JOURNAL OF APPROXIMATION THEORY **84**, 315–350 (1996)
ARTICLE NO. 0022

Unified Representations of Nonlinear Splines

ANDERS LINNÉR

Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115

Communicated by Carl de Boor

Received December 14, 1992; accepted in revised form February 5, 1995

Golomb and Jerome's framework is modified and extended. The new framework is more general since it also handles interpolants which are not allowed to "slide" at the nodes. The space of interpolants of variable length is shown to be a smooth manifold. If the length is fixed, and there are no nodes, then the space of interpolants is a manifold. When there is at least one node, and at least one node is not on the line segment between the endpoints, then the space of interpolants of fixed length is a smooth manifold. Sufficient conditions are given which ensure the space of interpolants continues to be a smooth manifold in the presence of additional constraints such as clamping and pinning. A new fundamental finite-dimensional equation is derived. When it is solved it yields all nonlinear splines, and every nonlinear spline appears in this way. An important feature is that the same symbolic equation is used for all possible combinations of the constraints considered. It is shown how to take the solutions of the fundamental equation and use them to express the corresponding nonlinear splines in terms of a pair of elliptic functions. An inequality is derived that specifies which elliptic function appears along each section of the spline. The nonlinear splines are in a unified way shown to be C^2 for all possible combinations of the constraints considered.

0. INTRODUCTION

0.1. *Motivation and Overview*

0.1.1. *Elastic Energy*

When a thin beam or wire is bent it stores elastic energy. According to the Euler–Bernoulli model, this elastic energy is proportional to the total squared curvature. It is a classical problem to find the "equilibria" of the elastic energy. A "natural" constraint is to assume the length fixed. A different constraint is to fix the locations of the endpoints. This is called pinning. A third constraint is to fix the tangent directions at the endpoints. This is known as clamping at the endpoints.

0021-9045/96 $18.00

0.1.2. *Interpolation and Sliding at the Nodes*

In the context of splines one also considers the additional constraint of requiring the wire to pass through a finite number of given points in some specific order. The points are called nodes when considered as points along the wire. Arbitrary curves satisfying this kind of constraint are called interpolants. There are two ways the constraint can be imposed. The first possibility is that a node is required to appear at a spot already marked on the wire before bending. The other possibility is that the node can be anywhere along the wire, as long as the nodes appear in the prescribed order along the wire. The second possibility corresponds ("physically") to the ability of the wire to slide at the node.

0.1.3. *Discontinuous Second Derivative when Clamping at the Nodes*

Given any node it is possible to fix the tangent direction at the node. The expressions given in Section 3.5 reveal a discontinuity which disappears only if the clamping is at the endpoints. A consequence of this is the failure of an equilibrium to have continuous second derivatives (unless the equilibrium is still an equilibrium when the clamping at the nodes are removed). Since we like to prove that the equilibria have continuous second derivatives, clamping is only assumed at one or both of the endpoints, or not at all.

0.1.4. *Critical Points*

In our framework the concept of equilibrium is replaced by the notion of critical point. A critical point is a point where a gradient vector field vanishes. It follows that a critical point need not correspond to an extremum. This is significant for us since there need not exist an extremum. When the length is not fixed, and the straight line segment is not admissible, there is no minimum of the total squared curvature; see Section 2.6.

0.1.5. *Equilibria in the Calculus of Variations*

It is common to use collections of necessary conditions to define the equilibria; see, for instance, [5]. In variational problems the Euler–Lagrange equation is a necessary condition, however, depending on the constraints there may also be "other" necessary conditions. In fact, it may not be obvious what the "other" necessary conditions are, nor that all necessary conditions have been considered. (A "classic" example which illustrates these issues is the "brachistochrone" problem with variable endpoints. The problem is to find the path of minimal time to be followed by an object which due to gravity is sliding without friction along the path. Assume the path must start somewhere along some given curve and end somewhere along a different given curve.)

0.1.6. *Avoiding the Issue of Necessary Conditions*

As the last paragraph suggests it is legitimate to ask (when attempting to solve a variational problem) if all necessary conditions have been considered. In our framework this question never arises. From our point of view there is a functional with a domain which is an infinite dimensional open manifold. Associated with the functional is a gradient vector field defined on the tangent spaces of the infinite dimensional manifold. The constraints single out subsets of the domain which for now we assume are smooth manifolds. The gradient vector field is projected onto the tangent spaces. The problem is to find all points where the projection vanishes.

0.1.7. *The Variational Problem Turned into a Computation*

It follows that if the subsets corresponding to the constraints are shown to be smooth manifolds, then the variational problem is completely computational. The gradient must be computed and its projection determined. Of course, one likes to give the points where the projected vector field vanishes as explicitly as possible. It is important to employ a process which does not introduce nor hide solutions. The computational process in the proof of Theorem 4.1 exhibits this feature. We are very careful to include every single "candidate," so none is missed. Once the "survivors" have been found, each is tested explicitly to show that the projected gradient vector field indeed vanishes.

0.2. *New Contributions*

We now begin to describe in nontechnical terms what is new in this paper. Many of the details are considered in Section 1, where a comparison is made with [5].

0.2.1. *The Issue of Variable Length*

It is perfectly reasonable to ask why this "classical" Euler–Bernoulli problem warrants yet another investigation. One of the questions, which seemed beyond the earlier methods employed, concerns the existence of critical points when the length is not fixed but clamping is imposed at both endpoints. Even when there are no nodes the answer turns out to be quite complicated. In [16] it is shown that there are either no critical points, a finite number of critical points, or a countably infinite number of critical points.

0.2.2. *A New Example that Illustrates the Possible Complexity of the Collection of Critical Points*

In Section 1.1.5 our main theorem 4.1 is specialized to the case with clamping at only one endpoint, and this offers another example of this

```
cn[x_] := JacobiCN[x,1/2]
sn[x_] := JacobiSN[x,1/2]
am[x_] := JacobiAmplitude[x,1/2]
kc := EllipticK[1/2]
ec := EllipticE[1/2]
e[x_] := EllipticE[am[x],1/2]
m[x_] := kc (2 e[x] - x)
n[x_] := 2 ArcSin[sn[x]/Sqrt[2]]
f[x_] := ArcTan[Sqrt[2] kc cn[x]/(m[x] + Pi/2)] + n[x]
Plot[{f[x],Pi/2},{x,0,12 kc},
        PlotStyle->Thickness[0.00001]]
```

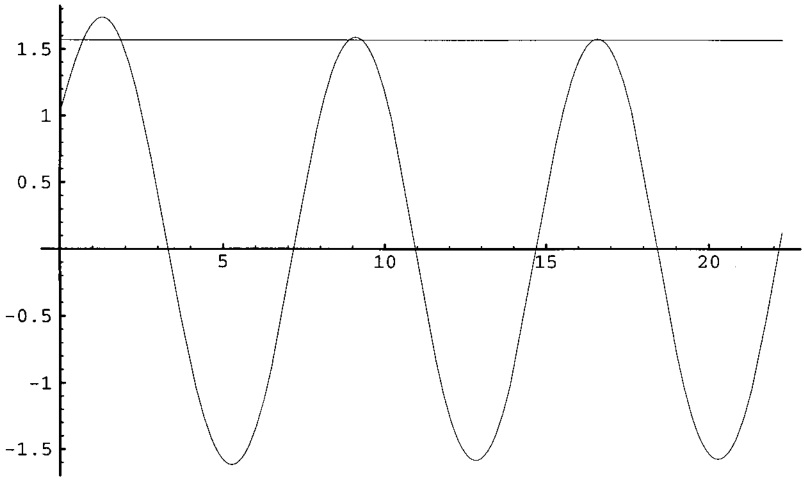

FIGURE 1

kind. As these examples illustrate, one cannot hope to find the critical points explicitly except in very special cases. The best one can do is to express the critical points in terms of known functions and a finite number of unknown parameters. The parameters are required to satisfy a system of equations and if this system is solved then the critical points are completely determined. In Fig. 1. each intersection between the line $\pi/2$ and the graph gives a different value of one particular parameter, and thus a different critical point. Figure 1 is discussed in detail in Section 1.1.5.

0.2.3. *A New Ambient Space Solves the Problem: Pinning at the Nodes*

For the time being think of nonlinear splines as critical points of the total squared curvature in the space of interpolants (the precise definition is given in Section 2.5). We introduce a new ambient space containing all sufficiently smooth curves. This space is a Cartesian product of three factors. The first two factors have been used before, by Langer and Singer in [10] and by the author in [14]. The third factor is new. Its purpose is to "track" the nodes and facilitate the solution of the case where no sliding is allowed at the nodes. Note that this case is avoided in the comprehensive paper [5].

0.2.4. *Properties of Smooth Infinite Dimensional Manifolds*

The ambient space is not quite a linear space but an open subset of a linear space. Curves which satisfy given constraints appear as subsets of the ambient space. The constraints, fixing length, clamping and pinning, "generically" produce subsets that are smooth submanifolds of the ambient space. Sufficient conditions are given in Proposition 3.2. A smooth submanifold means that there is a well-defined tangent space at each point of the subset. For us the tangent spaces are infinite dimensional but the corresponding normal spaces are finite dimensional. All normal spaces have the same dimension. There are no singularities such as cusps or self-intersections. When a smooth vector field on the ambient space is projected onto the tangent spaces of the submanifold, the resulting vector field is still smooth.

0.2.5. *No Critical Point is Overlooked*

The total square curvature is a functional defined on the ambient space. A crucial step is to compute the projection of its gradient vector field onto the tangent spaces of the constrained set. The ultimate goal is of course to determine all the points where the resulting tangent vector field vanishes. Since we insist on finding all critical points, the analysis of the well-known Euler–Lagrange equation is here extended to include the proof that all initial value problems can be solved using one or the other of the Jacobi elliptic functions *cn* and *dn*; see Proposition 3.3. The same issue is resolved in the special case $c=0$ in [16]. There it is shown that *cn* alone solve all initial value problems when $c=0$. Note that Proposition 3.3 also contains an inequality, which in terms of the data of the problem distinguishes between the cases solved by *cn* from the ones solved by *dn*. A related, but different, criterion is given by Brunnett in [2, Theorem 2, p. 47]. Analogous results are given in [3, Theorem 4, p. 6]. In the case $c=0$ see [4, Theorem 1.3.2, p. 9].

0.2.6. *The Fundamental Finite Dimensional Equation*

When all is put together, the main result, Theorem 4.1, presents a system of equations with the following properties. All constraints considered can be resolved using one and the same system of symbolic equations. Note that there are three kinds of symbols in this system. First there are the parameters associated with the solutions of the Euler–Lagrange equation: f_j as function names; p_j, q_j as moduli; and α_j, β_j as parameters in the arguments of the functions. Next there are the "multipliers" belonging to the constraints: the complex numbers λ_j for the nodes as points in the plane; the real numbers σ_j for the nodes as points on the nonlinear spline; and the real numbers μ for the length; μ_0, μ_1 for clamping. Finally there are the symbols associated with the data of the problem: the complex numbers $g_j = P_j - P_{j-1}$ for the location of the nodes in the plane; the real numbers $\tilde{s}_j$ for the location of the nodes along the wire; $\tilde{L}$ for the length; and θ_0, θ_1 for the clamping angles.

