Design of Optimal Sparse Feedback Gains via the Alternating Direction Method of Multipliers

Fu Lin, Makan Fardad, and Mihailo R. Jovanović

***Abstract*—We design sparse and block sparse feedback gains that minimize the variance amplification (i.e., the $\mathcal{H}_2$ norm) of distributed systems. Our approach consists of two steps. First, we identify sparsity patterns of feedback gains by incorporating sparsity-promoting penalty functions into the optimal control problem, where the added terms penalize the number of communication links in the distributed controller. Second, we optimize feedback gains subject to structural constraints determined by the identified sparsity patterns. In the first step, the sparsity structure of feedback gains is identified using the alternating direction method of multipliers, which is a powerful algorithm well-suited to large optimization problems. This method alternates between promoting the sparsity of the controller and optimizing the closed-loop performance, which allows us to exploit the structure of the corresponding objective functions. In particular, we take advantage of the separability of the sparsity-promoting penalty functions to decompose the minimization problem into sub-problems that can be solved analytically. Several examples are provided to illustrate the effectiveness of the developed approach.**

***Index Terms*—Alternating direction method of multipliers (ADMM), communication architectures, continuation methods, ℓ_1 minimization, optimization, separable penalty functions, sparsity-promoting optimal control, structured distributed design.**

I. Introduction

We develop methods for the design of *sparse* and *block sparse* feedback gains that minimize the variance amplification of distributed systems. Our approach consists of two steps. The first step, which can be viewed as a *structure identification step*, is aimed at finding sparsity patterns that strike a balance between the $\mathcal{H}_2$ performance and the sparsity of the controller. This is achieved by incorporating *sparsity-promoting* penalty functions into the optimal control problem, where the added sparsity-promoting terms penalize the number of communication links. We consider several sparsity-promoting penalty functions including the cardinality function and its convex relaxations. In the absence of sparsity-promoting terms, the solution to the standard $\mathcal{H}_2$ problem results in centralized controllers with dense feedback gains. By gradually increasing the weight on the sparsity-promoting penalty terms, the optimal feedback gain moves along a parameterized solution path from the centralized to the sparse gain of interest. This weight is increased until the desired balance between performance and sparsity is achieved. In the second step, in order to improve the $\mathcal{H}_2$ performance of the structured controller, we solve an optimal control problem subject to the feedback gain belonging to the identified structure.

We demonstrate that the *alternating direction method of multipliers* (ADMM) [1] provides an effective tool for the design of sparse distributed controllers whose performance is comparable to the performance of the optimal centralized controller. This method alternates between promoting the sparsity of the feedback gain matrix and optimizing the closed-loop $\mathcal{H}_2$ norm. The advantage of this alternating mechanism is threefold. First, it provides a flexible framework for incorporation of different penalty functions that promote sparsity or block sparsity. Second, it allows us to exploit the *separability* of the sparsity-promoting penalty functions and to decompose the corresponding optimization problems into sub-problems that can be solved *analytically*. These analytical results are immediately applicable to other distributed control problems where sparsity is desired. Finally, it facilitates the use of descent algorithms for $\mathcal{H}_2$ optimization, in which a descent direction can be formed by solving two Lyapunov equations and one Sylvester equation.

The ℓ_1 norm is widely used as a proxy for cardinality minimization in applied statistics, in sparse signal processing, and in machine learning; see [1]–[4]. In the controls community, recent work inspired by similar ideas includes [5]–[7]. In [5], an ℓ_0 induced gain was introduced to quantify the sparsity of the impulse response of a discrete-time system. In [6], the weighted ℓ_1 framework was used to design structured dynamic output feedback controllers subject to a given $\mathcal{H}_\infty$ performance. In [7], an ℓ_1 relaxation method was employed for the problem of adding a fixed number of edges to a consensus network.

Our presentation is organized as follows. We formulate the sparsity-promoting optimal control problem and compare several sparsity-promoting penalty functions in Section II. We present the ADMM algorithm, emphasize the separability of the penalty functions, and provide the analytical solutions to the sub-problems for both sparse and block sparse minimization problems in Section III. Several examples are provided in Section IV to demonstrate the effectiveness of the developed approach. We conclude with a summary of our contributions in Section V.

II. Sparsity-Promoting Optimal Control Problem

Consider the following control problem:

$$\begin{aligned} \dot{x} &= A\,x + B_1\,d + B_2\,u \\ z &= C\,x + D\,u \\ u &= -\,F\,x \end{aligned} \tag{1}$$

where d and u are the disturbance and control inputs, z is the performance output, $C = [\,Q^{1/2} \quad 0\,]^T$, and $D = [\,0 \quad R^{1/2}\,]^T$, with standard assumptions that (A, B_2) is stabilizable and $(A, Q^{1/2})$ is detectable. The matrix F is a state feedback gain, $Q = Q^T \geq 0$ and $R = R^T > 0$ are the state and control performance weights, and the closed-loop system is given by

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

The design of the optimal state feedback gain F, subject to structural constraints that dictate its zero entries, was recently considered by the authors in [8], [9]. Let the subspace $\mathcal{S}$ embody these constraints and let us assume that there exists a stabilizing $F \in \mathcal{S}$. References [8], [9] then search for $F \in \mathcal{S}$ that minimizes the $\mathcal{H}_2$ norm of the transfer function from d to z

$$\begin{aligned} &\text{minimize} \quad J(F) \\ &\text{subject to} \quad F \in \mathcal{S} \end{aligned} \tag{SH2}$$

where

$$J(F) = \begin{cases} \operatorname{trace}\left(B_1^T P(F) B_1\right), & F \text{ stabilizing} \\ \infty, & \text{otherwise.} \end{cases} \tag{3}$$

Manuscript received November 24, 2011; revised July 23, 2012; accepted March 20, 2013. Date of publication April 12, 2013; date of current version August 15, 2013. This paper appeared in part at the American Control Conference, 2012. This work was supported by the National Science Foundation under awards CMMI-0927509 and CMMI-0927720 and under CAREER Award CMMI-0644793. Recommended by Associate Editor G. E. Dullerud.

