica γ_p which minimizes F among all regular closed curves in R^2 having rotation index p and length one.

Of course, these elastica were already known from the classification Theorem 0.1; for $p \neq 0$, γ_p is the p-fold circular elastica, and with p=0 we have thus recovered the figure eight elastica by "topological" means and established that it globally minimizes F in its regular homotopy class.

Actually, a much shorter compactness argument would have sufficed (in place of Theorem 1.7) for the corollary. On the other hand, all immersions of S^1 into R^3 are known to be regularly homotopic (see [10]), so minimization of F in the non-planar case would yield only the circle. Thus, for the purpose of studying non-planar elastica the gradient technique is essential.

Even for the planar case the additional dynamical information which Theorem 1.7 provides is noteworthy:

<u>Corollary 1.9</u> The "curve straightening flow" induced by $-\nabla J_B$ deforms every (unit length) immersion $\gamma : S^1 \to R^2$ of rotation index one to the circle γ_1 in the infinite time limit. More generally, the trajectory through an immersion $\gamma : S^1 \to R^2$ of rotation index p converges to γ_p (in the case p=0 the trajectory may also converge to a multiple cover of γ_p).

<u>Proof</u>: This is immediate from Theorem 0.1, Theorem 1.7, and Theorem 4.1 of [8]. ///

The corresponding question for the "curve <u>shortening</u>" flow is still open (for that problem one begins with the arclength functional L rather than F and studies the dynamics of kN, the negative L^2-gradient of arclength). However, the convex case has been settled, that is, it is known that a convex curve curve becomes circular in the limit as it shrinks down to a point [2]. The argument depends on the fact that a convex curve remains convex under the curve shortening flow.

We conjecture that the curve straightening flow also preserves convexity. Indeed this is easy to prove if γ_0 happens to be symmetric with respect to the reflections about x and y axes, or with respect to rotation by π. For in this case the corresponding trajectory ω_w of $-\nabla J_B$ satisfies $\nabla J_B(\omega_w) = \nabla J(\omega_w)$, i.e., $\lambda_w \equiv 0$. Making use of this fact a simple computation shows that the curvature k of γ evolves according to $\frac{dk}{dw} = -k+2\pi$.

We conclude this section with some remarks concerning variations of Theorem 1.7:

1) Very little of the above has to be altered to cover the boundary value problem for elastica. The basic formulas and propositions remain essentially the same (but the proof of Proposition 1.6 becomes slightly more subtle). In particular, condition (C) holds for the following problem. Fix a length L, two unit vectors P,Q, and a point $\eta=(0,0,\ell)\in R^3$ with $\ell<L$. We then consider the functional $J_\eta:\Omega_\eta(P,Q)\to R$ where $\Omega_\eta(P,Q)=\{\omega\in H_1([0,L],S^2): \omega(0)=P$, $\omega(L)=Q$, and $\int_0^L\omega(t)dt=\eta\}$ and $J_\eta(\omega)=\int_0^L\langle\omega',\omega'\rangle dt$. Note that if ω is a critical point of J_η and if $\gamma:[0,1]\to R^3$ is an arclength parametrized curve satisfying $\gamma'=\omega$ then γ is an elastica of length L satisfying $\gamma'(0)=P, \gamma'(L)=Q$, and (after translation) $\gamma(0)=0$, $\gamma(L)=\eta$. One can also allow one or both of P,Q to be variable.

2) Even though condition (C) does not hold for either the arclength

functional L or the unconstrained total squared curvature functional F the combination $F^{\varepsilon}=F+\varepsilon L$, $\varepsilon>0$, is well behaved. In fact, given a regular closed curve $\gamma:[0,1]\to R^3$ we can parametrize γ by constant speed $L=L(\gamma)$ and associate with γ the pair $(\frac{\gamma'}{L}, L)\in \Omega_B\times R$. Then the functional $J_B^{\varepsilon}:\Omega_B\times R\to R$ defined by $J_B^{\varepsilon}(\omega,L)=\frac{1}{L}J_B(\omega)+\varepsilon L$ satisfies $F^{\varepsilon}(\gamma)=J_B^{\varepsilon}(\frac{\gamma'}{L}, L)$, and using Theorem 1.7 J_B^{ε} is easily shown to satisfy condition (C).

