

## HILBERT SPACE METHODS FOR CONTROL THEORETIC SPLINES: A UNIFIED TREATMENT

Y. ZHOU\*, M. EGERSTEDT†, AND C. MARTIN‡

**Abstract.** In this paper we give a basic derivation of smoothing and interpolating splines and through this derivation we show that the basic spline construction can be done through elementary Hilbert space techniques. Smoothing splines are shown to naturally separate into a filtering problem on the raw data and an interpolating spline construction. Both the filtering algorithm and the interpolating spline construction can be effectively implemented. We show that a variety of spline problems can be formulated into this common construction. By this construction we are also able to generalize the construction of smoothing splines to continuous data, a spline like filtering algorithm. Through the control theoretic approach it is natural to add multiple constraints and these techniques are developed in this paper.

**1. Introduction.** Estimation and smoothing for data sets that contain deterministic and random data present difficulties not present in purely random data sets. Yet such data sets are very common in practice and if the nature of the data is not respected conclusions may be drawn that have little relation to reality. In this paper we will present a unified treatment of such problems. We will extend the theory of smoothing splines to cover such situations. Some of the techniques that we will use have been developed in papers by Egerstedt and Martin, [4, 8, 14] and their colleagues. The main technical contribution of this paper will be to show that many of these problems can be cast as minimum norm problems in suitable a Hilbert spaces. This approach unifies a series of problems that have been solved by Egerstedt, Zhou, Sun and Martin, [17, 18, 19]. Furthermore, the approach of this paper gives a unified treatment of smoothing splines as developed by Wahba, [15], and the classical polynomial and exponential interpolating splines. The approach of this paper rests on the Hilbert space methods developed by Luenberger in [7].

The theory of smoothing splines is based on the premise that a datum,  $\alpha$  is the sum of a deterministic part,  $\beta$  and a random part  $\epsilon$ . It is assumed that  $\epsilon$  is the value of a random variable from some probability distribution. Smoothing splines are designed to approximate the deterministic part by minimizing the variance of the random part. Often the random variable comes from measurement error. In the following examples the random error comes either from measurement or from estimation based on incomplete data.

---

\*Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden. E-mail: yishao@math.su.se

†Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. E-mail: magnus@ece.gatech.edu

‡Clyde Martin, Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409-1042, USA. E-mail: clyde.f.martin@ttu.edu

EXAMPLE 1. *A very simple problem is to determine the volume of water contained in a playa lake in West Texas, [12]. These are transient water supplies that because of their formation are almost perfectly circular. If a transect is made across the center of the lake it is possible to obtain a fairly good estimate of the volume. At the boundary of the lake the depth of the water is 0cm. However the depth is measured by a graduate student wading through the lake and measuring the depth at a series of points. These measurement are quite random. The bottom of the lake is silted and so it is not clear where the bottom of the probe rests and the measurement is made by reading the depth of a marked probe. The data set then consists of two deterministic values at the boundary and a series of random numbers representing the depth at a series of predetermined points.*

EXAMPLE 2. *In population studies the census is taken every ten years and whether correct or not the values of the census are considered to be absolute for many purposes. Estimates are made of populations within a given city at irregular intervals between the censuses. Thus, if it is necessary to study the growth or decline of a city over a long period of time deterministic data is available at ten year intervals and estimated data is available at shorter and often irregular time periods. The data set consists of deterministic census data and estimated data with random error. Estimates such as the report by the State of California, [11], are a necessary and critical part of planning for governments.*

EXAMPLE 3. *For most individuals in the United States their home is the principle component of their financial portfolio. The question of the value of the portfolio is of interest in a variety of economic indicators, [9]. When the home is purchased there is a firm monetary value that can be measured and when the home is sold there is a firm value. In between the value is less certain. Almost every individual can give you an estimate of the value but unless a formal appraisal is done there may be a very large error in the estimate. This results in a data set with a few deterministic values, the purchase price, the selling price and formal appraisals and many random values that are estimates by the owner.*

These problems all have in common some data that can be assumed to be exact and some data that is subject to error. The goal of this paper is to find a common frame work to treat all such linear problems.

In this paper we will consider the problem of approximating discrete or continuous data using the dynamics of a linear controlled system. The system may have hard constraints such as boundary values and/or hard constraints at internal values. The data will be assume to noisy with known statistics. A contribution of this paper is to formulate these problems as a general class of minimum norm problems in Hilbert space. Egerstedt, Sun and Martin, [8, 14], have formulated interpolation problems as minimum norm problems but the general problems of smoothing spines have not to this point been so formulated. The advantage is more than conceptual in that

the smoothed data is immediately available as is the smooth functional approximation. Thus we will be able to split the problems into an estimation problem and a problem of finding the interpolating splines. Both of these problems can have fast implementations.

The outline of the paper is as follows. In Section 2 we state and solve the basic problem of smoothing splines using Hilbert space methods to solve an associated minimum norm problem. In Section 3 we state the basic algorithm for solving optimal control problems in Hilbert space as minimal norm problems. This algorithm is the basis for the entire paper. In Section 4 we show that interpolating splines can be constructed using Hilbert space techniques. More importantly in this section we extend the basic theory of interpolating splines to find interpolating splines with optimal initial data. In Hilbert space this is a trivial extension of the theory but not using other methods. In Section 5 we make a major extension of the theory to problem in which there are additional hard constraints. In one sense this amounts to changing the spline generator to general linear boundary value problems but we show that these additional hard constraints can be made to define the "constraint variety" and again handled as minimum norm problems. In this section we solve several classical problems. In Section 6 we consider the related problem of smoothing continuous data. The main conceptual result is that these problems are in reality no different than the problem of discrete data. The only complication is that they tend to involve a lot of integration. However we show that we can solve these filtering problems even when there are additional constraints that the filtered result must satisfy.

**2. Statement of the basic problem.** In this section we state the basic problem of smoothing splines and construct the solution. Here we show that the construction splits into two parts in a very natural way. Ultimately, this will allow the implementation of fast algorithms for smoothing spline constructions. The basic idea of the construction is to define a linear variety, in a Hilbert space, that is defined by the constraints. The data is then defined as a point in the Hilbert space and the optimization reduces to finding the point on the affine variety that is closest (in the sense of the norm in the Hilbert space) to the data point. We know that we can construct this point by finding the orthogonal complement of the linear variety that defines the affine variety and constructing the intersection of the affine variety with the orthogonal complement. In this process we follow Luenberger, [7].

**2.1. The definitions.** Let

$$(2.1) \quad \dot{x} = Ax + bu, \quad y = cx, \quad x(0) = x_0$$

be a controllable and observable linear system with initial data  $x(0) = x_0$ . We think of this system as the curve generator. As will be seen we achieve the smoothest

approximation if we impose the conditions for  $n \geq 2$

$$(2.2) \quad cb = cAb = cA^2b = \dots = cA^{n-2}b = 0$$

where  $n$  is the dimension of the system. We impose this condition to obtain maximal smoothness in the functions to be defined by equation (2.4). The initial data can be chosen as part of the optimization algorithm. However, when we consider two or multiple point value problems we will see that the initial data can be specified. Throughout we will use  $(\ )'$  to denote transpose of a matrix.

Let a data set be given as

$$D = \{(t_i, \alpha_i) : i = 1, \dots, N\}$$

and assume that  $t_i > 0$  and let  $T = t_N$ . We will refer to the points  $t_i$  as the “nodes.” Our goal is to find a control  $u(t)$  that minimizes

$$(2.3) \quad J(u, x_0) = \int_0^T u^2(t)dt + (\hat{y} - \hat{\alpha})'Q(\hat{y} - \hat{\alpha}) + x_0'Rx_0$$

where  $Q$  and  $R$  are positive definite matrices. It is not strictly necessary for these matrices to be positive definite. However, as in the case of the linear regulator, if they are not positive definite then other conditions must be imposed to ensure a unique solution. We will discuss this further in Section 2.2. The vector  $\hat{y}$  has components

$$y_i = y(t_i) = ce^{At_i}x_0 + \int_0^{t_i} ce^{A(t_i-s)}bu(s)ds$$

and the vector  $\hat{\alpha}$  has components  $\alpha_i$ .

