Input-Output Performance of Linear-Quadratic Saddle-Point Algorithms With Application to Distributed Resource Allocation Problems

Simpson-Porco, John W.; Poolla, Bala Kameshwar; Monshizadeh, Nima; Dorfler, Florian

[Link to publication in University of Groningen/UMCG research database](#)

Input–Output Performance of Linear–Quadratic Saddle-Point Algorithms With Application to Distributed Resource Allocation Problems

John W. Simpson-Porco ⬤, *Member, IEEE*, Bala Kameshwar Poolla ⬤, *Student Member, IEEE*,
Nima Monshizadeh ⬤, and Florian Dörfler ⬤, *Member, IEEE*

Abstract—**Saddle-point or primal-dual methods have recently attracted renewed interest as a systematic technique to design distributed algorithms, which solve convex optimization problems. When implemented online for streaming data or as dynamic feedback controllers, these algorithms become subject to disturbances and noise; convergence rates provide incomplete performance information, and quantifying input–output performance becomes more important. We analyze the input–output performance of the continuous-time saddle-point method applied to linearly constrained quadratic programs, providing explicit expressions for the saddle-point $\mathcal{H}_2$ norm under a relevant input–output configuration. We then proceed to derive analogous results for regularized and augmented versions of the saddle-point algorithm. We observe some rather peculiar effects—a modest amount of regularization significantly improves the transient performance, while augmentation does not necessarily offer improvement. We then propose a distributed dual version of the algorithm, which overcomes some of the performance limitations imposed by augmentation. Finally, we apply our results to a resource allocation problem to compare the input–output performance of various centralized and distributed saddle-point implementations and show that distributed algorithms may perform as well as their centralized counterparts.**

Index Terms—**Distributed algorithms, input-output performance, primal-dual dynamics, resource allocation, saddle-point methods, system norms.**

J. W. Simpson-Porco is with the Department of Electrical and Computer Engineering, University of Waterloo, ON N2L 3G1, Canada (e-mail: jwsimpson@uwaterloo.ca).

B. K. Poolla and F. Dörfler are with the Automatic Control Laboratory, Swiss Federal Institute of Technology Zürich, 8092 Zürich, Switzerland (e-mail: bpoolla@ethz.ch; dorfler@ethz.ch).

N. Monshizadeh is with the Engineering and Technology Institute, University of Groningen, 9747 AG Groningen, The Netherlands (e-mail: n.monshizadeh@rug.nl).

I. INTRODUCTION

S ADDLE-POINT methods are a class of continuous-time gradient-based algorithms for solving constrained convex optimization problems. Introduced in the early 1950s [1], [2], these algorithms are designed to seek the saddle points of the optimization problem's Lagrangian function. These saddle points are in one-to-one correspondence with the solutions of the first-order optimality (Karush–Kuhn–Tucker, KKT) conditions, and the algorithm, therefore, drives its internal state toward the global optimizer of the convex program; see [3]–[5] for convergence results.

Recently, these algorithms have attracted renewed attention, e.g., in the context of machine learning [6], in the control literature for solving *distributed* convex optimization problems [7], where agents cooperate through a communication network to solve an optimization problem with minimal or no centralized coordination. Applications of distributed optimization include utility maximization [3], congestion management in communication networks [8], and control in power systems [9]–[14]. While most standard optimization algorithms require centralized information to compute the optimizer, saddle-point algorithms often yield distributed strategies, in which agents perform state updates using only locally measured information and communication with some subset of other agents. We refer the reader to [4] and [15]–[20] for control-theoretic interpretations of these algorithms.

Rather than solving the optimization problem offline, it is desirable to run these distributed algorithms online as controllers, in feedback with system and/or disturbance measurements, to provide references so that the optimizer can be tracked in real time as operating conditions change. Such algorithms offer promise for online optimization, especially in scenarios with streaming data. However, when saddle-point methods are implemented online as controllers, they become subject to disturbances arising from fluctuating parameters and noise (the precise nature of these disturbances being application dependent). The standard method for assessing optimization algorithms—namely, convergence rate analysis—is also insufficient to capture the performance of the algorithm. Indeed, an algorithm with a fast convergence rate would be inappropriate for control applications if it responded poorly to disturbances during transients, or if it greatly amplifies measurement noise in a steady state.

The appropriate tool for measuring dynamic algorithm performance is instead the system norm, as commonly used in feedback system analysis to capture system response to exogenous disturbances. Recent work in this direction includes input-to-state-stability results [21], [22], finite $\mathscr{L}_2$-gain analysis [23], and the robust control framework proposed in [24] and [25]. The purpose of this paper is to continue this line of investigation. In particular, the case of saddle-point algorithms, applied to optimization problems with quadratic objective functions and linear equality constraints, leads to a very tractable instance of this analysis problem, where many basic questions can be asked and accurately answered. Following are the relevant questions.

1) How do saddle-point algorithms amplify disturbances, which may enter the objective function and/or equality constraints?

2) How does performance in the presence of disturbances change when the initial optimization problem is reformulated (e.g., dual or distributed formulations)?

3) How do standard modifications to optimization algorithms, such as regularization and Lagrangian augmentation, affect these results?

Contributions: The three main contributions of this paper are as follows.[1] First, in Section III, we consider the effect of disturbances on the saddle-point dynamics arising from linearly constrained convex quadratic optimization problems. We quantify the input–output performance of the method via the $\mathcal{H}_2$ system norm and—for a relevant input-output configuration—derive an explicit expression for the norm as a function of the algorithm parameters. In Theorem 3.1, we find that the squared $\mathcal{H}_2$ norm scales linearly with the number of disturbances to both the primal and dual variable dynamics.

Second, we study two common modifications to the Lagrangian optimization paradigm: regularization and augmentation. We show that regularization strictly improves the transient $\mathcal{H}_2$ performance of saddle-point algorithms (see Theorem 3.2). However, this improvement in performance is not usually monotone in the regularization parameter, and the system norm may achieve its global minimum at some finite regularization parameter value. For augmented Lagrangian saddle-point methods, we derive an explicit expression for the $\mathcal{H}_2$ norm (see Theorem 3.4); the results show that augmentation may either improve or deteriorate the $\mathcal{H}_2$ performance. For cases when standard augmentation deteriorates performance, we propose an augmented dual distributed saddle-point (ADD-SP) algorithm, which strictly improves performance (see Section IV-B).

Third and finally, in Section V, we apply our results to resource allocation problems, comparing and contrasting the input–output performance of the different algorithms we have considered. The results show that distributed implementations can perform equally well as centralized implementations, but

that significant performance differences can appear between the algorithms once augmentation is considered.

Taken together, these results provide fairly complete answers to questions 1–3 outlined in the introduction, for the class of problems considered. A similar study in an $\mathcal{H}_\infty$ or $\mathscr{L}_2$-gain framework requires a significantly different analysis and is outside the scope of this paper; see Section VI for open directions. As background, in Section II, we review saddle-point algorithms for the relevant class of optimization problems and then recall the basic facts about the $\mathcal{H}_2$ norm as a measure of input-output system performance.

Notation: The $n \times n$ identity matrix is I_n, $\mathbb{0}$ is a matrix of zeros of appropriate dimension, while $\mathbb{1}_n$ (respectively, $\mathbb{0}_n$) are n-vectors of all ones (respectively, zeros). If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $\frac{\partial f}{\partial x} : \mathbb{R}^n \to \mathbb{R}^n$ is its gradient. For $A \in \mathbb{R}^{n \times n}$, $A^\top$ is its transpose and $\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}$ is its trace. If $S \in \mathbb{R}^{r \times n}$ has full row rank, then $SS^\dagger = I_r$, where $S^\dagger = S^\top(SS^\top)^{-1}$ is the Moore–Penrose pseudoinverse of S. For a positive-semidefinite matrix $Q \succeq \mathbb{0}$, $Q^{\frac{1}{2}}$ is its square root. The symbol $\otimes$ denotes the Kronecker product. Given elements $\{a_i\}_{i=1}^{n}$ (scalars, vectors, or matrices), $\mathrm{col}(a_1, \ldots, a_n) = (a_1^\top, \ldots, a_n^\top)^\top$ denotes the vertically concatenation of the elements (assuming compatible dimensions), and $\mathrm{blkdiag}(a_1, a_2, \ldots, a_n)$ is a block matrix with the elements $\{a_i\}$ on the diagonals.

