Probability Distribution Estimation Using Control Theoretic Smoothing Splines

by

Janelle K. Charles, M. Sc., B. Sc.

A Dissertation

In

MATHEMATICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Approved

Dr. Clyde F. Martin, Chair

Dr. Kevin Long

Dr. Byungtae Seo

Dr. Bijoy Ghosh

Fred Hartmeister
Dean of the Graduate School

May, 2009

ACKNOWLEDGMENTS

I would like to express sincerest gratitude to everyone who has assisted me during my doctoral program.

I especially want to thank my advisor Dr. Clyde F. Martin for his guidance with my research. He was always available and willing to assist and provide words of advice. In addition, I would like to thank the other members of my dissertation committee–Dr. Kevin Long, Dr. Bigoy Ghosh and Dr. Byungtae Seo.

My heartfelt gratitude goes to my family and friends. Their concern and support has helped me stay focused throughout these difficult years. Most significantly, I would like to thank my mother, Cecily, for her unwavering love and encouragement. She has been a constant source of strength along my journey in obtaining this degree.

TABLE OF CONTENTS

Texas Tech University, Janelle K. Charles, May 2009

ABSTRACT

In this paper, we examine the relationship between optimal control and statistics. We explore the use of control theoretic smoothing splines in the estimation of continuous probability distribution functions defined on a finite interval $[0, T]$, where the data is summarized by empirical probability distributions. In particular, we consider the estimation of distributions of the form $e^{f(t)}$, where there is no restriction on the sign of $f(t)$. The construction of the optimal smoothed curve, $y(t)$, is based on the minimization of an integral cost function done through the application of the Hilbert Projection Theorem, which guarantees that a unique minimum exists.

Further spline construction is implemented in approximating cumulative distribution functions where the Hilbert space methodology is no longer applicable. This estimation is based on a process of iterative optimization through dynamic programming.

LIST OF FIGURES

Chapter I
INTRODUCTION

The purpose of this paper is to approximate the probability and cumulative distributions when they are given empirically. Our goal is to show that control theoretic smoothing splines have certain advantages over the conventional smoothing splines of statistics and provide an effective tool for this problem. Control theoretic smoothing splines have been developed for the last decade in control theory, mostly in the area of path planing [5]. We will show that the construction of such splines can be adapted to construct probability distributions.

In this paper we examine two spline constructions for probability distribution estimation. The first is to simply use the spline to directly approximate the empirical distribution. Here we must impose the constraint that the spline $y(t)$ is non-negative and that the area is equal to one. The area constraint is simple for this is just a linear constraint, which the theory handles easily. The non-negative constraint is harder and we spend considerable time attacking this problem. The second construction seeks an estimation for distributions of the form $e^{f(t)}$. Here two problems are evident. The first is that the distribution that we will approximate must be strictly positive. The second is that the area must be one. In this case, we have a nonlinear constraint that is very difficult to handle directly; however, we solve this problem in the paper.

Splines were developed as a tool for approximation in numerical analysis and were strictly interpolating. There the errors were assumed to be insignificant or nonexistent. Splines did not become important in statistics until it was realized that

you could penalize deviation from data points while penalizing curvature of the spline itself. In control theory, smoothing splines were developed for path planning when there was no requirement for an exact path. It was later realized that the control theoretic splines and smoothing splines of statistics shared a common framework. The primary difficulty in the construction of control theoretic splines is the numerical linear algebra. The size of the linear equations that must be solved grows proportional to the number of data points and the matrices are Grammians. This raises numerical stability issues. In this paper the creation of the empirical distribution is not of utmost concern and we will assume that it is drawn with no more than ten subintervals. The reason for this being to avoid computational instability.

Interpolating splines present fewer numerical issues than do smoothing splines but in most statistical applications, where noisy data is apparent, interpolation gives very little insight into the underlying distribution function from which the data was obtained. The construction of the smoothing spline $y(t)$ is such that the errors between the spline and the data possess good statistical properties, for instance that the residuals have small variance. Thus, the process of constructing a smoothing spline can be thought of as minimum variance problem. Significant work in the development of smoothing splines for statistics was done in [6], [11],[12],[14],[15] and significant work in control theoretic splines has been done in [2] and the many references there.

We will continue the development of smoothing splines in statistical applications from a control theory framework where our problem becomes one of minimization in an appropriate Hilbert space under specified constraints. The theory of control

theoretic smoothing splines is based on the important fact that given a closed linear subspace, V, of a Hilbert space and a point p not in that subspace there is a point $x \in V$ such that for every $y \in V$, $\|x - p\| \leq \|y - p\|$ and the point x is given by the intersection of V and $(V^{\perp} + p)$–the Hilbert Projection Theorem [8]. The problem is roughly modeled as finding a control $u \in L_2[0, T]$ that minimizes a cost functional of the form

$$J(u) = \lambda \int_0^T u^2(t)dt + \sum_{i=1}^{N} (y(t_i) - \alpha_i)^2$$

subject to the constraint that for each i

$$y(t_i) = \int_0^{t_i} ce^{A(t_i - s)}bu(s)ds.$$

We define the Hilbert space to be

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

with norm

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

We then define a linear subspace

$$V = \left\{ (u; y(t_1), \cdots, y(t_N)) : \forall i \; y(t_i) = \int_0^{t_i} ce^{A(t_i - s)}bu(s)ds \right\},$$

the data point is then given by

$$p = (0; \alpha_1, \cdots, \alpha_N)$$

and the Hilbert projection theorem is applied to find the optimal control u that generates the required spline. Variants of this basic procedure is used throughout the paper in probability distribution estimation.

In this paper, we will also focus on the use of monotone smoothing splines in continuous cumulative distribution function (CDF) estimation when given the empirical CDF defined on a specified interval [0,T]. Therefore, we will examine smoothing spline construction where the optimal curve preserves monotonicity. This property translates to the non-negativity constraint on the first derivative of the spline. In this case, we have a nonlinear constraint which is very difficult to handle directly; however, we show that this infinite dimensional problem can be translated and solved in a finite setting following the dynamic programming algorithm for second order systems as illustrated in [2] and [4]. In addition, we will require that the spline $y(t)$ has end conditions $y(0) = 0$ and $y(T) = 1$.

The outline for this paper is as follows. In chapter 2, we will discuss the process of constructing the control theoretic smoothing spline when no constraints are imposed on the control system. We will show, in chapter 4, that application of this construction can be used to determine distributions of the form $e^{f(t)}$ and consider a biometric application of these methods. In chapter 3, we will extend this process to obtain smoothing splines to estimate the probability distribution when we impose integral and point constraints on the system. In chapter 5, we develop the theory of cubic monotone smoothing splines in approximating cumulative distribution functions.

Chapter II

CONTROL THEORETIC SMOOTHING SPLINES

In this chapter, we describe the components of the control system and the data set which will be considered in this paper. Throughout this paper, we will assume a controllable and observable[1] system of the form

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

with initial data $x(0) = x_0$, where $x \in \mathbf{R}^n$, A, b, and, c are constant matrices of compatible dimension, and u and y are scalar functions. For the smoothest approximation, [2] and [9], we will further assume

$$cb = cAb = cA^2b = \cdots = cA^{n-2}b = 0. \tag{2.2}$$

In general, any system which satisfies (2.2) will have coefficient matrices given by $b' = (0, 0, \cdots, 1)$, $c = (1, 0, \cdots, 0)$, and

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ a_1 & a_2 & a_3 & a_4 & \cdots & a_n \end{pmatrix}.$$

[1] See any elementary control theory textbook.

The case where all of the a_is are zero produces polynomial splines. The solution to (2.1) is

$$y(t) = ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)ds. \tag{2.3}$$

Given the empirical distribution defined on a finite interval $[0, T]$, we choose a nodal value t_i in each interval and the value of the distribution at that point as the datum, thus creating the data set

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

and assume $\alpha_i \geq 0$, $0 \leq t_1 < \cdots < t_N \leq T$, and $N \leq 10$. Our goal is to use control theoretic smoothing spline techniques, through the Hilbert Projection Theorem [8], to determine $u^*(t)$ that minimizes the quadratic cost function

$$J(u; x_0) \;=\; \lambda \int_0^T u^2(t)dt + (\hat{y} - \hat{\alpha})'Q(\hat{y} - \hat{\alpha}) + x_0'Rx_0 \tag{2.4}$$

where Q=diag$\{\omega_i : i = 1, \ldots, N\}$ and R are positive definite matrices, the constant $\omega_i > 0$ reflect how important it is that $y(t_i)$ passes close to α_i and smoothing parameter $\lambda > 0$ controls the trade off between smoothness of the spline curve and goodness-of-fit, that is, closeness of this curve to the data. There has been substantial work in the selection of the optimal smoothing parameter [14] and thus, such estimation will be omitted from this paper. The vector $\hat{y} \in \mathbf{R}^N$ has components

$$\begin{aligned} y_i = y(t_i) &= ce^{At_i}x_0 + \int_0^T ce^{A(t_i-s)}bu(s)ds, \\ &= \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L , \end{aligned}$$

with the basis function

$$l_i(s) = \begin{cases} ce^{A(t_i - s)}b : \ t_i \geq s \\ 0 \qquad\qquad : \ t_i < s \end{cases}$$

$\beta_i = R^{-1}e^{A't_i}c'$, and the vector $\hat{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)'$. We define the inner products

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

Consider the Hilbert space

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

with inner product

$$\langle (v; w; z), (u; x_0; \hat{y}) \rangle = \lambda \int_0^T v(t)u(t)dt + z'Q\hat{y} + w'Rx_0.$$

We define the constraint variety, V in $\mathcal{H}$ as

$$V = \{(u; x_0; \hat{y}) \ : \ y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L, i = 1, \ldots, N\}.$$

We will now construct the orthogonal complement of V in $\mathcal{H}$. By definition of the orthogonal complement we have

$$\begin{aligned} V^{\perp} &= \{(v; w; z) : (v; w; z) \perp V\} \\ &= \{(v; w; z) : \ \forall \ (u; x_0; \hat{y}) \in V \ \langle (v; w; z), (u; x_0; \hat{y}) \rangle = 0\} \\ &= \left\{(v; w; z) : \lambda \langle v, u \rangle_L + \langle w, x_0 \rangle_R + \langle z, \hat{y} \rangle_Q = 0\right\}. \end{aligned}$$

Now,

$$\begin{aligned}\langle z;\hat{y}\rangle_Q &= \sum_{i=1}^{N}\langle z,e_i\rangle_Q\, y_i \\ &= \sum_{i=1}^{N}\langle z,e_i\rangle_Q\left(\langle\beta_i,x_0\rangle_R+\langle l_i,u\rangle_L\right) \\ &= \left\langle\sum_{i=1}^{N}\langle z,e_i\rangle_Q\,\beta_i,x_0\right\rangle_R+\left\langle\sum_{i=1}^{N}\langle z,e_i\rangle_Q\, l_i,u\right\rangle_L.\end{aligned}$$