It is convenient to define the functions

$$(2.4) \quad \ell_i(s) = \begin{cases} ce^{A(t_i-s)}b, & t_i \geq s, \\ 0, & t_i < s. \end{cases}$$

Note that if the assumption on zeros, equation (2.2), holds then  $\ell_i(s)$  is  $n - 2$  times continuously differentiable at  $t_i$ , i.e.

$$(2.5) \quad \ell_i^{(k)}(t) = \begin{cases} cA^k e^{A(t_i-s)}b, & t_i \geq s, \\ 0, & t_i < s. \end{cases}$$

As long as  $cA^k b = 0$  the  $\ell_i^{(k)}(t)$  is continuous, that is until  $k = n - 2$ . We can now write

$$y_i = ce^{At_i}x_0 + \int_0^T \ell_i(s)u(s)ds = ce^{At_i}x_0 + \langle \ell_i, u \rangle_L,$$

where  $\langle \ell_i, u \rangle_L := \int_0^T \ell_i(s)u(s)ds$ . Now let  $\beta_i := R^{-1}e^{A't_i}c'$ . Then,

$$\begin{aligned} y_i &= ce^{At_i}x_0 + \int_0^T \ell_i(s)u(s)ds \\ &= \langle \beta_i, x_0 \rangle_R + \langle \ell_i, u \rangle_L \end{aligned}$$

where we define the inner products

$$\langle x, w \rangle_R = x' R w \text{ and } \langle g, h \rangle_L = \int_0^T g(t) h(t) dt.$$

Note that if we carefully take the derivatives of  $y(t)$  when  $u = \ell_i(t)$  we have

$$y^{(2n-2)}(t) = \sum_{k=0}^{n-2} c A^{n-2+k} b \ell_i^{(n-2-k)}(t) + \int_0^T c A^{2n-1} e^{A(t-s)} b \ell_i(s) ds.$$

Now this derivative is continuous but the next derivative fails to be so. Thus this  $y$  is  $2n - 2$  times continuously differentiable everywhere and real analytic between the nodes.

## 2.2. The Hilbert space and the affine variety. Let

$$\mathcal{H} = L_2[0, T] \times \mathbb{R}^n \times \mathbb{R}^N$$

with norm

$$\|(u; x; d)\|^2 = \int_0^T u^2(t) dt + d' Q d + x' R x,$$

and corresponding inner product

$$\langle (u; x; d), (v; z; f) \rangle = \int_0^T u(t) v(t) dt + x' R z + d' Q f.$$

Note that elements of  $L_2[0, T]$  are equivalence classes of functions. As is usual we will work with representatives of each equivalence class. A data point in  $\mathcal{H}$  is denoted by  $p$ . We define the linear subspace of constraints,  $V_0$ , in  $\mathcal{H}$  as

$$V_0 = \{(u; x; d) : 0 = -d_i + \langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L\}.$$

We use the notation  $V_0$  for consistency with later notation. Note that  $V_0$  is of infinite dimension since it contains a copy of  $L_2[0, T]$  and is of finite co-dimension since it is the intersection of a finite number of co-dimension 1 subspaces. We will construct the orthogonal complement of  $V_0$  in  $\mathcal{H}$ . Also we note that given any pair  $(u; x)$  there is a corresponding  $d$ .

LEMMA 2.1. *The orthogonal complement of  $V_0$  in  $\mathcal{H}$  is*

$$V_0^\perp = \{(v; w; z) : w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i = 0, \quad v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i = 0\}.$$

*Proof.* By definition

$$V^\perp = \{(v; w; z) : \langle v, u \rangle_L + \langle z, d \rangle_Q + \langle w, x \rangle_R = 0, \forall (u; x; d) \in V\}.$$

Now we have

$$\begin{aligned}\langle z, d \rangle_Q &= \sum_{i=1}^N \langle z, e_i \rangle_Q d_i = \sum_{i=1}^N \langle z, e_i \rangle_Q [\langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L] \\ &= \left\langle \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i, x \right\rangle_R + \left\langle \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i, u \right\rangle_L.\end{aligned}$$

Therefore we have

$$\begin{aligned}0 &= \langle v, u \rangle_L + \langle w, x \rangle_R + \langle z, d \rangle_Q \\ &= \langle v, u \rangle_L + \langle w, x \rangle_R + \left\langle \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i, x \right\rangle_R + \left\langle \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i, u \right\rangle_L \\ &= \left\langle w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i, x \right\rangle_R + \left\langle v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i, u \right\rangle_L.\end{aligned}$$

From the definition of  $V_0$  we have that given a pair  $(u : x_0)$  there exists a  $d$  so that  $(u; x_0, d) \in V_0$ . Thus the above equality is true for  $u \in L_2[0, T]$  and for all  $x \in \mathbb{R}^n$  and as a consequence we must have

$$0 = w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i \quad \text{and} \quad 0 = v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i.$$

Note that the latter equality is in the sense of  $L_2[0, T]$ . The lemma follows.  $\square$

**2.3. The intersection of  $V_0 \cap (V_0^\perp + p)$ .** Before constructing the intersection two things must be verified. The first is that  $V_0$  is nonempty and the second is that  $V_0$  is closed. That  $V_0$  is nonempty is a consequence of the fact that every choice of  $u$  and  $x$  determines a triple in  $V_0$ . We state as a lemma the fact that  $V_0$  is closed.

LEMMA 2.2.  $V_0$  is a closed subspace of the Hilbert space  $\mathcal{H}$ .

*Proof.* Define the function with domain  $L_2[0, T] \times \mathbb{R}^n$  and range  $\mathbb{R}^N$  as

$$F_i((u; x)) = \langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L.$$

Note that  $F_i$  is continuous since it is defined in terms of the inner products and note that  $V_0$  is the graph of  $F$  where  $F$  is the function with components  $F_i$ . It then follows from the closed graph theorem that  $V_0$  is closed in  $\mathcal{H}$ .  $\square$

Since  $V_0$  is closed we have that the intersection of  $V_0$  and  $V_0^\perp + p$  consists of a single point. This point is the solution of the optimal control problem given by equations (2.1) and (2.3).

LEMMA 2.3. The intersection of  $V_0 \cap (V_0^\perp + p)$  is

$$V_0 \cap (V_0^\perp + p) = \left\{ \left( \sum_{i=1}^N \gamma_i \ell_i; \sum_{i=1}^N \rho_i \beta_i; (I + GQ + FQ)^{-1}(GQ + FQ)\hat{\alpha} \right) \right\},$$

where

$$\begin{aligned}\gamma_i &= \langle [I - (I + GQ + FQ)^{-1}(GQ + FQ)]\hat{\alpha}, e_i \rangle_Q, \\ \rho_i &= \langle [I - (I + GQ + FQ)^{-1}(GQ + FQ)]\hat{\alpha}, e_i \rangle_Q.\end{aligned}$$

*Proof.* Equating quantities from  $V_0$  and  $V_0^\perp + p$ , (Here  $p = (0; 0; \hat{\alpha})$  is the data point) we have from the definition of  $V_0$  and some rearrangement of terms

$$\begin{aligned}d_i &= \langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L \\ &= - \sum_{j=1}^N \langle z, e_j \rangle_Q \langle \beta_i, \beta_j \rangle_R - \sum_{j=1}^N \langle z, e_j \rangle_Q \langle \ell_i, \ell_j \rangle_L\end{aligned}$$

Now, equating  $d$  with  $\hat{y}$  and  $z$  with  $\hat{y} - \hat{\alpha}$ , we get

$$\begin{aligned}y_i &= - \sum_{j=1}^N \langle \hat{y} - \hat{\alpha}, e_j \rangle_Q \langle \beta_i, \beta_j \rangle_R - \sum_{j=1}^N \langle \hat{y} - \hat{\alpha}, e_j \rangle_Q \langle \ell_i, \ell_j \rangle_L \\ &= - e_i' GQ (\hat{y} - \hat{\alpha}) - e_i' FQ (\hat{y} - \hat{\alpha})\end{aligned}$$

where  $G$  is the Grammian of the  $\beta_i$ 's and  $F$  is the Grammian of the  $\ell_i$ s. Note that since the  $\ell_i$ s are linearly independent,  $F$  is invertible. In more compact form we have

$$\hat{y} = -(GQ + FQ)(\hat{y} - \hat{\alpha}),$$

or finally we have that

$$(2.6) \quad (I + GQ + FQ)\hat{y} = (GQ + FQ)\hat{\alpha}.$$

By rewriting  $I + GQ + FQ = (Q^{-1} + F + G)Q$  and since  $F$  and  $Q$  are positive definite and  $G$  is positive semi-definite the matrix  $(I + GQ + FQ)$  is invertible and we find  $\hat{y}$  as linear function of the data  $\hat{\alpha}$ . This  $\hat{y}$  is the optimal smoothed estimate of the data  $\hat{\alpha}$ . Using  $\hat{y}$  we can then calculate both the optimal control and the optimal initial condition using the defining equations of the orthogonal complement.

