

L^2 AND L^1 CONTROL THEORETIC SMOOTHING SPLINES

Alaeddin Malek^{1 §}, Mohammad Asadollahi²

^{1,2}Department of Applied Mathematics

Faculty of Mathematical Sciences

Tarbiat Modares University

P.O. Box 14115-134, Tehran, IRAN

Abstract: L^1 and L^2 control theoretic splines are effective for Gaussian noise in data since estimations are based on L^1 and L^2 optimization. Here, it is shown that the result is not robust against outliers for L^2 control theoretic splines. Numerical simulations for both $y(t)$ and dy/dt under Gaussian and Laplacian noise are given. It is shown that for meaningful sampling data (number of data more than 75) the L^1 control theoretic spline has better performance than L^2 control theoretic spline. Numerical results and graphs for minimum, maximum and mean errors are given.

AMS Subject Classification: 49M05, 37M10

Key Words: optimal control, smoothing splines, linear control, convex optimization, control theoretic splines

1. Introduction

In the recent years in order to approximate noisy data values, researchers in mathematics and statistics have used smoothing splines. They establish a method of fitting smooth curves determined by noisy data values, usually Gaussian or Laplacian assumption for noise. The smoothing spline minimized residual (sum of squared errors) plus an effort:

Received: July 16, 2016

© 2016 Academic Publications

[§]Correspondence author

$$J(u) := \lambda \int_0^T u^2(t) dt + \sum_{i=1}^m w_i (y(t_i) - y_i)^2, \quad (*)$$

where the data y_i are noisy data, w_i and $y(t_i)$ are weights and exact data values at time t_i , $i = 1, 2, \dots, m$, for $u(t) \in L^1[0, T]$ or $u(t) \in L^2[0, T]$ one might calculate the output $y_{L^1}(t)$, $y_{L^2}(t)$ from the input to a linear single input single output (SISO) system. Comparing these two different output curves based on the $L^1[0, T]$ and $L^2[0, T]$ norms is the motivation of this paper. In this paper, authors compared both $y_{L^1}(t)$, $y_{L^2}(t)$ and $dy_{L^1}(t)/dt$, $dy_{L^2}(t)/dt$ in order to recognize the effectuality of the L^2 and L^1 control theoretic smoothing splines when one face by Gaussian or Laplacian noise. In the year 1964 Schoenberg [16] used smoothing splines for the approximation of noisy data values. Reformulating this problem to an optimal control problem (*) for SISO systems was introduced by Egerstedt and Martin in year 2001 [3]. L^1 control theoretic smoothing spline was introduced by Nagahara and Martin [7].

The control theoretic splines are useful in the study of trajectory planning, mobile robots in, contour modeling of images and image processing [2]. The reminder of this article is organized as follows: In Section 2 we give the problem formulation, in Section 3 – L^2 control theoretic smoothing splines, in Section 4 – L^1 control theoretic smoothing splines, and numerical examples are included in Section 5, and Section 6 draws conclusions.

2. Problem Formulation

Consider a linear time-invariant, single-input or single-output system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \quad x(0) = x_0, \quad x_0 \in \mathbb{R}^n, \\ y(t) &= c^T x(t), \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}$ is the control input, $y \in \mathbb{R}$ is the plant output, $A \in \mathbb{R}^{n \times n}$ and $b, c \in \mathbb{R}^n$. Let us define $(t_1, y_1), (t_2, y_2), \dots, (t_m, y_m)$, where $y_1, y_2, \dots, y_m \in \mathbb{R}$ are noisy sample data on the time instants $0 < t_1 < t_2 < \dots < t_m = T$.

Assumption 1: (A, b) is controllable and (c^T, A) is observable.

We now formulate a cost function of the form

$$J(u) := \lambda \int_0^T u^2(t) dt + \sum_{i=1}^m w_i (y(t_i) - y_i)^2, \quad (2)$$

where $\lambda > 0$ is the regularization parameter that specifies the tradeoff between the smoothness of control $u(t)$ and the minimization of the squared empirical risk in the second term of (2), in which $w_i > 0$ is a weight for i -th squared loss $(y(t_i) - y_i)^2$.