Graphs and graph matrices: A graph is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E}_\mathrm{u})$, where $\mathcal{V}$ is the set of vertices (nodes) and $\mathcal{E}_\mathrm{u}$ is the set of undirected edges (unordered pairs of nodes). The set of neighbors of node $i \in \mathcal{V}$ is denoted by $\mathcal{N}(i)$. If a label $e \in \{1, \ldots, |\mathcal{E}_\mathrm{u}|\}$ and an arbitrary orientation are assigned to each edge, we can define a corresponding directed edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ with elements $e \sim (i, j) \in \mathcal{E}$. The *node-edge incidence matrix* $E \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ is defined componentwise as $E_{ke} = 1$ if node k is the source node of edge e and as $E_{ke} = -1$ if node k is the sink node of edge e, with all other elements being zero. If the graph is connected, then $\ker(E^\top) = \mathrm{Im}(\mathbb{1}_{|\mathcal{V}|})$. A graph is a *tree* (or *acyclic*) if it contains no cycles, and in this case, $\ker(E) = \{\mathbb{0}_{|\mathcal{E}|}\}$.

II. SADDLE-POINT METHODS AND $\mathcal{H}_2$ PERFORMANCE

A. Review of the Saddle-Point Method

We consider the constrained quadratic optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^{n_x}}{\mathrm{minimize}} \quad & J(x) := \frac{1}{2}x^\top Q x + x^\top c \\ \mathrm{subject\ to} \quad & Sx = W_b b \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^{n_x}$, $Q = Q^\top \succ \mathbb{0}$ is positive definite, $c \in \mathbb{R}^{n_x}$ and $b \in \mathbb{R}^{n_b}$ are parameter vectors, and $S \in \mathbb{R}^{n_r \times n_x}$ and $W_b \in \mathbb{R}^{n_r \times n_b}$ with $n_r < n_x$. We make the blanket assumption that S and W_b have full row rank, which simply means that the constraints $Sx = W_b b$ are not redundant. While the right-hand side $W_b b$ of the constraints is apparently overparameterized, this formulation is natural when considering particular problem instances. In the resource allocation problem of Section V, $W_b = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$ and b is a vector of demands; the product $W_b b$ is simply the total demand.

---

[1] A preliminary version of these results with application to power system control appeared in the conference article [26]. In contrast to the conference article, this paper reports the proofs of Theorems 3.1 and 3.4, studies the effect of regularization (see Theorem 3.2), studies both centralized and distributed dual saddle-point approaches (see Proposition 4.1 and Corollary 4.2), and studies the application of these results to resource allocation problems (see Section V).

The problem (1) describes only a subclass of the optimization problems to which saddle-point algorithms are applicable; more generally, one can consider strictly convex costs and convex inequality constraints as well. We restrict our attention to (1), as this case will allow linear time-invariant (LTI) system analysis techniques to be applied, and represent a large enough class of problems to yield some general insights. Intuitively, the performance of the saddle-point algorithm on (1) should indicate a "best" case performance for the general case, as the objective $J(x)$ is smooth and strongly convex, and (1) is free of hard inequality constraints. See Section VI for further discussion. Under these assumptions, the convex problem (1) has a finite optimum, the equality constraints are strictly feasible, and (1) may be equivalently studied through its Lagrange dual with a zero duality gap [27]. The Lagrangian $L : \mathbb{R}^{n_x} \times \mathbb{R}^{n_r} \to \mathbb{R}$ of (1) is

$$L(x,\nu) = \frac{1}{2}x^\top Q x + c^\top x + \nu^\top(Sx - W_b b) \qquad (2)$$

where $\nu \in \mathbb{R}^{n_r}$ is a vector of Lagrange multipliers. By strong duality, the KKT conditions

$$\frac{\partial L}{\partial x}(x,\nu) = \mathbb{0}_{n_x} \iff \mathbb{0}_{n_x} = Qx + S^\top \nu + c$$
$$\frac{\partial L}{\partial \nu}(x,\nu) = \mathbb{0}_{n_r} \iff \mathbb{0}_{n_r} = Sx - W_b b \qquad (3)$$

are necessary and sufficient for optimality. From these linear equations, the unique global optimizer $(x^\star, \nu^\star)$ is

$$\begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -Q^{-1}(S^\top \nu^\star + c) \\ -(SQ^{-1}S^\top)^{-1}(W_b b + SQ^{-1}c) \end{bmatrix}. \qquad (4)$$

While (4) is the exact solution to the optimization problem (1), its evaluation requires centralized knowledge of the matrices S, Q, W_b and the vectors c and b. If any of these parameters change or evolve over time, the optimizer should be recomputed. In many multiagent system applications, the cost matrix Q is diagonal or block-diagonal, and $J(x) = \sum_i \frac{q_i}{2}x_i^2 + c_i x_i$ is, therefore, a sum of local costs. Finally, the constraints encoded in S are often sparse, mirroring the topology of an interaction or communication network between agents. These factors motivate the solution of (1) in an online distributed fashion, where agents in the network communicate and cooperate to calculate the global optimizer.

A simple continuous-time algorithm to seek the optimizer is the *saddle-point* or *primal-dual* method [3]–[5], [28], [29]

$$\mathcal{T}_x \dot{x} = -\frac{\partial}{\partial x}L(x,\nu), \quad \mathcal{T}_\nu \dot{\nu} = +\frac{\partial}{\partial \nu}L(x,\nu)$$

which here reduces to the affine dynamical system

$$\mathcal{T}_x \dot{x} = -Qx - S^\top \nu - c \qquad (5a)$$
$$\mathcal{T}_\nu \dot{\nu} = Sx - W_b b \qquad (5b)$$

where $\mathcal{T}_x, \mathcal{T}_\nu \succ \mathbb{0}$ are positive-definite diagonal matrices of time constants. By construction, the equilibrium points of (5) are in one-to-one correspondence with the solutions of the KKT conditions (3), and the system is internally exponentially stable [5].

Lemma 2.1 (Global convergence to optimizer): The unique equilibrium point $(x^\star, \nu^\star)$ given in (4) of the saddle-point dynamics (5) is globally exponentially stable, with exponential convergence rate $\propto 1/\tau_{\max}$ where $\tau_{\max} = \max(\max_{i \in \{1,\dots,n\}} \mathcal{T}_{x,ii}, \max_{i \in \{1,\dots,r\}} \mathcal{T}_{\nu,ii})$.

Proof of Lemma 2.1: The proof of lemma follows by using the Lyapunov candidate $V(x,\nu) = 0.5(x - x^\star)^\top \mathcal{T}_x(x - x^\star) + 0.5(\nu - \nu^\star)^\top \mathcal{T}_\nu(\nu - \nu^\star) + \varepsilon(\nu - \nu^\star)^\top S\mathcal{T}_x(x - x^\star)$ for $\varepsilon > 0$, i.e., consider the Lyapunov candidate

$$P = \frac{1}{2}\begin{bmatrix} \mathcal{T}_x & \varepsilon \mathcal{T}_x S^\top \\ \varepsilon S \mathcal{T}_x & \mathcal{T}_\nu \end{bmatrix}$$

which is positive definite for sufficiently small ε. With A as the system matrix in (5), we compute that

$$A^T P + PA = -\begin{bmatrix} \Xi & \varepsilon QS^\top/2 \\ \varepsilon SQ/2 & \varepsilon SS^\top \end{bmatrix}$$

where $\Xi = Q - \frac{\varepsilon}{2}(\mathcal{T}_x S^\top \mathcal{T}_\nu^{-1} S + S^\top \mathcal{T}_\nu^{-1} S \mathcal{T}_x)$. Since $\varepsilon > 0$ and S has full row rank, $\varepsilon SS^\top$ is positive definite. Moreover, Ξ is positive definite if ε is sufficiently small. Standard Schur complement results then yield that $A^\top P + PA \prec \mathbb{0}$ if and only if

$$SS^\top - \frac{\varepsilon}{4}SQ\Xi^{-1}QS^\top > \mathbb{0}$$

which holds for ε sufficiently small as $\lim_{\varepsilon \to 0} \Xi = Q$.

As the convergence rate is dictated by the eigenvalues, it follows from a simple variant of [28, Th. 3.6] that the rate is $\propto \min_{ii}(\text{blkdiag}(\mathcal{T}_x^{-1}, \mathcal{T}_\nu^{-1}))$ or $\propto 1/\max(\max_{i \in \{1,\dots,n\}} \mathcal{T}_{x,ii}, \max_{i \in \{1,\dots,r\}} \mathcal{T}_{\nu,ii})$. ∎

With stability settled, in what follows, we will focus exclusively on quantifying transient input–output performance of (5) in the presence of exogenous disturbances.

B. System Performance in the $\mathcal{H}_2$ Norm

Consider the LTI system

$$\dot{x} = Ax + B\eta$$
$$z = Cx \qquad (6)$$

where η is the disturbance input signal and z is the performance output. With $x(0) = \mathbb{0}$, we denote the linear operator from η to z by G. If (6) is input–output stable, its $\mathcal{H}_2$ norm $\|G\|_{\mathcal{H}_2}$ is defined as

$$\|G\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi}\int_{-\infty}^{\infty} \text{Tr}(G(-j\omega)^\top G(j\omega))\,d\omega$$

where $G(j\omega) = C(j\omega I - A)^{-1}B$ is the frequency response of (6).