F. Lin and M. R. Jovanović are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: fu@umn.edu; mihailo@umn.edu).

M. Fardad is with the Department of Electrical Engineering and Computer Science, Syracuse University, NY 13244 USA (e-mail: makan@syr.edu).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2013.2257618

0018-9286 © 2013 IEEE

The matrix $P(F)$ in (3) denotes the closed-loop observability Gramian

$$P(F) = \int_0^\infty \mathrm{e}^{(A-B_2F)^T t}\,(Q + F^T RF)\,\mathrm{e}^{(A-B_2F)t}\,\mathrm{d}t \tag{4}$$

which can be obtained by solving the Lyapunov equation

$$(A - B_2F)^T P + P(A - B_2F) = -\left(Q + F^T RF\right). \tag{5}$$

While the communication architecture of the controller in (SH2) is *a priori* specified, in this note our emphasis shifts to identifying favorable communication structures without any prior assumptions on the sparsity patterns of the matrix F. We propose an optimization framework in which the sparsity of the feedback gain is directly incorporated into the objective function.

Consider the following optimization problem:

$$\text{minimize} \quad J(F) + \gamma\, g_0(F) \tag{6}$$

where

$$g_0(F) = \mathbf{card}\,(F) \tag{7}$$

denotes the cardinality function, i.e., the *number of nonzero elements* of a matrix. In contrast to problem (SH2), no structural constraint is imposed on F; instead, our goal is to promote sparsity of the feedback gain by incorporating the cardinality function into the optimization problem. The positive scalar γ characterizes our emphasis on the sparsity of F; a larger γ encourages a sparser F, while $\gamma = 0$ renders a centralized gain that is the solution of the standard LQR problem. For $\gamma = 0$, the solution to (6) is given by $F_c = R^{-1}B_2^T P$, where P is the unique positive definite solution of the algebraic Riccati equation, $A^T P + PA + Q - PB_2R^{-1}B_2^T P = 0$.

A. Sparsity-Promoting Penalty Functions

Problem (6) is a combinatorial optimization problem whose solution usually requires an intractable combinatorial search. In optimization problems where sparsity is desired, the cardinality function is typically replaced by the ℓ_1 norm of the optimization variable [10, Chapter 6]

$$g_1(F) = \|F\|_{\ell_1} = \sum_{i,j} |F_{ij}|. \tag{8}$$

Recently, a *weighted* ℓ_1 norm was used to enhance sparsity in signal recovery [4]

$$g_2(F) = \sum_{i,j} W_{ij}|F_{ij}| \tag{9}$$

where W_{ij} are non-negative weights. If W_{ij}'s are chosen to be inversely proportional to the magnitude of F_{ij}, i.e., $\{W_{ij} = 1/|F_{ij}|, F_{ij} \neq 0; W_{ij} = 1/\varepsilon, F_{ij} = 0,\ \ 0 < \varepsilon \ll 1\}$, then the weighted ℓ_1 norm and the cardinality function of F coincide, $\sum_{i,j} W_{ij}\,|F_{ij}| = \mathbf{card}\,(F)$. This scheme for the weights, however, cannot be implemented, since the weights depend on the unknown feedback gain. A reweighted algorithm that solves a sequence of weighted ℓ_1 optimization problems in which the weights are determined by the solution of the weighted ℓ_1 problem in the previous iteration was proposed in [4], [11]. This reweighted scheme was recently employed by the authors to design sparse feedback gains for a class of distributed systems [12], [13].

Both the ℓ_1 norm and its weighted version are *convex* relaxations of the cardinality function. On the other hand, we also examine utility of the *nonconvex sum-of-logs* function as a more aggressive means for promoting sparsity [4]

$$g_3(F) = \sum_{i,j} \log\left(1 + \frac{|F_{ij}|}{\varepsilon}\right), \qquad 0 < \varepsilon \ll 1. \tag{10}$$

Remark 1: Design of feedback gains that have block sparse structure can be achieved by promoting sparsity at the level of the *submatrices* instead of at the level of the individual elements. Let the feedback gain F be partitioned into submatrices $F_{ij} \in \mathbb{R}^{m_i \times n_j}$ that need not have the same size. The weighted ℓ_1 norm and the sum-of-logs can be generalized to matrix blocks by replacing the absolute value of F_{ij} in (9) and (10) by the Frobenius norm $\|\cdot\|_F$ of F_{ij}. Similarly, the cardinality function (7) should be replaced by $\sum_{i,j} \mathbf{card}\,(\|F_{ij}\|_F)$, where $\|F_{ij}\|_F$ does not promote sparsity within the F_{ij} block; it instead promotes sparsity at the level of submatrices.

B. Sparsity-Promoting Optimal Control Problem

Our approach to sparsity-promoting feedback design makes use of the above discussed penalty functions. In order to obtain state feedback gains that strike a balance between the quadratic performance and the sparsity of the controller, we consider the following optimal control problem

$$\text{minimize} \quad J(F) + \gamma\, g(F) \tag{SP}$$

where J is the square of the closed-loop $\mathcal{H}_2$ norm (3) and g is a sparsity-promoting penalty function, e.g., given by (7), (8), (9), or (10). When the cardinality function in (7) is replaced by (8), (9), or (10), problem (SP) can be viewed as a relaxation of the combinatorial problem (6)–(7), obtained by approximating the cardinality function with the corresponding penalty functions g.