To construct the optimal control  $u^*$  we have from Lemma 2.1 and the identifications above

$$\begin{aligned}u^*(t) &= - \sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \ell_i(t) \\ &= - \sum_{i=1}^N \langle (I + GQ + FQ)^{-1}(GQ + FQ)\hat{\alpha} - \hat{\alpha}, e_i \rangle_Q \ell_i(t) \\ &= \sum_{i=1}^N \langle [I - (I + GQ + FQ)^{-1}(GQ + FQ)]\hat{\alpha}, e_i \rangle_Q \ell_i(t)\end{aligned}$$

The construction of the optimal initial condition is carried out in a similar manner. Thus the lemma is proved.  $\square$

We summarize the results of the section with the following theorem.

**THEOREM 2.4.** *Let*

$$\dot{x} = Ax + bu, \quad y = cx$$

be a controllable and observable linear system with initial data  $x(0) = x_0$  and let a data set be given as

$$D = \{(t_i, \alpha_i) : i = 1, \dots, N\}$$

and assume that  $t_i > 0$  and let  $T = t_N$ . Let the cost function be given as

$$J(u, x_0) = \int_0^T u^2(t)dt + (\hat{y} - \hat{\alpha})'Q(\hat{y} - \hat{\alpha}) + x_0'Rx_0$$

where  $Q$  and  $R$  are positive definite matrices. The vector  $\hat{y}$  has components

$$y_i = y(t_i) = ce^{At_i}x_0 + \int_0^{t_i} ce^{A(t_i-s)}bu(s)ds$$

and the vector  $\hat{\alpha}$  has components  $\alpha_i$ . Minimizing  $J$  over  $u \in L_2[0, t]$  and  $x_0 \in \mathbb{R}^n$  we have that the optimal smoothed data is given by

$$(2.7) \quad \hat{y} = (I + GQ + FQ)^{-1}(GQ + FQ)\hat{\alpha},$$

the optimal control is given by

$$(2.8) \quad u = \sum_{i=1}^N \langle [I - (I + GQ + FQ)^{-1}(GQ + FQ)]\hat{\alpha}, e_i \rangle_Q \ell_i,$$

and the optimal initial condition is given by

$$(2.9) \quad x_0 = \sum_{i=1}^N \langle [I - (I + GQ + FQ)^{-1}(GQ + FQ)]\hat{\alpha}, e_i \rangle_Q \beta_i$$

**3. The basic algorithm.** In the above section we have formulated and solved a problem using an algorithm based on the development in Luenberger [7] and that is, in some sense, just an implementation of the projection theorem from the general theory of Hilbert space. This algorithm is extremely powerful. We will see in this paper many problems in optimal control that can be solved by using this algorithm. We will now state the algorithm with some explanation. We begin by describing the inputs and outputs of the algorithm.

*INPUTS:*

- a quadratic cost function in the control, possibly the initial data and in the data;
- a given set of constraints that include a linear control system and deterministic constraints on the solution of the control system and the initial data.

*OUTPUTS:*

- smoothing data  $\hat{y}$ ;
- optimal control  $u$ ;
- optimal initial data  $x_0$ .

**The algorithm**

1. Define the Hilbert space of the control, initial data and data as  $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$  where  $\mathcal{H}_1$  is the Hilbert space of the control,  $\mathcal{H}_2$  is the Hilbert space of the initial data, finite dimensional and  $\mathcal{H}_3$  is the Hilbert space of the data which may be finite or infinite dimensional. The norm is based on the cost functional.
2. Define the affine subvariety  $V_c$  of the constraints. In many of the applications  $c$  is replaced by the parameter from the problem, for example 0 or  $h$ .
3. Define the data as a point  $p$  in  $\mathcal{H}$ .
4. Verify that the variety is well defined in the Hilbert space. Verify that any point evaluations are well defined. For example if  $f \in L_2[0, T]$  then the value  $f(1)$  may not well defined unless it is defined in terms of the inner product.
5. Verify that the variety is closed. This is an essential step but usually in these problem is a consequence of the closed graph theorem.
6. Calculate the orthogonal complement of  $V_0$ . This step may or may not complicated. It is usually straight forward.
7. Calculate the intersection of  $(V_0^\perp + p) \cap V_c$ . This step can be complicated because it reduces to solving a system of Linear equations derived from the definitions of  $V_c$  and  $V_0^\perp$  or  $V_0^\perp + p$ . The equations can be a mix of integral equations and finite dimensional linear equations and may involve several parameters that must be eliminated.
8. The solution to the equations exist and is unique since we know that the intersection will contain a single point. This point is the optimal  $u$ ,  $x_0$  and the optimal output of the linear system.

This paper is an exercise in applying this algorithm to solve a series of important problems in the theory of control theoretic smoothing splines. There are of course other methods of solving these problems. However, no other method seems as straight forward and as intuitive. It is basically just a generalization of the problem from euclidian geometry of finding a point on a given line nearest to a given point in the plane—a problem from high school geometry. The process is described by Figure (1). Note that in some problems  $p = 0$  and these problems reduce to finding a point of minimum norm in an affine subvariety. See [7] for many such examples.

**4. Interpolating splines with initial data.** For interpolating splines we are required to find a control that drives the output  $y$  through the points in the data set

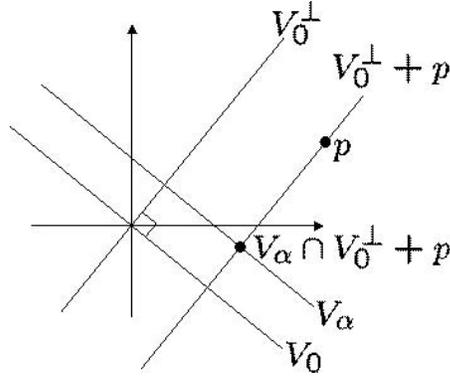


FIG. 1. The general process for finding the point on an affine variety ( $V_\alpha$ ) in a Hilbert space closest to a given data point ( $p$ ).

D. This can be expressed in terms of additional constraints of the form

$$\alpha_i = \langle \beta_i, x_0 \rangle_R + \langle \ell_i, u \rangle_L$$

for  $i = 1, \dots, N$ . The goal is to find a control and an initial condition that minimizes

$$J(u, x_0) = \int_0^T u^2(t) dt + x_0' R x_0$$

subject to the constraints. Just as for smoothing splines we define the Hilbert space to be

$$\mathcal{H} = L_2[0, T] \times \mathbb{R}^n.$$

Now the affine variety of constraints is given by

$$V_{\hat{\alpha}} = \{(u; x_0) : 0 = -\alpha_i + \langle \beta_i, x_0 \rangle_R + \langle \ell_i, u \rangle_L, \quad i = 1, \dots, N\}.$$

Here the goal is to find the point in  $V_{\hat{\alpha}}$  of minimum norm. The procedure is much the same as for smoothing splines. We first must verify that  $V_{\hat{\alpha}}$  is nonempty. This follows from the hypothesized controllability of the linear system. We construct  $V_0^\perp$  and construct the intersection

$$V_0^\perp \cap V_{\hat{\alpha}},$$

which consists of a single point, [7], provided that  $V_0$  is closed.