Another interpretation of $\|G\|_{\mathcal{H}_2}^2$ is as the steady-state variance of the output

$$\|G\|_{\mathcal{H}_2}^2 = \lim_{t \to \infty} \mathbb{E}[z^\top(t)z(t)] \qquad (7)$$

when each component of $\eta(t)$ is stochastic white noise with unit covariance (i.e., $\mathbb{E}[\eta(t)\eta(t')^\top] = \delta(t - t')I$). Therefore, $\|G\|_{\mathcal{H}_2}$ measures how much the output varies in the steady state under stochastic disturbances.

If the state matrix A is Hurwitz, then the $\mathcal{H}_2$ norm is finite and can be computed as [30, Ch. 4]

$$\|G\|_{\mathcal{H}_2}^2 = \mathrm{Tr}(B^\top X B) \tag{8}$$

where the observability Gramian $X = X^\top \succeq \mathbb{0}$ is the unique solution to the Lyapunov equation

$$A^\top X + XA + C^\top C = \mathbb{0}. \tag{9}$$

If the pair (C, A) is observable, then X is positive definite.

III. $\mathcal{H}_2$ PERFORMANCE OF SADDLE-POINT METHODS

We now subject the saddle-point dynamics (5) to disturbances in both the primal and dual equations. Specifically, we assume that the vectors b and c are subject to disturbances $\eta_b \in \mathbb{R}^{n_b}$ and $\eta_c \in \mathbb{R}^{n_x}$ and make the substitutions $b \mapsto b + t_b \eta_b$ and $c \mapsto c + t_c \eta_c$ in the saddle-point dynamics (5). The scalar parameters $t_b, t_c \geq 0$ parameterize the strength of the disturbances and will help us keep track of which terms in the resulting norm expressions arise from which disturbances. As an example, when we study distributed resource allocation problems in Section V, b will have the interpretation of a vector of demands for some resource, and η_b will, therefore, model a fluctuation or disturbance to this nominal demand.

After translating the nominal equilibrium point (4) of the system (5) to the origin,[2] we obtain the LTI system

$$\begin{bmatrix} \mathcal{T}_x \dot{\tilde{x}} \\ \mathcal{T}_\nu \dot{\tilde{\nu}} \end{bmatrix} = \begin{bmatrix} -Q & -S^\top \\ S & \mathbb{0} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix} - \begin{bmatrix} t_c I_{n_x} & \mathbb{0} \\ \mathbb{0} & t_b W_b \end{bmatrix} \begin{bmatrix} \eta_c \\ \eta_b \end{bmatrix} \tag{10a}$$

$$z = \begin{bmatrix} C_1 & \mathbb{0} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix} \tag{10b}$$

where $C_1 \in \mathbb{R}^{n_x \times n_x}$ is an output matrix.

As the system (10) is written in error coordinates, convergence to the saddle-point optimizer $(x^\star, \nu^\star)$ from (4) is equivalent to convergence of $(\tilde{x}(t), \tilde{\nu}(t))$ to the origin. How should we measure this convergence? A natural way is to use the cost matrix Q from the optimization problem (1) as a weighting matrix and to study the performance output $\|z(t)\|_2^2 = \tilde{x}(t)^\top Q \tilde{x}(t)$, which is obtained by choosing $C_1 = Q^{\frac{1}{2}}$. For example, in the context of the resource allocation problems in Section V, these weights describe the relative importance of the various resources.

Theorem 3.1 (Saddle-point performance): Consider the input–output saddle-point dynamics (10) with $Q = Q^\top \succ \mathbb{0}$ diagonal, and let $C_1 = Q^{\frac{1}{2}}$ so that $\|z(t)\|_2^2 = \tilde{x}^\top(t) Q \tilde{x}(t)$. Then, the squared $\mathcal{H}_2$ norm of the saddle-point system (10) is

$$\|G\|_{\mathcal{H}_2}^2 = \frac{t_c^2}{2} \mathrm{Tr}(\mathcal{T}_x^{-1}) + \frac{t_b^2}{2} \mathrm{Tr}(W_b^\top \mathcal{T}_\nu^{-1} W_b). \tag{11}$$

Proof of Theorem 3.1: We will directly construct the unique positive-definite observability Gramian; since the system is internally stable (see Lemma 2.1), this also indirectly establishes

observability [31, Exercise 4.8.1]. Assuming for the moment a block-diagonal observability Gramian $X = \mathrm{blkdiag}(X_1, X_2)$, the Lyapunov equation (9) yields the two equations

$$X_1 \mathcal{T}_x^{-1} Q + Q \mathcal{T}_x^{-1} X_1 - C_1^2 = \mathbb{0} \tag{12a}$$

$$X_2 \mathcal{T}_\nu^{-1} S - S \mathcal{T}_x^{-1} X_1 = \mathbb{0} \tag{12b}$$

with the third independent equation trivially being $\mathbb{0} = \mathbb{0}$. By inspection, the solution to (12a) is diagonal and given by

$$X_1 = \frac{1}{2} \mathcal{T}_x Q^{-1} C_1^2 = \frac{1}{2} \mathcal{T}_x$$

since Q is diagonal and $C_1 = Q^{\frac{1}{2}}$. Clearly, X_1 is positive definite and symmetric. Since S has full row rank, X_2 can be uniquely recovered from (12b) as

$$X_2 = S \mathcal{T}_x^{-1} X_1 S^\dagger \mathcal{T}_\nu = \frac{1}{2} S S^\dagger \mathcal{T}_\nu = \frac{1}{2} \mathcal{T}_\nu.$$

It follows that X_2 is positive definite, and therefore, $X = \frac{1}{2} \mathrm{blkdiag}(\mathcal{T}_x, \mathcal{T}_\nu)$ is the unique positive-definite solution to (9). Since X is block-diagonal, we find from (8) and (10) that

$$\|G\|_{\mathcal{H}_2}^2 = t_c^2 \mathrm{Tr}(\mathcal{T}_x^{-1} X_1 \mathcal{T}_x^{-1}) + t_b^2 \mathrm{Tr}(W_b^\top \mathcal{T}_\nu^{-1} X_2 \mathcal{T}_\nu^{-1} W_b)$$

from which the result follows. ∎

We make two key observations about the result (11). First, (11) is independent of both the cost matrix Q and the constraint matrix S; neither matrix has any influence on the value of the system norm. Second, the expression in (11) scales inversely with the time constants $\mathcal{T}_x$ and $\mathcal{T}_\nu$, which indicates an inherent tradeoff between convergence speed and input–output performance. As a special case of Theorem 3.1, suppose that $\mathcal{T}_x$ and $\mathcal{T}_\nu$ are multiples of the identity matrix, i.e., $\mathcal{T}_x = \overline{\tau}_x I_{n_x}$ and $\mathcal{T}_\nu = \overline{\tau}_\nu I_{n_r}$, and that $W_b = I_{n_r}$, meaning there is one disturbance for each constraint. Then, (11) reduces to

$$\|G\|_{\mathcal{H}_2}^2 = \frac{t_c^2}{2\overline{\tau}_x} n_x + \frac{t_b^2}{2\overline{\tau}_\nu} n_r \tag{13}$$

meaning that the squared $\mathcal{H}_2$ norm scales linearly in the number of disturbances to the primal dynamics and the number of disturbances to the dual dynamics. While this scaling is quite reasonable, the lack of tunable controller gains other than the time constants means that convergence speed and input–output performance are always conflicting objectives. Finally, we note that (13) can be immediately reinterpreted as a design equation for the time constants. That is, given a specified level $\gamma > 0$ of desired $\mathcal{H}_2$ performance, one has $\|G\|_{\mathcal{H}_2} \leq \gamma$ if

$$\min\{\overline{\tau}_x, \overline{\tau}_\nu\} \geq \frac{1}{\gamma^2} \left(\frac{t_c^2 n_x}{2} + \frac{t_b^2 n_r}{2} \right).$$

A. Performance of Regularized Saddle-Point Methods

A common variation of the Lagrangian optimization framework includes a quadratic penalty [32]–[34] on the dual variable ν in the Lagrangian (2). The so-called *regularized* Lagrangian assumes the form

$$L_{\mathrm{reg}}(x, \nu) = \frac{1}{2} x^\top Q x + c^\top x + \nu^\top (Sx - W_b b) - \frac{\epsilon}{2} \|\nu\|_2^2 \tag{14}$$

---

[2]In the remainder of the section, we assume that we have made the change of state variables $\tilde{x} = x - x^\star$ and $\tilde{\nu} = \nu - \nu^\star$, where $\tilde{x}$ and $\tilde{\nu}$ are the error coordinates and $x^\star$ and $\nu^\star$ refer to the equilibria of the saddle-point dynamics.

where $\epsilon > 0$ is small. The regularization term adds concavity to the Lagrangian and has been shown to increase the convergence rate of optimization algorithms. However, this regularization alters the equilibrium of the closed-loop system, which moves from the value in (4) to the new value

$$\begin{bmatrix} x_{\mathrm{reg}}^{\star} \\ \nu_{\mathrm{reg}}^{\star} \end{bmatrix} = \begin{bmatrix} -Q^{-1}(S^{\top}\nu_{\mathrm{reg}}^{\star} + c) \\ -(SQ^{-1}S^{\top} + \epsilon I_{n_r})^{-1}(W_b b + SQ^{-1}c) \end{bmatrix}. \quad (15)$$

The penalty coefficient ϵ is chosen to strike a balance between the convergence rate improvement and the deviation of $(x_{\mathrm{reg}}^{\star}, \nu_{\mathrm{reg}}^{\star})$ from $(x^{\star}, \nu^{\star})$. The Lagrangian (15) also admits a continuous-time saddle-point algorithm with *regularized saddle-point dynamics*

$$\begin{aligned} \mathcal{T}_x \dot{x} &= -Qx - S^{\top}\nu - c \\ \mathcal{T}_\nu \dot{\nu} &= Sx - W_b b - \epsilon\nu. \end{aligned} \quad (16)$$

Quite strikingly, we shall observe that a small regularization, which has a minor effect on the equilibrium, achieves a tremendous improvement in performance.

