



# Sparsity-promoting optimal control of systems with symmetries, consensus and synchronization networks<sup>☆</sup>



Xiaofan Wu<sup>a</sup>, Mihailo R. Jovanović<sup>b,\*</sup>

<sup>a</sup> Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, United States

<sup>b</sup> Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, United States

## ARTICLE INFO

### Article history:

Received 29 February 2016

Received in revised form 22 October 2016

Accepted 6 February 2017

### Keywords:

Alternating direction method of multipliers

Consensus and synchronization networks

Distributed systems

Regularization

Sparsity-promoting optimal control

Spatially-invariant systems

## ABSTRACT

Optimal control problems in systems with symmetries and consensus/synchronization networks are characterized by structural constraints that arise either from the underlying group structure or the lack of absolute measurements for part of the state vector. Our objective is to design controller structures and resulting control strategies that utilize limited information exchange between subsystems in large-scale networks. To obtain controllers with low communication requirements, we seek solutions to regularized versions of the  $\mathcal{H}_2$  optimal control problem. Non-smooth regularization terms are introduced to tradeoff network performance with sparsity of the feedback-gain matrix. In contrast to earlier results, our framework allows the state-space representations that are used to quantify the system's performance and sparsity of the static output-feedback controller to be expressed in different sets of coordinates. We show how alternating direction method of multipliers can be leveraged to exploit the underlying structure and compute sparsity-promoting controllers. In particular, for spatially-invariant systems, the computational complexity of our algorithm scales linearly with the number of subsystems. We also identify a class of optimal control problems that can be cast as semidefinite programs and provide an example to illustrate our developments.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

In large networks of dynamical systems centralized information processing may impose prohibitively expensive communication and computation burden [1,2]. This motivates the development of theory and techniques for designing distributed controller architectures that lead to favorable performance of large-scale networks. Recently, regularized versions of standard optimal control problems were introduced as a means for achieving this goal [3–6]. For example, in consensus and synchronization networks, it is of interest to achieve desired objective using relative information exchange between limited subset of nodes [7–18].

The objective of this paper is to design controllers that provide a desired tradeoff between the network performance and the sparsity of the static output-feedback controller. This is accomplished by regularizing the  $\mathcal{H}_2$  optimal control problem with a penalty on communication requirements in the distributed controller. In contrast to previous work [3–5], this regularization penalty reflects the

fact that sparsity should be enforced in a specific set of coordinates. In [3–5], the elements of the state-feedback gain matrix were taken to represent communication links. Herein, we present a unified framework where a communication link is a linear function of the elements of the output-feedback gain matrix.

The proposed framework addresses challenges that arise in systems with invariances and symmetries, as well as consensus and synchronization networks. For example, the block diagonal structure of spatially-invariant systems in the spatial frequency domain facilitates efficient computation of the optimal centralized controllers [1]. However, since the sparsity requirements are typically expressed in the physical space, it is challenging to translate them into frequency domain specifications. Furthermore, in wide-area control of power networks [19–21], it is desired to design the controllers that respect the structure of the original system: in both open- and closed-loop networks, only relative rotor angle differences between different generators are allowed to appear. To deal with these structural requirements, we introduce a coordinate transformation to eliminate the average mode and assure stabilizability and detectability of the remaining modes. Once again, it is desired to promote sparsity of the feedback gain in physical domain and it is challenging to translate these requirements in the transformed set of coordinates.

<sup>☆</sup> Financial support from the National Science Foundation under award ECCS-1407958 is gratefully acknowledged.

\* Corresponding author.

E-mail addresses: [wuxxx836@umn.edu](mailto:wuxxx836@umn.edu) (X. Wu), [mihailo@umn.edu](mailto:mihailo@umn.edu) (M.R. Jovanović).

We leverage the alternating direction method of multipliers (ADMM) [22] to exploit the structure of the corresponding objective functions in the regularized optimal control problem. ADMM alternates between optimizing closed-loop performance and promoting sparsity of the feedback gain matrix. The sparsity promoting step in ADMM has an explicit solution and the performance optimization step is solved using Anderson-Moore and proximal gradient methods. Our framework thus allows for performance and sparsity requirements to be expressed in different set of coordinates and facilitates efficient computation of sparse static output-feedback controllers.

For undirected consensus networks, the proposed approach admits a convex characterization. Furthermore, for systems with invariances and symmetries, transform techniques are utilized to gain additional computational advantage and improve efficiency. For example, by bringing the matrices associated with a state-space representation of a spatially-invariant system into a block-diagonal form, the regularized optimal control problem amounts to easily parallelizable task of solving a sequence of smaller, fully-decoupled problems. While the computational complexity of algorithms that do not exploit spatially invariant structure increases cubically with the number of subsystems, our algorithms exhibit a linear growth. After having identified a controller structure, the structured design step optimizes the network performance over the identified structure.

Our presentation is organized as follows. In Section 2, we provide motivating examples and formulate the generalized sparsity-promoting optimal control problem that we study in this paper. In Section 3, we identify a class of convex problems that can be cast as semidefinite programs. In Section 4, we leverage the alternating direction method of multipliers algorithm to exploit the structure of the corresponding objective functions and solve the regularized optimal control problem. In Section 5, we illustrate our developments using a synchronization network. We conclude the paper in Section 6.

## 2. Motivation and background

We consider a class of control problems

$$\begin{aligned}\dot{\hat{x}} &= \hat{A}\hat{x} + \hat{B}_1\hat{d} + \hat{B}_2\hat{u} \\ \hat{z} &= \hat{C}_1\hat{x} + \hat{D}\hat{u} \\ \hat{y} &= \hat{C}_2\hat{x} \\ \hat{u} &= -\hat{K}\hat{y}\end{aligned}\quad (1)$$

where  $\hat{x}$  is the state,  $\hat{d}$  and  $\hat{u}$  are the disturbance and control inputs,  $\hat{z}$  is the performance output, and  $\hat{y}$  is the measured output. The matrices  $\hat{C}_1$  and  $\hat{D}$  are given by  $[\hat{Q}^{1/2} \ 0]^*$  and  $[0 \ \hat{R}^{1/2}]^*$  with standard assumptions on stabilizability and detectability of pairs  $(\hat{A}, \hat{B}_2)$  and  $(\hat{A}, \hat{Q}^{1/2})$ . Here,  $(\cdot)^*$  denotes complex-conjugate transpose of a given matrix. The matrices  $\hat{Q} = \hat{Q}^* \geq 0$  and  $\hat{R} = \hat{R}^* > 0$  are the state and control performance weights, and the closed-loop system is given by

$$\begin{aligned}\dot{\hat{x}} &= (\hat{A} - \hat{B}_2\hat{K}\hat{C}_2)\hat{x} + \hat{B}_1\hat{d} \\ \hat{z} &= \begin{bmatrix} \hat{Q}^{1/2} \\ -\hat{R}^{1/2}\hat{K}\hat{C}_2 \end{bmatrix}\hat{x}.\end{aligned}\quad (2)$$

We assume that there is a stabilizing feedback gain matrix  $\hat{K}$ .

