ESAIM: Control, Optimisation and Calculus of Variations August 2003, Vol. 9, 553–562
DOI: 10.1051/cocv:2003026

STATISTICAL ESTIMATES FOR GENERALIZED SPLINES *

Magnus Egerstedt[1] and Clyde Martin[2]

Abstract. In this paper it is shown that the generalized smoothing spline obtained by solving an optimal control problem for a linear control system converges to a deterministic curve even when the data points are perturbed by random noise. We furthermore show that such a spline acts as a filter for white noise. Examples are constructed that support the practical usefulness of the method as well as gives some hints as to the speed of convergence.

Mathematics Subject Classification. 93-xx.

Received March 29, 2001. Revised November 27, 2002.

1. Introduction

We consider a data set of the form $\{(t_i, f(t_i) + \epsilon_i) : \quad i = 1, \ldots, N\}$, where $t_i,\ i = 1, \ldots, N$, are the fixed sample times, $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, and the ϵ_i's are independent, identically normally distributed random variables with mean 0 and variance σ^2. Our goal is to construct a curve using this data that is smooth and is produced from the output of a single-input, single-output dynamical system. The main result, presented in this paper, is that under quite general conditions the constructed curve converges to a deterministic curve that is a function of the curve $f(t)$. The reason why it does not converge exactly to $f(t)$ is that we want to be able to guarantee that the curve exhibits certain differentiability properties, which we achieve by minimizing the L_2-norm of the control signal, as shown in [6].

In this paper we continue the program, initiated in [9,11], of producing *smoothing splines* as computationally feasible solutions to optimal control problems. The types of relaxed interpolation problems that we investigate need to be solved for a number of different reasons. For instance, in air traffic control we need to be able to specify the position that the system will be in at a sequence of times. However, in most situations, it is not really crucial that we pass through these points exactly, but rather that we go reasonably close to them, while minimizing the cost functional. This is a desirable property for two apparent reasons. First of all, a small deviation from the prespecified point can result in a significant decrease in the cost, and secondly, when the data that we work with is noise contaminated, which is the case in this paper, it is not even desirable to interpolate through these points exactly. Inspired by [12] and [13], we can incorporate these aspects into our proposed method, and by not demanding exact interpolation, we end up with smoothing splines instead of the

Keywords and phrases. Optimal control, smoothing splines, linear systems, interpolation.

* *The first author gratefully acknowledges the support from the Swedish Foundation for Strategic Research through its Centre for Autonomous Systems. The second author was supported in part by NASA Grants NAG 2-902 and NAG 2-899 and by NSF Grant DMS 9628558.*

[1] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA;
e-mail: `magnus@ece.gatech.edu`

[2] Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA; e-mail: `martin@math.ttu.edu`

standard splines. Further generalizations of this work have been developed in a series of papers by Egerstedt and Martin and their collaborators in [1, 7, 9, 11, 14] but what is novel in this paper is that we focus explicitly on the convergence aspects of the problem. A somewhat similar approach for interpolating splines has been developed by Crouch and Silva–Leite and their collaborators in [2–5] with emphasis on deriving splines on non-Euclidean spaces.

We assume as given a linear, single-input, single-output control system

$$\begin{aligned} \dot{x} &= Ax + bu,\ x(0) = 0 \\ y &= cx, \end{aligned} \tag{1.1}$$

where $x \in \mathbb{R}^n,\ u, y, \in \mathbb{R}$, and A, b, c are constant matrices and vectors of compatible dimensions. We furthermore assume that the system (A, b, c) is both controllable and observable as well as $cb = 0$. This last condition assures smoothness of the splines, as shown in [11].

Three distinct cost functions will be central to the development of our theory, and for the sake of easy reference we introduce them already at this point:

Definition 1.1.

$$J_N^{\epsilon}(u) = \frac{1}{N}\sum_{i=1}^{N}(y(t_i) - f(t_i) - \epsilon_i)^2 + \lambda \int_0^T u(t)^2 \mathrm{d}t,$$

where $\lambda > 0$, $\epsilon_i \sim N(0, \sigma^2),\ i = 1, \ldots, N$, and $0 \le t_1 < \cdots < t_N \le T$.

