OPTIMAL SAMPLE TIME SELECTIONS FOR INTERPOLATION AND SMOOTHING

Florent Delmotte * **Magnus Egerstedt** *
Clyde Martin **

* {florent,magnus}@ece.gatech.edu
Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA 30332, USA
** martin@math.ttu.edu
Mathematics and Statistics,
Texas Tech University, Lubbock, TX 79409, USA

Abstract: A solution is presented to the problem of selecting sample points in an optimal fashion. These points are used for interpolation and smoothing procedures, and, in particular, we derive necessary optimality conditions for the sample points. An example is presented concerning generalized smoothing splines that illustrate the generality as well as the numerical feasibility of the proposed approach. *Copyright©2005 IFAC.*

Keywords: Smoothing, Interpolation, Optimal control, Switching times, Splines.

1. INTRODUCTION

In this paper, we consider the problem of selecting the data points in an optimal fashion for interpolating or smoothing procedures. In particular, we show how the solution to the interpolation problem can be related to trajectories from generated switched, autonomous dynamical systems, that explicitly depend on the sample times. Optimal timing control, based on classic variational techniques, will then be employed in order to find locally optimal sample times.

Note that the only available results for similar problems are given by the Tschebyscheff polynomials. Given a function $h(t) \in C^{N-1}(t_0, t_f)$, the unique polynomial P_{N-1} that interpolates the data points $h(t_1), \ldots, h(t_N)$ satisfies

$$|h(t) – P_{N-1}(t)| \leq \max_{t_0 \leq \xi \leq t_f} |h^N(\xi)| \max_{t_0 \leq \xi \leq t_f} \frac{\prod_{i=1}^{N} |\xi - t_i|}{N!}$$
$$\triangleq H(t_1, \ldots, t_N)$$

as shown in (Davis, 1975). Moreover, the solution to the problem $\min_{t_1,\ldots,t_N} H(t_1, \ldots, t_N)$ is given by the Tschebyscheff polynomials. However, this result only holds for exact polynomial interpolation. Since we are interested in solving a more general problem, with general curve and cost types, the Tschebyscheff polynomials will not provide much assistance.

The outline of this paper is as follows: in Section 2, we discuss the connection between interpolation (or smoothing) and optimal control. This is followed by a derivation of necessary optimality conditions on the sample times, together with the presentation of a gradient-based numerical algorithm, in Section 3. The paper concludes with an example in Section 4, in which optimal sample times are generated in the case of generalized smoothing splines.

2. SAMPLING, INTERPOLATION, AND SMOOTHING

The connection between data interpolation (and smoothing) and optimal control is a well-studied subject. For example, (Mangasarin *et al.*, 1969; Schumaker, 1981) showed how to relate a number of polynomial interpolation procedures, including the classic cubic splines, to certain optimal control problems. This line of thought was continued by (Wahba, 1990) in the field of statistical data smoothing , showing how polynomial smoothing splines resulted from the minimization of

$$\min \int_{t_0}^{t_f} \frac{d^k p(t)}{dt^k} dt + \sum_{i=1}^{N} (p(\tau_i) - \xi_i)^2 \omega_i, \qquad (1)$$

given $k \geq 0$, $\omega_i > 0, i = 1, \ldots, N$. Here, the resulting polynomial is given by $p(t) \in \Re$, and it should be noted that if $k = 2$, the resulting polynomial $p(t)$ is a smoothing cubic spline, while $k = 3$ gives a fourth order polynomial spline, and so on. In the expression above, the data points are given by $(\tau_1, \xi_1), \ldots, (\tau_N, \xi_N)$, with $t_0 < \tau_1 < \cdots < \tau_N < t_f$, where the measurements $\xi_i \in \Re$ are presumably obtained from $\xi_i = h(\tau_i) + \epsilon_i$, given some underlying curve $h(t)$, corrupted by measurement noise ϵ_i, $i = 1, \ldots, N$.

This view of curve fitting as an optimal control problem was expanded in (Martin *et al.*, 1997), where generalized interpolating splines were constructed through optimal control. In particular, the underlying dynamics was given by a controllable and observable, single-input single-output linear control system

$$\begin{aligned} &\dot{x} = Ax + Bu, \ x \in \Re^n, \ u \in \Re \\ &y = Cx, \ y \in \Re. \end{aligned} \qquad (2)$$

The resulting curve $y(t)$ would constitute a polynomial, trigonometric, exponential, or mixed spline, depending on the spectrum of A. This approach was moreover taken further in (Sun *et al.*, 2000), where generalized smoothing splines were obtained from

$$\min_{u \in L_2[t_0, t_f]} \int_{t_0}^{t_f} u^2(t) dt + \sum_{i=1}^{N} \omega_i (y(\tau_i) - \xi_i)^2, \qquad (3)$$

given the underlying system in Equation 2.