Our objective is to achieve a desired tradeoff between the  $\mathcal{H}_2$  performance of system (2) and the sparsity of a matrix that is related to the feedback gain matrix  $\hat{K}$  through a linear transformation  $\mathcal{T}(\hat{K})$ . To address this challenge we consider a regularized optimal control problem

$$\underset{\hat{K}}{\text{minimize}} \quad J(\hat{K}) + \gamma g(\mathcal{T}(\hat{K})) \quad (3)$$

where  $J(\hat{K})$  is the  $\mathcal{H}_2$  norm of system (2),  $\gamma$  is a positive regularization parameter, and  $g(\mathcal{T}(\hat{K}))$  is a sparsity-promoting regularization term (see Section 2.3 for details).

Linear transformation  $\mathcal{T}(\hat{K})$  of the feedback gain  $\hat{K}$  in (3) reflects the fact that *sparsity should be enforced in a specific set of coordinates*. This characterization is more general than the one considered in [3–5] where the sparsity-promoting optimal control was originally introduced and algorithms were developed. In contrast to [3–5], where it was assumed that the state-space model is given in physically meaningful coordinates, herein we only require that the states in (2) are related to these coordinates via a linear transformation  $\mathcal{T}$ . One such example arises in spatially invariant systems where the “spatial frequency” domain is convenient for minimizing quadratic performance objective [1], whereas sparsity requirements are naturally expressed in the physical domain. Another class of problems is given by consensus and synchronization networks where the absence of absolute measurements confines standard control-theoretic requirements to a subspace of the original state-space.

### 2.1. Problem formulation

As mentioned earlier, while it is convenient to formulate minimization of the quadratic performance index in terms of the feedback gain  $\hat{K}$ , it may be desirable to promote sparsity in a different set of coordinates. By introducing an additional optimization variable  $K$ , we bring (3) into the following form,

$$\begin{aligned}\underset{\hat{K}, K}{\text{minimize}} \quad & J(\hat{K}) + \gamma g(K) \\ \text{subject to} \quad & \mathcal{T}(\hat{K}) - K = 0,\end{aligned}\quad (4a)$$

where  $g(K)$  is a sparsity-promoting regularization term and  $\mathcal{T}$  is a linear operator. In the  $\mathcal{H}_2$  setting,  $J(\hat{K})$  is given by

$$J(\hat{K}) := \begin{cases} \text{trace} \left( (\hat{Q} + \hat{C}_2^* \hat{K}^* \hat{R} \hat{K} \hat{C}_2) \hat{X} \right), & \hat{K} \text{ stabilizing} \\ \infty, & \text{otherwise} \end{cases} \quad (4b)$$

where the closed-loop controllability Gramian  $\hat{X}$  satisfies the Lyapunov equation

$$(\hat{A} - \hat{B}_2\hat{K}\hat{C}_2)\hat{X} + \hat{X}(\hat{A} - \hat{B}_2\hat{K}\hat{C}_2)^* + \hat{B}_1\hat{B}_1^* = 0. \quad (4c)$$

Clearly, for any feasible  $\hat{K}$  and  $K$ , the optimal control problems (3) and (4a) are equivalent. We note that the linear constraint in (4a) is more general than the constraint considered in [3–5], where  $\hat{K} - K = 0$ . This introduces additional freedom in control design and broadens applicability of the developed tools.

In the set of coordinates where it is desired to promote sparsity, the closed-loop system takes the form

$$\begin{aligned}\dot{x} &= (A - B_2 K C_2)x + B_1 d \\ z &= \begin{bmatrix} Q^{1/2} \\ -R^{1/2} K C_2 \end{bmatrix} x,\end{aligned}\quad (5)$$

where  $K = \mathcal{T}(\hat{K})$ .

### 2.2. Examples

We next discuss several classes of problems that are encountered in applications. For each of these, the optimal control problem can be brought into the form (4) via a suitable change of coordinates.

### 2.2.1. Consensus and synchronization networks

Consensus and synchronization problems are of increasing importance in applications ranging from biology to computer science to power systems [7–21]. In each of these, it is of interest to reach an agreement or to achieve synchronization between the nodes in the network.

In consensus and synchronization networks with the state vector

$$x := [p^* \ q^*]^* \in \mathbb{R}^n$$

only relative differences between the components of the vector  $p(t) \in \mathbb{R}^N$  are allowed to enter into (5). This requirement imposes structural constraints on the matrices in (5), which are partitioned conformably with the partition of the state vector  $x$ ,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B_i = \begin{bmatrix} B_{ip} \\ B_{iq} \end{bmatrix}, \quad (6)$$

$$Q = \begin{bmatrix} Q_p & 0 \\ 0 & Q_q \end{bmatrix}, \quad K = [K_p \ K_q].$$

For  $C_2 = I$ , the restriction on the absence of the access to the absolute measurements of the components of the vector  $p$  translates into the following requirements

$$A_{11} \mathbb{1} = 0, \quad A_{21} \mathbb{1} = 0, \quad Q_p \mathbb{1} = 0, \quad K_p \mathbb{1} = 0 \quad (7)$$

where  $\mathbb{1}$  is the vector of all ones. Under these conditions, the closed-loop system (5) has an eigenvalue at zero and the corresponding eigenvector  $[\mathbb{1}^* \ 0^*]^*$  is associated with the average of the vector  $p$ ,  $\bar{p} := (1/N) \mathbb{1}^* p$ . If the pairs  $(A, B_2)$  and  $(A, Q^{1/2})$  are stabilizable and detectable on the subspace  $\mathcal{S}$ ,

$$\mathcal{S} := \begin{bmatrix} \mathbb{1} \\ 0 \end{bmatrix}^\perp = \begin{bmatrix} \mathbb{1}^\perp \\ \mathbb{R}^{n-N} \end{bmatrix}$$

a coordinate transformation  $\hat{x} := Tx$  can be introduced to eliminate the average mode  $\bar{p}$  from (5).