Our goal is to minimize the quadratic functional J over the Hilbert space of square integrable functions on the interval $[0, T]$ subject to the affine constraint

$$y(t) = c^T e^{At} x_0 + \int_0^t c^T e^{A(t-s)} b u(s) ds. \quad (3)$$

Let $\beta_i = c^T e^{At_i} x_0$, $\beta_i - y_i = \gamma_i$ and

$$g_{t_i}(t) = \begin{cases} c^T e^{A(t_i-t)} b & t \leq t_i \\ 0 & t > t_i \end{cases}, \quad (4)$$

where the t_i 's are the interpolation times.

We define a set of linear functionals as

$$L_{t_i}(u) = \int_0^T g_{t_i}(t) u(t) dt. \quad (5)$$

Now, (3)-(5) yield

$$y(t_i) = \beta_i + \int_0^T g_{t_i}(t) u(t) dt = \beta_i + L_{t_i}(u).$$

Thus (2) can be rewritten as:

$$J(u) = \lambda \int_0^T u^2(t) dt + \sum_{i=1}^m w_i (L_{t_i}(u) + \gamma_i)^2. \quad (6)$$

For $\varepsilon \in \mathbb{R}$ and h as an arbitrarily function in $L^2[0, T]$, we calculate the Frechet derivative in the form

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (J(u + \varepsilon h) - J(u)) \\ &= \sum_{i=1}^m 2w_i L_{t_i}(h) (L_{t_i}(u) + \gamma_i) + 2\lambda \int_0^T h(t) u(t) dt \\ &= 2 \int_0^T \left[\sum_{i=1}^m w_i g_{t_i}(t) (L_{t_i}(u) + \gamma_i) + \lambda u(t) \right] h(t) dt. \end{aligned} \quad (7)$$

To ensure that u is a minimum, a necessary condition requires that (7) vanishes, i.e.

$$K(u) = \sum_{i=1}^m w_i g_{t_i}(t) (L_{t_i}(u) + \gamma_i) + \lambda u(t) = 0. \quad (8)$$

In the following rewritten operator

$$K(u) = \int_0^T \left(\sum_{i=1}^m w_i g_{t_i}(t) g_{t_i}(s) \right) u(s) ds + \lambda u(t), \quad (9)$$

we show that K is one-to-one and onto- operator.

Lemma 1. *The set of functions $\{g_{t_i}(t) : i = 1, 2, \dots, m\}$ in (4) are linearly independent.*

Proof. The proof is obvious and relies on the fact that the different g_{t_i} 's vanish at instants times t_1, t_2, \dots, t_m . \square

Lemma 2. ([6]) *The operator K is one-to-one for all choices of $w_i > 0$ and $\lambda > 0$.*

Proof. Suppose $K(u_0) = 0$, which would imply that

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

hence for $L_{t_i}(u_0) = a_i$ we have

$$\sum_{i=1}^m w_i g_{t_i}(t) a_i + \lambda u_0(t) = 0.$$

This implies that any solution u_0 of $K(u_0) = 0$ is in the span of the set $\{g_{t_i}(t) : i = 1, 2, \dots, m\}$. Now consider a solution of the form

$$u_0(t) = \sum_{i=1}^m \theta_i g_{t_i}(t) \quad (10)$$

and from (9) evaluate $K(u_0)$ to derive

$$K(u_0) = \sum_{i=1}^m w_i g_{t_i}(t) L_{t_i}(u) \left(\sum_{j=1}^m \theta_j g_j(t) \right) + \lambda \sum_{i=1}^m \theta_i g_i(t) = 0.$$

Thus for each i ,

$$w_i \sum_{j=1}^m L_{t_i}(g_j) \theta_j + \lambda \theta_i = 0.$$

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

$$(WG + \lambda I)\theta = 0,$$

where W is the diagonal matrix of the weights w_i and $G = [g_{ij}]$ is the Grammian with $g_{ij} = L_{t_i}(g_j)$. Now consider the matrix $WG + \lambda I$ and multiply it on the left by W^{-1} , and consider the scalar

$$x^T(G + \lambda W^{-1})x = x^T G x + \lambda x^T W^{-1} x > 0$$

since both terms are positive. Thus for positive weights and positive λ the only solution is $\theta = 0$. \square

Remark 1. The matrix $G = [g_{ij}] \in \mathbb{R}^{m \times m}$ is the Grammian defined by

$$\begin{aligned} g_{ij} &= \langle g(t_i - \cdot), g(t_j - \cdot) \rangle \\ &= \int_0^T g(t_i - t)g(t_j - t)dt \quad , \quad i, j = 1, 2, \dots, m. \end{aligned} \tag{11}$$

Lemma 3. ([6]) For $w_i > 0$ and $\lambda > 0$ the operator K is onto.

