L^1 Control Theoretic Smoothing Splines

Masaaki Nagahara, *Member, IEEE,* and Clyde F. Martin, *Fellow, IEEE*

***Abstract*—In this paper, we propose control theoretic smoothing splines with L^1 optimality for reducing the number of parameters that describes the fitted curve as well as removing outlier data. A control theoretic spline is a smoothing spline that is generated as an output of a given linear dynamical system. Conventional design requires exactly the same number of base functions as given data, and the result is not robust against outliers. To solve these problems, we propose to use L^1 optimality, that is, we use the L^1 norm for the regularization term and/or the empirical risk term. The optimization is described by a convex optimization, which can be efficiently solved via a numerical optimization software. A numerical example shows the effectiveness of the proposed method.**

***Index Terms*—Control theoretic splines, smoothing splines, L^1 optimization, convex optimization.**

I. Introduction

The spline has been widely used in signal processing, numerical computation, statistics, etc. In particular, the *smoothing spline* gives a smooth curve that has the best fit to given noisy data [1], [2]. The smoothness is achieved by limiting the L^2 norm of the m-th derivative of the curve as well as minimizing the squared error (or empirical risk) between data and the curve.

The *control theoretic smoothing spline* [3] is generalization of the smoothing spline using control theoretic ideas, by which the spline curve is determined by the output of a linear dynamical system. It is shown in [4] that control theoretic splines give a richer class of smoothing curves relative to polynomial curves. Fig. 1 illustrates the idea of the control theoretic spline; given a finite number of data, the robot modeled by a dynamical system with transfer function $P(s)$ is driven by a control input $u(t)$ and draws a smooth curve $y(t)$ that fits to the data. The problem of the control theoretic spline is to find control $u(t)$ that gives an expected motion of the robot, based on the model $P(s)$ and the data set. Furthermore, the control theoretic spline has been proved to be useful for trajectory planning in [5], mobile robots in [6], contour modeling of images in [7], probability distribution estimation in [8], to name a few. For more applications and a rather complete theory of control theoretic splines, see [4].

Conventional design of control theoretic splines is based on L^2 optimization [3], and has two main drawbacks. One is that we need the same number of parameters as the data to represent the fitted curve. If the data set is big, then the number of parameters becomes crucial when for example the actuator system of the robot (see Fig. 1) has just a small area of memory. The other drawback is that the spline is not robust against outliers in observed data. In other words, conventional control theoretic splines are sensitive to outliers. To overcome these drawbacks, we propose to use L^1 optimality in the design. For reduction of the number of parameters, we utilize the *sparsity-promoting property* of the L^1 norm regularization, also known as LASSO (least absolute shrinkage and selection operator) [9], [10]. For robustness against outliers, we adopt the L^1 norm for the empirical risk minimization [11], assuming that the noise is Laplacian, heavier-tailed distribution than Gaussian that is assumed in conventional studies[1]. The problem is then described in convex optimization, which can be efficiently solved by numerical computation software, e.g. `cvx` on MATLAB [14], [15]. For numerical computation, we implement the design procedure on MATLAB programs with `cvx`, access [16] to obtain the programs. Based on the programs, we show a numerical example that illustrates the effectiveness of the proposed method.

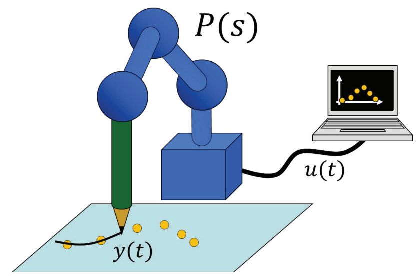

Fig. 1. Control theoretic spline as a robot $P(s)$ that draws a smooth curve $y(t)$ with a control input $u(t)$ based on given data.

The remainder of this article is organized as follows: Section II reviews the conventional L^2-optimal control theoretic spline and discusses drawbacks of the L^2 spline. Section III formulates the problem of the proposed L^1 spline to overcome drawbacks in the L^2 spline, and show a procedure to the solution. A numerical example is included in Section IV. Section V draws conclusions.

II. L^2 Control Theoretic Smoothing Splines

Consider a linear dynamical system P defined by

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= A\boldsymbol{x}(t) + \boldsymbol{b}u(t), \quad \boldsymbol{x}(0) = \boldsymbol{0} \in \mathbb{R}^n, \\ y(t) &= \boldsymbol{c}^\top \boldsymbol{x}(t), \quad t \geq 0 \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$. We assume $(A, \boldsymbol{b})$ is controllable and $(\boldsymbol{c}^\top, A)$ is observable[2]. For this system, suppose that a

[1]The idea of using a heavier-tailed loss function for control theoretic smoothing splines was first proposed in [12], [13].