As the parameter γ varies over $[0, +\infty)$, the solution of (SP) traces the trade-off path between the $\mathcal{H}_2$ performance J and the feedback gain sparsity g. When $\gamma = 0$, the solution is the centralized feedback gain F_c. We then slightly increase γ and employ an iterative algorithm – the alternating direction method of multipliers (ADMM) – initialized by the optimal feedback matrix at the previous γ. The solution of (SP) becomes sparser as γ increases. After a desired level of sparsity is achieved, we fix the sparsity structure and find the optimal structured feedback gain by solving the structured $\mathcal{H}_2$ problem (SH2).

Since the set of stabilizing feedback gains is in general not convex [14] and since the matrix exponential is not necessarily a convex function of its argument [10], J need not be a convex function of F. This makes it difficult to establish convergence to the global minimum of (SP). Even in problems for which we cannot establish the convexity of $J(F)$, our extensive computational experiments suggest that the algorithms developed in Section III provide an effective means for attaining a desired trade-off between the $\mathcal{H}_2$ performance and the sparsity of the controller.

III. Identification of Sparsity-Patterns Via ADMM

Consider the following constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & J(F) + \gamma\, g(G) \\ \text{subject to} \quad & F - G = 0 \end{aligned} \tag{11}$$

which is clearly equivalent to the problem (SP). The augmented Lagrangian associated with the constrained problem (11) is given by

$$\mathcal{L}_\rho(F, G, \Lambda) = J(F) + \gamma\, g(G) + \mathrm{trace}\left(\Lambda^T (F - G)\right) + \left(\frac{\rho}{2}\right)\|F - G\|_F^2$$

where Λ is the dual variable (i.e., the *Lagrange multiplier*), ρ is a positive scalar, and $\|\cdot\|_F$ is the Frobenius norm. By introducing an additional variable G and an additional constraint $F - G = 0$, we have simplified the problem (SP) by decoupling the objective function into two parts that depend on two different variables. As discussed below, this allows us to exploit the structures of J and g.

In order to find a minimizer of the constrained problem (11), the ADMM algorithm uses a sequence of iterations

$$F^{k+1} := \arg\min_{F} \mathcal{L}_\rho(F, G^k, \Lambda^k) \tag{12a}$$

$$G^{k+1} := \arg\min_{G} \mathcal{L}_\rho(F^{k+1}, G, \Lambda^k) \tag{12b}$$

$$\Lambda^{k+1} := \Lambda^k + \rho(F^{k+1} - G^{k+1}) \tag{12c}$$

until $\|F^{k+1} - G^{k+1}\|_F \leq \epsilon$ and $\|G^{k+1} - G^k\|_F \leq \epsilon$. In contrast to the *method of multipliers* [1], in which F and G are *minimized jointly*,

$$(F^{k+1}, G^{k+1}) := \arg\min_{F,\, G} \mathcal{L}_\rho(F, G, \Lambda^k)$$

ADMM consists of an F-minimization step (12a), a G-minimization step (12b), and a dual variable update step (12c). Note that the dual variable update (12c) uses a step-size equal to ρ, which guarantees that one of the dual feasibility conditions is satisfied in each ADMM iteration; see [1, Section 3.3].

ADMM brings two major benefits to the sparsity-promoting optimal control problem (SP):

- *Separability of g.* The penalty function g is *separable* with respect to the *individual* elements of the matrix. In contrast, the closed-loop $\mathcal{H}_2$ norm cannot be decomposed into componentwise functions of the feedback gain. By separating g and J in the minimization of the augmented Lagrangian $\mathcal{L}_\rho$, we can determine *analytically* the solution to the G-minimization problem via decomposition of (12b) into sub-problems that only involve *scalar* variables.
- *Differentiability of J.* The square of the closed-loop $\mathcal{H}_2$ norm J is a *differentiable* function of F [9]; this is in contrast to g which is a *non-differentiable* function. By separating g and J in the minimization of the augmented Lagrangian $\mathcal{L}_\rho$, we can utilize descent algorithms that rely on the differentiability of J to solve the F-minimization problem (12a).

We next provide the analytical expressions for the solutions of the G-minimization problem (12b) in Section III-A, describe a descent method to solve the F-minimization problem (12a) in Section III-B, present Newton's method to solve the structured problem (SH2) in Section III-C, and discuss the convergence of ADMM in Section III-D.

A. Separable Solution to the G-Minimization Problem (12b)

The completion of squares with respect to G in the augmented Lagrangian $\mathcal{L}_\rho$ can be used to show that (12b) is equivalent to

$$\text{minimize} \quad \phi(G) = \gamma\, g(G) + \left(\frac{\rho}{2}\right) \|G - V^k\|_F^2 \tag{13}$$

where $V^k = (1/\rho)\Lambda^k + F^{k+1}$. To simplify notation, we drop the superscript k in V^k throughout this section. Since both g and the square of the Frobenius norm can be written as a summation of componentwise functions of a matrix, we can decompose (13) into sub-problems expressed in terms of the *individual* elements of G. For example, if g is the weighted ℓ_1 norm, then $\phi(G) = \sum_{i,j} \left(\gamma\, W_{ij}\, |G_{ij}| + (\rho/2)(G_{ij} - V_{ij})^2\right)$. This facilitates the conversion of (13) to minimization problems that only involve *scalar* variables G_{ij}. By doing so, the solution of (13) can be determined *analytically* for the weighted ℓ_1 norm, the sum-of-logs, and the cardinality function.