Hence, substituting we get

$$\begin{aligned}0 &= \lambda\langle v,u\rangle_L+\langle w,x_0\rangle_R+\left\langle\sum_{i=1}^{N}\langle z,e_i\rangle_Q\,\beta_i,x_0\right\rangle_R+\left\langle\sum_{i=1}^{N}\langle z,e_i\rangle_Q\, l_i,u\right\rangle_L \\ &= \left\langle w+\sum_{i=1}^{N}\langle z,e_i\rangle_Q\,\beta_i,x_0\right\rangle_R+\left\langle\lambda v+\sum_{i=1}^{N}\langle z,e_i\rangle_Q\, l_i,u\right\rangle_L,\end{aligned}$$

where $\lambda>0$. Using the definition of V, we have that given a pair $(u;x_0)$ there exists y such that $(u;x_0;y)\in V$. Hence, the previous equality holds for all $u\in L_2[0,T]$ and $x_0\in\mathbf{R}^n$. Therefore,

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

and so the orthogonal complement of V in $\mathcal{H}$ is

$$V^{\perp}=\left\{(v;w;z)\ :\ w+\sum_{i=1}^{N}\langle z,e_i\rangle_Q\,\beta_i=0,\lambda v+\sum_{i=1}^{N}\langle z,e_i\rangle_Q\, l_i=0\right\}.$$

Let the data set be represented by a point in the Hilbert space, that is, $p=(0;0;\hat{\alpha})\in\mathcal{H}$. Following the Hilbert Projection Theorem [8], we will find that

the optimal control $u^*(t)$ is the unique point in the intersection $V \cap (V^\perp + p)$. This first requires that V is nonempty, which follows since every $u \in L_2[0,T]$ and $x_0 \in \mathbf{R}^n$ determines a triple in V, and secondly that V is closed is a direct consequence of the Closed Graph Theorem [8].

Now, $V \cap (V^\perp + p)$ gives

$$\begin{aligned} y_i &= \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L \\ &= \left\langle \beta_i, -\sum_{j=1}^{N} \langle z, e_j \rangle_Q \beta_j \right\rangle_R + \left\langle l_i, -\frac{1}{\lambda} \sum_{j=1}^{N} \langle z, e_j \rangle_Q l_j \right\rangle_L \\ &= -\sum_{j=1}^{N} \langle z, e_j \rangle_Q \langle \beta_i, \beta_j \rangle_R - \frac{1}{\lambda} \sum_{j=1}^{N} \langle z, e_j \rangle_Q \langle l_i, l_j \rangle_L . \end{aligned}$$

Hence replacing z with $\hat{y} - \hat{\alpha}$ we get

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

where F and G are the Grammians of $\beta_i' s$ and $l_i s$ respectively. Thus $V \cap (V^\perp + p)$ is

$$\begin{aligned} \hat{y} &= -FQ(\hat{y} - \hat{\alpha}) - \frac{1}{\lambda} GQ(\hat{y} - \hat{\alpha}) \\ &= -\left(FQ + \frac{1}{\lambda} GQ \right) (\hat{y} - \hat{\alpha}), \end{aligned}$$

or equivalently,

$$\left(I + FQ + \frac{1}{\lambda} GQ \right) \hat{y} = \left(FQ + \frac{1}{\lambda} GQ \right) \hat{\alpha}.$$

Clearly, the matrix $\left(I + FQ + \frac{1}{\lambda} GQ \right) = \left(Q^{-1} + F + \frac{1}{\lambda} G \right) Q$ is invertible. Therefore,

the optimal smoothed estimate of the data is

$$\hat{y} = \left(I + FQ + \frac{1}{\lambda} GQ \right)^{-1} \left(FQ + \frac{1}{\lambda} GQ \right) \hat{\alpha}.$$

The optimal control $u^*(t)$ is thus given by

$$\begin{aligned} u^*(t) &= -\frac{1}{\lambda} \sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \, l_i(t) \\ &= -\frac{1}{\lambda} \sum_{i=1}^{N} \left\langle \left(I + FQ + \frac{1}{\lambda} GQ \right)^{-1} \left(FQ + \frac{1}{\lambda} GQ \right) \hat{\alpha} - \hat{\alpha}, e_i \right\rangle_Q l_i(t) \\ &= -\frac{1}{\lambda} \sum_{i=1}^{N} \left\langle \left\{ \left(I + FQ + \frac{1}{\lambda} GQ \right)^{-1} \left(FQ + \frac{1}{\lambda} GQ \right) - I \right\} \hat{\alpha}, e_i \right\rangle_Q l_i(t) \end{aligned}$$

and proceeding similarly, the optimal initial condition is

$$\begin{aligned} x_0 &= -\sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \, \beta_i \\ &= -\sum_{i=1}^{N} \left\langle \left\{ \left(I + FQ + \frac{1}{\lambda} GQ \right)^{-1} \left(FQ + \frac{1}{\lambda} GQ \right) - I \right\} \hat{\alpha}, e_i \right\rangle_Q \beta_i. \end{aligned}$$

Therefore, the resulting smoothed spline takes the form

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

Using the classical cubic smoothing spline gives the coefficient matrices

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ and } c = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

An alternative to the cubic spline is the torsion spline with coefficient matrices

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ and } c = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

In this paper, we concentrate on the use of the classical cubic splines, second order systems, as they are the most natural for statistical applications. For these splines, we have basis functions

$$l_i(s) = \begin{cases} t_i - s : & t_i \geq s \\ 0 \quad\quad : & t_i < s \end{cases}$$

for $i = 1, \cdots, N$. The Grammian matrix G has components

$$\begin{aligned} G_{ij} = G_{ji} &= \int_0^{min(t_i,t_j)} l_i(s) l_j(s) ds = \int_0^{min(t_i,t_j)} (t_i - s)(t_j - s) ds, \text{ for } i \neq j \\ G_{ii} &= \int_0^T l_i^2(s) ds = \int_0^{t_i} (t_i - s)^2 ds, \text{ for } i = j, \end{aligned}$$

and the Grammian matrix $F = \beta\beta'$ where β is an $N \times n$ matrix with ith row given by $\beta_i = R^{-1} e^{A' t_i} c'$ for $i = 1 \cdots, N$.

Chapter III

SMOOTHING SPLINES WITH INTEGRAL CONSTRAINTS

In this chapter, we will consider the estimation of the probability distribution with a spline $y(t)$, where we impose the integral constraint

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

We will assume again that we are given the empirical distribution of the data and that it is defined on the finite interval $[0, T]$. Proceeding with the Hilbert Projection Theorem [8], we define the constraint variety

$$V_1 = \Bigg\{ (u; x_0; \hat{y}) \quad : \quad y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L, i = 1, \cdots, N, \\ \int_0^T \left[ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)dt \right] dt = 1 \Bigg\},$$

which implies that

$$V_0 = \Bigg\{ (u; x_0; \hat{y}) \quad : \quad y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L, i = 1, \cdots, N, \\ \int_0^T \left[ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)dt \right] dt = 0 \Bigg\}.$$

By the definition of $V_0^{\perp}$ we have

$$0 = \left\langle w + \sum_{i=1}^{N} \langle z, e_i \rangle_Q \beta_i, x_0 \right\rangle_R + \left\langle \lambda v + \sum_{i=1}^{N} \langle z, e_i \rangle_Q l_i, u \right\rangle_L$$

and the integral constraint in V_0 gives

$$\begin{aligned}
0 &= \int_0^T ce^{At}x_0dt + \int_0^T\int_0^t ce^{A(t-s)}bu(s)dsdt \\
&= \int_0^T ce^{At}x_0dt + \int_0^t\int_0^T ce^{A(t-s)}bdtu(s)ds \\
&= \int_0^T ce^{At}x_0dt + \int_0^T\int_s^T b'e^{A'(t-s)}bdtu(s)ds \\
&= \left\langle a\int_0^T R^{-1}e^{A't}c'dt, x_0 \right\rangle_R + \left\langle a\int_s^T b'e^{A'(t-s)}c'dt, u \right\rangle_L
\end{aligned}$$

for some constant a. Thus, the orthogonal complement of V_0 is

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

Consider the point $p = (0;0;\hat{\alpha}) \in \mathcal{H}$ representing the data, then the unique solution to $V_1 \cap (V_0^\perp + p)$ can be found by solving the system of linear equations

$$y_i = \langle \beta_i, x_0\rangle_R + \langle l_i, u\rangle_L \tag{3.1}$$
$$1 = \int_0^T ce^{At}x_0dt + \int_0^T\int_s^T b'e^{A'(t-s)}c'dtu(s)ds \tag{3.2}$$
$$x_0 = -\sum_{i=1}^N \langle \hat{y}-\hat{\alpha}, e_i\rangle_Q \beta_i + a\int_0^T R^{-1}e^{A't}c'dt \tag{3.3}$$
$$u = -\frac{1}{\lambda}\sum_{i=1}^N \langle \hat{y}-\hat{\alpha}, e_i\rangle_Q l_i + \frac{a}{\lambda}\int_s^T b'e^{A'(t-s)}c'dt \tag{3.4}$$

Now,

$$\begin{aligned}\langle \beta_i, x_0 \rangle_R &= \left\langle \beta_i, -\sum_{j=1}^{N} \langle \hat{y}-\hat{\alpha}, e_j \rangle_Q \beta_j + a\int_0^T R^{-1}e^{A't}c'dt \right\rangle_R \\ &= -\sum_{j=1}^{N} \langle \hat{y}-\hat{\alpha}, e_j \rangle_Q \langle \beta_i, \beta_i \rangle_R + a\int_0^T \beta_i' e^{A't}c'dt \\ &= -e_i'FQ(\hat{y}-\hat{\alpha}) + a\int_0^T \beta_i' e^{A't}c'dt\end{aligned}$$

and

$$\begin{aligned}\langle l_i, u \rangle_L &= -\frac{1}{\lambda}\sum_{j=1}^{N} \langle \hat{y}-\hat{\alpha}, e_j \rangle_Q \langle l_i, l_j \rangle_L + \left\langle l_i, \frac{a}{\lambda}\int_s^T b'e^{A'(t-s)}c'dt \right\rangle_L \\ &= -\frac{1}{\lambda}e_i'GQ(\hat{y}-\hat{\alpha}) + \frac{a}{\lambda}\int_0^T l_i(s)\int_s^T b'e^{A'(t-s)}c'dtds.\end{aligned}$$

