PERGAMON

Mathematical and Computer Modelling 33 (2001) 237–253

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.nl/locate/mcm

Optimal Control, Statistics and Path Planning

C. F. Martin* and Shan Sun
Department of Mathematics and Statistics
Texas Tech University, Lubbock, TX 79409, U.S.A.

M. Egerstedt†
Optimization and Systems Theory
Royal Institute of Technology
SE-100 44 Stockholm, Sweden

Abstract—In this paper, some of the relationships between optimal control and statistics are examined. In a series of earlier papers, we examined the relationship between optimal control and conventional splines and between optimal control and the statistical theory of smoothing splines. In this paper, we present a unified treatment of these two problems and extend the same framework to include the concept of "dynamic time warping", which is being seen as an important statistical tool as well as being of importance in physics. We show that these three major problems unite to give a satisfactory solution to the problem of trajectory or path planning.

Keywords—Optimal control, Data interpolation, Curve registration, Path planning.

1. INTRODUCTION

In this paper, we consider a series of eight problems that have their origins in optimal control theory, statistics, and numerical analysis. In the papers [1,2], we have exploited the fact that splines and linear optimal control theory as well as controllability are very closely related to the classical numerical theory of splines. In the papers [3,4], we have exploited the fact that optimal control theory and the statistical theory of smoothing splines are closely related. Our attention was drawn to the paper [5] and the book [6], from which we noted that the basic idea of curve registration or "dynamic time warping" were again very closely related to the problem of optimal control theory. In this paper, we give a unified treatment to the eight problems. This is not to say that we have effectively solved all eight problems, for some of the problems are of the nature that the main goal will be to develop effective numerical algorithms. The main contribution of this paper is to show that all eight of the problems are very similar in nature if not in solution.

In Problem 1, we show that the theory of interpolating splines is naturally considered as a problem of minimizing a quadratic cost functional subject to a set of linear constraints. For polynomial splines, that was the original concept in their development. For Problem 3, we show

*Supported in part by NSF Grants ECS 9705312 and 9720357 and NASA Grant NAG2-902. The support of KTH during the spring of 1998 is gratefully acknowledged.
†The support of the Swedish Foundation for Strategic Research through its Centre for Autonomous Systems is gratefully acknowledged.

0895-7177/01/$ - see front matter Typeset by $\mathcal{AMS}$-TEX
PII: S0895-7177(00)00241-7

that the theory of smoothing splines can be considered as being very close to the theory of interpolating splines, with the difference being that the linear constraints are included in the cost functional as a penalty term. The optimization problem is a straightforward problem of minimizing a quadratic cost functional over the space of square integrable functions. In Problems 2 and 4, we show that the problem of constructing splines that pass through intervals instead of points can be reduced to the problem of minimizing a quadratic cost functional subject to a set of inequality constraints. This problem is then reduced to the problem of solving a quadratic programming problem. In Problem 5, we show that the problem of constructing splines that are nondecreasing at the nodes is again reducible to the problem of minimizing a quadratic cost functional with inequality constraints, and hence, reducible to a quadratic programming problem. In Problem 6, we simply restate the curve registration problem of Li and Ramsay [5] as an optimal output tracking problem. Here, the problems become noticeably harder. The cost functional is now nonlinear as well as the differential constraint. In Problem 7, we formally state the output tracking problem and show that Problem 6 is a special case. Finally, in Problem 8 we state a version of the path planning problem. The statement of the problem involves the previous seven problems. We do not explicitly solve the problem, but we do present an algorithm that will at least produce a suboptimal problem. We suggest, based on the solutions of the first five problems, that it suffices to reduce the problem to a nonlinear optimization problem over a finite parameter space. This gives some support to ideas in [5,6] that one can use splines to solve the problem. We use a slightly different pseudo time function than was used in the two references, but it is clear that our formulation owes its existence to the development of Li, Ramsay, and Silverman.

2. NOTATION

In this section, we establish some notation concerning the control system and the data sets which we will consider. We will not consider the most complicated cases in this paper but will restrict ourselves to a fairly simple situation. We will only consider the univariate case in this paper.

We will assume as given throughout this paper a linear dynamical system of the form

$$\dot{x} = Ax + bu, \tag{1}$$

$$y = cx, \tag{2}$$

where $x \in \mathbb{R}^n$, u, and y are scalar functions and A, b, and c are constant matrices of compatible dimension. We will further assume that

$$cb = cAb = cA^2b = \cdots = cA^{n-2}b = 0. \tag{3}$$

This condition can be relaxed, but the exposition is simplified with this assumption. For the most part, we will be considering systems for which

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & & \cdots & \alpha_n \end{pmatrix},$$

$b^\top = (0, 0, \ldots, 0, 1)$, and $c = (1, 0, 0, \ldots, 0)$. Any system which satisfies the zero constraints of (3) is equivalent to a system with this form. The case in which all of the α_is are zero plays a particularly important role, for then the solutions to the differential equation are simply polynomials convolved with the control function u. Under this assumption, all of what follows in this paper reduces to the case of polynomial splines.

The data sets we consider in this paper are of two basic types—deterministic and random. For the trajectory planning problem, we usually consider the data to be given in a deterministic form; that is, the coordinates of the locations and times are given exactly. We denote by

$$DD = \{(\alpha_i, t_i) : t_1 < t_2 < \cdots < t_N\}, \tag{4}$$

where the DD stands for deterministic data and we denote by

$$SD = \{(f(t_i) + \epsilon_i, t_i) : t_1 < t_2 < \cdots < t_N\}, \tag{5}$$

where the ϵ_i are observed values of a random variable which in generally we assume is symmetric with mean 0. The term SD is to be interpreted as stochastic data. The data set SD is the set usually encountered in statistics.