LEMMA 4.1.  $V_0$  is closed.

*Proof.* Define  $F_i(x, w) = \langle \beta_i, x \rangle_R + \langle \ell_i, w \rangle_L$ . Now  $F_i$  is a continuous linear functional on the Hilbert space  $\mathcal{H}$  and hence the  $\ker(F_i)$  is a closed subset of  $\mathcal{H}$ . Now  $V_0$  is the intersection of a finite number of closed subsets and hence is closed.  $\square$

After some calculation we have

$$V_0^\perp = \{(v; w) : v = \sum_{i=1}^N \tau_i \ell_i, w = \sum_{i=1}^N \tau_i \beta_i\}.$$

After some more calculation we have that the optimal  $u$  is in fact given by

$$u = \sum_{i=1}^N e'_i (F + G)^{-1} \hat{\alpha} \ell_i,$$

and the optimal initial condition is given by

$$x_0 = \sum_{i=1}^N e'_i (F + G)^{-1} \hat{\alpha} \beta_i.$$

the matrices  $F$  and  $G$ , the vectors  $\beta_i$  and the elements of  $L_2[0, T]$ ,  $\ell_i(t)$  are as in the previous section. This is just a slight generalization of the construction given in [14] and hence the details are left out.

For cubic splines the classical construction reduces to solving a system of equations of the form  $\Lambda x = \rho$  where  $\Lambda$  is tridiagonal and of course this a much faster procedure. In [16] the construction of interpolating splines is reduced to solving banded matrices. However, in both cases additional constraints are required to make the problem have a unique solution. With the procedure developed here the additional constraints are unnecessary because of the optimization. Neither the classical cubic splines nor the procedure developed in [16] can easily handle the optimal initial data.

**5. Smoothing and estimation for problems with additional hard constraints.** In a series of papers Willsky and coauthors [1, 2] and Krener [6] considered an estimation problem based on a stochastic boundary value problem. In this section we consider a similar problem in which the smoothing spline is generated by linear system for which there are hard constraints. The constraints may occur as boundary values but they may also occur as fixed internal values or even as linear operator constraints on the solution. We will show that many of these problems can be formulated and solved with the machinery that we have established. The basic idea is that we have a data set in which each data point is of the form  $\alpha_i = f(t_i) + \epsilon_i$  where  $f(t_i)$  is deterministic and the  $\epsilon_i$  is the value of random variable. The goal is to produce a curve (the spline) that better approximates  $f(t)$ . This is, of course, a standard statistical assumption, [15].

**5.1. Two point boundary value problems.** We begin by considering a general boundary value problem. Let the boundary condition be given by

$$(5.1) \quad \Phi x(0) + \Psi x(T) = h,$$

where we let  $h \in \mathbb{R}^k$ . This, of course, includes the classical two point boundary value formulations and other problems of interest. We note that since

$$x(T) = e^{AT}x(0) + \int_0^T e^{A(T-s)}bu(s)ds,$$

the specific dependence on  $x(T)$  can be removed and the boundary constraint simply becomes

$$(5.2) \quad Px(0) + \Psi \int_0^T e^{A(T-s)}bu(s)ds = h,$$

where

$$P := \Phi + \Psi e^{AT}.$$

Note that if there is any solution to (5.1) then by the controllability hypothesis there is a solution to (5.2). We hypothesize that there is at least one solution of (5.1).

We now define the Hilbert space as

$$\mathcal{H} = L_2[0, T] \times \mathbb{R}^n \times \mathbb{R}^N$$

with norm

$$\|(u; x_0; y)\|^2 = \int_0^T u^2(t)dt + x_0'Rx_0 + y'Qy.$$

We define the constraint variety to be

$$V_h = \{(u; x; d) : d_i = \langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L, Px + \Psi \int_0^T e^{A(T-s)}bu(s)ds = h\}.$$

We first prove the following lemma.

LEMMA 5.1.  $V_0$  is a closed subspace of  $\mathcal{H}$ .

*Proof.* The mapping

$$(u; x) \rightarrow \Psi \int_0^T e^{A(T-s)}bu(s)ds + Px,$$

with domain  $L_2[0, T] \times \mathbb{R}^n$  is continuous and hence the subspace

$$W = \{(u, x) \in L_2[0, T] \times \mathbb{R}^n : Px + \Psi \int_0^T e^{A(T-s)}bu(s)ds = 0\}$$

is closed. Now the mapping from  $W$  to  $\mathbb{R}^N$  defined by

$$d_i = \langle \beta_i, x \rangle_R + \langle \ell_i, u \rangle_L$$

is continuous and again we appeal to the closed graph theorem to finish the proof.  $\square$

We now construct  $V_0^\perp$ .

LEMMA 5.2. For some  $\lambda \in \mathbb{R}^k$ ,

$$V_0^\perp = \{(v; w; z) : w = -\sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i + R^{-1}P'\lambda, v = -\sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i + (\Psi e^{A(T-t)})'\lambda\}.$$

*Proof.* The first part of the construction is exactly the same as in subsection 2.2 and from there we have

$$V_0^\perp = \{(v; w; z) : \langle w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i, x \rangle + \langle v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i, u \rangle = 0\}.$$

Now the relationship does not hold for all  $x$  and  $u$  but only for those  $x$  and  $u$  for which equation (5.2) holds. Multiplying by  $\lambda'$ ,  $\lambda \in \mathbb{R}^k$ , we can rewrite equation (5.2) as

$$(5.3) \quad \langle R^{-1}P'\lambda, x \rangle_R + \langle (\Psi e^{A(T-t)})'\lambda, u \rangle_L = 0.$$

From this we conclude that

$$w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i = R^{-1}P'\lambda,$$

and

$$v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i = (\Psi e^{A(T-t)})'\lambda,$$

and the lemma follows.  $\square$

It remains to construct the intersection  $V_h \cap (V_0^\perp + p)$  to find the optimal point. This construction is technically more complicated than the simple smoothing spline but the technique is identical.

The unique point in the intersection is defined as the solution of the following system of four equations in the unknowns  $u$ ,  $x_0$ ,  $y$  and  $\lambda$ , obtained by identifying  $x$  and  $w$  with  $x_0$ ,  $d$  with  $\hat{y}$ , and  $z$  with  $\hat{y} + \hat{\alpha}$ .

$$(5.4) \quad u = -\sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \ell_i + b'e^{A'(T-t)}\Psi'\lambda,$$

$$(5.5) \quad x_0 = -\sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \beta_i + R^{-1}(\Phi + \Psi e^{AT})'\lambda,$$

$$(5.6) \quad h = Px_0 + \int_0^T \Psi e^{A(T-s)}bu(s)ds,$$

$$(5.7) \quad y_i = \langle \beta_i, x_0 \rangle_R + \langle \ell_i, u \rangle_L.$$

We begin by eliminating  $x_0$  and  $u$  from equation (5.7) by substituting equations (5.4) and (5.5). After some manipulation we have

$$y_i = e'_i G(\hat{y} - \hat{\alpha}) - e'_i F(\hat{y} - \hat{\alpha}) + \beta'_i P'\lambda + \int_0^T \ell_i(s)b'e^{A'(T-s)}\Psi' ds \lambda.$$

Since  $\beta_i = R^{-1}e^{A't_i}c'$  let

$$\beta = R^{-1}(e^{A't_1}c', \dots, e^{A't_N}c') =: R^{-1}E$$

to obtain

$$\hat{y} = -G(\hat{y} - \hat{\alpha}) - F(\hat{y} - \hat{\alpha}) + E'R^{-1}P'\lambda + \Lambda\lambda,$$

where

$$\Lambda = \int_0^T l(s)b'e^{A'(T-s)}\Psi' ds.$$

We will now use equation (5.6) to obtain a second equation in  $\lambda$  and  $\hat{y}$ .

$$\begin{aligned} h = P & \left[ - \sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \beta_i + R^{-1}P'\lambda \right] + \\ & + \int_0^T \Psi e^{A(T-s)} b \left[ - \sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \ell_i + b'e^{A'(T-s)}\Psi'\lambda \right] ds. \end{aligned}$$