As we did with the standard saddle-point dynamics, we can shift the equilibrium point $(x_{\mathrm{reg}}^{\star}, \nu_{\mathrm{reg}}^{\star})$ of (16) to the origin and introduce disturbances to the parameters b and c, leading to the input–output model

$$\begin{aligned} \begin{bmatrix} \mathcal{T}_x \dot{\tilde{x}} \\ \mathcal{T}_\nu \dot{\tilde{\nu}} \end{bmatrix} &= \begin{bmatrix} -Q & -S^{\top} \\ S & -\epsilon I_{n_r} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix} - \begin{bmatrix} t_c I_{n_x} & \mathbb{0} \\ \mathbb{0} & t_b W_b \end{bmatrix} \begin{bmatrix} \eta_c \\ \eta_b \end{bmatrix} \\ z &= \begin{bmatrix} Q^{\frac{1}{2}} & \mathbb{0} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix} \end{aligned} \quad (17)$$

where we consider the time constant matrices $\mathcal{T}_x$ and $\mathcal{T}_\nu$ and disturbances η_b and η_c, as in (10).

Theorem 3.2 (Regularized saddle-point performance): Consider the input–output regularized saddle-point dynamics (17) denoted by G_{reg} with $Q = Q^{\top} \succ \mathbb{0}$ a diagonal matrix. Then, the squared $\mathcal{H}_2$ norm $\|G_{\mathrm{reg}}\|_{\mathcal{H}_2}^2$ of the system (17) is upper-bounded by (11) with strict inequality.

Proof of Theorem 3.2: We rewrite (17) in the standard state-space form $G_{\mathrm{reg}} := (A_{\mathrm{reg}}, B, C, \mathbb{0})$, where

$$\begin{bmatrix} \dot{\tilde{x}} \\ \dot{\tilde{\nu}} \end{bmatrix} = \underbrace{\begin{bmatrix} -\mathcal{T}_x^{-1}Q & -\mathcal{T}_x^{-1}S^{\top} \\ \mathcal{T}_\nu^{-1}S & -\mathcal{T}_\nu^{-1}\epsilon I_{n_r} \end{bmatrix}}_{A_{\mathrm{reg}}} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix}$$

$$- \underbrace{\begin{bmatrix} \mathcal{T}_x^{-1}t_c & \mathbb{0} \\ \mathbb{0} & \mathcal{T}_\nu^{-1}W_b t_b \end{bmatrix}}_{B} \begin{bmatrix} \eta_c \\ \eta_b \end{bmatrix}, \quad z = \underbrace{\begin{bmatrix} Q^{\frac{1}{2}} & \mathbb{0} \end{bmatrix}}_{C} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix}.$$

∎

One may verify that A_{reg} is Hurwitz, and that G_{reg} is observable. Consider the observability Gramian from (12), i.e., $X = \frac{1}{2}\mathrm{blkdiag}(\mathcal{T}_x, \mathcal{T}_\nu)$. An easy computation shows that

$$X A_{\mathrm{reg}} + A_{\mathrm{reg}}^{\top} X + C^{\top}C = \begin{bmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{0} & -\epsilon I_{n_r} \end{bmatrix} \preceq \mathbb{0}. \quad (18)$$

We conclude that X is a *generalized* observability Gramian for the regularized system G_{reg}. If X_ϵ is the true observability Gramian for G_{reg}, then $X_\epsilon \neq X$ and $X_\epsilon \preceq X$ [31, Ch. 4.7], and we conclude that

$$\|G_{\mathrm{reg}}\|_{\mathcal{H}_2}^2 = \mathrm{Tr}(B^{\top}X_\epsilon B) \leq \mathrm{Tr}(B^{\top}XB) = \|G\|_{\mathcal{H}_2}^2. \quad (19)$$

It remains only to show that the above inequality holds strictly. Proceeding by contradiction, assume that $\mathrm{Tr}(B^{\top}X_\epsilon B) = \mathrm{Tr}(B^{\top}XB)$, which implies that $\mathrm{Tr}(B^{\top}(X - X_\epsilon)B) = 0$. Since $X - X_\epsilon \succeq \mathbb{0}$, we may write $X - X_\epsilon = F^{\top}F$ for some matrix F and

$$0 = \mathrm{Tr}(B^{\top}(X - X_\epsilon)B) = \mathrm{Tr}(B^{\top}F^{\top}FB).$$

Since B has full row rank, this implies that F must be zero, and thus, $X = X_\epsilon$, which is a contradiction.

Corollary 3.3 (Regularized saddle-point performance with one constraint): Consider the case with one constraint ($n_r = 1$) and one constraint disturbance ($n_b = 1$) with uniform problem parameters $Q = qI_{n_x}, \mathcal{T}_x = \overline{\tau}_x I_{n_x}, \mathcal{T}_\nu = \overline{\tau}_\nu I_{n_r}$ for scalars $q, \overline{\tau}_x, \overline{\tau}_\nu > 0$ and $W_b = 1$. Then, we have

$$\|G\|_{\mathcal{H}_2}^2 - \|G_{\mathrm{reg}}\|_{\mathcal{H}_2}^2 = \alpha_\epsilon t_c^2 + \gamma_\epsilon t_b^2 \quad (20)$$

where $s = \|S\|_2$ and

$$\alpha_\epsilon = \frac{\epsilon s^2}{2(\epsilon q + s^2)(\epsilon\overline{\tau}_x + q\overline{\tau}_\nu)}$$

$$\gamma_\epsilon = \frac{\epsilon(\overline{\tau}_x q\epsilon + q^2\overline{\tau}_\nu + \overline{\tau}_x s^2)}{2\overline{\tau}_\nu(\epsilon q + s^2)(\epsilon\overline{\tau}_x + q\overline{\tau}_v)}.$$

Proof of Corollary 3.3: Let $\Delta := X - X_\epsilon$, where $X = \frac{1}{2}\mathrm{blkdiag}(\mathcal{T}_x, \mathcal{T}_\nu)$ is the observability Gramian of (10) with $C_1 = Q^{\frac{1}{2}}$, and X_ϵ is the observability Gramian of G_{reg}. As noted in the proof of Theorem 3.2, $X \succeq X_\epsilon$, and thus, the matrix Δ is positive semidefinite. Clearly, Δ satisfies

$$A_{\mathrm{reg}}^{\top}(X - \Delta) + (X - \Delta)A_{\mathrm{reg}} + C^{\top}C = \mathbb{0}.$$

By (18), this reduces to

$$A_{\mathrm{reg}}^{\top}\Delta + A_{\mathrm{reg}}\Delta + \begin{bmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \epsilon I_{n_r} \end{bmatrix} = \mathbb{0}. \quad (21)$$

Then, it is easy to see that

$$\|G_{\mathrm{reg}}\|_{\mathcal{H}_2}^2 = \|G\|_{\mathcal{H}_2}^2 - \|G_\epsilon\|_{\mathcal{H}_2}^2 \quad (22)$$

where $\|G\|_{\mathcal{H}_2}^2$ is as in (11), and $G_\epsilon(s) := C_\epsilon(sI - A_{\mathrm{reg}})^{-1}B$ with $C_\epsilon = \begin{bmatrix} \mathbb{0} & \epsilon I_{n_r} \end{bmatrix}$. Therefore, the improvement in the $\mathcal{H}_2$ norm performance is equal to the squared $\mathcal{H}_2$ norm of the axillary system given by G_ϵ. Next, we calculate the $\mathcal{H}_2$ norm of G_ϵ, which requires computing the observability Gramian from (21). Consider the matrix