Solving the differential equation, we have

$$y(t) = ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)\,ds. \tag{6}$$

It is convenient to set $x_0 = 0$ since the initial data can be absorbed into the data. We now define a one parameter family of functions which are basic to the rest of the paper. Let

$$g_t(s) = \begin{cases} ce^{A(t-s)}b, & t > s, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where we consider s to be the independent variable and t a parameter. We now define a linear functional in terms of $g_t(s)$ as

$$L_t(u) = \int_0^\top g_t(s)u(s)\,ds. \tag{8}$$

With this notation, we have what will be a fundamental relationship for this paper

$$y(t) = L_t(u). \tag{9}$$

We will use the following functional in Section 6. Define

$$D^k L(u) = \int_0^\top \frac{d^k g_t}{dt^k}(s)u(s)\,ds, \tag{10}$$

and note that the derivative is well defined provided that $k < n-2$. With this notation, we have

$$\frac{d^k}{dt^k}y(t) = D^k L_t(u). \tag{11}$$

3. INTERPOLATING SPLINES

In this section, we consider the problem of constructing a control law $u(t)$ that drives the output function $y(t)$ through a set of data points at prescribed times. We will construct u so that the resulting output curve is piecewise smooth and generalizes the classical concept of polynomial spline. We consider the data set DD. The interpolating conditions can then be expressed as

$$y(t_i) = L_{t_i}(u), \qquad i = 1, \ldots, N. \tag{12}$$

There are, of course, infinitely many control laws that will satisfy these constraints. The problem is to have a scheme that will select a unique control law and will choose one that has some physical meaning. Linear quadratic optimal control provides a convenient tool for this selection and the main object of this paper is to show that optimal control plays a natural role in this problem and the other problems of this paper. For the purposes of this paper, we choose the very simple cost functional

$$J(u) = \int_0^\top u^2(s)\,ds. \tag{13}$$

It is possible to increase the complexity of the cost functional, and this has been done quite successfully in references [7,8]. We must specify from what set the control is to be chosen. The first problem then becomes as follows.

Problem 1.

$$\min_u J(u),$$

subject to the N constraints

$$y\left(t_i\right) = L_{t_i}(u),$$

with

$$u \in L_2[0, T].$$

The problem is solved using techniques elucidated in [9]. We construct the orthogonal complement to the linear subspace defined by

$$\operatorname{span}\left\{u(s) : L_{t_i}(u) = 0\right\}.$$

It is a rather trivial exercise to determine that this set is the same as the set spanned by the functions $g_{t_i}(s)$. Thus, the optimal control is of the form

$$u(s) = \sum_{i=1}^{N} \alpha_i g_{t_i}(s). \tag{14}$$

Substituting this into the equations defining the affine variety, we have then a set of equations

$$\begin{aligned} y\left(t_1\right) &= \alpha_1 L_{t_1}\left(g_{t_1}\right) + \cdots + \alpha_N L_{t_1}\left(g_{t_N}\right), \\ y\left(t_2\right) &= \alpha_1 L_{t_2}\left(g_{t_1}\right) + \cdots + \alpha_N L_{t_2}\left(g_{t_N}\right), \\ &\vdots \\ y\left(t_N\right) &= \alpha_1 L_{t_N}\left(g_{t_1}\right) + \cdots + \alpha_N L_{t_N}\left(g_{t_N}\right). \end{aligned}$$

We can write this in matrix form as

$$\hat{y} = G\hat{\alpha},$$

and it should be noted that the matrix G is a Grammian, and hence has the potential to be very poorly conditioned. The advantage is that we can immediately see that there is a unique solution since the $g_{t_i}(s)$s are linearly independent. The conditioning can be greatly improved by replacing the functions $g_{t_i}(s)$ with a set of functions that are nonzero only on intervals of the form $[t_i, t_{i+n}]$. This procedure is outlined in [3]. The advantage is that it reduces the matrix G to a banded matrix (tridiagonal in the case that $n = 2$) which somewhat simplifies the solution. It is not clear how the overall stability of the numerical procedure is affected by this transformation.

It should be noted that although the spline techniques are often said to be nonparametric, in fact, they are equivalent to a finite dimensional parametric problem. Also note that in the case that the matrix A is nilpotent, i.e., that the parameters are zero, the above construction is just the ordinary polynomial spline.

There are several ways to construct splines to solve the basic problem. A totally different construction that is much better conditioned is developed in [1]. That construction develops the banded structure directly but has the disadvantage of not carrying over to the more general problems of this paper.

4. INTERPOLATING SPLINES WITH CONSTRAINTS

In this section, we consider a somewhat different problem that has considerable application and can be solved in much the same manner as the classical interpolating spline. The problem we consider is when instead of data points through which the system must be driven, we require that the system be driven through intervals. We state the problem formally.

Problem 2.

$$\min_u J(u),$$

subject to the constraints

$$a_i \leq L_{t_i}(u) \leq b_i, \qquad i = 1, \ldots, N$$

and

$$u \in L_2[0, T].$$

In the survey by Wegman and Wright [10], this type of spline is discussed. In this section, we show that this type of spline can be recovered with standard optimal control techniques similar to what we used in the last section taken together with mathematical programming.

We first note that because of linearity, the set of controls that satisfy the constraints is closed and convex.

Lemma 4.1. *The set of controls that satisfy the constraints of Problem 2 is a closed and convex subset of* $L^2[0, T]$.

Proof. The proof is elementary and is probably found in any number of elementary textbooks. Suppose that for some finite number M we have controls $u_k(s)$ which satisfy the constraints of Problem 2. Then, for each i, we have

$$a_i = \sum_{k=1}^{N} \alpha_k a_i \leq \sum_{k=1}^{N} \alpha_k L_{t_i}(u_k) \leq \sum_{k=1}^{N} \alpha_k b_i = b_i,$$

where

$$\sum_{k=1}^{N} \alpha_k = 1$$

and

$$\alpha_k > 0.$$

Now consider

$$\sum_{k=1}^{N} \alpha_k L_{t_i}(u_k) = L_{t_i}\left(\sum_{k=1}^{N} \alpha_k u_k\right).$$

Thus, the convex sum of controls satisfies the constraints if the individual controls satisfy the constraints. On the other hand, assume that $\{u_i\}$ is a sequence of controls that each satisfy the constraints. Passing the limit through the integral, because of compactness of the interval $[0, T]$, it follows that the limit also satisfies the constraints. The lemma follows.