We make the following observation:

$$\sum_{i=1}^N \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \beta_i = \sum_{i=1}^N \beta_i e_i' Q (\hat{y} - \hat{\alpha}) = R^{-1}EQ(\hat{y} - \hat{\alpha}).$$

We now define

$$M = \sum_{i=1}^N \int_0^T \Psi e^{A(T-s)} b \ell_i(s) e_i' ds Q,$$

and hence

$$\sum_{i=1}^N \int_0^T \Psi e^{A(T-s)} b \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \ell_i(s) ds = M(\hat{y} - \hat{\alpha}).$$

Using these two constructions we then have

$$(5.8) \quad h = P(-R^{-1}EQ(\hat{y} - \hat{\alpha})) + PR^{-1}P'\lambda - M(\hat{y} - \hat{\alpha}) + \Psi\Gamma\Psi'\lambda,$$

where  $\Gamma$  is the controllability Grammian

$$\Gamma = \int_0^T e^{A(T-s)} b b' e^{A'(T-s)} ds.$$

By combining these two expressions linking  $\hat{y}$  and  $\lambda$  gives the following linear equation system

$$(5.9) \quad \begin{pmatrix} I + (G + F)Q & -E'R^{-1}P' - \Lambda \\ PR^{-1}EQ - M & PR^{-1}P' + \Psi\Gamma\Psi' \end{pmatrix} \begin{pmatrix} \hat{y} \\ \lambda \end{pmatrix} = \begin{pmatrix} (G + F)Q\hat{\alpha} \\ h + PR^{-1}EQ + M\hat{\alpha} \end{pmatrix}$$

where  $\Gamma$  is the controllability Grammian

$$\Gamma = \int_0^T e^{A(T-s)} b b' e^{A'(T-s)} ds.$$

Using equation (5.9) we can solve for  $\hat{y}$  and for  $\lambda$ . These values can be used in equations (5.4) and (5.5) to uniquely determine the optimal control and the optimal initial condition. As before we see that the optimal estimate of the data is obtained independently of the control.

REMARK: *The matrix  $E$  is a Grammian like matrix that determines if the initial data can be recovered from sampled observational data, i.e. if  $\dot{x} = Ax$ ,  $x(0) = \alpha$ ,  $y = cx$  and the output is sampled at a set of discrete points  $t_i$  then the output is recoverable from these observations if and only if  $E$  has full rank. Thus  $E$  plays the same role as the observability Grammian. There are no known necessary and sufficient conditions for  $E$  to have full rank. This problem was studied originally by Smith and Martin and was reported in [13]. It is also interesting that the controllability Grammian arises in the formulation of the equation (5.8). The reason for the controllability Grammian to appear is more obvious when one considers the simpler problem of optimally moving between affine subspaces. This problem is studied in [20].*

**5.2. Multiple point constraints.** In this case we have a hard constraint of the form

$$\Phi_1 x(r_1) + \cdots + \Phi_k x(r_k) = h$$

and the data set

$$D = \{(t_i, \alpha_i) : i = 1, \dots, N\}$$

and we assume without loss of generality that

$$\{r_i : i = 1, \dots, k\} \cup \{t_i : i = 1, \dots, N\} = \emptyset.$$

We again make the assumption that there exist at least one set of vectors  $a_i$  such that

$$\Phi_1 a_1 + \cdots + \Phi_k a_k = h.$$

We construct the variety of constraints and note that we can replace  $x(r_i)$  with

$$e^{Ar_i} x(0) + \int_0^{r_i} e^{A(r_i-s)} b u(s) ds.$$

Thus the constraint depends only on  $u$  and  $x_0$ . We use the Hilbert space

$$\mathcal{H} = L_2[0, T] \times \mathbb{R}^n \times \mathbb{R}^N.$$

The constraint variety is

$$V_h = \{(u; x_0; \hat{y}) : y_i = \langle \beta_i, x_0 \rangle + \langle \ell_i, u \rangle_L, \sum_{i=1}^k \Phi_i e^{Ar_i} x_0 + \sum_{i=1}^k \int_0^T \Phi_i \ell_{r_i}(s) u(s) ds = h\}.$$

As before we construct the orthogonal complement to  $V_0$  and then determine the intersection

$$V_h \cap (V_0^\perp + (0; 0; \hat{\alpha})).$$

We leave this construction to the reader.

**5.3. Examples.** In this section we will present some examples of problems that fit this generalized boundary value formulation. We let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad T = 1$$

$$t_1 = 0.2, \quad t_2 = 0.3, \quad t_3 = 0.5, \quad t_4 = 0.7, \quad t_5 = 0.8$$

$$\hat{\alpha} = \begin{pmatrix} 0.8 & 0.2 & 0.5 & 1 & 0.3 \end{pmatrix}$$

$$Q = 10^4 I_5, \quad R = 10^4 I_2, \quad (I_p = p \times p \text{ identity matrix}).$$

EXAMPLE 4 (*Periodic splines*).

We first study the situation when we insist that  $x(0) = x(T)$ . In this case we have that  $\Phi = -\Psi = I_2$ , while  $h = 0$ . The solution is depicted in Figure 2.

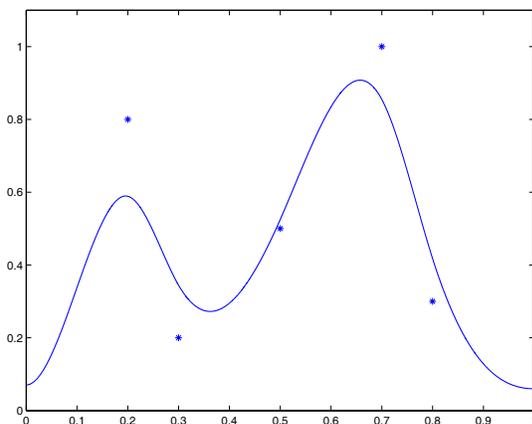


FIG. 2. *Periodic splines*: Here the boundary value is given by  $x(0) = x(T)$ . Depicted are  $y(t)$  (solid) and  $\alpha_i$ ,  $i = 1, \dots, 4$  (stars).

EXAMPLE 5 (*Two point boundary value problems*).

We now let the boundary constraint be encoded by  $\Phi = \begin{pmatrix} 1 & 1 \end{pmatrix}$ ,  $\Psi = -\Phi$ ,  $h = 1$ , which implies that the boundary values are given by the set

$$\{(x_0, x_T) \mid (1, 1)x_0 - (1, 1)x_T = 1\}.$$

The solution is given in Figure 3.

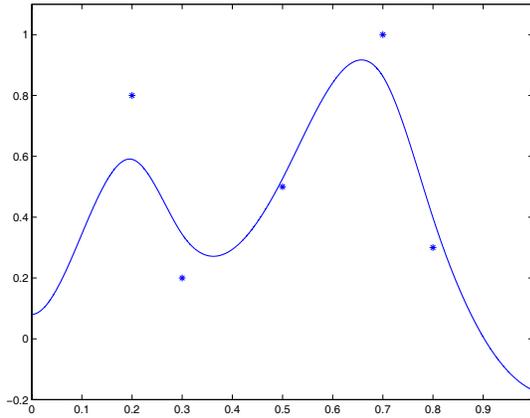


FIG. 3. Boundary value problem:  $(1, 1)x(0) - (1, 1)x(T) = 1$ .