Definition 1.2.

$$J_N(u) = \frac{1}{N}\sum_{i=1}^{N}(y(t_i) - f(t_i))^2 + \lambda \int_0^T u(t)^2 \mathrm{d}t,$$

where $\lambda > 0$ and $0 \le t_1 < \cdots < t_N \le T$.

Definition 1.3.

$$J(u) = \int_0^T (y(t) - f(t))^2 + \lambda u(t)^2 \mathrm{d}t,$$

where $\lambda > 0$.

We will prove that as N approaches infinity the solution to the problem of minimizing J_N^{ϵ} with respect to $u \in L_2[0, T]$, under the dynamics $\dot{x} = Ax + bu,\ y = cx,\ x(0) = 0$, converges to the solution of the corresponding optimal control problem with cost function $J(u)$. The interpretation of this is that, in the limit, *i.e.* as the number of sample points goes to infinity, the optimal control procedure removes the error from the solution. Thus the generalized smoothing splines act as filters for data corrupted by Gaussian noise.

The proof of this convergence theorem will constitute the main part of this paper: first we show, in Section 3, that if the data consists of noise alone then in the limit the spline approaches the zero function. This takes full advantage of the fact that the noise is assumed to be identically normally distributed and that the different samples are independent. We then show, in Section 4, that the optimal solution of J_N^{ϵ} can be decomposed uniquely as the sum of the pure noise term and the solution of the optimal control problem with cost function J_N. As we have already shown in [11], the optimal solution of J_N converges to the optimal solution of J, and hence the main theorem follows.

2. Basic results

The material in this section is contained in an equivalent form in [9, 11] and is included for the convenience of the reader. We begin by deriving the form of the solution to the optimal control problem under cost function J_N^{ϵ}.

We first establish some notation. Let

$$g_s(t) = \begin{cases} ce^{A(s-t)}b & s-t \geq 0 \\ 0 & s-t < 0, \end{cases}$$

which allow us to define the linear functional

$$G_s(u) = \int_0^T g_s(t)u(t)\mathrm{d}t,$$

which gives $y(s) = G_s(u)$ since $x(0) = 0$. We can rewrite the cost function $J_N^\epsilon(u)$ using the functionals G_{t_i} as

$$J_N^\epsilon(u) = \frac{1}{N}\sum_{i=1}^{N}(G_{t_i}(u) - f(t_i) - \epsilon_i)^2 + \lambda\int_0^T u(t)^2\mathrm{d}t. \tag{2.1}$$

In this form J_N^ϵ is an explicit quadratic functional of u and we find the minimizing u by calculating the Gateaux derivative of J_N^ϵ with respect to u and setting the result equal to zero. Performing this calculation we get that the optimal u^* satisfies

$$\frac{1}{N}\sum_{i=1}^{N}(G_{t_i}(u^*) - f(t_i) - \epsilon_i)g_{t_i}(u^*) + \lambda u^* = 0.$$

To solve this for u^*, one critical observation is that the expression

$$(G_{t_i}(u^*) - f(t_i) - \epsilon_i)$$

is constant and hence the optimal control u^* can be expressed as a linear combination of the basis functions g_{t_i}. We have thus reduced the optimal control problem from an infinite dimensional problem to a finite dimensional problem.