Hence, substituting equations (3.3) and (3.4) into (3.1) yields the components

$$y_i = -e_i'FQ(\hat{y}-\hat{\alpha}) - \frac{1}{\lambda}e_i'GQ(\hat{y}-\hat{\alpha}) + aM_i,$$

where

$$M_i = \int_0^T \beta_i' e^{A't}c'dt + \frac{1}{\lambda}\int_0^T l_i(s)\int_s^T b'e^{A'(t-s)}c'dtds,$$

or equivalently

$$\hat{y} = -FQ(\hat{y}-\hat{\alpha}) - \frac{1}{\lambda}GQ(\hat{y}-\hat{\alpha}) + aM,$$

which reduces to the optimal smoothed data that satisfies the integral constraint

$$\hat{y} = (I + FG + \frac{1}{\lambda}GQ)^{-1}(FQ\hat{\alpha} + \frac{1}{\lambda}GQ\hat{\alpha} + aM).$$

Substituting equations (3.3) and (3.4) into (3.2) we have

$$\begin{aligned}
1 &= \int_0^T ce^{A\nu}\left(-\sum_{i=1}^{N}\langle\hat{y}-\hat{\alpha},e_i\rangle_Q\,\beta_i + a\int_0^T R^{-1}e^{A't}c'dt\right)d\nu \\
&\quad + \int_0^T\int_s^T b'e^{A'(\nu-s)}c'd\nu\left(-\frac{1}{\lambda}\sum_{i=1}^{N}\langle\hat{y}-\hat{\alpha},e_i\rangle_Q\,l_i + \frac{a}{\lambda}\int_s^T b'e^{A'(t-s)}c'dt\right)ds \\
&= \int_0^T ce^{A\nu}\left(-\sum_{i=1}^{N}\langle\hat{y}-\hat{\alpha},e_i\rangle_Q\,\beta_i + a\int_0^T R^{-1}e^{A't}c'dt\right)d\nu \\
&\quad -\frac{1}{\lambda}\sum_{i=1}^{N}\int_0^T\int_s^T b'e^{A'(\nu-s)}c'd\nu\,\langle\hat{y}-\hat{\alpha},e_i\rangle_Q\,l_i(s)ds \\
&\quad +\frac{a}{\lambda}\int_0^T\left(\int_s^T b'e^{A'(\nu-s)}c'd\nu\right)^2 ds.
\end{aligned}$$

For simplicity, we define

$$\begin{aligned}
M1 &= -\frac{1}{\lambda}\sum_{i=1}^{N}\int_0^T\int_s^T b'e^{A'(\nu-s)}c'd\nu l_i(s)ds e_i'Q \\
M2 &= \frac{1}{\lambda}\int_0^T\left(\int_s^T b'e^{A'(\nu-s)}c'd\nu\right)^2 ds + \int_0^T ce^{A\nu}\left(\int_0^T R^{-1}e^{A't}c'dt\right)d\nu \\
M3 &= -\sum_{i=1}^{N}\int_0^T ce^{A\nu}\beta_i d\nu e_i'Q,
\end{aligned}$$

so that equation (3.2) reduces to $1 = M1(\hat{y}-\hat{\alpha}) + aM2 + M3(\hat{y}-\hat{\alpha})$. Therefore, combining the reduced forms of 3.1 and 3.2, relating $\hat{y}$ and a, we obtain an $(N+1)$ dimension system of linear equations

$$\begin{aligned}
1 + (M1+M3)\hat{\alpha} &= (M1+M3)\hat{y} + aM2 \\
\left(-FQ - \frac{1}{\lambda}GQ\right)\hat{\alpha} &= (-I - FQ - \frac{1}{\lambda}GQ)\hat{y} + aM
\end{aligned}$$

which in equivalent matrix form is

$$\begin{pmatrix} M1+M3 & M2 \\ -I-FQ-\frac{1}{\lambda}GQ & M \end{pmatrix} \begin{pmatrix} \hat{y} \\ a \end{pmatrix} = \begin{pmatrix} 1+(M1+M3)\hat{\alpha} \\ (-FQ-\frac{1}{\lambda}GQ)\hat{\alpha} \end{pmatrix}$$

The matrix on the left is invertible due to the uniqueness and existence of the solution of the minimum norm problem as guaranteed by the Hilbert Projection Theorem [8]. Once this system of equations is solved, the values for $\hat{y}$ and a can be substituted into equations (3.3) and (3.4) to determine the optimal control function u^* and initial data x_0.

In the previous spline construction, there is no guarantee that the spline would be nonnegative on the intervals, see Figure 3.1.

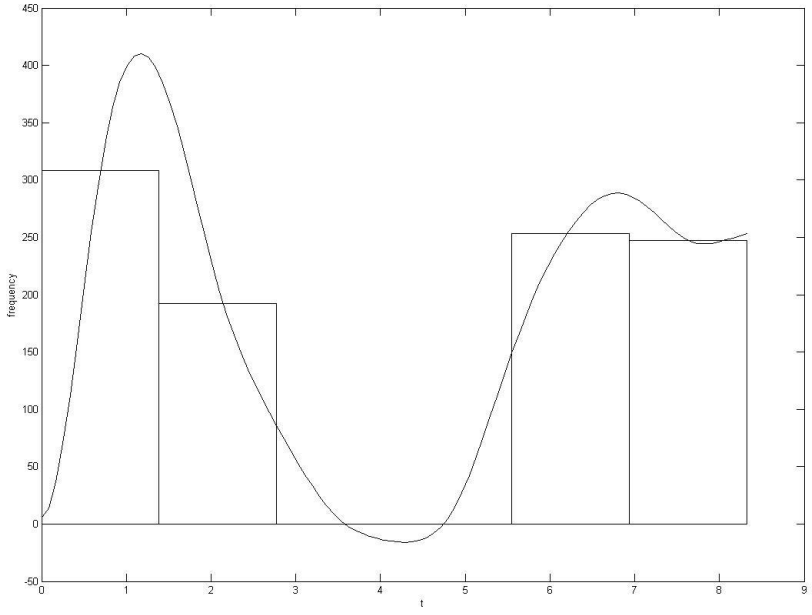

Figure 3.1. Optimal spline satisfying integral constraint

This problem becomes most evident when the empirical distribution is zero on an interval, that is $\alpha_j = 0$ for some t_j. In Figure 3.1, we observe that the optimal smoothing spline satisfies the integral constraint $\int_0^T y(t)dt = 1$ and closely approximates the empirical distribution, but the non-negativity property of probability distribution is not satisfied.

Staying within the minimum norm framework with linear constraints, one solution to this problem involves looking at the spline to determine where it becomes negative and impose the constraint that the derivative at these points is zero. That is, we determine locations t_1' and t_2' such that

$$y(t_1') = y(t_2') = 0 \text{ with } y(t_i') < 0 \text{ for } t_i' \in (t_1', t_2'),$$

then, we impose the point constraints

$$\dot{y}(t_1') = 0 \text{ and } \dot{y}(t_2') = 0,$$

where

$$\dot{y}(t_i) = \int_0^T \dot{l}_i(s)u(s)ds + cAe^{At_i}x_0 \text{ and } \dot{l}_i(s) = \begin{cases} cAe^{A(t_i-s)}b : & t_i \geq s \\ 0 : & t_i < s \end{cases}$$

Thus, the constraint variety is

$$V_1 = \Bigg\{ (u; x_0; \hat{y}) \;\; : \;\; y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L, i = 1, \cdots, N, \\ \int_0^T \left[ce^{At}x_0 + \int_0^t ce^{A(t-s)}bu(s)dt \right] dt = 1, \dot{y}(t_1') = 0, \dot{y}(t_2') = 0 \Bigg\},$$

so that

$$V_0 = \Bigg\{ (u; x_0; \hat{y}) \;\; : \;\; y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L, i = 1, \cdots, N, \\ \int_0^T \left[ce^{At} x_0 + \int_0^t ce^{A(t-s)} bu(s) dt \right] dt = 0, \dot{y}(t_1') = 0, \dot{y}(t_2') = 0 \Bigg\}.$$

Therefore, if we let $p = (0; 0; \hat{\alpha})$ represent the data point in $\mathcal{H}$ then applying the Hilbert Projection Theorem [8] yields a unique solution to the minimum norm problem of finding the optimal control since the spaces V_1 and $V_0^{\perp}$ are closed and nonempty. We observe that the space $V_1 \cap (V_0^{\perp} + p)$ becomes less simple as the number of constraints increases as summarized in the following equations

$$u = -\frac{1}{\lambda} \sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \, l_i(s) + \frac{a_1}{\lambda} \int_s^T b' e^{A'(t-s)} c' dt + \frac{a_2}{\lambda} k_1 + \frac{a_3}{\lambda} k_2 \tag{3.5}$$

$$x_0 = -\sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \, \beta_i + a_1 \int_0^T R^{-1} e^{A't} c' dt + a_2 R^{-1} e^{A' t_1'} A' c' \\ + a_3 R^{-1} e^{A' t_2'} A' c' \tag{3.6}$$

$$y_i = \langle \beta_i, x_0 \rangle_R + \langle l_i, u \rangle_L \tag{3.7}$$

$$1 = \int_0^T ce^{At} x_0 dt + \int_0^T \int_s^T b' e^{A'(t-s)} c' dt u(s) ds \tag{3.8}$$

$$0 = \langle k_1, u \rangle_L + \left\langle R^{-1} e^{A' t_1'} A' c', x_0 \right\rangle_R \tag{3.9}$$

$$0 = \langle k_2, u \rangle_L + \left\langle R^{-1} e^{A' t_2'} A' c', x_0 \right\rangle_R \tag{3.10}$$

where, for $i = 1, 2$, we define

$$k_i(s) = \begin{cases} cAe^{A(t_i' - s)} b : \; t_i' \geq s \\ 0 \qquad\qquad\quad : \; t_i' < s \end{cases}$$