1) Weighted ℓ_1 Norm: The unique solution to (13) is given by the *soft thresholding* operator (e.g., see [1, Section 4.4.3])

$$G_{ij}^\star = \begin{cases} \left(1 - \frac{a}{|V_{ij}|}\right) V_{ij}, & |V_{ij}| > a \\ 0, & |V_{ij}| \leq a \end{cases} \tag{14}$$

where $a = (\gamma/\rho) W_{ij}$. For given V_{ij}, $G_{ij}^\star$ is obtained by moving V_{ij} towards zero with the amount $(\gamma/\rho)W_{ij}$. In particular, $G_{ij}^\star$ is set to zero if $|V_{ij}| \leq (\gamma/\rho)W_{ij}$, implying that a more aggressive scheme for driving $G_{ij}^\star$ to zero can be obtained by increasing γ and W_{ij} and by decreasing ρ.

2) Cardinality Function: The unique solution to (13) is given by the *truncation* operator

$$G_{ij}^\star = \begin{cases} V_{ij}, & |V_{ij}| > b \\ 0, & |V_{ij}| \leq b \end{cases} \tag{15}$$

where $b = \sqrt{2\gamma/\rho}$. For given V_{ij}, $G_{ij}^\star$ is set to V_{ij} if $|V_{ij}| > \sqrt{2\gamma/\rho}$ and to zero if $|V_{ij}| \leq \sqrt{2\gamma/\rho}$.

3) Sum-of-Logs Function: As shown in [15], the solution to (13) is given by

$$G_{ij}^\star = \begin{cases} 0, & \Delta \leq 0 \text{ or } \{\Delta > 0 \text{ and } r^+\} \\ r^+ V_{ij}, & \Delta > 0 \text{ and } r^- \leq 0 \text{ and } 0 < r^+ \leq 1 \\ G^0, & \Delta > 0 \text{ and } 0 \leq r^\pm \leq 1 \end{cases} \tag{16}$$

where

$$\Delta = (|V_{ij}| + \varepsilon)^2 - 4\left(\frac{\gamma}{\rho}\right)$$
$$r^\pm = \frac{|V_{ij}| - \varepsilon \pm \sqrt{\Delta}}{2\,|V_{ij}|} \tag{17}$$

and $G^0 := \arg\min\{\phi_{ij}(r^+ V_{ij}), \phi_{ij}(0)\}$. For fixed ρ and ε, (16) is determined by the value of γ. For small γ, (16) resembles the soft thresholding operator and for large γ, it resembles truncation operator.

Remark 2: In block sparse design, g is determined by

$$\left\{\sum_{i,j} W_{ij}\|G_{ij}\|_F; \sum_{i,j} \mathbf{card}\left(\|G_{ij}\|_F\right); \sum_{i,j} \log\left(1 + \frac{\|G_{ij}\|_F}{\varepsilon}\right)\right\}$$

and the minimizers of (13) are obtained by replacing the absolute value of V_{ij} in (14), (15), and (17) with the Frobenius norm $\|\cdot\|_F$ of the corresponding block submatrix V_{ij}.

B. Anderson-Moore Method for the F-Minimization Problem (12a)

We next employ the Anderson-Moore method to solve the F-minimization problem (12a). The advantage of this algorithm lies in its fast convergence (compared to the gradient method) and in its simple implementation (compared to Newton's method); e.g., see [9], [16], [17]. When applied to the F-minimization problem (12a), this method requires the solutions of two Lyapunov equations and one Sylvester equation in each iteration. We next recall the first and second order derivatives of J; for related developments, see [17].

Proposition 1: The gradient of J is determined by

$$\nabla J(F) = 2\,(RF - B_2^T P)\, L$$

where L and P are the controllability and observability Gramians of the closed-loop system

$$(A - B_2F)L + L(A - B_2F)^T = -\,B_1B_1^T \tag{NC-L}$$
$$(A - B_2F)^T P + P(A - B_2F) = -\,(Q + F^T RF). \tag{NC-P}$$

The second-order approximation of J is determined by

$$J(F + \tilde{F}) \approx J(F) + \langle \nabla J(F), \tilde{F}\rangle + \left(\frac{1}{2}\right) \langle H(F, \tilde{F}), \tilde{F}\rangle$$

where $H(F,\tilde{F})$ is the linear function of $\tilde{F}$

$$H(F,\tilde{F}) = 2\left((R\tilde{F} - B_2^T\tilde{P})\,L + (RF - B_2^T P)\,\tilde{L}\right)$$

and $\tilde{L}$, $\tilde{P}$ are the solutions of the following Lyapunov equations:

$$\begin{aligned}(A - B_2F)\,\tilde{L} + \tilde{L}\,(A - B_2F)^T &= B_2\tilde{F}L + (B_2\tilde{F}L)^T\\ (A - B_2F)^T\,\tilde{P} + \tilde{P}\,(A - B_2F) &= (PB_2 - F^TR)\,\tilde{F}\\ &\quad + \tilde{F}^T\,(B_2^TP - RF).\end{aligned}$$

By completing the squares with respect to F in the augmented Lagrangian $\mathcal{L}_\rho$, we obtain the following equivalent problem to (12a)

$$\text{minimize}\quad \varphi(F) = J(F) + \left(\frac{\rho}{2}\right)\|F - U^k\|_F^2$$

where $U^k = G^k - (1/\rho)\Lambda^k$. Setting $\nabla\varphi := \nabla J + \rho(F - U^k)$ to zero yields the necessary conditions for optimality

$$2\left(RF - B_2^TP\right)L + \rho\left(F - U^k\right) = 0 \qquad \text{(NC-F)}$$

where L and P are determined by (NC-L) and (NC-P).