To achieve the goal of eliminating the average mode,  $\bar{p} := (1/N) \mathbb{1}^* p$ , we introduce the following coordinate transformation

$$\underbrace{\begin{bmatrix} p \\ q \end{bmatrix}}_x = \underbrace{\begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix}}_{T^+} \underbrace{\begin{bmatrix} \psi \\ q \end{bmatrix}}_{\hat{x}} + \underbrace{\begin{bmatrix} \mathbb{1} \\ 0 \end{bmatrix}}_{\bar{p}}$$

where the columns of the matrix  $U \in \mathbb{R}^{N \times (N-1)}$  form an orthonormal basis for the subspace  $\mathbb{1}^\perp$ . For example, the columns of  $U$  can be obtained from the  $(N-1)$  eigenvectors of the matrix  $Q_p$  corresponding to the non-zero eigenvalues. Using properties of the matrix  $U$

$$U^* U = I, \quad U U^* = I - (1/N) \mathbb{1} \mathbb{1}^*, \quad U^* \mathbb{1} = 0,$$

we equivalently have

$$\underbrace{\begin{bmatrix} \psi \\ q \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} U^* & 0 \\ 0 & I \end{bmatrix}}_T \underbrace{\begin{bmatrix} p \\ q \end{bmatrix}}_x.$$

This change of coordinates brings the closed-loop system (5) into the form (2) which does not contain the average mode  $\bar{p}$ . The matrices in (2) are given by

$$\hat{A} := T A T^+, \quad \hat{B}_i := T B_i, \quad \hat{C}_2 := C_2 T^+$$

$$\hat{Q} := T^{+*} Q T^+, \quad \hat{R} := R$$

with  $\hat{u} = u$ ,  $\hat{d} = d$ ,  $\hat{z} = z$ . Finally, we note that the feedback gain matrices  $\hat{K}$  and  $K$  are related by the transformation matrix  $T$

$$K = \mathcal{T}(\hat{K}) = \hat{K} T \Leftrightarrow \hat{K} = K T^+,$$

which has the right inverse  $T^+, T T^+ = I$ . In consensus and synchronization networks, the rows of the matrix  $T$  form an orthonormal basis and we thus have  $T^+ = T^*$ .

We next provide particular examples that can be described by (2) and (5) with structural constraints (7).

**Swing equation.** In power networks, the swing equation is used to characterize energy exchange between generators [23]. After linearization around a stationary operating point, the swing equation reduces to

$$M \ddot{p} + D \dot{p} + L_p p = d + u \quad (8)$$

where  $p$  is the vector of rotor angles,  $M$  and  $D$  are diagonal matrices of inertia and damping coefficients, and  $L_p$  is the Laplacian matrix that describes the interaction topology. By setting  $q := \dot{p}$ , (8) is brought into the state-space form (5)–(6) with

$$A = \begin{bmatrix} 0 & I \\ -M^{-1} L_p & -M^{-1} D \end{bmatrix}, \quad B_i = \begin{bmatrix} 0 \\ M^{-1} \end{bmatrix}. \quad (9)$$

Since  $L_p \mathbb{1} = 0$ , the structural restrictions (7) are satisfied if  $Q_p \mathbb{1} = 0$  and  $K_p \mathbb{1} = 0$ .

**Single-integrator consensus networks.** Networks in which each node updates a scalar variable  $p_i$  using relative information exchange with its neighbors can be obtained from (8) by setting the matrix  $M$  to zero; e.g., see [7]. In this case, the matrices in the state-space model (5)–(6) simplify to  $A = -D^{-1} L_p$  and  $B_1 = B_2 = D^{-1}$ .

**Power systems.** Models of power networks account for the dynamics of generators, control devices, and algebraic equations that describe load flow, stators, and electronic circuits. Control actions are typically executed using generator excitation via power system stabilizers (PSS), governor control, or power electronics (FACTS). In addition to the rotor angles  $p$  and frequencies  $v := \dot{p}$ , additional states  $r$  that account for fast electrical devices are needed to describe the dynamics of the entire system. After linearization at a stationary operating point, the state-space model can be written in the form (5)–(6) by defining  $q := [v^* \ r^*]^*$  with

$$A_{11} := 0, \quad A_{12} := [I \ 0],$$

$$A_{21} := \begin{bmatrix} -M^{-1} L_p \\ A_{rp} \end{bmatrix}, \quad A_{22} := \begin{bmatrix} -M^{-1} D & A_{qr} \\ A_{rq} & A_{rr} \end{bmatrix}.$$

Since only differences between rotor angles of different generators enter into the original nonlinear differential equations, this property is shared by the linearized set of equations, thereby implying  $A_{21} \mathbb{1} = 0$ . Furthermore, in the absence of the access to the absolute rotor angle measurements both the matrix  $A$  in (6) and its closed-loop equivalent in (5) have an eigenvalue at zero with the corresponding eigenvector  $[\mathbb{1}^* \ 0^*]^*$ . Such formulation has been recently utilized in [21].

### 2.2.2. Spatially-invariant systems

For systems with invariances and symmetries, transform techniques can be used to bring a large-scale analysis and design problems into a parametrized family of smaller problems. One such class is given by spatially invariant systems that evolve over a discrete spatially-periodic domain (e.g., a one-dimensional circle or a multi-dimensional torus). In this case, the matrices in (5) are block circulant matrices and the application of the discrete Fourier transform (DFT) in the spatially invariant directions brings them into a block-diagonal form (2). As shown in [1], the optimal centralized controllers for spatially invariant systems with quadratic performance indices are also spatially invariant; thus, in the transformed domain they also take the block-diagonal form. Consequently, determining the optimal centralized controller amounts to easily parallelizable task of solving a sequence of smaller, fully-decoupled optimal control problems. While there are no guarantees that

structured optimal controllers for spatially-invariant systems are also spatially invariant, we show that significant computational advantage can be gained by restricting synthesis to a class of spatially-invariant controllers.

For spatially-invariant systems (5) with block-circulant matrices, the application of DFT

$$\hat{x} = T x, \quad \hat{u} = T u, \quad \hat{d} = T d, \quad \hat{z} = T z,$$

brings the closed-loop system (5) to the form (2) with block-diagonal matrices  $\hat{A}$ ,  $\hat{B}_1$ ,  $\hat{B}_2$ ,  $\hat{C}_2$ ,  $\hat{Q}$ ,  $\hat{R}$ , and  $\hat{K}$ . Here,  $T$  is the discrete Fourier matrix and the feedback gain matrices are related via a linear transformation [24],

$$K = \mathcal{T}(\hat{K}) = T^* \hat{K} T.$$

### 2.3. Sparsity-promoting penalty functions

We briefly describe two classes of sparsity-promoting penalty functions. More sophisticated penalties can also be introduced; see [21] for examples in power networks.