The existence and uniqueness of the optimal control follows from standard theorems; see for example [9].

Theorem 4.1. *There exists a unique control function* $u(t)$ *that satisfies the constraints and minimizes* $J(u)$.

Let

$$\hat{a} = (a_1, a_2, \ldots, a_m)^\top$$

and

$$\hat{b} = (b_1, b_2, \ldots, b_m)^\top.$$

We form the associated optimal control problem

$$\max_{\lambda, \gamma} \min_u H(u, \lambda, \gamma), \tag{15}$$

subject to the constraint

$$\lambda \geq 0, \qquad \gamma \geq 0,$$

where

$$H(u, \lambda, \gamma) = \frac{1}{2}\int_0^{\top} u^2(t)\, dt + \sum_{i=1}^{N} \lambda_i \left(a_i - L_{t_i}(u)\right) + \sum_{i=1}^{N} \gamma_i \left(L_{t_i}(u) - b_i\right), \tag{16}$$

and

$$\gamma^{\top} = \left(\gamma_1, \gamma_2, \ldots, \gamma_N\right)^{\top},$$

and

$$\lambda^{\top} = \left(\lambda_1, \lambda_2, \ldots, \lambda_N\right)^{\top}.$$

We minimize the function H over u assuming that λ and γ are fixed. The minimum is achieved at the point where the Gateaux derivative of H, with respect to u, is zero. This is found by calculating

$$\lim_{\tau \to 0} \frac{1}{\tau}\left(H(u + \tau v, \lambda, \gamma) - H(u, \lambda, \gamma)\right) = \int_0^{\top} \left(u(t) - \sum_{i=1}^{N} \lambda_i g_{t_i}(t) + \sum_{i=1}^{N} \gamma_i g_{t_i}(t)\right) v(t)\, dt.$$

Setting this equal to zero and solving, we have the condition that the optimal u^* is given by

$$u^*(t) = \sum_{i=1}^{N} \lambda_i g_{t_i}(t) - \sum_{i=1}^{N} \gamma_i g_{t_i}(t) = \lambda^{\top} g - \gamma^{\top} g, \tag{17}$$

where

$$g^{\top} = \left(g_1(t), g_2(t), \ldots, g_N(t)\right)^{\top}.$$

We now eliminate u^* from the Hamiltonian by substituting to obtain

$$\begin{aligned}
H\left(u^*, \lambda, \gamma\right) &= \frac{1}{2}\int_0^{\top} \left(\left(\lambda^{\top} - \gamma^{\top}\right) g\right)^2 \, dt + \sum_{i=1}^{N} \lambda_i \left(a_i - L_{t_i}\left(u^*\right)\right) - \sum_{i=1}^{N} \gamma_i \left(L_{t_i}\left(u^*\right) - b_i\right) \\
&= \frac{1}{2}(\lambda - \gamma)^{\top} G(\lambda - \gamma) + \lambda^{\top} a - \gamma^{\top} b - \sum_{i=1}^{N} \lambda_i L_{t_i}\left(u^*\right) + \sum_{i=1}^{N} \gamma_i L_{t_i}\left(u^*\right) \\
&= \frac{1}{2}(\lambda - \gamma)^{\top} G(\lambda - \gamma) + \lambda^{\top} a - \gamma^{\top} b - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j L_{t_i}\left(g_{t_j}\right) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_j \gamma_i L_{t_i}\left(g_{t_j}\right) \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \gamma_j L_{t_i}\left(g_{t_j}\right) - \sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_j \gamma_i L_{t_i}\left(g_{t_j}\right) \\
&= \frac{1}{2}(\lambda - \gamma)^{\top} G(\lambda - \gamma) + \lambda^{\top} a - \gamma^{\top} b - \lambda^{\top} G \lambda + \lambda^{\top} G \gamma + \lambda G \gamma - \gamma^{\top} G \gamma \\
&= \frac{1}{2}(\lambda - \gamma)^{\top} G(\lambda - \gamma) + \lambda^{\top} a - \gamma^{\top} b - (\lambda - \gamma)^{\top} G(\lambda - \gamma) \\
&= -\frac{1}{2}(\lambda - \gamma)^{\top} G(\lambda - \gamma) + \lambda^{\top} a - \gamma^{\top} b.
\end{aligned}$$

We can write $H(u^*, \lambda, \gamma)$ in a form suitable for use in quadratic programming packages in the following manner:

$$H\left(u^*, \lambda, \gamma\right) = \frac{-1}{2}\begin{pmatrix} \lambda & \gamma \end{pmatrix}^{\top} \begin{pmatrix} G & -G \\ -G & G \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} + \begin{pmatrix} \lambda & \gamma \end{pmatrix}^{\top} \begin{pmatrix} a \\ -b \end{pmatrix}. \tag{18}$$

Thus, to find the optimal u, we need only solve the quadratic programming problem

$$\max_{\lambda, \gamma} H\left(u^*, \lambda, \gamma\right), \tag{19}$$

subject to positivity constraints on λ and γ.

In general, the control laws constructed by this technique are going to drive the control close to the end points of the interval. A better control law can be obtained by penalizing the control for deviating from the center of the interval. The derivation is essentially the same in concept, but it is a bit more complex in calculation. We leave the calculation to Section 6.

5. SMOOTHING SPLINES

In many problems, to insist that the control drives the output through the points of the data set is overly restrictive and in fact leads to control laws that produce wild excursions of the output between the data points. This phenomena was observed by Wahba, and she developed a theory of smoothing splines that at least partial corrected this problem. In this section, we will develop a theory of smoothing splines based on the same optimal control techniques used to produce interpolating splines. We will penalize the control for missing the data point instead of imposing hard constraints. We formulate this as Problem 3.

PROBLEM 3. *Let*

$$J(u) = \sum_{i=1}^{N} w_i \left(L_{t_i}(u) - \alpha_i\right)^2 + \rho \int_0^{\top} u(t)^2 \, dt.$$

The problem is simply

$$\min_u J(u)$$

for

$$u \in L_2[0, T].$$

The constants w_i are assumed to be strictly positive as is ρ. The choice of the parameters w_i and ρ are important. They control the rate of convergence of the optimal control as the number of data points goes to infinity. This is discussed in detail in [4]. The choice of ρ for fixed data sets is an important issue and is discussed at length in the monograph of Wahba [11].