Starting with a stabilizing feedback F, the Anderson-Moore method solves the two Lyapunov equations (NC-L) and (NC-P), and then solves the Sylvester equation (NC-F) to obtain a new feedback gain $\bar{F}$. In other words, it alternates between solving (NC-L) and (NC-P) for L and P with F being fixed and solving (NC-F) for F with L and P being fixed. It can be shown that the difference between two consecutive steps $\tilde{F} = \bar{F} - F$ forms a descent direction of φ; see [9] for a related result. Thus, line search methods [18] can be employed to determine step-size s in $F + s\tilde{F}$ to guarantee closed-loop stability and the convergence to a stationary point of φ.

Remark 3 (Closed-Loop Stability): Since the $\mathcal{H}_2$ norm is well defined for causal, strictly proper, stable closed-loop systems, we set J to infinity if $A - B_2F$ is not Hurwitz. Furthermore, J is a smooth function that increases to infinity as one approaches the boundary of the set of stabilizing gains [9]. Thus, the decreasing sequence of $\{\varphi(F^i)\}$ ensures that $\{F^i\}$ are stabilizing gains.

C. Solving the Structured $\mathcal{H}_2$ Problem

We next turn to the $\mathcal{H}_2$ problem subject to structural constraints on the feedback gain. Here, we fix the sparsity patterns $F \in \mathcal{S}$ identified using ADMM and then solve (SH2) to obtain the optimal feedback gain that belongs to $\mathcal{S}$. This procedure, commonly used in optimization [10, Section 6.3.2], can improve the performance of sparse feedback gains resulting from the ADMM algorithm.

As noted in Remark 3, the sparse feedback gains obtained in ADMM are stabilizing. This feature facilitates the use of descent algorithms (e.g., Newton's method) to solve (SH2). Given an initial gain $F^0 \in \mathcal{S}$, a decreasing sequence of the objective function $\{J(F^i)\}$ is generated by updating F according to $F^{i+1} = F^i + s^i\,\tilde{F}^i$; here, s^i is the step-size and $\tilde{F}^i \in \mathcal{S}$ is the Newton direction that is determined by the minimizer of the second-order approximation of the objective function (3). Equivalently, $\tilde{F}^i \in \mathcal{S}$ is the minimizer of $\Phi(\tilde{F}) := (1/2)\,\langle H(\tilde{F}) \circ I_{\mathcal{S}}, \tilde{F}\rangle + \langle \nabla J \circ I_{\mathcal{S}}, \tilde{F}\rangle$ where *structural identity* $I_{\mathcal{S}}$ of subspace $\mathcal{S}$ (under entry-wise multiplication $\circ$ of two matrices) is used to characterize structural constraints

$$\begin{aligned}I_{\mathcal{S}\,ij} &= \begin{cases}1, & \text{if } F_{ij} \text{ is a free variable}\\ 0, & \text{if } F_{ij} = 0 \text{ is required}\end{cases}\\ \Rightarrow F \circ I_{\mathcal{S}} &= F \quad \text{for} \quad F \in \mathcal{S}.\end{aligned}$$

To compute Newton direction, we use the conjugate gradient method that does not require forming or inverting the large Hessian matrix explicitly; see [18, Chapter 5]. It is noteworthy that techniques such as the negative curvature test [18, Section 7.1] can be employed to guarantee the descent property of the Newton direction; consequently, line search methods, such as the Armijo rule [18, Section 3.1], can be used to generate a decreasing sequence of J.

D. Convergence of ADMM

For convex problems the convergence of ADMM to the global minimizer follows from standard results [1]. For nonconvex problems, where convergence results are not available, extensive computational experience suggests that ADMM works well when the value of ρ is sufficiently large [19], [20]. This is attributed to the quadratic term $(\rho/2)\|F - G\|_F^2$ that tends to locally convexify the objective function for sufficiently large ρ; see [21, Chapter 14.5].

For problem (SP) with g determined by the weighted ℓ_1 norm (9), we next show that when ADMM converges, it converges to a critical point of (SP). For a convergent point $(F^\star, G^\star, \Lambda^\star)$ of the sequence $\{F^k, G^k, \Lambda^k\}$, (12c) simplifies to $F^\star - G^\star = 0$. Since $F^\star$ minimizes $\mathcal{L}_\rho(F, G^\star, \Lambda^\star)$ and since $G^\star$ minimizes $\mathcal{L}_\rho(F^\star, G, \Lambda^\star)$, we have $\{0 = \nabla J(F^\star) + \Lambda^\star,\ 0 \in \gamma\,\partial g(G^\star) - \Lambda^\star\}$ where ∂g is the subdifferential of the convex function g in (9). Therefore, $(F^\star, G^\star)$ satisfies the necessary conditions for the optimality of (SP) and ADMM converges to a critical point of (SP).

IV. Examples

We next use three examples to illustrate the utility of the approach developed in Section III. The identified sparsity structures result in *localized* controllers in all three cases. Additional information about these examples, along with MATLAB source codes, can be found at

www.ece.umn.edu/~mihailo/software/lqrsp/

A. Mass-Spring System

For a mass-spring system with N masses on a line, let p_i be the displacement of the ith mass from its reference position and let the state variables be $x_1 := [\,p_1\ \ldots\ p_N\,]^T$ and $x_2 := \dot{x}_1$. For unit masses and spring constants, the state-space representation is given by (1) with

$$A = \begin{bmatrix} O & I\\ T & O\end{bmatrix}, \qquad B_1 = B_2 = \begin{bmatrix} O\\ I\end{bmatrix}$$

where T is an $N \times N$ tridiagonal Toeplitz matrix with -2 on its main diagonal and 1 on its first sub- and super-diagonal, and I and O are $N \times N$ identity and zero matrices. The state performance weight Q is the identity matrix and the control performance weight is $R = 10I$.