[2]For controllability and observability of a linear system, see e.g. [17, Chap. 9].

M. Nagahara is with Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan; email: nagahara@ieee.org (corresponding author)

C. F. Martin is with is with Department of Mathematics & Statistics, Texas Tech University, Texas, USA; email: clyde.f.martin@ttu.edu

data set

$$\mathcal{D} := \{(t_1, y_1), (t_2, y_1), \dots (t_N, y_N)\}$$

is given, where $t_1, \dots, t_N$ are sampling instants which satisfy $0 < t_1 < t_2 < \cdots < t_N =: T$, and $y_1, \dots, y_N$ are noisy sampled data of the output of (1). The objective here is to find control $u(t)$, $t \in [0, T]$ for the dynamical system (1) such that $y(t_i) \approx y_i$ for $i = 1, \dots, N$. For this purpose, the following *quadratic* cost function has been introduced in [3]:

$$J(u) := \lambda \int_0^T |u(t)|^2 \mathrm{d}t + \sum_{i=1}^N w_i |y(t_i) - y_i|^2, \tag{2}$$

where $\lambda > 0$ is the regularization parameter that specifies the tradeoff between the smoothness of control $u(t)$ defined in the first term of (2) and the minimization of the squared empirical risk in the second term. Also, $w_i > 0$ is a weight for i-th squared loss $|y(t_i) - y_i|^2$. Then the problem of L^2 control theoretic smoothing spline is formulated as follows:

Problem 1 (L^2 control theoretic smoothing spline): Find control $u(t)$ that minimizes 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 [3], [4]

$$u^*(t) = \sum_{i=1}^N \theta_i^* g(t_i - t), \tag{3}$$

where $g(\cdot)$ is defined by

$$g(\tau) := \begin{cases} \boldsymbol{c}^\top e^{A\tau} \boldsymbol{b}, & \text{if } \tau \in [0, T], \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Note that $\boldsymbol{c}^\top e^{A\tau}\boldsymbol{b}$ in $g(\tau)$ is the impulse response of the dynamical system (1). The optimal coefficients $\theta_1^*, \dots, \theta_N^*$ are given by

$$\boldsymbol{\theta}^* := \left[\theta_1^*, \dots, \theta_N^*\right]^\top = (\lambda I + WG)^{-1} W \boldsymbol{y}, \tag{5}$$

where

$$W := \operatorname{diag}(w_1, \dots, w_N), \quad \boldsymbol{y} := [y_1, \dots, y_N]^\top. \tag{6}$$

The matrix $G = [G_{ij}] \in \mathbb{R}^{N \times N}$ in (5) is the Grammian defined by

$$\begin{aligned} G_{ij} &= \langle g(t_i - \cdot), g(t_j - \cdot) \rangle \\ &= \int_0^T g(t_i - t) g(t_j - t) \mathrm{d}t, \quad i, j = 1, \dots, N. \end{aligned} \tag{7}$$

An advantage of the L^2 control theoretic smoothing spline is that the optimal control can be computed offline via equation (5). However, the formula indicates that if the data size N is large, so is the number of base functions in $u^*(t)$, as shown in (3). This becomes a drawback if we have only a small memory or a 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 [13], since the squared empirical risk in (2) assumes that the additive noise is Gaussian. To solve these problems, we adopt L^1 optimality for the design of spline.

III. L^1 Control Theoretic Smoothing Splines

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

Lemma 1: Assume that control $u(t)$ is given by

$$u(t) = \sum_{i=1}^N \theta_i g(t_i - t), \tag{8}$$

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

$$y(t) = \sum_{i=1}^N \theta_i \langle g(t - \cdot), g(t_i - \cdot) \rangle, \quad t \in [0, T]. \tag{9}$$

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

$$y(t_j) = \sum_{i=1}^N \theta_i G_{ij}. \tag{10}$$

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

$$\begin{aligned} y(t) = \int_0^t \boldsymbol{c}^\top e^{A(t-\tau)} \boldsymbol{b} u(\tau) \mathrm{d}\tau &= \int_0^T g(t - \tau) u(\tau) \mathrm{d}\tau \\ &= \langle g(t - \cdot), u \rangle \end{aligned}$$

Substituting (8) into the above equation gives (9). Then, from the definition of G_{ij} in (7), we immediately have (10). ■