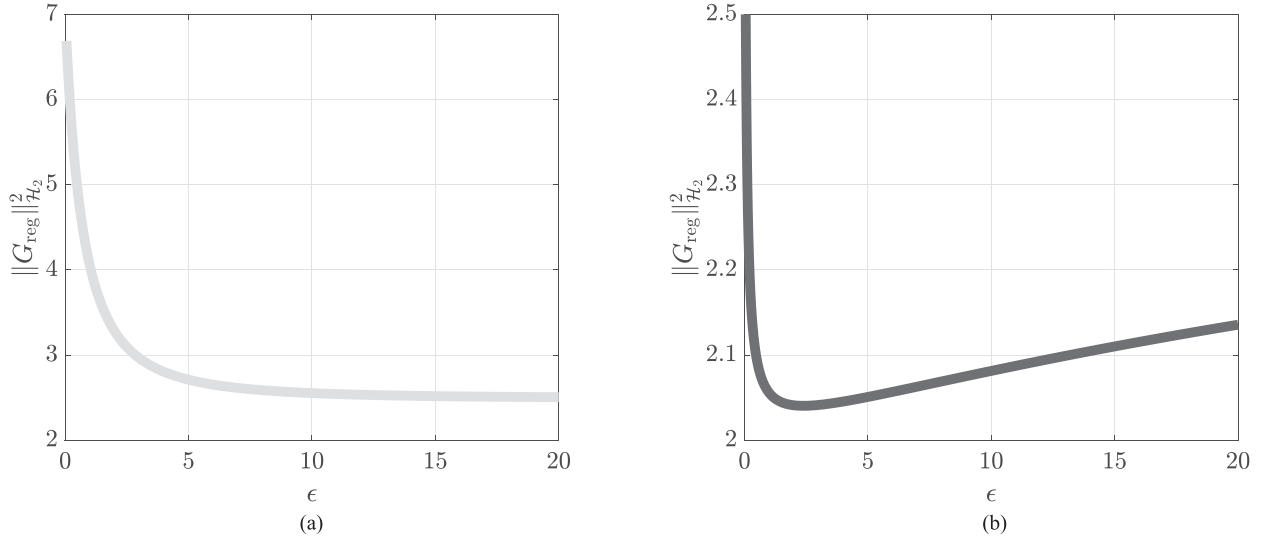$$\overline{\Delta} = \begin{bmatrix} \alpha S^{\top}S & \beta S^{\top} \\ \beta S & \gamma \end{bmatrix} \quad (23)$$

Fig. 1. System norm of regularized dynamics as a function of ϵ, for $S = [0.82\ 0.90\ 0.13\ 0.91\ 0.63]$, $\overline{\tau}_x = \overline{\tau}_\nu = t_c = 1$, and $Q = qI_5$. (a) $q = 3$ and $t_b = 3$. (b) $q = 0.05$ and $t_b = 1$.

where α, β, and γ are constant and positive. A straightforward calculation shows that by choosing

$$
\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \dfrac{s^2}{\overline{\tau}_x} & \dfrac{\epsilon}{\overline{\tau}_\nu} + \dfrac{q}{\overline{\tau}_x} & -\dfrac{1}{\overline{\tau}_\nu} \\ \dfrac{q}{\overline{\tau}_x} & -\dfrac{1}{\overline{\tau}_\nu} & 0 \\ 0 & \dfrac{\overline{\tau}_\nu s^2}{\overline{\tau}_x \epsilon} & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \dfrac{\overline{\tau}_v}{2} \end{bmatrix} \quad (24)
$$

the matrix $\overline{\Delta}$ is a solution to the Lyapunov equation (21). Given the fact that A_{reg} is Hurwitz, this solution is unique, and the matrix $\Delta = \overline{\Delta}$ is the observability Gramian of the system given by G_ϵ. The proof is completed by calculating the inverse in (24), using (8), and noting $\mathrm{Tr}(S^\top S) = s^2$. ∎

Corollary 3.3 quantifies the performance improvement resulting from the regularization in the special case of a single constraint and uniform parameters. For sufficiently large ϵ, this improvement is approximated by $\frac{t_b^2}{2\overline{\tau}_x}$, which coincides with the second term in the right-hand side of (13). This means that, as expected, the constraints do not contribute to the $\mathcal{H}_2$ norm in case the penalty term in the regularized Lagrangian (14) tends to infinity. On the other hand, for $\epsilon \to 0^+$, the $\mathcal{H}_2$ norm of the regularized dynamics (17) clearly converges to the $\mathcal{H}_2$ norm of (10). In general, the improvement in the input–output performance obtained due to regularization is not a monotonic function of the regularization parameter ϵ. To illustrate this, Fig. 1 plots the right-hand side of (20) as a function of ϵ. It is noteworthy that for both the plots, even a modest ϵ improves the performance. Depending on the specific values of the parameters, the maximum performance gain may be achieved as $\epsilon \to \infty$ [see Fig. 1(a)] or at a finite value of ϵ [see Fig. 1(b)].

B. Performance of Augmented Saddle-Point Methods

Another option for improving the $\mathcal{H}_2$ performance of saddle-point methods is to return to the Lagrangian function (2) and instead consider the *augmented Lagrangian* [35]

$$
L_{\mathrm{aug}}(x, \nu) \triangleq L(x, \nu) + \frac{\rho}{2}\|Sx - W_b b\|_2^2 \quad (25)
$$

where we have incorporated the squared constraint $Sx - W_b b = \mathbb{0}_{n_r}$ into the Lagrangian with a gain $\rho \geq 0$. One way to interpret this is that the term $\frac{\rho}{2}\|Sx - W_b b\|_2^2$ adds additional convexity to the Lagrangian in the x variable.

It follows that (x, ν) is a saddle point of $L_{\mathrm{aug}}(x, \nu)$ if and only if it is a saddle point of $L(x, \nu)$, and hence, the optimizer is unchanged. Applying the saddle-point method to the augmented Lagrangian $L_{\mathrm{aug}}(x, \nu)$, we obtain the *augmented saddle-point dynamics*

$$
\begin{aligned}
\mathcal{T}_x \dot{x} &= -(Q + \rho S^\top S)x - S^\top \nu - c + \rho S^\top W_b b \\
\mathcal{T}_\nu \dot{\nu} &= Sx - W_b b.
\end{aligned} \quad (26)
$$

One may verify that as before, the unique stable equilibrium point of (26) is given by (4). We again consider disturbances η_b and η_c and make the substitution $b \mapsto b + t_b \eta_b$ and $c \mapsto c + t_c \eta_c$. After translating the equilibrium point to the origin, we obtain the multiple-input multiple-output system

$$
\begin{bmatrix} \mathcal{T}_x \dot{\tilde{x}} \\ \mathcal{T}_\nu \dot{\tilde{\nu}} \end{bmatrix} = \begin{bmatrix} -(Q + \rho S^\top S) & -S^\top \\ S & \mathbb{0} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix}
$$
$$
- \begin{bmatrix} t_c I_{n_x} & -\rho t_b S^\top W_b \\ \mathbb{0} & t_b W_b \end{bmatrix} \begin{bmatrix} \eta_c \\ \eta_b \end{bmatrix}, \quad (27)
$$
$$
z = \begin{bmatrix} Q^{\frac{1}{2}} & \mathbb{0} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{\nu} \end{bmatrix}.
$$

The additional term $-\rho S^\top S$ in the dynamics (27) complicates the solution of the Lyapunov equation, and we require additional assumptions to obtain an explicit formula. We consider the parametrically uniform case, where $Q = qI_{n_x}$, $\mathcal{T}_x =$

$\overline{\tau}_x I_{n_x}$, and $\mathcal{T}_\nu = \overline{\tau}_\nu I_{n_r}$ for scalars $q, \overline{\tau}_x, \overline{\tau}_\nu > 0$. The next result appeared in [26] without proof.

Theorem 3.4 (Augmented saddle-point performance with uniform parameters): Consider the input–output augmented saddle-point dynamics (27), denoted by G_{aug} under the above assumptions, with performance output $\|z(t)\|_2^2 = \tilde{x}^\top(t) Q\tilde{x}(t) = q\|\tilde{x}(t)\|_2^2$. Then, the squared $\mathcal{H}_2$ norm of the augmented saddle-point system (27) for identically weighted disturbances $W_b = I_{n_r}$ is

$$\|G_{\text{aug}}\|_{\mathcal{H}_2}^2 = \frac{t_c^2}{2\overline{\tau}_x}(n_x - n_r) + \left(\frac{t_b^2}{2\overline{\tau}_\nu} + \frac{t_c^2}{2\overline{\tau}_x}\right)\sum_{i=1}^{n_r}\frac{q}{q+\rho\sigma_i^2}$$
$$+ \frac{t_b^2}{2\overline{\tau}_x}\sum_{i=1}^{n_r}\frac{q\rho^2\sigma_i^2}{q+\rho\sigma_i^2} \quad (28)$$

where σ_i is the ith nonzero singular value of S.