We calculate the Gateaux derivative of J in the form

$$\lim_{\alpha \to 0} \frac{1}{\alpha} \left(J(u + \alpha h) - J(u)\right) \tag{20}$$

$$= \sum_{i=1}^{N} 2 w_i L_{t_i}(h) \left(L_{t_i}(u) + \alpha_i\right) + 2\rho \int^{\top} h(t) u(t) \, dt \tag{21}$$

$$= 2 \int_0^{\top} \left[\sum_{i=1}^{N} w_i g_{t_i}(t) \left(L_{t_i}(u) + \alpha_i\right) + \rho u(t)\right] h(t) \, dt. \tag{22}$$

Now to ensure that u is a minimum, we must have that the Gateaux derivative vanishes but this can only happen if

$$\sum_{i=1}^{N} w_i g_{t_i}(t) \left(L_{t_i}(u) + \alpha_i\right) + \rho u(t) = 0. \tag{23}$$

We now consider the operator

$$T(u) = \sum_{i=1}^{N} w_i g_{t_i}(t) L_{t_i}(u) + \rho u(t). \tag{24}$$

We can rewrite this operator in the following form:

$$T(u) = \int_0^{\top} \left(\sum_{i=1}^{N} w_i g_{t_i}(t) g_{t_i}(s)\right) u(s) \, ds + \rho u(t). \tag{25}$$

Our goal is to show that the operator T is one to one and onto which we now do.

LEMMA 5.1. *The operator T is one to one for all choices of $w_i > 0$ and $\rho > 0$.*

PROOF. Suppose $T(u_0) = 0$. Equation (24) directly gives that

$$\sum_{i=1}^{N} w_i g_{t_i}(t) L_{t_i}(u_0) + \rho u_0(t) = 0,$$

and hence that

$$\sum_{i=1}^{N} w_i g_{t_i}(t) a_i + \rho u_0(t) = 0,$$

where a_i is the constant $L_{t_i}(u_0)$. This implies that any solution u_0 of $T(u_0) = 0$ is in the span of the set $\{g_{t_i}(t) : i = 1, \ldots, N\}$. Now consider a solution of the form

$$u_0(t) = \sum_{i=1}^{N} \tau_i g_{t_i}(t)$$

and evaluate $T(u_0)$ to obtain

$$\sum_{i=1}^{N} w_i g_{t_i}(t) L_{t_i}\left(\sum_{j=1}^{N} \tau_j g_{t_j}(t)\right) + \rho \sum_{i=1}^{N} \tau_i g_{t_i}(t) = 0.$$

Thus, for each i,

$$w_i \sum_{j=1}^{N} L_{t_i}\left(g_{t_j}\right) \tau_j + \rho \tau_i = 0.$$

The coefficient τ is then the solution of a set of linear equations of the form

$$(DG + \rho I)\tau = 0,$$

where D is the diagonal matrix of the weights w_i and G is the Grammian with $g_{ij} = L_{t_i}(g_{t_j})$. Now consider the matrix $DG + \rho I$ and multiply on the left by D^{-1} and consider the scalar

$$x^\top \left(G + \rho D^{-1}\right) x = x^\top G x + \rho x^t D^{-1} x > 0,$$

since both terms are positive. Thus, for positive weights and positive ρ, the only solution is $\tau = 0$.

It remains to show that the operator T is onto.

LEMMA 5.2. *For $\rho > 0$ and $w_i > 0$, the operator T is onto.*

PROOF. Suppose T is not onto. Then there exists a nonzero function f such that

$$\int_0^\top f(t) T(u)(t)\, dt = 0,$$

for all u. We have after some manipulation

$$\begin{aligned}\int_0^\top f(t) T(u)(t)\, dt &= \int_0^\top \left[\int_0^\top \sum_{i=0}^{N} w_i g_{t_i}(t) g_{t_i}(s) f(t)\, dt + \rho f(s)\right] u(s)\, ds \\ &= 0,\end{aligned}$$

and hence that

$$\int_0^\top \sum_{i=0}^{N} w_i g_{t_i}(t) g_{t_i}(s) f(t)\, dt + \rho f(s) = 0.$$

By the previous lemma, the only solution of this equation is $f = 0$, and hence T is onto.

We have proved the following proposition.

Proposition 5.1. *The functional*

$$J(u) = \sum_{i=1}^{N} w_i \left(L_{t_i}(u) + \alpha_i\right)^2 + \rho \int_0^{\top} u^2(t)\, dt$$

has a unique minimum.

We now use (23) to find the optimal solution. As in the proof that T is one to one, we look for a solution of the form

$$u(t) = \sum_{i=1}^{N} \tau_i g_{t_i}(t).$$

Substituting this into (23), we have upon equating coefficients of the $g_{t_i}(t)$, the system of linear equations

$$(DG + \rho I)\tau = D\gamma. \tag{26}$$

As in the proof of the lemma, the coefficient matrix is invertible, and hence the solution exists and is unique.

The resulting curve $y(t)$ is a spline. The major difference is that the nodal points are determined by the optimization instead of being predetermined. Inverting the matrix $DG + \rho I$ is not trivial. Since it is a Grammian, we can expect it to be badly conditioned. However, by using the techniques in [3], the conditioning can be improved. It is still not clear what happens to the overall numerical stability of the problem.

6. SMOOTHING SPLINES WITH CONSTRAINTS

In this section, we consider two different problems. The first problem which we will consider is a rather straight forward extension of Problem 2. The derivation though is significantly more complex. This is stated as Problem 4. The resulting spline is of practical importance. The second class of problems which we will consider in this paper is very important. It is often the case that there is something known about the underlying curve, i.e., in SD we may have some prior knowledge about the function f. For example, if the data represent growth data on a child from age three months to seven years, we can be reasonably assured that the function f is monotone increasing and the resulting spline must also be monotone increasing if the curve is to have any credibility. There are also cases in which the underling curve is convex or perhaps even more information is known so that the curve must have other shape parameters. In this section, we will show that the techniques we have developed for optimal control can be used to formulate and solve a version of these problems, although we do not yet have a suitable construction for monotone splines.