By this lemma, the error $y(t_i) - y_i$ is given by

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

or equivalently

$$\begin{bmatrix} y(t_1) - y_1 \\ \vdots \\ y(t_N) - y_N \end{bmatrix} = G\boldsymbol{\theta} - \boldsymbol{y}, \tag{11}$$

where $\boldsymbol{\theta} := [\theta_1, \dots, \theta_N]^\top$ and $\boldsymbol{y}$ is given in (6). Based on this, we consider the following optimization problem:

Problem 2 (L^1-optimal spline coefficients): Find $\boldsymbol{\theta} \in \mathbb{R}^N$ that minimizes

$$J_p(\boldsymbol{\theta}) := \eta \|\boldsymbol{\theta}\|_1 + \|W(G\boldsymbol{\theta} - \boldsymbol{y})\|_p^p, \tag{12}$$

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

The regularization term, $\|\boldsymbol{\theta}\|_1$, is for sparsity of coefficients $\theta_1, \dots, \theta_N$, as used in LASSO [9], [10]. Also, small $\|\boldsymbol{\theta}\|_1$ leads to small L^1 norm of control u since from (8) we have

$$\int_0^T |u(t)| \mathrm{d}t \le C \|\boldsymbol{\theta}\|_1,$$

for some constant $C > 0$. On the other hand, the empirical risk term, $\|W(G\boldsymbol{\theta} - \boldsymbol{y})\|_p^p$, 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 $J_p(\boldsymbol{\theta})$ is convex in $\boldsymbol{\theta}$.

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 the optimization with $p = 2$, we can adopt FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) [18], which is an extension of Nesterov's work [19] to achieve the convergence rate $O(1/k^2)$ at k-th iteration. On the other hand, 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 [14], [15].

Remark 1: The optimization is related to the following signal subspace

$$V := \left\{ u \in L^2[0,T] : u = \sum_{i=1}^{N} \theta_i g(t_i - \cdot),\ \theta_i \in \mathbb{R} \right\}.$$

That is, we seek the optimal control u in V such that the coefficients minimize (12). Note that $\{g(t_1 - \cdot), \ldots, g(t_N - \cdot)\}$ is a basis of V due to the controllability and observability of system (1).

Remark 2: Although we have assumed that the initial state $\boldsymbol{x}$ is $\boldsymbol{0}$, we can also set the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ as a design variable in a similar manner. In this case, the output $y(t)$ becomes

$$y(t_j) = \boldsymbol{c}^\top e^{At_j} \boldsymbol{x}_0 + \sum_{i=1}^{N} \theta_i G_{ij}, \quad j = 1, 2, \ldots, N,$$

and the optimization is formulated by

$$\min_{\boldsymbol{x}_0, \boldsymbol{\theta}} \left\{ \eta \|\boldsymbol{\theta}\|_1 + \|W(H\boldsymbol{x}_0 + G\boldsymbol{\theta} - \boldsymbol{y})\|_p^p \right\}, \qquad (13)$$

where $H := [e^{A^\top t_1}\boldsymbol{c}, \ldots, e^{A^\top t_N}\boldsymbol{c}]^\top$. This is also a convex optimization problem and can be efficiently solved via numerical optimization softwares.

Remark 3: The choice of parameters η and w_i influences the performance of curve fitting. The regularisation parameter η controls the trade-off between the sparsity and fidelity of the solution; a larger η leads to a sparser solution (i.e. more θ_i's are zero) while a smaller η leads to a smaller empirical risk. On the other hand, w_i may be chosen to be larger if the data y_i contains smaller error. These parameters should be chosen by trial and error (e.g. cross-validation [10]).

IV. Numerical Example

In this section, we show a numerical example that illustrates the effectiveness of the proposed L^1 control theoretic smoothing spline. We set the dynamical system P with transfer function

$$P(s) = \frac{1}{s^3 + 1}.$$

State-space matrices for $P(s)$ are given by

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

We assume the original curve is given by

$$y_{\text{orig}}(t) = \sin(2t) + 1.$$

The sampling instants are given by

$$t_i = 0.1 + 0.01(i - 1), \quad i = 1, 2, \ldots, 501,$$

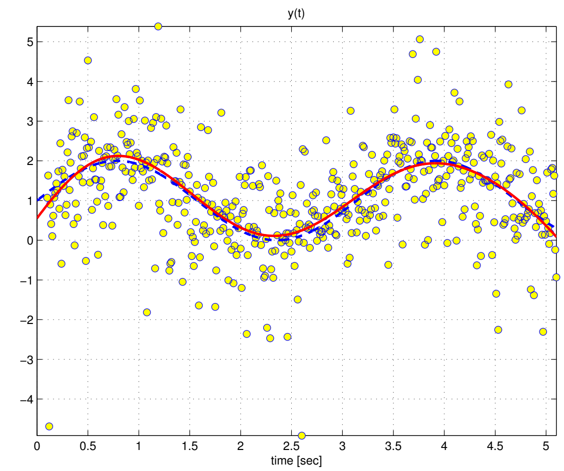

Fig. 2. Simulation result of L^1 spline: original curve (dashed line), observed data (circles), fitted curve (solid line).