0.2.7. *How the Fundamental Equation is used*

If a constraint is imposed, then its corresponding multiplier is unknown and the value of the constrained quantity is known. Conversely, if a constraint is missing then the multiplier is zero but the constrained quantity is now unknown. It is shown in Section 5 how to use the same system of equations in order to handle all combinations of constraints. The symbolic relationships are fixed for all combinations of the constraints. When the constraints are changed there is a transition between symbols that have a known value and the symbols that are unknown. A key to the analysis of nonlinear splines is to understand this fundamental system of equations.

0.2.8. *Unification Made Possible by the Use of the New Ambient Space*

The unification of the treatment of nonlinear splines subject to different constraints is successful primarily because we use a single ambient manifold. Note that every nonlinear spline must correspond to a solution of the fundamental system of equations. Conversely, if there is a solution of the fundamental system of equations and this solution is compatible with the restrictions imposed by the inequality of Theorem 4.1, then the corresponding critical point must be a nonlinear spline satisfying all the constraints imposed.

0.2.9. *Previous Cases as Special Cases of Theorem* 4.1

Theorem 4.1 shows how the infinite dimensional problem of finding all the nonlinear splines, in the space of all interpolants subject to some specific set of constraints, is reduced to a finite dimensional problem. One cannot in general expect further reductions. There is, however, one case which is completely explicit. Assume that there are no nodes (so $n = 0$), no length constraint and no clamping. The fundamental system of equations of Theorem 4.1 in this case simplifies and can be completely solved. If clamping is imposed at one endpoint then the system simplifies somewhat, but not to the point where it can be completely solved. Recall Fig. 1 and the intersections of the two graphs. It is too much to ask to have all points of intersection explicitly. (An analysis of the asymptotic behavior shows there is a countably infinite number of solutions.) The first of these cases is considered in [16]. Both cases are now very special subcases of Theorem 4.1.

0.3. *Organization*

The paper is organized as follows:

0. Introduction.
1. Background.
2. The setup.
3. Derivatives, gradients, manifolds and the differential equation.
4. The fundamental equation and all constraints imposed.
5. Not all constraints imposed.

Some of this work has been presented at the NIU classical analysis seminar and the author would like to take this opportunity to thank all the participants. The author sincerely appreciates all the input given to him by Carl de Boor. A special acknowledgement also goes to David A. Singer for a valuable suggestion and, as always, inspiring the author in a time of despair. Last but not least, I hope I can be as strong and courageous as Dorothy Leah Johnson, my *favorite* kidney donor.

1. Background

1.1. *Relationship to the Work by Golomb and Jerome*

1.1.1. *Definition and Admissible Variations*

We now sketch in more detail how this paper relates to the comprehensive paper [5] by Golomb and Jerome. The first difference appears in the definition of a nonlinear spline. Golomb and Jerome use a variational definition, which in their abstract is given by $\delta \int_0^s \kappa^2(s)\, ds = 0$, in terms of the total squared curvature. To be precise one should also specify the domain over which the variation takes place. This is particularly important when the imposition of constraints makes the domain nonaffine in which case one can proceed easily only if a tangent space can be shown to exist at all points of interest in the domain. We do give conditions which guarantee such existence.

1.1.2. *Difficulties with Variable Length*

If the length of the curve is variable, a certain technical difficulty arises when curves are parametrized according to arclength (note the variable upper limit in $\delta \int_0^s \kappa^2(s)\, ds = 0$.) There is no single fixed function space on which to carry out the analysis. To circumvent this difficulty we write the domain as a product of a function space and a factor representing the length. As mentioned in Section 0.2.3, this technique has been used before.

The use of variational methods in search of critical points in a space of curves of arbitrary length is extensive. The total squared curvature functional has been used to find geodesics. See [12–14] for the geodesic problem and [15] for a survey.

1.1.3. *Allowing Pinning*

If the total length is fixed, pinning corresponds to fixing the length between consecutive nodes. We introduce a new factor representing the parameters associated with the nodes. For us pinning is allowed, but not required. If the total length is variable, then pinning at a node fixes the ratio between the length "before" the node and the length "after" the node. Neither kind of pinning is considered in [5].

1.1.4. *Existence of Nonlinear Splines with No Length Restrictions*

We quote from [5, p. 422], "the existence of such extremals interpolating n points in general position, and whether they are local minima or not, remains an open question ... ." Our main result shows that the answer to this question is determined by whether the fundamental finite-dimensional equation has solutions or not. Every solution corresponds to a nonlinear spline and conversely. Examples show that anything can happen. There may be no solutions, any finite number of solutions, or an infinite number of solutions. These possibilities occur already for curves with no nodes; see [16]. To answer the open question one must find all solutions of the fundamental equation. Only in very special cases can such a solution be given explicitly. As the next example illustrates, explicit solutions are not expected in general. In Section 1.1.6 we discuss local minima.

1.1.5. *A Simple Example Illustrating Existence*

Consider all planar curves of any length starting at $(0, 0)$ and ending at $(1, 0)$. Assume the tangent direction at $(1, 0)$ is given as an angle with respect to the x-axis. What is the range of angles so that there are corresponding nonlinear splines, starting at $(0, 0)$ ending at $(1, 0)$, such that their tangent direction at $(1, 0)$ agrees with the given angle? Note that the tangent direction at the origin is not fixed, so this problem is different than the ones considered in [16, 2]. Also, recall Section 0.1.5 and the example of the difficulties associated with variable endpoints.

An application of Theorem 4.1 yields the following; see Section 5.4. The range is approximately $[-1.7378, 1.7378]$ in radians. Note that 1.7 is bigger than $\pi/2$. Moreover, it follows that if the prescribed angle is 1.8 radians (say), then there are no nonlinear splines satisfying the boundary conditions.

If the angle is in $[-1.7378, 1.7378]$, but not in $[-\pi/2, \pi/2]$, then there is only a finite positive number of nonlinear splines satisfying the boundary conditions. If the prescribed angle is in $[-\pi/2, \pi/2]$, then there is a countably infinite number of nonlinear splines. Figure 1 illustrates the case $\pi/2$. An analysis of the asymptotic behavior shows that the local maxima in Fig. 1 reach above $\pi/2$ a countably infinite number of times. The limiting value of the local maxima is $\pi/2$.

1.1.6. *Four of the Corresponding Critical Curves*

Figure 2 illustrates the "first" four critical curves. Two of the curves have vertical tangent direction at the origin. Both are critical without the tangential constraint at $(1, 0)$, and both are saddle points of the total squared curvature for the following reason. Add two straight line segments at both endpoints and make sure they are tangent to the curve (the result is a sufficiently smooth curve). The constraints are still satisfied and the total squared curvature is unchanged. The new curve is not a member of the countable list of critical points given in [16]. It follows that the gradient is not zero in the space without tangential constraints. The new curve is also in the space of curves subject to the tangential constraint. It is again not a member of the "new" countable list of critical points. By shortening the two line segments, one gets a sequence of curves arbitrarily close to the critical curve such that at each curve the gradient is not zero. If one follows the negative gradient trajectory through any one of these curves, new curves with lower total squared curvature are found instantly. Conversely, if the straight line segment added at the origin is replaced by a piece shaped like a question mark, one gets a curve with larger total squared curvature. Again it is possible to adjust the shape of the question mark to get a sequence approaching the longer spline with strictly larger total squared curvature.

The first part of the reasoning above is not applicable to the other pair of critical curves. Note that the straight line segment between $(0, 0)$ and $(1, 0)$ is not admissible when the angular constraint is imposed. Without the constraint the line segment is a global minimum. If the angular constraint is zero rather than $\pi/2$, then the line segment is admissible. As the constrained angle is increased above zero, the critical point initially corresponding to the line segment must "move" somewhere. We suspect the

```
tc = Pi/2
lc := ac Sin[tc - n[ac - kc]]/(Sqrt[2] cn[ac - kc])
lc := ac Cos[tc - n[ac - kc]]/(2 (e[ac - kc] + ec) - ac)
enc := ac/lc (2 (e[ac - kc] + ec) - ac)
g[s_] := lc/ac (2 (e[ac s - kc] + ec) - ac s - I Sqrt[2] *
                cn[ac s - kc]) Exp[ I (tc - n[ac - kc])]
ac = FindRoot[f[x] == tc, {x,0,1}][[1,2]] + kc
ac = FindRoot[f[x] == tc, {x,2,3}][[1,2]] + kc
ac = FindRoot[f[x] == tc, {x,8,9}][[1,2]] + kc
ac = FindRoot[f[x] == tc, {x,9,10}][[1,2]] + kc
ParametricPlot[{Re[g[s]],Im[g[s]]},{s,0,1},
                PlotStyle->Thickness[0.00001],
                AspectRatio->Automatic]
```

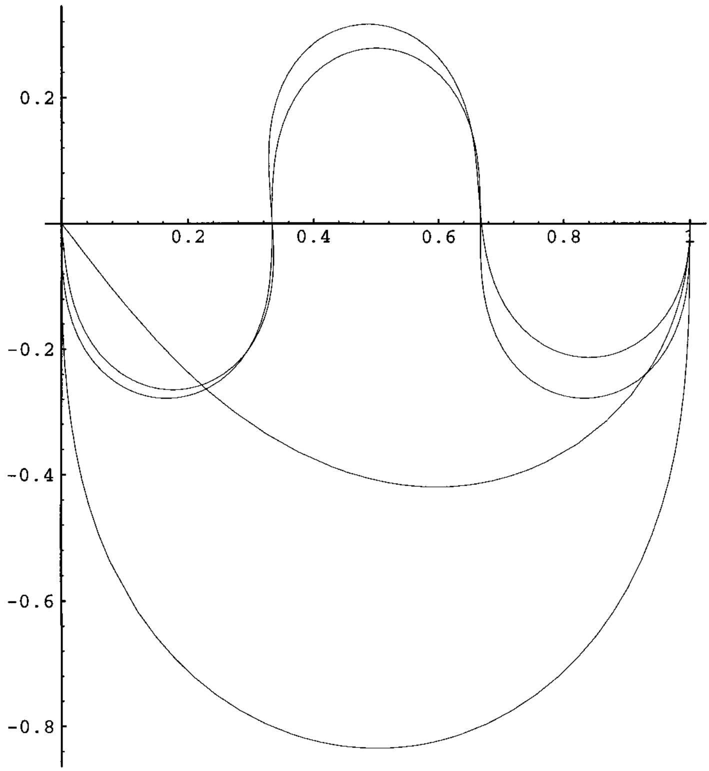

FIGURE 2

shortest curve in Fig. 2. is the "new" critical point when the angular constraint is increased all the way to $\pi/2$. Based on this we conjecture that the shortest curve is a local minimum. We show in Section 2.6 that the infimum of the total squared curvature is zero, so it follows that the shortest curve cannot be a global minimum.

The figures were produced with the help of Mathematica. The code is given in Fig. 1 and Fig. 2. The code is completely symbolic except for the numerical "FindRoot" function. The same formulas work for different constrained angles by changing "tc = Pi/2." Note how the curve is expressed as a complex function. The length can be computed using the formula for "lc." For the energy use the formula for "enc." The lengths are 1.406996, 2.18844, 2.10987, and 2.18844, where the second and the fourth have a vertical tangent direction at both endpoints. The corresponding energies are 2.64508, 2.87108, 25.8139, and 25.8397.