We can now use the fact that

$$u^*(t) = \sum_{i=1}^{N}\rho_i g_{t_i}(t) \tag{2.2}$$

to get u^* by calculating the optimal values of the parameters ρ_i. Equation (2.2) gives, after some straight forward manipulations, the following optimal cost

$$J_N^\epsilon(u^*) = \frac{1}{N}\left(\rho^T G^2\rho - 2(\hat{f}+\hat{\epsilon})^T G\rho + \sum_{i=1}^{N}(f(t_i) - \epsilon_i)^2\right) + \lambda\rho^T G\rho. \tag{2.3}$$

Here, G is the Grammian $[g_{ij}]_{ij}$ where

$$g_{ij} = \int_0^T g_{t_i}(t)g_{t_j}(t)\mathrm{d}t,$$

and ρ is the vector of unknown parameters ρ_i, $\hat{f}$ is the vector of functional values $f(t_i)$, and $\hat{\epsilon}$ is the vector of errors $\epsilon_i,\ i = 1,\ldots,N$. Since J_N^ϵ is quadratic in ρ we can calculate the minimum value by taking the derivative with respect to ρ. This directly leads to the following equation.

$$\frac{1}{N}\big(G\rho - (\hat{\epsilon}+\hat{f})\big) + \lambda\rho = 0, \tag{2.4}$$

or in other words

$$(G + \lambda NI)\rho = (\hat{\epsilon}+\hat{f}). \tag{2.5}$$

Now, the Grammian G is positive semi-definite (see for example [6]) and hence $G + \lambda NI$ is positive definite. Thus equation (2.5) has a unique solution. For future reference we state this as a lemma.

Lemma 2.1. *The matrix $G + \lambda NI$ is positive definite and hence the optimal solution of J_N^ϵ is uniquely determined by the parameters ρ. The optimal ρ depends only on the values of $\hat{f}, N, \lambda$ and $\hat{\epsilon}$.*

We now calculate the response to the optimal control. Since $y(t_i) = G_{t_i}(u^*)$ we can substitute the solution of equation (2.5) for u^*, which, after some manipulations, gives the following simple expression

$$y(t_i) = e_i^T G\rho,$$

where e_i is the i:th unit vector in $\mathbb{R}^N$.

3. The data is pure noise

In this section we will show that if the data consists solely of independent, normally distributed, zero-mean random noise then, as the number of samples goes to infinity, the optimal control that minimizes J_N^ϵ goes to zero as well.

We will derive our results under the assumption that $f(t) \equiv 0$, *i.e.* that the data points are pure noise, and we first note that $y(t_i)$ is a linear function of the noise $\hat{\epsilon}$ and hence the $y(t_i)$'s are normally distributed random variables for $i = 1, \ldots, N$. By linearity of the expectation it inherits the zero-mean property of the noise, and we state this fact as a remark.

Remark 3.1. If y is the output of the linear system (1.1) driven by the optimal control that solves J_N^ϵ when $f = 0$, then the sample mean goes to zero as N goes to infinity, *i.e.*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N y(t_i) = 0.$$

The second main step is to show that the sample variance converges to zero as well.

Lemma 3.2. *If y is the output of the linear system (1.1) driven by the optimal control that solves J_N^ϵ when $f = 0$, then the sample variance goes to zero as N goes to infinity, i.e.*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N y(t_i)^2 = 0.$$

Proof. We note that $J_N^\epsilon(u_N^*) \le J_N^\epsilon(0)$ due to the fact that u_N^* is optimal. (Here we explicitly use the subscript N to draw attention to the fact that N sample points are used in the optimal control problem J_N^ϵ.) We thus have

$$\begin{aligned} J_N^\epsilon(u_N^*) &= \frac{1}{N}\sum_{i=1}^N (y(t_i) - \epsilon_i)^2 + \lambda\int_0^T u_N^*(t)^2 \mathrm{d}t \\ &= \frac{1}{N}\sum_{i=1}^N y(t_i)^2 - \frac{2}{N}\sum_{i=1}^N \epsilon_i y(t_i) + \frac{1}{N}\sum_{i=1}^N \epsilon_i^2 + \lambda\int_0^T u_N^*(t)^2\mathrm{d}t \\ &\le \frac{1}{N}\sum_{i=1}^N \epsilon_i^2, \end{aligned}$$

which gives that

$$\frac{1}{N}\sum_{i=1}^N y(t_i)^2 \le \frac{2}{N}\sum_{i=1}^N y(t_i)\epsilon_i. \tag{3.1}$$

