Automatica 113 (2020) 108679

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

CLOT norm minimization for continuous hands-off control[☆]

Masaaki Nagahara[a,b,*], Debasish Chatterjee[b], Niharika Challapalli[c], Mathukumalli Vidyasagar[d,1]

[a] Institute of Environmental Science and Technology, The University of Kitakyushu, Hibikino 1-1, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan
[b] Systems and Control Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra, 400076, India
[c] University of Texas at Dallas, Richardson, TX, 75080, USA
[d] Indian Institute of Technology Hyderabad, Kandi, Telangana, 502285, India

ARTICLE INFO

Article history:
Received 7 October 2017
Received in revised form 9 September 2019
Accepted 15 October 2019
Available online 14 November 2019

Keywords:
Optimal control
Convex optimization
Sparsity
Maximum hands-off control
Bang-off-bang control

ABSTRACT

In this paper, we propose optimal control that is both sparse and continuous, unlike previously proposed alternatives to maximum hands-off control. The maximum hands-off control is the L^0-optimal (or sparsest) control among all feasible controls that are bounded by a specified value and transfer the state from a given initial state to the origin within a fixed time duration. The proposed control is obtained via minimization of the CLOT (Combined L-One and Two) norm of the control input along with the constraints on the state variable. The constraints on the state variable ensures that the states are not blown up while achieving the optimal control. By using the non-smooth maximum principle, we prove that the CLOT-norm optimal control is unique, and it is continuous in the time variable. We show by numerical simulations that the CLOT control is continuous unlike L^0-optimal control (or maximum hands-off control) and much sparser (i.e. has longer time duration on which the control equals 0) than the conventional EN (elastic net) control, which is a convex combination of L^1 and squared L^2 norms.

1. Introduction

Sparsity has recently emerged as an important topic in signal/image processing, machine learning, statistics, etc. If $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ are specified with $m < n$, then the equation $y = Ax$ is under-determined and has infinitely many solutions for x if A has rank m. Finding the sparsest solution (that is, the solution with the fewest number of nonzero elements) can be formulated as

$$\min_z \|z\|_0 \text{ subject to } Az = b.$$

However, this problem is NP hard, as shown in Natarajan (1995). Therefore other approaches have been proposed for this problem. This area of research is known as *sparse regression*. One of the most popular is LASSO (Tibshirani, 1996), also referred to as forgetting (Ishikawa, 1996), or basis pursuit (Chen, Donoho, & Saunders, 1999), in which the ℓ^0-norm is replaced by the ℓ^1-norm. Thus the problem becomes

$$\min_z \|z\|_1 \text{ subject to } Az = b.$$

The advantage of LASSO is that it is a convex optimization problem and therefore very large problems can be solved efficiently, for example by using the Matlab-based package `cvx` (Grant & Boyd, 2014). Moreover, under mild technical assumptions, the LASSO-optimal solution has no more than m nonzero components (Osborne, Presnell, & Turlach, 2000). However, the exact location of the nonzero components is very sensitive to the vector y. To overcome this deficiency, another approach known as the Elastic Net was proposed in Zou and Hastie (2005), where the ℓ^1 norm in LASSO is replaced by a weighted sum of ℓ^1 and squared ℓ^2 norms. This leads to the optimization problem

$$\min_z \lambda_1\|z\|_1 + \lambda_2\|z\|_2^2 \text{ subject to } Az = b,$$

[☆] The material in this paper was partially presented at the 20th IFAC World Congress of the International Federation of Automatic Control, July 9–14, 2017, Toulouse, France. This paper was recommended for publication in revised form by Associate Editor Kok Lay Teo under the direction of Editor Ian R. Petersen.

* Corresponding author at: Institute of Environmental Science and Technology, The University of Kitakyushu, Hibikino 1-1, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan.

E-mail addresses: nagahara@ieee.org (M. Nagahara), dchatter@iitb.ac.in (D. Chatterjee), niharika15c@gmail.com (N. Challapalli), m.vidyasagar@iith.ac.in (M. Vidyasagar).

[1] This work was carried out when the author was with the University of Texas at Dallas.

https://doi.org/10.1016/j.automatica.2019.108679

where λ_1 and λ_2 are positive weights such that $\lambda_1 + \lambda_2 = 1$. It is shown in Zou and Hastie (2005, Theorem 1) that the EN formulation gives the *grouping effect*; If two columns of the matrix A are highly correlated, then the corresponding components of the solution for x have nearly equal values. This ensures that the solution for x is not overly sensitive to small changes in y. The name "elastic net" is meant to suggest a stretchable fishing net that retains *all the big fish*.

During the past decade and a half, another research area known as *compressed sensing* has witnessed a great deal of interest. In compressed sensing, the matrix A is not specified; rather, the user gets to choose the integer m (known as the number of measurements), as well as the matrix A. The objective is to choose the matrix A as well as a corresponding *decoder* map $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ such that, the unknown vector x is sparse and the measurement vector y equals Ax, then $\Delta(Ax) = x$ for all sufficiently sparse vectors x. More generally, if measurement vector $y = Ax + \eta$ where η is the measurement noise, and the vector x is nearly sparse (but not exactly sparse), then the recovered vector $\Delta(Ax+\eta)$ should be sufficiently close to the true but unknown vector x. This is referred to as *robust sparse recovery*. Minimizing the ℓ_1-norm is among the more popular decoders. See the books by Elad (2010), Eldar and Kutyniok (2012) and Foucart and Rauhut (2013) for the theory and some applications. Due to its similarity to the LASSO formulation of Tibshirani (1996), this approach to compressed sensing is also referred to as LASSO.