Proof. Suppose K is not onto. Then, for all u there exists a nonzero function f such that $\int_0^T f(t)K(u)(t)dt = 0$. We have, after some manipulations, that

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

and hence

$$\int_0^T \sum_{i=0}^m w_i g_{t_i}(t)g_{t_i}(s)f(t)dt + \lambda f(s) = 0.$$

The only solution of this equation is $f = 0$ and hence K is onto. \square

Lemma 4. ([6]) $J(u)$ is convex in u .

Proof. Since J is closed, quadratic function in u , convexity follows immediately since $\lambda > 0$, $W > 0$.

From the convexity, the results on existence and uniqueness now follow from standard infinite dimensional optimization. \square

We have thus proved the following proposition.

Proposition 1. *The following functional has a unique minimum*

$$J(u) = \sum_{i=1}^m w_i (L_{t_i}(u) + \gamma_i)^2 + \lambda \int_0^T u^2(t) dt.$$

3. L^2 Control Theoretic Smoothing Splines

Problem 1. The problem of L^2 control theoretic smoothing spline is formulated as follows:

Find control $u(t)$ that minimize the cost $J(u)$ in (2) subject to the state-space equation in (1). The Optimal control $u = u^*$ that minimizes $J(u)$ is given by (10) of the form

$$u^*(t) = \sum_{i=1}^m \theta_i g_{t_i}(t), \quad (12)$$

where $g_{t_i}(t)$ is defined by (4).

The optimal coefficients $\theta_1^*, \theta_2^*, \dots, \theta_N^*$ are as:

$$\theta^* = [\theta_1^*, \theta_2^*, \dots, \theta_m^*]^T = (WG + \lambda I)^{-1} y, \quad (13)$$

where

$$y := [y_1, y_2, \dots, y_m]^T. \quad (14)$$

An advantage of the L^2 control theoretic smoothing spline is that the optimal control can be computed offline via equation (13). However, the formula indicates that if the data size N is large, so is the number of base functions in $u^*(t)$, as it is shown in (12). This becomes a drawback if we have only a small memory or simple actuator for drawing a curve with the optimal control $u^*(t)$.

Another drawback is that the L^2 spline is not robust at all against outliers, as reported in [7], since the squared empirical risk in (2) is measured by L^2 norm.

This is based on the assumption that the additive noise is Gaussian. However, there may exist outliers in data, which may be ignored under the Gaussian assumption of noise, the regression may be very sensitive. For overcome to these drawbacks we adopt L^1 optimality for the design of spline.

4. L^1 Control Theoretic Smoothing Splines

Before formulating the design problem of L^1 spline, we prove the following lemma:

Lemma 5. ([6]) Assume that control $u(t)$ is given by

$$u(t) = \sum_{i=1}^m \theta_i g_{t_i}(t) \quad (15)$$

for some $\theta_i \in \mathbb{R}$, $i = 1, 2, \dots, m$. Then we have

$$y(t) = \sum_{i=1}^m \theta_i \langle g(t - \cdot), g(t_i - \cdot) \rangle, \quad t \in [0, T]. \quad (16)$$

In particular, for $j = 1, 2, \dots, m$, we have

$$y(t_j) = \sum_{i=1}^m \theta_i G_{ij}. \quad (17)$$

Proof. If $u(t) = 0$ for $t < 0$, then the solution of (1) is given by

$$y(t_j) = \int_0^t c^T e^{A(t-s)} b u(s) ds = \int_0^T g_{t_i}(s) u(s) ds = \langle g(t - \cdot), u \rangle.$$

Substituting (15) into the above equation, gives (16). Then, from the definition of G_{ij} in (11), we immediately have (17).