1.2. *Constraints and Manifolds of Interpolants*

1.2.1. *Sufficient Conditions That Guarantee Subsets Are Manifolds*

Let n be a nonnegative integer. Suppose $P_0, ..., P_{n+1}$ are $n+2$, not necessarily distinct points in the plane. It is shown how to represent interpolators in a way that facilitates the analysis associated with additional constraints. Examples of such constraints include: fixing the total length of the curve, pinning, clamping. If a curve is pinned at the nodes P_i and P_{i+1}, then the length of the curve between the two nodes is in a fixed proportion of the total length. If a curve is clamped at P_i, then the tangent direction of the curve is fixed at P_i.

The space of interpolators is shown to be a smooth manifold. The subset of interpolators of fixed length is also a smooth manifold, provided the P_i's are not all on the same line. If the length is fixed and the interpolator is pinned at all nodes P_i, then the corresponding subset of interpolators is a smooth manifold, provided the pinning is such that the length of the interpolator between any P_i, P_{i+1} exceeds the distance $|P_{i+1} - P_i|$. If the length is not fixed and the interpolator is pinned at all nodes, then the associated subset is a smooth manifold, provided the ratio between the length of the interpolator from P_i to P_{i+1} and the distance $|P_{i+1} - P_i|$ is different for each i. It is also shown that if the interpolator is clamped at P_0 and/or P_{n+1}, then in each of the above cases the resulting subset is still a smooth manifold.

1.2.2. *Necessary Conditions Are Not Known*

When none of the sufficient conditions are satisfied, it is an open question whether the subset is a manifold or not. The proof of Proposition 3.2

shows that the subset of curves satisfying the constraints is an intersection of manifolds, where the intersection is not transversal. Such an intersection can yield a manifold of the "right" dimension. A finite dimensional example is given by the intersection of the "sheets" $y=x^2$ and $y=-x^2$ in three-dimensional space. In general the subset need not be a manifold. For instance, cusps can be created, $z=x^2-y^3$ and $z=0$, as well as "self-intersections," $z=x^2-y^2$ and $z=0$.

1.3. *Existence of Minimizers*

If the length is fixed, there always exists a nonlinear spline satisfying the constraints; see [7, 8]. When the length is variable, we show in Section 2.6 why the total squared curvature is not globally minimized whenever the straight line segment does not satisfy the constraints; see also [1].

If a positive constant times the length is added to the functional, then there is a global minimizer. This constant is the tension parameter used by Brunnett in [2]. As the examples in Section 2.6 show, the functional considered in [2] has no global minimum when the tension parameter is not positive. There may be no solutions to the variational problem; see Theorem 4.1 in [16] for precise conditions in the case of no tension. It is not difficult to modify Theorem 4.1 so that it includes the tension parameter.

1.4. *Existence of Multipliers*

A different issue concerns the application of Lagrange's multiplier rule. The existence of multipliers is by no means guaranteed. The Lagrange multiplier theorem requires a complete function space. In this paper we modify and expand the approach used by Golomb and Jerome in [5]. The domain consists of a product of factors where the infinite dimensional factor is the largest possible Sobolev space on which the total squared curvature is still defined. A Hilbert space structure is used to project gradients onto the tangent spaces and the normal spaces. At the nonlinear splines the projection of the gradient of the total squared curvature vanishes. This definition includes local minimizers among others. The normal space is spanned by gradients associated with the constraints. The normal component of the gradient of the total squared curvature is expressed as a linear combination of the spanning vectors. The associated scalars are the Lagrange multipliers.

1.5. *A Unified Reduction to a Finite Dimensional Problem*

If a nonlinear spline is sought from scratch, an infinite dimensional problem must be solved. The main contribution of this paper is to reduce the infinite dimensional problem to a finite dimensional problem. This reduction is done in a unified way for all constraints. As discussed in

Sections 0.2.6–0.2.8, the fundamental finite dimensional equation is symbolically the same for any combination of the constraints considered. All nonlinear splines must correspond to solutions of the fundamental equation. Conversely, every solution corresponds to a nonlinear spline. Note that the variational problem is given a firm foundation. Nonlinear splines are defined geometrically as points in an infinite dimensional space where a certain vector field vanishes. The question regarding the existence of nonlinear splines is reduced to a question about existence of solutions to a finite dimensional equation. The geometric approach also reveals why the nonlinear splines are C^2 or better, no matter which constraints are imposed. This unified treatment incorporates all cases considered in [2, 5, 6, 8, 9, 11].

1.6. *Real Sobolev Spaces, Complex Numbers*

The desire to deal with all combinations of constraints causes serious technical and notational difficulties. The use of complex numbers to represent the curves, and also some of the multipliers, alleviates some of the difficulties. Note that the Sobolev space we use is a real vector space, so this use of the complex numbers is not standard; see also [16].

1.7. *No Solution Is Missed*

To reduce the problems from infinite to finite dimensions a well-known ordinary differential equation must be solved. It is easy to check that the Jacobi elliptic functions cn and dn satisfy the differential equation. It is considerably more difficult to prove that there are no other functions that satisfy the differential equation. Proposition 3.3 contains this result and gives an inequality that distinguishes between the two possibilities. Without this result it is not obvious how to prove that *every* nonlinear spline is given by a solution of the finite dimensional fundamental equation. As mentioned in 0.2.5, Brunnett has resolved the same issue in a slightly different context.

2. The Setup

2.1. *Interpolants*

Consider the standard Euclidean plane and suppose $P_0, ..., P_{n+1}$ are $n+2$ given points. The integer n is nonnegative and the points P_i are not necessarily distinct. A curve passing through the points P_i, in the given order, is called an interpolant. The initial point of the curve is thus P_0, and the curve ends at P_{n+1}. The n points P_1 through P_n are referred to as nodes. Note that curves are not assumed to be representable as graphs of real-valued functions on some interval, so in particular, an interpolant may intersect itself. We assume the curves are rectifiable so the length is defined. The length is not constrained and can be any positive real number.

2.2. *Functionals and Domains*

For sufficiently smooth curves γ, let κ be the curvature. The Euler–Bernoulli elastic energy is given by $F(\gamma) := \int_\gamma \kappa^2$. Note that F requires the existence of second derivatives (a.e.), and the second derivatives must be square integrable. The set of all such curves can be given a Hilbert space structure. The result is the Sobolev space $W_2^2 = H^2$. A given curve may be parametrized in many ways. In our context each curve is parametrized so the speed is equal to the length of the curve, the domain is therefore the interval $[0, 1] =: I$. This permits us to introduce an auxiliary function from which the curve can be recovered by integration. The (a.e.) derivative of the auxiliary function is required to be square integrable in the sense of Lebesgue. The corresponding Sobolev space is $W_2^1 = H^1 =: H$. There are several inner products available. We let $\dot{v}$ denote the (a.e.) derivative of v, and choose

$$\langle v, w \rangle_H = v(0)\, w(0) + \int_I \dot{v}\dot{w},$$

because it leads to the least complicated formulas.

2.3. *Parametrizations*

Note that F is invariant under changes of the parametrization of γ. We like the domain representing the interpolants to have the property that the nonlinear splines are isolated critical points of F. It is therefore necessary to further restrict to curves parametrized in a unique way. A common choice is to parametrize by arclength (see [5]). When curves of variable length are considered this choice causes the parameter interval to be variable. The difference between the final parameter and the initial parameter is equal to the length, so there is no single function space representing all variable length interpolants.

A related issue concerns the possibility of allowing the length of the interpolant to vary between consecutive P_i's. If the total length is fixed, this corresponds to letting the curve slide at the P_i's. Note that the parameters corresponding to the nodes are bounded above by the total length. When the curve is parametrized by arclength, and the total length is variable, such a bound causes difficulties. The constraining relationship between the total length and the parameters associated with the nodes, is not readily expressed in a form suitable for analysis.

Suppose that $\gamma: I \to R^2$ is a continuously differentiable curve of length L. There is a continuous function $\theta: I \to R$ such that $\dot{\gamma}(s) = L(\cos\theta(s), \sin\theta(s))$. Note that $\theta(s)$ is the angle between the x-axis and the tangent of γ at the point $\gamma(s)$. If θ is differentiable then the curvature of γ is given by $\kappa = \dot{\theta}/L$. In terms of θ and L the Euler–Bernoulli energy F is proportional to $J(\theta, L) = (1/2L)\int_I \dot{\theta}^2$ (one L disappears due to the change of parameter, and the $\frac{1}{2}$ simplifies formulas).

2.4. *The Space of Interpolants*

In order to deal with the values of the parameters associated with the $(n>0)$ nodes, we introduce the space $\Delta^n = \{\bar{s} \in R^n \mid \bar{s} = (s_1, ..., s_n), 0 < s_1 < \cdots < s_n < 1\}$. The length is an element of the positive reals R^+. The space of all sufficiently smooth curves is represented by $H \times R^+ \times \Delta^n$. If there are no nodes, so $n = 0$, replace $H \times R^+ \times \Delta^n$ by $H \times R^+$. The space $H \times R^+ \times \Delta^n$ is an open manifold which is almost a vector space. The space of sufficiently smooth interpolants is a subset of $H \times R^+ \times \Delta^n$. If more constraints are imposed the corresponding subset is smaller. We consider the following additional constraints: fixing length, pinning, and clamping. To illustrate we let $(\theta, L, \bar{s})$ be some element of $H \times R^+ \times \Delta^n$. To fix the length means to require $L = \tilde{L}$ for a given $\tilde{L} \in R^+$. Similarly, pinning at each node corresponds to fixing the value of $\bar{s}$. If an interpolant is pinned at each node, then the proportion of lengths between consecutive P_i's is fixed. If, in addition, the total length is fixed, then these lengths between consecutive P_i's are also fixed. If an interpolant is not pinned at a node, then the curve can "slide" at the node. Finally, clamping corresponds to fixing $\theta(0)$ and $\theta(1)$ (or only one of the two), so the tangent directions are given at the endpoints. Recall from Section 0.1.3 that clamping at the nodes is not considered.

At each point, the manifold $H \times R^+ \times \Delta^n$ has a tangent space which is identified with $H \times R \times R^n$. Let $v = (v_\theta, v_L, v_{\bar{s}}) \in H \times R \times R^n$, and similarly w, be tangent vectors. With the use of the standard Euclidean dot product, the manifold $H \times R^+ \times \Delta^n$ becomes a Riemannian manifold with inner products

$$\langle v, w \rangle = \langle v_\theta, w_\theta \rangle_H + v_L w_L + v_{\bar{s}} \cdot w_{\bar{s}}$$

on the tangent spaces. To see that it is necessary to distinguish between $H \times R^+ \times \Delta^n$ and its tangent spaces, consider a variation of curves of decreasing length. The tangent vector associated with the variation must have a negative length component, and there is no such element in $H \times R^{+} \times \Delta^{n}$. A similar remark can be made regarding Δ^{n} and R^{n}.

2.5. *Definition of Nonlinear Splines*

Consider a constraint represented by a submanifold. At each point the tangent space of $H \times R^{+} \times \Delta^{n}$ splits into a subspace tangent to the submanifold, and its orthogonal complement. A real-valued smooth differentiable functional defined on $H \times R^{+} \times \Delta^{n}$ determines a continuous linear map defined on the tangent space at each point. This linear map is the derivative. The Riesz representation of the derivative is the gradient. At each point the gradient is an element of the tangent space $H \times R \times R^{n}$. The derivative meanwhile is an element of the dual. The gradient splits into a tangential and a normal part. If at some point the tangential part vanishes, the point is called a critical point. An element in a submanifold, corresponding to interpolants with possibly additional constraints, is called a nonlinear spline if the tangential component of the gradient of J vanishes.