Proof of Theorem 3.4: Under the given assumptions, the system (27) simplifies to

$$\begin{bmatrix}\dot{\tilde{x}} \\ \dot{\tilde{\nu}}\end{bmatrix} = \begin{bmatrix} -\frac{1}{\overline{\tau}_x}(qI_{n_x} + \rho S^\top S) & -\frac{1}{\overline{\tau}_x}S^\top \\ \frac{1}{\overline{\tau}_\nu}S & \mathbb{0} \end{bmatrix}\begin{bmatrix}\tilde{x} \\ \tilde{\nu}\end{bmatrix}$$
$$- \begin{bmatrix} \frac{t_c}{\overline{\tau}_x}I_{n_x} & -\frac{t_b\rho}{\overline{\tau}_x}S^\top W_b \\ \mathbb{0} & \frac{t_b}{\overline{\tau}_\nu}W_b \end{bmatrix}\begin{bmatrix}\eta_c \\ \eta_b\end{bmatrix}, z = \begin{bmatrix} q^{\frac{1}{2}}I_{n_x} & \mathbb{0} \end{bmatrix}\begin{bmatrix}\tilde{x} \\ \tilde{\nu}\end{bmatrix}.$$

Let $S = U\Sigma V^\top$ be the singular value decomposition of S, where $U \in \mathbb{R}^{n_r \times n_r}$ and $V \in \mathbb{R}^{n_x \times n_x}$ are both orthogonal matrices, $\Sigma = \begin{bmatrix} \overline{\Sigma} & \mathbb{0}_{n_r \times (n_x - n_r)} \end{bmatrix}$, and $\overline{\Sigma} \in \mathbb{R}^{n_r \times n_r}$ is the diagonal matrix of nonzero singular values. Consider now the invertible change of variables $\hat{x} = V^\top\tilde{x}$, $\hat{\nu} = U^\top\tilde{\nu}$. In these new coordinates, the dynamics become

$$\begin{bmatrix}\dot{\hat{x}} \\ \dot{\hat{\nu}}\end{bmatrix} = \begin{bmatrix} -\frac{1}{\overline{\tau}_x}(qI_{n_x} + \rho\Sigma^\top\Sigma) & -\frac{1}{\overline{\tau}_x}\Sigma^\top \\ \frac{1}{\overline{\tau}_\nu}\Sigma & \mathbb{0} \end{bmatrix}\begin{bmatrix}\hat{x} \\ \hat{\nu}\end{bmatrix}$$
$$- \begin{bmatrix} \frac{t_c}{\overline{\tau}_x}V^\top & -\frac{t_b\rho}{\overline{\tau}_x}V^\top S^\top W_b \\ \mathbb{0} & \frac{t_b}{\overline{\tau}_\nu}U^\top W_b \end{bmatrix}\begin{bmatrix}\eta_c \\ \eta_b\end{bmatrix}, y = \begin{bmatrix} q^{\frac{1}{2}}V & \mathbb{0} \end{bmatrix}\begin{bmatrix}\hat{x} \\ \hat{\nu}\end{bmatrix}.$$

We now show the observability of the pair (C, A); note that this is equivalent to observability of $(C^\top C, A)$. First note that $\ker(C^\top C)$ is spanned by $\begin{bmatrix} \mathbb{0}_{n_x}^\top & \nu^\top \end{bmatrix}^\top$. Now, suppose that $\begin{bmatrix} \mathbb{0}_{n_x}^\top & \nu^\top \end{bmatrix}^\top$ is an eigenvector of A with eigenvalue λ:

$$\begin{bmatrix} -\frac{1}{\overline{\tau}_x}(qI_{n_x} + \rho\Sigma^\top\Sigma) & -\frac{1}{\overline{\tau}_x}\Sigma^\top \\ \frac{1}{\overline{\tau}_\nu}\Sigma & \mathbb{0} \end{bmatrix}\begin{bmatrix} \mathbb{0}_{n_x} \\ \nu \end{bmatrix} = \lambda\begin{bmatrix} \mathbb{0}_{n_x} \\ \nu \end{bmatrix}.$$

Since by stability $\text{Re}(\lambda) < 0$, the above relation only holds for $\nu = \mathbb{0}_{n_r}$, which shows observability by the eigenvector test. Assuming a block-diagonal observability Gramian $X = \text{blkdiag}(X_1, X_2)$, the Lyapunov equation (9) yields the two

independent equations

$$X_1\frac{1}{\overline{\tau}_x}(qI_{n_x} + \rho\Sigma^\top\Sigma) + \frac{1}{\overline{\tau}_x}(qI_{n_x} + \rho\Sigma^\top\Sigma)X_1 = qI_{n_x} \quad (29a)$$

$$X_2\frac{1}{\overline{\tau}_\nu}\Sigma - \Sigma\frac{1}{\overline{\tau}_x}X_1 = \mathbb{0} \quad (29b)$$

where we have used the fact that $V^\top V = I_{n_x}$. Noting that $\Sigma^\top\Sigma = \text{blkdiag}(\overline{\Sigma}, \mathbb{0}_{(n_x - n_r)\times(n_x - n_r)})$, we find by inspection that the solution to (29a) is diagonal and given by

$$X_{1ii} = \frac{1}{2}\frac{q}{q+\rho\sigma_i^2}\overline{\tau}_x, \qquad i \in \{1, \ldots, n_r\}$$

$$X_{1ii} = \frac{1}{2}\overline{\tau}_x, \qquad i \in \{n_r + 1, \ldots, n_x\}.$$

Observe that X_1 is positive definite and symmetric. The matrix equation (29b) admits a solution X_2 if and only if $\ker(\Sigma) \subseteq \ker(\Sigma X_1)$, which holds in this case since X_1 is diagonal. The lower block X_2 can be uniquely recovered from (29b) as

$$X_2 = \overline{\tau}_x^{-1}(\Sigma X_1\Sigma^\top)(\Sigma\Sigma^\top)^{-1}\overline{\tau}_\nu.$$

A straightforward calculation shows that this is equivalent to the component formula

$$X_{2ii} = \frac{1}{2}\frac{q}{q+\rho\sigma_i^2}\overline{\tau}_\nu, \quad i \in \{1, \ldots, n_r\}.$$

It follows that X_2 is diagonal and positive definite, and therefore, $X = \text{blkdiag}(X_1, X_2)$ is the unique positive-definite solution. A calculation using (8) now shows that

$$\|G_{\text{aug}}\|_{\mathcal{H}_2}^2 = \frac{t_c^2}{\overline{\tau}_x^2}\text{Tr}\left(VX_1V^\top\right) + \frac{t_b^2}{\overline{\tau}_\nu^2}\text{Tr}\left(W_b^\top UX_2U^\top W_b\right)$$
$$+ \frac{t_b^2\rho^2}{\overline{\tau}_x^2}\text{Tr}\left(W_b^\top SVX_1V^\top S^\top W_b\right).$$

For the special case of one disturbance per constraint, i.e., $W_b = I_{n_r}$, the result (28) follows by applying the cyclic property of the trace operation. ∎

Under the assumed restrictions on parameters, Theorem 3.4 generalizes Theorem 3.1, since, when $\rho = 0$, the expression (28) reduces to (13). Consider now the dependence of the expression (28) on the augmentation gain ρ. First, in the case when $t_b = 0$ (meaning the vector b is not subject to disturbances), then as $\rho \to \infty$, the expression (28) reduces to only the first term: in this case, augmentation unambiguously improves input–output performance. In particular, note that a more favorable scaling than in (11) is achieved when n_r is comparable to n_x; see the resource allocation problem in Section V.[3] On the other hand, if $t_b \neq 0$, then as ρ becomes large, the second term in the expression vanishes, while the third term grows without bound. Therefore, a large augmentation gain will lead to poor input–output performance. This behavior is explained by examining (26): the vector b enters the primal dynamics multiplied by ρ, and hence, any noise in b is amplified as ρ grows. To remedy this

---

[3]As an observation, we note that even when $S^\top$ is a sparse matrix, $S^\top S$ typically will not be, and hence, the augmented dynamics (26) may not be immediately implementable as a distributed algorithm. A notable exception is when S is the transposed incidence matrix of a sparse graph, which gives $S^\top S$ as the corresponding sparse graph Laplacian; this will occur in Section V.

deficiency in the augmented approach, the next section exploits a dual formulation of the optimization problem (1).

IV. DUAL AND DISTRIBUTED DUAL METHODS FOR IMPROVED SADDLE-POINT ALGORITHM PERFORMANCE

This section develops an approach to overcome the performance issues of augmented Lagrangian methods observed in Theorem 3.4 when disturbances enter the constraints. To focus on these problematic disturbances, in this section, we ignore possible disturbances to the vector c and set $t_c = 0$. Section IV-A contains a quick examination of dual ascent, before proceeding to a distributed dual formulation in Section IV-B.