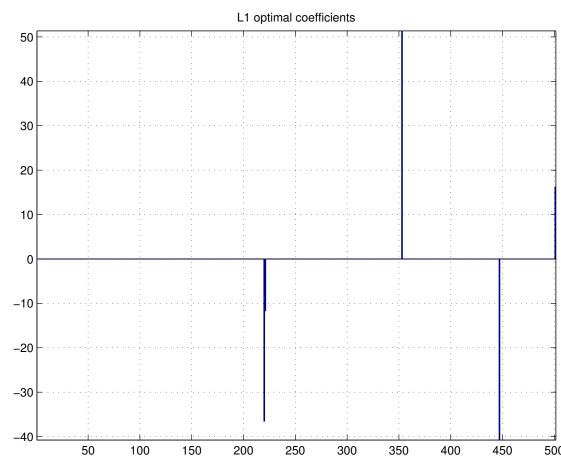

Fig. 3. Coefficients of L^1 spline

that is, the data are sampled at rate 100 [Hz] (100 samples per second) from initial time $t_1 = 0.1$. The observed data $y_1, y_2, \ldots, y_{501}$ are assumed to be disturbed by additive Laplacian noise with mean 0 and variance 1. See Fig. 2 for the original curve $y_{\text{orig}}(t)$ and the observed data $y_1, y_2, \ldots, y_{501}$.

For these data, we compute the optimal coefficients of the L^1 control theoretic spline with $p = 1$ corresponding to Laplacian noise. The design parameters are $\eta = 0.01$ and $w_i = 1$ for all i (i.e. all elements have equal weight). We assume that the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ is also a design variable, that is, we solve optimization (13).

Fig. 2 shows the resulting fitted curve $y(t)$ computed with the L^1-optimal control $u(t)$. We can see that the data are considerably disturbed by Laplacian noise, but the reconstructed curve well fits the original curve. To see the sparsity property of the L^1-optimal coefficients, we plot the value of the coefficients in Fig. 3. As shown in this figure, the L^1-optimal coefficients are quite sparse. In fact, the number of coefficients whose absolute values are greater than 0.001 is

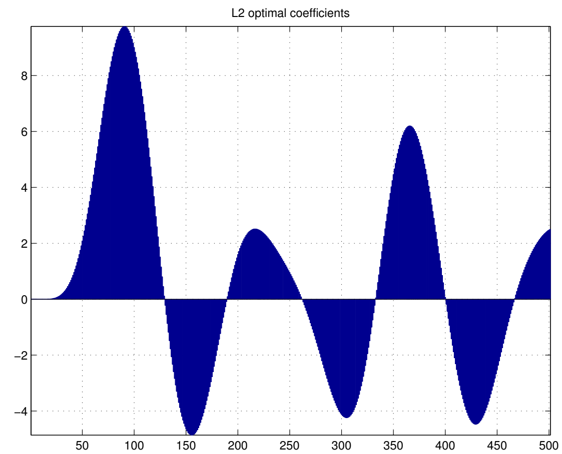

Fig. 4. Coefficients of L^2 spline

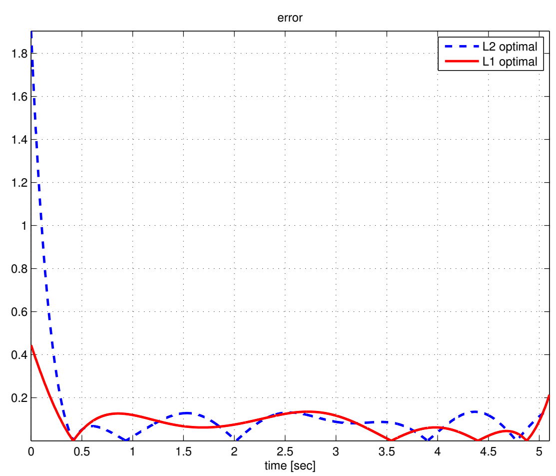

Fig. 5. Error between original curve $y(t)$ and fitted curve by L^1 spline (solid line) and L^2 spline (dashed line)

just 5 out of 501 coefficients. On the other hand, we show the L^2-optimal coefficients with $\lambda = 0.0001$, see equation (2), in Fig. 4. This figure indicates that the coefficients are not sparse at all and the L^2 spline requires almost all the base functions to represent the fitted curve. Note that the reconstructed curve by the L^2 spline also well fits the original curve as shown in Fig. 5, which shows the error between the original curve and the fitted curves. This figure shows that the L^2 spline is almost comparable with the L^1 splines[3].