We now use the fact that $y(t_i) = e_i^T G\rho$ and concentrate on the right hand side of the inequality in (3.1). We have

$$\frac{2}{N}\sum_{i=1}^{N} y(t_i)\epsilon_i = \frac{2}{N}\hat{\epsilon}^T G\rho.$$

We furthermore know that $G + \lambda NI$ is positive definite, and hence that ρ is given by

$$\rho = \left(\frac{G}{\lambda N} + I\right)^{-1} \frac{\hat{\epsilon}}{\lambda N},$$

which gives that

$$\frac{2}{N}\hat{\epsilon}G\rho = \frac{2}{N}\hat{\epsilon}G\left(\frac{G}{\lambda N} + I\right)^{-1} \frac{\hat{\epsilon}}{\lambda N} = \sum_{i=1}^{N}\sum_{j=1}^{N}\zeta_{ij}\frac{\epsilon_i\epsilon_j}{N^2}.$$

Here we recall that $G = [g_{ij}]_{ij}$, where

$$g_{ij} = \int_0^{\min\{t_i,t_j\}} g_{t_i}(t)g_{t_j}(t)\mathrm{d}t$$

and hence there exists a uniform bound on all such inner products no matter where the points t_i and t_j are located. Thus, since $\frac{G}{\lambda N}$ goes to zero as N goes to infinity we have that ζ_{ij} must be uniformly bounded, which in turn implies that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y(t_i)^2 = 0,$$

due to the nature of the noise. □

We can now state as a theorem the main result of this section.

Theorem 3.3. *The mean and variance of the the random variables $y(t_i)$ are both equal to zero.*

4. The main result

In this section we will prove the following theorem:

Theorem 4.1. *Let the t_i's be such that*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g(t_i) = \int_0^T g(t)\mathrm{d}t$$

for every continuous function g. Let $h(t)$ be the optimal response to the optimal control problem with cost function

$$J(u) = \int_0^T (y(t) - f(t))^2\mathrm{d}t + \lambda\int_0^T u(t)^2\mathrm{d}t.$$

Let $y_N^\epsilon(t)$ be the optimal solution to the optimal control problem

$$J_N^\epsilon(u) = \frac{1}{N}\sum_{i=1}^{N}(y(t_i) - f(t_i) - \epsilon_i)^2 + \lambda\int_0^T u(t)^2\mathrm{d}t.$$

Then

$$\lim_{N\to\infty} y_N^\epsilon(t) = h(t),$$

in probability.

Proof. The proof relies on the fact that we can decompose the optimal solution of J_N^ϵ into the sum of the optimal solution of J_N and the optimal solution of the pure noise problem. This decomposition follows from the fact that the optimal solution of the pure noise problem is given by the solution of

$$(G + \lambda NI)\hat{\rho}_1 = \hat{\epsilon}$$

and that the optimal solution of J_N is given by

$$(G + \lambda NI)\hat{\rho}_2 = \hat{f}.$$

It is immediate that the optimal solution of J_N^ϵ is given by

$$(G + \lambda NI)(\hat{\rho}_1 + \hat{\rho}_2) = \hat{\epsilon} + \hat{f}.$$

Since the matrix $G + \lambda NI$ is positive definite the solution is unique and since y_N^ϵ depends linearly on ρ and we have shown, in Theorem 3.3 that $y_N^\epsilon \to 0$ in probability in the pure noise case, the theorem would follow if we could show that the optimal solution of J_N goes to the optimal solution of J. But, in [11] this was shown under the additional assumptions on the sample times as stated in this theorem, which concludes the proof. □

This result is already known to hold for standard smooth splines [8], and can thus be thought of as a more general version of that result.

5. Examples

Here we present two examples of the convergence of the splines in the presence of noise. Even though we work with a fairly simple example it is to be thought of as evidence for the practical usefulness of our proposed method. It also sheds some light on the convergence speed, as we will see.