A. Centralized Dual Ascent

To begin, we return to the Lagrangian (2) of the optimization problem (1) and compute

$$x^\star(\nu) = \underset{x \in \mathbb{R}^n}{\arg\min}\, L(x, \nu) = -Q^{-1}(c + S^\top \nu).$$

It follows quickly that the dual function $\Phi(\nu)$ is given by

$$\Phi(\nu) = \min_{x \in \mathbb{R}^n} L(x, \nu)$$
$$= -\frac{1}{2}\nu^\top S Q^{-1} S^\top \nu - \nu^\top (S Q^{-1} c + W_b b) - \frac{1}{2} c^\top Q^{-1} c. \quad (30)$$

With the primal variable eliminated, a possible approach is to simply maximize the dual function $\Phi(\nu)$ via gradient ascent. Introducing disturbance inputs $b \mapsto b + t_b \eta_b$ and performance outputs similar to before, and shifting the unique equilibrium point to the origin, one quickly obtains the input–output dual ascent dynamics

$$\mathcal{T}_\nu \dot{\tilde{\nu}} = -S Q^{-1} S^\top \tilde{\nu} - t_b W_b \eta_b$$
$$z = -Q^{-\frac{1}{2}} S^\top \tilde{\nu} \quad (31)$$

where $\mathcal{T}_\nu \succ \mathbb{0}$. The performance of (31) is characterized by the following result.

Proposition 4.1 (Dual ascent performance): The $\mathcal{H}_2$ norm of the input–output dual ascent dynamics (31) is given by

$$\|G\|_{\mathcal{H}_2}^2 = \frac{t_b^2}{2} \mathrm{Tr}(W_b^\top \mathcal{T}_\nu^{-1} W_b).$$

Proof of Proposition 4.1: The Lyapunov equation (9) for this problem takes the form

$$-X \mathcal{T}_\nu^{-1} S Q^{-1} S^\top - S Q^{-1} S^\top \mathcal{T}_\nu^{-1} X + S Q^{-1} S^\top = \mathbb{0}$$

from which we find the unique solution $X = \frac{1}{2}\mathcal{T}_\nu \succ \mathbb{0}$. With $B = [\, t_b \mathcal{T}_\nu^{-1} W_b \,]$, the result follows by applying (8). ∎

Comparing the result of Proposition 4.1 to the unaugmented saddle-point result of Theorem 3.1, we observe that the terms proportional to t_b^2 are identical. Therefore, when considering algorithm performance with disturbances entering the constraints, the primal-dual and pure dual-ascent algorithms achieve identical performance.

B. Distributed Dual Augmented Lagrangian Method

Building off the dual function (30), we now derive a modified augmented Lagrangian algorithm, which can overcome the performance issues posed by disturbances affecting the vector b. The particulars of the derivation below are tailored toward distributed solutions, which will be discussed further in Section V in the context of distributed resource allocation. With this application in mind, we will focus on the case, where $n_b = n_x$, so that the ith component of the disturbance b can be uniquely associated with the ith primal variable x_i; this assumption can be relaxed in the derivation below as long as one uniquely assigns components of b and the associated columns of W_b to a particular agent. We partition each of the following matrices according to their columns:

$$S = \begin{bmatrix} s_1 & s_2 & \cdots & s_{n_x} \end{bmatrix}, W_b = \begin{bmatrix} w_1 & w_2 & \cdots & w_{n_x} \end{bmatrix}.$$

With this partitioning, one may quickly see that for $Q = \mathrm{diag}(q_1, \ldots, q_{n_x})$, the dual function (30) may be written as

$$\Phi(\nu) = \sum_{i=1}^{n_x} \underbrace{\left[-\frac{1}{2q_i} \nu^\top s_i s_i^\top \nu - \nu^\top \left(\frac{c_i}{q_i} s_i + w_i b_i \right) - \frac{c_i^2}{2q_i} \right]}_{:= \widetilde{\Phi}_i(\nu)}.$$

The dual function appears to separate into a sum, except for the common multiplier ν. To complete the separation, for each $i \in \{1, \ldots, n_x\}$, we introduce a local copy $\nu^i \in \mathbb{R}^{n_r}$ of the vector of Lagrange multipliers $\nu \in \mathbb{R}^{n_r}$ and require that $\nu^i = \nu^j$ for all $i, j \in \{1, \ldots, n_x\}$. To enforce these so-called agreement constraints, let $E \in \mathbb{R}^{n_x \times |\mathcal{E}|}$ be the oriented node-edge incidence matrix [36, Ch. 8] of a weakly connected acyclic[4] graph $\mathcal{G} = (\{1, \ldots, n_x\}, \mathcal{E})$, where $\mathcal{E}$ is the set of oriented edges. The dual problem $\mathrm{maximize}_{\nu \in \mathbb{R}^{n_r}}\, \Phi(\nu)$ is then equivalent to the constrained problem[5]

$$\underset{\boldsymbol{\nu} \in \mathbb{R}^{(n_x n_r)}}{\mathrm{minimize}} \quad J_{\mathrm{dual}}(\boldsymbol{\nu}) := -\sum_{i=1}^{n_x} \widetilde{\Phi}_i(\nu^i) \quad (32)$$

$$\text{subject to} \quad (E^\top \otimes I_{n_r})\boldsymbol{\nu} = \mathbb{0}_{(|\mathcal{E}| n_r)}$$

where $\boldsymbol{\nu} = \mathrm{col}(\nu^1, \nu^2, \ldots, \nu^{n_x}) \in \mathbb{R}^{(n_x n_r)}$ is a column vector. Since the graph is acyclic, $E^\top$ has full row rank and, therefore, satisfies our assumption concerning the constraint matrix (see Section II). The key observation now is that the parameters b do not enter into the equality constraints of the optimization problem (32); this permits the full application of augmented Lagrangian techniques for improving the $\mathcal{H}_2$ performance of the saddle-point algorithm. Building the augmented Lagrangian

---

[4]The acyclic assumption implies that $\mathrm{rank}(E^\top) = |\mathcal{E}|$, in line with our assumption from Section II-A that the constraint matrix has full row rank. This assumption can be relaxed at the expense of more complex stability/performance proofs.

[5]Another equivalent formulation can be obtained by using the Laplacian matrix $E E^\top = \mathrm{L} = \mathrm{L}^\top \in \mathbb{R}^{n_x \times n_x}$ of the graph, and using instead the constraint $(\mathrm{L} \otimes I_{n_r})\boldsymbol{\nu} = \mathbb{0}_{(n_x n_r)}$. This formulation is sometimes preferable for multiagent implementations, the analysis of which requires only small modifications from the present analysis. We focus instead on formulations involving the incidence matrix.

(25) for the problem (32), we have

$$L_{\mathrm{dual}}^{\mathrm{aug}} = -\sum_{i=1}^{n_x} \widetilde{\Phi}_i(\nu^i) + \boldsymbol{\mu}^\top (E^\top \otimes I_{n_r})\boldsymbol{\nu} + \frac{\rho}{2}\boldsymbol{\nu}^\top (\mathrm{L} \otimes I_{n_r})\boldsymbol{\nu}$$

where $\boldsymbol{\mu} = \mathrm{col}(\mu^1, \mu^2, \ldots, \mu^{|\mathcal{E}|}) \in \mathbb{R}^{(|\mathcal{E}|n_r)}$ is a stacked vector of Lagrange multipliers $\mu^\ell \in \mathbb{R}^{n_r}$ for $\ell \in \{1, \ldots, |\mathcal{E}|\}$, and $\mathrm{L} = \mathrm{L}^\top = EE^\top \in \mathbb{R}^{n_x \times n_x}$ is the Laplacian matrix of the undirected graph $\mathcal{G}_{\mathrm{u}}$, obtained by ignoring the orientation of the edges in $\mathcal{G}$; we let $\mathcal{N}(i)$ denote the neighbors of vertex i in the graph $\mathcal{G}_{\mathrm{u}}$. Applying the saddle-point method, the dynamics may be written block-componentwise as

$$
\begin{aligned}
\overline{\tau}_\nu \dot{\nu}^i = &-\frac{s_i s_i^\top}{q_i}\nu^i - \left(\frac{c_i}{q_i}s_i + w_i b_i\right) \\
&- \sum_{j:(i,j)\in\mathcal{E}}\mu^{ij} + \sum_{j:(j,i)\in\mathcal{E}}\mu^{ji} \\
&- \rho\sum_{j\in\mathcal{N}(i)}(\nu^i - \nu^j), \quad i \in \{1, \ldots, n_x\}
\end{aligned}
\tag{33}
$$

$$\overline{\tau}_\mu \dot{\mu}^{ij} = \nu^i - \nu^j, \quad (i,j) \in \mathcal{E}.$$

The algorithm (33) is distributed, in that the ith update equation requires only the local parameters s_i, w_i, b_i, c_i, and q_i along with communicated state variables ν^j and μ^{ij}, which come from adjacent nodes and edges in the graph $\mathcal{G}$. We refer to this dynamical system as the ADD-SP dynamics. Following (31), we equip this system with disturbance inputs and performance outputs

$$b_i \mapsto b_i + t_b \eta_i, \quad z_i = -q_i^{-\frac{1}{2}} s_i^\top \tilde{\nu}^i, \quad i \in \{1, \ldots, n_x\}. \tag{34}$$

While a closed-form expression for the $\mathcal{H}_2$ norm of this system is difficult to compute, we can state the following comparative result.