We use the weighted ℓ_1 norm as the sparsity-promoting penalty function, where we follow [4] and set the weights W_{ij} to be inversely proportional to the magnitude of the solution $F^\star$ of (SP) at the previous value of γ, $W_{ij} = 1/(|F_{ij}^\star| + \varepsilon)$. This places larger relative weight on smaller feedback gains and they are more likely to be dropped in the sparsity-promoting algorithm. Here, $\varepsilon = 10^{-3}$ is introduced to have well-defined weights when $F_{ij}^\star = 0$.

The optimal feedback gain at $\gamma = 0$ is computed from the solution of the algebraic Riccati equation. As γ increases, the number of nonzero sub- and super-diagonals of both position $F_p^\star$ and velocity $F_v^\star$ gains decreases; see Fig. 1(a) and (b). Eventually, both $F_p^\star$ and $F_v^\star$ become diagonal matrices. It is noteworthy that diagonals of both position and velocity feedback gains are nearly constant except for masses that are close to the boundary; see Fig. 1(c) and (d).

After sparsity structures of controllers are identified by solving (SP), we fix sparsity patterns and solve structured $\mathcal{H}_2$ problem (SH2) to obtain the optimal *structured* controllers. Comparing the sparsity level

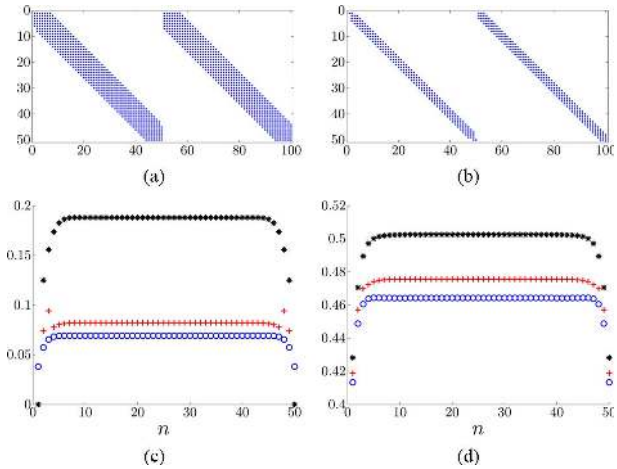
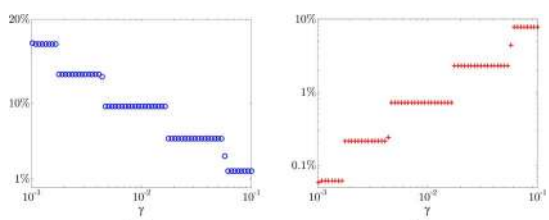


Fig. 1. Sparsity patterns of $F^\star = [F_p^\star \;\; F_v^\star] \in \mathbb{R}^{50\times 100}$ for the mass-spring system obtained using weighted ℓ_1 norm with (a) $\gamma = 10^{-4}$ and (b) $\gamma = 0.0105$. As γ increases, the number of nonzero sub- and super-diagonals of $F_p^\star$ and $F_v^\star$ decreases. The diagonals of (c) $F_p^\star$ and (d) $F_v^\star$ for different values of γ: 10^{-4} ($\circ$), 0.0281 ($+$), and 0.1 ($*$). The diagonals of the centralized position and velocity gains are almost identical to ($\circ$). (a) $\gamma = 10^{-4}$. (b) $\gamma = 0.0105$. (c) . (d) .



γ	0.01	0.04	0.10
$\mathbf{card}\,(F^\star)/\mathbf{card}\,(F_c)$	9.4%	5.8%	2.0%
$(J(F^\star) - J(F_c))/J(F_c)$	0.8%	2.3%	7.8%

(c)

Fig. 2. (a) Sparsity level and (b) the performance degradation of $F^\star$ compared to the centralized gain F_c for mass-spring system. (c) Sparsity vs. performance: using 2% of nonzero elements, $\mathcal{H}_2$ performance of $F^\star$ is only 7.8% worse than performance of F_c. (a) $\mathbf{card}\,(F^\star)/\mathbf{card}\,(F_c)$. (b) $(J(F^\star) - J(F_c))/J(F_c)$. (c) .

and the performance of these controllers to those of the centralized controller F_c, we see that using only a *fraction* of nonzero elements, the sparse feedback gain $F^\star$ achieves $\mathcal{H}_2$ performance comparable to the performance of F_c; see Fig. 2. In particular, using about 2% of nonzero elements, $\mathcal{H}_2$ performance of $F^\star$ is only about 8% worse than that of F_c.

B. Network With 100 Unstable Nodes

Let $N = 100$ nodes be randomly distributed with a uniform distribution in a square region of 10×10 units. Each node is an unstable second order system coupled with other nodes through the exponentially decaying function of the Euclidean distance $\alpha(i,j)$ between them [22]

$$\begin{bmatrix} \dot{x}_{1i} \\ \dot{x}_{2i} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} + \sum_{j\neq i} \mathrm{e}^{-\alpha(i,j)} \begin{bmatrix} x_{1j} \\ x_{2j} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (d_i + u_i)$$

with $i = 1, \ldots, N$. The performance weights Q and R are set to identity matrices.