Therefore, substituting (3.5) and (3.6) into (3.7) gives components

$$\begin{aligned}
\langle \beta_i, x_0 \rangle_R &= -\sum_{j=1}^{N} \langle \hat{y} - \hat{\alpha}, e_j \rangle_Q \langle \beta_i, \beta_j \rangle_R + a_1 \int_0^T \beta_i' e^{A't} c' dt \\
&\quad + a_2 \beta_i' e^{A' t_1'} A' c' + a_3 \beta_i' e^{A' t_2'} A' c' \\
&= -e_i' F Q (\hat{y} - \hat{\alpha}) + e_i' a_1 P1 + e_i' a_2 P2 + e_i' a_3 P3 \\
\langle l_i, u \rangle_L &= -\frac{1}{\lambda} \sum_{j=1}^{N} \langle \hat{y} - \hat{\alpha}, e_j \rangle_Q \langle l_i, l_j \rangle_L + \frac{a_1}{\lambda} \int_0^T l_i(s) \int_s^T b' e^{A'(t-s)} c' dt ds \\
&\quad + \frac{a_2}{\lambda} \int_0^T l_i(s) k_1(s) ds + \frac{a_3}{\lambda} \int_0^T l_i(s) k_2(s) ds \\
&= -\frac{1}{\lambda} e_i' G Q (\hat{y} - \hat{\alpha}) + e_i' a_1 F1 + e_i' a_2 F2 + e_i' a_3 F3
\end{aligned}$$

which implies that the vector of smoothed data is given by

$$\hat{y} = -FQ(\hat{y} - \hat{\alpha}) - \frac{1}{\lambda} GQ(\hat{y} - \hat{\alpha}) + a_1(P1 + F1) + a_3(P2 + F2) + a_3(P3 + F3). \quad (3.11)$$

Using the definitions for M1, M2, and M3 given previously, the components of equation (3.8) are

$$\begin{aligned}
\int_0^T c e^{A\nu} x_0 d\nu &= \int_0^T c e^{A\nu} \left(-\sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \beta_i \right) d\nu + a_3 \int_0^T c e^{A\nu} R^{-1} e^{A' t_2'} A' c' d\nu \\
&\quad + \int_0^T c e^{A\nu} \left(a_1 \int_0^T R^{-1} e^{A't} c' dt + a_2 R^{-1} e^{A' t_1'} A' c' \right) d\nu \\
&= M3(\hat{y} - \hat{\alpha}) + a_1 \int_0^T c e^{A\nu} \int_0^T R^{-1} e^{A't} c' dt d\nu \\
&\quad + a_2 \int_0^T c e^{A\nu} R^{-1} e^{A' t_1'} A' c' d\nu + a_3 \int_0^T c e^{A\nu} R^{-1} e^{A' t_2'} A' c' d\nu
\end{aligned}$$

and defining

$$g(s) = \int_s^T b' e^{A'(\nu - s)} c' d\nu$$

yields

$$\begin{aligned} \int_0^T g(s)u(s)ds &= \frac{a_1}{\lambda}\int_0^T g^2(s)ds + \frac{a_2}{\lambda}\int_0^T g(s)k_1(s)ds \\ &\quad + \frac{a_3}{\lambda}\int_0^T g(s)k_2(s)ds + M1(\hat{y} - \hat{\alpha}). \end{aligned}$$

Thus, equation (3.8) becomes

$$\begin{aligned} 1 &= M1(\hat{y} - \hat{\alpha}) + M3(\hat{y} - \hat{\alpha}) + a_2 \int_0^T \left(c e^{A\nu} R^{-1} e^{A' t_1'} A' c' d\nu + \frac{1}{\lambda} g(s) k_1(s) \right) ds \\ &\quad + a_3 \int_0^T \left(c e^{A\nu} R^{-1} e^{A' t_2'} A' c' d\nu + \frac{1}{\lambda} g(s) k_2(s) \right) ds + a_1 M2 \\ 1 &= M1(\hat{y} - \hat{\alpha}) + M3(\hat{y} - \hat{\alpha}) + a_1 M2 + a_2 M4 + a_3 M5 \end{aligned} \tag{3.12}$$

Substituting equations (3.5) and (3.6) into (3.9) gives components

$$\begin{aligned} \langle k_1, u \rangle_L &= -\frac{1}{\lambda} \sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q \langle k_1, l_i \rangle_L + \frac{a_1}{\lambda} \int_0^T k_1(s) g(s) ds \\ &\quad + \frac{a_2}{\lambda} \langle k_1, k_1 \rangle_L + \frac{a_3}{\lambda} \langle k_1, k_2 \rangle_L \\ \left\langle R^{-1} e^{A' t_1'} A' c', x_0 \right\rangle_R &= a_1 c A e^{A t_1'} \int_0^T R^{-1} e^{A' t} c' dt + a_2 c A e^{A t_1'} R^{-1} e^{A' t_1'} A' c' \\ &\quad + a_3 c A e^{A t_1'} R^{-1} e^{A' t_2'} A' c' - \sum_{i=1}^{N} \langle \hat{y} - \hat{\alpha}, e_i \rangle_Q c A e^{A t_1'} \beta_i. \end{aligned}$$

Then, if we let $\gamma_i = cAe^{At_i'}$ we get

$$\begin{aligned} 0 &= \sum_{i=1}^{N}\left(-\frac{1}{\lambda}\left\langle k_1, l_i\right\rangle_L e_i' - \gamma_1\beta_i e_i'\right)Q\hat{y} + \sum_{i=1}^{N}\left(\frac{1}{\lambda}\left\langle k_1, l_i\right\rangle_L e_i' + \gamma_1\beta_i e_i'\right)Q\hat{\alpha} \\ &\quad + a_1\left(\frac{1}{\lambda}\int_0^T k_1(s)g(s)ds + \gamma_1\int_0^T R^{-1}e^{A't}c'dt\right) + a_2\left(\frac{1}{\lambda}\left\langle k_1, k_1\right\rangle_L + \gamma_1 R^{-1}{\gamma_1}'\right) \\ &\quad + a_3\left(\frac{1}{\lambda}\left\langle k_1, k_2\right\rangle_L + \gamma_1 R^{-1}{\gamma_1}'\right) \\ &= A1\hat{y} + A2 + a_1 A3 + a_2 A4 + a_3 A5. \end{aligned} \tag{3.13}$$

Similarly, substituting equations (3.5) and (3.6) into (3.10) gives

$$\begin{aligned} 0 &= \sum_{i=1}^{N}\left(-\frac{1}{\lambda}\left\langle k_2, l_i\right\rangle_L e_i' - \gamma_2\beta_i e_i'\right)Q\hat{y} + \sum_{i=1}^{N}\left(\frac{1}{\lambda}\left\langle k_2, l_i\right\rangle_L e_i' + \gamma_2\beta_i e_i'\right)Q\hat{\alpha} \\ &\quad + a_1\left(\frac{1}{\lambda}\int_0^T k_2(s)g(s)ds + \gamma_2\int_0^T R^{-1}e^{A't}c'dt\right) + a_2\left(\frac{1}{\lambda}\left\langle k_2, k_1\right\rangle_L + \gamma_2 R^{-1}{\gamma_2}'\right) \\ &\quad + a_3\left(\frac{1}{\lambda}\left\langle k_2, k_2\right\rangle_L + \gamma_2 R^{-1}{\gamma_2}'\right) \\ &= B1\hat{y} + B2 + a_1 B3 + a_2 B4 + a_3 B5. \end{aligned} \tag{3.14}$$

Therefore, to find the optimal control and initial data we need to solve the $(N+3)$-dimension system of linear equations represented in equations (3.11) through (3.14) given in equivalent matrix form

$$\Phi\begin{pmatrix} \hat{y} \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} (-FQ - \frac{1}{\lambda}GQ)\hat{\alpha} \\ 1 + (M1 + M3)\hat{\alpha} \\ -A2 \\ -B2 \end{pmatrix}$$

where Φ is an $(N+3)\times(N+3)$ constant matrix defined as

$$\Phi = \begin{pmatrix} -I - FQ - \frac{1}{\lambda}GQ & P1 + F1 & P2 + F2 & P3 + F3 \\ M1 + M3 & M2 & M4 & M5 \\ A1 & A3 & A4 & A5 \\ B1 & B3 & B4 & B5 \end{pmatrix}.$$

The fact that matrix Φ is invertible follows by showing the linear independence of l_is and $\dot{l}_i$s, that is, for constants d_1, d_2, $d_1 l_i(s) + d_2 \dot{l}_i(s) = 0$ implies $d_1 = d_2 = 0$. If we consider the sum $d_1 c e^{A(t_i - s)} b + d_2 cAe^{A(t_i - s)} b = (d_1 c + d_2 cA) e^{A(t_i - s)} b = 0$. Thus, we have that $(d_1 c + d_2 cA) = 0$, which when multiplied by $A^{n-2}b$ gives $d_1 cA^{n-2}b + d_2 cA^{n-1}b = 0$. Now, by assumption 2.2 we have $d_2 cA^{n-1}b = 0$ so that $d_2 = 0$ and it follows similarly that d_1=0. Since Φ is invertible, the unique solution to the minimum norm problem can be obtained by furthermore substituting $\hat{y}, a_1, a_2, a_3$ into equations (3.5) and (3.6).

We note here that this derivation was based on the assumption that the spline obtained imposing only the integral constraint possesses exactly two zeros namely t_1' and t_2'. However, this process becomes much more tedious as the number of zeros increases. In fact, the use of classical cubic splines can produce at most three zeros between any pair of nodal points, thus creating problems of matrix inversion up to a size of $N + 1 + 3(N-1)$, where N is the total number of data points. Moreover, this construction may not produce horizontal tangents (zero point derivatives) at the roots of $y(t)$ if these are locations of inflection points shown in Figure 3.2.

A more global construction, which we consider in chapter 5, involves imposing the constraint that $\dot{y}(t) \geq 0$ on the entire domain. This technique is based on the

formulation of the unique optimal control discussed in [3]. Here, we impose the additional constraints that $y(0) = 0$ and $y(T) = 1$. This construction converts the minimization problem in Hilbert space to a dynamic programming problem. However, we will show that methods developed in the next chapter can be used to obtain the non-negativity constraint via some interval reweighting strategy.

