Proximal algorithms for large-scale statistical modeling and sensor/actuator selection

Armin Zare, *Member, IEEE*, Hesameddin Mohammadi, *Student Member, IEEE*, Neil K. Dhingra, *Member, IEEE*, Tryphon T. Georgiou, *Fellow, IEEE*, and Mihailo R. Jovanović, *Fellow, IEEE*

***Abstract*—Several problems in modeling and control of stochastically-driven dynamical systems can be cast as regularized semi-definite programs. We examine two such representative problems and show that they can be formulated in a similar manner. The first, in statistical modeling, seeks to reconcile observed statistics by suitably and minimally perturbing prior dynamics. The second seeks to optimally select a subset of available sensors and actuators for control purposes. To address modeling and control of large-scale systems we develop a unified algorithmic framework using proximal methods. Our customized algorithms exploit problem structure and allow handling statistical modeling, as well as sensor and actuator selection, for substantially larger scales than what is amenable to current general-purpose solvers. We establish linear convergence of the proximal gradient algorithm, draw contrast between the proposed proximal algorithms and alternating direction method of multipliers, and provide examples that illustrate the merits and effectiveness of our framework.**

***Index Terms*—Actuator selection, sensor selection, sparsity-promoting estimation and control, method of multipliers, nonsmooth convex optimization, proximal algorithms, regularization for design, semi-definite programming, structured covariances.**

I. Introduction

Convex optimization has had tremendous impact on many disciplines, including system identification and control design [1]–[7]. The forefront of research points to broadening the range of applications as well as sharpening the effectiveness of algorithms in terms of speed and scalability. The present paper focuses on two representative control problems, statistical control-oriented modeling and sensor/actuator selection, that are cast as convex programs. A range of modern applications require addressing these over increasingly large parameter spaces, placing them outside the reach of standard solvers. A contribution of the paper is to formulate such problems as regularized semi-definite programs (SDPs) and to develop customized optimization algorithms that scale favorably with size.

Modeling is often seen as an inverse problem where a search in parameter space aims to find a parsimonious representation of data. For example, in the control-oriented modeling of fluid flows, it is of interest to improve upon dynamical equations arising from first-principles (e.g., linearized Navier-Stokes equations), in order to accurately replicate observed statistical features that are estimated from data. To this end, a perturbation of the prior model can be seen as a feedback gain that results in dynamical coupling between a suitable subset of parameters [8]–[10]. On the flip side, active control of large-scale and distributed systems requires judicious placement of sensors and actuators which again can be viewed as the selection of a suitable feedback or Kalman gain. In either modeling or control, the selection of such gain matrices must be guided by optimality criteria as well as simplicity (low rank or sparse architecture). We cast both types of problems as optimization problems that utilize suitable convex surrogates to handle complexity. The use of such surrogates is necessitated by the fact that searching over all possible architectures is combinatorially prohibitive.

Applications that motivate our study require scalable algorithms that can handle large-scale problems. While the optimization problems that we formulate are SDP representable, e.g., for actuator selection, worst-case complexity of generic solvers scales as the sixth power of the sum of the state dimension and the number of actuators. Thus, solvers that do not exploit the problem structure cannot cope with the demands of such large-scale applications. This necessitates the development of customized algorithms that are pursued herein.

Our presentation is organized as follows. In Section II, we describe the modeling and control problems that we consider, provide an overview of literature and the state-of-the-art, and highlight the technical contribution of the paper. In Section III, we formulate the *minimum energy covariance completion* (control-oriented modeling) and *sensor/actuator selection* (control) problems as nonsmooth SDPs. In Section IV, we present a customized Method of Multipliers (MM) algorithm for covariance completion. An essential ingredient of MM is the Proximal Gradient (PG) method. We also use the PG method for sensor/actuator selection and establish its convergence rate. In Section V, we offer two motivating examples for actuator selection and covariance completion and discuss computational experiments. We conclude with a brief summary of the results and future directions in Section VI.

II. Motivating applications and contribution

We consider dynamical systems with additive stochastic disturbances. In the first instance, we are concerned with a modeling problem where the statistics are not consistent with

Financial support from the National Science Foundation under Awards CMMI 1739243, ECCS 1509387, 1708906, 1809833, and 1839441, and the Air Force Office of Scientific Research under FA9550-16-1-0009, FA9550-17-1-0435, and FA9550-18-1-0422 is gratefully acknowledged.

A. Zare is with the Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75219. H. Mohammadi and M. R. Jovanović are with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089. N. K. Dhingra is with Numerica Corporation, Fort Collins, CO 80528. T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697. E-mails: armin.zare@utdallas.edu, hesamedm@usc.edu, neil.k.dh@gmail.com, tryphon@uci.edu, mihailo@usc.edu.

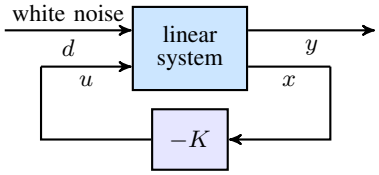


Fig. 1. A feedback connection of an LTI system with a static gain matrix that is designed to account for the sampled steady-state covariance X.

a prior model that is available to us. In that case, we seek to modify our model in a parsimonious manner (a sparse and structured perturbation of the state matrix) so as to account for the partially observed statistics. In the second, we are concerned with the control of such stochastic dynamics via a collection of judiciously placed sensors and actuators. Once again, the architecture of the (now) control problem calls for the selection of sparse matrix gains that effect control and estimation. These problems are explained next.

A. Statistical modeling and covariance completion

It is well-established that the linearized Navier-Stokes (NS) equations driven by stochastic excitation can account for qualitative [11]–[15] and quantitative [9], [10] features of shear flows. The value of such models has been to provide insights into the underlying physics as well as to guide control design. A significant recent step in this direction was to recognize [9] that *colored-in-time* excitation can account for features of the flow field that *white* noise in earlier literature cannot [16]. Furthermore, it has been pointed out that the effect of colored-in-time excitation is *equivalent* to white-in-time excitation together with a *structural perturbation of the system dynamics* [8], [9]. Such structural perturbations may reveal salient dynamical couplings between variables and, thereby, enhance understanding of basic physics [9, Section 6.1]; see [10] for a review of covariance completion problems and its relevance in stochastic dynamical modeling of turbulent flows.

These insights and reasoning motivate an optimal state-feedback synthesis problem [17] to identify dynamical couplings that bring consistency between the model and the observed statistics. Model parsimony dictates a penalty on the complexity of structural perturbations and leads to an optimization problem that involves a composite cost function

$$f(X,K) \;+\; \gamma\, g(K) \tag{1}$$

subject to stability of the system in Fig. 1. Here, X denotes a state covariance matrix and K is a state-feedback matrix. The function $f(X,K)$ penalizes variance and control energy while $g(K)$ is a sparsity-promoting regularizer which penalizes the number of nonzero rows in K; sparsity in the rows of K amounts to a reduced number of feedback couplings that modify the system dynamics. In addition, state statistics may be partially known, in which case a constraint $X_{ij} = G_{ij}$ for $(i,j) \in \mathcal{I}$ is added, where the entries of G represent known entries of X for indices in $\mathcal{I}$.

The resulting *minimum-control-energy covariance completion problem* can be cast as an SDP which, for small-size problems, is readily solvable using standard software. A class of similar problems have been proposed in the context of stochastic control [18]–[21] and of output covariance estimation [22], [23] which, likewise and for small-size, are readily solvable by standard software.

B. Sensor and actuator selection

The selection and proper placement of sensors/actuators impacts the performance of closed-loop control systems; making such a choice is a nontrivial task even for systems of modest size. Previous work on actuator/sensor placement either relies on heuristics or on greedy algorithms and convex relaxations.

The benefit of a particular sensors/actuator placement is typically quantified by properties of the resulting observability/controllability and the selection process is guided by indicators of diminishing return in performance near optimality [24], [25]. However, metrics on the performance of Kalman filters and other control objectives have been shown to lack supermodularity [26], [27], which hampers the effectiveness of greedy approaches in sensor/actuator selection.

The literature on different approaches includes convex formulations for sensor placement in problems with linear measurements [28], maximizing the trace of the Fisher information under constraints when dealing with correlated measurement noise [29], and a variation of optimal experiment design for placing measurement units in power networks [30]. Actuator selection via genetic algorithms has also been explored [31]. Finally, a non-convex formulation of the joint sensor and actuator placement was advanced in [32], [33] and was recently applied to the linearized Ginzburg-Landau equation [34].

Herein, we cast our placement problem as one of optimally selecting a subset of potential sensors or actuators which, in a similar manner as our earlier modeling problem, involves the minimization of a nonsmooth composite function as in (1). More specifically, we utilize the sparsity-promoting framework developed in [35]–[37] to enforce block-sparse structured observer/feedback gains and select sensors/actuators.

The algorithms developed in [37] have been used for sensor selection in target tracking [38] and in periodic sensor scheduling in networks of dynamical systems [39]. However, they were developed for general problems, without exploiting a certain hidden convexity in sensor/actuator selection. Indeed, for the design of row-sparse feedback gains, the authors of [40] introduced a convex SDP reformulation of the problem formulated in [37]. Inspired by [37], the authors of [41] extended the SDP formulation to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ sensor/actuator placement problems for discrete time LTI systems. Their approach utilizes standard SDP-solvers with re-weighted ℓ_1-norm regularizers. In the present paper, we integrate several of these ideas. In particular, we borrow group-sparsity regularizers from statistics [42] and develop efficient customized proximal algorithms for the resulting SDPs.

C. Main contribution

In the present paper, we highlight the structural similarity between statistical modeling and sensor/actuator selection, and develop a unified algorithmic framework for handling large-scale problems. Proximal algorithms are utilized to address the non-differentiability of the sparsity-promoting term $g(K)$

in the objective function. We exploit the problem structure, implicitly handle the stability constraint on state covariances and controller gains by expressing one in terms of the other, and develop a customized proximal gradient algorithm that scales with the third power of the state-space dimension. We prove linear convergence for the proximal gradient algorithm with fixed step-size and propose an adaptive step-size selection method that can improve convergence. We also discuss initialization techniques and stopping criteria for our algorithms, and provide numerical experiments to demonstrate the effectiveness of our approach relative to existing methods.

III. Problem formulation

Consider a linear time-invariant (LTI) system with state-space representation

$$\begin{aligned} \dot{x} &= A\,x + B\,u + d \\ y &= C\,x \end{aligned} \tag{2}$$

where $x(t) \in \mathbb{C}^n$ is the state vector, $y(t) \in \mathbb{C}^p$ is the output, $u(t) \in \mathbb{C}^m$ is the control input, and $d(t)$ is a white stochastic process with zero-mean and the covariance matrix $V \succ 0$, $\mathbf{E}(d(t)d^*(\tau)) = V\delta(t-\tau)$. Here, $\mathbf{E}$ is the expected value, $B \in \mathbb{C}^{n\times m}$ is the input matrix with $m \leq n$, $C \in \mathbb{C}^{p\times n}$ is the output matrix, and the pair (A, B) is controllable. The choice of the state-space is motivated by spatially distributed systems where the application of the spatial Fourier transform naturally leads to complex-valued quantities in (2); e.g., see [43].

We consider two specific applications, one that relates system identification and covariance completion, and another that focuses on actuator selection in a control problem. Both can be cast as the problem to select a stabilizing state-feedback control law, $u = -Kx$, that utilizes few input degrees of freedom in the sense that the matrix K has a large number of zero rows. At the same time, the closed-loop system

$$\dot{x} = (A - B\,K)\,x + d$$

shown in Fig. 1 is consistent with partially available state-correlations and/or is optimal in a quadratic sense.

More specifically, if

$$X := \lim_{t\to\infty} \mathbf{E}\,(x(t)\,x^*(t))$$

denotes the stationary state-covariance of the controlled system, the pertinent quadratic cost is

$$\begin{aligned} f(X,K) &:= \operatorname{trace}(Q\,X + K^*R\,K\,X) \\ &= \lim_{t\to\infty} \mathbf{E}\,(x^*(t)\,Q\,x(t) + u^*(t)\,R\,u(t)) \end{aligned} \tag{3}$$

whereas $Q = Q^* \succ 0$ and $R = R^* \succ 0$ specify penalties on the state and control input, respectively. Both stability of the feedback dynamics and consistency with the state covariance X reduce to an algebraic constraint on K and X, namely,

$$(A - B\,K)X + X(A - B\,K)^* + V = 0. \tag{4}$$

Finally, the number of non-zero rows of K can be seen as the number of active degrees of freedom of the input $u = -Kx$. The choice of such a K, with few non-zero rows is sought via minimization of a non-smooth composite objective function in Problem 1, where

$$g(K) := \sum_{i=1}^{n} w_i \,\|\mathrm{e}_i^* K\|_2 \tag{5}$$

is a regularizing term that promotes row-sparsity of K [42], w_i are positive weights, and e_i is the ith unit vector in $\mathbb{R}^m$.

Problem 1: Minimize $f(X,K) + \gamma\, g(K)$, subject to (4), $X \succ 0$, and, possibly, constraints on the values of specified entries of X, $X_{ij} = G_{ij}$ for $(i,j) \in \mathcal{I}$, where a set of pairs $\mathcal{I}$ and the entries G_{ij} are given.

In this problem, $\gamma > 0$ specifies the importance of sparsity, and $\mathcal{I}$ specifies indices of available covariance data. A useful variant of the constraint on the entries of X, when, e.g., statistics of output variables are estimated, can be expressed as

$$(CXC^*)_{ij} = G_{ij} \quad \text{for } (i,j) \in \mathcal{I} \tag{6}$$

We next explain how Problem 1 relates to the two aforementioned topics of covariance completion and actuator selection.