Until recently the situation was that LASSO achieves robust sparse recovery in compressed sensing, but did not achieve the grouping effect in sparse regression. On the flip side, EN achieves the grouping effect, but it was not known whether it achieves robust sparse recovery. A recent paper (Ahsen, Challapalli, & Vidyasagar, 2017) sheds some light on this problem. It is shown in Ahsen et al. (2017) that EN *does not achieve* robust sparse recovery. To achieve both the grouping effect in sparse regression as well as robust sparse recovery in compressed sensing, Ahsen et al. (2017) have proposed the CLOT (Combined L-One and Two) formulation:

$$\min_z \lambda_1 \|z\|_1 + \lambda_2 \|z\|_2 \quad \text{subject to} \quad Az = b,$$

where $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 + \lambda_2 = 1$. The difference between EN and CLOT is the ℓ^2 norm term; EN has the squared ℓ^2 norm while CLOT has the pure ℓ^2 norm. This slight change leads to both the grouping effect and robust sparse recovery, as shown in Ahsen et al. (2017).

In parallel with these advances in sparse regression and recovery of unknown sparse vectors, sparsity techniques have also been applied to control. Sparsity-promoting optimization has been applied to networked control in Nagahara, Quevedo, and Østergaard (2014), where quantization errors and data rate can be reduced at the same time by sparse representation of control packets. Other examples of control applications include optimal controller placement by Casas, Clason and Kunisch (2012), Clason and Kunisch (2012) and Fardad, Lin, and Jovanović (2011), design of feedback gains by Lin, Fardad, and Jovanovic (2013) and Polyak, Khlebnikov, and Shcherbakov (2013), state estimation by Charles, Asif, Romberg, and Rozell (2011), and sparse control of partial differential equations (Casas, Herzog and Wachsmuth, 2012; Herzog, Stadler, & Wachsmuth, 2012), to name a few.

More recently, a novel control called the *maximum hands-off control* has been proposed in Nagahara, Quevedo, and Nešić (2016) for *continuous-time* systems. The maximum hands-off control is the L^0-optimal control (the control that has the minimum support length) among all feasible controls that are bounded by a fixed value and transfer the state from a given initial state to the origin within a fixed time duration. Such a control is effective for reduction of electricity or fuel consumption; an electric/hybrid vehicle shuts off the internal combustion engine (i.e. hands-off control) when the vehicle is stopped or the speed is lower than a preset threshold; see Chan (2007) for example. Railway vehicles also utilize hands-off control, often called *coasting control*, to cut electricity consumption; see Liu and Golovitcher (2003) for details. In Nagahara et al. (2016), the authors have proved the theoretical relation between the maximum hands-off control and the L^1 optimal control under the assumption of normality. Also, important properties of the maximum hands-off control have been proved in Ikeda and Nagahara (2016) for the convexity of the value function, and in Chatterjee, Nagahara, Quevedo, and Mallikarjuna Rao (2016) for necessary conditions of optimality, and in Ikeda, Nagahara, and Ono (2017) for the discrete-valued control.

In general, the maximum hands-off control is a bang-off-bang control taking values of ± 1 and 0. Such control is called a discrete-valued control, for which parametrization methods have been proposed in Lee, Teo, Rehbock, and Jennings (1999) and Wu, Teo, and Rehbock (2008). In some applications, discrete-valued control is preferable. However, for some other applications, such a discontinuity property is not desirable. To obtain a continuous but still sparse control, Nagahara et al. (2016) have proposed to use a combined L^1 and *squared* L^2 minimization, like EN mentioned above. Let us call this control an EN control. As in the case of EN in the vector optimization, the EN control often shows much less sparsity (i.e. has a larger L^0 norm) than the maximum hands-off control. Then, in Challapalli, Nagahara, and Vidyasagar (2017), we have proposed to use the CLOT norm, a convex combination of L^1 and *non-squared* L^2 norms. The minimum CLOT-norm control is called the CLOT control. In Challapalli et al. (2017), we have shown by numerical simulation that the CLOT control is continuous and much sparser (i.e. has longer time duration on which the control equals 0) than the conventional EN control.

In Nagahara et al. (2016), both the LASSO and EN approaches to hands-off control are solved in continuous-time. It is shown, using Pontryagin's maximum principle, that the LASSO solution is bang-off-bang, while the EN solution is continuous. However, the CLOT formulation cannot be addressed via Pontryagin's principle. Therefore it is not clear whether the resulting optimal control is continuous. The main contribution of this paper is to show the continuity property of the CLOT control. For the analysis, we equivalently transform the original optimal control problem into a non-smooth optimal control problem, and employ the non-smooth version of Pontryagin's maximum principle by Clarke (2013). This paper is based on the authors' conference papers (Challapalli et al., 2017; Nagahara & Chatterjee, 2019).

The remainder of this article is organized as follows. In Section 2, we formulate the control problem considered in this paper. The existence, uniqueness, and continuity of the CLOT-norm optimal control are proved in Section 3. In Section 4, we propose a discretization method to numerically compute the optimal control, and also consider additional state constraints for the optimal control problem. In Section 5, we show a numerical example to illustrate the advantages of the CLOT control compared with the maximum hands-off control and the EN control. We present conclusions in Section 6.