2.6. *Examples of Nonexistence*

If the length is not fixed there may not be a minimum. Suppose $n=0$ and $P_0 = P_1 = (0, 0)$. Consider circles touching the origin. As the radius increases the Euler–Bernoulli energy decreases to zero. Since only a straight line segment has zero energy, there is no minimum. By Theorem 3.1 in [16], there are no nonlinear splines in this case.

Now change P_1 to $(1, 0)$ and introduce clamping by requiring the tangent direction to be some given angle at P_1. Using the same radius, put two semicircles and two quarter circles together in two question mark mirror images. Connect the free ends of the semicircles using a straight line segment. Draw a straight line segment from P_1 with the given tangent direction. The segment is to end at a point, where the corresponding vector from the origin is perpendicular to the segment. Place one end of the previous figure at the origin and the other at the end of the last line segment. Make sure the tangent directions are the same. An arbitrarily large radius gives an energy arbitrarily close to zero. Again there is no minimum.

More generally, if the tangent directions are given at both P_0 and P_1, put a circular arc of radius r with the given tangent direction at P_0. Arrange so that the end is parallel to the x-axis and points in the negative direction. Do the same thing at P_1, except arrange so it points in the positive direction and make sure the radius is at least r. Connect as before

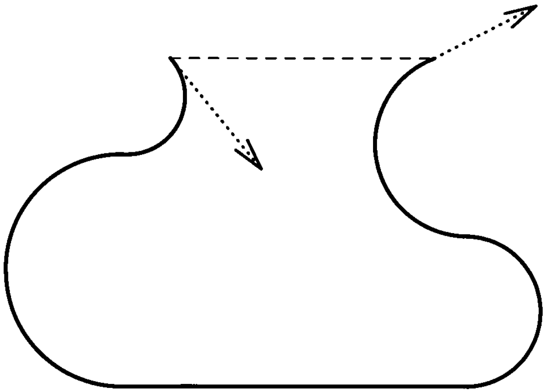

FIGURE 3

using two semi-circles of radius r or more, and a straight line segment; see Fig. 3. Increasing r decreases the energy to zero.

Note that all the piecewise defined examples are curves in W_2^2. The set of smooth C^∞ curves is dense in W_2^2. Since F is smooth it follows that there are smooth curves with arbitrary small energy.

2.7. *Ensuring Existence*

There are two ways to ensure existence of minimizers: (1) fix the length, (2) add a positive multiple of the length to F. The nonexistence of minimizers does not preclude the existence of critical points. In the case of a constrained tangent direction at P_1, there may be a countably infinite number of critical points, a finite number of critical points, or no critical points. By varying the tangent direction, examples of each kind can be exhibited using the functional J, see 1.1.5 in the introduction.

2.8. *Smoothness and Constraint Removal*

Recall that the interpolants are assumed to be W_2^2, so the second derivative is only assumed to be in L^2. In the proof of Theorem 4.1 it is shown that for any combination of the constraints considered, each nonlinear spline is in fact C^2. It is also worth pointing out that, by proving Theorem 4.1 with all constraints imposed, it is possible to take care of all cases with *fewer* constraints. The method can simplistically be summarized as: "when a constraint is removed, set the corresponding multiplier to zero and turn the value of the constraint into an unknown." Note that the number of equations stays the same and so does the number of unknowns. The details are given in Section 5.

3. Derivatives, Gradients, Manifolds and the Differential Equation

3.1. *Smoothness and Complex Numbers*

The Euler–Bernoulli functional is to be considered on the collection of all sufficiently smooth curves γ with given end points P_0, P_{n+1} which pass, in order, through given points $P_1, ..., P_n$. Since we are only dealing with planar curves, it is convenient to interpret the points in R^2 as complex numbers and, correspondingly, parametrize the curves in the form

$$\gamma\colon [0, 1] \to C\colon s \mapsto P_0 + L \int_0^s e^{i\theta(t)}\, dt.$$

In these terms, "sufficiently smooth" means that the tangent indicatrix, θ, of such a curve is in

$$H := W_2^1([0, 1]),$$

the space of all absolutely continuous real functions on $[0, 1]$ with first derivative in L_2. The intent is to restrict attention to those triples

$$(\theta, L, \bar{s}) \in H \times R^+ \times \Delta^n$$

for which

$$P_0 + L \int_0^{s_j} e^{i\theta} = P_j, \qquad j = 1, ..., n+1$$

(with $s_0 := 0$, $s_{n+1} := 1$). In this way, the domain of the functional J becomes the set

$$H \times R^+ \times \Delta^n$$

with tangent spaces

$$H \times R \times R^n.$$

To simplify the notation, we occasionally write $p = (\theta, L, \bar{s}) \in H \times R^+ \times \Delta^n$ and $v = (v_\theta, v_L, v_{\bar{s}}) \in H \times R \times R^n$.

Recall that a regular curve is a differentiable curve with nonzero speed everywhere. It is a standard fact that by changing the parametrization any regular curve can be represented by a curve $\gamma\colon I \to C$ such that $|\dot{\gamma}(s)| \equiv L$, where $L > 0$ is the length of the curve.

3.2. *The Directional Derivative*

We are interested in the following functional

$$J\colon H \times R^+ \times \Delta^n \to R$$

given by

$$J(\theta, L, \bar{s}) = \frac{1}{2L} \int_I \dot{\theta}^2.$$

It has the directional derivative

$$DJ(p)\, v = DJ(\theta, L, \bar{s})(v_\theta, v_L, v_{\bar{s}}) = \frac{1}{L} \int_I \dot{\theta} \dot{v}_\theta - \frac{v_L}{L} J(p)$$

in the direction $(v_\theta, v_L, v_{\bar{s}}) \in H \times R \times R^n$.

3.3. *The Gradient*

On any Hilbert space X with inner product $\langle\ ,\ \rangle_X$, we define the gradient of a functional $F\colon X \to R$ by

$$DF(x)\, v = \langle \nabla F(x), v \rangle_X$$

for all x and v in X. Since the tangent space to $H \times R^+ \times \Delta^n$ is a Hilbert space there is an element $\nabla J(p)$ such that

$$DJ(p)\, v = \langle \nabla J(p), v \rangle$$

for all p in $H \times R^+ \times \Delta^n$ and all v in $H \times R \times R^n$. The gradient is given by

$$\nabla J(p) = \nabla J(\theta, L, \bar{s}) = \left(\frac{1}{L} (\theta(s) - \theta(0)),\ -\frac{1}{L} J(p),\ 0 \right).$$

3.4. *Common Types of Interpolants*

The ambient space $H \times R^+ \times \Delta^n$ represents the space of all sufficiently smooth curves in our application. The interpolants we are interested in are subsets of $H \times R^+ \times \Delta^n$. As it turns out the common types of interpolants are represented by closed submanifolds, except for some exceptional cases. The various subsets are given as level sets of certain smooth maps. If the derivatives of the maps are surjective, the implicit function theorem is applicable, and the level sets are closed submanifolds.

The space of interpolants with n nodes corresponds to the intersection of the zero sets of $n+1$ complex-valued functionals. If the curves are pinned at the interior nodes, $\bar{s}$ must be prescribed. If the length is fixed, $L \in R^+$ is given. It is also possible to prescribe the angles at the endpoints (clamping). There are no doubt many other possibilities, but in this paper we focus on the cases mentioned and their combinations.

The given constraints are all of the form

$$\Lambda(p) = 0,$$

with each Λ a map into R or into $R^2 \sim C$. For example, in order to enforce the interpolation conditions, we introduce the $n+1$ complex-valued *gap functionals*

$$\Lambda_j \colon (\theta, L, \bar{s}) \mapsto L \int_{s_{j-1}}^{s_j} e^{i\theta} - (P_j - P_{j-1}), \qquad j = 1, ..., n+1.$$

Also, if the curve is to be clamped at t, i.e., if $\theta(t)$ is to take a prescribed value, θ_t, for some t, we use the functional

$$\Lambda_{\theta_t, t} \colon (\theta, L, \bar{s}) \mapsto \theta(t) - \theta_t.$$

The set over which J is to be restricted is then describable as the intersection of sets of the form

$$\Omega := \Lambda^{-1}\{0\}.$$

Correspondingly, at a critical point for J, the gradient of J must be normal to the tangent space of each constraint at that point.

Note that the intersection of two manifolds is not necessarily a manifold, unless the intersection is transversal. It is one purpose of this paper to show that, except for certain explicitly identified circumstances, all constraints have corresponding subsets which are manifolds that intersect transversally at each point. This requires the construction of the tangent space associated with Λ at a point in Ω. Note that this space is infinite dimensional. The constraints we are interested in are such that the orthogonal complement to the tangent space is finite dimensional. To represent elements of this space we need to compute derivatives and gradients. If the map is into C, we compute the transpose.

3.5. *Angle Functionals*

The directional derivative is

$$D\Lambda_{\theta_t, t}(p)(v_\theta, v_L, v_{\bar{s}}) = v_\theta(t)$$

The gradient is

$$\nabla \Lambda_{\theta_t, t}(p) = (f, 0, 0)$$

with $\dot{f} = \chi([0, t])$ and $f(0) = 1$. Here χ is the characteristic function, with values one at points in the set, and zero elsewhere. Note that the gradient is independent of p so, for fixed t, we get a constant vector field on $H \times R \times R^n$.

The two cases, $t = 0$ and $t = 1$, associated with clamping, give the linearly independent gradients $(1, 0, 0)$ and $(1 + s, 0, 0)$, respectively. Note that these are the only values of t where the gradient is smooth. If the nonlinear splines are to be at least C^2 smooth, the expressions used in the proof of Theorem 4.1 cannot contain gradients which are not themselves at least C^1. The only exception occurs when the associated Lagrange multiplier is zero. However, if the multiplier is zero then the constraint can be removed altogether. It is for this reason we only consider clamping at the endpoints and not at the nodes.

3.6. *Gap Functionals and Their Directional Derivative*

We now compute the directional derivative of each gap functionals. We use the notation $v_{\bar{s}} =: (v^1, ..., v^n)$.

The directional derivative is

$$\begin{aligned} &D\Lambda_j(\theta, L, \bar{s})(v_\theta, v_L, v_{\bar{s}}) \\ &\quad = v_L \int_{s_{j-1}}^{s_j} e^{i\theta} + Lv^j e^{i\theta(s_j)} - Lv^{j-1} e^{i\theta(s_{j-1})} + iL \int_{s_{j-1}}^{s_j} v_\theta e^{i\theta} \\ &\quad = v_L \int_{s_{j-1}}^{s_j} e^{i\theta} + Lv^j e^{i\theta(s_j)} - Lv^{j-1} e^{i\theta(s_{j-1})} \\ &\qquad + iL \left\{ v_\theta(s_j) \int_{s_{j-1}}^{s_j} e^{i\theta} - \int_{s_{j-1}}^{s_j} \dot{v}_\theta \int_{s_{j-1}}^{s} e^{i\theta} \right\}. \end{aligned}$$

Let $\Omega_j = \Lambda_j^{-1}(0)$ and note the following important fact.