A. Covariance completion and model consistency

In many problems, it is often the case that a model is provided for a given process which, however, is inconsistent with new data. In such instances, it is desirable to revise the dynamics by a suitable perturbation to bring compatibility between model and data. The data in our setting consists of statistics in the form of a state covariance X for a linear model

$$\dot{x} = A\,x + d \tag{7}$$

with white noise input d.

We postulate and deal with a further complication when the data is incomplete. More specifically, we allow X to be only partially known. Such an assumption is motivated by fluid flow applications that rely on the linearized NS equations [9]. In this area both the numerical and experimental determination of all entries of X is often prohibitively expensive. Thus, the problem to bring consistency between data and model can be cast in the form of Problem 1, where we seek a completion of the missing entries of X along with a perturbation $\Delta := -BK$ of the system dynamics (7), into

$$\dot{x} = (A + \Delta)\,x + d.$$

The assumed structure of Δ is without loss of generality, and the choice of B may incorporate added insights into the strength and directionality of possible couplings between state variables. It should be noted that a full-rank matrix B that allows the perturbation signal Kx to manipulate all degrees of freedom can lead to the complete cancellation of the original dynamics A; see [8, Section III] for details. Then, when seeking a suitable perturbation, it is also natural to impose a penalty on the average quadratic size of signals Kx. This brings us into the setting of Problem 1, where the choice of most suitable perturbation is determined by the optimization criterion. Once again, the row-sparsity promoting penalty $g(K)$ impacts the choice of feedback couplings that need to be introduced to modify the dynamical generator A [17].

B. Actuator selection

As is well-known, the unique optimal control law that

minimizes the steady-state variance (3) of system (2) is a static state-feedback $u = -Kx$. The optimal gain K and the corresponding state covariance X can be obtained by minimizing $f(X, K)$, over $K \in \mathbb{C}^{m\times n}$, and positive definite $X = X^* \in \mathbb{C}^{n\times n}$. The solution can also be obtained by solving an algebraic Riccati equation arising from the KKT conditions of this optimization problem. In general, K is populated by non-zero entries, implying that all "input channels" (i.e., all entries of u) would be active. Since the columns of B encode the effect of individual "input channels", representing location of actuators, a subselection that is affected by the row-sparsity promoting regularizer in Problem 1, amounts to actuator selection amongst available options. A dual formulation can be cast to address sensor selection and can be approached in a similar manner; see Appendix A.

C. Change of variables and SDP representation

The constraint $X \succ 0$ in Problem 1 allows for a standard change of variables $Y := KX$ to replace K in $f(X, K) = \mathrm{trace}\,(QX + K^*RKX)$. This yields the function

$$f(X, Y) \;=\; \mathrm{trace}\,\big(Q\,X \,+\, Y^*RYX^{-1}\big) \tag{8}$$

which is jointly convex in (X, Y). Further, the row-sparsity of K is equivalent to the row-sparsity of Y [40]. This observation leads to the convex reformulation of Problem 1 (incorporating the more general version of constraints (6)) as follows.

Problem 2: Minimize $f(X, Y) + \gamma \sum_i w_i \|\mathrm{e}_i^* Y\|_2$ over a Hermitian matrix $X \in \mathbb{C}^{n\times n}$ and $Y \in \mathbb{C}^{m\times n}$, subject to:

$$\begin{aligned} A\,X \,+\, X\,A^* \,-\, B\,Y \,-\, Y^*B^* + V \;&=\; 0 \\ (1 \,-\, \delta)\,[\,(CXC^*) \circ E \,-\, G\,] \;&=\; 0 \\ X \;&\succ\; 0 \end{aligned}$$

where

$$\delta \;=\; \begin{cases} 0, & \text{for covariance completion} \\ 1, & \text{for actuator selection.} \end{cases}$$

The symbol $\circ$ denotes elementwise matrix multiplication, and E is the structural identity matrix,

$$E_{ij} \;=\; \begin{cases} 1, & \text{if } G_{ij} \text{ is available} \\ 0, & \text{if } G_{ij} \text{ is unavailable.} \end{cases}$$

As explained earlier, the matrices A, B, C, G, and V are problem data. From the solution of Problem 2, the optimal feedback gain matrix can be recovered as $K = YX^{-1}$. We note that the optimization of f can be expressed as an SDP. Specifically, the Schur complement can be used to characterize the epigraph of $\mathrm{trace}\,\big(RYX^{-1}Y^*\big)$ via the convex constraint

$$\begin{bmatrix} W & R^{1/2}Y \\ Y^*R^{1/2} & X \end{bmatrix} \;\succeq\; 0$$

and $\mathrm{trace}\,(W)$, where W is a matrix variable and the joint convexity of $\mathrm{trace}\,\big(RYX^{-1}Y^*\big)$ in (X, Y) follows [4].

We also note that although the row-sparsity patterns of Y and K are equivalent, the weights w_i are not necessarily the same in the respective expressions in Problems 1 and 2. In practice, the weights are iteratively adapted to promote row-sparsity; see Section IV-G. Problem 2 can be solved efficiently using general-purpose solvers for small number of variables. To address larger problems, we next exploit the structure and develop optimization algorithms based on the proximal gradient algorithm and the method of multipliers.

IV. Customized algorithms

In this section, we describe the steps through which we solve Problem 2, identify the essential input channels in B, and subsequently refine the solutions based on the identified sparsity structure. For notational compactness, we write the linear constraints in Problem 2 as

$$\begin{aligned} \mathcal{A}_1(X) \,-\, \mathcal{B}(Y) + V \;&=\; 0 \\ (1 \,-\, \delta)\,[\,\mathcal{A}_2(X) \,-\, G\,] \;&=\; 0 \end{aligned}$$

where the linear operators $\mathcal{A}_1$: $\mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$, $\mathcal{A}_2$: $\mathbb{C}^{n\times n} \to \mathbb{C}^{p\times p}$ and $\mathcal{B}$: $\mathbb{C}^{m\times n} \to \mathbb{C}^{n\times n}$ are given by

$$\begin{aligned} \mathcal{A}_1(X) \;&:=\; A\,X \,+\, X\,A^* \\ \mathcal{A}_2(X) \;&:=\; (CXC^*) \circ E \\ \mathcal{B}(Y) \;&:=\; B\,Y + Y^*B^*. \end{aligned}$$

A. Elimination of variable X

For any Y, there is a unique X that solves the equation

$$\mathcal{A}_1(X) \,-\, \mathcal{B}(Y) \,+\, V \;=\; 0 \tag{9}$$

if and only if the matrices A^* and $-A$ do not have any common eigenvalues [44]. When this condition holds, we can express the variable X as an affine function of Y,

$$X(Y) \;=\; \mathcal{A}_1^{-1}(\mathcal{B}(Y) \,-\, V) \tag{10}$$

and restate Problem 2 as

$$\begin{aligned} \underset{Y}{\text{minimize}} \quad & f(Y) \,+\, \gamma\, g(Y) \\ \text{subject to} \quad & (1 \,-\, \delta)\,[\,\mathcal{A}_2(X(Y)) \,-\, G\,] \;=\; 0 \\ & X(Y) \;\succ\; 0. \end{aligned} \tag{11}$$

The smooth part of the objective function in (11) is given by

$$f(Y) \;:=\; \mathrm{trace}\,\big(Q\,X(Y) \,+\, Y^*RYX^{-1}(Y)\big) \tag{12}$$

and the regularizing term is

$$g(Y) \;:=\; \sum_{i\,=\,1}^{n} w_i\, \|\mathrm{e}_i^* Y\|_2. \tag{13}$$

Since optimization problem (11) is equivalent to Problem 2 constrained to the affine equality (10), it remains convex.

When the matrix A is Hurwitz, expression (10) can be cast in terms of the well-known integral representation,

$$X(Y) \;=\; \int_0^\infty \mathrm{e}^{At}\,(V \,-\, B\,Y \,-\, Y^*B^*)\,\mathrm{e}^{A^*t}\,\mathrm{d}t.$$

Even for unstable open-loop systems, the operator $\mathcal{A}_1$ is invertible if the matrices A^* and $-A$ do not have any common eigenvalues. In our customized algorithms, we numerically evaluate the action of $\mathcal{A}_1^{-1}$ on the current iterate by solving the corresponding Lyapunov equation which requires making the following assumption.

Assumption 1: The operator $\mathcal{A}_1$ is invertible.

Appendix B provides a method to handle cases where this assumption does not hold.

B. Proximal gradient method for actuator selection

The proximal gradient (PG) method generalizes gradient descent to composite minimization problems in which the objective function is the sum of a differentiable and non-differentiable component [45], [46]. It is most effective when the proximal operator associated with the nondifferentiable component is easy to evaluate; many common regularization functions, such as the ℓ_1 penalty, nuclear norm, and hinge loss, satisfy this condition. Herein, we present details of a customized variant of the PG method for solving (11) with $\delta = 1$. In Algorithm 1, we follow the recommendations of [45], [47] for choosing the step-size and stopping criterion.

The PG method for solving (11) with $\delta = 1$ is given by

$$Y^{k+1} \;:=\; \mathbf{prox}_{\beta_k g}\big(Y^k \,-\, \alpha_k \nabla f(Y^k)\big) \tag{14}$$

where Y^k is the kth iterate, $\alpha_k > 0$ is the step-size, and $\beta_k := \gamma\alpha_k$. The proximal operator of a real-valued proper, closed, convex function h is defined as [48]

$$\mathbf{prox}_h(V) \;:=\; \underset{Y}{\operatorname{argmin}} \;\left(h(Y) \,+\, \frac{1}{2}\,\|Y \,-\, V\|_F^2\right). \tag{15}$$

where $\|\cdot\|_F$ is the Frobenius norm. For the row-sparsity regularizer, the proximal operator of the function βg is determined by the soft-thresholding operator which acts on the rows of the matrix V,

$$\mathcal{S}_\beta(\mathrm{e}_i^* V) \;=\; \begin{cases} (1 - \beta w_i/\|\mathrm{e}_i^* V\|_2)\,\mathrm{e}_i^* V, & \|\mathrm{e}_i^* V\|_2 \;>\; \beta w_i \\ 0, & \|\mathrm{e}_i^* V\|_2 \;\leq\; \beta w_i. \end{cases}$$

Proximal update (14) results from a local quadratic approximation of f at iteration k, i.e.,

$$\begin{aligned} Y^{k+1} \;:=\; \underset{Y}{\operatorname{argmin}} \;\; & f(Y^k) \,+\, \big\langle \nabla f(Y^k), Y - Y^k\big\rangle \,+ \\ & \frac{1}{2\alpha_k}\,\|Y \,-\, Y^k\|_F^2 \,+\, \gamma\, g(Y) \end{aligned} \tag{16}$$

followed by a completion of squares that brings the problem into the form of (15) with $h := \gamma\alpha_k g$. Here, $\langle\cdot,\cdot\rangle$ denotes the standard matricial inner product $\langle M_1, M_2\rangle := \operatorname{trace}(M_1^* M_2)$ and the expression for the gradient of $f(Y)$ is provided in Appendix C.

1) Initialization and choice of step-size in (14)*:* The PG algorithm is initialized with $Y^0 = K^0 X^0$, where K^0 is a stabilizing feedback gain and X^0 is the corresponding covariance matrix that satisfies (4). The optimal centralized controller resulting from the solution of the algebraic Riccati equation provides a stabilizing initial condition and the closed-loop stability is maintained via step-size selection in subsequent iterations of Algorithm 1. At each iteration of the PG method, we determine the step-size α_k via an adaptive Barzilai-Borwein (BB) initial step-size selection [47], i.e.,

$$\alpha_{k,0} \;=\; \begin{cases} \alpha_m & \text{if } \alpha_m/\alpha_s \;>\; 1/2 \\ \alpha_s \,-\, \alpha_m/2 & \text{otherwise} \end{cases} \tag{17}$$

followed by backtracking to ensure closed-loop stability

$$X(Y^{k+1}) \;\succ\; 0 \tag{18a}$$

and sufficient descent of the objective function $f(Y) + \gamma g(Y)$ resulting from

$$\begin{aligned} f(Y^{k+1}) \;\leq\; & f(Y^k) \,+\, \big\langle \nabla f(Y^k), Y^{k+1} - Y^k\big\rangle \,+ \\ & \frac{1}{2\alpha_k}\,\|Y^{k+1} - Y^k\|_F^2. \end{aligned} \tag{18b}$$

Similar strategies as (18b) were used in [45, Section 3]. Here, the "steepest descent" step-size α_s and the "minimum residual" step-size α_m are given by,

$$\begin{aligned} \alpha_s \;&=\; \frac{\big\langle Y^k - Y^{k-1}, Y^k - Y^{k-1}\big\rangle}{\big\langle Y^k - Y^{k-1}, \nabla f(Y^k) - \nabla f(Y^{k-1})\big\rangle} \\ \alpha_m \;&=\; \frac{\big\langle Y^k - Y^{k-1}, \nabla f(Y^k) - \nabla f(Y^{k-1})\big\rangle}{\big\langle \nabla f(Y^k) - \nabla f(Y^{k-1}), \nabla f(Y^k) - \nabla f(Y^{k-1})\big\rangle}. \end{aligned}$$

If $\alpha_s < 0$ or $\alpha_m < 0$, the step-size from the previous iteration is used; see [47, Section 4.1] for additional details.