We use the weighted ℓ_1 norm as the penalty function with the weights given in Section IV-A. As γ increases, the underlying communication

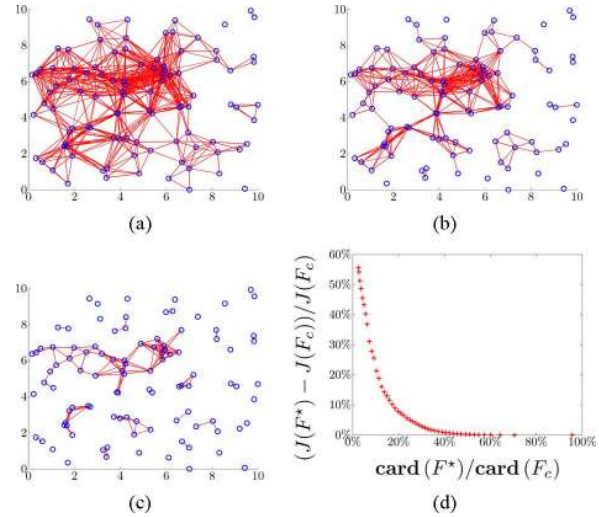


Fig. 3. (a)–(c) Localized communication graphs of distributed controllers obtained by solving (SP) for different values of γ for the network with 100 nodes. Note that the communication graph does not have to be connected since the nodes are dynamically coupled to each other and allowed to measure their own states. (d) The optimal trade-off curve between the $\mathcal{H}_2$ performance degradation and the sparsity level of $F^\star$ compared to the centralized gain F_c. (a) $\gamma = 12.6$. (b) $\gamma = 26.8$. (c) $\gamma = 68.7$. (d).

graphs gradually become *localized*; see Fig. 3(a)–(c). With about 8% of nonzero elements of F_c, $\mathcal{H}_2$ performance of $F^\star$ is about 28% worse than performance of the centralized gain F_c. Fig. 3(d) shows the optimal trade-off curve between the $\mathcal{H}_2$ performance and the feedback gain sparsity.

We note that the truncation of the centralized controller could result in a non-stabilizing feedback matrix [22]. In contrast, our approach gradually modifies the feedback gain and increases the number of zero elements, which plays an important role in preserving the closed-loop stability.

C. Block Sparsity: A Bio-Chemical Reaction Example

Consider a network of $N = 5$ systems coupled through the following dynamics:

$$\dot{x}_i = [A]_{ii}\, x_i - \frac{1}{2}\sum_{j=1}^{N} (i-j)\,(x_i - x_j) + [B_1]_{ii}\, d_i + [B_2]_{ii}\, u_i$$

where $[\,\cdot\,]_{ij}$ denotes the ijth block of a matrix and

$$[A]_{ii} = \begin{bmatrix} -1 & 0 & -3 \\ 3 & -1 & 0 \\ 0 & 3 & -1 \end{bmatrix},$$

$$[B_1]_{ii} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad [B_2]_{ii} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}.$$

The performance weights Q and R are set to identity matrices. Systems of this form arise in bio-chemical reactions with a cyclic negative feedback [23].

We use the weighted sum of Frobenius norms as the sparsity-promoting penalty function and we set the weights W_{ij} to be inversely proportional to the Frobenius norm of the solution $F_{ij}^\star$ to (SP) at the previous value of γ, i.e., $W_{ij} = 1/(\|F_{ij}^\star\|_F + \varepsilon)$ with $\varepsilon = 10^{-3}$. As γ increases, the number of *nonzero blocks* in F decreases. Fig. 4 shows sparsity patterns of feedback gains and the corresponding communication graphs resulting from solving (SP) with sparse and block sparse penalty functions. Setting γ to values that yield the same number of

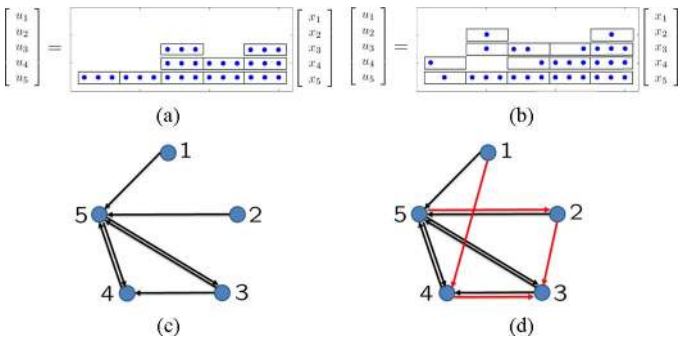


Fig. 4. Sparse feedback gains obtained by solving (SP) using (a) the weighted sum of Frobenius norms with $\gamma = 3.6$ and (b) the weighted ℓ_1 norm (9) with $\gamma = 1.3$. Here, $F \in \mathbb{R}^{5\times 15}$ is partitioned into 25 blocks $F_{ij} \in \mathbb{R}^{1\times 3}$. Both feedback gains have the same number of nonzero elements (indicated by dots) and close $\mathcal{H}_2$ performance (less than 1% difference), but different number of nonzero blocks (indicated by boxes). Communication graphs of (c) the block sparse feedback gain in (a), and (d) the sparse feedback gain in (b) (red color highlights the additional links). An arrow pointing from node i to node j indicates that i uses measurements from j.

nonzero elements in these feedback gains results in the block sparse feedback gain with a smaller number of nonzero blocks. In particular, the first two rows of the block sparse feedback gain in Fig. 4(a) are identically equal to zero (indicated by blank space). This means that the subsystems 1 and 2 do not need to be actuated. Thus, the communication graph determined by the block sparse feedback gain has fewer links; cf. Fig. 4(c) and (d).