PROPOSITION 3.1. *If $p \in \Omega_j$ then $D\Lambda_j(p)$ is onto.*

Note that Ω_j depends on P_j and P_{j-1}. It is much easier to show $D\Lambda_j(p)$ is onto when P_j and P_{j-1} are distinct. In any case the proof is almost identical to the proof given in [16, Proposition 2.2].

Remark. An application of the implicit function theorem shows that Ω_j is a closed submanifold of $H \times R^+ \times \Delta^n$. It is at this stage important that all of H, R^+, and Δ are complete "locally," so that the contraction mapping principle applies.

3.7. *The Transpose of the Derivative of the Gap Functionals*

We now focus on the spaces normal to tangent spaces $T\Omega_j$. Note that, as a consequence of Proposition 3.1 and a standard fact in functional analysis, the normal spaces are given by the image of the transpose of DA_j. It is important to observe that, in our context, the set of complex numbers is considered as a vector space over R. If z and w are complex numbers then we use the real inner product $\langle z, w\rangle = \mathrm{Re}(z\bar{w})$, where $\bar{w}$ denotes the complex conjugate of w.

The transpose of DA_j is a map into $H \times R \times R^n$, and it has three components which simplify as follows when restricted to Ω_j:

$$(DA_j^T(\theta, L, \bar{s}))_\theta = \begin{cases} (s+1)\, i(P_j - P_{j-1}), & 0 \leqslant s \leqslant s_{j-1} \\ -iL \int_{s_{j-1}}^{s} \int_{s_{j-1}}^{u} e^{i\theta} + (s+1)\, i(P_j - P_{j-1}), & s_{j-1} \leqslant s \leqslant s_j \\ -iL \int_{s_{j-1}}^{s_j} \int_{s_{j-1}}^{u} e^{i\theta} + (s_j+1)\, i(P_j - P_{j-1}), & s_j \leqslant s \leqslant 1 \end{cases}$$

$$(DA_j^T(\theta, L, \bar{s}))_L = (P_j - P_{j-1})/L,$$

$$(DA_j^T(\theta, L, \bar{s}))_{\bar{s}} = (0, \ldots, 0, -Le^{i\theta(s_{j-1})}, Le^{i\theta(s_j)}, 0, \ldots, 0).$$

Here the nonzero terms appear at the $(j-1)$th and jth position (or one of them not at all when $j=1$ or $j=n+1$).

It is straightforward to check that the transpose satisfies the required relationship determined by the inner product. Note that to get the value of the transpose at the complex number λ, simply take the inner product of the previous expressions with λ, using the real inner product mentioned earlier.

3.8. *Transversality*

In the presence of nodes, the space of splines is represented by the intersection of the Ω_j's. In the case of fixed length, clamping, or pinning, yet another collection of closed submanifolds is created, and their intersections with the Ω_j's must be analyzed. To see that the subsets corresponding to fixed length, clamping, and pinning, indeed, are closed submanifolds of $H \times R^+ \times \Delta^n$, it suffices to consider the real valued functionals $L - \tilde{L}$, $\theta(t) - \theta_t$, and $s_j - \tilde{s}_j$, and their zero level sets. It is not hard to show that their directional derivatives are onto, and, as before, the claim follows from the implicit function theorem. Each considered constraint thus by itself corresponds to a closed submanifold. Now we like to give sufficient conditions to ensure that all the different closed submanifolds intersect transversally.

PROPOSITION 3.2. *Assume that both endpoints are fixed and, possibly, clamped. Then, each of the following conditions ensures that the corresponding subsets of $H \times R^+ \times \Delta^n$ are submanifolds*:

I. *No interior node is pinned and, either*

Ia. *the length is variable; or else*

Ib. *the length is fixed but not all the P_i lie on the same straight line.*

II. *All interior nodes are pinned and, either*

IIa. *the length is variable and no two fractions $|P_j - P_{j-1}|/(\tilde{s}_j - \tilde{s}_{j-1})$ are the same; or else*

IIb. *the length is fixed and $|P_j - P_{j-1}|/(\tilde{s}_j - \tilde{s}_{j-1})$ is always different from the length for any j.*

Proof. First consider case Ia. We show that the Ω_j's intersect transversally by proving the linear independence of the images under $D\Lambda_j^T$. Suppose $\sum_{j=1}^{n+1} D\Lambda_j^T(\lambda_j) = 0$ for arbitrary λ_j's. Differentiating twice in the θ-component of $D\Lambda_j^T$ results in $\langle -iLe^{i\theta(s)}, \lambda_j \rangle = 0$ for all s in (s_{j-1}, s_j). It follows that either $\theta(s) \equiv \theta_j$, a constant on (s_{j-1}, s_j), or $\lambda_j = 0$. From the $\bar{s}$-component we have the n conditions $\langle Le^{i\theta(s_j)}, \lambda_j \rangle = \langle Le^{i\theta(s_j)}, \lambda_{j+1} \rangle$. Combining this with the previous, we conclude that if one of the λ_j's are zero, all are zero. Suppose next that all λ_j's are different from zero. Then $\theta(s) = \theta$ is a constant on all of $[0, 1]$ because by the above we know that it is piecewise constant, but since $\theta(s)$ is also in H it must be continuous. The condition from the $\bar{s}$-component then implies that $\langle Le^{i\theta}, \lambda_j \rangle$ are equal for all j. Since $P_j - P_{j-1} = L(s_j - s_{j-1})\, e^{i\theta}$, the previous conditions combined with the fact that $\sum_{j=1}^{n+1} \langle (P_j - P_{j-1})/L, \lambda_j \rangle = 0$ from the L-component, creates a contradiction to the assumption on the λ_j's. The only possibility left is that $\lambda_j = 0$ for all j.

If other constraints are present, the previous reasoning must be modified. Clamping results in the addition of $\mu_0 + \mu_1(1+s)$ to the θ-component. Taking two derivatives will annihilate this linear term and the argument above will again show that all the λ_j's equal zero. It right away follows that $\mu_0 = \mu_1 = 0$.

Fixing L or $\bar{s}$ requires imposing additional restrictions on the P_j's. If the length is constrained then there is an additional term μ in the length component. Since everything else stays the same, we again conclude that all λ_j's equal zero, provided not all P_j's are on the same line, because this contradicts the fact that $\langle Le^{i\theta}, \lambda_j \rangle$ are equal for all j. It then follows that $\mu = 0$ and $\mu_0 = \mu_1 = 0$ as in the case of clamping.

Fixing $\bar{s}$ introduces constants σ_1 through σ_n in the n equations associated with the $\bar{s}$ component and the previous line of reasoning therefore crumbles. We may still, however, assert that the only way we can have nonzero

λ_j's is if $P_j - P_{j-1} = L(\tilde{s}_j - \tilde{s}_{j-1})\, e^{i\theta_j}$ for any such j. Taking magnitudes on both sides will generate the remaining conditions in the proposition. Note that in the case of variable length it is a priori possible that all but one of the λ_j's equal zero, but in this case the condition in the length component gives a contradiction.

Remark. It is an open question whether any of these conditions are necessary. The simplest open case is all curves of fixed length with exactly one node placed on the line segment between the two endpoints. It is hard to see why this would not be a manifold but the question is open.

3.9. *Jacobi Elliptic Functions*

As a preparation for the rest of the paper, we fix the notation and collect needed facts concerning the Jacobi elliptic functions. We will use sn, cn, dn as well as the elliptic integrals K and E. The definitions are given by

$$\mathrm{sn}^{-1}(s) = \int_0^s \frac{du}{\sqrt{1-u^2}\,\sqrt{1-p^2u^2}}$$

with an odd continuous periodic extension of sn to all of R. The modulus p is a number in $(0, 1)$. We have

$$\mathrm{cn}(x)^2 + \mathrm{sn}(x)^2 = 1$$

for any real x, hence cn is even, continuous and $\mathrm{cn}(0) = 1$. We also define

$$\mathrm{dn}(x)^2 + p^2\,\mathrm{sn}(x)^2 = 1$$

with dn even, continuous and $\mathrm{dn}(0) = 1$. Integrals of different kinds are given by

$$K = \int_0^1 \frac{ds}{\sqrt{1-s^2}\,\sqrt{1-p^2s^2}}$$

$$E(x) = \int_0^x \mathrm{dn}(u)^2\, du.$$

We also recall the derivative formulas

$$\mathrm{sn}'(x) = \mathrm{cn}(x)\,\mathrm{dn}(x)$$

$$\mathrm{cn}'(x) = -\mathrm{sn}(x)\,\mathrm{dn}(x)$$

$$\mathrm{dn}'(x) = -p^2\,\mathrm{cn}(x)\,\mathrm{sn}(x).$$

We need the following antiderivatives:

$$\int \operatorname{cn}(x)\,dx = \frac{1}{p}\sin^{-1}(p\operatorname{sn}(x))$$

$$\int \operatorname{dn}(x)\,dx = \operatorname{Sign}(\operatorname{cn}(x))\sin^{-1}(\operatorname{sn}(x)).$$

3.10. *The Differential Equation*

The nonlinear splines are characterized by the fact that they are solutions of the differential equation about to be given. As it turns out later, each interval (s_{j-1}, s_j) will be associated with its own boundary value problem. The purpose of the following proposition is to specify exactly which functions can appear as solutions to the boundary value problem. Note that in the special case $c=0$ all solutions are given by cn.

PROPOSITION 3.3. *Given any real numbers φ_0, $\dot{\varphi}_0$, and c the unique solution of the initial value problem,*

$$\ddot{\varphi}(s) + \frac{\varphi(s)^3}{2} + c\varphi(s) = 0, \qquad \varphi(\tilde{s}) = \varphi_0, \qquad \dot{\varphi}(\tilde{s}) = \dot{\varphi}_0$$

is given by either

I. $\varphi(s) = A\operatorname{cn}(\alpha(s-\tilde{s})+\beta)$ *of modulus* $p \in [0, 1]$ *with* $A\operatorname{cn}(\beta) = \varphi_0$, $-A\alpha\operatorname{sn}(\beta)\operatorname{dn}(\beta) = \dot{\varphi}_0$, *and* $A^2 = 4p^2\alpha^2$, $c = \alpha^2(1-2p^2)$, *or*

II. $\varphi(s) = A\operatorname{dn}(\alpha(s-\tilde{s})+\beta)$ *of modulus* $p \in [0, 1]$ *with* $A\operatorname{dn}(\beta) = \varphi_0$, $-A\alpha p^2\operatorname{sn}(\beta)\operatorname{cn}(\beta) = \dot{\varphi}_0$, *and* $A^2 = 4\alpha^2$, $c = \alpha^2(p^2-2)$.

If $(4c+\varphi_0^2)\,\varphi_0^2 + 4\dot{\varphi}_0^2 \geqslant 0$ *then the solution is given by case* I; *otherwise the solution is given by case* II.

Remark. It follows that if $\varphi_0 = 0$ or $4c+\varphi_0^2 \geqslant 0$ then the solution is given by case I, otherwise the solution is given either by case I or by case II. This fact is used in Theorem 4.1, because there this last inequality can be expressed solely in terms of multipliers and the given data.

Proof. The case $c=0$ with $p^2 = \frac{1}{2}$ is taken care of in [16], so we assume that $c \neq 0$. It may be verified, using the derivative formulas, that the suggested solutions indeed solve the initial value problem as long as the corresponding four relations are satisfied. Now we must show that given arbitrary φ_0, $\dot{\varphi}_0$, and $c \neq 0$ the constants A, α, β and the modulus $p \in [0, 1]$ can be chosen so that the above conditions are satisfied.