This method of using smoothing generalized splines has for instance been applied by (Kano *et al.*, 1993) to the production of calligraphic Japanese characters . The idea is to let the data points encode the position of the paint brush as well as the thickness of the stroke at strategically selected points, with quite remarkable results. However, it is not, as of yet, clear how exactly these points should be selected. In this paper we address this problem of optimal sample point selection through optimal timing control.

From a general point-of-view, we can assume that the underlying system dynamics is given by $\dot{x} = F(x, u)$, where $x \in \Re^n, u \in \Re^q$, and F is smooth. Independent of interpolating or smoothing procedure, the resulting control law will in general depend on time t as well as the sample times $\tau = (\tau_1, \ldots, \tau_N)^T \in \Re^N$. Moreover, u will not be smooth (or even continuous) at the sample times, while it will be smooth for all other times. We can thus let the resulting optimal control law be given by

$$u(t, \tau) = G_i(t, \tau), \quad \forall t \in [\tau_{i-1}, \tau_i),$$

where $i = 1, \ldots, N+1$, $\tau_0 = t_0$ and $\tau_{N+1} = t_f$. In other words, the now autonomous yet switched system is given by

$$\dot{x} = F(x, G_i(t, \tau)) \triangleq f_i(x, t, \tau), \quad \forall t \in [\tau_{i-1}, \tau_i).$$

Moreover, if we assume that the data points are generated from an underlying curve $h(t) \in \Re$, and if we let the output from the dynamical system be $y(t) = g(x(t)) \in \Re$, we can define $L(x(t), t)$ as

$$L(x(t), t) = (g(x(t)) - h(t))^2,$$

and try to minimize the following cost

$$\min_{\tau} J(\tau) = \min_{\tau} \int_{t_0}^{t_f} L(x(t), t) dt,$$

subject to

$$\begin{aligned} &\dot{x} = f_i(x, t, \tau), \ t \in [\tau_{i-1}, \tau_i), \ i = 1, \ldots, N+1 \\ &x(t_0) = x_0. \end{aligned}$$

This general, optimal timing control problem will be addressed in the next section, followed by a discussion about how the results should be used when producing generalized, smoothing splines. It should, however, be noted already at this point that if the system dynamics $f_i(x, t, \tau)$ did not depend on the switching times τ explicitly, the optimal timing control problem would be solved (Egerstedt *et al.*, 2003; Shaikh *et al.*, 2002; Sussmann, 2000; Xu *et al.*, 2002). In fact, in (Egerstedt *et al.*, 2003), it was found that the gradient of the cost was given by

$$\frac{dJ}{d\tau_i} = \lambda(\tau_i)(f_i(x(\tau_i)) - f_{i+1}(x(\tau_i))),$$

given the costate equation

$$\begin{aligned} &\dot{\lambda} = -\frac{\partial L}{\partial x} - \lambda \frac{\partial f_i}{\partial x}, \ t \in [\tau_{i-1}, \tau_i) \\ &\lambda(t_f) = 0. \end{aligned}$$

Hence, the task undertaken in this paper is to extend this result to the case when f_i does in fact depend on τ.

3. OPTIMAL TIMING CONTROL

As before, consider the autonomous, switched dynamical system

$$\begin{aligned} \dot{x} &= f_i(x,t,\tau),\ t \in [\tau_{i-1}, \tau_i) \\ x(t_0) &= x_0, \end{aligned} \tag{4}$$

where $\{f_i(x,t,\tau)\}_{i=1}^{N+1}$ is a given sequence of smooth mappings from $\Re^n \times \Re \times \Re^N$ to $\Re^n$. Moreover, let L be a smooth function from $\Re^n \times \Re \to \Re$, and let as before the cost J be given by

$$J(\tau) = \int_{t_0}^{t_f} L(x(t),t)dt.$$

Note that J may very well be non-convex which means that only local optima can be expected to be obtained from gradient-based algorithms, which will be the case in this paper. The computation of the gradient of J with respect to the switching times is in fact the contribution in this section, and it will be based on the classic variational approach where the dynamical constraints are adjoined to the cost function via the costate variable λ.