2. Problem formulation

Let us consider a continuous-time linear time-invariant system described by

$$\frac{dx}{dt}(t) = Ax(t) + Bu(t), \;\; t \geq 0, \;\; x(0) = \xi. \tag{1}$$

Here we assume that $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, and the initial state $x(0) = \xi$ is fixed and given. The control objective is to drive the state $x(t)$ from $x(0) = \xi$ to the origin at time $T > 0$, that is

$$x(T) = 0. \tag{2}$$

We limit the control $u(t)$ to satisfy

$$\|u\|_\infty \leq 1. \tag{3}$$

If the system (1) is controllable and the final time T is larger than the optimal time $T^*(\xi)$, the minimum time to steer the state $x(t)$ in (1) from $x(0) = \xi$ to the origin by a control satisfying (3), then there exists at least one $u(t) \in L^\infty[0, T]$ that satisfies equations (1), (2), and (3). Let us call such a control a *feasible* control, and we denote by $\mathcal{U}$ the set of feasible controls.

The problem of the maximum hands-off control is then described by

$$\underset{u}{\text{minimize}} \quad \|u\|_0 \quad \text{subject to} \quad u \in \mathcal{U}. \tag{4}$$

Here $\|u\|_0$ is the L^0 norm of u defined by the length of the support set of u, that is, $\|u\|_0 \triangleq \mu_L(\text{supp}(u))$, where $\text{supp}(u)$ is the support of the signal u, and μ_L is the Lebesgue measure on $\mathbb{R}$.

The L^0 problem (4) is very hard to solve since the L^0 cost function is non-convex and discontinuous. For this problem, Nagahara et al. (2016) have shown that the L^0 optimal control in (4) is equivalent to the following L^1 optimal control:

$$\underset{u}{\text{minimize}} \quad \|u\|_1 \quad \text{subject to} \quad u \in \mathcal{U}, \tag{5}$$

if the plant is normal, that is, if the system (1) is controllable and the matrix A is nonsingular. Let us call the L^1 optimal control as the *LASSO control*. If the plant is normal, then the LASSO control is in general a *bang-off-bang* control that is piecewise constant taking values in $\{0, \pm 1\}$. The discontinuity of the LASSO solution may be undesirable in real applications, and a smoothed control is also proposed in Nagahara et al. (2016) by

$$\underset{u}{\text{minimize}} \quad \|u\|_1 + \lambda\|u\|_2^2 \quad \text{subject to} \quad u \in \mathcal{U}, \tag{6}$$

where $\lambda > 0$ is a design parameter for smoothness. Let us call this control the *EN (elastic net) control*. In Nagahara et al. (2016), it is proved that the solution of (6) is a continuous function on $[0, T]$.

While the EN control is continuous, it is shown by numerical experiments that the EN control is not sometimes sparse. This is an analogy of the EN for finite-dimensional vectors that EN does not achieve robust sparse recovery. Borrowing the idea of CLOT in Ahsen et al. (2017), we define the CLOT optimal control problem by

$$\underset{u}{\text{minimize}} \quad \|u\|_1 + \lambda\|u\|_2 \quad \text{subject to} \quad u \in \mathcal{U}. \tag{7}$$

We call this optimal control the *CLOT control*.

3. Continuity of CLOT control

In this section, we show the existence and continuity of CLOT control, and to this end, we stipulate that:

Assumption 1.

(1) The initial state is not zero, that is, $x(0) = \xi \neq 0$.
(2) The horizon length T is strictly greater than the minimum time $T^*(\xi)$.
(3) The pair (A, B) is controllable.

Then we have the following theorem.

Theorem 2. *Under Assumption 1, there exists a unique continuous optimal control that solves (7). In particular, there exist a vector $\pi_1 \in \mathbb{R}^n$ and a scalar $\pi_2 < 0$ such that the optimal control u^* is given by*

$$u^*(t) = Q_{\pi_2}\left(B^\top e^{-A^\top t}\pi_1\right), \tag{8}$$

where

$$Q_b(a) \triangleq \begin{cases} -1, & \text{if } \ a < -1 + 2b \\ -\frac{a+1}{2b}, & \text{if } \ -1 + 2b \leq a < -1 \\ 0, & \text{if } \ -1 \leq a \leq 1 \\ -\frac{a-1}{2b}, & \text{if } \ 1 < a \leq 1 - 2b \\ 1, & \text{if } \ 1 - 2b < a \end{cases} \tag{9}$$

From Theorem 2 (whose proof is given in the Appendix), we see that the CLOT control is continuous but not necessarily differentiable. We also note that there is no essential change in the result of Theorem 2 for multi-input LTI systems.

4. Discretization

Since the CLOT control problem (7) is infinite dimensional, we discretize it to obtain a finite dimensional problem for numerical computation. For this, we adopt the time discretization.

First, we divide the time interval $[0, T]$ into N subintervals, $[0, T] = [0, h) \cup \cdots \cup [(N-1)h, Nh]$, where h is the discretization step (or the sampling period) such that $T = Nh$. We assume that the control $u(t)$ in (1) are constant over each subinterval, that is, we consider a *zero-order hold* input for the control. On the discretization grid, $t = 0, h, \ldots, Nh$, the continuous-time system (1) is described as

$$\hat{x}_{k+1} = A_d\hat{x}_k + B_d\hat{u}_k, \quad k = 0, 1, \ldots, N-1, \tag{10}$$

where $\hat{x}_k \triangleq x(kh)$, $\hat{u}_k \triangleq u(kh)$, and

$$A_d \triangleq e^{Ah}, \quad B_d \triangleq \int_0^h e^{At}B dt. \tag{11}$$

Define the control vector $\hat{u} \triangleq [\hat{u}_0, \hat{u}_1, \ldots, \hat{u}_{N-1}]^\top$. Then the set $\mathcal{U}$ of continuous-time feasible controls is discretized by

$$\mathcal{U}_N \triangleq \left\{\hat{u} \in \mathbb{R}^N : A_d^N\xi + \Phi_N\hat{u} = 0, \ \|\hat{u}\|_\infty \leq 1\right\}, \tag{12}$$

where $\Phi_N \triangleq [A_d^{N-1}B_d, A_d^{N-2}B_d, \ldots, B_d]$.