Initially, let the system that we work with be given by

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

Since it is a second order nilpotent system, the spline is a standard cubic spline. (See for example [14].) The underlying curve $f(t)$ is taken to be a sin-function. The noise is furthermore taken to have zero mean and variance 0.05.

In Figure 1 we use four sample points and we see that the tracking is not too bad even with this few points. In other experiments we have noted that if we have one point in the interval $(0, .3)$, one point near $T = 1$, and one point on the right hand side of the curve the tracking is reasonable. As would be expected, if all four points are taken close to zero the spline fails to track at all.

In Figure 2 we use 20 sample points with the same error structure. Here we have quite good tracking and we see that there is a deviation occurring between the spline and the curve close to zero. This is because it takes a little bit of time for the system to start tracking the underlying curve under the assumption that $x(0) = 0$.

In Figure 3 we use 70 sample points. By now the tracking is quite good. There is little deviation between the spline of Figure 2 and the spline of Figure 3. The only difference is occurring near the maximal point of the underlying curve.

Now, consider the case when A is given by

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix},$$

in which case we have an exponential spline, as shown in [14]. By repeating the experiment on this new system, we get the results shown in Figures 4–6.

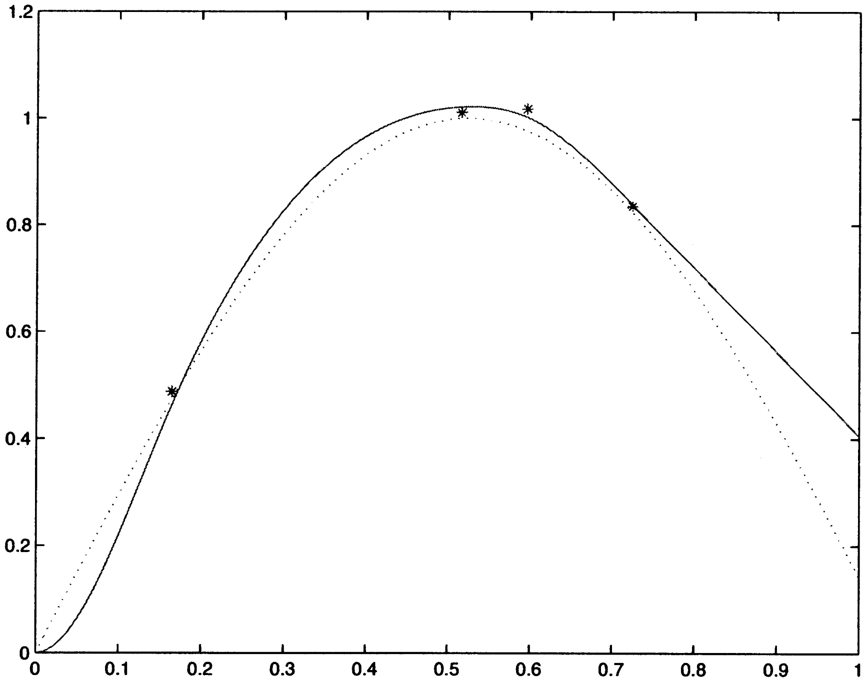

FIGURE 1. Tracking with four points using cubic splines.