**5.4. Integral constraints.** In many applications ranging from statistics to medical there are constraints of the form

$$\int_0^1 y(t) dt = 1.$$

We will consider a simple problem with  $\dot{x} = Ax + bu$  and a data set  $D = \{(t_i, \alpha_i) : i = 1, \dots, N\}$  and we will further assume that each  $\alpha_i > 0$ . Our constraint variety is given by

$$\begin{aligned} V_1 &= \{(u; x_0; y) : y_i = \langle \beta_i, x_0 \rangle + \langle \ell_i, u \rangle, 1 = \int_0^T y(t) dt, y(t) \\ &= ce^{At}x_0 + \int_0^t e^{A(t-s)}bu(s)ds\}. \end{aligned}$$

As per the algorithm we compute  $V_0^\perp$ . The definition of the orthogonal complement gives

$$V_0^\perp = \{(v; w, z) : \langle v, u \rangle + \langle w, x_0 \rangle_R + \langle z, y \rangle_q = 0, \forall (u; x_0, y) \in V_0\}.$$

Using the defining relation after some calculation using the first relationship in the definition of  $V_0$  we have

$$(5.10) \quad \langle v + \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i, u \rangle + \langle w + \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i, x_0 \rangle = 0.$$

From the second and third defining relations for  $V_0$  we have

$$\begin{aligned} 0 &= \int_0^T (e^{At}x_0 + \int_0^t e^{A(t-s)}bu(s)ds) dt \\ &= \int_0^T e^{At} dt x_0 + \int_0^T \int_s^T e^{A(t-s)} b dt u(s) ds \end{aligned}$$

Multiplying both sides by  $\lambda$  we have

$$(5.11) \quad 0 = \langle \int_0^T R^{-1} e^{A't} dt \lambda, x_0 \rangle + \langle \int_s^T b' e^{A(t-s)} dt \lambda, u \rangle.$$

Now using equations (5.10) and  $V_0^\perp$ .

$$V_0^\perp = \{ (v : w; z) : v = - \sum_{i=1}^N \langle z, e_i \rangle_Q \ell_i + \int_s^T b' e^{A(t-s)} dt \lambda, \\ w = - \sum_{i=1}^N \langle z, e_i \rangle_Q \beta_i + \int_0^T R^{-1} e^{A't} dt \lambda \}.$$

Let  $p = (0; 0; \alpha)$  where  $\alpha$  is the vector of data.

Then in order to construct  $(V_0^\perp + p) \cap V_1$  we must solve the following four equations.

$$(5.12) \quad y_i = \langle \beta_i, x_0 \rangle + \langle \ell_i, u \rangle$$

$$(5.13) \quad 1 = \int_0^T e^{At} dt x_0 + \int_0^T \int_s^T e^{A(t-s)} b dt u(s) ds$$

$$(5.14) \quad u = - \sum_{i=1}^N \langle y + \alpha, e_i \rangle_Q \ell_i + \int_s^T b' e^{A(t-s)} dt \lambda$$

$$(5.15) \quad x_0 = - \sum_{i=1}^N \langle y + \alpha, e_i \rangle_Q \beta_i + \int_0^T R^{-1} e^{A't} dt \lambda$$

The procedure for solving these four equations is exactly the same as before and we leave the details to the reader. Use equations (5.14) and (5.15) to eliminate  $x_0$  and  $u$  from equations (5.12) and (5.13). This results in a pair of equations for  $\lambda$  and the optimal  $y$ . Solve this system and substitute these values to obtain the optimal  $u$  and  $x_0$ . After some calculation the problem of finding the optimal  $y$  and  $\lambda$  reduces to solving a matrix equation. The entries in the matrix must be calculated separately and involve some integration that can be done using standard quadrature algorithms. One  $y$  and  $\lambda$  are found then  $u$  and  $x_0$  are found by substituting into equations (5.14) and (5.15).

**6. Continuous data.** In the previous sections we developed machinery for constructing smoothing splines with discrete data, both deterministic and random. However there are many problems for which the data is continuous. EKG, ECG and EMG are primary examples for which there is a continuous data stream and often there are aspects of the data that are hidden by complexity of the stream. Smoothing splines are an excellent tool to recover such data. In [5] smoothing splines based on B-splines were used to recover long term trends in the stock market. In [17] it was shown that discrete smoothing splines are well approximated by continuous filters which amounts to using continuous data. In this section we will develop a Hilbert space approach for smoothing continuous data with and without additional deterministic discrete data.

**6.1. The linear quadratic regulator problem.** To establish the technique we will solve the linear quadratic regulator problem and then will use the construction for the smoothing problems. This construction is basically the same as is given in Doolin and Martin [3]. We are given a cost function

$$J(u) = \int_0^T x'(t)Qx(t) + u^2(t)dt$$

and a controllable linear system

$$\dot{x} = Ax + bu, \quad x(0) = x_0$$

and we assume that  $x_0$  is given. We also assume that the matrix  $Q$  is positive definite. We define a Hilbert space

$$\mathcal{H} = L_2[0, T] \times L_2^n[0, T]$$

with norm

$$\|(u; x)\|^2 = \int_0^T x'(t)Qx(t) + u^2(t)dt,$$

where  $L_2^n[0, T] := \underbrace{L_2[0, T] \times \cdots \times L_2[0, T]}_{n \text{ times}}$ . Let the constraint variety be defined as

$$V_{x_0} = \{(u; x) : x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}bu(s)ds\}.$$

Note again that  $V_0$  is closed by the closed graph theorem. It is seen, as for the discrete case, that we minimize the cost function by finding the point of minimum norm in  $V_{x_0}$ . Thus we construct the orthogonal complement of  $V_0$ . We have

$$V_0^\perp = \{(v; w) : \int_0^T x'(t)Qw(t) + u(t)v(t)dt = 0, \forall (u; x) \in V_0\}.$$

Using this definition we have

$$\begin{aligned} 0 &= \int_0^T x'(t)Qw(t) + u(t)v(t)dt \\ &= \int_0^T w'(t)Q\left(\int_0^t e^{A(t-s)}bu(s)ds\right) + u(t)v(t)dt \\ &= \int_0^T \int_s^T w'(t)Q(e^{A(t-s)}bu(s))dtds + \int_0^T u(s)v(s)ds \\ &= \int_0^T \left(\int_s^T w'(t)Qe^{A(t-s)}bdt + v(s)\right)u(s)ds \end{aligned}$$

Thus we have

$$V_0^\perp = \{(v; w) : v(s) = -\int_s^T b'e^{A'(t-s)}Qw(t)dt\}.$$

In order to find the intersection  $V_{x_0} \cap V_0^\perp$  we must solve the following system of two integral equations.

$$(6.1) \quad x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}bu(s)ds$$

$$(6.2) \quad u(s) = - \int_s^T b'e^{A'(t-s)}Qx(t)dt$$

To solve this system we let  $u = -b'\lambda$  and from equation (6.2) we have

$$\lambda = \int_t^T e^{A'(r-t)}Qx(r)dr$$

and from equation (6.1) we have

$$x(t) = e^{At}x_0 - \int_0^t e^{A(t-s)}bb'\lambda ds.$$

Differentiating these two equations we have the standard Hamiltonian formulation of the optimal control problem.

$$(6.3) \quad \frac{d}{dt} \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix} = \begin{pmatrix} A & -bb' \\ -Q & -A' \end{pmatrix} \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix}, \quad \begin{pmatrix} x(0) \\ \lambda(T) \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}.$$

The solution of the equation then is done by introducing the Riccati transform. See for example [3] or any elementary control text.

**6.2. Continuous data.** As before we use a linear system

$$\dot{x} = Ax + bu, \quad y = cx, \quad x(0) = x_0$$

as the spline generator. Without loss of generality we assume that  $b'b = 1$ . We assume we are given as our data a function

$$f(t) = g(t) + \epsilon(t),$$

where  $g(t)$  is true underlying function and  $\epsilon(t)$  is a represents random error. We do not need to know how the function  $\epsilon$  is generated. We further assume that  $f$  and  $g$  and hence  $\epsilon$  are square integrable.

We seek to minimize the following cost function. This problem is solved in great detail in [17].

$$J(u, x_0) = \int_0^T u^2(t)dt + x_0'Rx_0 + \int_0^T (y(t) - f(t))^2 dt.$$

We define a Hilbert space  $\mathcal{H} = L_2[0, T] \times \mathbb{R}^n \times L_2[0, T]$  with inner product

$$\langle (u; x; g), (v; w; h) \rangle = \int_0^T u(t)v(t)dt + x'Rw + \int_0^T g(t)h(t)dt,$$

and norm

$$\|(u; x; g)\|^2 = \int_0^T u^2(t)dt + x'Rx + \int_0^T g^2(t)dt.$$

Define the constraint variety to be

$$V_{x_0} = \{(u(t), x_0, y(t)) : y(t) = ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)ds\}$$

and the data point to be  $p = (0; 0; f(t))$ . Note that  $V_0$  is closed. The optimization problem is seen to be solved by finding the point in  $V_{x_0}$  that minimizes the distance to the data point  $p$ .