Next, by the zero-order hold assumption, the L^1 norm of u is discretized as

$$\|u\|_1 = \sum_{k=0}^{N-1}\int_{kh}^{(k+1)h} |\hat{u}_k| dt = \|\hat{u}\|_1 h. \tag{13}$$

In the same way, we obtain approximation of the L^2 norm of u as $\|u\|_2^2 = \|\hat{u}\|_2^2 h$.

Finally, the CLOT optimal control problem (7) is discretized as

$$\underset{\hat{u} \in \mathbb{R}^N}{\text{minimize}} \quad h\|\hat{u}\|_1 + \sqrt{h}\lambda\|\hat{u}\|_2 \quad \text{subject to} \quad \hat{u} \in \mathcal{U}_N. \tag{14}$$

The optimization problem is convex and can be efficiently solved by numerical software packages such as `cvx` with Matlab; see Grant and Boyd (2014) for details. The computational time for both EN and CLOT are almost similar (see Section 5). Also, the problems can be efficiently solved by ADMM (Alternating Direction Method of Multipliers) algorithm (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011).

We can also consider additional state constraints to (14) to ensure that ℓ_2 norm of the state at any given instant does not blow up. The constraint is the ℓ_2 norm of the state vector at any given time should not exceed a specified threshold θ, that is,

$$\|\hat{x}_k\|_2 \leq \theta, \qquad k \in \{1, 2, \ldots, N-1\}, \tag{15}$$

where $\hat{x}_k$ is the discrete-time state at time instant k as defined in (10). Note that if θ is larger than the maximum value, say $l_{\max}$, of $\|\hat{x}_k\|_2$, $k \in \{1, \ldots, N-1\}$, then the optimization problem is still

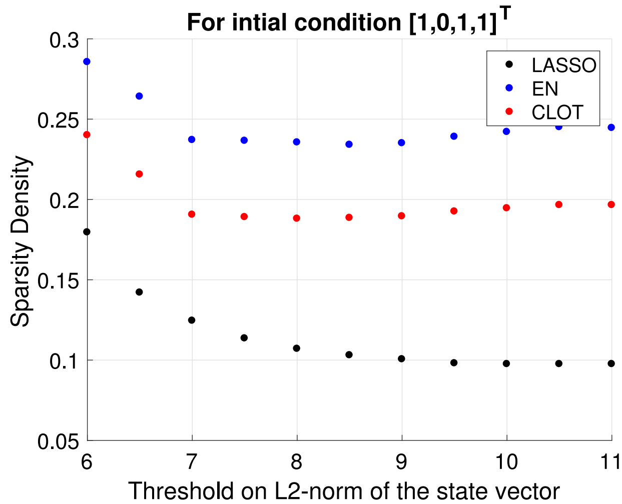

Fig. 1. State threshold vs. sparsity density.

unconstrained. Therefore, to make the constraint (15) effective, θ should be strictly less than $l_{\max}$.

The constrained optimization problem becomes

$$\begin{aligned} &\underset{\hat{u}\in\mathbb{R}^N}{\text{minimize}} \quad h\|\hat{u}\|_1 + \sqrt{h}\lambda\|\hat{u}\|_2 \\ &\text{subject to} \quad \hat{u} \in \mathcal{U}_N,\ \|\hat{x}_k\|_2 \le \theta,\ k \in \{1,\dots,N-1\}. \end{aligned} \tag{16}$$

We summarize the relationship between (16) and (4) as follows. The set of feasible solutions to the problems (16) is a subset of $\mathcal{U}_N$, the set of feasible solutions to (14). Hence the optimal value of (16) is larger than or equal to that of (14). Problem (14) is a discretized problem of (7). If we take $\lambda = 0$, then (7) is equivalent to (5). The equivalence between (5) and (4) is shown in Nagahara et al. (2016) under the assumption of normality, which holds when the system (1) is controllable and matrix A is non-singular.

5. Numerical example

In this section we present a numerical example of the CLOT optimal control of a linear plant, and compare the results with those with LASSO and EN optimizations.

We here consider a 4th order linear system with state–space matrices

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{17}$$

The horizon length is $T = 20$ and the initial state is $x(0) = [1, 0, 1, 1]^\top$. We take the regularization parameter $\lambda = 1$. With these parameters, we compute three optimal controls: LASSO, EN, and CLOT, with state constraint (15) across the range of the threshold parameter $6 \le \theta \le 10$. Fig. 1 shows the sparsity density of each control, defined by $\|u^*\|_0/T$, the fraction of time where the control is nonzero.

From Fig. 1, it is clearly noted that CLOT control input is more sparse than that of EN and less sparse compared to that of LASSO. Also, we measured the number of FLOPS (Floating Point Operations) for these optimizations using MATLAB function FLOPS.[2] Table 1 shows the average number of FLOPS exhausted in the numerical optimizations with the same platform of CVX in MATLAB. We can observe that they are also the same.