**Elementwise sparsity.** The weighted  $\ell_1$ -norm,

$$g(K) := \sum_{i,j} W_{ij} |K_{ij}| \quad (10)$$

is a commonly used proxy for enhancing elementwise sparsity of the matrix  $K$  [25]. The non-negative weights  $W_{ij}$  provide additional flexibility relative to the standard  $\ell_1$ -regularization. An iterative reweighting method was introduced in [25] to provide better approximation of the non-convex cardinality function. In the  $m$ th iteration, the weights  $W_{ij}$  are set to be inversely proportional to the absolute value of  $K_{ij}$  in the previous iteration,

$$W_{ij}^m = 1/(|K_{ij}^{m-1}| + \epsilon)$$

where  $0 < \epsilon \ll 1$  guards against  $K_{ij} = 0$ .

**Block sparsity.** By selecting  $g(K)$  to penalize the Frobenius norm of the  $ij$ th block of the matrix  $K$ ,

$$g(K) := \sum_{i,j} W_{ij} \|K_{ij}\|_F$$

sparsity can be enhanced at the level of submatrices. In the iterative reweighting algorithm, the absolute value should be replaced by the Frobenius norm of  $K_{ij}$ .

### 3. Class of convex problems

For an undirected consensus network in which each node updates a scalar value  $p_i$ , we next show that the sparsity-promoting optimal control problem can be formulated as an SDP. The closed-loop system (5) with

$$A := -L_p, \quad B_1 = B_2 := I, \quad C_2 = I, \quad K := L_k$$

can be written as

$$\begin{aligned} \dot{p} &= -(L_p + L_k)p + d \\ z &= \begin{bmatrix} Q^{1/2} \\ -R^{1/2} L_k \end{bmatrix} p \end{aligned} \quad (11)$$

where the symmetric positive semi-definite matrices  $L_p$  and  $L_k$  satisfy  $L_p \mathbf{1} = 0$ ,  $L_k \mathbf{1} = 0$ . These two Laplacian matrices contain information about the interconnection structure of the open-loop system and the controller.

The  $\ell_1$ -regularized  $\mathcal{H}_2$  optimal control problem can be formulated as

$$\underset{L_k}{\text{minimize}} \quad J(L_k) + \gamma \|W \circ L_k\|_{\ell_1}. \quad (12)$$

Here,  $\circ$  denotes elementwise matrix multiplication and the solution to the algebraic Lyapunov equation

$$(L_p + L_k)P + P(L_p + L_k) = Q + L_k R L_k$$

determines the  $\mathcal{H}_2$  of the closed-loop system,  $J(L_k) = \text{trace}(P)$ . It is readily shown that the stability of (11) on the subspace  $\mathbf{1}^\perp$  amounts to positive-definiteness of the matrix  $(L_p + L_k)$  on  $\mathbf{1}^\perp$ . Under this condition, we can rewrite  $J(L_k)$  as

$$\begin{aligned} J(L_k) &= \text{trace} \left( (L_p + L_k)^\dagger (Q + L_k R L_k) \right) \\ &= \frac{1}{2} \text{trace} \left( \left( L_p + L_k + \frac{1}{N} \mathbf{1} \mathbf{1}^T \right)^{-1} (Q + L_k R L_k) \right) \end{aligned}$$

where  $(L_p + L_k)^\dagger$  denotes the Moore–Penrose pseudoinverse of  $(L_p + L_k)$ , and cast the sparsity-promoting optimal control problem (12) as an SDP via the Schur complement,

$$\begin{aligned} &\underset{Y, Z, L_k}{\text{minimize}} \quad \frac{1}{2} \text{trace}(Y) + \gamma \mathbf{1}^T Z \mathbf{1} \\ &\text{subject to} \quad \begin{bmatrix} Y & \begin{bmatrix} Q^{1/2} \\ R^{1/2} L_k \end{bmatrix} \\ \begin{bmatrix} (\cdot)^* \\ L_p + L_k + \frac{1}{N} \mathbf{1} \mathbf{1}^T \end{bmatrix} & \geq 0 \end{bmatrix} \quad (13) \\ &L_k \mathbf{1} = 0 \\ &-Z \leq W \circ L_k \leq Z. \end{aligned}$$

For small size problems, the resulting SDP formulation can be solved efficiently using available SDP solvers.

In addition to optimal edge design in undirected consensus networks, several other classes of problems admit convex characterizations: a class of optimal synchronization problems [17], optimal actuator/sensor selection [26,27], symmetric modifications of symmetric systems [3,28], and diagonal modifications of positive systems [29].

### 4. Design of controller structure

We next develop an algorithm, based on the Alternating Direction Method of Multipliers (ADMM), to solve the sparsity-promoting optimal control problem (4),

$$\begin{aligned} &\underset{\hat{K}, K}{\text{minimize}} \quad J(\hat{K}) + \gamma g(K) \\ &\text{subject to} \quad \mathcal{T}(\hat{K}) - K = 0. \end{aligned}$$

As we describe next, the introduction of the linear constraint in (4a) in conjunction with utilization of the ADMM algorithm allows us to exploit the respective structures of the objective functions  $J$  and  $g$  in (4).