2. The Second Variation of J_B

Following the notation of the previous section we associate with any $\omega\in\Omega_B$ the unique number $\lambda>0$ and unit vector y satisfying $\nabla J(\omega) = \nabla J_B(\omega) + \lambda\nabla Y(\omega)$ (unless $\nabla J(\omega)=\nabla J_B(\omega)$, in which case we set $\lambda=0$). If ω happens to be a critical point of $J_B:\Omega_B \to R$ then ω is smooth and has at each point a velocity vector $V=\omega'$, speed $v=|V|$, and Frenet frame $\{T,N\}$. Letting R be the Riemann curvature tensor on the standard sphere S^2 (with sign convention $\langle R(U,W)W,U\rangle\geq 0$), we have the following formula for the second variation of J_B:

Proposition 2.1 Let ω be a critical point of J_B and let U,W be in $T_\omega\Omega_B$. Then

$$D^2J_B(\omega)(U,W) = -\int_0^1\langle U, \frac{D^2W}{\partial t^2} + R(W,V)V - \lambda\langle y,\omega\rangle W\rangle dt$$
$$= -\int_0^1\langle U, \frac{D^2W}{\partial t^2} + v^2\langle W,N\rangle N - \lambda\langle y,\omega\rangle W\rangle dt.$$

Proof: To prove the first formula (from which the second follows trivially) we consider a two parameter variation $\bar\omega(t)=\omega(u,w)(t)$ of ω within Ω_B and corresponding variation fields $\bar U$, $\bar W$ and velocity field $\bar V$ extending U,W,V, respectively.

The first variation of J_B at $\bar\omega$ in the $\bar U$ direction is simply $DJ_B(\bar\omega)\bar U=\int_0^1\langle\frac{D\bar U}{\partial t},\frac{D\bar\alpha}{\partial t}\rangle dt$ where $\bar\alpha=\nabla J(\bar\omega)$. But to get our second variation formula we must take advantage of the fact that the vector $\lambda\bar\beta=\lambda\nabla Y(\bar\omega)$ is always orthogonal to Ω_B and thus we can write instead $DJ_B(\bar\omega)\bar U = \int_0^1\langle\frac{D\bar U}{\partial t},\frac{D\bar\alpha}{\partial t}-\lambda\frac{D\bar\beta}{\partial t}\rangle dt.$

Before taking another derivative we should clear up a significant notational ambiguity. The vector y and the number λ are fixed, i.e., belong to $\omega(t)$ rather than $\bar\omega(t)$; thus, the above formula is not simply another way of writing the identity

$DJ_B(\bar{\omega})\bar{U} = \langle \bar{U}, \nabla J_B(\bar{\omega})\rangle_{\bar{\omega}}$.

Since $0=\alpha-\lambda\beta=\bar{\alpha}-\lambda\bar{\beta}|_{(0,0)}$ we can now write (using formulas (1.1)-(1.4))

$$D^2 J_B(\omega)(U,W) = \int_0^1 \frac{d}{dw}\left\langle \frac{D\bar{U}}{\partial t}, \frac{D\bar{\alpha}}{\partial t} - \lambda\frac{D\bar{\beta}}{\partial t}\right\rangle dt\Big|_{v=w=0} = \int_0^1 \left\langle \frac{DU}{\partial t}, \frac{D}{\partial w}(\bar{V}+\bar{x}-\lambda(\bar{z}+\bar{\chi}))\right\rangle dt\Big|_{w=0}$$

$$= \int_0^1 \left\langle \frac{DU}{\partial t}, \frac{D\bar{V}}{\partial w}\right\rangle - \left\langle U, \frac{D}{\partial t}\frac{D}{\partial w}(\bar{x}-\lambda(\bar{z}+\bar{\chi}))\right\rangle dt\Big|_{w=0}$$

[Note that integration by parts for the second term is valid because the fields $\bar{x}$, $\bar{z}$, $\bar{\chi}$ are all in H_1 along each $\bar{\omega}$. On the other hand, $\bar{V}$ is known to be in H_1 only along the critical point ω, so partial integration of the first term has to wait until w is set equal to zero].

$$= \int_0^1 \left\langle \frac{DU}{\partial t}, \frac{DW}{\partial t}\right\rangle - \langle U, R(V,W)(x-\lambda(z+\chi))\rangle dt - \int_0^1 \left\langle U, \lambda\frac{D}{\partial w}(y-\langle y,\bar{\omega}\rangle\bar{\omega})\right\rangle dt\Big|_{w=0}.$$