2 https://mathworks.com/matlabcentral/fileexchange/50608.

Table 1
Average number of FLOPS.

	LASSO	EN	CLOT
# of FLOPS	311,997	312,538	312,001

6. Conclusions

In this article, we have proposed the CLOT control that minimizes the weighted sum of L^1 and L^2 norms among feasible controls, to obtain a continuous control signal that is sparser than the EN control introduced in Nagahara et al. (2016). We have shown a discretization method, by which the CLOT optimal control problem can be solved via finite-dimensional convex optimization. We have shown that the CLOT control solution shows continuity unlike L^0-optimal control.

We have also introduced the state constraints to obtain the optimal control, to ensure the states does not blow up in order to get the optimal control. A numerical example has been shown to illustrate the advantage of the CLOT control. Future work includes extension of CLOT norm minimization to time-variant systems.

Acknowledgments

The research of MN was supported in part by JSPS, Japan KAKENHI Grant Numbers JP15H02668 and JP16KK0134, and the Ministry of Internal Affairs and Communications of Japan under the Strategic Information and Communications R&D Promotion Programme, 2017–2019, No. 3620. The research of MV and NC was supported by the US National Science Foundation, USA under Award No. ECCS-1306630, the Cancer Prevention and Research Institute of Texas (CPRIT), USA under award No. RP140517, and a grant from the Department of Science and Technology, Government of India.

Appendix A. Necessary conditions by non-smooth maximum principle

Here we give necessary conditions for the optimality of CLOT control. The CLOT optimal control problem is described by

$$\begin{aligned} &\underset{u}{\text{minimize}} \quad \int_0^T |u(t)|dt + \lambda\sqrt{\int_0^T |u(t)|^2 dt} \\ &\text{subject to} \\ &\quad \frac{dx}{dt}(t) = Ax(t) + Bu(t) \\ &\quad x(0) = \xi,\ x(T) = 0 \\ &\quad |u(t)| \le 1, \quad t \in [0,T]. \end{aligned} \tag{A.1}$$

To obtain necessary conditions, we transform the optimal control problem (A.1). Define

$$\begin{aligned} &z_1(t) \triangleq x(t) \\ &z_2(t) \triangleq \int_0^t |u(\tau)|^2 d\tau, \quad z_2(0) = 0, \quad t \in [0,T]. \end{aligned} \tag{A.2}$$

It is easily shown that the CLOT optimal control problem in (A.1) is equivalent to

$$\begin{aligned} &\underset{u}{\text{minimize}} \quad \int_0^T |u(t)|dt + \lambda\sqrt{z_2(T)} \\ &\text{subject to} \\ &\quad \frac{dz_1}{dt}(t) = Az_1(t) + Bu(t) \\ &\quad \frac{dz_2}{dt}(t) = |u(t)|^2 \\ &\quad z_1(0) = \xi,\ z_1(T) = 0,\ z_2(0) = 0 \\ &\quad |u(t)| \le 1, \quad t \in [0,T]. \end{aligned} \tag{A.3}$$

Then we apply the non-smooth maximum principle by Clarke (2013, Theorem 22.26). Let u^* be the optimal control and $z^*(t) \triangleq [z_1^*(t)^\top, z_2^*(t)]^\top$ be the optimal state trajectory by $u^*(t)$. The Hamiltonian with abnormal multiplier $\eta \in \{0, 1\}$ is given by

$$H^\eta(p, z, u) = p_1^\top(Az_1 + Bu) + p_2^\top |u|^2 - \eta|u|, \quad \text{(A.4)}$$

where $p(t) \triangleq [p_1(t)^\top p_2(t)]^\top$ is the co-state for (A.3). The necessary conditions are as follows.

- The non-triviality condition:

$$(\eta, p(t)) \neq (0, 0), \quad \text{for all } t \in [0, T]. \quad \text{(A.5)}$$

- The adjoint equation:

$$-\frac{dp}{dt}(t) = \partial_z H^\eta\big(p(t), z^*(t), u^*(t)\big) = \begin{bmatrix} A^\top p_1(t) \\ 0 \end{bmatrix}$$

or

$$\frac{p_1}{dt}(t) = -A^\top p_1(t), \quad \frac{p_2}{dt}(t) = 0.$$

From this, p_1 is an exponential function, that is,

$$p_1(t) = e^{-A^\top t}\pi_1, \quad t \in [0, T], \quad \text{(A.6)}$$

where $\pi_1 \in \mathbb{R}^n$, and p_2 is a constant function, that is

$$p_2(t) = \pi_2 \in \mathbb{R}, \quad t \in [0, T]. \quad \text{(A.7)}$$

- The transversality condition:

$$\begin{aligned}&\big(p(0), -p(T)\big) \\ &\quad = \eta \partial_L \ell\big(z^*(0), z^*(T)\big) + N_E^L\big(z^*(0), z^*(T)\big),\end{aligned} \quad \text{(A.8)}$$

where ℓ is the boundary cost function

$$\ell\big((\xi_1, \xi_1'), (\xi_2, \xi_2')\big) \triangleq \lambda\sqrt{|\xi_2'|} \quad \text{for } (\xi_i, \xi_i') \in \mathbb{R}^d \times \mathbb{R},$$

E is the boundary constraint set

$$E = \Big\{ \big((\xi_1, \xi_1'), (\xi_2, \xi_2')\big) \in (\mathbb{R}^d \times \mathbb{R})^2 \ \Big| \ \xi_1 = \xi, \xi_2 = 0, \xi_1' = 0 \Big\},$$