2) Stopping criterion: We employ a combined condition that terminates the algorithm when either the relative residual

$$r_r^{k+1} \;=\; \frac{\|r^{k+1}\|}{\max\{\|\nabla f(Y^{k+1})\|, \|(\hat{Y}^{k+1} - Y^{k+1})/\alpha_k\|\} \,+\, \epsilon_r},$$

or the normalized residual

$$r_n^{k+1} \;=\; \frac{\|r^{k+1}\|}{\|r^1\| \,+\, \epsilon_n}$$

are smaller than a desired tolerance. Here, ϵ_r and ϵ_n are small positive constants, the residual is defined as

$$r^{k+1} \;:=\; \nabla f(Y^{k+1}) \,+\, (\hat{Y}^{k+1} \,-\, Y^{k+1})/\alpha_k$$

and $\hat{Y}^{k+1} := Y^k - \alpha_k \nabla f(Y^k)$. While achieving a small r_r guarantees a certain degree of accuracy, its denominator nearly vanishes when $\nabla f(x^\star) = 0$, which happens when $0 \in \partial g(Y^\star)$. In such cases, $\|r_n\|$ provides an appropriate stopping criterion; see [47, Section 4.6] for additional details.

Algorithm 1 Customized PG Algorithm

input: A, B, V, Q, R, $\gamma > 0$, positive constants ϵ_r, ϵ_n, tolerance ϵ, and backtracking constant $c \in (0, 1)$.

initialize: $k = 0$, $\alpha_{0,0} = 1$, $r_r^0 = 1$, $r_n^0 = 1$, choose $Y^0 = K^0 X^0$ where K^0 is a stabilizing feedback gain with corresponding covariance matrix X^0.

while: $r_r^k > \epsilon$ or $r_n^k > \epsilon$

- compute α_k: largest feasible step in $\{c^j \alpha_{k,0}\}_{j=0,1,\ldots}$ such that Y^{k+1} satisfies (18)
- compute r_r^{k+1} and r_n^{k+1}
- $k = k + 1$
- choose $\alpha_{k,0}$ based on (17)

endwhile

output: ϵ-optimal solutions, Y^{k+1} and $X(Y^{k+1})$.

C. Convergence of the proximal gradient algorithm

We next analyze the convergence of the PG algorithm for the strongly convex nonsmooth composite optimization problem,

$$\underset{Y}{\text{minimize}} \quad f(Y) \,+\, \gamma\, g(Y). \tag{19}$$

The PG algorithm (14) with suitable step-size converges with the linear rate $O(\rho^k)$ for some $\rho \in (0,1)$ if: (i) the function f is strongly convex and smooth (i.e., it has a Lipschitz continuous gradient) *uniformly over the entire domain*; and (ii) the function g is proper, closed, and convex [49, Theorem 10.29]. In problem (11), however, condition (i) does not hold over the function domain

$$\mathcal{D}_s := \{Y \in \mathbb{C}^{m\times n} \,|\, \mathcal{A}_1(X(Y)) - \mathcal{B}(Y) = -V,\ X(Y) \succ 0\} \tag{20}$$

corresponding to stabilizing feedback gains $K = YX^{-1}$. To address this issue, we exploit the coercivity [48, Definition 11.10] of common regularization functions and establish linear convergence of the PG method for a class of problems (19) in which the function f satisfies the following assumption.

Assumption 2: For all scalars a, the proper closed convex function f defined over an open convex domain $\mathcal{D}$ has

(i) compact sublevel sets $\mathcal{D}(a) := \{Y \in \mathcal{D} \,|\, f(Y) \leq\, a\}$;

(ii) an L_a-Lipschitz continuous gradient over $\mathcal{D}(a)$;

(iii) a strong convexity modulus $\mu_a > 0$ over $\mathcal{D}(a)$.

Proposition 1 establishes linear convergence of the PG algorithm with sufficiently small fixed step-size. Proofs of all technical results presented here are provided in Appendix D.

Proposition 1: Let the function g be coercive, proper, closed, and convex and let the function f in (19) satisfy conditions (i) and (ii) in Assumption 2. Then, for any initial condition $Y^0 \in \mathcal{D}$ the iterates $\{Y^k\}$ of the PG algorithm (14) with step-size $\alpha \in [0, 1/L_a]$ remain in the sublevel set $D(a)$, with $a > f(Y^0) + \gamma\,(g(Y^0) - g(Y))$, for all Y. Furthermore, if condition (iii) in Assumption 2 also holds, then

$$\|Y^{k+1} \,-\, Y^\star\|_F^2 \;\leq\; (1 \,-\, \mu_a\alpha)\,\|Y^k \,-\, Y^\star\|_F^2 \tag{21}$$

where $Y^\star$ is the globally optimal solution of (19).

We next establish strong-convexity and smoothness for the function f in (12) over its sublevel sets. These properties allow us to invoke Proposition 1 and prove linear convergence for the PG algorithm applied to problem (11) with $\delta = 1$.

Proposition 2: The function f in (12) with the convex domain $\mathcal{D}_s$ given by (20) satisfies Assumption 2.

Our main result is presented in Theorem 1.

Theorem 1: For any stabilizing initial condition $Y^0 \in \mathcal{D}_s$, the iterates of the PG algorithm (14) with step-size $\alpha \in [0, 1/L_a]$ applied to problem (11) with $\delta = 1$ satisfy (21), where μ_a and L_a are the strong convexity modulus and smoothness parameter of the function f over $\mathcal{D}(a)$ with $a > f(Y^0) + \gamma\, g(Y^0)$.

Proof: In addition to being proper, closed, and convex, it is straightforward to verify that the function g given by (13) is coercive, i.e.,

$$\lim_{\|Y\|_F \,\to\, +\infty} g(Y) \;=\; +\infty.$$

Moreover, from the nonnegativity of $g(Y)$, it follows that $a > f(Y^0) + \gamma\,(g(Y^0) - g(Y))$ for all Y. Thus, the result follows from combining Propositions 1 and 2. ∎

Remark 1: Proposition 1 proves that the PG algorithm with fixed step-size $\alpha \in (0, 1/L_a]$ converges at the linear rate $O((1 - \mu_a\alpha)^k)$. A linear rate $O(\rho^k)$ with $\rho = 1 - \min\{1/(\sqrt{2}L_a), c/L_a\}$ can also be guaranteed using the adaptive step-size selection method of Section IV-B1; see Appendix E.

The next lemma provides an expression for the smoothness parameter of the function f over its sublevel sets. We note that this parameter depends on problem data.

Lemma 1: Over any non-empty sublevel set $\mathcal{D}(a)$, the gradient $\nabla f(Y)$ is Lipschitz continuous with parameter

$$L_a \;=\; \frac{2\,\lambda_{\max}(R)}{\nu}\left(1 \,+\, \frac{\sqrt{a}\,\|\mathcal{A}_1^{-1}\mathcal{B}\|_2}{\sqrt{\nu\,\lambda_{\min}(R)}}\right)^2 \tag{22a}$$

where the positive scalar

$$\nu \;:=\; \frac{\lambda_{\min}^2(V)}{4\,a}\left(\frac{\|A\|_2}{\sqrt{\lambda_{\min}(Q)}} \,+\, \frac{\|B\|_2}{\sqrt{\lambda_{\min}(R)}}\right)^{-2} \tag{22b}$$

gives the lower bound $\nu I \preceq X(Y)$ on the covariance matrix.

Remark 2: While Lemma 1 provides an expression for the smoothness parameter, we have recently established an explicit expression for the strong convexity modulus [50]

$$\mu_a = \frac{2\,\lambda_{\min}(R)\lambda_{\min}(Q)}{\Big(a^{1/2} + a^2\|\mathcal{B}\|_2\big(\lambda_{\min}(Q)\lambda_{\min}(V)\sqrt{\nu\lambda_{\min}(R)}\big)^{-1}\Big)^2}.$$

Based on Theorem 1, the explicit expressions for parameters L_a and μ_a determine a theoretical bound of $1 - \mu_a/L_a$ on the linear convergence rate of the PG algorithm with step-size $\alpha_k = 1/L_a$. It should be noted that this bound depends on the initial condition Y^0 and problem data.

D. Method of multipliers for covariance completion

We handle the additional constraint in the covariance completion problem by employing the Method of Multipliers (MM). MM is the dual ascent algorithm applied to a smooth variant of the dual problem and it is widely used for solving constrained nonlinear programming problems [51]–[53].

The MM algorithm for constrained optimization problem (11) with $\delta = 0$ is given by,

$$Y^{k+1} \;:=\; \underset{Y}{\text{argmin}}\ \mathcal{L}_{\rho_k}(Y;\,\Lambda^k) \tag{23a}$$

$$\Lambda^{k+1} \;:=\; \Lambda^k \,+\, \rho_k\left(\mathcal{A}_2(X(Y^{k+1})) \,-\, G\right) \tag{23b}$$

where $\mathcal{L}_\rho$ is the associated augmented Lagrangian,

$$\mathcal{L}_\rho(Y;\Lambda) \;=\; f(Y) \,+\, \gamma\, g(Y) \,+\, \langle\Lambda, \mathcal{A}_2(X(Y)) \,-\, G\rangle \,+\, \frac{\rho}{2}\,\|\mathcal{A}_2(X(Y)) \,-\, G\|_F^2$$

$\Lambda \in \mathbb{C}^{p\times p}$ is the Lagrange multiplier and ρ is a positive scalar. The algorithm terminates when the primal and dual residuals are small enough. The primal residual is given as

$$\Delta_p \;=\; \|\mathcal{A}_2(X(Y^{k+1})) \;-\; G\|_F \tag{24a}$$

and the dual residual corresponds to the stopping criterion on subproblem (23a)

$$\Delta_d \;=\; \min\{r_r, r_n\} \tag{24b}$$

where the relative and normal residuals, r_r and r_n, are described in Section IV-B.

1) Solution to the Y-minimization problem (23a)*:* For fixed $\{\rho_k, \Lambda^k\}$, minimizing the augmented Lagrangian with respect to Y amounts to finding the minimizer of $\mathcal{L}_{\rho_k}(Y;\Lambda^k)$ subject to $X(Y) \succ 0$. Since $g(Y)$ is nonsmooth, we cannot use standard gradient descent methods to find the update Y^{k+1}. However, similar to Section IV-B, a PG method can be used to solve this subproblem iteratively

$$Y^{j+1} \;=\; \mathbf{prox}_{\beta_j g}\left(Y^j \;-\; \alpha_j \nabla F(Y^j)\right) \tag{25}$$

where j is the inner PG iteration counter, $\alpha_j > 0$ is the step-size, $\beta_j := \alpha_j\gamma$, and $F(Y)$ denotes the smooth part of the augmented Lagrangian $\mathcal{L}_{\rho_k}(Y;\Lambda^k)$,

$$F(Y) \;:=\; f(Y) \;+\; \left\langle \Lambda^k, \mathcal{A}_2(X(Y)) \;-\; G\right\rangle \;+\; \frac{\rho_k}{2}\,\|\mathcal{A}_2(X(Y)) \;-\; G\|_F^2.$$

The expression for the gradient of $F(Y)$ is provided in Appendix F. Similar to Section IV-B, we combine BB step-size initialization with backtracking to satisfy conditions (18).

2) Lagrange multiplier update and choice of step-size in (23b)*:* Customized MM for covariance completion is summarized as Algorithm 2. We follow the procedure outlined in [53, Algorithm 17.4] for the adaptive update of ρ_k. This procedure allows for inexact solutions of subproblem (23a) and a more refined update of the Lagrange multiplier Λ through the adjustment of convergence tolerances on Δ_p and Δ_d. Note that standard convergence results for MM depend on the level of accuracy in solving subproblem (23a) [51, Sections 5.3 and 5.4]. While we establish linear convergence of the PG algorithm for solving this subproblem, we relegate a detailed convergence analysis for the MM algorithm to future work.

E. Computational complexity

Computation of the gradient in both algorithms involves evaluation of X from Y based on (10), a matrix inversion, and solution to the Lyapunov equation. Each of these take $O(n^3)$ operations as well as an $O(mn^2)$ matrix-matrix multiplication. The proximal operator for the function g amounts to computing the 2-norm of all m rows of a matrix with n columns, which takes $O(mn)$ operations. These steps are embedded within an iterative backtracking procedure for selecting the step-size α. If the step-size selection takes q_1 inner iterations the total computation cost for a single iteration of the PG algorithm is $O(q_1 n^3)$. On the other hand, if it takes q_2 iterations for the PG method to converge, the total computation cost for a single iteration of our customized MM algorithm is $O(q_1 q_2 n^3)$. In practice, the backtracking constant c is chosen such that $q_1 < 50$. The computational efficiency of the PG algorithm relative to standard SDP solvers whose worst-case complexity is $O(n^6)$ is thus evident. However, in MM, q_2 depends on the required level of accuracy in solving (23a). While there is a clear trade-off between this level of accuracy and the number of MM steps, careful analysis of such effects is beyond the scope of the current paper. Nonetheless, in Section V-B, we demonstrate that relative to ADMM and SDPT3, customized MM can provide significant speedup.