In order to find this point we first construct the orthogonal complement of  $V$  in  $\mathcal{H}$ . Now by definition we have

$$V_0^\perp = \{(v; w; z) \in \mathcal{H} : \langle u, v \rangle + x'_0Rw + \langle y, z \rangle = 0\}.$$

Using this definition and the definition of  $y$  we have the following

$$\begin{aligned} 0 &= \langle u, v \rangle + x'_0Rw + \langle y, z \rangle_\beta \\ &= \int_0^T u(s)v(s)ds + x'_0Rw + \int_0^T (e^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s))dsz(t)dt \\ &= \int_0^T u(s)v(s)ds + x'_0Rw + \int_0^T (ce^{At}x_0z(t)dt + \int_0^T \int_s^T ce^{A(t-s)}bz(t)dtu(s)ds) \\ &= \int_0^T u(s)[v(s) + \int_s^T ce^{A(t-s)}bz(t)dt] + x'_0R[w + \int_0^T e^{A't}c'z(t)dt] \end{aligned}$$

From this we conclude that

$$(6.4) \quad V_0^\perp = \{(v; w; z) : v(s) = - \int_s^T ce^{A(t-s)}bz(t)dt, w = - \int_0^T e^{A't}c'z(t)dt\}.$$

To find the point of intersection of

$$(V^\perp + p) \cap V$$

we must solve the following three equations.

$$(6.5) \quad y(t) = ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)ds$$

$$(6.6) \quad u(s) = - \int_s^T ce^{A(t-s)}b(y(t) + f(t))dt$$

$$(6.7) \quad x_0 = - \int_0^T e^{A't}c'(y(t) + f(t))dt$$

Now as in the solution of the regulator problem we let

$$(6.8) \quad u = -b'\lambda$$

and from equation (6.5) and (6.8) we have

$$x(t) = e^{At}x_0 - \int_0^t e^{A(t-s)}bb'\lambda ds$$

and from equation (6.6) we have

$$\lambda = \int_s^T b'be^{A'(t-s)}c'cx(t)dt + \int_s^T b'be^{A'(t-s)}c'f(t)dt,$$

using our assumption that  $b'b = 1$  we have the equation

$$\lambda = \int_s^T e^{A'(t-s)}c'cx(t)dt + \int_s^T e^{A'(t-s)}c'f(t)dt.$$

Differentiating these two equations we have the following system of equations

$$(6.9) \quad \frac{d}{dt} \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix} = \begin{pmatrix} A & -bb' \\ -c'c & -A' \end{pmatrix} \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix} - \begin{pmatrix} 0 \\ c' \end{pmatrix} f(t), \quad \begin{pmatrix} x(0) \\ \lambda(T) \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}$$

Thus we have a forced Hamiltonian that governs the solution. This was noted and solved in [14]. It remains to find the optimal  $x_0$ . This will be done by calculating  $y(t)$  from the solution of the forced Hamiltonian. We make a change of variables defined by

$$\begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} I & 0 \\ P(t) & I \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}$$

Making the change of variables we have

$$(6.10) \quad \frac{d}{dt} \begin{pmatrix} x(t) \\ w(t) \end{pmatrix} = \begin{pmatrix} A + b'bP(t) & -bb' \\ R(t) & -(A + bb'P(t))' \end{pmatrix} \begin{pmatrix} x(t) \\ w(t) \end{pmatrix} - \begin{pmatrix} 0 \\ c' \end{pmatrix} f(t),$$

where

$$(6.11) \quad R(t) = \dot{P} + PA + A'P + Pbb'P - c'c.$$

We now set  $R(t) = 0$  and let  $P(T) = 0$  so that  $w(T) = 0$ . Thus we have the standard Riccati equation for the linear optimal control problem of minimizing

$$J(u) = \int_0^T x'c'cx + ubb'udt.$$

We have by solving this system of differential equations the optimal  $y(t)$  and  $u(t)$ . it remains to find the optimal initial condition. We first solve and store the solution to the Riccati equation  $P(t)$ . We are guaranteed a unique solution since the system was assumed to be observable and controllable. We then integrate the linear system

$$\dot{w} = -(A + bb'P(t))'w - c'f$$

to obtain

$$w(t) = - \int_t^T \phi(s, t) c' f(s) ds$$

and therefore we have

$$x(t) = - \int_0^t \phi(t, v) \int_v^T \phi(r, v) c' f(r) dr dv + \phi(t, 0) x_0$$

and hence

$$(6.12) \quad y(t) = - \int_0^t c\phi(t, v) \int_v^T \phi(r, v) c' f(r) dr dv + c\phi(t, 0) x_0$$

From equation (6.6) we have

$$(6.13) \quad \begin{aligned} u(s) = & \int_s^T \int_0^s \int_v^T c e^{A(t-s)} b c \phi(s, v) \phi(r, v) c' f(r) dr dv dt \\ & - \int_s^T c e^{A(t-s)} b f(t) dt - \int_s^T c e^{A(t-s)} b c \phi(s, 0) x_0 dt. \end{aligned}$$

Thus we see that  $u$  feeds back the data function  $f$  and the optimal  $x_0$  and from equation (6.7) we see that the optimal initial data is likewise a function of the data.

$$x_0 = - \int_0^T e^{A't} c' \left( - \int_0^t c\phi(t, v) \int_v^T \phi(r, v) c' f(r) dr dv + f(t) \right) dt - \int_0^T e^{A't} c' c\phi(t, 0) dt x_0$$

and hence we have

$$\begin{aligned} & \left( \int_0^T e^{A't} c' c\phi(t, 0) dt + I \right) x_0 \\ & = - \int_0^T e^{A't} c' \left( - \int_0^t c\phi(t, v) \int_v^T \phi(r, v) c' f(r) dr dv + f(t) \right) dt. \end{aligned}$$

The matrix multiplier of  $x_0$  is invertible since the solution is guaranteed to exist and to be unique from the Hilbert space formulation.

**6.3. Deterministic constraints.** In this subsection we consider the above problem with additional constraints at specific points,

$$\sum_{i=0}^M \Phi_i x(t_i) = h$$

and we assume that  $t_M = T$ . This includes various forms of the boundary value problem ( $M = 1$ ), various forms of the initial value problem ( $M = 0$ ) and the problems with constraints at internal points. We are given a measured function and the desire is to approximate it by  $y(t)$  subject to the constraints. It is possible to add other constraints such as constraints on integrals of  $y$ . In this case the operators  $\Phi$  are integral operators on  $x(t)$ . However in this paper we will assume that the  $\Phi$  are

matrices. We must verify that there is a solution  $(a_1, \dots, a_M)$  of the constraint equation. If there is a solution then by controllability there is a control  $u$  that drives the system through the points.

We first note that the constraint  $\sum_{i=0}^M \Phi_i x(t_i) = h$  can be reduced to a constraint on  $u$  and  $x_0$ .

$$\begin{aligned}
 h &= \sum_{i=0}^M \Phi_i (e^{At_i} x_0 + \int_0^{t_i} e^{A(t_i-s)} b u(s) ds) \\
 &= \left( \sum_{i=0}^M \Phi_i e^{At_i} \right) x_0 + \int_0^T \left( \sum_{i=0}^M \Phi_i e^{A(t_i-s)} I_{[0, t_i)}(s) b \right) u(s) ds \\
 (6.14) \quad &= \Phi x_0 + \int_0^T \Lambda(t) b u(t) dt
 \end{aligned}$$

where

$$\Phi = \sum_{i=0}^M \Phi_i e^{At_i}$$

and

$$\Lambda(s) = \sum_{i=0}^M \Phi_i e^{A(t_i-s)} I_{[0, t_i)}(s),$$

where  $I_{[0, t_i)}(s)$  equals 1 if  $s \in [0, t_i)$  and 0 otherwise. We see that this problem has the same complexity as the two point boundary value problem.