By Lemma 5, the error $y(t_i) - y_i$ is given by

$$y(t_i) - y_i = \sum_{i=1}^m \theta_i G_{ij} - y_i, \quad j = 1, 2, \dots, m,$$

or equivalently,

$$\begin{bmatrix} y(t_1) - y_1 \\ \vdots \\ y(t_m) - y_m \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1m} \\ G_{21} & G_{22} & \cdots & G_{2m} \\ \vdots & \vdots & & \vdots \\ G_{m1} & G_{m2} & \cdots & G_{mm} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}. \quad (18)$$

□

Based on this, we consider the following optimization problem:

Problem 2. (The problem of L^1 control theoretic smoothing spline: L^1 -optimal spline coefficients):

Find $\theta \in \mathbb{R}^m$ where $\theta := [\theta_1, \theta_2, \dots, \theta_m]^T$ that minimizes

$$J_p(\theta) = \eta \|\theta\|_1 + \|W(G\theta - y)\|_p^p, \quad (19)$$

where $\eta > 0$ and $p \in \{1, 2\}$, [11].

The regularization term, $\|\theta\|_1$, is for sparsity of coefficients $\theta_1, \theta_2, \dots, \theta_m$, as used in regression LASSO [12]. Also, small $\|\theta\|_1$ leads to small L^1 norm of control u since from (15) we have

$$\int_0^T |u(t)| dt \leq C \|\theta\|_1,$$

for some constant $C > 0$. On the other hand, the empirical risk term, $\|W(G\theta - y)\|_p^p$, is for the fidelity to the data. For $p = 1$, additive noise is assumed to be Laplacian, a heavy tailed distribution, to take outliers into account, while $p = 2$ is related to Gaussian noise. In each case, cost function $J_p(\theta)$ is convex in θ . Unlike L^2 spline, the solution to the optimization in Problem 2 cannot be represented in a closed form. However, by using a numerical optimization algorithm we can obtain an approximated solution within a reasonable time. For example for $p = 1$, there is no algorithm achieving such a rate, but the optimization is still convex and we can use an efficient convex optimization software, such as cvx on MATLAB [13].

5. Numerical Example

In this section, by some numerical simulations we show the effectiveness of the proposed L^1 control theoretic smoothing spline against L^2 control theoretic smoothing spline. In what follows let us assume that the dynamical system $P(s)$ is given by transfer function $P(s) = 1/s^3$, state-space matrices for $P(s)$ are given by

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad c = [0 \quad 0 \quad 1].$$

Mean and variance of the Laplacian noise and the original curve are given in Table 1 for four different examples.

Example	$y_{orig}(t)$	Noise		Figures
		Mean	Variance	
1	$1 + \cos(3t)$	0	1	1&2
2	$1 + \cos(3t)$	0	1	3&4
3	$1 + \cos(3t)$	0	2	5&6
4	$\sin(2t)$	0	1	7&8

Table 1

This table shows that for $t_1 = 0.1$ in Example 1 data are sampled at rate 10 [Hz] ($t_i = 0.1 + 0.1(i - 1)$), while in Examples 2, 3 and 4, data samples are at rate 100 [Hz] ($t_i = 0.1 + 0.01(i - 1)$). In Example 1, (t_i, y_i) , $i = 1, 2, \dots, 51$, are the noisy sample data, while in Examples 2, 3 and 4, (t_i, y_i) , $i = 1, 2, \dots, 501$, are the noisy sample data. Note that noisy sample data are given by additive Laplacian noise with mean 0 and variance 1 or 2.

In this paper, the optimal coefficients θ_i for L^1 control theoretic smoothing spline with $p = 1$ (see (19)) is computed using convex optimization code. Note that $p = 1$ corresponds to the Laplacian noise, that is used in this paper. The design parameters is given by $\eta = 0.01$ and the weights are all equal and are fixed to 1.

First, for $y(t)$, L^2 control theoretic smoothing splines are compared with L^1 control theoretic smoothing splines (see Table 2 and Figures 1a–2d). Second, for dy/dt , L^1 and L^2 control theoretic smoothing splines are compared (see Table 3 and Figures 3a–4d).