Algorithm 2 Customized MM Algorithm

input: A, B, C, E, G, V, $\gamma > 0$, and tolerances ϵ_p and ϵ_d.
initialize: $k = 0$, $\rho_0 = 1$, $\rho_{\max} = 10^9$, $\epsilon_0 = 1/\rho_0$, $\eta_0 = \rho_0^{-0.1}$, choose $Y^0 = K^0X^0$ where K^0 is a stabilizing feedback gain with corresponding covariance matrix X^0.
for $k = 0, 1, 2, \ldots$
  solve (23a) using a similar PG algorithm to Algorithm 1 such that $\Delta_d \le \epsilon_k$.
  if $\Delta_p \le \eta_k$
    if $\Delta_p \le \epsilon_p$ and $\Delta_d \le \epsilon_d$
      stop with approximate solution Y^{k+1}
    else
      $\Lambda^{k+1} \;=\; \Lambda^k \;+\; \rho_k\left(\mathcal{A}_2(X(Y^{k+1})) \;-\; G\right)$
      $\rho_{k+1} \;=\; \rho_k, \quad \eta_{k+1} \;=\; \max\{\eta_k\,\rho_{k+1}^{-0.9}, \epsilon_p\}$
      $\epsilon_{k+1} \;=\; \max\{\epsilon_k/\rho_{k+1}, \epsilon_d\}$
    endif
  else
    $\Lambda^{k+1} \;=\; \Lambda^k$
    $\rho_{k+1} \;=\; \{5\rho_k, \rho_{\max}\}, \quad \eta_{k+1} \;=\; \max\{\rho_{k+1}^{-0.1}, \epsilon_p\}$
    $\epsilon_{k+1} \;=\; \max\{1/\rho_{k+1}, \epsilon_d\}$
  endif
endfor
output: optimal solutions, Y^{k+1} and $X(Y^{k+1})$.

F. Comparison with other methods

One way of dealing with the lack of differentiability of the objective function in (11) is to split the smooth and nonsmooth parts over separate variables and to add an additional equality constraint to couple these variables. This allows for the minimization of the augmented Lagrangian via the Alternating Direction Method of Multipliers (ADMM) [54].

In contrast to splitting methods, the algorithms considered in this paper use the PG method to solve the nonsmooth problem in terms of the primal variable Y, thereby avoiding the necessity to update additional auxiliary variables and their corresponding Lagrange multipliers. Moreover, it is important to note that the performance of augmented Lagrangian-based methods is strongly influenced by the choice of ρ. In contrast to ADMM, there are principled adaptive rules for updating the step-size ρ_k in MM. Typically, in ADMM, either a constant step-size is used or the step-size is adjusted to keep the norms of primal and dual residuals within a constant factor of one another [54]. Our computational experiments demonstrate that

the customized proximal algorithms considered in this paper significantly outperform ADMM.

Remark 3: In [55], a customized ADMM algorithm was proposed for solving the optimal sensor and actuator selection problems. In this, the structural Lyapunov constraint on X and Y is dualized via the augmented Lagrangian. While this approach does not rely on the invertibility of operator $\mathcal{A}_1$ (cf. (10)), it involves subproblems that are difficult to solve. Furthermore, as we show in Section V, it performs poorly in practice, especially for large-scale systems. This is because of higher computational complexity ($O(n^5)$ per iteration) of the ADMM algorithm developed in [55].

G. Iterative reweighting and polishing

To obtain sparser structures at lower values of γ, we follow [56] and implement a reweighting scheme in which we run the algorithms multiple times for each value of γ and update the weights as $w_i^{j+1} = 1/(\|\mathrm{e}_i^* Y^j\|_2 + \epsilon)$. Here, Y^j is the solution in the jth reweighting step and the small parameter ϵ ensures that the weights are well-defined.

After we obtain the solution to problem (11), we conduct a *polishing* step to refine the solution based on the identified sparsity structure. For this, we consider the system

$$\dot{x} \;=\; (A \,-\, B_{\rm sp}\, K)\, x \;+\; d$$

where the matrix $B_{\rm sp} \in \mathbb{C}^{n\times q}$ is obtained by eliminating the columns of B corresponding to the identified row sparsity structure of Y, and q denotes the number of retained input channels. For this system, we solve optimization problem (11) with $\gamma = 0$. This step allows us to identify the optimal matrices $Y \in \mathbb{C}^{q\times n}$ and $K \in \mathbb{C}^{q\times n}$ for a system with a lower number of input channels.

V. Computational experiments

We provide two examples to demonstrate the utility of the optimization framework for optimal actuator selection and covariance completion problems and highlight the computational efficiency of our customized algorithms.

A. Actuator selection

The Swift-Hohenberg equation is a partial differential equation that has been widely used as a model for studying pattern formations in hydrodynamics and nonlinear optics [57]. Herein, we consider the linearized Swift-Hohenberg equation around its time independent spatially periodic solution [58]

$$\begin{aligned}\partial_t\, \psi(t,\xi) \;&=\; -\left(\partial_x^2 \,+\, 1\right)^2 \psi(t,\xi) \,-\, c\,\psi(t,\xi) \,+\, f\,\psi(t,\xi) \\ &\quad +\, u(t,\xi) \,+\, d(t,\xi)\end{aligned}$$

with periodic boundary conditions on a spatial domain $\xi \in [0,\, 2\pi]$. Here, the state $\psi(t,\xi)$ denotes the fluctuation field, $u(t,\xi)$ is a spatio-temporal control input, $d(t,\xi)$ is a zero-mean additive white noise, c is a constant bifurcation parameter, and we assume that $f(\xi) := \alpha\cos(\omega\xi)$ with $\alpha \in \mathbb{R}$. Finite dimensional approximation using the spectral collocation method yields the following state-space representation

$$\dot{\psi} \;=\; A\,\psi \,+\, u \,+\, d. \tag{26}$$

TABLE I
Comparison of different algorithms (in seconds) for different number of discretization points n and $\gamma = 10$.

n	CVX	PG	ADMM
32	12.39	6.2	362.4
64	268.11	51.9	4182.6
128	8873.3	875.8	–
256	–	3872.1	–

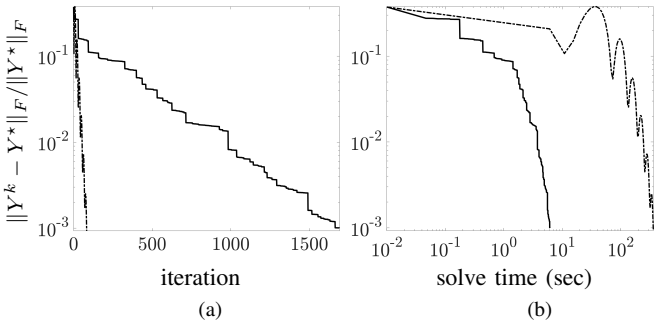


Fig. 2. Convergence curves showing performance of PG ($-$) and ADMM ($-\,\cdot\,-$) vs. (a) the number of outer iterations; and (b) solve times for the Swift-Hohenberg problem with $n = 32$ discretization points and $\gamma = 10$. Here, $Y^\star$ is the optimal value for Y.

For $c = -0.2$, $\alpha = 2$, and $\omega = 1.25$, the linearized dynamical generator has two unstable modes. We set $Q = I$ and $R = 10I$ and solve the actuator selection problem (problem (11) with $\delta = 1$) for 32, 64, 128 and 256 discretization points and for various values of the regularization parameter γ. For $\gamma = 10$, Table V-A compares the proposed PG algorithm against SDPT3 [59] and the ADMM algorithm of [55]. Both PG and ADMM were initialized with $Y^0 = K_c X_c$, where K_c and X_c solve the algebraic Riccati equation which specifies the optimal centralized controller. This choice guarantees that $X(Y^0) \succ 0$. All algorithms were implemented in Matlab and executed on a 2.9 GHz Intel Core i5 processor with 16 GB RAM. The parser CVX [60] was used to call the solver SDPT3. The algorithms terminate when an iterate achieves a certain distance from optimality, i.e., $\|X^k - X^\star\|_F/\|X^\star\|_F < \epsilon$ and $\|Y^k - Y^\star\|_F/\|Y^\star\|_F < \epsilon$. The choice of $\epsilon = 10^{-3}$ guarantees that the value of the objective function is within 0.01% of optimality. For $n = 256$, CVX failed to converge. In this case, iterations are run until the relative or normalized residuals defined in Section IV-B2 become smaller than 10^{-2}.

For $n = 128$ and 256, ADMM did not converge to desired accuracy in reasonable time. Typically, the ADMM algorithm of [55] computes low-accuracy solutions quickly but obtaining higher accuracy requires precise solutions to subproblems. The iterative reweighting scheme of Section IV-G can be used to improve the sparsity patterns that are identified by such low-accuracy solutions. Nonetheless, Fig. 2 shows that even for larger tolerances, PG is faster than ADMM.

As γ increases in Problem 2, more and more actuators are dropped and the performance degrades monotonically. For $n =$

64, Fig. 3(a) shows the number of retained actuators as a function of γ and Fig. 3(b) shows the percentage of performance degradation as a function of the number of retained actuators. Figure 3(b) also illustrates that for various numbers of retained actuators, the solution to convex optimization problem (11) with $\delta = 1$ consistently yields performance degradation that is no larger than the performance degradation of a greedy algorithm (that drops actuators based on their contribution to the $\mathcal{H}_2$ performance index). For example, the greedy algorithm leads to 24.6% performance degradation when 30 actuators are retained whereas our approach yields 20% performance degradation for the same number of actuators. This greedy heuristic is summarized in Algorithm 3, where S is the set of actuators and $f(S)$ denotes the performance index resulting from the actuators within the set S. When the individual subproblems for choosing fixed numbers of actuators can be executed rapidly, greedy algorithms provide a viable alternative. There has also been recent effort to prove the optimality of such algorithms for certain classes of problems [61]. However, in our example, the greedy algorithm does not always provide the optimal set of actuators with respect to the $\mathcal{H}_2$ performance index. Relative to the convex formulation, similar greedy techniques yield suboptimal sensor selection for a flexible aircraft wing [7, Section 5.2].

The absence of the sparsity promoting regularizer in Problem 2 leads to the optimal centralized controller which can be obtained from the solution to the algebraic Riccati equation. For $n = 64$, Figs. 4(a) and 4(b) show this centralized feedback gain and the two norms of its rows, respectively. For $\gamma = 0.4$, 21 of 64 possible actuators are retained and the corresponding optimal feedback gain matrix and row norms are shown in Figs. 4(c) and 4(d). Figure 4(d) also shows that a truncation of the centralized feedback gain matrix based on its row-norms (marked by blue $*$ symbols) yields a different subset of actuators than the solution to Problem 2.

Algorithm 3 A greedy heuristic for actuator selection

input: A, B, V, Q R.

initialize: $\Pi \leftarrow \{1, \ldots, m\}$.

while: $|\Pi| > 0$ and $f(S) < \infty$

$$e^* = \underset{e \in \Pi}{\operatorname{argmin}}\, f(\Pi) - f(\Pi \backslash \{e\})$$

$$\Pi \leftarrow \Pi \backslash \{e\}$$

endwhile

output: the set of actuators represented by the set Π.

B. Covariance completion

We provide an example to demonstrate the utility of our approach for the purpose of completing partially available second-order statistics of a three-dimensional channel flow. In an incompressible channel-flow, the dynamics of infinitesimal fluctuations around the parabolic mean velocity profile, $\bar{\mathbf{u}} = [\, U(x_2) \;\; 0 \;\; 0 \,]^T$ with $U(x_2) = 1 - x_2^2$, are governed by the Navier-Stokes equations linearized around $\bar{\mathbf{u}}$. The streamwise,

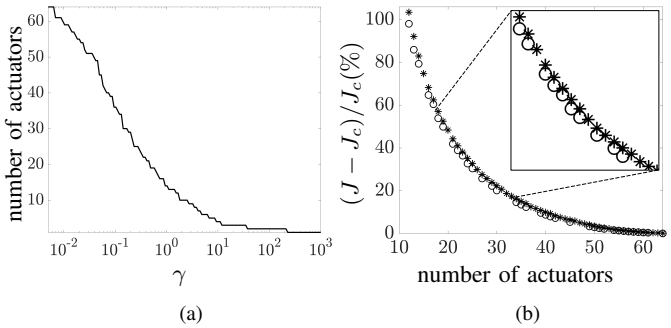


Fig. 3. (a) Number of actuators as a function of the sparsity-promoting parameter γ; and (b) performance comparison of the optimal feedback controller resulting from the regularized actuator selection problem ($\bigcirc$) and from the greedy algorithm ($*$) for the Swift-Hohenberg problem with $n = 64$.