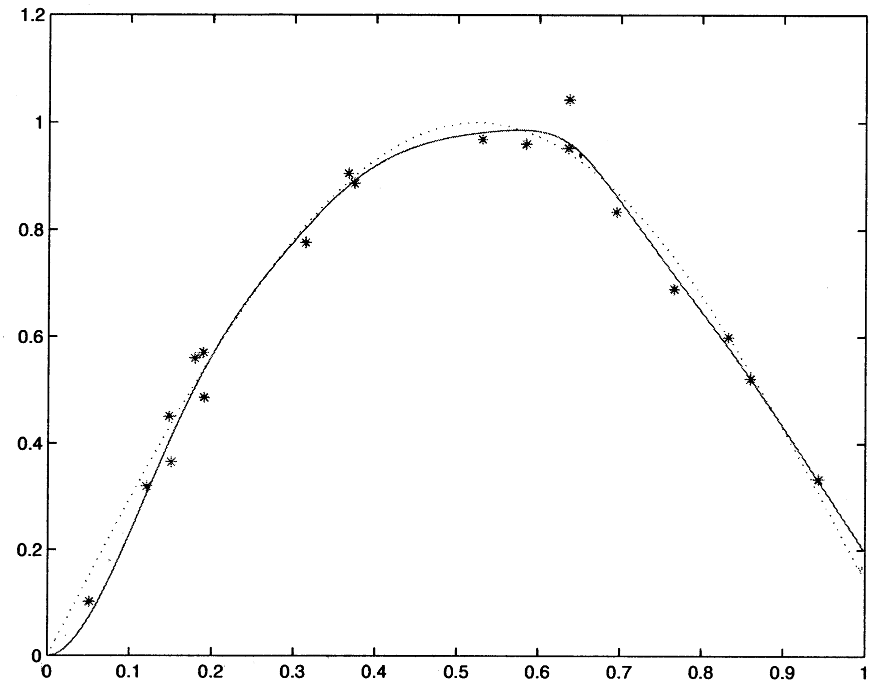

FIGURE 2. Tracking with 20 points using cubic splines.

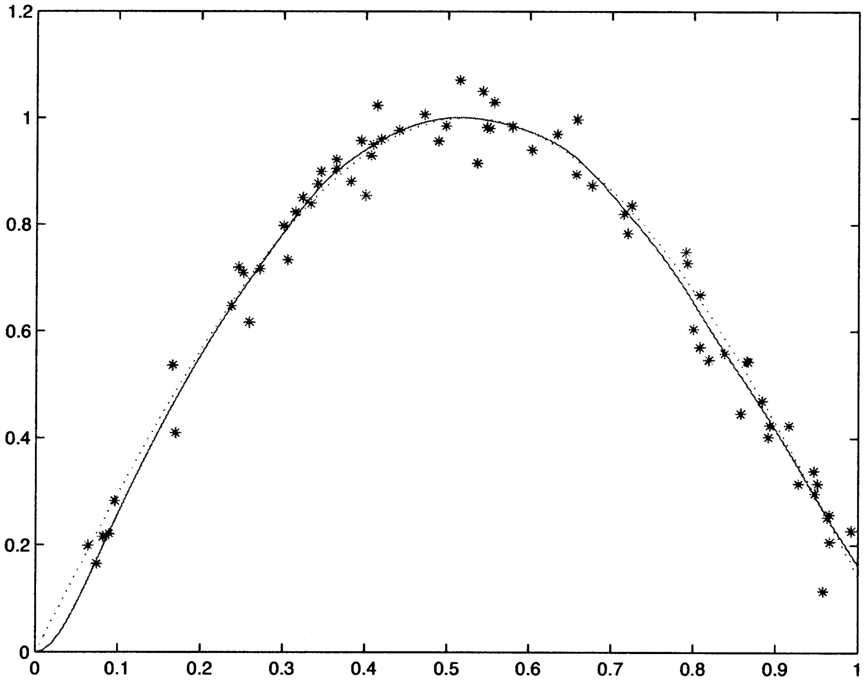

FIGURE 3. Tracking with 70 points using cubic splines.