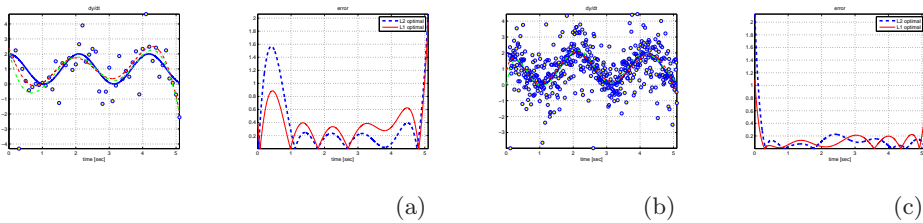


Figure 1: (a) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (b) Error between fitted curve by L^2 spline and fitted curve by L^1 spline, (c) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (d) Error between fitted curve by L^2 spline and fitted curve by L^1 spline.

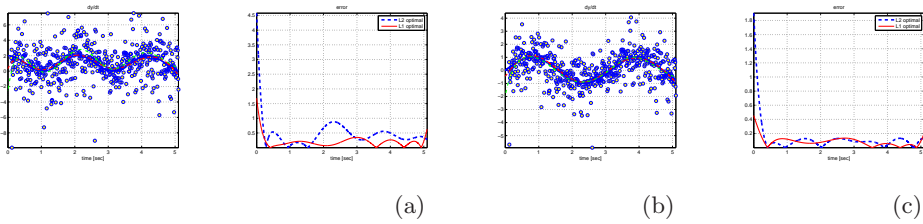


Figure 2: (a) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (b) Error between fitted curve by L^2 spline and fitted curve by L^1 spline, (c) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (d) Error between fitted curve by L^2 spline and fitted curve by L^1 spline.

Figures 1a, 1c, 2a and 2c illustrate the original curves y_{orig} and their fitted curves by L^1 and L^2 control smoothing splines for Examples 1, 2, 3 and 4, respectively. Figures 1b, 1d, 2b and 2d show the error between L^1 and L^2 control smoothing splines for Examples 1, 2, 3 and 4, respectively.

In Fig. 1a, it is shown that even for small number of sampling data 51, the reconstructed curves based on L^1 and L^2 control smoothing splines fits the original curves. Fig. 1b shows the error between original curve $1 + \cos(3t)$ and the fitted curves. In Table 2 one can observe that the fitted curves by L^1 control smoothing spline has less mean error in all of the cases however for small number of sampling data (less than 75) minimum error L^2 control smoothing spline is less than the minimum error for L^1 control smoothing spline, as it was expected.

In Figs. 1c and 1d for Example 2 in comparing with Figs. 2c and 2d for Example 4, show that the mean error of L^1 control smoothing spline is less than L^2 control smoothing spline, even if the original curves are different, i.e., they are chosen to be $1 + \cos(3t)$ and $\sin(2t)$, respectively (see Table 2). Several original functions are examined and to the knowledge of authors, the L^1 control smoothing spline gives better performance than L^2 control smoothing spline.

In Table 3, it is shown that for dy/dt , L^1 control smoothing spline gives better performance than L^2 control smoothing spline for large number of sampling data while Laplacian noise is used (see Figures 3 and 4).

	L^1 Errors			L^2 Errors		
	Min Error	Mean Error	Max Error	Min Error	Mean Error	Max Error
Ex 1	3.9267e-04	0.3759	1.9397	1.7250e-04	0.3760	2.0620
Ex 2	5.7627e-05	0.1252	1.0928	5.6152e-05	0.1394	2.1278
Ex 3	4.2735e-05	0.2012	1.6819	1.5821e-06	0.4693	4.5782
Ex 4	7.2212e-06	0.0849	0.4538	3.8729e-05	0.1231	1.9036

Table 2: L^1 and L^2 Errors for different Examples 1, 2, 3 and 4 for $y(t)$.