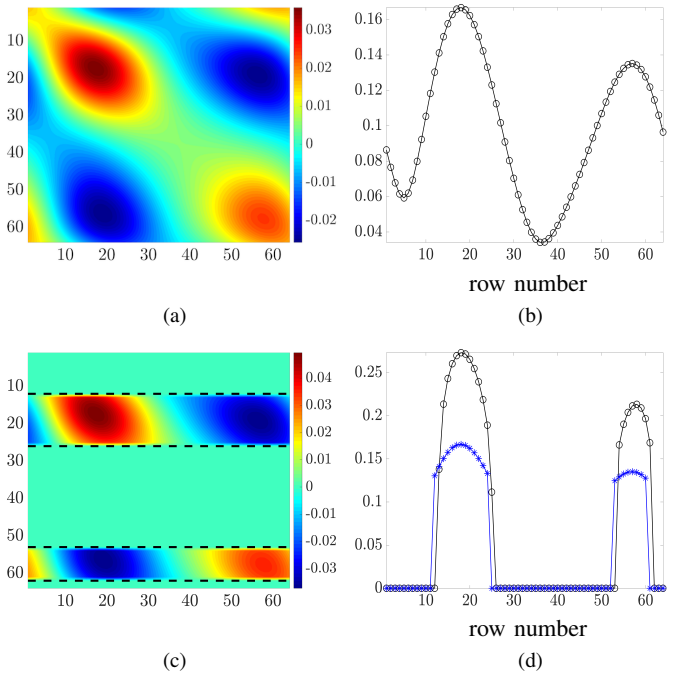
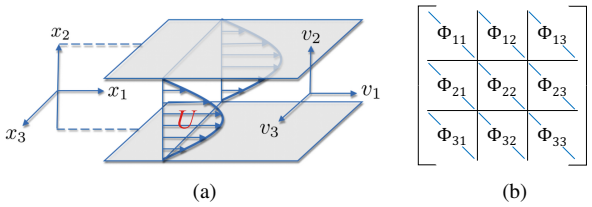


Fig. 4. (a) Optimal centralized feedback gain matrix and (b) its row-norms corresponding to the Swift-Hohenberg dynamics (26) with $n = 64$. (c) The optimal feedback gain matrix and (d) its row-norms ($\circ$) resulting from solving Problem 2 with $\delta = 1$ and $\gamma = 0.4$ in which the rows between the dashed lines have been retained and polished via optimization. The result of truncating the centralized feedback gain matrix based on its row-norms is shown using blue $*$ symbols.

wall-normal, and spanwise coordinates are represented by x_1, x_2, and x_3, respectively; see Fig. 5(a) for geometry. Finite dimensional approximation via application of the Fourier transform in horizontal dimensions (x_1 and x_3) and spatial discretization of the wall-normal dimension (x_2) using N collocation points, yields the state-space representation

$$\begin{aligned} \dot{\boldsymbol{\psi}}(\mathbf{k}, t) &= A(\mathbf{k})\, \boldsymbol{\psi}(\mathbf{k}, t) \;+\; \xi(\mathbf{k}, t) \\ \mathbf{v}(\mathbf{k},\, t) &= C(\mathbf{k})\, \boldsymbol{\psi}(\mathbf{k}, t). \end{aligned} \tag{27a}$$

Here, $\boldsymbol{\psi} = [\, v_2^T \;\; \eta^T \,]^T \in \mathbb{C}^{2N}$ is the state of the linearized model, v_2 and $\eta = \partial_{x_3} v_1 - \partial_{x_1} v_3$ are the normal velocity and vorticity, the output $\mathbf{v} = [\, v_1^T \;\; v_2^T \;\; v_3^T \,]^T \in \mathbb{C}^{3N}$ denotes the fluctuating velocity vector, ξ is a stochastic forcing dis-



(a) (b)

Fig. 5. (a) Geometry of a three-dimensional pressure-driven channel flow. (b) Structure of the matrix $\Phi = \lim_{t\to\infty} \mathbf{E}\,(\mathbf{v}(t)\mathbf{v}^*(t))$, where Φ_{ij} denotes the cross-correlation matrix of components v_i and v_j of the velocity vector $\mathbf{v}$ across the discretization points in the wall-normal direction. Available diagonal entries of the blocks in the velocity covariance matrix Φ determine correlations at the same discretization point.

turbance, $\mathbf{k} = [\,k_1 \;\; k_3\,]^T$ denotes the vector of horizontal wavenumbers, and the input matrix is the identity $I_{2N\times 2N}$. The dynamical matrix $A \in \mathbb{C}^{2N\times 2N}$ and output matrix $C \in \mathbb{C}^{3N\times 2N}$ are described in [13].

We assume that the stochastic disturbance ξ is generated by a low-pass filter with state-space representation

$$\dot{\xi}(\mathbf{k},t) \;=\; -\xi(\mathbf{k},t) \,+\, w(t) \qquad (27b)$$

where w denotes a zero-mean white process with identity covariance matrix. The steady-state covariance of system (27) can be obtained as the solution to the Lyapunov equation

$$\tilde{A}\,\Sigma \,+\, \Sigma\,\tilde{A}^* \,+\, \tilde{B}\,\tilde{B}^* \;=\; 0$$

$$\tilde{A} \;=\; \begin{bmatrix} A & I \\ O & -I \end{bmatrix},\; \tilde{B} \;=\; \begin{bmatrix} 0 \\ I \end{bmatrix},\; \Sigma \;=\; \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^* & \Sigma_{22} \end{bmatrix}.$$

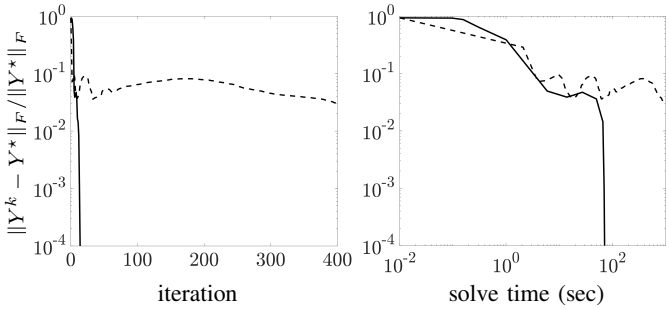
For any $\mathbf{k}$, the matrix $\Sigma_{11} = \lim_{t\to\infty} \mathbf{E}\,(\boldsymbol{\psi}(t)\boldsymbol{\psi}^*(t))$ denotes the steady-state covariance of system (27a) and is related to the steady-state covariance matrix of the output $\mathbf{v}$ via $\Phi(\mathbf{k}) = C(\mathbf{k})\Sigma_{11}(\mathbf{k})C^*(\mathbf{k})$. Figure 5(b) shows the structure of the output covariance matrix Φ.

In this example, we assume that all one-point velocity correlations, i.e., the diagonal entries of all submatrices Φ_{ij} in Fig. 5(b), are known. Owing to experimental and computational limitations, one-point correlations are easier to measure and compute than two-point spatial correlations [10]. While the colored-in-time input process ξ enters across all channels, not all input channels equally impact the state statistics Σ_{11} as the input to state gain differs across different inputs. Herein, we seek a minimal set of input channels with dominant contribution that can lead to a parsimonious perturbation $A - BK$ of the system dynamics. The identified structure represents important feedback mechanisms that are responsible for generating the available statistics when the system is driven by white noise d. Finally, we note that due to the parameterization of system dynamics (27) over wavenumbers $\mathbf{k}$, modification BK also depends on $\mathbf{k}$.

Computational experiments are conducted for a flow with Reynolds number 10^3, the wavenumber pair $(k_1, k_3) = (0, 1)$, for various number of collocation points N in the wall-normal direction (state dimension $n = 2N$), $R = I$, $Q = 0$, and for various values of the regularization parameter γ. Moreover, we assume that system (2) is driven by white process d with covariance $V = I$. We initialize Algorithm 2 with

TABLE II
COMPARISON OF DIFFERENT ALGORITHMS (IN SECONDS) FOR DIFFERENT NUMBER OF DISCRETIZATION POINTS N AND $\gamma = 10$.

N	CVX	MM	ADMM
11	9.3	0.19	3.10
21	97.67	5.6	113.4
31	900	7.19	574.44
51	–	34.76	–
101	–	146.51	–

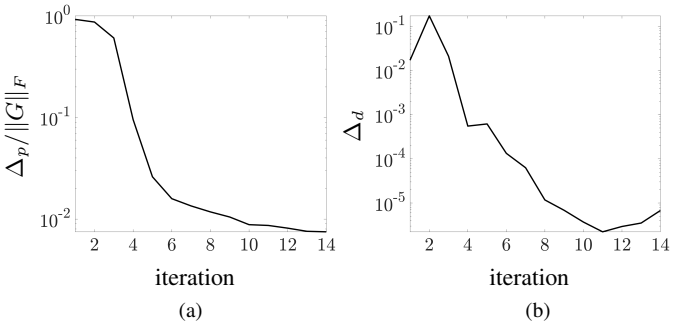


(a) (b)

Fig. 6. Convergence curves showing performance of MM (−) and ADMM (−−) versus (a) the number of outer iterations; and (b) solve times for $N = 31$ collocation points in the normal direction x_2 and $\gamma = 10$. Here, $Y^\star$ is the optimal value for Y.

the optimal centralized controller, $Y^0 := K_c X_c$. Our MM algorithm is compared against SDPT3 and ADMM where CVX is used to call SDPT3. When CVX can compute the optimal solution of Problem 2, for each method, iterations are run until the solutions are within 5% of the CVX solution. For larger problems, iterations are run until the primal and dual residuals satisfy certain tolerances; ϵ_p, $\epsilon_d = 10^{-2}$. For $\gamma = 10$, Table V-B compares various methods based on run times (sec). For $N = 51$ and 101, CVX failed to converge and ADMM did not converge in a reasonable time. Clearly, MM outperforms ADMM. This can also be deduced from Fig. 6, which shows convergence curves for 14 steps of MM and 500 steps of ADMM for $N = 31$ and $\gamma = 10$. For this example, Fig. 7 shows the convergence of MM based on the normalized primal residual $\Delta_p/\|G\|_F$ and the dual residual Δ_d in (24).

We now focus on $N = 51$ collocation points and solve Problem 2 for various values of γ. Since $B = I$, the number



(a) (b)

Fig. 7. Performance of MM for the fluids example with $N = 31$ collocation points in the normal direction x_2 and $\gamma = 10$. (a) normalized primal residual; and (b) dual residual based on (24).

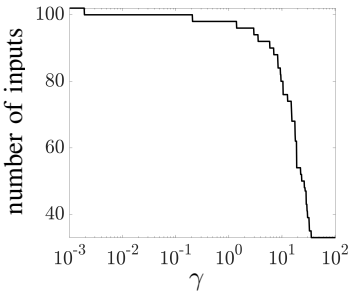


Fig. 8. The γ-dependence of the number of input channels that are retained after solving problem (11) for the channel flow problem with $m = 102$ inputs.

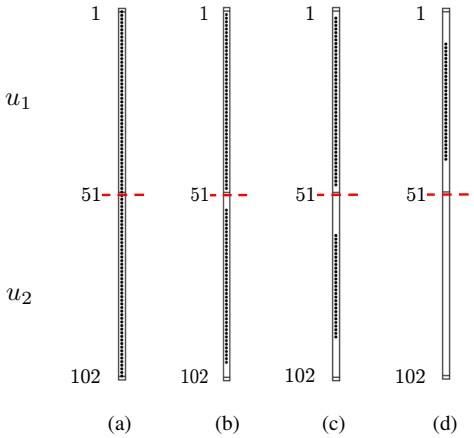


Fig. 9. Active input channels in $u \in \mathbb{C}^{102}$ (black dots) corresponding to row-sparsity of Y^* in problem (11) with (a) $\gamma = 0$, (b) $\gamma = 0.1$, (c) $\gamma = 10$, and (d) $\gamma = 100$, for the channel flow problem.

of inputs u in this case is $m = 102$. Figure 8 shows the γ-dependence of the number of retained input channels that result from solving Problem 2. As γ increases, more and more input channels are dropped. A feature of our framework is that the solution $Y^\star$ determines which inputs in u play a role in matching the available statistics in a way that is consistent with the underlying dynamics. Figure 9 shows the input channels that are retained via optimization for different values of γ. This figure illustrates the dominant role of input channels that enter the dynamics of normal velocity v_2 and away from the boundaries of the channel. In favor of brevity, we do not expand on the physical interpretations of such findings.

Figures 10(b,d) show the streamwise, and the streamwise/normal two-point correlation matrices (Φ_{11} and Φ_{12} in Fig. 5(b)) resulting from solving (11) with $\gamma = 100$. Even though only one-point velocity correlations along the main diagonal of these matrices were used in Problem 2, we observe reasonable recovery of off-diagonal terms of the full two-point velocity correlation matrices and 82% of the original output covariance matrix Φ is recovered. This quality of completion is consistently observed for various values of γ that do not result in the elimination of the critical input channels in the direction of normal velocity, and is an artifact of including the Lyapunov constraint in our formulation. This allows us to simultaneously retain the relevance of the system dynamics

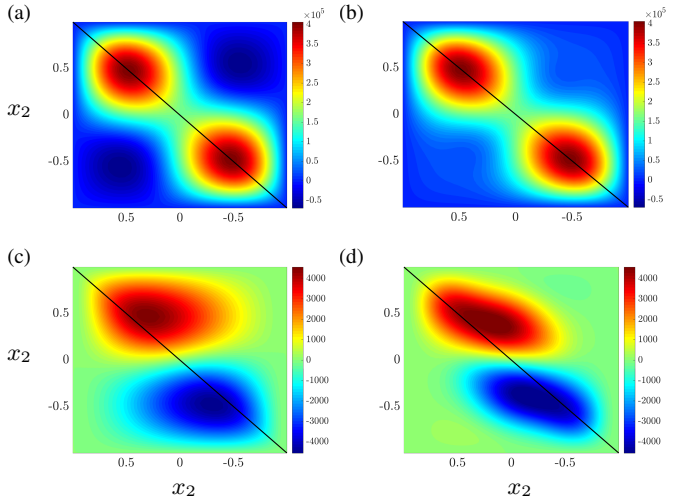


Fig. 10. True covariance matrices of the output velocity field (a, c), and covariance matrices resulting from solving problem (11) (b, d) with $\gamma = 100$ and $N = 51$. (a, b) Streamwise Φ_{11}, and (c, d) streamwise/normal Φ_{12} two-point correlation matrices at $\mathbf{k} = (0, 1)$. One-point correlation profiles that are used as problem data are marked along the main diagonals.

and match the partially available statistics of the underlying dynamical system. Additional details regarding the stochastic modeling of turbulent flow statistics and the importance of predicting two-point velocity correlations can be found in [9].

VI. Concluding Remarks

We have examined two problems that arise in modeling and control of stochastically driven dynamical systems. The first addresses the modeling of second-order statistics by a parsimonious perturbation of system dynamics, while the second deals with the optimal selection of sensors/actuators for estimation/control purposes. We have shown that both problems can be viewed as the selection of suitable feedback gains, guided by similar optimality metrics and subject to closed-loop stability constraints. We cast both problems as optimization problems and use convex surrogates from group-sparsity paradigm to address combinatorial complexity of searching over all possible architectures. While these are SDP representable, the applications that drive our research give rise to the need for scalable algorithms that can handle large problem sizes. We develop a unified algorithmic framework to address both problems using proximal methods. Our algorithms allow handling statistical modeling, as well as sensor and actuator selection, for substantially larger scales than what is amenable to current general-purpose solvers.