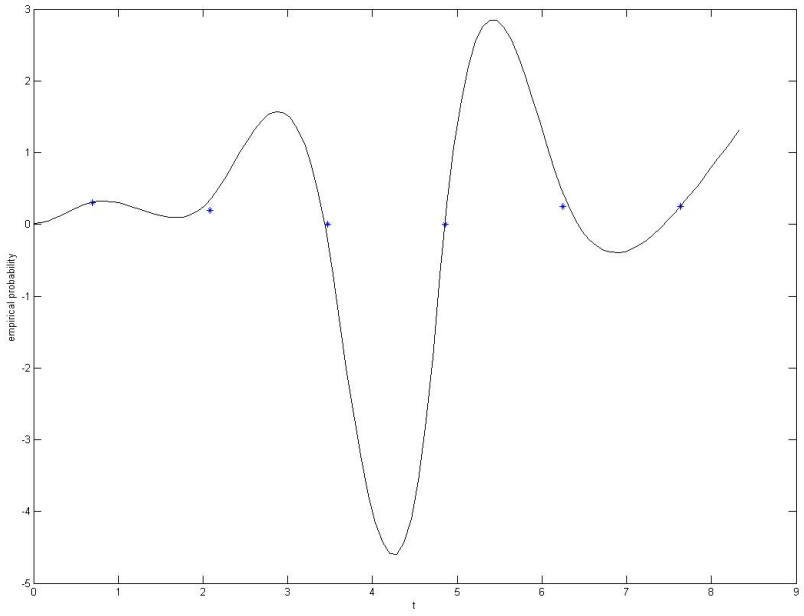

Figure 3.2. Optimal spline with integral and derivative point constraints

Chapter IV

SMOOTHING SPLINES WITH NONLINEAR CONSTRAINTS

In this chapter, we construct smoothing splines when the constraints are known to be nonlinear. It is of particular importance to obtain smoothing splines for estimating distributions belonging to the exponential family, that is, we consider probability distributions of the form $e^{f(t)}$. Here, we encounter two difficulties in the spline construction. The first is that the estimated distribution function must be positive and the second that area must be one, that is, $\int_{-\infty}^{\infty} e^{f(t)}dt = 1$. This is a nonlinear constraint which the Hilbert space construction described in the previous chapters is not suitable for. As such, we concentrate on the spline construction when we restrict the domain of the exponential distribution from the infinite interval to a finite interval. Therefore, we construct a smoothing spline $e^{y(t)}$ with domain $[0, T]$ such that $e^{y(t)} > 0$ and $\int_0^T e^{y(t)}dt = 1$.

We consider distribution estimation based on a simulation of data from a normal distribution[1]. We draw the empirical distribution by shifting these values such that they lie in the interval $[0, T]$, where T is the largest simulated value. We split the interval $[0, T]$ into N subintervals each of which is assumed to have relative frequency, $\tau_i > 0$. This assumption is in keeping with the basic property that exponential distributions are strictly positive. In order to find the approximate distribution function, we take the logarithm of these empirical probabilities. Therefore, we are faced with estimation of the function $f(t)$. There is no restriction on the sign of $f(t)$ and so we can implement smoothing spline techniques developed

[1]For real numbers μ and $\sigma > 0$, the normal distribution function with mean μ and variance σ^2 is given by $\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

in chapter 2 for this estimation. The data for our smoothing spline construction is thus given by

$$D = \{(t_i, \alpha_i = \ln \tau_i) \ : i = 1, \cdots N\},$$

where t_i is the nodal value chosen from interval i. The resulting smoothing spline that approximates this data is shown in Figure 4.1. Since we are able obtain smooth curve $y(t)$ that estimates $f(t)$ then the exponential estimate $e^{y(t)}$ will also be smooth. Moreover, this approximation will be strictly positive and we have thus taken care of the first difficulty in this construction.

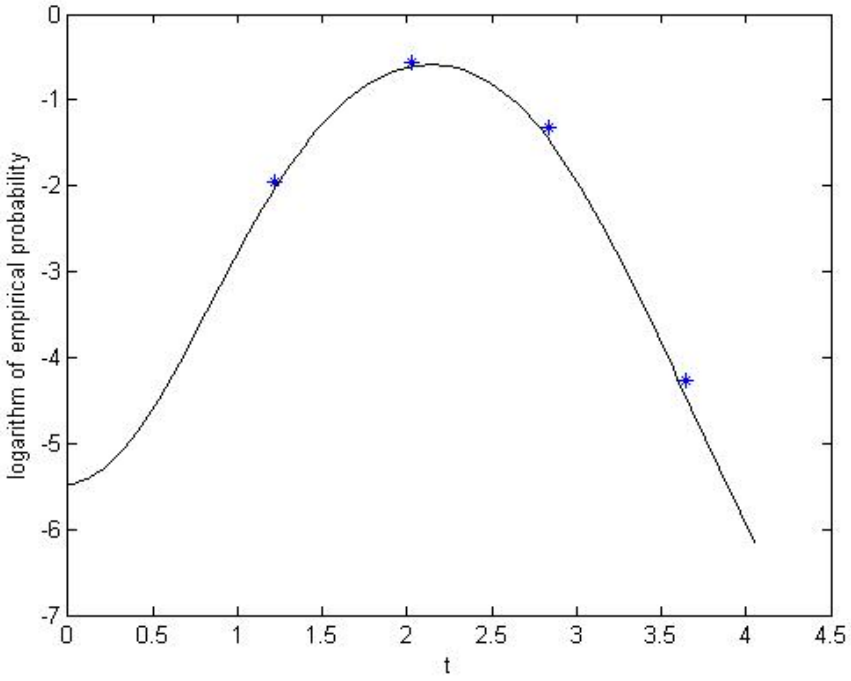

Figure 4.1. Optimal spline for estimating $f(t)$ in the distribution $e^{f(t)}$

Having solved the positivity constraint in this construction, it remains to verify our integral constraint $\int_0^T e^{y(t)}dt = 1$. This area of one is clearly not expected as we

have restricted the infinite domain to a finite interval $[0, T]$. However, if we let

$$\int_0^T e^{y(t)} dt = k,$$

then by renormalizing we get that the smoothing spline which satisfies the integral constraint is given by $\frac{1}{k} e^{y(t)}$ shown in Figure 4.2 and we have

$$\int_0^T \frac{1}{k} e^{y(t)} dt = \int_0^T e^{y(t) - \ln k} dt = 1.$$

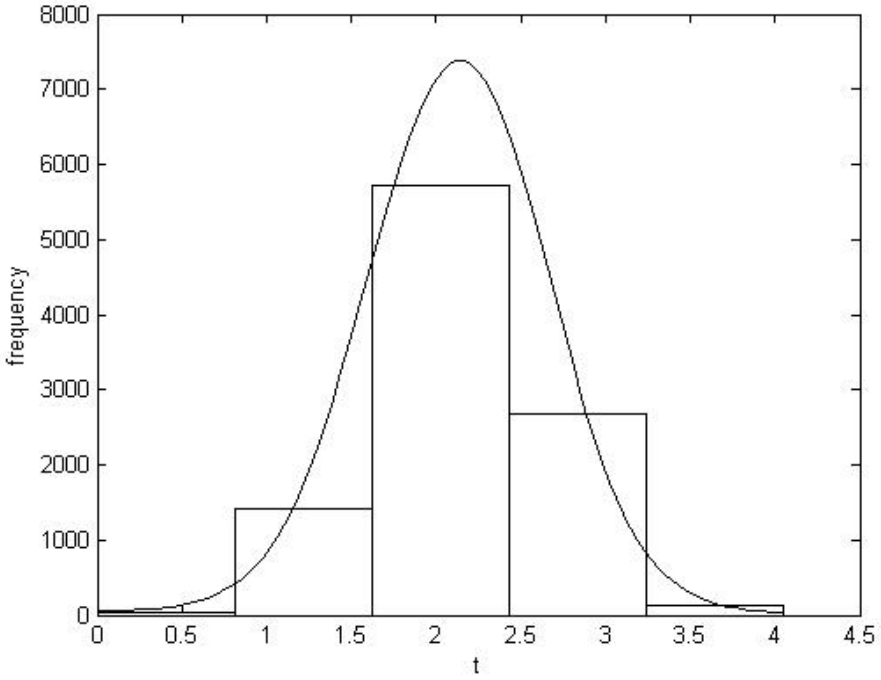

Figure 4.2. Renormalized optimal spline for estimating distribution $e^{f(t)}$

4.1 Biometrical Example

As a biometrical application of the smoothing spline estimation discussed in this chapter, we consider the estimation for the distribution of patient deaths, see Figure 4.3, from simulated data. The observed data represents 600 subjects treated for a disease based on the survival model

$$x_{t+1} = x_t + .05 + .15\varepsilon_t.$$

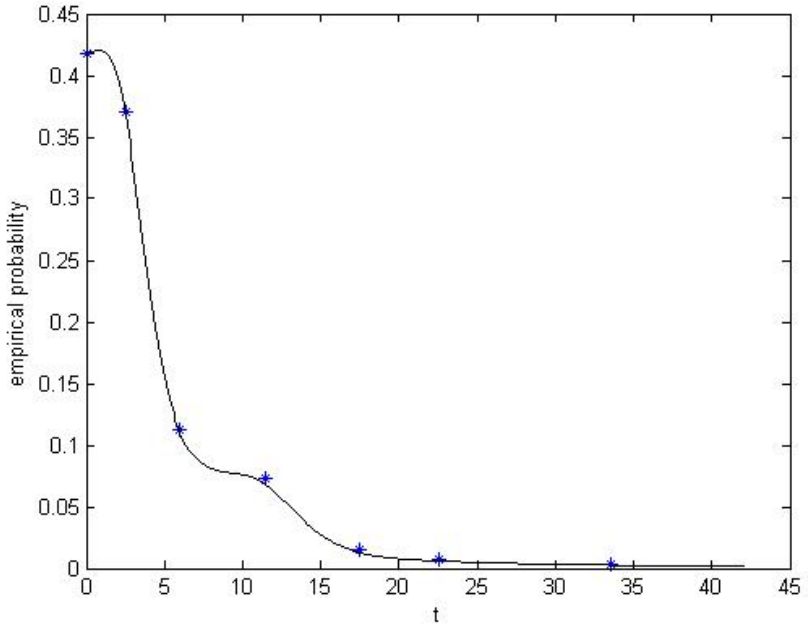

Figure 4.3. Optimal spline estimating distribution of deaths from simulated data

In this model, $x(0)$ is assumed to be uniformly distributed between 0 and 0.15, ε_t is normally distributed with zero mean, and x_{t+1} represents the stage of the subject's disease in the next time point. Measurements for each subject is terminated once the subject has either been cured or has died. Such survival models can be used to

determine the fraction of the population that will survive pass a particular time, or the rate at which those who do survive eventually die.

4.2 Reweighting Distributions with Zero Probability on Intervals

In this section, we describe reweighting strategies implemented for solving the problem of probability distribution estimation when there are intervals of zero probability. This estimation problem was addressed in Chapter 3, however, we discovered that imposing the non-negativity constraint could not be solved using the Hilbert space construction for our example. Here, we will show how methods within this chapter can be used for estimation.

Consider the interval j, in the empirical distribution, such that $\alpha_j = 0$ with $\alpha_{j+1} > 0$ or $\alpha_{j-1} > 0$. We select one of these positive probabilities, say α_{j-1}, and distribute its weight equally amongst the intervals $j-1$ and j. Now that all probabilities are strictly positive, we can use the methods of this chapter to estimate the distribution function $f(t)$ since $f(t) = e^{\ln(f(t))}$. This process can be easily extended to multiple adjacent intervals having zero probability as is the case shown in Figure 4.4. Here, the approximation handles the problem of non-negativity and has area of one, however, we observe that the estimates for intervals of zero probability are significantly larger than zero.