	L^1 Errors			L^2 Errors		
	Min Error	Mean Error	Max Error	Min Error	Mean Error	Max Error
Ex 1	3.7560e-04	0.8175	2.7660	2.9793e-04	0.7453	3.5864
Ex 2	1.3737e-05	0.0828	0.2453	3.3330e-05	0.1218	2.2062
Ex 3	1.7462e-05	0.1620	0.5202	6.0957e-07	0.2811	4.2969
Ex 4	8.0341e-06	0.0828	0.8083	9.6813e-06	0.1223	1.9350

Table 3: L^1 and L^2 Errors for different Examples 1, 2, 3 and 4 for dy/dt .

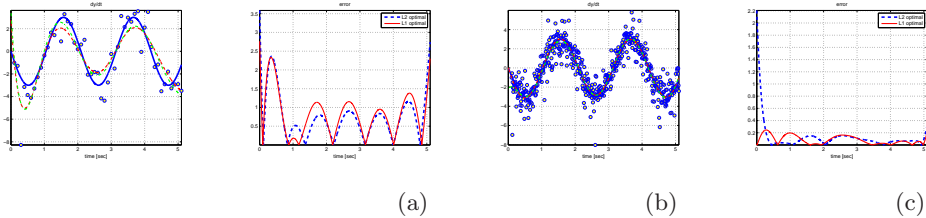


Figure 3: (a) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (b) Error between fitted curve by L^2 spline and fitted curve by L^1 spline, (c) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (d) Error between fitted curve by L^2 spline and fitted curve by L^1 spline.

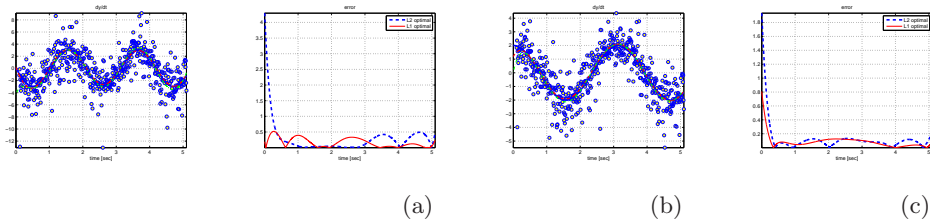


Figure 4: (a) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (b) Error between fitted curve by L^2 spline and fitted curve by L^1 spline, (c) Sampled data (circles), original curve (solid line), fitted curve by L^2 (dash-dotted line), fitted curve by L^1 (dashed line), (d) Error between fitted curve by L^2 spline and fitted curve by L^1 spline.

References

- [1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004).
- [2] M. Egerstedt, C. Martin, *Control Theoretic Splines*, Princeton University Press (2010).
- [3] M. Egerstedt, C. Martin, Optimal trajectory planning and smoothing splines, *Automatica* (2001), 1057-1064.
- [4] M. Nagahara, C. Martin, Monotone smoothing splines using general linear systems, *Asian Journal of Control* (2013), 461-468.
- [5] D. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York (1969).
- [6] S. Sun, M. Egerstedt and C. Martin, Control theoretic smoothing splines, *Automatic Control, IEEE Trans.* (2010), 2271-2279.
- [7] M. Nagahara, C. Martin, L^1 -optimal splines for outlier rejection, *59th World Statistics Congress* (2013), 1137-1142.
- [8] G. Wahba, *Spline Models for Observational Data*, SIAM (1990).

- [9] Z. Zhang, J. Tomlinson and C. Martin, Splines and linear control theory, *Acta Math. Appl.*, **49** (1997), 1-34.
- [10] W. Rugh, *Linear Systems Theory*, Princeton Hall (1996).
- [11] M. Nagahara, C. Martin, L^1 -control theoretic smoothing splines, *Signal Processing. IEEE Trans.* (2012).
- [12] R. Tibshirani, Splines and Linear Control Theory, *J. R. Statist. Soc.*, **58** (1996), 267-288.
- [13] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, <http://cvxr.com/cvx> (2011).
- [14] M. Egerstedt, C. Martin, Optimal trajectory planning and smoothing splines, *Automatica*, **37**, No 7 (2001), 1057-1064.
- [15] H. Kano, M. Egerstedt, H. Fujioka, S. Takahashi and C. Martin, Periodic smoothing splines, *Automatica*, **44**, No 1 (2008), 185-192.
- [16] I.J. Schoenberg, Spline functions and the problem of graduation, *Proc. Natl. Acad. Sci. USA*, **52** (1964).