First square $-A\alpha\,\mathrm{sn}(\beta)\,\mathrm{dn}(\beta)=\dot{\varphi}_0$ and $-A\alpha p^2\,\mathrm{sn}(\beta)\,\mathrm{cn}(\beta)=\dot{\varphi}_0$, respectively, and use suitable identities for the elliptic functions, together with the remaining three conditions in each case to eliminate all but the modulus. We get two functions $F(p^2)=f(p^2)/4(1-2p^2)^2$ and $G(p^2)=g(p^2)/4(p^2-2)^2$, where

$$f(p^2)=(4p^2c-\varphi_0^2(1-2p^2))(4c(1-p^2)+\varphi_0^2(1-2p^2))$$
$$g(p^2)=(4c-\varphi_0^2(p^2-2))(\varphi_0^2(p^2-2)-4c(1-p^2)).$$

Note that $f(0)=f(1)=-\varphi_0^2(4c+\varphi_0^2)$ and $f(\frac{1}{2})=4c^2$. Also observe that $F(p^2)$ is symmetric about $p^2=\frac{1}{2}$. When $c>0$ we have $p^2<\frac{1}{2}$ and when $c<0$ we have $p^2>\frac{1}{2}$. Given p^2, the relations used between the constants may be used to first find α, then A, and finally β. Since the initial value problem has a unique solution it follows that there can only be at the most one solution to $F(p^2)=\dot{\varphi}_0^2$ and $G(p^2)=\dot{\varphi}_0^2$. If $\varphi_0=0$ we see that F is onto the nonnegative reals because it is singular at $p^2=\frac{1}{2}$. If $\varphi_0\neq 0$ but $4c+\varphi_0^2\geqslant 0$ then $F(0)=F(1)$ is negative so again F takes on all nonnegative values. Next suppose that $\varphi_0\neq 0$ and $4c+\varphi_0^2<0$. Since $F(p^2)$ is smooth and 1-1 on $(\frac{1}{2}, 1]$, its minimum must be at $p^2=1$. Note that $G(0)<0$ and $G(1)=F(1)=-(c+\varphi_0^2/4)\,\varphi_0^2$ so it follows that the two types of solutions together account for all possible values $\dot{\varphi}_0$. The inequality in the Proposition is essentially the statement $F(1)\leqslant\dot{\varphi}_0^2$. Once the modulus is determined simply retrace the steps and determine the values of the constants eliminated above.

Remark. In what follows the boundary values are not given explicitly but rather in terms of the multipliers. It follows that both types of solutions may occur in the spline. In terms of computational efficiency this is unfortunate since in the worst case one would have to examine all 2^{n+1} possibilities.

4. The Fundamental Equation and All Constraints Imposed

In this section we consider the case of curves of fixed length which are *not* allowed to slide at the nodes (if any). No sliding is allowed at the endpoints, and if the directions are given at the endpoints there will be two multipliers present (μ_0 and μ_1), along with two extra conditions. Without such clamping simply set the two multipliers equal to zero and drop the two extra conditions.

If the curves considered are periodic, so that the two endpoints are the same and the directions at the endpoints are identical, then we have $\mu_0+\mu_1=0$ and only one extra condition.

In between the nodes the curve is given by either one of a pair of Jacobi elliptic functions and two sets of parameters α_j and β_j. Both the "amplitudes" A_j, and the moduli p_j are related to the α_j's. All these quantities are tied together with the complex multipliers λ_j, real multipliers σ_j, μ, and in the case of clamping also with μ_0 and μ_1. The following theorem gives necessary and sufficient conditions for these parameters and multipliers so that the corresponding curves are splines satisfying all the constraints. It is assumed that $s_0 = 0$ and $s_{n+1} = 1$. Conditions for when the constraints result in subsets of $H \times R^+ \times \Delta^n$ which are closed submanifolds are given in Proposition 3.2. To simplify many of the expressions we introduce $n+1$ complex numbers by $g_j := P_j - P_{j-1}$.

THEOREM 4.1. *Given $n+2$ points $P_0, ..., P_{n+1}$ and a fixed $\tilde{L} > 0$ and a fixed $\tilde{s} = (\tilde{s}_1, ..., \tilde{s}_n) \in \Delta^n$. Assume these data are such that the corresponding constrained set is a submanifold of $H \times R^+ \times \Delta^n$. In the case of clamping also assume that the two real numbers θ_0 and θ_1 are given. Let $g_j := P_j - P_{j-1}$. The element $(\theta, L, \bar{s}) \in H \times R^+ \times \Delta^n$, such that $L \int_{s_{j-1}}^{s_j} e^{i\theta} = g_j$ for $j = 1, ..., n+1$, with $L = \tilde{L}$ and $\bar{s} = \tilde{s}$ (and $\theta(0) = \theta_0$, $\theta(1) = \theta_1$ if clamping) is a critical point of the functional $J(\theta, L, \bar{s}) = (1/2L) \int_0^1 \dot{\theta}(s)^2 \, ds$ when restricted to the constrained set if and only if $\theta_j = \theta|_{(\tilde{s}_{j-1}, \tilde{s}_j)}$ satisfies $\dot{\theta}_j(s) = A_j f_j(\alpha_j(s - \tilde{s}_{j-1}) + \beta_j)$, with the function f_j one* (i) *or possibly the other* (ii) *of the Jacobi elliptic functions* cn (i) *or* dn (ii). *We have* (i) *for sure if, with $C_j = \mu\tilde{L} + \sum_{k=j}^{n} (1 - \tilde{s}_k) \sigma_k - \sum_{k=1}^{j-1} \tilde{s}_k \sigma_k$,*

$$\left(\mu_0 + \sum_{k=1}^{j-1} \langle ig_k, \lambda_k \rangle \right)^2 \geqslant \frac{4C_j}{\tilde{L}}$$

or $\mu_0 + \sum_{k=1}^{j-1} \langle ig_k, \lambda_k \rangle = 0$; otherwise we may have case (ii).

In each case (i), *the modulus and the "amplitude" are given by*

$$p_j^2 = \frac{1}{2}\left(1 + \frac{\tilde{L}C_j}{\alpha_j^2}\right)$$

$$A_j^2 = 4p_j^2\alpha_j^2.$$

In each case (ii), *we have*

$$p_j^2 = 2 - \frac{\tilde{L}C_j}{\alpha_j^2}$$

$$A_j^2 = 4\alpha_j^2.$$

We let $q_j = p_j$ in case (i) *and $q_j = 1$ in case* (ii). *The parameters α_j, β_j, the elliptic moduli p_j, together with the complex multipliers λ_j and the real multipliers μ_0 and μ_1, are required to satisfy*

$$\text{I.}\quad \sum_{j=1}^{n+1} |g_j|\,|\lambda_j| \sin(\arg(g_j) - \arg(\lambda_j)) = \mu_0 + \mu_1$$

$$\text{IIa.}\quad \pm\alpha_j 2q_j f_j(\beta_j) = \tilde{L}\left\{\mu_0 + \sum_{k=1}^{j-1} \langle ig_k, \lambda_k\rangle\right\}$$

$$\text{IIb.}\quad \pm\alpha_j 2q_j f_j(\alpha_j(\tilde{s}_j - \tilde{s}_{j-1}) + \beta_j) = \tilde{L}\left\{\mu_0 + \sum_{k=1}^{j} \langle ig_k, \lambda_k\rangle\right\}$$

$$\text{IIIa.}\quad \alpha_j^2 p_j^2 = \tilde{L}^2 q_j^2\,|\lambda_j|$$

$$\text{IIIb.}\quad \pm 2\sin^{-1}(q_j \operatorname{sn}(\alpha_j(\tilde{s}_j - \tilde{s}_{j-1}) + \beta_j)) + \arg(\lambda_j) = \pm 2\sin^{-1}(q_{j+1}\operatorname{sn}(\beta_{j+1})) + \arg(\lambda_{j+1})$$

$$\text{IIIc.}\quad \pm 2\sin^{-1}(q_1 \operatorname{sn}(\beta_1)) + \arg(\lambda_1) = \theta_0$$

$$\text{IIId.}\quad \pm 2\sin^{-1}(q_{n+1}\operatorname{sn}(\alpha_{n+1}(1 - \tilde{s}_n) + \beta_{n+1})) + \arg(\lambda_{n+1}) = \theta_1$$

$$\text{IV(i).}\quad 2\tilde{L}\{E(\alpha_j(\tilde{s}_j - \tilde{s}_{j-1}) + \beta_j) - E(\beta_j)\} - \alpha_j\tilde{L}(\tilde{s}_j - \tilde{s}_{j-1}) = \alpha_j\,|g_j| \cos(\arg(g_j) - \arg(\lambda_j))$$

$$\text{IV(ii).}\quad 2\tilde{L}\{E(\alpha_j(\tilde{s}_j - \tilde{s}_{j-1}) + \beta_j) - E(\beta_j)\} + \alpha_j\tilde{L}(p_j^2 - 2)(\tilde{s}_j - \tilde{s}_{j-1}) = \alpha_j p_j^2\,|g_j| \cos(\arg(g_j) - \arg(\lambda_j)).$$

Remark 1. Note that there are $2(n+1)$ unknown parameters α_j and β_j. The moduli p_j correspond to another $n+1$ unknown. The complex multipliers λ_j yet another $2(n+1)$ unknowns. Including the real multipliers μ_0 and μ_1, we get a total of $5(n+1)+2$ unknowns. The number of equations is given by: $\mathrm{I} = 1$, $\mathrm{II} = 2(n+1)$, $\mathrm{IIIa} = (n+1)$, $\mathrm{IIIb} = n$, $\mathrm{IV} = (n+1)$, which with IIIc, IIId add up to $5(n+1)+2$.

Remark 2. Equations I–IV and the equations for the p_j^2's is the "fundamental finite-dimensional equation" we referred to in Section 0.2.6.

It is also worth pointing out separately the following basic fact which is shown in the proof of the theorem.

COROLLARY 4.2. *If $(\theta, L, \bar{s}) \in H \times R^+ \times \Delta^n$ is critical subject to the constraints, then $\dot{\theta}$ is continuous.*

It follows that a nonlinear spline of the above kind, when regarded as a curve in the plane, is at least C^2.