Observe that $\frac{d}{dw}(y-\langle y,\bar{\omega}\rangle\bar{\omega}) = -\langle y,\bar{w}\rangle\bar{\omega} - \langle y,\bar{\omega}\rangle\bar{w}$, and the covariant derivative just takes the tangential part of this. Noting also that $x-\lambda(z+\chi)=-V$, the formula follows. ///

To prove our main result on instability it will be important to observe that the self-adjoint operator naturally associated to the above Hessian is of the form identity plus compact:

<u>Proposition 2.2</u> Let ω be a critical point of J_B. Then the self-adjoint operator $K: T_\omega\Omega_B \to T_\omega\Omega_B$ satisfying $D^2 J_B(\omega)(U,W) = \langle U, W+K(W)\rangle_\omega$ is compact.

<u>Proof</u>: From Proposition 2.1 we see that K satisfies $\langle U,K(W)\rangle_\omega = \langle U,-v^2\langle W,N\rangle N+\lambda\langle y,\omega\rangle W\rangle_0$. Therefore, $\langle U, K(W)\rangle_\omega^2 \le C\langle U,U\rangle_0\langle W,W\rangle_0$ where C depends only on ω. From Section 1 we know that any X in $T_\omega\Omega_B$ satisfies $\langle X,X\rangle_0 \le 4\langle X,X\rangle_\omega$

so we have, setting $U=K(W)$, $\langle K(W),K(W)\rangle_\omega \le 4C\langle W,W\rangle_0$. Since K is linear and since the Hilbert space $T_\omega\Omega_B$ is compactly embedded in H_0 it follows that K is compact. ///

For concrete computations we write W in terms of the Frenet frame for ω, $W=fT+gN$. Letting k_ω denote the geodesic curvature of $\omega\subset S^2$ (so $\frac{DT}{\partial t}=vk_\omega N$) and substituting into Prop.2.1 one easily obtains:

<u>Proposition 2.3</u> $$D^2J_B(\omega)(W,W)=-\int_0^1 f[f''-2vk_\omega g'-(v^2k_\omega^2+\lambda\langle y,\omega\rangle)f] +g[g''+2vk_\omega f'-(v^2(k_\omega^2-1)+\lambda\langle y,\omega\rangle)g]dt.$$

Since we are interested in the stability of elasticae it is still more convenient for most applications to write the formula in terms of the curvature k and torsion τ of the elastica $\gamma:I\to R^3$

<u>Proposition 2.4</u> $$D^2J_B(\omega)(W,W)=-\int_0^1 f[f''-2\tau g'+(\tfrac{1}{2}(k^2-\mu)-\tau^2)f] +g[g''+2\tau f'+(\tfrac{1}{2}(3k^2-\mu)-\tau^2)g]dt,$$

where μ is given by formula (0.2).

<u>Proof</u>: Noting that $v=k$ and the Frenet frame $\{T_\gamma,N_\gamma,B_\gamma\}$ for γ is given by $T_\gamma=\omega$, $N_\gamma=T$, $B_\gamma=N$, one easily checks that $vk_\omega=\tau$. Taking $\frac{D^2}{\partial t^2}$ of the identity $\alpha=\lambda\beta$ gives $v'T+v^2k_\omega N=-\lambda\hat{y}$, hence $(\frac{dk}{ds})N_\gamma+k\tau B_\gamma=-\lambda(y-\langle y,T_\gamma\rangle T_\gamma)$, where s is the arclength parameter on γ From Section 0 we know that $J_0=(k^2-\mu)T_\gamma+2(\frac{dk}{ds})N_\gamma+2k\tau B_\gamma$ extends to a constant field on R^3. It follows by comparison with the previous line that $-\frac{J_0}{2}=\lambda y$. The formula is now easily obtained from Proposition 2.3 using the substitutions $v=k$, $vk_\omega=\tau$, $\langle\lambda y,\omega\rangle=\frac{(\mu-k^2)}{2}$

In applying Propositions 2.3 and 2.4 one must keep in mind that f and g are not arbitrary functions vanishing at 0 and 1; for W is supposed to be tangent to Ω_B, i.e., $\int_0^1 W dt=0$. Of course, this complicates the analysis of the already complicated looking

$D^2J_B(\omega)$. Fortunately, the formula simplifies drastically in certain special cases of interest, e.g., if γ is planar, if $\lambda=0$, if we let one of f or g be zero. Each of these possibilities will occur in one of the following