3.1 Gradient Computation

We have

$$J_0 = \sum_{i=1}^{N+1}\left(\int_{\tau_{i-1}}^{\tau_i} (L(x(t),t) + \lambda(f_i(x(t),t,\tau) - \dot{x}))\, dt\right).$$

Now, by $i(t)$ we understand $i(t) = j$ when $t \in [\tau_{j-1}, \tau_j)$, and we consider the variation in J due to a perturbation in τ_k only. Hence, we replace $f_i(x,t,\tau)$ with $f_i(x,t,\tau_k)$ for the sake of notational ease. The variation is obtained through the small perturbation $\tau_k \to \tau_k + \epsilon\theta_k$, where $\epsilon \ll 1$, which results in the state variation $x \to x + \epsilon\eta$. The cost function for the perturbed system is

$$\begin{aligned} J_\epsilon = &\int_{t_0}^{\tau_k} \left[L(x+\epsilon\eta, t) + \lambda(f_{i(t)}(x+\epsilon\eta, t, \tau_k+\epsilon\theta_k) - \dot{x} - \epsilon\dot{\eta})\right] dt \\ &+ \int_{\tau_k}^{\tau_k+\epsilon\theta_k} [L(x+\epsilon\eta, t) + \lambda(f_k(x+\epsilon\eta, t, \tau_k+\epsilon\theta_k) - \dot{x} - \epsilon\dot{\eta})] dt \\ &+ \int_{\tau_k+\epsilon\theta_k}^{t_f} \left[L(x+\epsilon\eta, t) + \lambda(f_{i(t)}(x+\epsilon\eta, t, \tau_k+\epsilon\theta_k) - \dot{x} - \epsilon\dot{\eta})\right] dt \end{aligned}$$

A first order approximation of the continuously differentiable functions f_i and L gives

$$\begin{aligned} J_\epsilon - J_0 = &\int_{t_0}^{t_f}\left(\frac{\partial L}{\partial x}\epsilon\eta + \lambda\left(\frac{\partial f_{i(t)}}{\partial x}\epsilon\eta + \frac{\partial f_{i(t)}}{\partial \tau_k}\epsilon\theta_k - \epsilon\dot{\eta}\right)\right) dt \\ &+ \int_{\tau_k}^{\tau_k+\epsilon\theta_k} \lambda(f_k(x,t,\tau_k) - f_{k+1}(x,t,\tau_k))dt. \end{aligned}$$

By following the development in (Egerstedt *et al.*, 2003), we chose the continuous costate

$$\begin{aligned} \dot{\lambda} &= -\frac{\partial L}{\partial x} - \lambda\frac{\partial f_i}{\partial x},\ t \in [\tau_{i-1}, \tau_i) \\ \lambda(t_f) &= 0, \end{aligned} \tag{5}$$

which, through integration by parts, simplifies the variation $\delta J = (J_\epsilon - J_0)/\epsilon$ to the following

$$\delta J = \left(\int_{t_0}^{t_f} \lambda\frac{\partial f_{i(t)}}{\partial \tau_k}dt + \lambda(\tau_k)(f_k - f_{k+1})|_{t=\tau_k}\right)\theta_k.$$

We thus have that the k:th component of the gradient of J with respect to τ is given by

$$\frac{dJ}{d\tau_k} = \int_{t_0}^{t_f} \lambda\frac{\partial f_{i(t)}}{\partial \tau_k}dt + \lambda(\tau_k)(f_k - f_{k+1})|_{t=\tau_k} \tag{6}$$

which allows us to use gradient-based algorithms for selecting locally optimal sample points in our interpolation and smoothing problems, as we will see in the next section. But, we start by a brief discussion of the numerical aspects of this approach.

3.2 Gradient Descent

The reason why the formula derived in the previous paragraphs is particularly easy to work with is that it gives us access to a very straight-forward numerical algorithm.

For each iteration k, let $\tau(k)$ be the set of switching times, and compute the following:

(1) Compute $x(t)$ forward in time on $[t_0, t_f]$ by integrating Equation 4 from $x(t_0) = x_0$.
(2) Compute $\lambda(t)$ backward in time from t_f to t_0 by integrating Equation 5 from $\lambda(t_f) = 0$.
(3) Use Equation 6 to compute $dJ/d\tau(\tau(k))$.
(4) Update τ as

$$\tau(k+1) = \tau(k) - l(k)\left(\frac{dJ}{d\tau}(\tau(k))\right)^T,$$

where $l(k)$ is the stepsize, e.g. given by the Armijo algorithm (Armijo, 1966).
(5) Repeat.