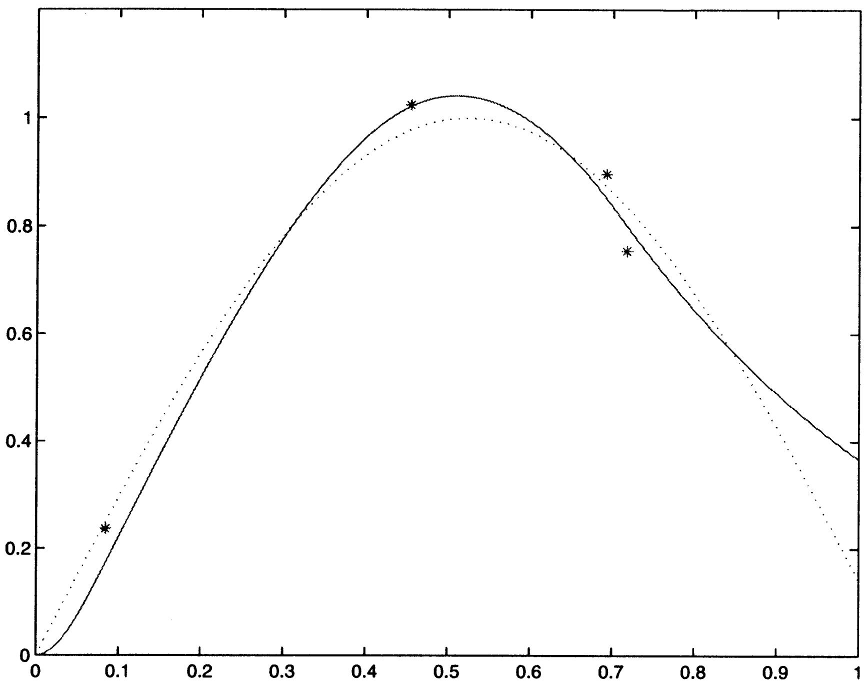

FIGURE 4. Tracking with four points using exponential splines.

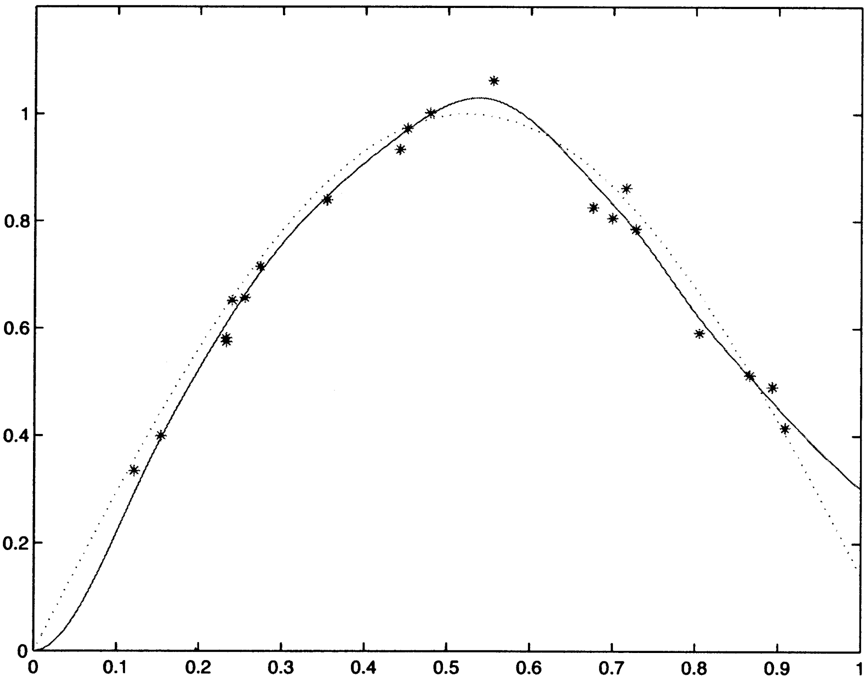

FIGURE 5. Tracking with 20 points using exponential splines.