Examples: 1) Let γ_p be the p-fold circular elastica (lying in the x,y-plane) having curvature $2\pi p$. Thus, $\omega_p=\gamma_p'$ is the p-fold geodesic $\omega_p=(\cos(2\pi pt),\sin(2\pi pt),0)$ having velocity $2\pi p$. Since ω_p is already a critical point of the unconstrained functional J we have $\lambda=0$. So Proposition 2.3 gives
$D^2J_B(\omega_p)(W,W)=-\int_0^1 ff'' + gg'' + (2\pi p)^2g^2dt.$

Let us first set $g\equiv 0$. Then $D^2J_B(\omega_p)(W,W)=\int_0^1(f')^2dt>0$, corresponding to the fact that γ_p is stable with respect to planar variations.

Now let $f\equiv 0$, and write $g(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty} a_n\cos(2\pi nt)+b_n\sin(2\pi nt)$.

Then the tangency condition $\int_0^1 Wdt$ is simply $a_0=0$. Substitution now leads to $D^2J_B(\omega_p)(W,W)=\sum_{n=1}^{\infty}(2\pi)^2(a_n^2+b_n^2)(n^2-p^2)$. For p=1 this quantity is positive (ignoring pure rotations of S^2), and for p>1 this quantity can be made negative (simply choose $a_n=0$ $\forall n$, $b_1=1$, and $b_n=0$ for n>1). This corresponds to the fact that γ_1 is stable (even for non-planar variations) while γ_p, p>1, is unstable as an elastica in R^3.

2) In this example we show that an elastica consisting of at least two full turns of a helix is unstable as a solution to the boundary value problem (see Remark 1 at the end of Section 1). A unit speed parametrization $\gamma:[0,L]\to R^3$ for exactly two turns of a helix of curvature k and torsion τ is given by $\gamma(t)=(\frac{k}{a^2}\cos(at), \frac{k}{a^2}\sin(at), \frac{\tau}{a}t)$, where $a^2=k^2+\gamma^2$, $L=\frac{4\pi}{a}$.

The corresponding two-fold circle $\omega(t)=\gamma'(t)=\frac{k}{a}(-\sin(at),\cos(at),\frac{\tau}{k}$ is a critical point of the functional $J_\eta:\Omega_\eta(P,P)\to R$ (here $P=\omega(0)=(0,\frac{k}{a},\frac{\tau}{a})$ and $\eta=(0,0,\frac{4\pi\tau}{a^2})$).

Proposition 2.4 applies to J_η if we simply replace 1 with L in the upper limit of integration. For helices the modulus p is zero (see Figure 1) so formulas (0.1) and (0.2) imply $\mu=k^2-2\tau^2$. It follows that the second variation of J_η at ω in the direction $W=fT+gN$ can be written $D^2J_\eta(\omega)(W,W)=\int_0^L (f')^2+(g')^2-k^2g^2-4\tau f'g\, d$

Since the unit tangent and normal vectors to ω are given by $T=(-\cos(at),-\sin(at),0)$ and $N=\frac{1}{a}(\tau\sin(at),-\tau\cos(at),k)$, the condition $\int_0^L Wdt=0$ for tangency of W to Ω_η can be written as three equations $\int_0^L f\cos(at)dt=\frac{\tau}{a}\int_0^L g\sin(at)dt$, $\int_0^L f\sin(at)dt=-\frac{\tau}{a}\int_0^L g\cos(a$ and $\int_0^L gdt=0$.

The instability of ω (hence of γ) can now be seen by checking that , for example, the choice $W=4\tau(\cos(\frac{a}{2}t)-1)T-a\sin(\frac{a}{2}t)N$ gives $D^2J_\eta(\omega)(W,W)=-\frac{3\pi a}{2}(5\tau^2+k^2)<0$ and is compatible with the tangency conditions.

For the helix one can also determine explicitly all Jacobi fields and after a lengthy computation one obtains a transcendenta equation describing the occurrence of conjugate points. It can be seen in this way that the distribution of conjugate points depends in a rather complicated way on k and τ but that in any case a conjugate point will occur after less than two turns of the helix.

3) We show here that the figure eight elastica is unstable in R^3. Considering a normal variation $W=gN$, we see from Proposition 2.4 that $D^2J_B(\omega)(W,W)=-\int_0^1 g(g''+\frac{1}{2}(3k^2-\mu)g)dt$.