We begin by formulating the following problem.

Problem 4. *Let the cost functional be defined by*

$$J(u) = \frac{1}{2} \int_0^{\top} u^2(t)\, dt + \frac{1}{2} \sum_{i=1}^{N} \left(L_{t_i}(u) - \rho_i\right)^2,$$

where

$$\rho_i = \frac{a_i + b_i}{2}.$$

The problem is then to

$$\min_u J(u),$$

subject to the constraints of Problem 2.

Define the Hamiltonian by

$$\begin{aligned} H(u,\lambda,\gamma) &= \frac{1}{2}\int_0^{\top} u^2(t)\,dt + \frac{1}{2}\sum_{i=1}^{N}\left(L_{t_i}(u)-\rho_i\right)^2 \\ &+ \sum_{i=1}^{N}\lambda_i\left(a_i - L_{t_i}(u)\right) + \sum_{i=1}^{N}\gamma_i\left(L_{t_i}(u)-b_i\right). \end{aligned} \tag{27}$$

As before, we want to minimize H with respect to u and maximize with respect to λ and γ. Calculating the Gateaux derivative of H with respect to u, we find

$$\begin{aligned} &\lim_{\alpha\to 0}\frac{1}{\alpha}\left(H(u+\alpha v,\lambda,\gamma)-H(u,\lambda,\gamma)\right) \\ &= \int_0^{\top}\left[u + \sum_{i=1}^{N} g_{t_i}\left(L_{t_i}(u)-\rho_i\right) - \sum_{i=1}^{N}\lambda_i g_{t_i} + \sum_{i=1}^{N}\gamma_i g_{t_i}\right] v(t)\,dt. \end{aligned} \tag{28}$$

Setting this equal to 0, we find the condition that

$$u + \sum_{i=1}^{N} g_{t_i}\left(L_{t_i}(u)-\rho_i\right) - \sum_{i=1}^{N}\lambda_i g_{t_i} + \sum_{i=1}^{N}\gamma_i g_{t_i} = 0. \tag{29}$$

Thus, we see that we must have the optimal u as a linear combination of the g_{t_i}s,

$$u^*(t) = \sum_{i=1}^{N}\tau_i g_{t_i}(t). \tag{30}$$

This has the effect of reducing the nonparametric problem to a problem of calculation of parameters in a finite dimensional space.

Substituting u^* into H, we have

$$H(\tau,\lambda,\gamma) = \frac{1}{2}\tau^{\top}G\tau + \frac{1}{2}\tau^{\top}G^2\tau - \tau^{\top}G(a+b) + \lambda^{\top}a - \tau^{\top}G\lambda + \tau^{\top}G\gamma - \gamma^{\top}b + k, \tag{31}$$

where k is a constant that does not effect the location of the optimal point. The problem is now the following:

$$\max_{\lambda,\gamma}\min_{\tau} H(\tau,\lambda,\gamma), \tag{32}$$

subject to positivity constraints on λ and γ.

Calculating the derivative of H with respect to τ, we have

$$\frac{\partial H}{\partial \tau} = G\tau^* + G^2\tau^* - G(a+b) - G\lambda + G\gamma = 0, \tag{33}$$

where τ^* is optimal. Solving for τ^*, we have

$$\tau^* = (I+G)^{-1}(a+b+\lambda-\gamma). \tag{34}$$

Now substituting this into H, we have

$$\begin{aligned}
H(\tau^*, \lambda, \gamma) &= \frac{1}{2}(a+b+\lambda-\gamma)^\top G(I+G)^{-2}(a+b+\lambda-\gamma) \\
&\quad + \frac{1}{2}(a+b+\lambda-\gamma)^\top G^2(I+G)^{-2}(a+b+\lambda-\gamma) \\
&\quad - (a+b+\lambda-\gamma)^\top (I+G)^{-1}G(a+b) \\
&\quad - (a+b+\lambda-\gamma)^\top (I+G)^{-1}G\lambda \\
&\quad + (a+b+\lambda-\gamma)^\top (I+G)^{-1}G\gamma + \lambda^\top a - \gamma^\top b + k \\
&= \frac{1}{2}(a+b+\lambda-\gamma)^\top G(I+G)^{-1}(a+b+\lambda-\gamma) \\
&\quad - (a+b+\lambda-\gamma)^\top G(I+G)^{-1}(a+b+\lambda-\gamma) \\
&\quad + \lambda^\top a - \gamma^\top b + k \\
&= \frac{-1}{2}(a+b+\lambda-\gamma)^\top G(I+G)^{-1}(a+b+\lambda-\gamma) \\
&\quad + (\lambda-\gamma)^\top (I+G)^{-1}G(a+b) + k + k_2 + \lambda^\top a - \gamma^\top b \\
&= \frac{-1}{2}\begin{pmatrix}\lambda & \gamma\end{pmatrix}^\top \begin{pmatrix} (I+G)^{-1}G & -(I+G)^{-1}G \\ -(I+G)^{-1}G & (I+G)^{-1}G \end{pmatrix} \begin{pmatrix}\lambda \\ \gamma\end{pmatrix} \\
&\quad + (\lambda, \gamma)^\top \begin{pmatrix} (I+G)^{-1}G(a+b)+a \\ -(I+G)^{-1}G(a+b)-b \end{pmatrix}.
\end{aligned}$$

It remains to solve the quadratic programming problem

$$\begin{aligned}
\max_{\lambda,\gamma} \Bigg\{ &\frac{-1}{2}\begin{pmatrix}\lambda & \gamma\end{pmatrix}^\top \begin{pmatrix} (I+G)^{-1}G & -(I+G)^{-1}G \\ -(I+G)^{-1}G & (I+G)^{-1}G \end{pmatrix} \begin{pmatrix}\lambda \\ \gamma\end{pmatrix} \\
&+ (\lambda, \gamma)^\top \begin{pmatrix} (I+G)^{-1}G(a+b)+a \\ -(I+G)^{-1}G(a+b)-b \end{pmatrix} \Bigg\}.
\end{aligned} \tag{35}$$

We have found in numerical simulation that this problem is easily solved and produces splines which are quite well behaved. Although the quadratic programming problem is more complicated in terms of the matrices, these formulations seem to offer enough improvement over the formulation of Problem 2 to be worthwhile.