An alternative reweighting strategy involves assigning frequencies close to zero, say 10^{-10}, to intervals that have zero empirical probability as shown in Figure 4.5. The problem with this estimation procedure comes about because we are using natural logarithms of the empirical probabilities, as such, the value $\ln 10^{-10}$ goes toward $-\infty$. Thus, the spline closely approximates the intervals with zero probability but produces a poor fit to other intervals.

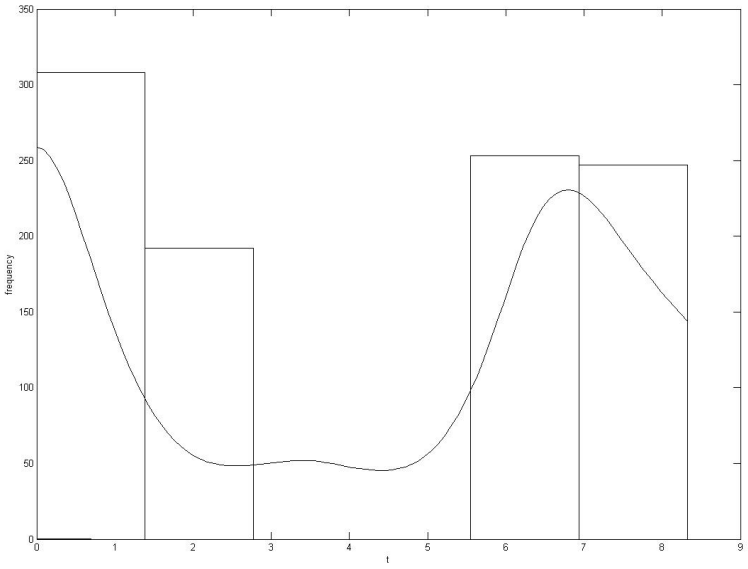

Figure 4.4. Optimal spline $y(t)$ obtained by redistributing probabilities.

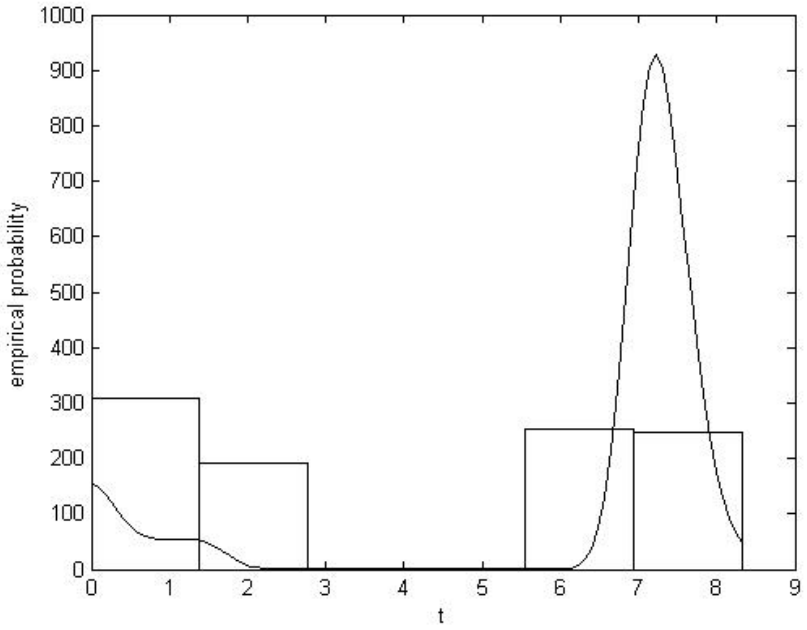

Figure 4.5. Optimal spline $y(t)$ obtained by replacing zero probabilities with 10^{-10}.

Chapter V

MONOTONE SPLINES

The type of smoothing spline construction considered so far in this paper may not be enough, especially in cases where one requires the distribution to have certain characteristics such as monotonicity. This feature corresponds to an infinite dimensional,non-negativity constraint on the first derivative of the curve. In this chapter, we will focus on the use of monotone cubic smoothing splines in cumulative distribution function, CDF, estimation on a specified interval [0,T]. Given the empirical probability distribution defined on $[0, T]$ subdivided into $N \leq 10$ intervals, we select a nodal value t_i in each subinterval and the relative frequency of the interval τ_i. Thus, we consider the data set

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

where $\alpha_i = \sum_{j=1}^{i} \tau_j \geq 0$, $0 < t_1 < \cdots < t_N \leq T$. Moreover, the approximating spline $y(t)$ needs to satisfy the following constraints

$$y \in C^1[0, T], \dot{y}(t) \geq 0, y(0) = 0 \text{ and } y(T) = 1.$$

The requirement that $y \in C^1[0, T]$ is necessary as the derivative of the CDF results in the continuous probability density function from which the data was sampled. As we have discovered in chapter 3, inequality constraints are quite difficult to achieve via the Hilbert space construction. In fact, proceeding with the Hilbert space construction when no constraints on the derivative are imposed on the curve,

produces CDF approximations as shown in Figure 5.1 which clearly does not solve the monotonicity problem. Hence, following [2] and [4], we will show that our approximation can be obtained numerically, in a finite setting, with a dynamic programming algorithm in the case of second order systems.

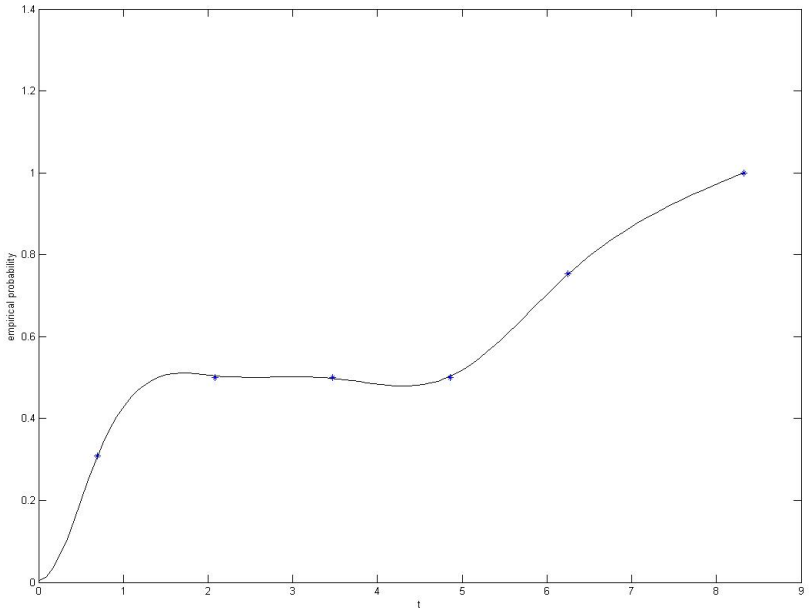

Figure 5.1. Optimal CDF with no derivative constraint imposed

5.1 Smoothing Splines with Derivative Constraints

In this section, we describe the monotone cubic spline construction using the Hilbert space methods and via Lagrange multipliers. Here, we assume without loss of generality that $x_0 = 0$ in (2.1), since the initial data can be absorbed into the data [9]. Thus, our goal is to obtain a control function $u^* \in L_2[0, T]$ that minimizes

the cost

$$J(u) = \lambda \int_0^T u^2(t)dt + \sum_1^N \omega_i \left(\int_0^T l_i(s)u(s)ds - \alpha_i \right)^2. \tag{5.1}$$

5.1.1 Hilbert Space Spline Construction

This spline construction seeks to obtain finite reparametrization of the minimization problem using similar Hilbert space methods discussed in previous chapters. We assume that the nodes $0 < t_1 < \cdots < t_N = T$ and begin this construction by considering the interval $[0, t_1)$. The problem thus reduces to fitting a curve $y(t)$ between the points (0,0) and (t_1, α_1) under the given constraints $y(0) = 0 = \delta_1$, $y(t_1) = \delta_2$, $\delta_1 \leq \delta_2$, and $\dot{y}(t) \geq 0$ for all $t \in [0, t_1)$. We define the constraint variety as

$$V_{\delta_2} = \left\{ u : \int_0^{t_1} (t_1 - s)u(s)ds = \delta_2 \right\}$$

and the orthogonal complement of V_0 is

$$V_0^{\perp} = \left\{ v : \int_0^{t_1} v(s)u(s)ds = 0 \forall u \in V_0 \right\}.$$

The optimal control that solves this problem is given by

$$u^*(s) = \begin{cases} k_1(t_1 - s) : s \in [0, t_1) \\ 0 \qquad\qquad : otherwise \end{cases}$$

where from the initial conditions we get

$$k_1 = \frac{\delta_2 - \delta_1}{\int_0^{t_1} (t_1 - s)^2 ds}.$$

Then, for the remaining intervals, that is $[t_j, t_{j+1})$ for $j = 1, \cdots, N-1$, we want to fit a curve $y(t)$ between the points (t_j, α_j) and (t_{j+1}, α_{j+1}) under the constraints $y(t_j) = \delta_{j+1}$, $y(t_{j+1}) = \delta_{j+2}$, $\delta_{j+2} \geq \delta_{j+1}$ and $\dot{y}(t) \geq 0$. Proceeding as before gives the control function

$$u^*_{j+1}(s) = \begin{cases} k_{j+1}(t_{j+1} - s) : s \in [t_j, t_{j+1}) \\ 0 \qquad\qquad\qquad : otherwise \end{cases}$$

where $k_{j+1} = \frac{\delta_{j+2} - \delta_{j+1}}{\int_{t_j}^{t_{j+1}} (t_{j+1} - s)^2 ds}$, for $j = 1, \cdots, N-1$. From this construction we get that the optimal spline that approximates the data is

$$y(t) = \begin{cases} k_1 \int_0^t (t-s)(t_1 - s) ds & : t \in [0, t_1) \\ k_{j+1} \int_{t_j}^t (t-s)(t_{j+1} - s) ds + \delta_{j+1} & : t \in [t_j, t_{j+1}), j = 1, \cdots, N-1 \\ \delta_{N+1} & : t = t_N = T. \end{cases}$$