∂_L is the limiting subdifferential, and $N_E^L(\zeta_1, \zeta_2)$ is the limiting normal cone to E at $(\zeta_1, \zeta_2) \in (\mathbb{R}^{d+1})^2$. From the given boundary conditions it follows that

$$N_E^L\big(z^*(0), z^*(T)\big) = (\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \{0\}).$$

Moreover, under Assumption 1, it follows that no optimal control is a.e. 0; consequently, $z_2^*(T) > 0$, leading to

$$\partial_L \ell\big(z^*(0), z^*(T)\big) = \Big((0, 0), \Big(0, \tfrac{\lambda}{2\sqrt{z_2^*(T)}}\Big)\Big).$$

Substituting the two sets above into (A.8) with (A.7) leads to

$$p_2(T) = \pi_2 = -\frac{\eta\lambda}{2\sqrt{z_2^*(T)}}. \quad \text{(A.9)}$$

- The Hamiltonian maximization condition:

$$u^*(t) \in \arg\max_{u \in [-1,1]} \big\{p_1(t)^\top Bu + p_2(t)|u|^2 - \eta|u|\big\}. \quad \text{(A.10)}$$

From (A.6) and (A.7), this can be written as

$$u^*(t) \in \arg\max_{u \in [-1,1]} \big\{\pi_1^\top e^{-At} Bu + \pi_2|u|^2 - \eta|u|\big\}. \quad \text{(A.11)}$$

Appendix B. Proof of Theorem 2

First, we observe that the feasible set of (7) is non-empty by Assumption 1. Consequently, Bressan and Piccoli (2007, Theorem 5.2.2) guarantees the existence of an optimal control that solves (7).

Second, we show uniqueness of CLOT control. The feasible set $\mathcal{U}$ is a non-empty, closed, and convex set since it is the intersection of two non-empty, closed, and convex sets, $\{u : \Phi u + e^{AT}\xi = 0\}$ and $\{u : \|u\|_\infty \leq 1\}$, where Φ is the linear operator defined by

$$\Phi u \triangleq \int_0^T e^{A(T-t)} Bu(t) dt.$$

Suppose that u' and u'' are two distinct solutions of (7). We claim that $\frac{1}{2}(u' + u'')$ belongs to the admissible set and yields a strictly lower cost than both u' and u'', which leads to a contradiction to the optimality of u' and u''.

Before proceeding to prove our claim, we pause for a moment to recall that if two vectors v_1, v_2 satisfy $\|v_1 + v_2\|_2 = \|v_1\|_2 + \|v_2\|_2$, then $v_2 = \alpha v_1$ for some $\alpha \geq 0$. Indeed, the purported equality is equivalent to $\langle v_1, v_2\rangle = \|v_1\|_2 \|v_2\|_2$, and the equality case of Schwarz inequality yields the assertion at once. Suppose now that $v_1, v_2 \in \{u : \Phi u + e^{AT}\xi = 0\}$ satisfy $\|v_1 + v_2\|_2 = \|v_1\|_2 + \|v_2\|_2$. Since the preceding argument shows that $v_2 = \alpha v_1$ for some $\alpha \geq 0$, and the case $\alpha = 0$ does not arise because otherwise $v_1, v_2 \notin \{u : \Phi u + e^{AT}\xi = 0\}$, we further strengthen the range of admissible α to $\alpha > 0$. However, then

$$\begin{aligned}\Phi v_1 + e^{AT}\xi = 0 &= \Phi v_2 + e^{AT}\xi \\ &= \Phi(\alpha v_1) + \alpha e^{AT}\xi + (1-\alpha)e^{AT}\xi \\ &= \alpha\big(\Phi v_1 + e^{AT}\xi\big) + (1-\alpha)e^{AT}\xi \\ &= (1-\alpha)e^{AT}\xi,\end{aligned}$$

which means either $\alpha = 1$ or $e^{AT}\xi = 0$. Since e^{AT} is nonsingular for any $T > 0$, the second option $e^{AT}\xi = 0$ contradicts the assumption (Assumption 1) that $\xi \neq 0$. Therefore, $\alpha = 1$ is the only solution, which corresponds to the situation that $v_2 = v_1$. Consequently, if v_1 and v_2 are distinct, then $\|v_1 + v_2\|_2 < \|v_1\|_2 + \|v_2\|_2$.

To continue with the proof of our claim, note that convexity of $\mathcal{U}$ shows that $\frac{1}{2}(u' + u'')$ belongs to $\mathcal{U}$, and

$$\begin{aligned}&\left\|\frac{1}{2}(u' + u'')\right\|_1 + \lambda\left\|\frac{1}{2}(u' + u'')\right\|_2 \\ &\quad < \frac{1}{2}\big(\|u'\|_1 + \|u''\|_1\big) + \frac{\lambda}{2}\big(\|u'\|_2 + \|u''\|_2\big) \\ &\quad = \|u'\|_1 + \lambda\|u'\|_2 \quad \text{since } u', u'' \text{ solve (7)},\end{aligned}$$

where we have employed the triangle inequality for the $\|\cdot\|_1$-norm and the preceding argument for the $\|\cdot\|_2$-norm. However, this result contradicts optimality of u', and uniqueness of CLOT optimal solutions follows.