The next problem that we consider is the problem of constructing monotone splines. This problem, as we discussed in the beginning of this section, is very important for many practical applications. We will do less than construct monotone splines here, although it is possible to extend the techniques we are using to produce an infinite dimensional quadratic programming problem that produces monotone splines. We do not make that extension in this paper, but restrict ourselves to ensuring that the spline in nondecreasing at each node. This problem has a significant increase in difficulty over the problem we have considered to this point.

Problem 5. *Let*

$$J(u) = \rho \int_0^\top u^2(t)\,dt + \sum_{i=1}^N w_i \left(L_{t_i}(u) - \alpha_i\right)^2,$$

and let a set of constraints be imposed as

$$DL_{t_i}(u) \ge 0, \qquad i = 1, \ldots, N.$$

The problem then becomes simply

$$\min_u J(u),$$

subject to the constraints and with

$$u \in L_2[0, T].$$

We define the Hamiltonian to be

$$H(u,\lambda) = J(u) + \sum_{i=1}^{N} \frac{w_i}{2}\left(L_{t_i}(u) - \alpha_i\right)^2 - \sum_{i=1}^{N} \lambda_i D L_{t_i}(u). \tag{36}$$

As before, the idea is to minimize H over u and to maximize H over all positive λ. The scheme is to construct the control that minimizes H as a function of λ and to use this parameterized control to convert the Hamiltonian to a function of a finite set of parameters. The resulting Hamiltonian will then be minimized with respect to a subset of the parameters and the Hamiltonian will be reduced to a function of the λ alone. Then the problem reduces to a quadratic programming problem which can be solved using standard software.

We will use the notation

$$h_{t_i}(s) = \frac{d}{dt} g_{t_i}(s)$$

for

$$\left.\frac{d}{dt} g_t(s)\right|_{t=t_i},$$

and we can then rewrite the Hamiltonian as

$$H(u,\lambda) = \frac{1}{2}\int_0^{\top}\left(u^2(s) - \sum_{i=1}^{N}\lambda_i \frac{d}{dt} g_{t_i}(s) u(s)\right) ds + \sum_{i=1}^{N} w_i\left(L_{t_i}(u) - \alpha_i\right)^2. \tag{37}$$

We calculate the Gateaux derivative of H with respect to u to obtain

$$D_u H(u,\lambda)(w) = \int_0^{\top}\left(u(s) - \sum_{i=1}^{N}\lambda_i h_{t_i}(s) + \sum_{i=1}^{N} w_i\left(L_{t_i}(u) - \alpha_i\right) g_{t_i}(s)\right) w(s)\, ds, \tag{38}$$

and thus the optimal u must satisfy

$$u(s) - \sum_{i=1}^{N}\lambda_i h_{t_i}(s) + \sum_{i=1}^{N} w_i\left(L_{t_i}(u) - \alpha_i\right) g_{t_i}(s) = 0. \tag{39}$$

Thus, the optimal u can be represented as

$$u(s) = \sum_{i=1}^{N}\lambda_i h_{t_i}(s) + \sum_{i=1}^{N}\tau_i g_{t_i}(s), \tag{40}$$

and the representation is unique provided that for each i, the functions h_{t_i} and g_{t_i} are linearly independent. This reduces to the condition $A^{n-1} \neq 0$ and with the conditions of equation (3) that $A \neq 0$. So we may assume that without loss of generality, the representation is unique.

We now substitute u into H to obtain the Hamiltonian as a function of τ and λ. We first establish some notation which we will need in order to simplify the formulation of the Hamiltonian. Let

$$H = (h_{ij}), \qquad h_{ij} = \int_0^{\top} h_{t_i}(s) h_{t_j}(s)\, ds,$$

$$K = (k_{ij}), \qquad k_{ij} = \int_0^{\top} h_{t_i}(s) g_{t_j}(s)\, ds,$$

$$G = (g_{ij}), \qquad g_{ij} = \int_0^{\top} g_{t_i}(s) g_{t_j}(s)\, ds.$$

We substitute the expression for u into $H(u,\lambda)$ and obtain after considerable simplification

$$\begin{aligned} H(\tau,\lambda) = \frac{1}{2}\left[\lambda^\top H\lambda + \lambda^\top K\tau + \tau^\top K^\top\lambda + \tau^\top G\tau\right] - \lambda^\top H\lambda - \lambda^\top K\tau \\ + \sum_{i=1}^{N}\frac{w_i}{2}\left[\lambda^\top K^\top e_i e_i^\top K\lambda + \tau^\top G e_i e_i^\top G\tau + \alpha_i^2 + 2\lambda^\top K^\top e_i e_i^\top G\tau\right. \\ \left. - 2\alpha_i\lambda^\top K^\top e_i - 2\alpha_i \tau G e_i\right]. \end{aligned} \tag{41}$$

We rewrite this after further simplification as

$$\begin{aligned} H(\tau,\lambda) = -\frac{1}{2}\lambda^\top H\lambda + \frac{1}{2}\tau^\top G\tau + \frac{1}{2}\lambda^\top K^\top DK\lambda + \frac{1}{2}\tau^\top GDG\tau \\ + \frac{1}{2}\alpha^\top D\alpha + \lambda^\top KDG\tau - \lambda^\top K^\top D\alpha - \tau^\top GD\alpha, \end{aligned} \tag{42}$$

where D is the diagonal matrix of weights w_i.