The problem then becomes one of determining the $(N+1)$-vector δ that minimizes the cost

$$\begin{aligned} J(\delta) &= \lambda k_1 \int_0^{t_1} u(s) ds + \lambda \sum_1^{N-1} \int_{t_j}^{t_{j+1}} u_{j+1}(s) ds + \sum_1^N \omega_j (\delta_{j+1} - \alpha_j)^2 \\ &= \frac{\lambda(\delta_2 - \delta_1)}{\int_0^{t_1} (t_1 - s)^2 ds} + \lambda \sum_1^{N-1} \frac{\delta_{j+2} - \delta_{j+1}}{\int_{t_j}^{t_{j+1}} (t_{j+1} - s)^2 ds} \int_{t_j}^{t_{j+1}} (t_{j+1} - s) ds \\ &\quad + \sum_1^N \omega_j (\delta_{j+1} - \alpha_j)^2 \\ &= (\delta_2 - \delta_1) a_1 + \sum_1^{N-1} (\delta_{j+2} - \delta_{j+1}) a_{j+1} + \sum_1^N b_j (\delta_{j+1} - \alpha_j)^2 \end{aligned}$$

subject to $\delta_{j+2} - \delta_{j+1} \geq 0, \delta_1 = 0,$ and $\delta_{N+1} = 1$. We may write this problem in

equivalent matrix form

$$\min_{\delta} f'\delta + \frac{1}{2}\delta' H\delta,$$

where $f' = [-a_1, (a_1 - a_2 - 2\alpha_1 b_1), \cdots, (a_{N-1} - a_N - 2\alpha_N b_N), (a_N - 2\alpha_N b_N)]$, H is a $(N+1) \times (N+1)$ matrix with components H =diag$\{0, 2\omega_1, \cdots, 2\omega_N\}$ and constants

$$a_{j+1} = \begin{cases} \dfrac{\lambda \int_0^{t_1} (t_1 - s)ds}{\int_0^{t_1} (t_1 - s)^2} & : j = 0 \\ \dfrac{\lambda \int_{t_j}^{t_{j+1}} (t_{j+1} - s)ds}{\int_{t_j}^{t_{j+1}} (t_{j+1} - s)^2 ds} & : j = 1, \cdots, N-1. \end{cases}$$

We have thus reduced the problem to a quadratic programming problem that can be solved using standard software. Using this optimization routine produces curves as shown in Figure 5.2.

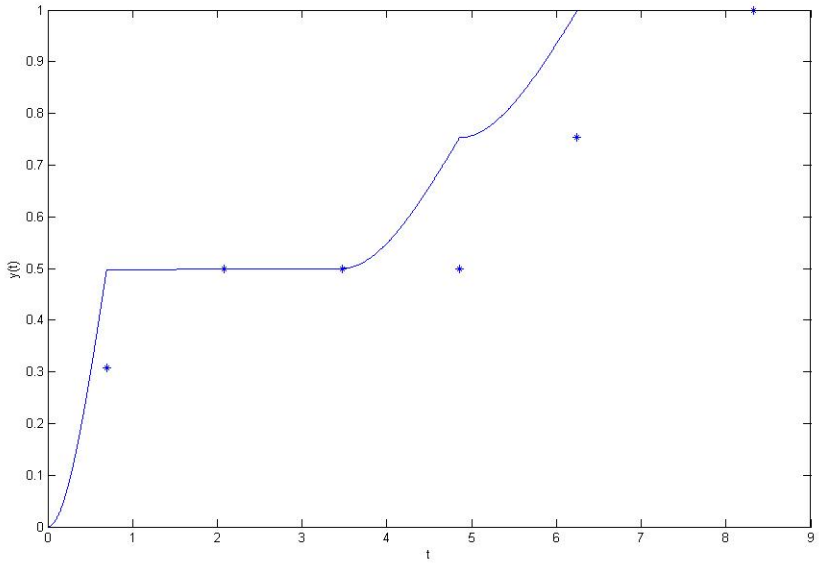

Figure 5.2. Optimal spline with equal weight assigned to all way points.

Here we obtain an approximation which is monotonically increasing and satisfies

the end conditions $y(0) = 0$ and $y(T) = 1$, however, the spline is not differentiable everywhere on $[0, T]$.

5.1.2 Lagrangian Spline Construction

Based on the cost function defined in (5.1) we can form the associated Lagrangian

$$L(u, \nu) = \lambda \int_0^T u^2(t)dt - \int_0^T \dot{y}(t)d\nu(t) + \sum_1^N \omega_i \left(\int_0^T l_i(s)u(s)ds - \alpha_i \right)^2 \quad (5.2)$$

where $\dot{y}(t) = \int_0^t u(s)ds \geq 0$, $\nu \in BV[0, T]$, the space of functions of bounded variation on $[0, T]$, which is the dual space of $C[0, T]$ [8]. Integrating the Stieltjes integral (5.2) by parts yields

$$\int_0^T \int_0^t u(s)ds d\nu = \int_0^T u(t)\nu(T)dt - \int_0^T u(t)\nu(t)dt$$

so the Lagrangian becomes

$$L(u, \nu) = \lambda \int_0^T u^2(t)dt - \int_0^T (\nu(T) - \nu(t))u(t)dt + \sum_1^N \omega_i \left(\int_0^T l_i(s)u(s)ds - \alpha_i \right)^2.$$

Thus, the optimal curve [2] is determined by solving the problem

$$\max_{\nu \geq 0} \inf_u L(u, \nu). \quad (5.3)$$

It is shown in [2] and [4] that the set of control functions which solves this optimization problem exists and is unique. Moreover, due to the convexity of the problem, we can replace the inf in (5.3) with min. As such, we may obtain the optimal control function by calculating the Fréchet differential of L with respect to

u. Following [2], we let $L_\nu(u) = L(u, \nu)$, then for $h \in L_2[0, T]$

$$\partial L_\nu(u, h) = \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left(L_\nu(u + \varepsilon h) - L_\nu(u) \right).$$

Now the terms in the differential simplify as follows

$$\frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left(\int_0^T (\nu(T) - \nu(t))((u + \varepsilon h)(t) - u(t))dt \right) = \int_0^T (\nu(T) - \nu(t))h(t)dt$$

$$\begin{aligned} \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left(\lambda \int_0^T \left((u + \varepsilon h)^2(t) - u^2(t) \right) dt \right) &= \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \lambda \int_0^T (2\varepsilon u(t)h(t) + \varepsilon^2 h^2(t))dt \\ &= 2\lambda \int_0^T u(t)h(t)dt. \end{aligned}$$

For simplicity we define

$$K = \int_0^T l_i(t)u(t)dt - \alpha_i$$

$$\begin{aligned} \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left\{ \left(\int_0^T l_i(u + \varepsilon h)dt - \alpha_i \right)^2 - K^2 \right\} &= \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left\{ \left(K + \varepsilon \int_0^T l_i h dt \right)^2 - K^2 \right\} \\ &= \frac{1}{\varepsilon} \lim_{\varepsilon \to 0} \left(2K\varepsilon \int_0^T l_i h dt + \varepsilon^2 \int_0^T l_i h dt \right) \\ &= 2K \int_0^T l_i(t)ht(t)dt. \end{aligned}$$

Therefore, the Fréchet differential of $L_\nu(u)$ reduces to

$$\int_0^T \left(2\lambda u(t) - (\nu(T) - \nu(t)) \right) h(t)dt + 2 \sum_1^N \omega_i \int_0^T K l_i(t)h(t)dt.$$

Hence, the differential is zero for all $h \in L_2[0,T]$ whenever

$$2\lambda u^*(t) + 2\sum_i^N \omega_i \left(\int_0^T l_i(t)u^*(t)dt - \alpha_i \right) l_i(t) - (\nu(T) - \nu(t)) = 0.$$

The above equation is true especially for the optimal $\nu = \nu^*$, which gives

$$2\lambda u^*(t) + 2\sum_i^N \omega_i \left(\int_0^T l_i(t)u^*(t)dt - \alpha_i)l_i(t \right) - C_t = 0,$$

where $C_t = \nu^*(T) - \nu^*(t) \geq 0$ from the non-negativity constraint on ν^* whenever $\dot{y}(t) > 0$ [3]. For the second order system, $l_i(t)$ is linear in t for $i = 1, \cdots, N$, and so the above equation gives that $u^*(t)$ has to be piecewise linear. Based on the definition of $l_i(t)$, we have that the optimal control changes at the specified way points and whenever C_t changes, that is $\dot{y}(t) = 0$ [3]. Also, if $\dot{y}(t) = 0$ on an interval, then $u^*(t) = 0$. Therefore the optimal control is a piecewise linear function for all $t \in [0,T]$.

To determine the optimal control u^* that minimizes our cost using this Lagrangian method for spline construction, requires first determining the optimal function $\nu^* \in BV[0,T]$. This increases the difficulty of obtaining a solution to our problem and for this reason we have chosen to go no further with such construction.

5.2 Dynamic Programming

In this section, we illustrate the reformulation of the monotone problem in a finite setting that can be handled easily. Given the nodal values t_i and t_{i+1}, the positions y_i and y_{i+1}, and the corresponding derivatives $\dot{y}_i$ and $\dot{y}_{i+1}$, our problem can be stated as follows: *How do we drive the system* $\dot{x} = Ax + bu, y = cx$ *between* $(y_i, \dot{y}_i)$

and $(y_{i+1}, \dot{y}_{i+1})$ *with a piecewise linear control function that changes whenever* $\dot{y}(t) = 0$ *such that* $\dot{y}(t) \geq 0 \forall t \in [t_i, t_{i+1}]$ *while minimizing the quadratic cost function 5.1?* Dividing the cost function (5.1) into an interpolation and smoothing part yields the optimal value function [1]

$$\begin{cases} \hat{S}_i(y_i, \dot{y}_i) = \min\limits_{y_{i+1} \geq y_i, \dot{y}_{i+1} \geq 0} \{\lambda V_i(y_i, \dot{y}_i, y_{i+1}, \dot{y}_{i+1}) + \hat{S}_{i+1}(y_{i+1}, \dot{y}_{i+1})\} \\ \qquad\qquad +\omega_i(y_i - \alpha_i)^2, i = 0, \cdots, N-1 \\ \hat{S}_N(y_N, \dot{y}_N) = \omega_N(y_N - \alpha_N)^2 \end{cases}$$

subject to $\sum_0^N (y_{i+1} - y_i) = 1$, which is equivalent to $y(T) = 1$, where $V_i(y_i, \dot{y}_i, y_{i+1}, \dot{y}_{i+1})$ is the cost for driving the system between the points $(y_i, \dot{y}_i)$ and $(y_{i+1}, \dot{y}_{i+1})$ on the interval $[t_i, t_{i+1}]$, while keeping the derivative nonnegative.

The optimal solution is thus found by determining $\hat{S}(0, 0)$ where we let $\omega_0 = 0$ and α_0 be an arbitrary number. Solving this dynamic programming problem reduces to determining the function $V_i(y_i, \dot{y}_i, y_{i+1}, \dot{y}_{i+1})$, which is equivalent to finding the $2 \times N$ variables $y_1, \cdots, y_N, \dot{y}_1, \cdots, \dot{y}_N$. This is the finite reparametrization of the infinite dimensional problem.