In this work, we promote row sparsity by penalizing a weighted sum of row norms of the feedback gain matrix. While we note that iterative reweighting [56] can improve the row-sparsity patterns determined by this approach, the efficacy of more refined approximations, namely low-rank inducing norms [62], [63], for which proximal operators can be efficiently computed, is a subject of future research. Moreover, we will investigate solving these problems via primal-dual algorithms based on the proximal augmented Lagrangian [64], [65], and proximal Newton-type methods [66], [67].

ACKNOWLEDGMENTS

We thank Meisam Razaviyayn for useful discussions.

APPENDIX

A. Sensor selection

Consider the LTI system

$$\begin{aligned} \dot{x} &= A_s\, x \,+\, d \\ y &= C\, x \,+\, \eta \end{aligned}$$

where y denotes measurement data which is corrupted by additive white noise η. If (A, C) is observable, the observer

$$\dot{\hat{x}} \;=\; A_s\, \hat{x} \,+\, L\, C\, (x \,-\, \hat{x}) \,+\, L\, \eta$$

provides an estimate $\hat{x}$ of the state x, where L is the observer gain. When $A_s - LC$ is Hurwitz, the zero-mean estimate of x is given by $\hat{x}$. The Kalman gain minimizes the steady-state variance of $x - \hat{x}$, it is obtained by solving a Riccati equation, and, in general, has no particular structure and uses all available measurements.

Designing a Kalman filter which uses a subset of the available sensors is equivalent to designing a column-sparse Kalman gain matrix L. Based on this, the optimal sensor selection problem can be addressed by solving the following regularized optimization problem

$$\begin{aligned} \underset{L,\, X}{\text{minimize}} \;\; & \operatorname{trace}\left(XV_d \,+\, L\, V_\eta L^* X\right) \,+\, \gamma \sum_{i\,=\,1}^{n} w_i\, \|L\, \mathrm{e}_i\|_2 \\ \text{subject to} \;\; & (A_s - L\, C)^* X \,+\, X(A_s - L\, C) + C^* C \;=\; 0 \\ & X \;\succ\; 0 \end{aligned} \tag{28}$$

where γ, w_i, e_i are as described in Problem 1, $V_d \succ 0$ is the covariance of d, and $V_\eta \succ 0$ is the covariance of η. By setting the problem data in Problem 1 to

$$\begin{aligned} A &= A_s^*, \qquad B \;=\; C^*, \qquad Q \;=\; V_d \\ V &= C^* C, \qquad R \;=\; V_\eta \end{aligned}$$

the solution to problem (28) can be obtained from the solution to the actuator selection problem as X and $L = K^*$.

B. Non-invertibility of $\mathcal{A}_1$

In cases where the matrix X cannot be expressed via (10), since (A, B) is a controllable pair we can center the design variable around a stabilizing controller K_0, i.e., by letting $K := K_0 + K_1$, where K_0 is held fixed and K_1 is the design variable. Based on this, the change of variables introduced in Section III-C yields $Y = K_0 X + K_1 X := K_0 X + Y_1$ and $X(Y_1) = \hat{\mathcal{A}}_1^{-1}(\mathcal{B}(Y_1) - V)$ with

$$\hat{\mathcal{A}}_1(X) \;:=\; (A - BK_0)\, X \;+\; X\, (A - BK_0)^*\,. \tag{29}$$

The resulting optimization problem,

$$\begin{aligned} \underset{Y_1}{\text{minimize}} \;\; & f(Y_1) \,+\, \gamma\, g(Y_1 + K_0 X(Y_1)) \\ \text{subject to} \;\; & (1 \,-\, \delta)\, [\, \mathcal{A}_2(X(Y_1)) \,-\, G\,] \;=\; 0 \\ & X(Y_1) \;\succ\; 0 \end{aligned}$$

involves a nonsmooth term g which is not separable in Y_1, and the smooth term is given by

$$\begin{aligned} f(Y_1) \;:=\; & \operatorname{trace}\left(Q\, X(Y_1)\right) \,+ \\ & \operatorname{trace}\left((Y_1 + K_0 X(Y_1))^*\, R\, (Y_1 + K_0 X(Y_1))\, X^{-1}\right). \end{aligned}$$

Although convex, $g(Y_1 + K_0 X(Y_1))$ does not have an easily computable proximal operator, making it difficult to apply algorithms that are based on proximal methods.

In this case, one may begin with an input matrix B^0 such that the pair (A, B^0) is stabilizable and the nonzero columns of B^0 correspond to a subset of input channels $\mathcal{I}$ that *always* remain active. It would thus be desired to search over input channels from the complement of $\mathcal{I}$ via the following optimization problem

$$\begin{aligned} \underset{Y_1}{\text{minimize}} \;\; & f(Y_1) \,+\, \gamma\, \hat{g}(Y_1) \\ \text{subject to} \;\; & (1 \,-\, \delta)\, [\, \mathcal{A}_2(X(Y_1)) \,-\, G\,] \;=\; 0 \\ & X(Y_1) \;\succ\; 0. \end{aligned}$$

The operator $\hat{\mathcal{A}}_1$ in (29) would now be defined using B^0 and the fixed feedback gain matrix K_0 that abides the row-sparsity structure corresponding to $\mathcal{I}$. The regularization term $\hat{g}(Y_1) := \sum_{i\,\notin\,\mathcal{I}} w_i \|\mathrm{e}_i^* Y_1\|_2$ is used to impose row-sparsity on the *remaining* input channels $i \notin \mathcal{I}$ and has an easily computable proximal operator, thus facilitating the use of proximal methods. It is noteworthy that this approach may also be employed to obtain an operator $\hat{\mathcal{A}}_1$ which is better conditioned than $\mathcal{A}_1$.

The alternative approach would be to avoid this problem altogether by not expressing X as a function of Y and directly dualizing the Lyapunov constraint on X and Y via augmented Lagrangian based methods, e.g., ADMM [55]. However, as we show in Section V, such approaches do not lead to algorithms that are computationally efficient for large problems.

C. Gradient of $f(Y)$ in (14)

To find $\nabla f(Y)$ in (14), we expand $f(Y + \epsilon \tilde{Y})$ around Y for the variation $\epsilon \tilde{Y}$, and collect first-order terms in ϵ. We also account for the variation of X as a result of the variation of Y from

$$(X + \epsilon\, \tilde{X})^{-1} \;=\; X^{-1} \;-\; \epsilon\, X^{-1} \tilde{X}\, X^{-1} \;+\; o(\epsilon)$$

and the linear dependence of $\tilde{X}$ on $\tilde{Y}$, i.e., $\tilde{X} = \mathcal{A}_1^{-1}(\mathcal{B}(\tilde{Y}))$. Here, $o(\epsilon)$ contains higher-order terms in ϵ. Thus, at the kth iteration, the gradient of f with respect to Y is given by,

$$\nabla f(Y^k) \;=\; 2\, R\, Y^k X^{-1} \;-\; 2\, B^*(W_2 \,-\, W_1)$$

where W_1 and W_2 solve the Lyapunov equations

$$\begin{aligned} A^* W_1 \;+\; W_1 A \;+\; X^{-1} Y^{k*} R\, Y^k X^{-1} \;&=\; 0 \\ A^* W_2 \;+\; W_2 A \;+\; Q \;&=\; 0 \end{aligned}$$

and X^{-1} denotes the inverse of $X(Y^k)$.

D. Proofs of Section IV-C

1) Proof of Proposition 1: Without loss of generality, let $\gamma = 1$ and $a = b - c$, where $b = f(Y^0) + g(Y^0)$ and $c < g(Y)$

is a lower bound on the function g. Consider the sublevel set

$$\mathcal{E}(b) \;:=\; \{\, Y \in \mathcal{D} \mid f(Y) + g(Y) \;\leq\; b \,\}.$$

It is easy to verify that $Y^0 \in \mathcal{E}(b) \subset \mathcal{D}(a)$. For a given $Y \in \mathcal{E}(b)$, let $P\colon \mathbb{R}^+ \to \mathbb{C}^{m\times n}$ be defined as

$$P(\alpha) \;=\; \mathbf{prox}_{\alpha g}(Y \;-\; \alpha \nabla f(Y))\,.$$

In what follows, we show that $P(\alpha) \in \mathcal{E}(b)$ for all $\alpha \in [0, 1/L_a]$, with L_a being the Lipschitz continuity parameter of $\nabla f(Y)$ over the sublevel set $\mathcal{D}(a)$. Since $P(0) = Y$, this holds trivially for $\alpha = 0$. For $\alpha > 0$, consider the quadratic function $l_\alpha : \mathbb{C}^{m\times n} \to \mathbb{R}$,

$$l_\alpha(\hat{Y}) \;:=\; f(Y) + \left\langle \nabla f(Y), \hat{Y} \;-\; Y \right\rangle + \frac{1}{2\alpha}\,\|\hat{Y} \;-\; Y\|^2$$

which satisfies

$$f(\hat{Y}) \;\leq\; l_\alpha(\hat{Y}) \tag{30}$$

for all $\hat{Y} \in \mathcal{D}(a)$ and $\alpha \in (0, 1/L_a]$. Inequality (30) follows from the L_a-Lipschitz continuity of $\nabla f(Y)$ over $\mathcal{D}(a)$ (Descent Lemma). Moreover, by definition,

$$P(\alpha) \;=\; \underset{\hat{Y}\,\in\,\mathbb{C}^{m\times n}}{\operatorname{argmin}}\; l_\alpha(\hat{Y}) + g(\hat{Y}) \tag{31}$$

and $l_\alpha(Y) = f(Y)$, which yields

$$l_\alpha(P(\alpha)) + g(P(\alpha)) \;\leq\; f(Y) + g(Y) \;\leq\; b \tag{32}$$

for all positive α. We next show that $P(\alpha) \in \mathcal{D}(a)$ for all $\alpha \in (0, 1/L_a]$, which allows us to substitute $P(\alpha)$ for $\hat{Y}$ in (30) and complete the proof by combining (30) and (32).

Since the functions g and $\|\cdot\|^2$ are coercive, it follows from [48, Theorem 26.20] that the map $P(\alpha)$ is continuous. Let $\alpha_1 \in (0, +\infty]$ be the smallest scalar such that $f(P(\alpha_1)) \geq a$. Such α_1 exists and $f(P(\alpha_1)) = a$ because the set $\mathcal{D}$ is open, the function $f(P(\alpha))$ is continuous, and $f(P(0)) = f(Y) < a$. We next show that $\alpha_1 > 1/L_a$. For the sake of contradiction, suppose $\alpha_1 \leq 1/L_a$. By substituting $P(\alpha_1)$ for $\hat{Y}$ in Eq. (30), using (32), and $c < g(P(\alpha_1))$, we arrive at

$$a \;=\; f(P(\alpha_1)) \;\leq\; l_{\alpha_1}(P(\alpha_1)) \;<\; b - c$$

which contradicts with $a = b - c$. Thus, $\alpha_1 > 1/L_a$ and $P(\alpha) \in \mathcal{D}(a)$ for all $\alpha \in [0, 1/L_a]$. Furthermore, based on this, substituting $P(\alpha)$ in (30) and utilizing (32) gives

$$f(P(\alpha)) + g(P(\alpha)) \;\leq\; b$$

which in turn implies $P(\alpha) \in \mathcal{E}(b)$.

Based on the fact that we can restrict the domain of the optimization problem (19) to the sublevel set $\mathcal{D}(a)$, the rest of the proof about the convergence rate follows from the proof of [49, Theorem 10.29].

2) Proof of Proposition 2: It is straightforward to verify that the set $\mathcal{D}_s$ is open. We first utilize previously established properties of the set of stabilizing feedback gains to prove that the sublevel sets $\mathcal{D}(a)$ of the function $f(Y)$ are compact. We then prove that for any convex compact set $\mathcal{C} \subset \mathcal{D}_s$ there exist a strong convexity modulus $\mu > 0$ and a smoothness parameter $L > 0$ for $f(Y)$ over $\mathcal{C}$.

Consider the function $Y(K) := KX(K)$ where K belongs to the set of stabilizing feedback gains $\mathcal{K}_s$ and $X(K) \succ 0$ is the unique solution to the algebraic Lyapunov equation (4). The function $X(K)$ is continuous and the sublevel sets of the function $f(Y(K))$

$$\mathcal{K}(a) \;:=\; \{K \in \mathcal{K}_s \mid f(Y(K)) \;\leq\; a\}$$

are compact [68]. Since the sublevel set $\mathcal{D}(a)$ is the image of the compact set $\mathcal{K}(a)$ under the continuous map $Y(K)$, it follows that $\mathcal{D}(a)$ is also compact.

The next lemma provides an expression for the second-order approximation of the function $f(Y)$.

Lemma 2: The Hessian of the function $f(Y)$ satisfies

$$\left\langle \tilde{Y}, \nabla^2 f(Y; \tilde{Y}) \right\rangle \;=\; 2\,\|R^{\frac{1}{2}}(\tilde{Y} - YX^{-1}\mathcal{M}(\tilde{Y}))X^{-\frac{1}{2}}\|_F^2$$

where $X = \mathcal{A}_1^{-1}(\mathcal{B}(Y) - V)$ and $\mathcal{M}(\tilde{Y}) := \mathcal{A}_1^{-1}(\mathcal{B}(\tilde{Y}))$.