#### 4.1. Structure design via ADMM

The structure of feedback gains that strike a balance between quadratic performance of the system and sparsity of the controller is designed via ADMM. The ADMM algorithm starts by introducing the augmented Lagrangian

$$\begin{aligned} \mathcal{L}_\rho(\hat{K}, K, \Lambda) &= J(\hat{K}) + \gamma g(K) + \langle \Lambda, \mathcal{T}(\hat{K}) - K \rangle \\ &\quad + \frac{\rho}{2} \langle \mathcal{T}(\hat{K}) - K, \mathcal{T}(\hat{K}) - K \rangle \end{aligned}$$

where  $\Lambda$  is the Lagrange multiplier,  $\rho$  is a positive scalar, and  $\langle \cdot, \cdot \rangle$  is the standard inner product between two matrices. Instead of minimizing the augmented Lagrangian jointly with respect to  $\hat{K}$  and  $K$ , ADMM uses a sequence of iterations [22],

$$\begin{aligned} \hat{K}^{k+1} &= \underset{\hat{K}}{\text{argmin}} \quad \mathcal{L}_\rho(\hat{K}, K^k, \Lambda^k) \\ K^{k+1} &= \underset{K}{\text{argmin}} \quad \mathcal{L}_\rho(\hat{K}^{k+1}, K, \Lambda^k) \\ \Lambda^{k+1} &= \Lambda^k + \rho (\mathcal{T}(\hat{K}^{k+1}) - K^{k+1}) \end{aligned}$$

until primal and dual residuals are smaller than specified thresholds,

$$\|\mathcal{T}(\hat{K}^{k+1}) - K^{k+1}\|_F \leq \epsilon_p, \quad \|K^{k+1} - K^k\|_F \leq \epsilon_d.$$

It is readily shown that  $\hat{K}$ -minimization step amounts to the quadratically-augmented minimization of  $J(\hat{K})$ ,

$$\hat{K}^{k+1} := \operatorname{argmin}_{\hat{K}} \left( J(\hat{K}) + \frac{\rho}{2} \|\mathcal{T}(\hat{K}) - H^k\|_F^2 \right)$$

where  $H^k := K^k - (1/\rho) \Lambda^k$ . Similarly, using completion of squares, the  $K$ -minimization problem can be brought into the following form

$$K^{k+1} := \operatorname{argmin}_K \left( \gamma g(K) + \frac{\rho}{2} \|K - V^k\|_F^2 \right)$$

with  $V^k := \mathcal{T}(\hat{K}^{k+1}) + (1/\rho) \Lambda^k$ . Thus, updating  $K$  requires computation of the proximal operator of the function  $g$ .

#### 4.1.1. $K$ -minimization step

For elementwise sparsity, the objective function in the  $K$ -minimization step takes separable form,

$$\sum_{i,j} \left( \gamma W_{ij} |K_{ij}| + \frac{\rho}{2} (K_{ij} - V_{ij}^k)^2 \right),$$

and the update of  $K$  is obtained via convenient use of the soft-thresholding operator,

$$K_{ij}^{k+1} = \begin{cases} (1 - a/|V_{ij}^k|) V_{ij}^k & |V_{ij}^k| > a \\ 0 & |V_{ij}^k| \leq a \end{cases}$$

where  $a := (\gamma/\rho) W_{ij}$ . This analytical update of  $K$  is independent of the quadratic performance index  $J$ . Similarly, for block sparsity, the minimizer is determined by

$$K_{ij}^{k+1} = \begin{cases} (1 - a/\|V_{ij}^k\|_F) V_{ij}^k & \|V_{ij}^k\|_F > a \\ 0 & \|V_{ij}^k\|_F \leq a \end{cases}$$

where  $K_{ij}$  and  $V_{ij}$  are the corresponding submatrices.

#### 4.1.2. $\hat{K}$ -minimization step

Finding the solution to the  $\hat{K}$ -minimization problem represents the biggest challenge to solving the sparsity-promoting optimal control problem (4) via ADMM. In what follows, we introduce two methods to solve the  $\hat{K}$ -minimization problem: the Anderson–Moore method and the proximal gradient method.

**Anderson–Moore method.** For the  $\mathcal{H}_2$  optimal control problem (4), the optimality conditions in the  $\hat{K}$ -minimization step are given by

$$(\hat{A} - \hat{B}_2 \hat{K} \hat{C}_2) \hat{X} + \hat{X} (\hat{A} - \hat{B}_2 \hat{K} \hat{C}_2)^* = -\hat{B}_1 \hat{B}_1^* \quad (\text{NC-X})$$

$$(\hat{A} - \hat{B}_2 \hat{K} \hat{C}_2)^* \hat{P} + \hat{P} (\hat{A} - \hat{B}_2 \hat{K} \hat{C}_2) = -(\hat{Q} + \hat{C}_2^* \hat{K}^* \hat{R} \hat{C}_2) \quad (\text{NC-P})$$

$$2(\hat{R} \hat{K} \hat{C}_2 - \hat{B}_2^* \hat{P}) \hat{X} \hat{C}_2^* + \rho \mathcal{T}^\dagger(\mathcal{T}(\hat{K}) - H^k) = 0 \quad (\text{NC-K})$$

where  $\mathcal{T}^\dagger$  is the adjoint of the operator  $\mathcal{T}$ ,

$$\langle K, \mathcal{T}(\hat{K}) \rangle = \langle \mathcal{T}^\dagger(K), \hat{K} \rangle.$$

The unknowns in this system of nonlinear matrix-valued equations are the feedback gain  $\hat{K}$  as well as the controllability and observability Gramians  $\hat{X}$  and  $\hat{P}$  of the closed-loop system (2). These equations can have multiple solutions, each of which is a stationary point of the  $\hat{K}$ -minimization problem. In general, it is not known how many stationary points exist or how to find all of them.

The Anderson–Moore method solves the above system of equations in an iterative fashion. In each iteration, the algorithm starts with a stabilizing feedback matrix  $\hat{K}$  and solves two Lyapunov

equations and one Sylvester equation. Specifically, it first solves (NC-X) and (NC-P) for controllability and observability Gramians  $\hat{X}$  and  $\hat{P}$  with  $\hat{K}$  being fixed. Then the Sylvester equation (NC-K) is solved for  $\hat{K}$  with  $\hat{X}$  and  $\hat{P}$  being fixed.

For consensus and synchronization problems discussed in Section 2.2.1, we have

$$K = \mathcal{T}(\hat{K}) = \hat{K} T \Leftrightarrow \hat{K} = \mathcal{T}^\dagger(K) = K T^+$$

with  $T T^+ = I$ . If the control weight  $\hat{R}$  is given by a scaled version of the identity matrix

$$\hat{R} = r I, \quad r > 0.$$

Sylvester equation (NC-K) can be explicitly solved for  $\hat{K}$ ,

$$\hat{K} = \left( 2 \hat{B}_2^* \hat{P} \hat{X} \hat{C}_2^* + \rho H^k T^+ \right) \left( 2 r \hat{C}_2 \hat{X} \hat{C}_2^* + \rho I \right)^{-1}.$$

Following [5], we can show that the difference between two consecutive updates of  $\hat{K}$  forms a descent direction for

$$L(\hat{K}) := J(\hat{K}) + \frac{\rho}{2} \|\mathcal{T}(\hat{K}) - H^k\|_F^2.$$

In conjunction with backtracking, this can be used to determine step-size to guarantee closed-loop stability and convergence to a stationary point of  $L(\hat{K})$ .