The squared curvature of the figure eight can be written

$k^2(t)=\alpha cn^2(rt+K,p)$; here we have translated the argument rt by K, the complete elliptic integral of the first kind. For the figure eight we know also that $w=p$ (see Figure 1) so formulas (0.1) and (0.2) give $r^2=\frac{\alpha}{4p^2}$, $\mu=\alpha(1-\frac{1}{2p^2})$. Setting $u=rt+K$ and $h(u)=rg(t)$, the above integrand becomes $h(u)\cdot Lh(u)$, where L is the Lamé operator given by $Lh=h''+(4p^2+1-6p^2 sn^2 u)h$.

Note that the figure eight closes up in one period of k, i.e., after u changes by $4K$. Thus, we can write $D^2 J_B(\omega)(W,W)=-\frac{1}{r}\int_K^{3K} h(u)\cdot Lh(u)du$. Therefore, it suffices to find a positive eigenvalue of the Lamé operator L belonging to an eigenfunction h which vanishes at endpoints and integrates to zero (for then the corresponding g will satisfy $g(1)=g(0)=0$, as well as the tangency condition $0=(\int_0^1 gdt)N=\int_0^1 Wdt$).

Consider the function $h(u)=cn(u)dn(u)$, where $dn(u)$ is the elliptic function $dn(u)=\sqrt{1-p^2 sn^2 u}$. Then $h(K)=h(3K)=0$, and since $h(u+2K)=-h(u)$ we also know that h integrates to 0. Finally, using the standard formulas $\frac{d}{du}sn\,u=cn\,u\;dn\,u$, $\frac{d}{du}cn\,u=-sn\,u\;dn\,u$, and $\frac{d}{du}dn\,u=-p^2 cn\,u\;sn\,u$, one readily obtains $Lh=3p^2h$.

In the next section we will be applying the minimax principle to the restriction of J_B to a certain submanifold of Ω_B consisting of curves having additional symmetry. We will need to know that the multiply covered geodesics ω_p are stable with respect to the corresponding symmetric variations (even though, as seen in the first example, they are unstable with respect to general variations in Ω_B).

To be specific, let us take as our base point P a point on the equator of S^2, let $q \geq p > 0$ be integers, let $m = p+q$, and let G be the group generated by R_θ, the rotation about the z-axis by an angle $\theta = \frac{2\pi p}{p+q} = \frac{2\pi p}{m}$. The submanifold of Ω_B which we wish to consider can now be defined by $\overline{\Omega}_B = \{\omega \in \Omega_B : \omega(t+\frac{1}{m}) = R_\theta(\omega(t)) \text{ for all } t \in I\}$. We note that the tangent space to $\overline{\Omega}_B$ at ω is descibed by

$$T_\omega \overline{\Omega}_B = \{W \in T_\omega \Omega_B : W(t+\tfrac{1}{m}) = DR_\theta \cdot W(t),\ t \in I\}$$
$$= \{(fT+gN) \in T_\omega \Omega_B : f(t+\tfrac{1}{m}) = f(t),\ g(t+\tfrac{1}{m}) = g(t),\ t \in I\}.$$

<u>Proposition 2.5</u> Let $\overline{J}_B$ be the restriction of J_B to $\overline{\Omega}_B$. Then ω_p and ω_{-q} are strict relative minima for $\overline{J}_B$. In fact the Hessian of $\overline{J}_B$ is positive definite at ω_p and ω_{-q}, i.e., for $r = p, -q$ and for all W in $T_{\omega_r}\overline{\Omega}_B$, $D^2\overline{J}_B(\omega_r)(W,W) \geq \frac{1}{m^2}\langle W,W\rangle_{\omega_r}$.

<u>Proof</u>: Since ω_r is already a critical point of the "extended" functional J_B, we can simply take the second variation formula for J_B and apply it to vectors W in $T_{\omega_r}\overline{\Omega}_B$.