Proof: For any $Y \in \mathcal{D}_s$ and $X = \mathcal{A}_1^{-1}(\mathcal{B}(Y) - V)$, the function $f(X, Y)$ in Problem 2 reduces to $f(Y)$. The second-order approximation of $f(Y)$ is determined by

$$f(Y + \tilde{Y}) \;\approx\; f(Y) + \left\langle \nabla f(Y), \tilde{Y} \right\rangle + \frac{1}{2}\left\langle \tilde{Y}, \nabla^2 f(Y; \tilde{Y}) \right\rangle$$

where the matrix $\nabla^2 f(Y; \tilde{Y})$ depends linearly on $\tilde{Y}$.

The gradient $\nabla f(X, Y)$ can be found by expanding $f(X + \epsilon\,\tilde{X}, Y + \epsilon\,\tilde{Y})$ around the ordered pair (X, Y) for the variation $(\epsilon\tilde{X}, \epsilon\tilde{Y})$ and collecting first-order terms in ϵ. This yields,

$$\begin{aligned} \nabla_X f(X, Y) \;&=\; Q \;-\; X^{-1}Y^*RYX^{-1} \\ \nabla_Y f(X, Y) \;&=\; 2\,RYX^{-1}. \end{aligned}$$

To find the Hessian, we expand $\nabla f(X + \epsilon\,\tilde{X}, Y + \epsilon\,\tilde{Y})$,

$$\begin{aligned} \nabla_X f(X \,+\, \epsilon\tilde{X}, Y) \;-\; \nabla_X f(X, Y) \;&=\; \epsilon N_1 \,+\, o(\epsilon) \\ \nabla_X f(X, Y \,+\, \epsilon\tilde{Y}) \;-\; \nabla_X f(X, Y) \;&=\; \epsilon N_2 \,+\, o(\epsilon) \\ \nabla_Y f(X \,+\, \epsilon\tilde{X}, Y) \;-\; \nabla_Y f(X, Y) \;&=\; \epsilon N_3 \,+\, o(\epsilon) \\ \nabla_Y f(X, Y \,+\, \epsilon\tilde{Y}) \;-\; \nabla_Y f(X, Y) \;&=\; \epsilon N_4 \,+\, o(\epsilon) \end{aligned}$$

where the matrices

$$\begin{aligned} N_1 &:=\; X^{-1}Y^*RYX^{-1}\tilde{X}X^{-1} \,+\, X^{-1}\tilde{X}X^{-1}Y^*RYX^{-1} \\ N_2 &:=\; -X^{-1}\tilde{Y}^*RY\,X^{-1} \,-\, X^{-1}Y^*R\tilde{Y}X^{-1} \\ N_3 &:=\; -2RYX^{-1}\tilde{X}\,X^{-1} \\ N_4 &:=\; 2R\tilde{Y}X^{-1} \end{aligned}$$

depend linearly on $\tilde{X}$ and $\tilde{Y}$. Thus, we arrive at

$$\begin{aligned} &\left\langle (\tilde{X}, \tilde{Y}), \nabla^2 f(X, Y; \tilde{X}, \tilde{Y}) \right\rangle \\ &\qquad =\; \left\langle \tilde{X}, N_1 + N_2 \right\rangle \,+\, \left\langle \tilde{Y}, N_3 + N_4 \right\rangle \\ &\qquad\qquad =\; 2\,\|R^{\frac{1}{2}}(\tilde{Y} \,-\, YX^{-1}\tilde{X})X^{-\frac{1}{2}}\|_F^2. \end{aligned}$$

The result follows from $\mathcal{A}_1(\tilde{X}) = \mathcal{B}(\tilde{Y})$. ∎

Let us define $\zeta\colon \mathcal{D}_s \times \mathcal{S}_1 \to \mathbb{R}$ as

$$\zeta(Y, \tilde{Y}) \;=\; \left\langle \tilde{Y}, \nabla^2 f(Y, \tilde{Y}) \right\rangle$$

where $\mathcal{S}_1 := \{\tilde{Y} \in \mathbb{C}^{m\times n} \,|\, \|\tilde{Y}\|_F = 1\}$. To establish strong convexity of $f(Y)$ and Lipschitz continuity of its gradient over

a compact set $\mathcal{C}$, we find a positive lower bound μ and an upper bound L on ζ, $\mu \leq \zeta(Y,\tilde{Y}) \leq L$, for all $(Y,\tilde{Y}) \in \mathcal{C} \times \mathcal{S}_1$.

Using the expression in Lemma 2, it is straightforward to show that the function ζ is continuous. From the continuity of $\zeta(Y,\tilde{Y})$ and the compactness of $\mathcal{C} \times \mathcal{S}_1$, it follows that ζ is bounded on $\mathcal{C} \times \mathcal{S}_1$. This implies the existence of an upper bound L. To find a positive lower bound, let $(Y_o,\tilde{Y}_o)$ be a minimizer of the function $\zeta(Y,\tilde{Y})$ over the set $\mathcal{C} \times \mathcal{S}_1$. The existence of $(Y_o,\tilde{Y}_o)$ follows from the compactness of $\mathcal{C} \times \mathcal{S}_1$ and the continuity of the function ζ. We next show that $\mu := \zeta(Y_o,\tilde{Y}_o) > 0$.

Suppose, for the sake of contradiction, that $\zeta(Y_o,\tilde{Y}_o) = 0$. From Lemma 2, we have

$$\tilde{Y}_o \;=\; K_o\,\tilde{X}_o \tag{33}$$

where $K_o = Y_o X_o^{-1}$, $X_o = X(Y_o)$, and

$$\tilde{X}_o \;=\; \mathcal{M}(\tilde{Y}_o). \tag{34}$$

Combining (34) and the Lyapunov equation (9) yields

$$\mathcal{A}_1(X_o + \tilde{X}_o) \,-\, \mathcal{B}(Y_o + \tilde{Y}_o) \;=\; -V. \tag{35}$$

From (33), we also have

$$Y_o \,+\, \tilde{Y}_o \;=\; K_o\,(X_o \,+\, \tilde{X}_o). \tag{36}$$

Substituting for $Y_o + \tilde{Y}_o$ in (35) from (36), we arrive at

$$\mathcal{A}_1(X_o + \tilde{X}_o) \,-\, \mathcal{B}\left(K_o\,(X_o + \tilde{X}_o)\right) \;=\; -V.$$

Consequently, both X_o and $X_o + \tilde{X}_o$ solve the Lyapunov equation with stabilizing feedback gain K_o, which is a contradiction. Thus, $\zeta(Y_o,\tilde{Y}_o)$ is positive. This completes the proof.

3) Proof of Lemma 1: We first show that the positive definite matrix $X = \mathcal{A}_1(\mathcal{B}(Y) - V)$ satisfies

$$\nu I \;\preceq\; X \tag{37}$$

with ν given by (22b). Let v be the normalized eigenvector corresponding to the smallest eigenvalue of X. Multiplying Lyapunov equation (9) from left and right by v^* and v gives

$$\begin{aligned} v^*(DX^{\frac{1}{2}} \,+\, X^{\frac{1}{2}}D^*)\,v \;&=\; \sqrt{\lambda_{\min}(X)}\;v^*(D \,+\, D^*)\,v \\ &=\; -v^*V\,v \end{aligned}$$

where $D := AX^{1/2} - BYX^{-1/2}$. We thus have

$$\lambda_{\min}(X) \;=\; \frac{(v^*V\,v)^2}{(v^*(D \,+\, D^*)\,v)^2} \;\geq\; \frac{\lambda_{\min}^2(V)}{4\,\|D\|_2^2} \tag{38}$$

where we have applied the Cauchy-Schwarz inequality on the denominator. For $Y \in \mathcal{D}(a)$, we have

$$\mathrm{trace}\left(Q\,X \,+\, Y^*RYX^{-1}\right) \;\leq\; a.$$

This inequality along with $\mathrm{trace}\,(Q\,X) \geq \lambda_{\min}(Q)\,\|X^{1/2}\|_F^2$ and $\mathrm{trace}\left(RYX^{-1}Y^*\right) \geq \lambda_{\min}(R)\,\|YX^{-1/2}\|_F^2$ yields

$$\|X^{1/2}\|_F^2 \;\leq\; a/\lambda_{\min}(Q) \tag{39a}$$

$$\|YX^{-1/2}\|_F^2 \;\leq\; a/\lambda_{\min}(R). \tag{39b}$$

Combination of the triangle inequality, submultiplicative property of the 2-norm, and (39) leads to

$$\|D\|_2 \;\leq\; \sqrt{a}\left(\frac{\sigma_{\max}(A)}{\sqrt{\lambda_{\min}(Q)}} \,+\, \frac{\sigma_{\max}(B)}{\sqrt{\lambda_{\min}(R)}}\right). \tag{40}$$

Inequality (37), with ν given by (22b), follows from combining (38) and (40).

We now show that L_a given by (22a) is a Lipschitz continuity parameter of ∇f. Form (39b) and (37), we have

$$\|YX^{-1}\|_F^2 \;\leq\; \frac{a}{\lambda_{\min}(R)\lambda_{\min}(X)} \;\leq\; \frac{a}{\nu\,\lambda_{\min}(R)}. \tag{41}$$

This allows us to upper bound the quadratic form provided in Lemma 2,

$$\left\langle \tilde{Y}, \nabla^2 f(Y,\tilde{Y})\right\rangle \;=\; 2\,\|R^{\frac{1}{2}}(\tilde{Y} \,-\, YX^{-1}\mathcal{M}(\tilde{Y}))X^{-\frac{1}{2}}\|_F^2.$$

In particular, for $Y \in \mathcal{D}(a)$ and $\tilde{Y}$ with $\|\tilde{Y}\|_F = 1$, we have

$$\begin{aligned} &2\,\|R^{\frac{1}{2}}(\tilde{Y} \,-\, YX^{-1}\mathcal{M}(\tilde{Y}))X^{-\frac{1}{2}}\|_F^2 \\ &\quad\leq\; 2\,\lambda_{\max}(R)\lambda_{\max}(X^{-1})\|\tilde{Y} \,-\, YX^{-1}\mathcal{M}(\tilde{Y})\|_F^2 \\ &\quad\leq\; 2\,\lambda_{\max}(R)\lambda_{\max}(X^{-1})\big(\|\tilde{Y}\|_F \,+\, \|YX^{-1}\mathcal{M}(\tilde{Y})\|_F\big)^2 \\ &\quad\leq\; \frac{2\,\lambda_{\max}(R)}{\nu}\left(1 \,+\, \frac{\sqrt{a}\|\mathcal{M}\|_2}{\sqrt{\nu\,\lambda_{\min}(R)}}\right)^2 \;=\; L_a \end{aligned}$$

where the last inequality follows from (37), (41), and the submultiplicative property. This completes the proof.

E. Linear convergence with adaptive step-size selection

We show that iterates $\{Y^k\}$ of the PG algorithm with the backtracking scheme of Section IV-B1 remain in $\mathcal{D}(a)$ and achieve linear convergence. The main challenge in proving the first part of Proposition 1 is to show that (30) holds for $\hat{Y} = P(\alpha_k)$, $\alpha_k \in (0, 1/L_a]$, where $P(\alpha_k)$ is given by (31). However, condition (18b) is itself equivalent to (30) with $\hat{Y} = P(\alpha_k)$. Thus, from the proof of Proposition 1, it is easy to verify that the iterates $\{Y^k\} \subset \mathcal{D}(a)$ and

$$\|Y^{k+1} \,-\, Y^\star\|_F^2 \;\leq\; (1 \,-\, \mu_a\alpha_k)\,\|Y^k \,-\, Y^\star\|_F^2. \tag{42}$$

Here, we show that the adaptive backtracking method generates a sequence $\{\alpha_k\}$ that is lower bounded by a fixed positive scalar. Together with (42), this lower bound yields linear convergence for the PG method with backtracking.

As we discussed in the proof of Proposition 1, the step-size $\alpha_k = 1/L_a$ satisfies conditions (18). Thus, backtracking from a constant initial step-size $\alpha_{k,0}$ would result in a step-size $\alpha_k \geq \min\{\alpha_{k,0}, c/L_a\}$, where c is the backtracking parameter in Algorithm 1. While the initialization $\alpha_{k,0}$ proposed by (17) is not constant, we show that $\alpha_{k,0} \geq 1/(\sqrt{2}L')$, for any

$$L' \;\geq\; \|\Delta_2\|_F/\|\Delta_1\|_F \tag{43}$$

where $\Delta_1 := Y^k - Y^{k-1}$ and $\Delta_2 := \nabla f(Y^k) - \nabla f(Y^{k-1})$. Assuming $\langle \Delta_1, \Delta_2\rangle > 0$, the steepest descent and minimum residual step-sizes are given by $\alpha_s = \|\Delta_1\|_F^2/\langle \Delta_1, \Delta_2\rangle$ and $\alpha_m = \langle \Delta_1, \Delta_2\rangle / \|\Delta_2\|_F^2$, respectively. If $\alpha_m/\alpha_s > 1/2$, then

$\sqrt{2}\,\langle \Delta_1, \Delta_2\rangle > \|\Delta_1\|\|\Delta_2\|$, which yields

$$\alpha_{k,0} \;=\; \frac{\langle \Delta_1, \Delta_2\rangle}{\|\Delta_2\|_F^2} \;>\; \frac{\|\Delta_1\|_F}{\sqrt{2}\|\Delta_2\|_F} \;\geq\; \frac{1}{\sqrt{2}\,L'}.$$

On the other hand, if $\alpha_m/\alpha_s \leq 1/2$, then $\sqrt{2}\,\langle \Delta_1, \Delta_2\rangle \leq \|\Delta_1\|_F\|\Delta_2\|_F$, which yields

$$\alpha_{k,0} \;=\; \frac{\|\Delta_1\|_F^2}{\langle \Delta_1, \Delta_2\rangle} - \frac{\langle \Delta_1, \Delta_2\rangle}{2\|\Delta_2\|_F^2} \;\geq\; \frac{3}{2\sqrt{2}}\frac{\|\Delta_1\|_F}{\|\Delta_2\|_F} \;\geq\; \frac{3}{2\sqrt{2}\,L'}.$$

Since Y^k, $Y^{k-1} \in \mathcal{D}(a)$, inequality (43) holds with $L' = L_a$ the Lipschitz continuity factor of $\nabla f(Y)$ over $\mathcal{D}(a)$. Thus, the resulting step-size satisfies $\alpha_k \geq \min\{1/(\sqrt{2}L_a), c/L_a\}$.