We now calculate the Gateaux derivative of H with respect to τ and obtain

$$D_\tau H(\tau,\lambda)(w) = w^\top\left(G\tau + GDG\tau + GDK^\top\lambda - G\alpha\right). \tag{43}$$

Setting this equal to zero, we have that the optimal τ must satisfy the equation

$$(G + GDG)\tau = G\alpha - GDK\lambda, \tag{44}$$

or equivalently since G is invertible

$$\tau = \left(D^{-1} + G\right)^{-1}(\alpha - K\lambda). \tag{45}$$

It is clear that when we substitute this into H, we have a quadratic function of λ and so the problem is reduced to solving a quadratic programming problem. Some simplification is possible, but the overall form of the matrices involved seem to be quite messy. In general,

$$H(\lambda) = \lambda^\top F_1\lambda + F_2\lambda + F_3\alpha. \tag{46}$$

It is easy enough to generalize this construction to include higher order derivative constraints. The only problem that arises is to ensure that the representation of the optimal control at (40) is unique. In order to generalize the construction to linear combinations of derivatives and even to different linear combinations at different points, this obstruction becomes quite severe. We would anticipate that some of the same problems arise here as do in the case of Birkhoff interpolation. This remains an active area of investigation.

This construction does not guarantee that the spline function is monotone, but only that the function is nondecreasing at each node. In numerical experiments, we have found that by adding points we can create monotone splines using this construction. We have not at this point proved that the addition of a finite number of points suffices to produce monotone splines.

7. DYNAMIC TIME WARPING

Traditionally, statistics has dealt with discrete data sets. However, most statisticians would agree that information is sometimes lost when data is considered to be point or vector data. In longitudinal studies, it is clear that it is the record of an event that is important, not the individual measurements. For example, if one is studying the growth of individuals in an isolated community, it is not the heights at yearly intervals that are of interest but the curve that represents the growth of an individual over a sequence of years. These matters are discussed at length in the seminal book of Ramsay and Silverman [6]. This book makes a very convincing argument for the study of curves as opposed to discrete data sets.

Often when studying curves, it is not clear that the independent variable (which we will refer to as time) is well defined. In the paper by Li and Ramsay [5], several examples are considered which make this point quite well. The first and second author of this paper have seen this problem when trying to construct weight curves for premature babies—the time of conception is seldom known exactly, and different ethnic stocks may have different growth curves. This leads to the problem of "dynamic time warping" or "curve registration" in order to compare curves that have different bases of time. In this section, we will follow the development of Li and Ramsay [5] and the development of Ramsay and Silverman [6]. We will show that their formulation is equivalent to the problem of optimal output tracking in control theory, and then will give a formulation that is somewhat better behaved from an optimization viewpoint.

Consider a set of curves

$$DC = \{f_i(t) : i = 1, \dots, N\}\,,$$

and assume that there is some commonality among the curves; i.e., they are all growth curves. Choose one such curve, say

$$f_0(t).$$

The choice of this curve is discussed in some detail in [6]. Let

$$T = \{x_\alpha(t) : x_\alpha(t) \text{ is "time-like"}\}\,.$$

What we mean by "time-like" is quite vague. We would like for the functions to be at least almost monotone and convexity would be a good property although at this point we do not want to impose too many conditions. We can pose the problem in the following manner:

$$\min_\alpha \|f_0(t) - f_i\left(x_\alpha(t)\right)\|\,.$$

This problem, although elegant in its simplicity, is too general to solve.

In [5], the set of pseudotimes is constructed in a very clever and insightful manner. They impose the conditions that

$$\frac{\ddot{x}}{\dot{x}} = u(t),$$

with $u(t)$ being "small", the idea being that this will make the curvature of the time $x(t)$ small and that the resulting function $x(t)$ would be time-like. They produce very convincing results using this technique. We can reformulate the problem in the context of optimal control. (We emphasize that we are only reformulating their problem.)

Problem 6. *(See [5].)*

$$\min_u \int_0^\top u^2(t) + \left(f_i(t) - f_0\left(x(t)\right)\right)^2\, dt,$$

subject to the constraint

$$\ddot{x} = u(t)\dot{x}.$$

The question of initial data is problem dependent, and we will leave a complete discussion to a later paper. The problem can be solved with or without the initial data being given. See the treatise of Polak [12] for a very complete treatment of these various cases.

Consider a control system with output, of the form

$$\begin{aligned} \dot{x}(t) &= f\left(x(t)\right) + u(t)g(t), \\ y(t) &= h\left(x(t)\right), \end{aligned} \tag{47}$$

and let a curve $z(t)$ be given which we would like to follow as closely as possible with the curve $y(t)$. The classical model following problem is to construct u so that the distance between y and z is minimized. This problem is often considered in the asymptotic sense, but in reality it is the finite time domain that is the most important in almost all applications.

Problem 7.

$$\min_u \int_0^\top u^2(t) + (z(t) - y(t))^2 \, dt,$$

subject to the constraint that

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + u(t)g(t), \\ y(t) &= h(x(t)). \end{aligned}$$

It is clear that Problem 5 is a special case of Problem 6 by taking

$$z = \begin{pmatrix} x \\ \dot{x} \end{pmatrix},$$

and then noting that

$$\dot{z} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} z + u \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} z.$$

The solution to this problem is given by the solution to the corresponding Euler-Lagrange equations which in this case is a nonlinear two point boundary value problem. The nonlinearity comes from the fact that the differential equation has a nonlinearity, i.e., in Problem 5 the equation is bilinear and in Problem 6 the equation is nonlinear affine. However, even if we choose a linear constraint, the Euler-Lagrange equations are nonlinear because the output function is a nonlinear function of the state. These problems are in general only solvable by numerical methods. Here the book of Polak [12] becomes very useful.

8. TRAJECTORY PLANNING

The trajectory planning problem is a fundamental problem in aeronautics, robotics, biomechanics, and many other areas of application in control theory. The problem comes in two distinct versions. The most general version is of interest for autonomous vehicles or autonomous movement in general. There the complete route is not known in advance and must be planned "on-line". This problem is far beyond what we can do with the relatively simple tools we have developed here. The version of the problem we will consider in this section is in contrast quite simple. We are given a sequence of target points and target times, and we are required to be close to the point at some time close to the target time. This is typical of such problems as the path planning in air traffic control and many problems in industrial robotics. The problem of being close in space is nicely solved by Problems 2–4, but the problem of being close in time has been difficult to resolve. The concept of "dynamic time warping" seems to be the tool that can resolve this problem. Neither the problem nor its solution are trivial, and it is unfortunate that there does not appear to be an analytic solution. The following formulation seems to be the best that can be done at the moment.