Proof. In order for $(\theta, L, \bar{s})$ to be critical, the gradient vector field associated with J, when projected onto the tangent space of the submanifold corresponding to the constraints, must vanish. When this is expressed explicitly, we get three kinds of conditions, one for each component in the tangent space of $H \times R^+ \times \Delta^n$. Assume that $p = (\theta, \tilde{L}, \tilde{s}) \in H \times R^+ \times \Delta^n$ is critical and that it satisfies the constraints

$$\tilde{L} \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} e^{i\theta} = g_j \tag{1}$$

and possibly also

$$\theta(0) = \theta_0 \tag{2}$$

$$\theta(1) = \theta_1. \tag{3}$$

The element $(\theta, \tilde{L}, \tilde{s}) \in H \times R^+ \times \Delta^n$ being critical subject to the constraints is equivalent to the existence of complex numbers λ_j, and real numbers σ_j, μ, μ_0, and μ_1 such that

$$\frac{1}{\tilde{L}} (\theta(s) - \theta(0)) + \mu_0 + \mu_1 (s+1)$$

$$= \sum_{j=1}^{n+1} \begin{cases} \langle -(s+1)\, ig_j, \lambda_j \rangle, & 0 \leqslant s \leqslant \tilde{s}_{j-1}, \\ \left\langle i\tilde{L} \int_{\tilde{s}_{j-1}}^{s} \int_{\tilde{s}_{j-1}}^{u} e^{i\theta} - (s+1)\, ig_j, \lambda_j \right\rangle, & \tilde{s}_{j-1} \leqslant s \leqslant \tilde{s}_j, \\ \left\langle i\tilde{L} \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} \int_{\tilde{s}_{j-1}}^{u} e^{i\theta} - (\tilde{s}_j + 1)\, ig_j, \lambda_j \right\rangle, & \tilde{s}_j \leqslant s \leqslant 1, \end{cases} \tag{4}$$

$$\sum_{j=1}^{n+1} \langle g_j, \lambda_j \rangle = J(p) - \mu \tilde{L} \tag{5}$$

$$\langle \tilde{L} e^{i\theta(\tilde{s}_j)}, \lambda_j \rangle - \langle \tilde{L} e^{i\theta(\tilde{s}_j)}, \lambda_{j+1} \rangle + \sigma_j = 0. \tag{6}$$

Next we use these equations to deduce necessary conditions.

If we let $s = 0$ in (4) we get

$$\mu_0 + \mu_1 = -\sum_{j=1}^{n+1} \langle ig_j, \lambda_j \rangle. \tag{7}$$

When (7) is spelled out we get I.

Note that the right-hand side of (4) is differentiable and it follows that a critical θ must be differentiable.

Differentiating gives us

$$\frac{1}{\tilde{L}}\,\dot{\theta}(s)+\mu_1=\sum_{j=1}^{n+1}\begin{cases}\langle -ig_j,\lambda_j\rangle, & 0\leqslant s\leqslant \tilde{s}_{j-1},\\ \left\langle i\tilde{L}\int_{\tilde{s}_{j-1}}^{s} e^{i\theta}-ig_j,\lambda_j\right\rangle, & \tilde{s}_{j-1}\leqslant s\leqslant \tilde{s}_j,\\ 0, & \tilde{s}_j\leqslant s\leqslant 1.\end{cases} \tag{8}$$

Observe that (8) and (1) show that a critical θ must have $\dot{\theta}$ continuous, which proves Corollary 4.2.

Also note that by a standard bootstrap argument θ must be C^∞ on each interval $(\tilde{s}_{j-1},\tilde{s}_j)$. On any such interval the right-hand side is C^1, which forces the left-hand side to be C^2, but then the right-hand side is C^3 and so on and so forth.

Let θ_j denote the restriction of θ to this interval. If we differentiate we get

$$\frac{\ddot{\theta}_j}{\tilde{L}}=\langle i\tilde{L}e^{i\theta_j},\lambda_j\rangle. \tag{9}$$

Since $\dot{\theta}_j$ is an integrating factor we can integrate to get

$$\frac{\dot{\theta}_j^2}{2\tilde{L}}=\langle \tilde{L}e^{i\theta_j},\lambda_j\rangle+C_j. \tag{10}$$

If we combine (10) for all j and integrate from 0 to 1 and (5), (1) are used, we see that $\sum_{j=1}^{n+1} C_j(\tilde{s}_j-\tilde{s}_{j-1})=\mu\tilde{L}$.

The conditions (6) show that $C_j-C_{j+1}=\sigma_j$, so, with the previous equation, we have $n+1$ linear equations for the C_j's in terms of the σ_j's.

The solutions are given by $C_j=\mu\tilde{L}+\sum_{k=j}^{n}(1-\tilde{s}_k)\,\sigma_k-\sum_{k=1}^{j-1}\tilde{s}_k\sigma_k$.

Differentiating (9) and combining with (10) gives us

$$\dddot{\theta}_j(s)+\frac{\dot{\theta}_j(s)^3}{2}-\tilde{L}C_j\dot{\theta}_j(s)=0. \tag{11}$$

According to Proposition 3.3 in Section 3.10, all the solutions of (11) have the form:

$$\dot{\theta}_j(s)=A_jf_j(\alpha_j(s-\tilde{s}_{j-1})+\beta_j), \tag{12}$$

where f_j is either cn if $\dot{\theta}_j(\tilde{s}_{j-1})^2\geqslant 4\tilde{L}C_j$ or $\dot{\theta}(\tilde{s}_{j-1})=0$, or possibly dn in the other case.

The "amplitudes" are in each case given by

$$A_j^2=4p_j^2\alpha_j^2,\qquad A_j^2=4\alpha_j^2. \tag{13}$$

The moduli in the respective cases are given by

$$-\tilde{L}C_j = \alpha_j^2(1-2p_j^2), \qquad -\tilde{L}C_j = \alpha_j^2(p_j^2-2). \tag{14}$$

Using (7) and (8) we extract the values

$$\dot{\theta}(\tilde{s}_j) = \tilde{L}\left\{\mu_0 + \sum_{k=1}^{j} \langle ig_k, \lambda_k\right\} \tag{15}$$

and we have $(n+1)$ boundary value problems, one for each θ_j.

When (13) and (14) are expressed at both endpoints of $(\tilde{s}_{j-1}, \tilde{s}_j)$ for each j we get the conditions II.

Recalling from 3.9 the formula for the antiderivatives of cn, dn and using (12) we get

$$\theta_j(s) = \pm 2\sin^{-1}(q_j \operatorname{sn}(\alpha_j(s-\tilde{s}_{j-1})+\beta_j)) + D_j, \tag{16}$$

where $q_j = p_j$ in case of cn and $q_j = 1$ in case of dn. To be precise, the sign of cn should also enter into (16) in case (ii) when dn is used. The same sign will appear as a consequence of taking a square root of a square below. Instead of dealing with both the sign function and an absolute value we permit ourselves to be sloppy and leave out both. The cancellation will take place but this point did at one time confuse the author. The antiderivative formula in 3.9 is typically not stated with the sign of cn as we have indicated in 3.9.

To simplify the display of some of the following equations we let $x := (\alpha_j(s-\tilde{s}_{j-1})+\beta_j)$ and $y := (\alpha_j(\tilde{s}_j-\tilde{s}_{j-1})+\beta_j)$.

With the use of algebra and elliptic identities the two cases give us

$$e^{i\theta_j} = (\operatorname{dn}(x)^2 - p_j^2\operatorname{sn}(x)^2 \pm i2p_j\operatorname{sn}(x)\operatorname{dn}(x))\, e^{iD_j} \tag{17a}$$

$$e^{i\theta_j} = (\operatorname{cn}(x)^2 - \operatorname{sn}(x)^2 \pm i2\operatorname{sn}(x)\operatorname{cn}(x))\, e^{iD_j}. \tag{17b}$$

If (12) is differentiated and (9) is rewritten using (17a), (17b) we get an identity which can only be satisfied for all $s \in (\tilde{s}_{j-1}, \tilde{s}_j)$ if $D_j = \arg(\lambda_j)$ and $\alpha_j^2 p_j^2 = \tilde{L}^2 q_j^2\, |\lambda_j|$.

To see this, it is convenient to argue under two different circumstances. First suppose that there is an $s \in (\tilde{s}_{j-1}, \tilde{s}_j)$ at which sn vanishes. At such an s we must have $\langle i\tilde{L}e^{iD_j}, \lambda_j\rangle = 0$. Using this fact for all $s \in (\tilde{s}_{j-1}, \tilde{s}_j)$ the conclusion follows.

In the contrary case, sn vanishes nowhere in $(\tilde{s}_{j-1}, \tilde{s}_j)$. In case (i) we can divide by $2p_j\operatorname{sn}(x)\operatorname{dn}(x)$ and use the fact that $\operatorname{cn}(x)^2$ and $2p_j\operatorname{sn}(x)\operatorname{dn}(x)$ are linearly independent to again deduce that $\langle i\tilde{L}e^{iD_j}, \lambda_j\rangle = 0$, so again the conclusion follows.

In case (ii) we first dispose of when for some s we have $\mathrm{cn}(x)=0$ in the same way as previously. In the remaining circumstances we can divide by $2\,\mathrm{sn}(x)\,\mathrm{cn}(x)$, and use linear independence the same way as before. This gives us IIIa and the continuity of θ give us IIIb in the theorem. In the case of clamping we get IIIc, IIId.

The constraints (1) are responsible for the conditions IV which result from an integration of (17a), (17b) over $s \in (\tilde{s}_{j-1}, \tilde{s}_j)$. Note that the second term appearing in (17a), (17b) can be integrated explicitly. The resulting term can be rewritten with the help of II and IIIa so that we have

$$\frac{|g_j|}{\tilde{L}}\, e^{i(\arg(g_j)-\arg(\lambda_j))}$$

$$= \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} \mathrm{dn}(x)^2 - p_j^2\,\mathrm{sn}(x)^2\,ds - \frac{i}{\tilde{L}\,|\lambda_j|}\langle ig_j, \lambda_j\rangle \tag{18a}$$

$$\frac{|g_j|}{\tilde{L}}\, e^{i(\arg(g_j)-\arg(\lambda_j))}$$

$$= \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} \mathrm{cn}(x)^2 - \mathrm{sn}(x)^2\,ds - \frac{i}{\tilde{L}\,|\lambda_j|}\langle ig_j, \lambda_j\rangle \tag{18b}$$

The imaginary parts of both sides of the equations are seen to be equal. The real parts must also be equal. Using an elliptic identity the integral term can be rewritten in terms of a standard elliptic integral, and IV follows. This concludes the proof that I–IV all are necessary conditions.

In order to prove sufficiency we assume that there are parameters α_j, β_j, together with multipliers λ_j, σ_j, μ, and possibly μ_0, μ_1 satisfying I–IV of the theorem. Next we define θ in terms of α_j, β_j by

$$\theta_j(s) = \pm 2\sin^{-1}(q_j\,\mathrm{sn}(x)) + \arg(\lambda_j). \tag{19}$$

First we observe that θ is well defined because if (19) is differentiated IIa, IIb imply that $\dot{\theta}$ is continuous.

Next we show that the constraints are satisfied. The definition of θ_j yields

$$e^{i\theta_j} = (\mathrm{dn}(x)^2 - p_j^2\,\mathrm{sn}(x)^2 \pm i2p_j\,\mathrm{sn}(x)\,\mathrm{dn}(x))\,e^{i\arg(\lambda_j)} \tag{20a}$$

$$e^{i\theta_j} = (\mathrm{cn}(x)^2 - \mathrm{sn}(x)^2 \pm i2\,\mathrm{sn}(x)\,\mathrm{cn}(x))\,e^{i\arg(\lambda_j)}. \tag{20b}$$

If (20a), (20b) is integrated over $(\tilde{s}_{j-1}, \tilde{s}_j)$ and IV is used we see that (1) is satisfied. In the case of clamping IIIc, IIId imply (2) and (3).