As in Example 1 we set $W = fT+gN$ and consider the Fourier series for $g(t)$, the only difference being that $g(t)$ is now periodic with period $\frac{1}{m}$:
$g(t) = \sum_{n=1}^{\infty} a_n \cos(2\pi mnt) + b_n \sin(2\pi mnt)$. This time substitution gives

$$D^2\overline{J}_B(\omega_r)(W,W) = \int_0^1 (f')^2 dt + \sum_{n=1}^{\infty} (2\pi)^2 (a_n^2+b_n^2)(m^2n^2-r^2) \geq \frac{1}{m^2}\int_0^1 (f')^2+(g')^2 dt = \frac{1}{m^2}\langle W,W\rangle$$

Taylor's theorem now implies that ω_r is a strict local minimum for $\overline{J}_B$. //

3. The Minimax Argument

A simple description of the minimax principle begins with a connected manifold M, a smooth function $f:M\to R$, and two local minima $x,y\in M$ for f. Intuitively, one expects to obtain a third critical point $z\in M$ by considering the set Λ of all continuous paths $\xi:I\to M$ joining x to y and setting $\mathrm{Minimax}(f,\Lambda)= \inf_{\xi\in\Lambda}\sup_{t\in I} f(\xi(t))$; the number $\mathrm{Minimax}(f,\Lambda)$ should be a critical value of f belonging to some unstable, i.e., "saddle type" critical point $z\in M$.

In this section we set $f=\overline{J}_B$, $x=\omega_p$, and $y=\omega_{-q}$. It is not hard to see that x and y are G-equivariantly homotopic among balanced curves in S^2, so it makes sense to let M be the component of $\overline{\Omega}_B$ containing x and y. We wish to see now that the above picture is valid in our case.

To begin with, the conclusion that $\mathrm{Minimax}(f,\Lambda)$ is indeed a critical value of f is justified since f satisfies condition (C), precisely the condition which enables one to extend to the infinite dimensional setting the key lemma on deforming M downward via the flow of $-\nabla f$ (see [8] for a general discussion of the minimax principle). Actually, Theorem 1.7 asserts only that condition (C) holds for J_B, not its restriction f. But observe that J_B is always tangent to the submanifold $M\subset\Omega_B$; thus, unlike the restriction of J to Ω_B, the restriction of J_B to M trivially preserves condition (C). The same observation also implies that the resulting critical point z of f is in fact a critical point of J_B (one can also view this as a simple case of the principle of symmetric criticality [9]).

Still we have gained nothing unless we can show that this minimax critical point is not x or y. For this it would suffice of course to show that $f(z)=\text{Minimax}(f,\Lambda)$ is greater than both $f(x)$ and $f(y)$. Such a statement would be obvious if we knew that there existed neighborhoods O_x, O_y of x,y, respectively, such that $f(u)>f(x)$ for all $u\in\partial O_x$ and $f(u)>f(y)$ for all $u\in\partial O_y$. But this follows easily from Proposition 2.5; for given the positive definiteness of D^2f at x and y one can either appeal to the Morse-Palais lemma (since x,y must be non-degenerate) or one can argue directly from Taylor's theorem.

Theorem 3.1 For each pair of integers $q\geq p>0$ there exists a non-circular closed elastica $\gamma_{p,q}$ in R^3 which is G-symmetric and G-equivariantly regularly homotopic to γ_p, the p-fold circular elastica, and G is the group generated by rotation about the z axis by an angle $\theta=\frac{2\pi p}{p+q}$. For distinct relatively prime pairs p,q the $\gamma_{p,q}$ are geometrically distinct nonplanar elastic curves

Proof: The theorem follows essentially from the above observations together with the correspondence between critical points of $J_B:\Omega_B\to R$ and closed elastic curves in R^3. However, one must make the following additional observations. First, the minimax critical point z is not itself a geodesic since x,y are the only geodesics in M. Second, the only non-circular closed elastica in the plane is a figure eight curve having a Z_2 symmetry (see Theorem 0.1), so z can be planar only in the case p=q. ///

Comparison with Theorem 0.1 shows that we have thus recove all closed elastic curves in R^3 by our symmetrical minimax argument. The goal now is to show that all of these critical points are of saddle type. In finite dimensions such a conclusion would follow automatically from the minimax argument, but in the infin dimensional case one has to be careful.

It would suffice to show that these minimax critical points are all non-degenerate in the sense of [7], but the whole point is that we are trying to avoid a detailed analysis of the apparently complicated Hessian for non-planar elastic curves. Fortunately, the conclusion of Proposition 2.2 is precisely the hypothesis of the splitting theorem of Gromoll and Meyer for degenerate critical points [3] (we are indebted to Tony Tromba for calling our attention to their theorem).