We define an auxiliary system which we will use as the pseudotime. We have chosen the system to be linear rather than the more complicated nonlinear system of Li and Ramsay. Let

$$F = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$g^\top = (0, 1)$, and $h = (1, 0)$. We then consider the system

$$\begin{aligned} \dot{z} &= Fz + gv, \\ w &= hz, \end{aligned} \tag{48}$$

with 0 initial data. We then have the output w represented as

$$w(t) = t + \int_0^t (t - s)v(s) \, ds,$$

so that if u is small w is indeed "time-like".

We now formulate the path planning problem in the following manner.

Problem 8. *Let*

$$J(u,v)=\frac{1}{2}\sum_{i=1}^{N} w_i\left(L_{w(t_i)}(u)-\alpha_i\right)^2+\frac{1}{2}\int_0^{\top}\left(u^2(t)+\rho v^2(t)\right)\,dt+\frac{1}{2}\sum_{i=1}^{N}\tau_i\left(w\left(t_i\right)-t_i\right)^2,$$
$$\min_v \min_u J(u,v),$$

subject to the constraints

$$\dot{z}=Fz+gv,$$
$$w=hz,$$

and

$$\dot{w}\left(t_i\right)\geq 0,$$

with both

$$u,v\in L_2[0,T].$$

There is no really clean solution to this problem, but we are able to present an algorithm which gives at least a suboptimal solution. We first minimize with respect to u (this is just Problem 3) and we find that the optimal u is of the form

$$u(t)=\sum_{i=1}^{N}\tau_i g_{w(t_i)}(t), \tag{49}$$

where the τ_i are chosen as in the solution to Problem 3. Recall that the vector τ satisfies the matrix equation

$$(DG+\rho I)\tau=D\alpha,$$

and hence the τ_i are functions of the as of yet nonoptimal $w(t_i)$. However, this choice of u is optimal for any choice of the $w(t_i)$. Substituting u into the functional J, we reduce J to a function of v alone. We are then faced with a highly nonlinear functional to be minimized subject to a differential equation constraint. That is, we now have an optimal control problem which a nonlinear cost functional and a very simple linear control system as the constraint. It is possible to write down the Euler-Lagrange equations for this problem, but they are somewhat intimidating. Instead we opt for a suboptimal solution that has a chance of being calculated using good numerical optimization procedures.

We assume that the control v will have the form

$$v(t)=\sum_{i=1}^{N}\gamma_i\left(t_i-t\right)_+, \tag{50}$$

where the function $(t_i-t)_+$ is the standard function of polynomial splines that is zero when $t>t_i$ and is t_i-t otherwise, and the γ_i are to be determined. Calculating $w(t)$ with this choice of v, we see that w is a cubic spline with nodes at the t_i. Substituting w into the cost functional, we have that J is now a function of the N parameters γ_i. We have reduced the problem to the a finite dimensional optimization problem. At this point, we have not taken into account the inequality constraints on the derivatives of the $w(t_i)$, and this is done by introducing the Hamiltonian of Problem 5. Because of the nonquadratic nature of the cost functional, the quadratic programming problem has become a nonlinear programming problem, and as such presents more difficult numerical problems.

In a future paper, we will develop numerical algorithms for the solution of the problem.

9. CONCLUSIONS

In this paper, eight different problems originating from optimal control theory, statistics, and numerical analysis were investigated. We showed that these problems could be addressed within a unified framework, based on the relationship between optimal control and conventional or smoothing splines.

The first five problems concerned the minimization of a quadratic cost functional subject to a set of linear constraints, ranging from exact interpolation or penalized deviations from the nodes, to splines that pass through intervals or are nondecreasing at the nodes. For these problems, we were able to come up with explicit solutions.

In Problems 6 and 7, the curve registration problem was addressed by extending our optimal control formulation to include the concept of "dynamic time warping". The last problem concerned trajectory planning where we, given a set of target points and target times, wanted to be close to the points at times close to the target times. We were able to formulate these last problems within our optimal control framework in order to unite them with the previous problems into one unified theory.

REFERENCES

1. Z. Zhang, J. Tomlinson and C. Martin, Splines and linear control theory, *Acta Appl. Math.* **49** (1), 1–34, (1997).
2. Z. Zhang, J. Tomlinson, P. Enqvist and C. Martin, Linear control theory, splines and approximation, In *Computation and Control III*, (Edited by K. Bowers and J. Lund), pp. 269–288, Birkhäuser, Boston, MA, (1995).
3. M. Egerstedt and C. Martin, Trajectory planning for linear control systems with generalized splines, In *Proceedings of MTNS*, Padova, Italy, (1998).
4. S. Sun, M. Egerstedt and C. Martin, Control theoretic smoothing splines, *IEEEE Tran. Aut. Conf.* (to appear).
5. X. Li and J. Ramsay, Curve registration, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** (2), 351–363, (1998).
6. J. Ramsay and B. Silverman, *Functional Data Analysis*, Springer Series in Statistics, New York, (1997).
7. N. Agwu and C. Martin, Optimal control of dynamic systems: Application to spline approximations, *Applied Mathematics and Computation* **97**, 99–138, (1998).
8. N. Agwu, Optimal control of dynamic systems and its application to spline approximation, Ph.D. Dissertation, Texas Tech University, (1996).
9. D. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, (1969).
10. E. Wegman and I. Wright, Splines in statistics, *J. Amer. Statist. Assoc.* **78**, 351–365, (1983).
11. G. Wahba, Spline models for observational data, In *CBMS-NSF Regional Conference Series in Applied Mathematics*, Volume 59, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, (1990).
12. E. Polak, Optimization. Algorithms and Consistent Approximations, In *Applied Mathematical Sciences*, Vol. 124, Springer-Verlag, New York, (1997).