Before we prove criticality, we state a crucial relationship between the moduli p_j, the multipliers λ_j, σ_j, μ, and the length $\tilde{L}$:

$$\text{(i)}\quad 2p_j^2 = 1 + \frac{C_j}{\tilde{L}\,|\lambda_j|}, \qquad \text{(ii)}\quad \frac{2}{p_j^2} = 1 + \frac{C_j}{\tilde{L}\,|\lambda_j|}. \tag{21}$$

With the use of (21) we get

$$\frac{\dot{\theta}_j^2}{2\tilde{L}} = (\tilde{L}\,|\lambda_j| + C_j)\,\mathrm{cn}(x)^2 \tag{22a}$$

$$\frac{\dot{\theta}_j^2}{2\tilde{L}} = (\tilde{L}\,|\lambda_j| + C_j)\,\mathrm{dn}(x)^2. \tag{22b}$$

Using elliptic identities we also get

$$\begin{aligned} \mathrm{dn}(x)^2 - p^2\,\mathrm{sn}(x)^2 &= 2p^2\,\mathrm{cn}(x)^2 + (1-2p^2) \\ &= \left(1 + \frac{C_j}{\tilde{L}\,|\lambda_j|}\right)\mathrm{cn}(x)^2 - \frac{C_j}{|\tilde{L}\lambda_j|} \end{aligned} \tag{23a}$$

$$\begin{aligned} \mathrm{cn}(x)^2 - \mathrm{sn}(x)^2 &= \frac{2}{p^2}\,\mathrm{dn}(x)^2 - \left(\frac{2}{p^2} - 1\right) \\ &= \left(1 + \frac{C_j}{\tilde{L}\,|\lambda_j|}\right)\mathrm{dn}(x)^2 - \frac{C_j}{\tilde{L}\,|\lambda_j|}. \end{aligned} \tag{23b}$$

In order to prove criticality we must show (4)–(6), and we begin with (6). With the help of (20a), (20b) we get

$$\langle e^{i\theta_j(s_j)}, \lambda_j\rangle = |\lambda_j|\,(\mathrm{dn}(y)^2 - p_j^2\,\mathrm{sn}(y)^2) \tag{24a}$$

$$\langle e^{i\theta_j(s_j)}, \lambda_j\rangle = |\lambda_j|\,(\mathrm{cn}(y)^2 - \mathrm{sn}(y)^2) \tag{24b}$$

$$\langle e^{i\theta_{j+1}(s_j)}, \lambda_{j+1}\rangle = |\lambda_{j+1}|\,(\mathrm{dn}(\beta_{j+1})^2 - p_j^2\,\mathrm{sn}(\beta_{j+1})^2) \tag{25a}$$

$$\langle e^{i\theta_{j+1}(s_j)}, \lambda_{j+1}\rangle = |\lambda_{j+1}|\,(\mathrm{cn}(\beta_{j+1})^2 - \mathrm{sn}(\beta_{j+1})^2). \tag{25b}$$

Using (24a), (24b), IIa, IIb, and IIIa we see that the right-hand sides of (24a), (24b) and (25a), (25b) are equal so (6) follows.

Next we prove (5). If J is expressed in terms of f_j we have

$$J(p) = \frac{1}{2\tilde{L}} \sum_{j=1}^{n+1} A_j^2 \int_{\tilde{s}_{j-1}}^{\tilde{s}_j} f_j(x)^2\,ds. \tag{26}$$

Combining (26) with (22a), (22b) we see that we get an expression which can be rewritten using (23a), (23b) backwards so that if the real part of (19a), (19b) is used backwards (5) follows. Note that (18a), (18b) are a consequence of the constraint (1).

It is straightforward to show (9) which when antidifferentiated gives us

$$\frac{\dot{\theta}_j(s)}{\tilde{L}} - \frac{\dot{\theta}_j(\tilde{s}_{j-1})}{\tilde{L}} = \left\langle i\tilde{L}\int_{\tilde{s}_{j-1}}^{s} e^{i\theta_j}, \lambda_j \right\rangle.$$

The $(n+1)$ subintervals can now be combined and with the help of (7) (which really is I) we get (8). Another antidifferentiation and one more application of (7) shows (4). This concludes the proof of the theorem.

5. Not All Constraints Imposed

The remarks preceding Theorem 4.1 regarding clamping and periodicity apply everywhere in this section. Note that in all the cases below we again have that $\dot{\theta}$ is *continuous*.

5.1. *Pinned but with Variable Length*

Consider the case of curves of variable length which are *not* allowed to slide at the interior nodes (if any). Theorem 4.1 and its proof must be modified in the following way. Replace $\tilde{L}$ by L everywhere and let $\mu = 0$ everywhere. Note in this case that there are $2(n+1)$ parameters α_j and β_j. The elliptic moduli p_j correspond to $n+1$ unknowns and the complex multipliers λ_j yet another $2(n+1)$. There are n real multipliers σ_j and two real multipliers μ_0 and μ_1. With the length L we get a total of $6(n+1)+2$ unknowns. As in the remark following Theorem 4.1 we have $5(n+1)+2$ equations and in this case we also include the $n+1$ equations for the p_j^2 for a total of $6(n+1)+2$.

5.2. *Fixed Length but No Pinning*

Consider the case of curves of fixed length which are allowed to slide at the interior nodes (if any). Theorem 4.1 and its proof must be modified in the following way. Replace $\tilde{s}$ by $\bar{s}$ everywhere and let $\sigma_j = 0$ everywhere. Note that $C_j = \mu\tilde{L}$ for all j's. It follows that the inequality changes to $(\mu_0 + \sum_{k=1}^{j-1} \langle ig_k, \lambda_k \rangle)^2 \geqslant 4\mu$ so if $\mu \leqslant 0$ only the cn is present along the spline.

Note that there are $2(n+1)$ unknown parameters α_j and β_j. The $\bar{s}$ corresponds to n unknown and the elliptic moduli p_j another $n+1$. There are $2(n+1)$ unknown due to the complex multipliers λ_j. With the real multipliers μ, μ_0, and μ_1 we get a total of $6(n+1)+2$. This is equal to the number of equations since as in 5.1 we include the $n+1$ equations for the p_j^2.

5.3. *Variable Length and No Pinning*

Consider the case of curves of variable length which are allowed to slide at the interior nodes (if any). Theorem 4.1 and its proof must be modified in the following way. Replace $\tilde{L}$ by L and $\tilde{s}$ by $\bar{s}$ everywhere. Let $\mu=0$ and let $\sigma_j=0$ everywhere. Note that $C_j=0$ for all j's. This implies that $p_j^2=\frac{1}{2}$ for all j's and only the cn function is needed.

This last case is clearly the simplest, but remember that the existence of critical points is not guaranteed when the length is variable.

5.4. *Variable Length, No Interior Nodes and Clamping at One Endpoint*

This last section illustrates how Theorem 4.1 is used to deal with the special case discussed in Section 1.1.5. Since there are no interior nodes we let $n=0$ and consider curves between $P_0=(0,0)$ and $P_1=(1,0)$. Variable length implies that $\mu=0$. No clamping at P_0 implies that $\mu_0=0$. As in 5.3 we need only consider cn functions of modulus $p^2=\frac{1}{2}$. In Eq. IIa the summation index varies over an empty set so $\mu_0=0$. This implies $\mathrm{cn}(\beta)=0$. We choose $\beta=-K$. When Eqs. I and IIb are combined we get $\pm\sqrt{2}\alpha\,\mathrm{cn}(\alpha-K)=L\,|\lambda|\sin(\arg(\lambda))$. Equation IIIa becomes $\alpha^2=L^2\,|\lambda|$ and IIId, $\pm 2\sin^{-1}(1/\sqrt{2}\,\mathrm{sn}(\alpha-K))+\arg(\lambda)=\theta_1$. Both IIIb and IIIc are without content. Finally, Eq. IV becomes $2L(E(\alpha-K)-E(K))-\alpha L=\alpha\cos(\arg(\lambda))$.

All the unknown but α can be eliminated. To see how, divide the sin equation by the cos equation. Next replace the coefficient in front of tan using IIIa and note how $|\lambda|$ drops out. Finally, use IIId to replace $\arg(\lambda)$. When this final equation is rewritten with everything but θ_1 to the left of the equal sign, the left side is equal to $f[x]$ in Fig. 1. Once α is determined all other quantities, such as the length, the value of the functional and the curve itself, can be computed explicitly; see Fig. 2.

As an example consider the length formula. Use the sin equation, get rid of $|\lambda|$ and $\arg(\lambda)$ using IIIa and IIId, and solve for L. A similar computation yields a second formula for the length. It is used in case the denominator of the first formula is zero. Once the length is determined the value of the functional can be found by integrating $\dot{\theta}^2$. To get the curve itself, integrate $\dot{\theta}$ first and then $e^{i\theta}$. The constants of integration have values so the constraints are satisfied.

References

1. G. Birkhoff and C. de Boor, Piecewise polynomial interpolation and approximation, *in* "Approximation of Functions" (H. L. Garabedian, Ed.), pp. 164–190, Elsevier, New York/Amsterdam, 1965.
2. G. Brunnett, A new characterization of plane elastica, *in* "Mathematical Methods in Computer Aided Geometric Design II" (T. Lyche and L. Schumaker, Eds.), pp. 43–56, Academic Press, New York, 1992.
3. G. Brunnett, "The Curvature of Plane Elastic Curves," Tech. Report NPS-MA-93-013, Naval Postgraduate School, 1993.
4. G. Brunnett, Properties of minimal energy splines, *in* "Curve and Surve Design" (H. Hagen, Ed.), pp. 3–22, SIAM, Philadelphia, 1992.
5. M. Golomb and J. W. Jerome, Equilibria of the curvature functional and manifolds of nonlinear interpolating spline curves, *SIAM J. Math. Anal.* **13** (1982), 421–458.
6. B. K. P. Horn, The curve of least energy, *ACM Trans. Math. Software* **9** (1983), 441–460.
7. J. W. Jerome, Minimization problems and linear and nonlinear spline functions. I. Existence, *SIAM J. Numer. Anal.* **10** (1973), 808–819.
8. J. W. Jerome, Smooth interpolating curves of prescribed length and minimum curvature, *Proc. Amer. Math. Soc.* **51** (1975), 62–66.
9. E. D. Jou and W. Han, Minimal-energy splines. I. Plane curves with angle constraints, *Math. Methods Appl. Sci.* **13** (1990), 351–372.
10. J. Langer and D. A. Singer, Curve-straightening and a minimax argument for closed elastic curves, *Topology* **24** (1985), 75–88.
11. E. H. Lee and G. E. Forsythe, Variational study of nonlinear spline curves, *SIAM Rev.* **15** (1973), 120–133.
12. A. Linnér, Some properties of the curve straightening flow in the plane, *Trans. Amer. Math. Soc.* **314**, No. 2 (1989), 605–617.
13. A. Linnér, Steepest descent used as a tool to find critical points of $\int k^2$ defined on curves in the plane with arbitrary types of boundary conditions, *in* "Geometric Analysis and Computer Graphics," MSRI Publ., Vol. 17, pp. 127–138, Springer-Verlag, New York/Berlin, 1991.
14. A. Linnér, Curve-straightening in closed Euclidean submanifolds, *Commun. Math. Phys.* **138** (1991), 33–49.
15. A. Linnér, Curve-straightening, *in* "Proc. Symp. Pure Math.," Vol. 54, III, pp. 451–458, Amer. Math. Soc., Providence, RI, 1993.
16. A. Linnér, Existence of free non-closed Euler–Bernoulli elastica, *Nonlinear Anal.* **21**, No. 8 (1993), 575–593.