**Proximal gradient method.** Proximal gradient method provides an alternative approach to solving the  $\hat{K}$ -minimization step. It is based on a simple quadratic approximation of the quadratic objective function  $J(\hat{K})$  around current inner iterate  $\hat{K}^m$ ,

$$J(\hat{K}) \approx J(\hat{K}^m) + \langle \nabla J(\hat{K}^m), \hat{K} - \hat{K}^m \rangle + \frac{1}{2\alpha_m} \|\hat{K} - \hat{K}^m\|_F^2$$

where  $\alpha_m$  denotes the step-size and

$$\nabla J(\hat{K}^m) = 2(\hat{R} \hat{K}^m \hat{C}_2 - \hat{B}_2^* \hat{P}^m) \hat{X}^m \hat{C}_2^*.$$

Using completion of squares, the  $\hat{K}$ -minimization step can be written as

$$\hat{K}^{m+1} = \operatorname{argmin}_{\hat{K}} \left( \frac{1}{2\alpha_m} \left\| \hat{K} - \left( \hat{K}^m - \alpha_m \nabla J(\hat{K}^m) \right) \right\|_F^2 + \frac{\rho}{2} \|\mathcal{T}(\hat{K}) - H^k\|_F^2 \right)$$

and the optimality condition is given by

$$\frac{1}{\alpha_m} \left( \hat{K} - \left( \hat{K}^m - \alpha_m \nabla J(\hat{K}^m) \right) \right) + \rho \mathcal{T}^\dagger(\mathcal{T}(\hat{K}) - H^k) = 0. \quad (14)$$

For consensus and synchronization networks,  $\mathcal{T}(\hat{K}) = \hat{K} T$ , and we have an explicit update for  $\hat{K}$ ,

$$\hat{K}^{m+1} = \frac{1}{1 + \alpha_m \rho} \left( \hat{K}^m - \alpha_m \nabla J(\hat{K}^m) + \alpha_m \rho H^k T^+ \right).$$

The proximal gradient algorithm converge with rate  $O(1/m)$  if  $\alpha_m$  is smaller than the reciprocal of the Lipschitz constant of  $\nabla J$  [30]. Since the Lipschitz constant is difficult to determine explicitly, we adjust  $\alpha_m$  via backtracking procedure that we describe next. Furthermore, to enhance the speed of convergence, we initialize the step-size using the Barzilai–Borwein (BB) method [31],

$$\alpha_m^0 = \frac{\|\hat{K}^m - \hat{K}^{m-1}\|_F^2}{\langle \hat{K}^{m-1} - \hat{K}^m, \nabla J(\hat{K}^{m-1}) - \nabla J(\hat{K}^m) \rangle}.$$

The BB method approximates the Hessian with a scaled version of the identity matrix and it represents an effective heuristics for improving convergence rate. The initial step-size  $\alpha_m^0$  is adjusted via

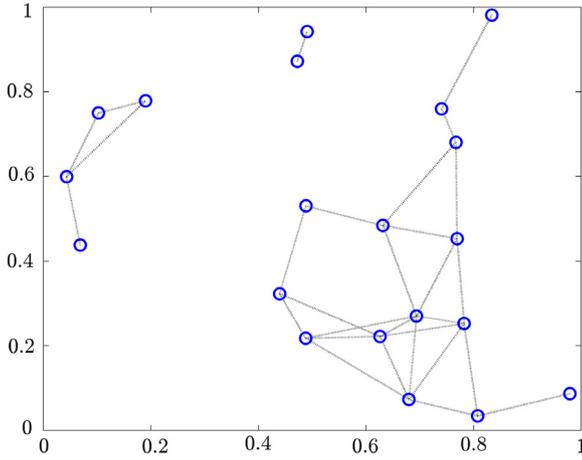


Fig. 1. Topology of a disconnected plant network with 3 clusters and 20 nodes.

backtracking to guarantee closed-loop stability and to make sure that,

$$J(\hat{K}^{m+1}) \leq J(\hat{K}^m) + \langle \nabla J(\hat{K}^m), \hat{K}^{m+1} - \hat{K}^m \rangle + \frac{1}{2\alpha_m} \|\hat{K}^{m+1} - \hat{K}^m\|_F^2.$$

The proximal gradient method terminates when

$$\|\nabla L(\hat{K}^m)\|_F \leq \epsilon_{\hat{K}}.$$

**Remark 1.** For spatially-invariant systems, the computational complexity of each  $\hat{K}$ -minimization step is  $O(Nn_s^3)$ . Here,  $N$  denotes the number of subsystems and  $n_s$  is the number of states in each subsystem. This should be compared and contrasted to  $O(n^3)$  complexity, with  $n = Nn_s$ , for systems without spatially-invariant structure.

**Polishing step.** After having identified the sparsity pattern  $S_p$  via ADMM, we optimize the network performance over the identified structure,

$$\begin{aligned} & \underset{\hat{K}}{\text{minimize}} && J(\hat{K}) \\ & \text{subject to} && \mathcal{T}(\hat{K}) \in S_p. \end{aligned} \quad (15)$$

We fix the sparsity pattern of  $\hat{K}$  and solve the optimal control problem (15) via an ADMM algorithm, where the  $\hat{K}$ -minimization step is the same as in Section 4.1.2 while the  $K$ -minimization step is computed by projecting the new  $K = \hat{K}T$  onto the convex set  $S_p$ . This polishing step is used to further improve performance of sparse feedback gains resulting from the structure design step.

## 5. Case study: Synchronization network

Twenty nodes in an undirected disconnected network shown in Fig. 1 are randomly distributed in a unit square. The nodes form three clusters and the network dynamics are described by the swing equation (8). The state-space model is given by (5) with  $C_2 = I$  and the matrices  $A$ ,  $B_1$ , and  $B_2$  determined by (9). The graph Laplacian  $L_p$  is obtained based on the proximity of the nodes: two nodes are connected if their distance is not greater than 0.25. The control objective is to minimize performance metric that penalizes angular kinetic energy and the mean square deviation from the angle average.