In summary, we can say by the simulation that the proposed L^1 control theoretic smoothing spline can effectively reduce the effect of noise in data and also give sufficiently sparse representation for the fitted curve.

V. Conclusion

In this paper, we have proposed the L^1 control theoretic smoothing splines for noise reduction and sparse representation. The design is formulated as coefficient optimization with an L^1 regularized term and an L^1 or L^2 empirical risk term, which can be efficiently solved by numerical computation softwares. A numerical example has been shown to illustrate the effectiveness of the proposed L^1 spline.

Future work may include extension to constrained splines as proposed in [20], and extension to sparse feedback control as discussed in [21], [22].

[3] Another example in [20] shows that an L^1 spline outperforms an L^2 spline in view of outlier rejection.

Acknowledgment

This research is supported in part by the JSPS Grant-in-Aid for Scientific Research (C) No. 24560543, MEXT Grant-in-Aid for Scientific Research on Innovative Areas No. 26120521, and an Okawa Foundation Research Grant.

References

[1] G. Kimeldorf and G. Wahba, “Some results on Tchebycheffian spline functions,” *Journal Math. Anal. Appl.*, vol. 33, no. 1, pp. 82–95, 1971.
[2] G. Wahba, *Spline Models for Observational Data*. SIAM, 1990.
[3] S. Sun, M. B. Egerstedt, and C. F. Martin, “Control theoretic smoothing splines,” *IEEE Trans. Autom. Control*, vol. 45, no. 12, pp. 2271–2279, Dec. 2000.
[4] M. Egerstedt and C. Martin, *Control Theoretic Splines*. Princeton University Press, 2010.
[5] M. Egerstedt and C. F. Martin, “Optimal trajectory planning and smoothing splines,” *Automatica*, vol. 37, no. 7, pp. 1057–1064, 2001.
[6] S. Takahashi and C. F. Martin, “Optimal control theoretic splines and its application to mobile robot,” in *Proc. of the 2004 IEEE CCA*, 2004, pp. 1729–1732.
[7] H. Kano, M. Egerstedt, H. Fujioka, S. Takahashi, and C. F. Martin, “Periodic smoothing splines,” *Automatica*, vol. 44, no. 1, pp. 185–192, 2008.
[8] J. K. Charles, S. Sun, and C. F. Martin, “Cumulative distribution estimation via control theoretic smoothing splines,” in *Three Decades of Progress in Control Sciences*. Springer, 2010, pp. 83–92.
[9] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *J. R. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[10] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*. Springer, 2011.
[11] B. Schölkopf and A. J. Smola, *Learning with Kernels*. The MIT Press, 2002.
[12] M. Nagahara, “Robust design of control theoretic splines via support vector regression,” in *Proc. 55th Annual Conference of the Institute of Systems, Control and Information Engineers (ISCIE)*, 2011, (in Japanese).
[13] M. Nagahara and C. F. Martin, “L^1-optimal splines for outlier rejection,” in *Proc. 59th World Statistics Congress*, Aug. 2013, pp. 1137–1142.
[14] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21,” http://cvxr.com/cvx, Apr. 2011.
[15] ——, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Splinger, 2008, vol. 371, pp. 95–110.
[16] http://www-ics.acs.i.kyoto-u.ac.jp/mat/spline/.
[17] W. J. Rugh, *Linear Systems Theory*. Prentice Hall, 1996.
[18] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
[19] Y. Nesterov, “A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$,” *Doklady AN USSR*, vol. 269, pp. 543–547, 1983, (translated as Soviet Math. Docl.
[20] M. Nagahara and C. F. Martin, “Monotone smoothing splines using general linear systems,” *Asian Journal of Control*, vol. 5, no. 2, pp. 461–468, Mar. 2013.
[21] M. Nagahara, D. E. Quevedo, and J. Østergaard, “Sparse packetized predictive control for networked control over erasure channels,” *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 1899–1905, Jul. 2014.
[22] M. Nagahara, D. E. Quevedo, and D. Nešić, “Maximum hands-off control and L^1 optimality,” in *Proc. of 52nd IEEE Conference on Decision and Control (CDC)*, Dec. 2013, pp. 3825–3830.