Note that this method will only converge to a local minimum. But, as we will see, it can still give quite significant reductions in cost.

3.3 Example - Linear Approximations

In this example, we try to approximate a continuous function $h : [t_0, t_f] \to \mathbb{R}$ by a function x such that for $i = 1, ..., N+1$ and $\forall t \in [\tau_{i-1}, \tau_i)$

$$x(t) = h(\tau_{i-1}) + (t - \tau_{i-1})\frac{h(\tau_i) - h(\tau_{i-1})}{\tau_i - \tau_{i-1}},$$

where $\tau_0 = t_0$ and $\tau_{N+1} = t_f$. This autonomous switched system is simpler than the general case considered previously since the derivative function

$\dot{x}(t) = f_i(x(t), \tau)$ on $[\tau_{i-1}, \tau_i)$ here only depends on τ_{i-1} and τ_i, i.e.

$$\dot{x}(t) = f_i(\tau_{i-1}, \tau_i) = \frac{h(\tau_i) - h(\tau_{i-1})}{\tau_i - \tau_{i-1}} \quad \text{on } [\tau_{i-1}, \tau_i).$$

We now apply the developed algorithm to the problem of determining $\tau_1, ..., \tau_N$ in order to minimize the cost function

$$J(\tau)) = \int_{\tau_0}^{\tau_N+1} (h(t) - x(t))^2 dt.$$

Figure 1 shows how the algorithm converges. The following parameters were used:

$$\begin{cases} h(t) = 5\sin(\frac{2\pi t}{300}) + 3\sin(\frac{2\pi t}{100}) + \frac{t^2}{20000} - \frac{t}{50} \\ [t_0, t_f] = [0, 200], \quad N = 4 \quad \text{and} \quad l = 1. \end{cases}$$

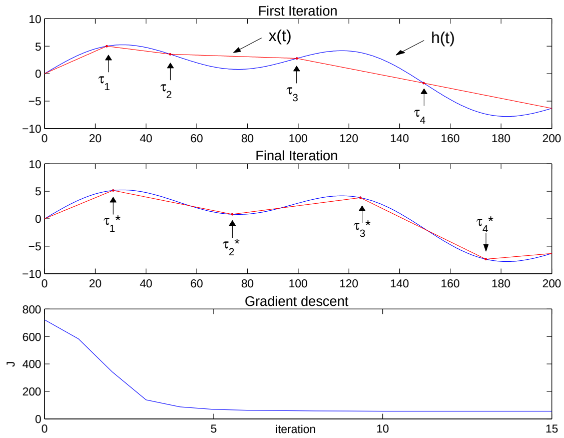

Fig. 1. Optimal linear approximation

The lowest figure in Figure 1 shows how fast the algorithm converges. The optimal solution is reached after a very few iterations, in spite of a "bad" initial guess and a constant step size l.

4. APPLICATION TO SMOOTHING SPLINES

As discussed previously, we can view the smoothing problem as a problem of finding the optimal control that drives the output of a given linear control system close to given data points. In particular, given the dynamics in Equation 2, the (unique) optimal solution to the problem in Equation 3 was in (Sun *et al.*, 2000) found to be

$$u(t) = \gamma(t)^T (I + \mathcal{W}\Gamma)^{-1} \mathcal{W}\xi,$$

where

$$\mathcal{W} = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_N \end{pmatrix}, \ \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{pmatrix} = \begin{pmatrix} h(\tau_1) \\ h(\tau_2) \\ \vdots \\ h(\tau_N) \end{pmatrix},$$

$$\gamma(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \vdots \\ \gamma_N(t) \end{pmatrix}, \ \gamma_i(t) = \begin{cases} Ce^{A(\tau_i - t)}B & \text{if } t \le \tau_i \\ 0 & \text{otherwise,} \end{cases}$$

and where the Grammian Γ is given by

$$\Gamma = \int_{t_0}^{t_f} \gamma(s)\gamma(s)^T ds \in \Re^{N \times N}.$$

Note that the definition of the basis functions $\gamma_i(t)$ imply that u may be discontinuous at τ_i. In fact, we could define a new set of basis functions