Communication links in the controller graph that result from elementwise sparsity-promoting algorithm with iterative re-weighting are illustrated in Fig. 2. Since local frequency measurements are readily available, the diagonal elements of the frequency

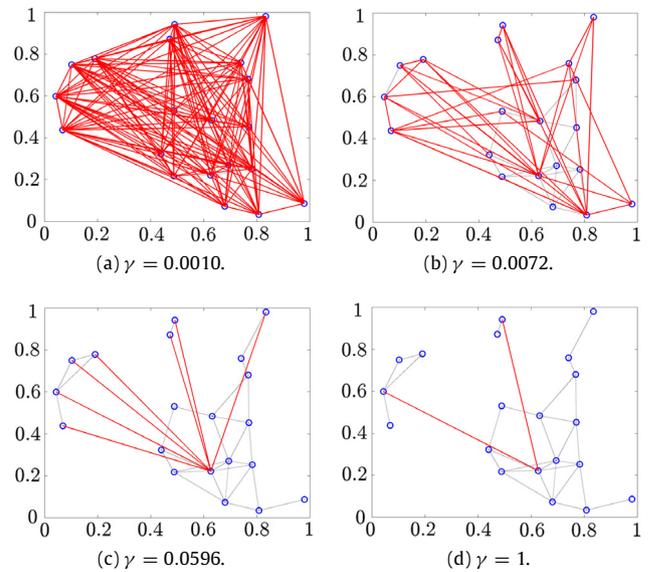


Fig. 2. Topology of controller network for different values of  $\gamma$ . Edges in the controller network are marked with red lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

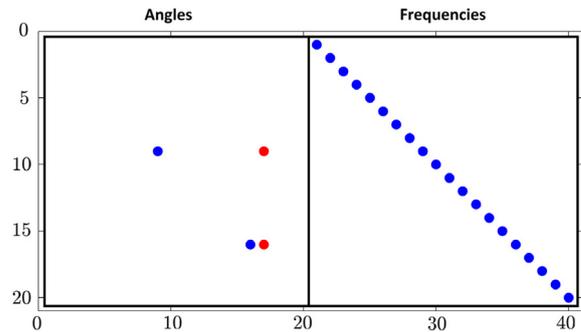


Fig. 3. Sparsity pattern of  $K$  for  $\gamma = 1$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

feedback gains are not penalized in the function  $g$ . The red lines mark the identified communication links (of either the rotational angles or the frequencies) between the nodes. As we increase  $\gamma$ , the controller graph becomes sparser. For  $\gamma = 1$ , there are only two long-range links that connect two small clusters to the large cluster of nodes. The controller connects the originally disconnected graph by adding links between different clusters, thereby guaranteeing synchronization and optimizing the desired performance metric.

The structure of the feedback gain  $K$  for  $\gamma = 1$  is shown in Fig. 3. The blue dots denote local feedback gains and the red dots identify information that needs to be communicated between different nodes. Since the frequency feedback gain matrix is diagonal, only local frequency measurements are required to form the control action. The two red dots correspond to the two red long-range links in Fig. 2(d) and they indicate that the controllers of the two furthest nodes require access to the angle measurements of the node in the center of the domain. Dropping either of these two links would yield a disconnected closed-loop network and synchronization would not be achieved.

Compared to the optimal centralized controller, the sparse controller with structure shown in Fig. 2(d) compromises performance by 8.81%; see Fig. 4. In contrast, the sparse controller in Fig. 2(c) has five additional long-range links and degrades performance by only 3.4%.

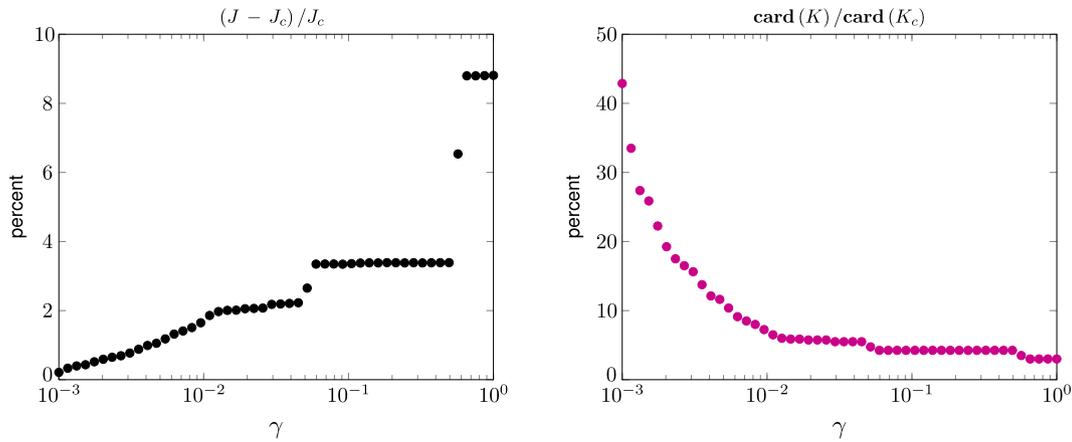


Fig. 4. Performance vs sparsity comparison with respect to the optimal centralized controller  $K_c$  for 50 logarithmically-spaced points  $\gamma \in [10^{-3}, 1]$ .

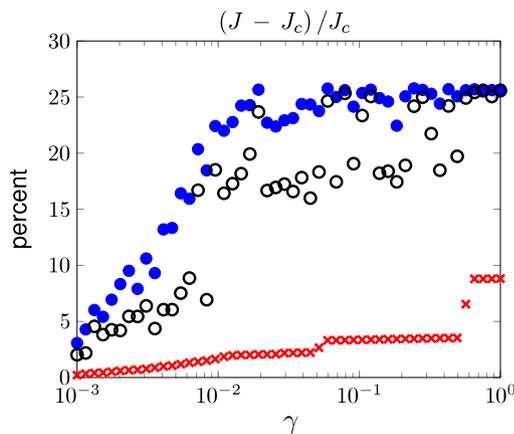


Fig. 5. Performance degradation comparison of  $K$  resulting from our framework ( $\times$ ) to the average of 100 feedback matrices ( $\bullet$ ) of random sparsity patterns with same sparsity level for each  $\gamma$ . The best closed-loop performance achieved by the heuristic method is marked by black hollow circles ( $\circ$ ).

Finally, for each value of the regularization parameter  $\gamma$  we compare and contrast performance of the controller with the sparsity pattern shown in Fig. 3 with a heuristic strategy that has the same sparsity level. The diagonal elements of the frequency feedback gain are always non-zero and sparsity pattern for the rest of the elements is randomly selected. For each sparsity level, we randomly select 100 off-diagonal patterns and optimize feedback gains over the fixed structure. Fig. 5 compares the average closed-loop performance degradation of this heuristic strategy to our method. For each sparsity level, our approach yields smaller performance loss and offers significant advantages for sparser control architectures (i.e., larger values of  $\gamma$ ).