Now we prove the continuity of the CLOT control. We begin with the following lemma:

Lemma 3. *Under Assumption 1, we have $\eta = 1$.*

Proof. To prove this, we assume $\eta = 0$ and derive contradiction. If $\eta = 0$, then $\pi_2 = 0$ from (A.9). Also, we have $\pi_1 \neq 0$ from (A.5) and (A.6). From (A.11), we have

$$u^*(t) = \arg\max_{u \in [-1,1]} \big\{\pi_1^\top e^{-At} Bu\big\} = \operatorname{sign}\big(\pi_1^\top e^{-At} B\big).$$

Since $\pi_1 \neq 0$ and (A, B) is controllable, we have $\pi_1^\top e^{-At} B \neq 0$ for almost all $t \in [0, T]$. It follows that

$$|u^*(t)| = 1,$$

for almost all $t \in [0, T]$. With this optimal control, we have

$$J(u^*) = \int_0^T |u^*(t)| dt + \lambda\sqrt{\int_0^T |u^*(t)|^2 dt} = T + \lambda\sqrt{T}. \quad \text{(B.1)}$$

On the other hand, if we consider the minimum-time control $u^\dagger$ with minimum-time $T^*(\xi)$, the following control is a feasible control (i.e. $u \in \mathcal{U}$):

$$u^\ddagger(t) \triangleq \begin{cases} u^\dagger(t), & \text{if } 0 \le t \le T^*(\xi) \\ 0, & \text{if } T^*(\xi) < t \le T \end{cases}$$

and the value of the cost function is

$$J(u^\ddagger) = T^*(\xi) + \lambda\sqrt{T^*(\xi)} < T + \lambda\sqrt{T} = J(u^*).$$

This contradicts the optimality of u^*, and hence $\eta \neq 0$, or $\eta = 1$. □

Now we prove Theorem 2. From Lemma 3 and (A.11), we have

$$u^*(t) \in \arg\max_{u\in[-1,1]} \left\{\pi_1^\top e^{-At}Bu + \pi_2|u|^2 - |u|\right\}.$$

Since $\lambda > 0$, we have $\pi_2 < 0$. Then, it is easily shown that the optimal control is given by (8) with (9). The two functions $t \mapsto B^\top e^{-A^\top t}\pi_1$ and $Q(\cdot)$ are continuous, and therefore, so is their composition. Our proof is complete.

References

Ahsen, Mehmet Eren, Challapalli, Niharika, & Vidyasagar, Mathukumalli (2017). Two new approaches to compressed sensing exhibiting both robust sparse recovery and the grouping effect. *Journal of Machine Learning Research (JMLR)*, *18*(54), 1–24.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, *3*(1), 1–122.

Bressan, Alberto, & Piccoli, Benedetto (2007). *Applied Mathematics (Book 2), Introduction to the mathematical theory of control*. AIMS.

Casas, E., Clason, C., & Kunisch, K. (2012). Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM Journal on Control and Optimization*, *50*, 1735–1752.

Casas, Eduardo, Herzog, Roland, & Wachsmuth, Gerd (2012). Approximation of sparse controls in semilinear elliptic equations. In Lirkov Ivan, Svetozar Margenov, & Jerzy Waśniewski (Eds.), *Large-scale scientific computing* (pp. 16–27). Berlin, Heidelberg: Springer Berlin Heidelberg.

Challapalli, Niharika, Nagahara, Masaaki, & Vidyasagar, Mathukumalli (2017). Continuous hands-off control by CLOT norm minimization. In *Proc. of the 20th IFAC world congress* (pp. 15019–15024).

Chan, C. (2007). The state of the art of electric, hybrid, and fuel cell vehicles. *Proceedings of the IEEE*, *95*(4), 704–718.

Charles, Adam, Asif, M. S., Romberg, J., & Rozell, Christopher (2011). Sparsity penalties in dynamical system estimation. In *45th Annual conference on information sciences and systems (CISS)* (pp. 1–6).

Chatterjee, Debasish, Nagahara, Masaaki, Quevedo, Daniel E., & Mallikarjuna Rao, K. S. (2016). Characterization of maximum hands-off control. *Systems & Control Letters*, *94*, 31–36.

Chen, S. S., Donoho, D. L., & Saunders, M. A. (1999). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, *20*(1), 33–61.

Clarke, F. (2013). *Graduate texts in mathematics*: *vol. 264*, *Functional analysis, calculus of variations and optimal control*. London: Springer.

Clason, C., & Kunisch, K. (2012). A measure space approach to optimal source placement. *Computational Optimization and Applications*, *53*, 155–171.

Elad, M. (2010). *Sparse and redundant representations*. Springer.

Eldar, Y. C., & Kutyniok, G. (2012). *Compressed sensing: Theory and applications*. Cambridge University Press.

Fardad, M., Lin, Fu, & Jovanović, M. R. (2011). Sparsity-promoting optimal control for a class of distributed systems. In *American control conference (ACC), 2011* (pp. 2050–2055).

Foucart, Simon, & Rauhut, Holger (2013). *A mathematical introduction to compressive sensing*. Birkhäuser.

Grant, Michael, & Boyd, Stephen (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

Herzog, R., Stadler, G., & Wachsmuth, G. (2012). Directional sparsity in optimal control of partial differential equations. *SIAM Journal on Control and Optimization*, *50*(2), 943–963.

Ikeda, T., & Nagahara, M. (2016). Value function in maximum hands-off control for linear systems. *Automatica*, *64*, 190–195.

Ikeda, T., Nagahara, M., & Ono, S. (2017). Discrete-valued control of linear time-invariant systems by sum-of-absolute-values optimization. *IEEE Transactions on Automatic Control*, *62*(6), 2750–2763.