Under specified assumptions, in [2] and [4], the cost in the optimal value function reduces to

$V_i(y_i, \dot{y}_i, y_{i+1}, \dot{y}_{i+1}) =$

$$\begin{cases} 4\frac{\dot{y}_i(t_{i+1}-t_i)^2 - 3(y_{i+1}-y_i)(t_{i+1}-t_i)(\dot{y}_i+\dot{y}_{i+1}) + 3(y_{i+1}-y_i)^2 + (t_{i+1}-t_i)^2\dot{y}_{i+1}^2}{(t_{i+1}-t_i)^3}, \\ \text{if } y_{i+1} - y_i \geq \chi(t_{i+1} - t_i, \dot{y}_i, \dot{y}_{i+1}), \\ \dfrac{4(\dot{y}_{i+1}^{3/2} + \dot{y}_i^{3/2})^2}{9(y_{i+1} - y_i)}, \text{ if } y_{i+1} - y_i < \chi(t_{i+1} - t_i, \dot{y}_i, \dot{y}_{i+1}) \end{cases}$$

where $\chi(t_{i+1} - t_i, \dot{y}_i, \dot{y}_{i+1}) = \frac{t_{i+1}-t_i}{3}(\dot{y}_i + \dot{y}_{i+1} - \sqrt{\dot{y}_i \dot{y}_{i+1}})$ and $t_0 = y_0 = \dot{y}_0 = 0$. Using the dynamic programming algorithm for CDF approximation yields optimal curves shown in Figures 5.3 and 5.4. In Figure 5.3, we return to the estimation problem introduced in Chapter 3, that of finding the approximation of the probability distribution that is zero on some specified interval(s). This characteristic translates to CDF estimation which is constant on the corresponding interval.

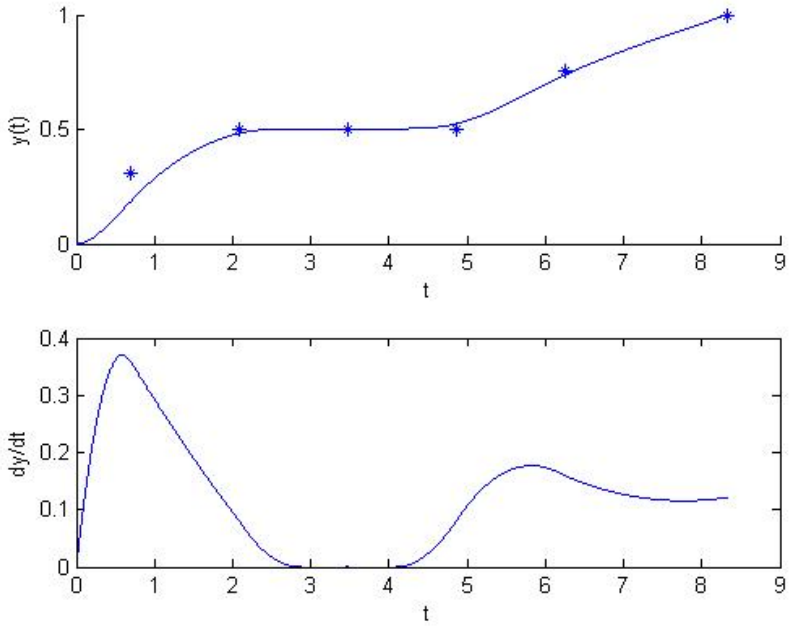

Figure 5.3. Optimal CDF with constant on interval via dynamic programming

The results depicted in Figure 5.4 show the distribution estimation where the data is simulated from the density

$$f(x) = \begin{cases} \dfrac{x^{11}(1-x)^{12}e^{-(x-.75)^2/.2}}{\int_0^1 x^{11}(1-x)^{12}e^{-(x-.75)^2/.2}dx}, x \in [0,1] \\ 0, \text{ otherwise} \end{cases}$$

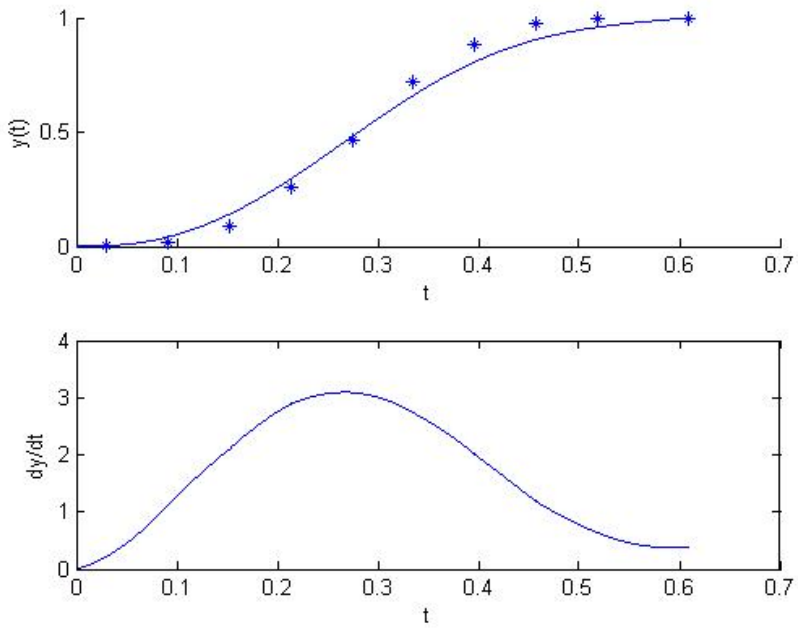

Figure 5.4. Optimal CDF estimate using dynamic programming

In addition to the CDF approximation, by taking the derivative of the optimal we have an estimate for the continuous density function from which the data was sampled. The dynamic programming algorithm implemented for CDF estimation produces a spline $y(t)$ satisfies all required constraints of our problem, that is,

$$y(0) = 0, y(T) = 1, \dot{y}(t) \geq 0, \text{ and } y \in C^1[0, T].$$

Chapter VI

CONCLUSION

The control theoretic splines discussed in this paper present methods for probability and cumulative distribution estimation. The term *control theoretic spline* refers to the way the spline function is constructed, that is, as the output of a dynamical system driven by a control function [7]. Our goal was to find smooth approximations to these distribution functions when given the empirical distributions. The smoothing spline construction for probability distribution estimation was a basic application of the Hilbert Projection Theorem [8]. This theorem was implemented to find the optimal control function which drives the smoothed data closest to the empirical probabilities given, while minimizing the cost functional. Thus, in our probability distribution estimations, we have essentially generalized the problem in Euclidean geometry of finding a point on a given line closest to a given point in the plane.

In chapters 2 through 4, we discussed the cubic smoothing spline construction based on different constraints imposed on the linear control system (2.1). We have shown that these constructions deal with two problems in the estimations of probability distributions. The first is that the spline approximating the distribution has to be nonnegative (strictly positive in the case of distributions of the form $e^{f(t)}$), and the second that the area has to equal one. The smoothing splines described in this paper satisfied our main goal which was to find curves that closely approximate the empirical distributions with minimum variance on the residuals. In addition, we have shown that this minimization process reduces the nonparametric problem to a

problem of calculation of parameters in a finite dimensional space.

The Hilbert space construction discussed was obtained by solving systems of linear equations based on the specified constraints. In many instances, one may want the optimal spline found to possess certain characteristics such as monotonicity in the estimation of cumulative distribution functions. This property corresponds to non-negativity constraints on the first derivative. For this inequality constraint on the derivatives, the type of construction based on the Hilbert Projection Theorem is not enough. In chapter 5, we have, however, obtained solutions to such problems by exploiting a finite reparametrization of the problem, leading to a dynamic programming formulation developed in [2] and [4]. Later methods of monotone spline construction include work shown in [10]. For the second order system considered, the monotone cubic splines converge quadratically to the probability distribution function. We expect much faster convergence rates using monotone quintic splines, however, this construction is yet to be developed.

BIBLIOGRAPHY

[1] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. 1, Athena Scientific, Belmont, MA, 1995.

[2] M. Egerstedt and C.F. Martin, *Control Theoretic Splines: Optimal Control, Statistics,and Path Planning*, 1st ed, Princeton University Press, (in press).

[3] M. Egerstedt and C.F. Martin, *Optimal control and monotone smoothing splines*, Lecture Notes in Control and Information Systems, Vol. 295, Springer, Berlin, 2004, p. 279-294.

[4] M. Egerstedt and C.F. Martin, *Monotone smoothing splines*, Proceedings of the Mathematical Theory of Networks and Sytems, Perpignan, France, 2000.

[5] M. Egerstedt and C.F. Martin, *Trajectory planning for linear control systems with generalized splines*, Proceedings of the Mathematical Theory of Networks and Systems, Padova, Italy, 1998.

[6] R.L. Eubank, *Nonparametric Regression and Spline Smoothing*, Vol. 157, Statistics: Textbooks and Monographs, Marcel Dekker, Inc., New York, NY, 1999.

[7] U.T. Jönsson, C.F. Martin, and Y. Zhou, *Trajectory planning for systems with a multiplicative stochastic uncertainty*, International Journal of Control, Vol. 77, No. 8, 2004, p.713-722.

[8] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York-London-Sydney, 1969.

[9] C. F. Martin, S. Sun and M. Egerstedt, *Optimal control, statistics, and path planning*, Mathematics and Computer Modeling, Vol. 33, 2001, p. 237 - 253.

[10] M.C. Meyer, *Inference using shape-restricted regression splines*, Annals of Applied Statistics, Vol. 2, No. 3, 2008, p.1013-1033.

[11] B.W. Silverman, *Some aspects of the spline smoothing approach to nonparametric regression curve fitting*, J. Royal Statist, Soc. B, Vol. 47, No. 1, 1985, p.1-52.

[12] B.W. Silverman, *Spline smoothing: the equivalent variable kernel method*. Ann. Statist., Vol. 12, No. 3,1984, p. 898-916.

[13] S. Sun, M. Egerstedt and C. F. Martin,*Control theoretic smoothing splines*,IEEE Transactions on Automatic Control, Vol. 45, No. 12, 2000, p. 2271-2279.

[14] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[15] E.J. Wegman and I.W. Wright,*Splines in statistics*, Journal of American Statistical Association, Vol. 78, No.382, 1983, p. 351-365.

[16] Y. Zhou, M. Egerstedt and C.F. Martin, *Hilbert space methods for control theoretic splines: A unified treatment*,Communications in Information and Systems, Vol. 6, No.1, 2006, p. 55-82.

[17] Y.Zhou, W. Dayawansa, and C.F. Martin, *Control theoretic smoothing splines are approximate linear filters*,Communications in Information and Systems, Vol. 4, No. 3, 2004, p. 253-272.