$$\zeta_i(t) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ Ce^{A(\tau_i - t)}B \\ Ce^{A(\tau_{i+1} - t)}B \\ \vdots \\ Ce^{A(\tau_N - t)}B \end{pmatrix}, \ t \in [\tau_{i-1}, \tau_i), \ i = 1, \ldots, N,$$

with $\zeta_{N+1} = 0$. Hence we have the new system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ &= Ax + B\zeta_i^T(t, \tau)(I + \mathcal{W}\Gamma(\tau))^{-1}\mathcal{W}\xi(\tau) \\ &\triangleq f_i(x, t, \tau), t \in [\tau_{i-1}, \tau_i) \\ y &= Cx \triangleq g(x), \end{aligned}$$

that is on the prescribed form.

Now, in order to be able to apply the gradient-based optimization methods, we need to obtain expressions for $\partial L / \partial x$, $\partial f_i / \partial x$, and $\partial f_i / \partial \tau_k$. If we, as before, let L be given by $(y(t) - h(t))^2$, we get for $i = 1, \ldots, N+1$

$$\begin{aligned} \frac{\partial f_i}{\partial x} &= A \\ \frac{\partial L}{\partial x} &= 2C(Cx(t) - h(t)) \\ \frac{\partial f_i}{\partial \tau_k} &= B\gamma(t)^T (I + \mathcal{W}\Gamma)^{-1} \mathcal{W}\delta\xi_k \\ &+ B\delta\gamma_k(t)^T \mathcal{W}(I + \mathcal{W}\Gamma)^{-1}\xi \\ &- B\gamma(t)^T (I + \mathcal{W}\Gamma)^{-1} \mathcal{W} \frac{\partial \Gamma}{\partial \tau_k} (I + \mathcal{W}\Gamma)^{-1} \mathcal{W}\xi, \end{aligned}$$

where

$$\delta\xi_k \triangleq \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{\partial h}{\partial t}(\tau_k) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow k\text{:th position}$$

$$\delta\gamma_k(t) \triangleq \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{\partial \gamma_k}{\partial \tau_k}(t) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow k\text{:th position}$$

$$\frac{\partial \gamma_k}{\partial \tau_k}(t) = \begin{cases} CAe^{A(\tau_k - t)}B & \text{if } t \leq \tau_k \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial \Gamma}{\partial \tau_k} = \int_{t_0}^{t_f} \left(\gamma(s)\delta\gamma_k(s)^T + \delta\gamma_k(s)\gamma(s)^T\right) ds.$$

Note that for this system, $(f_k - f_{k+1})|_{t=\tau_k} = 0$, which simplifies the derivative of the cost to

$$\frac{dJ}{d\tau_k} = \int_{t_0}^{t_f} \lambda \frac{\partial f_{i(t)}}{\partial \tau_k} dt.$$

4.1 Example

In this paragraph we apply this method to the system

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},\ B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, C = (1, 0, 0),$$

which gives the standard cubic smoothing spline.

Results from applying the gradient descent method using the Armijo step-size over 40 iterations is shown in Figures 2-4. In that example, the underlying curve was given by $h(t) = \sin(5t)$, and four sample times where selected with $\omega_i = 1$, $i = 1, \ldots, 4$.

5. CONCLUSIONS

In this paper we presented a method where variational techniques were employed in order to select optimal sample points for interpolation and smoothing applications. This method moreover resulted in a numerically straightforward algorithm that was put to use within the context of generalized smoothing splines. As such,, it produced results that go well beyond the previously known results on Tschebyscheff polynomials.

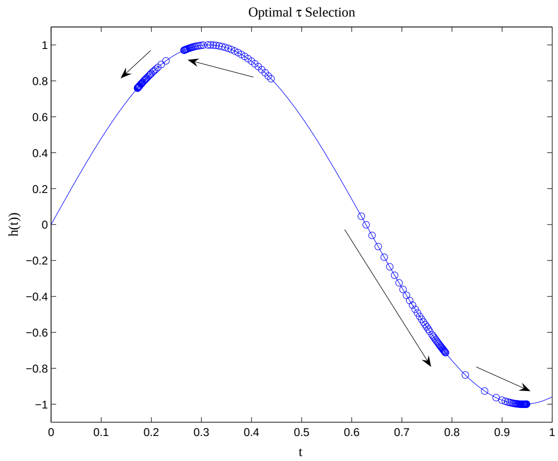

Fig. 2. The movements of the sample times is illustrated when creating smoothing splines for the underlying curve $h(t) = \sin(5t)$.