## 6. Concluding remarks

In this paper, we have considered a regularized version of the standard  $\mathcal{H}_2$  optimal control problem where the regularization term serves as a proxy for inducing sparsity. We consider a class of systems in which state-space representation is not cast in the set of coordinates in which it is desired to enhance sparsity. This setup arises in systems with invariances and symmetries as well as in consensus and synchronization networks. We achieve desired performance with fewer communication links in the controller network by promoting sparsity of the feedback gain matrix. Alternating direction method of multipliers allows for performance

and sparsity requirements to be expressed in different set of coordinates and facilitates efficient computation. A synchronization network is provided as an illustrative example to demonstrate the value of our developments. Recently, our generalized sparsity-promoting optimal control framework was used to suppress inter-area oscillations in power networks [21].

## References

- [1] B. Bamieh, F. Paganini, M.A. Dahleh, Distributed control of spatially invariant systems, *IEEE Trans. Automat. Control* 47 (7) (2002) 1091–1107.
- [2] M. Rotkowitz, S. Lall, A characterization of convex problems in decentralized control, *IEEE Trans. Automat. Control* 51 (2) (2006) 274–286.
- [3] M. Fardad, F. Lin, M.R. Jovanović, Sparsity-promoting optimal control for a class of distributed systems, in: *Proceedings of the 2011 American Control Conference*, 2011, pp. 2050–2055.
- [4] F. Lin, M. Fardad, M.R. Jovanović, Sparse feedback synthesis via the alternating direction method of multipliers, in: *Proceedings of the 2012 American Control Conference*, 2012, pp. 4765–4770.
- [5] F. Lin, M. Fardad, M.R. Jovanović, Design of optimal sparse feedback gains via the alternating direction method of multipliers, *IEEE Trans. Automat. Control* 58 (9) (2013) 2426–2431.
- [6] N. Matni, V. Chandrasekaran, Regularization for design, *IEEE Trans. Automat. Control* 61 (12) (2016) 3991–4006.
- [7] M. Mesbahi, M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*, Princeton University Press, 2010.
- [8] L. Xiao, S. Boyd, S.-J. Kim, Distributed average consensus with least-mean-square deviation, *J. Parallel Distrib. Comput.* 67 (1) (2007) 33–46.
- [9] A. Ghosh, S. Boyd, A. Saberi, Minimizing effective resistance of a graph, *SIAM Rev.* 50 (1) (2008) 37–66.
- [10] W. Ren, Synchronization of coupled harmonic oscillators with local interaction, *Automatica* 44 (2008) 3195–3200.
- [11] D. Zelazo, M. Mesbahi, Edge agreement: Graph-theoretic performance bounds and passivity analysis, *IEEE Trans. Automat. Control* 56 (3) (2011) 544–555.
- [12] B. Bamieh, M.R. Jovanović, P. Mitra, S. Patterson, Coherence in large-scale networks: dimension dependent limitations of local feedback, *IEEE Trans. Automat. Control* 57 (9) (2012) 2235–2249.
- [13] F. Dörfler, F. Bullo, Synchronization and transient stability in power networks and non-uniform Kuramoto oscillators, *SIAM J. Control Optim.* 50 (2012) 1616–1642.
- [14] A. Mauroy, P. Sacré, R.J. Sepulchre, Kick synchronization versus diffusive synchronization, in: *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 7171–7183.
- [15] M. Arcak, Synchronization and pattern formation in diffusively coupled systems, in: *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 7184–7192.
- [16] D. Zelazo, S. Schuler, F. Allgöwer, Performance and design of cycles in consensus networks, *Systems Control Lett.* 62 (1) (2013) 85–96.
- [17] M. Fardad, F. Lin, M.R. Jovanović, Design of optimal sparse interconnection graphs for synchronization of oscillator networks, *IEEE Trans. Automat. Control* 59 (9) (2014) 2457–2462.
- [18] X. Wu, M.R. Jovanović, Sparsity-promoting optimal control of consensus and synchronization networks, in: *Proceedings of the 2014 American Control Conference*, 2014, pp. 2948–2953.

- [19] F. Dörfler, M.R. Jovanović, M. Chertkov, F. Bullo, Sparse and optimal wide-area damping control in power networks, in: *Proceedings of the 2013 American Control Conference*, 2013, pp. 4295–4300.
- [20] F. Dörfler, M.R. Jovanović, M. Chertkov, F. Bullo, Sparsity-promoting optimal wide-area control of power networks, *IEEE Trans. Power Syst.* 29 (5) (2014) 2281–2291.
- [21] X. Wu, F. Dörfler, M.R. Jovanović, Input-output analysis and decentralized optimal control of inter-area oscillations in power systems, *IEEE Trans. Power Syst.* 31 (3) (2016) 2434–2444.
- [22] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.* 3 (1) (2011) 1–124.
- [23] P. Kundur, *Power System Stability and Control*, McGraw-Hill, 1994.
- [24] D.M. Zoltowski, N.K. Dhingra, F. Lin, M.R. Jovanović, Sparsity-promoting optimal control of spatially-invariant systems, in: *Proceedings of the 2014 American Control Conference*, 2014, pp. 1261–1266.
- [25] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted  $\ell_1$  minimization, *J. Fourier Anal. Appl.* 14 (2008) 877–905.
- [26] B. Polyak, M. Khlebnikov, P. Shcherbakov, An LMI approach to structured sparse feedback design in linear control systems, in: *Proceedings of the 2013 European Control Conference*, 2013, pp. 833–838.
- [27] N.K. Dhingra, M.R. Jovanović, Z.Q. Luo, An ADMM algorithm for optimal sensor and actuator selection, in: *Proceedings of the 53rd IEEE Conference on Decision and Control*, 2014, pp. 4039–4044.
- [28] N.K. Dhingra, M.R. Jovanović, Convex synthesis of symmetric modifications to linear systems, in: *Proceedings of the 2015 American Control Conference*, 2015, pp. 3583–3588.
- [29] N.K. Dhingra, M. Colombino, M.R. Jovanović, On the convexity of a class of structured optimal control problems for positive systems, in: *Proceedings of the 2016 European Control Conference*, 2016, pp. 825–830.
- [30] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2 (1) (2009) 183–202.
- [31] J. Barzilai, J.M. Borwein, Two-point step size gradient methods, *IMA J. Numer. Anal.* 8 (1) (1988) 141–148.