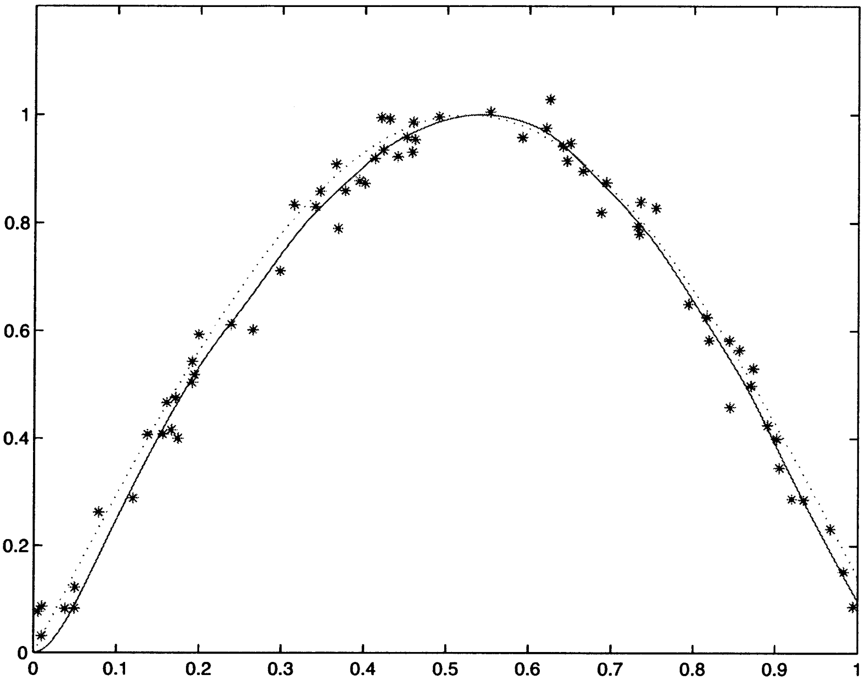

FIGURE 6. Tracking with 70 points using exponential splines.

References

[1] N. Agwu and C. Martin, Optimal Control of Dynamic Systems: Application to Spline Approximations. *Appl. Math. Comput.* **97** (1998) 99-138.

[2] M. Camarinha, P. Crouch and F. Silva–Leite, Splines of Class C^k on Non-Euclidean Spaces. *IMA J. Math. Control Inform.* **12** (1995) 399-410

[3] P. Crouch and J.W. Jackson, Dynamic Interpolation for Linear Systems, in *Proc. of the 29th. IEEE Conference on Decision and Control*. Hawaii (1990) 2312-2314

[4] P. Crouch, G. Kun and F. Silva–Leite, Generalization of Spline Curves on the Sphere: A Numerical Comparison, in *Proc. CONTROLO'98, 3rd Portuguese Conference on Automatic control*. Coimbra, Portugal (1998).

[5] P. Crouch and F. Silva–Leite, The Dynamical Interpolation Problem: On Riemannian Manifolds, Lie Groups and Symmetric Spaces. *J. Dynam. Control Systems* **1** (1995) 177-202.

[6] M. Egerstedt and C. Martin, Optimal Trajectory Planning and Smoothing Splines. *Automatica* **37** (2001).

[7] M. Egerstedt and C. Martin, Monotone Smoothing Splines, in *Proc. of MTNS*. Perpignan, France (2000).

[8] D. Nychka, Splines as Local Smoothers. *Ann. Statist.* **23** (1995) 1175-1197.

[9] C. Martin, M. Egerstedt and S. Sun, Optimal Control, Statistics and Path Planning. *Math. Comput. Modeling* **33** (2001) 237-253.

[10] R.C. Rodrigues, F. Silva–Leite and C. Simões, Generalized Splines and Optimal Control, in *Proc. ECC'99*. Karlsruhe, Germany (1999).

[11] S. Sun, M. Egerstedt and C. Martin, Control Theoretic Smoothing Splines. *IEEE Trans. Automat. Control* **45** (2000) 2271-2279.

[12] G. Wahba, *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia (1990).

[13] E.J. Wegman and I.W. Wright, Splines in Statistics. *J. Amer. Statist. Assoc.* **78** (1983).

[14] Z. Zhang, J. Tomlinson and C. Martin, Splines and Linear Control Theory. *Acta Math. Appl.* **49** (1997) 1-34.