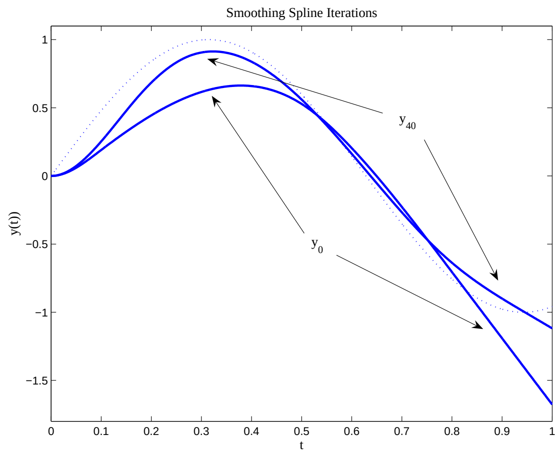

Fig. 3. The smoothing splines (solid) obtained at the first and the 40th iterations are depicted together with the underlying curve (dotted).

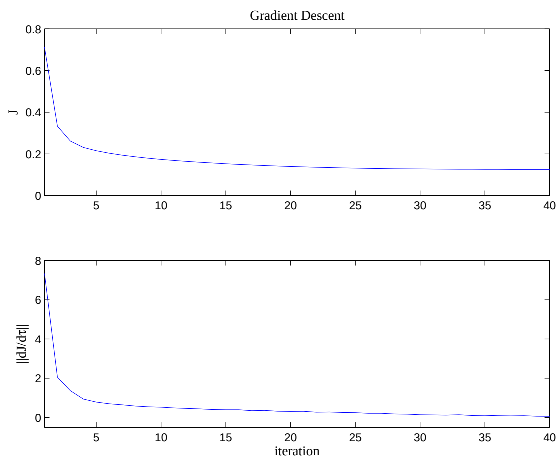

Fig. 4. The evolution of $J(\tau(k))$ together with $\|dJ/d\tau(\tau(k))\|$ is shown as a function of k, $k = 1, \ldots, 40$.

REFERENCES

Armijo L., (1966). Minimization of Functions Having Lipschitz Continuous First-Partial Derivatives. Pacific Journal of Mathematics, Vol. 16, ppm. 1-3.

Davis P.J., (1975). *Interpolation and Approximation*. Dover, New York, NY.

Egerstedt M., Y. Wardi, and F. Delmotte, (2003). Optimal Control of Switching Times in Switched Dynamical Systems. *Proc. 42nd Conference on Decision and Control*, Maui, HI, 2138-2143.

Kano H., M. Egerstedt, H. Nakata, and C.F. Martin, (1993). B-Splines and Control Theory. *Applied Mathematics and Computation*, Vo. 145, No. 2-3, pp. 265-288.

Mangasarin O.L. and L.L. Schumaker, (1969). Splines via Optimal Control. *Approximation with Special Emphasis on Spline Functions*, I.J. Schoenberg (Ed.), Academic Press, New York, NY.

Martin C.F., J. Tomlinson, and Z. Zhang, (1997). Splines and Linear Control Theory. *Acta Applicandae Mathematicae*, Vol. 92, No. 437, pp. 107-116.

Schumaker L.L., (1981). *Spline Functions: Basic Theory.* John Wiley and Sons, New York, NY.

Shaikh M.S. and P. Caines, (2002). On Trajectory Optimization for Hybrid Systems: Theory and ALgorithms for Fixed Schedules. *IEEE Conference on Decision and Control*, Las Vegas, NV.

Sun S., M. Egerstedt, and C.F. Martin, (2000). Control Theoretic Smoothing Splines. *IEEE Transactions on Automatic Control*, Vol. 45, No. 12, pp. 2271-2279.

Sussmann H.J., (2000). Set-Valued Differentials and the Hybrid Maximum Principle. *IEEE Conference on Decision and Control*, Vol. 1, pp. 558-563.

Wahba G., (1990). Spline Models for Observational Data. *Society for Industrial and Applied Mathematics*.

Xu X. and P. Antsaklis, (2002). Optimal Control of Switched Autonomous Systems. *IEEE Conference on Decision and Control*, Las Vegas, NV.