As preparation for quoting the splitting theorem we use a local chart ψ about our minimax critical point z to pull back f to an open set Q in a Hilbert space: $h=f\circ\psi^{-1}:Q\subset H\to R$. We might as well assume that ψ takes z to the origin in H and that $h(0)=0$. By the chain rule it follows that the Hessian of h at 0 has the same kind of representation in terms of a compact operator (which we will still call K).

Now set $N=\ker(Id+K)$, a finite dimensional subspace since Id+K is Fredholm, and let E be the orthogonal complement of N. Then the splitting theorem asserts the existence of an origin preserving diffeomorphism Φ of some neighborhood of the origin of H, an orthogonal projection $P:E\to E$, and an origin preserving smooth map $j:N\to E$ such that, for small $(v,w)\in E\times N=H$,

$$h\circ\Phi(v,w)= \|Pv\|^2 - \|(Id-P)v\|^2 + h(j(w),w).$$

We now assume z is a local minimum for f and seek a contradiction. Thus, we might as well assume that $h\circ\Phi$ is non-negative. Since we also have $h(j(0),0)=0$ it follows that P is in our case just the identity, so $h\circ\Phi= \|v\|^2 +h(j(w),w)$. From the classification theorem we know that z is an isolated critical point of f, so 0 is an isolated critical point of $h\circ\Phi$.

Therefore, there exists an $\varepsilon>0$ such that $h\circ\Phi(v,w)>0$ on the set $O_\varepsilon-\{(0,0)\}$, where $O_\varepsilon=\{(v,w)\in H: \|v\|\le\varepsilon, \|w\|\le\varepsilon\}$. In particular, $h(j(w),w)>0$ for $\|w\|=\varepsilon$. But N is finite dimensional so compactness implies the existence of δ such that $h(j(w),w)\ge\delta>0$. This of course is precisely the condition (transferred back to the original functional f) which would allow us to obtain a still higher critical value for f corresponding to a fourth critical point in M -- a contradiction to the classification theorem. Thus we have established

<u>Theorem 3.2</u> The only relative minimum for $J_B:\Omega_B\to R$ is a prime geodesic. Hence, the only stable closed elastica in R^3 is a circle, once covered.

<u>Corollary</u> Let M be R^2 or R^3. Then there exists (up to similarity) a unique stable closed elastica in each regular homotopy class of immersed circles in M.

<u>Proof</u>: For $M=R^3$ this is just the above theorem. For $M=R^2$ one combines Theorem 0.1 with Example 2.1 and the fact that multiply covered figure eight elasticae are unstable even with respect to planar variations (Theorem 3.2 and Example 2.3 give instability for figure eights in R^3 only, but a special argument can be made for the planar case). /

Finally, we remark that although the curve straightening flow itsel[illegible] does not realistically describe any physical process (as far as we know), [illegible] instability theorem does have physical consequences. It implies, in particu[illegible] that a knotted springy wire cannot rest in stable equilibrium without poin[illegible] of self-contact — an experimentally observable fact. This fact leads to a rather curious "topologically constrained" variational problem; what actual[illegible] happens if one forms a knot in a piece of springy wire? Experiments yield some beautiful curves with impressive symmetry (e.g., for the figure eight knot or the Chinese button knot).

REFERENCES

[1] R. BRYANT and P. GRIFFITHS, Reduction of order for constrained variational problems and $\int \frac{k^2}{2} ds$, Amer. Journ. of Math. (to appear).

[2] M. GAGE, Curve shortening makes convex curves circular, (1983) Invent. Math. (to appear).

[3] D. GROMOLL and W. MEYER, On differentiable functions with isolated critical points, Topology 8 (1969),361-369.

[4] W. KLINGENBERG, Lectures on Closed Geodesics, Springer-Verlag, Heidelberg , (1978).

[5] J. LANGER and D. SINGER, On the total squared curvature of closed curves in space forms,(1983) to appear.

[6] ________________________, Knotted elastic curves in R^3, (1983) to appear.

[7] R. PALAIS, Morse theory on Hilbert manifolds, Topology 2 (1963),299-340.

[8] _________, Critical point theory and the minimax principle, Global Analysis, Proc. Symp. Pure Math. XV (1970), 185-212.

[9] _________, The principle of symmetric criticality, Commun. Math. Phys. 69 (1979), 19-30.

[10] S. SMALE, Regular curves on Riemannian manifolds, Trans. Amer. Math. Soc. 87 (1958), 492-512.

[11] A. TROMBA, A general approach to Morse theory, J. Diff. Geo. 12 (1977), no. 1, 47-85.