Ishikawa, Masumi (1996). Structural learning with forgetting. *Neural Networks*, *9*(3), 509–521.

Lee, H. W. J., Teo, K. L., Rehbock, V., & Jennings, L. S. (1999). Brief control parametrization enhancing technique for optimal discrete-valued control problems. *Automatica*, *35*(8), 1401–1407.

Lin, F., Fardad, M., & Jovanovic, M. R. (2013). Design of optimal sparse feedback gains via the alternating direction method of multipliers. *IEEE Transactions on Automatic Control*, *58*(9), 2426–2431.

Liu, Rongfang, & Golovitcher, Iakov M. (2003). Energy-efficient operation of rail vehicles. *Transportation Research Part A: Policy and Practice*, *37*(10), 917–932.

Nagahara, M., & Chatterjee, D. (2019). Continuity of the combined L^1-L^2 optimal control for linear systems. In *5th Indian Control Conference (ICC)* (pp. 506–509). Delhi.

Nagahara, M., Quevedo, D. E., & Nešić, D. (2016). Maximum hands-off control: a paradigm of control effort minimization. *IEEE Transactions on Automatic Control*, *61*(3), 735–747.

Nagahara, M., Quevedo, D. E., & Østergaard, J. (2014). Sparse packetized predictive control for networked control over erasure channels. *IEEE Transactions on Automatic Control*, *59*(7), 1899–1905.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, *24*(2), 227–234.

Osborne, Michael R., Presnell, Brett, & Turlach, Berwin A. (2000). On the LASSO and its dual. *Journal of Computational and Graphical Statistics*, *9*, 319–337.

Polyak, B., Khlebnikov, M., & Shcherbakov, P. (2013). An lmi approach to structured sparse feedback design in linear control systems. In *European control conference (ECC)* (pp. 833–838).

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.

Wu, C. Z., Teo, K. L., & Rehbock, V. (2008). A filled function method for optimal discrete-valued control problems. *Journal of Global Optimization*, *44*(2), 213.

Zou, Hui, & Hastie, Trevor (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B.*, *67*, 301–320.

Masaaki Nagahara received the bachelor's degree in engineering from Kobe University in 1998, and the master's degree and the Doctoral degree in informatics from Kyoto University in 2000 and 2003, respectively.

He is currently a Full Professor with the Institute of Environmental Science and Technology, The University of Kitakyushu. He has also been a Visiting Professor with Indian Institute of Technology Bombay since 2017. His research interests include control theory, machine learning, and sparse modeling.

He received Transition to Practice Award in 2012 and George S. Axelby Outstanding Paper Award in 2018 from IEEE Control Systems Society.

He also received Young Authors Award in 1999, Best Paper Award in 2012, and Best Book Authors Award in 2016, from SICE, and Best Tutorial Paper Award in 2014 from IEICE Communications Society.

He is a senior member of IEEE, and a member of IEICE, SICE, ISCIE, and JSAI.

Debasish Chatterjee received his Ph.D. in Electrical & Computer Engineering from the University of Illinois at Urbana-Champaign in 2007. He worked at ETH Zurich as a postdoc before joining IIT Bombay in 2011. His research interests are in constrained control with emphasis on computational tractability, geometric techniques in control, and applied probability. He serves as an Associate Editor of Automatica and the IFAC Journal of Systems & Control, and as an Editor of the International Journal of Robust and Nonlinear Control.

Niharika Challapalli was born in India in 1992. She received the B.Tech.(Honors) from the Indian Institute of Technology Bhubaneswar, Bhubaneswar, India in 2013, the M.Tech. from the Indian Institute of Technology Madras, Chennai, India in 2015 and the Ph.D. from the University of Texas at Dallas, Dallas, Texas, USA in 2018, all in Electrical Engineering. She is currently with Qualcomm Inc., San Diego, California, USA working on extended reality (XR) applications. Her research interests include compressed sensing, optimal control, distributed control and optimization, statistical analysis, machine learning and deep learning techniques for XR applications.

Mathukumalli Vidyasagar was born in Guntur, India on September 29, 1947. He received the B.S., M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin in Madison, in 1965, 1967 and 1969 respectively. Between 1969 and 1989, he was a Professor of Electrical Engineering at Marquette University, Concordia University, and the University of Waterloo. In 1989 he returned to India as the Director of the newly created Centre for Artificial Intelligence and Robotics (CAIR) in Bangalore. In 2000 he moved to the Indian private sector, and joined India's largest software company, Tata Consultancy Services, as an Executive Vice President, and was located in Hyderabad. In 2009 he retired from TCS and joined the Erik Jonsson School of Engineering & Computer Science at the University of Texas at Dallas, as a Cecil & Ida Green Chair in Systems Biology Science. He retired from UT Dallas at the end of 2017, and joined the Indian Institute of Technology Hyderabad, where he is a SERB (Science and Engineering Research Board) Distinguished Fellow and is a Distinguished Professor.

His current research interests are in the areas of machine learning, and compressed sensing. On the applications front, he is interested in applying ideas from machine learning to problems in computational biology with emphasis on cancer.

Vidyasagar has received a number of awards in recognition of his research contributions, including Fellowship in The Royal Society, the world's oldest scientific academy in continuous existence, the IEEE Control Systems (Technical Field) Award, the Rufus Oldenburger Medal of ASME, the John R. Ragazzini Education Award from AACC, and others. He is the author of twelve books and about 150 papers in peer-reviewed journals.