V. Conclusion

We design sparse and block sparse state feedback gains that optimize the $\mathcal{H}_2$ performance of distributed systems. The design procedure consists of a structure identification step and a "polishing" step. In the identification step, we employ the ADMM algorithm to solve the sparsity-promoting optimal control problem, whose solution gradually moves from the centralized gain to the sparse gain of interest as our emphasis on the sparsity-promoting penalty term is increased. In the polishing step, we use Newton's method in conjunction with a conjugate gradient scheme to solve the minimum variance problem subject to the identified sparsity constraints.

Although we focus on the $\mathcal{H}_2$ performance, the developed framework can be extended to design problems with other performance indices. We emphasize that the analytical solutions to the G-minimization problem are independent of the assigned performance index. Consequently, the G-minimization step in ADMM for (SP) with alternative performance indices can be done exactly as in Section III-A. Thus, ADMM provides a flexible framework for sparsity-promoting optimal control problems of the form (SP).

We have recently employed ADMM for selection of an *a priori* specified number of leaders in order to minimize the variance of stochastically forced dynamic networks [24], for creation of new social links to maximize public awareness in social networks [25], and for identification of sparse representations of consensus networks [26]. We also aim to extend the developed framework to the observer-based sparse optimal feedback design. Our results on the identification of classes of convex optimal control problems will be reported elsewhere.

Acknowledgment

The authors would like to thank S. P. Boyd for inspiring discussions on ADMM, group lasso, and cardinality minimization, and R. Glowinski for his insightful comments on convergence of ADMM.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–124, 2011.

[2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2009.

[4] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl*, vol. 14, pp. 877–905, 2008.

[5] S. Schuler, C. Ebenbauer, and F. Allgöwer, "ℓ_0-system gain and ℓ_1-optimal control," in *Preprints 18th IFAC World Congr.*, 2011, pp. 9230–9235.

[6] S. Schuler, P. Li, J. Lam, and F. Allgöwer, "Design of structured dynamic output-feedback controllers for interconnected systems," *Int. J. Control*, vol. 84, no. 12, pp. 2081–2091, 2011.

[7] D. Zelazo, S. Schuler, and F. Allgöwer, "Performance and design of cycles in consensus networks," *Syst. Control Lett.*, vol. 62, no. 1, pp. 85–96, 2013.

[8] M. Fardad, F. Lin, and M. R. Jovanović, "On the optimal design of structured feedback gains for interconnected systems," in *Proc. 48th IEEE Conf. Decision Control*, Shanghai, China, 2009, pp. 978–983.

[9] F. Lin, M. Fardad, and M. R. Jovanović, "Augmented Lagrangian approach to design of structured optimal state feedback gains," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2923–2929, Dec. 2011.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[11] M. S. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," *Ann. Oper. Res.*, vol. 152, pp. 341–365, 2007.

[12] M. Fardad, F. Lin, and M. R. Jovanović, "Sparsity-promoting optimal control for a class of distributed systems," in *Proc. Amer. Control Conf.*, San Francisco, CA, USA, 2011, pp. 2050–2055.

[13] M. Fardad, F. Lin, and M. R. Jovanović, "On the optimal synchronization of oscillator networks via sparse interconnection graphs," in *Proc. Amer. Control Conf.*, Montréal, QC, Canada, 2012, pp. 4777–4782.

[14] P. L. D. Peres and J. C. Geromel, "An alternate numerical solution to the linear quadratic problem," *IEEE Trans. Autom. Control*, vol. 39, pp. 198–202, 1994.

[15] F. Lin, "Structure Identification and Optimal Design of Large-Scale Networks of Dynamical Systems," Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, USA, 2012.

[16] P. M. Mäkilä and H. T. Toivonen, "Computational methods for parametric LQ problems – A survey," *IEEE Trans. Autom. Control*, vol. 32, no. 8, pp. 658–671, 1987.

[17] T. Rautert and E. W. Sachs, "Computational design of optimal output feedback controllers," *SIAM J. Optim.*, vol. 7, no. 3, pp. 837–852, 1997.

[18] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer, 2006.

[19] M. Fortin and R. Glowinski, *Augmented Lagrangian Method: Application to the Numerical Solution of Boundary-Value Problems*. Amsterdam, The Netherlands: North-Holland, 1983.

[20] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Philadelphia, PA: SIAM, 1989.

[21] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. New York: Springer, 2008.

[22] N. Motee and A. Jadbabaie, "Optimal control of spatially distributed systems," *IEEE Trans. Automat. Control*, vol. 53, no. 7, pp. 1616–1629, Jul. 2008.

[23] M. R. Jovanović, M. Arcak, and E. D. Sontag, "A passivity-based approach to stability of spatially distributed systems with a cyclic interconnection structure," *IEEE Trans. Autom. Control: Special Issue on Syst. Biol.*, vol. 53, no. 1, pp. 75–86, Jan. 2008.

[24] F. Lin, M. Fardad, and M. R. Jovanović, "Algorithms for leader selection in large dynamical networks: Noise-corrupted leaders," in *Proc. 50th IEEE Conf. Decision Control Eur. Control Conf.*, Orlando, FL, 2011, pp. 2932–2937.

[25] M. Fardad, X. Zhang, F. Lin, and M. R. Jovanović, "On the optimal dissemination of information in social networks," in *Proc. 51th IEEE Conf. Decision Control*, Maui, HI, 2012, pp. 2539–2544.

[26] N. Dhingra, F. Lin, M. Fardad, and M. R. Jovanović, "On identifying sparse representations of consensus networks," in *Proc. 3rd IFAC Wkshp Distrib. Estim. Control Netw. Syst.*, Santa Barbara, CA, 2012, pp. 305–310.