Thus we have the constraint variety defined as

$$V_h = \{(u; x_0; y) : y(t) = ce^{At} x_0 + \int_0^t ce^{A(t-s)} b u(s) ds, h = \Phi x_0 + \int_0^T \Lambda(t) b u(t) dt\}.$$

Following the previous sections we find

$$\begin{aligned}
 V_0^\perp &= \{(v; w; z) : v(s) = - \int_s^T ce^{A(t-s)} b z(t) dt + b' \Lambda' \lambda, \\
 &w = - \int_0^T e^{A't} c' z(t) dt + R^{-1} \Phi' \lambda\}.
 \end{aligned}$$

To find the intersection  $(V_0^\perp + p) \cap V_h$  we solve the following system of four equations.

$$(6.15) \quad y(t) = ce^{At} x_0 + \int_0^t ce^{A(t-s)} b u(s) ds$$

$$(6.16) \quad h = \Phi x_0 + \int_0^T \Lambda(t) b u(t) dt$$

$$(6.17) \quad u(t) = - \int_t^T ce^{A(r-t)} b (y(r) + f(r)) dr + b' \Lambda'(t) \lambda$$

$$(6.18) \quad x_0 = - \int_0^T e^{A't} c' (y(t) + f(t)) dt + R^{-1} \Phi' \lambda$$

Letting

$$(6.19) \quad u(t) = -b'\gamma(t).$$

We have after some manipulation of equations (6.15) and (6.17)

$$(6.20) \quad x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}bb'\gamma(s)ds,$$

$$(6.21) \quad \gamma(t) = - \int_t^T e^{A'(r-t)}c'(y(r) + f(r))dr + \Lambda'(t).\lambda$$

Differentiating equations (6.20) and (6.21) we have the system of differential equations

$$(6.22) \quad \frac{d}{dt} \begin{pmatrix} x(t) \\ \gamma(t) \end{pmatrix} = \begin{pmatrix} A & -bb' \\ -c'c & -A' \end{pmatrix} \begin{pmatrix} x(t) \\ \gamma(t) \end{pmatrix} - \begin{pmatrix} 0 \\ c' \end{pmatrix} f(t)$$

with boundary conditions

$$x(0) = x_0 \text{ and } \gamma(T) = 0.$$

As in the previous subsection we make the Riccati transform,

$$\begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} I & 0 \\ P(t) & I \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix},$$

to obtain

$$(6.23) \quad \frac{d}{dt} \begin{pmatrix} x(t) \\ w(t) \end{pmatrix} = \begin{pmatrix} A + b'bP(t) & -bb' \\ 0 & -(A + bb'P(t))' \end{pmatrix} \begin{pmatrix} x(t) \\ w(t) \end{pmatrix} - \begin{pmatrix} 0 \\ c' \end{pmatrix} f(t),$$

where

$$(6.24) \quad \dot{P} = -PA - A'P - bb'P + c'c, \quad P(T) = 0.$$

From this point the problems is exactly the same as for the two point boundary value problem and we leave the final construction as an exercise for the reader.

**7. Conclusion.** In this paper we have established a common framework for interpolating splines and smoothing splines via a Hilbert space approach to control theoretic splines. We have demonstrated that control theoretic splines can be used to solve a wide variety of problems. While Willsky and colleagues, [1, 2], and Krener, [6], developed beautiful machinery based on very sophisticated stochastic analysis to solve estimation problems based on stochastic two point boundary value problems we have shown that the same problems have elegant and simple solutions based on control theoretic splines. Smoothing splines have along and illustrious history in statistics thanks to the pioneering work of Grace Wahba, [15], and interpolating splines date back to Shoenberg's seminal paper in 1946, [10] in theory and to the early 1960s in

practice. We have been able to show that there is a common framework. Furthermore we have been able to solve a series of problems that involve constraints on the spline and/or the data.

We recognize that not all of the interesting spline problems fall into this Hilbert space formulation. If there are inequality constraints on the spline or its derivative on entire intervals then we are forced to use other formulations. For example, the approach of this paper is not really suitable for monotone smoothing splines or for splines that are required to be in certain interval at the node points. These problems require more sophisticated algorithms and have been studied by Egerstedt, Martin, Sun and Zhou in several papers, for example in [8, 14, 19].

#### REFERENCES

- [1] M. B. ADAMS, A. S. WILLSKY, AND B. C. LEVY, *Linear estimation of boundary value stochastic processes. I. The role and construction of complementary models*, IEEE Trans. Automat. Control 29:9(1984), pp. 803–811.
- [2] M. B. ADAMS, A. S. WILLSKY, AND B. C. LEVY, *Linear estimation of boundary value stochastic processes. II. 1-D smoothing problems*, IEEE Trans. Automat. Control 29:9(1984), pp. 811–821.
- [3] B. DOOLIN AND C. MARTIN, *Differential Geometry for Control Engineers*, Marcel Decker, Inc. New York, 1990.
- [4] M. EGERSTEDT AND C. MARTIN, *Optimal control and monotone smoothing splines*, New trends in nonlinear dynamics and control, and their applications, 279–294, Lecture Notes in Control and Inform. Sci., 295, Springer, Berlin, 2003.
- [5] H. KANO, H. NAKATA, AND C. MARTIN, *Optimal curve fitting and smoothing using normalized B-splines: A tool for studying complex systems*, to appear, Applied Mathematics and Computation.
- [6] A. J. KRENER, *Boundary value linear systems*, Systems analysis (Conf., Bordeaux, 1978), pp. 149–165, Astrisque, 75-76, Soc. Math. France, Paris, 1980.
- [7] D. G. LUENBERGER, *Optimization by vector space methods*, John Wiley & Sons, Inc., New York-London-Sydney, 1969.
- [8] C. F. MARTIN, S. SUN, AND M. EGERSTEDT, *Optimal control, statistics and path planning*, Computation and Control, VI (Bozeman, MT, 1998). Math. Comput. Modelling 33:1-3(2001), pp. 237–253.
- [9] R. F. MARTIN, *Consumption, durable goods, and transaction costs*, Dissertation, University of Chicago, 2002.
- [10] I. SCHOENBERG, *Contributions to the problem of approximation of equidistant data by analytic functions*, Quart. Appl. Math. 4(1946), pp. 45-49.
- [11] D. SHEYA, M. MARTINDALE, AND L. GAGE, State of California, Department of Finance, Historical City/County Population Estimates, 1991-1998, with 1990 Census Counts. Sacramento, California, May 1998.
- [12] L. M. SMITH, *Playas of the Great Plains*. University of Texas Press, Austin, 2003.
- [13] J. SMITH AND C. MARTIN, *Approximation, interpolation and sampling*, Differential geometry: the interface between pure and applied mathematics (San Antonio, Tex., 1986), 227–252, Contemp. Math., 68, Amer. Math. Soc., Providence, RI, 1987.
- [14] S. SUN, M. B. EGERSTEDT, AND C. F. MARTIN, *Control theoretic smoothing splines*, IEEE Trans. Automat. Control, 45:12(2000), pp. 2271–2279.

- [15] G. WAHBA, *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, 59. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
- [16] Z. ZHANG, J. TOMLINSON, AND C. MARTIN, *Splines and linear control theory*, Acta Applicandae Mathematicae, 49(1997), pp. 1–34.
- [17] Y. ZHOU, M. DAYAWANSA, AND C. MARTIN, *Control theoretic smoothing splines are approximate linear filters*, Commun. Inf. Syst., 4:3(2004), pp. 253–272.
- [18] Y. ZHOU AND C. MARTIN, *A regularized solution to the Birkhoff interpolation problem*, Commun. Inf. Syst., 4(2004), pp. 89–102.
- [19] Y. ZHOU, M. EGERSTED, AND C. MARTIN, *Optimal approximation of functions*, Commun. Inf. Syst., 1:1(2001), pp. 101–112.
- [20] Y. ZHOU, M. EGERSTED, AND C. MARTIN, *Optimal trajectories between affine subspaces*, submitted.