F. Gradient of $F(Y)$ in (25)

Similar to Appendix C, we expand $F(Y + \epsilon\,\tilde{Y})$ around Y for the variation $\epsilon\tilde{Y}$, and collect first-order terms in ϵ. At the kth iteration, the gradient of F with respect to Y is given by,

$$\nabla F(Y^k) \;=\; 2\,Y^k X^{-1} \;-\; 2\,B^*(W_2 \,+\, \rho_k\,W_3 \,-\, W_1),$$

where W_1, W_2, and W_3 solve the Lyapunov equations

$$\begin{aligned} A^*W_1 \,+\, W_1A \,+\, X^{-1}Y^{k*}Y^kX^{-1} \;&=\; 0 \\ A^*W_2 \,+\, W_2A \,+\, \mathcal{A}_2^\dagger\left(\Lambda^k\right) \;&=\; 0 \\ A^*W_3 \,+\, W_3A \,+\, \mathcal{A}_2^\dagger\left(\mathcal{A}_2\left(X(Y^k)\right) \,-\, G\right) \;&=\; 0 \end{aligned}$$

Here, X^{-1} denotes the inverse of $X(Y^k)$ and the adjoint of the operator $\mathcal{A}_2$ is given by $\mathcal{A}_2^\dagger(\Lambda) := C^*\,(E \circ \Lambda)\,C$.

REFERENCES

[1] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994.

[2] G. E. Dullerud and F. Paganini, *A course in robust control theory: a convex approach*. New York: Springer-Verlag, 2000.

[3] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the 2001 American Control Conference*, 2001, pp. 4734–4739.

[4] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[5] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proceedings of the 2004 American Control Conference*, 2004, pp. 3273–3278.

[6] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1235–1256, 2009.

[7] M. R. Jovanović and N. K. Dhingra, "Controller architectures: tradeoffs between performance and structure," *Eur. J. Control*, vol. 30, pp. 76–91, July 2016.

[8] A. Zare, Y. Chen, M. R. Jovanović, and T. T. Georgiou, "Low-complexity modeling of partially available second-order statistics: theory and an efficient matrix completion algorithm," *IEEE Trans. Automat. Control*, vol. 62, no. 3, pp. 1368–1383, March 2017.

[9] A. Zare, M. R. Jovanović, and T. T. Georgiou, "Colour of turbulence," *J. Fluid Mech.*, vol. 812, pp. 636–680, February 2017.

[10] A. Zare, T. T. Georgiou, and M. R. Jovanović, "Stochastic dynamical modeling of turbulent flows," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 3, May 2020, in press; doi:10.1146/annurev-control-053018-023843; also arXiv:1908.09487.

[11] B. F. Farrell and P. J. Ioannou, "Stochastic forcing of the linearized Navier-Stokes equations," *Phys. Fluids A*, vol. 5, no. 11, pp. 2600–2609, 1993.

[12] B. Bamieh and M. Dahleh, "Energy amplification in channel flows with stochastic excitation," *Phys. Fluids*, vol. 13, no. 11, pp. 3258–3269, 2001.

[13] M. R. Jovanović and B. Bamieh, "Componentwise energy amplification in channel flows," *J. Fluid Mech.*, vol. 534, pp. 145–183, July 2005.

[14] R. Moarref and M. R. Jovanović, "Model-based design of transverse wall oscillations for turbulent drag reduction," *J. Fluid Mech.*, vol. 707, pp. 205–240, September 2012.

[15] W. Ran, A. Zare, M. J. P. Hack, and M. R. Jovanović, "Stochastic receptivity analysis of boundary layer flow," *Phys. Rev. Fluids*, vol. 4, no. 9, p. 093901 (28 pages), September 2019.

[16] M. R. Jovanović and B. Bamieh, "Modelling flow statistics using the linearized Navier-Stokes equations," in *Proceedings of the 40th IEEE Conference on Decision and Control*, 2001, pp. 4944–4949.

[17] A. Zare, M. R. Jovanović, and T. T. Georgiou, "Perturbation of system dynamics and the covariance completion problem," in *Proceedings of the 55th IEEE Conference on Decision and Control*, 2016, pp. 7036–7041.

[18] A. Hotz and R. E. Skelton, "Covariance control theory," *Int. J. Control*, vol. 46, no. 1, pp. 13–32, 1987.

[19] K. Yasuda, R. E. Skelton, and K. M. Grigoriadis, "Covariance controllers: A new parametrization of the class of all stabilizing controllers," *Automatica*, vol. 29, no. 3, pp. 785–788, 1993.

[20] K. M. Grigoriadis and R. E. Skelton, "Alternating convex projection methods for covariance control design," *Int. J. Control*, vol. 60, no. 6, pp. 1083–1106, 1994.

[21] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part II," *IEEE Trans. Automat. Control*, vol. 61, no. 5, pp. 1170–1180, 2016.

[22] F. Lin and M. R. Jovanović, "Least-squares approximation of structured covariances," *IEEE Trans. Automat. Control*, vol. 54, no. 7, pp. 1643–1648, July 2009.

[23] M. Zorzi and A. Ferrante, "On the estimation of structured covariance matrices," *Automatica*, vol. 48, no. 9, pp. 2145–2151, 2012.

[24] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, 2016.

[25] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie, "Minimal actuator placement with bounds on control effort," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 67–78, 2016.

[26] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, 2017.

[27] A. Olshevsky, "On (non)supermodularity of average control energy," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1177–1181, 2018.

[28] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[29] S. Liu, S. P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3509–3522, 2016.

[30] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, 2012.

[31] J. L. Rogers, "A parallel approach to optimum actuator selection with a genetic algorithm," in *AIAA Guidance, Navigation, and Control Conference*, 2000, pp. 14–17.

[32] S. Kondoh, C. Yatomi, and K. Inoue, "The positioning of sensors and actuators in the vibration control of flexible systems," *JSME Int. J., Ser. III*, vol. 33, no. 2, pp. 145–152, 1990.

[33] K. Hiramoto, H. Doki, and G. Obinata, "Optimal sensor/actuator placement for active vibration control using explicit solution of algebraic Riccati equation," *J. Sound Vib.*, vol. 229, no. 5, pp. 1057–1075, 2000.

[34] K. K. Chen and C. W. Rowley, "$\mathcal{H}_2$ optimal actuator and sensor placement in the linearised complex Ginzburg-Landau system," *J. Fluid Mech.*, vol. 681, pp. 241–260, 2011.

[35] M. Fardad, F. Lin, and 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.

[36] F. Lin, M. Fardad, and 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.

[37] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2426–2431, September 2013.

[38] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, pp. 845–848, 2012.

[39] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3055–3068, 2014.

[40] B. Polyak, M. Khlebnikov, and 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.

[41] U. Münz, M. Pfister, and P. Wolfrum, "Sensor and actuator placement for linear systems based on $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimization," *IEEE Trans. Automat. Control*, vol. 59, no. 11, pp. 2984–2989, 2014.

[42] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[43] B. Bamieh, F. Paganini, and M. A. Dahleh, "Distributed control of spatially invariant systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1091–1107, July 2002.

[44] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.

[45] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[46] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.

[47] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," arXiv:1411.3406, 2014.

[48] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[49] A. Beck, *First-Order Methods in Optimization*. SIAM, 2017, vol. 25.

[50] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović, "Global exponential convergence of gradient methods over the nonconvex landscape of the linear quadratic regulator," in *Proceedings of the 58th IEEE Conference on Decision and Control*, Nice, France, 2019, pp. 7474–7479.

[51] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. New York: Academic Press, 1982.

[52] D. P. Bertsekas, *Nonlinear programming*. Belmont, MA: Athena Scientific, 1999.

[53] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 2006.

[54] 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. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[55] N. K. Dhingra, M. R. Jovanović, and 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.

[56] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.

[57] M. C. Cross and P. C. Hohenberg, "Pattern formation outside of equilibrium," *Rev. Mod. Phys.*, vol. 65, no. 3, p. 851, 1993.

[58] J. Burke and E. Knobloch, "Localized states in the generalized Swift-Hohenberg equation," *Phys. Rev. E*, vol. 73, no. 5, p. 056211, 2006.

[59] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3-a MATLAB software package for semidefinite programming, version 1.3," *Optim. Methods Softw.*, vol. 11, no. 1-4, pp. 545–581, 1999.

[60] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[61] T. Summers, "Actuator placement in networks using optimal control performance metrics," in *Proceedings of the 55th IEEE Conference on Decision and Control*, 2016, pp. 2703–2708.

[62] C. Grussler, A. Zare, M. R. Jovanović, and A. Rantzer, "The use of the $r*$ heuristic in covariance completion problems," in *Proceedings of the 55th IEEE Conference on Decision and Control*, 2016, pp. 1978–1983.

[63] C. Grussler, "Rank reduction with convex constraints," Ph.D. dissertation, Lund University, 2017.

[64] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Trans. Automat. Control*, vol. 64, no. 7, pp. 2861–2868, July 2019.

[65] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "A second order primal-dual method for nonsmooth convex composite optimization," *IEEE Trans. Automat. Control*, 2017, submitted; also arXiv:1709.01610.

[66] J. D. Lee, Y. Sun, and M. A. Saunders, "Proximal Newton-type methods for minimizing composite functions," *SIAM J. Optim.*, vol. 24, no. 3, pp. 1420–1443, 2014.

[67] L. Stella, A. Themelis, and P. Patrinos, "Forward-backward quasi-Newton methods for nonsmooth optimization problems," *Comput. Optim. Appl.*, vol. 67, no. 3, pp. 443–487, 2017.

[68] H. T. Toivonen, "A globally convergent algorithm for the optimal constant output feedback problem," *Int. J. Control*, vol. 41, no. 6, pp. 1589–1599, 1985.

Armin Zare (S'08–M'17) received the BSc in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2010 and the MSEE and PhD degree in Electrical Engineering from the University of Minnesota, Minneapolis, MN, USA, in 2016. He is an Assistant Professor of Mechanical Engineering at the University of Texas at Dallas, Richardson, TX, USA. He was a Post-doctoral Research Associate in the Ming Hsieh Department of Electrical and Computer Engineering at the University of Southern California, Los Angeles, from 2017 to 2019. His primary research interests are in the modeling and control of complex fluid flows using tools from optimization and systems theory. He was a recipient of the Doctoral Dissertation Fellowship at the University of Minnesota in 2015 and a finalist for the Best Student Paper Award at the American Control Conference in 2014.

Hesameddin Mohammadi (S'17) received the MSc degree from Arizona State University, Tempe, AZ, USA, in 2017 and the BSc degree from Sharif University of Technology, Tehran, Iran, in 2015, both in Mechanical Engineering. He is currently pursuing the PhD degree in the Ming Hsieh Department of Electrical and Computer Engineering at the University of Southern California, Los Angeles, CA, USA. His primary research interests are in large-scale optimization, control, and inference problems.

Neil K. Dhingra (S'10–M'19) received his PhD in Electrical and Computer Engineering from the University of Minnesota, Twin Cities in 2017 where he developed tools for regularized optimization and studied the design of structured controllers for complex large-scale systems. He received his BSE in Electrical Engineering from the University of Michigan, Ann Arbor, in 2010 and has worked with NASA JPL, NASA AFRC, and the WIMS Center. Dr. Dhingra is currently a Research Scientist at Numerica Corporation in Fort Collins, CO.

Tryphon T. Georgiou (M'79–SM'99–F'00) received the Diploma in Mechanical and Electrical Engineering from the National Technical University of Athens, Greece, in 1979 and the PhD degree from the University of Florida, Gainesville, FL, USA, in 1983. He is currently a Chancellor's Professor at the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA. Earlier, he served on the faculty of Florida Atlantic University from 1983 to 1986, Iowa State University from 1986 to 1989, and the University of Minnesota from 1989 to 2016.

He is a recipient of the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society for the years 1992, 1999, 2003, and 2017, a Fellow of the Institute of Electrical and Electronic Engineers (IEEE) and the International Federation of Automatic Control (IFAC), and a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).

Mihailo R. Jovanović (S'00–M'05–SM'13–F'19) received the PhD degree from the University of California at Santa Barbara in 2004. He is a Professor in the Ming Hsieh Department of Electrical and Computer Engineering and the Founding Director of the Center for Systems and Control at the University of Southern California, Los Angeles, CA. He was a faculty member in the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities, MN, from 2004 until 2017, and has held visiting positions with Stanford University and the Institute for Mathematics and its Applications.

Prof. Jovanović is a Fellow of the American Physical Society (APS) and the Institute of Electrical and Electronic Engineers (IEEE). He received a CAREER Award from the National Science Foundation in 2007, the George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society in 2013, and the Distinguished Alumni Award from the Department of Mechanical Engineering at UC Santa Barbara in 2014. Papers of his students were finalists for the Best Student Paper Award at the American Control Conference in 2007 and 2014.