Corollary 4.2 (ADD-SP performance): Consider the ADD-SP dynamics (33), denoted by G_{ADD}, with disturbance inputs $\eta(t)$ and performance output $z(t)$ as in (34), under the assumptions in Theorem 3.4. Then, the squared $\mathcal{H}_2$ norm of the system (33), (34) satisfies the upper bound

$$\|G_{\mathrm{ADD}}\|_{\mathcal{H}_2}^2 \le \frac{t_b^2}{2\overline{\tau}_\nu}\mathrm{Tr}(\mathcal{W}_b^\top \mathcal{W}_b) \tag{35}$$

where $\mathcal{W}_b = \mathrm{blkdiag}(w_1, w_2, \ldots, w_{n_x}) \in \mathbb{R}^{(n_x n_r)\times n_x}$. Moreover, (35) is satisfied with equality if and only if $\rho = 0$.

Proof of Corollary 4.2: To begin, we note that

$$\sum_{i=1}^{n_x}\widetilde{\Phi}_i(\nu^i) = -\frac{1}{2}\boldsymbol{\nu}^\top \mathcal{S}Q^{-1}\mathcal{S}^\top \boldsymbol{\nu} - \boldsymbol{\nu}^\top \mathcal{S}Q^{-1}c - \boldsymbol{\nu}^\top \mathcal{W}_b b$$

where $\mathcal{S} = \mathrm{blkdiag}(s_1, s_2, \ldots, s_{n_x}) \in \mathbb{R}^{(n_x n_r)\times n_x}$. In vector notation after shifting the equilibrium to the origin, the system (33), (34) can be written as

$$\overline{\tau}_\nu \dot{\tilde{\boldsymbol{\nu}}} = -\mathcal{S}Q^{-1}\mathcal{S}^\top \tilde{\boldsymbol{\nu}} - t_b \mathcal{W}_b \eta_b - (E \otimes I_{n_r})\tilde{\boldsymbol{\mu}} - \rho(\mathrm{L} \otimes I_{n_r})\tilde{\boldsymbol{\nu}}$$

$$\overline{\tau}_\mu \dot{\tilde{\boldsymbol{\mu}}} = (E^\top \otimes I_{n_r})\tilde{\boldsymbol{\nu}}$$

$$z = -Q^{-\frac{1}{2}}\mathcal{S}^\top \tilde{\boldsymbol{\nu}}$$

where $\eta = \mathrm{col}(\eta_1, \ldots, \eta_{n_x}) \in \mathbb{R}^{n_x}$. In state space, this translates to

$$
\begin{bmatrix} \dot{\tilde{\boldsymbol{\nu}}} \\ \dot{\tilde{\boldsymbol{\mu}}} \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{1}{\overline{\tau}_\nu}\mathcal{S}Q^{-1}\mathcal{S}^\top - \frac{1}{\overline{\tau}_\nu}\rho(\mathrm{L} \otimes I_{n_r}) & -\frac{1}{\overline{\tau}_\nu}(E \otimes I_{n_r}) \\ \frac{1}{\overline{\tau}_\mu}(E^\top \otimes I_{n_r}) & \mathbb{0} \end{bmatrix}}_{A_{\mathrm{ADD}}}
$$

$$
\times \begin{bmatrix} \tilde{\boldsymbol{\nu}} \\ \tilde{\boldsymbol{\mu}} \end{bmatrix}
$$

$$
- \underbrace{\begin{bmatrix} \frac{t_b}{\overline{\tau}_\nu}\mathcal{W}_b \\ \mathbb{0} \end{bmatrix}}_{B_{\mathrm{ADD}}}\eta_b, \quad z = \underbrace{\begin{bmatrix} -Q^{-\frac{1}{2}}\mathcal{S}^\top & \mathbb{0} \end{bmatrix}}_{C_{\mathrm{ADD}}}\begin{bmatrix} \tilde{\boldsymbol{\nu}} \\ \tilde{\boldsymbol{\mu}} \end{bmatrix}. \tag{36}
$$

As in our previous results, it can be verified that A_{ADD} is Hurwitz, and that $(A_{\mathrm{ADD}}, C_{\mathrm{ADD}})$ is observable. Consider the observability Gramian candidate $X_{\mathrm{ADD}} = \frac{1}{2}\mathrm{blkdiag}(\overline{\tau}_\nu I_{(n_x n_r)}, \overline{\tau}_\mu I_{(|\mathcal{E}|n_r)})$. Straightforward algebra shows that

$$X_{\mathrm{ADD}} A_{\mathrm{ADD}} + A_{\mathrm{ADD}}^\top X_{\mathrm{ADD}} + Q_{\mathrm{ADD}} = \begin{bmatrix} -\rho(\mathrm{L} \otimes I_{n_r}) & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}$$

$$\preceq \mathbb{0} \tag{37}$$

where $Q_{\mathrm{ADD}} = C_{\mathrm{ADD}}^\top C_{\mathrm{ADD}}$. We conclude that X_{ADD} is a generalized observability Gramian for the system G_{ADD}. Furthermore, if X_t is the true observability Gramian for G_{ADD}, then as in Theorem 3.2, we have $X_t \preceq X_{\mathrm{ADD}}$ and therefore

$$\|G_{\mathrm{ADD}}\|_{\mathcal{H}_2}^2 \le \mathrm{Tr}(B_{\mathrm{ADD}}^\top X_{\mathrm{ADD}} B_{\mathrm{ADD}}) = \frac{t_b^2}{2\overline{\tau}_\nu}\mathrm{Tr}(\mathcal{W}_b^\top \mathcal{W}_b). \tag{38}$$

When $\rho = 0$, X_{ADD} is the exact observability Gramian and

$$\|G_{\mathrm{ADD}}\|_{\mathcal{H}_2}^2 = \frac{t_b^2}{2\overline{\tau}_\nu}\mathrm{Tr}(\mathcal{W}_b^\top \mathcal{W}_b). \tag{39}$$

To complete the proof, it remains to show that (39) in fact implies that $\rho = 0$. Suppose that (39) holds, and define $\Delta := X_{\mathrm{ADD}} - X_t \succeq \mathbb{0}$. Then, following a similar argument as in the proof of Theorem 3.2, one may show that $\Delta B_{\mathrm{ADD}} = \mathbb{0}$. From (37) and the fact that X_t is the actual observability Gramian of the system, we may subtract equations to obtain

$$\Delta A_{\mathrm{ADD}} + A_{\mathrm{ADD}}^\top \Delta = \begin{bmatrix} -\rho(\mathrm{L} \otimes I_{n_r}) & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}.$$

Now, by multiplying each side of the equality above from the left by $B_{\mathrm{ADD}}^\top$ and from the right by B_{ADD}, we find that

$$\mathbb{0} = \rho\mathcal{W}_b^\top (\mathrm{L} \otimes I_{n_r})\mathcal{W}_b$$

$$= \rho\mathcal{W}_b^\top (E \otimes I_{n_r})(E^\top \otimes I_{n_r})\mathcal{W}_b = \rho M^\top M$$

where $M = (E^\top \otimes I_{n_r})\mathcal{W}_b$. Since the graph $\mathcal{G}$ is connected and W_b is square and of full rank, it always holds that $M \ne \mathbb{0}$, and therefore, we conclude that $\rho = 0$. ∎

The point of interest from Corollary 4.2 is that the bound on the $\mathcal{H}_2$ performance is *independent* of ρ. In particular then, as ρ becomes large, the norm does not grow without bound,

which resolves the issue observed in the result of Theorem 3.4. When applied to the resource allocation problem in Section V, we will, in fact, be able to make the stronger statement that the norm is a strictly decreasing function of ρ, and augmentation can, therefore, be used successfully to improve saddle-point algorithm performance.

V. APPLICATION TO RESOURCE ALLOCATION PROBLEMS

We now apply the results from the previous subsections to a particular class of problems. As a special case of the problem (1), consider the *resource allocation problem*

$$
\begin{aligned}
\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & \sum_{i=1}^{n} \frac{1}{2} q_i x_i^2 + c_i x_i \\
\text{subject to} \quad & \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} d_i
\end{aligned}
\tag{40}
$$

where $q_i > 0$, $c_i \in \mathbb{R}$, and $d_i \in \mathbb{R}$. Comparing (40) to (1), we have $Q = \operatorname{diag}(q_1, \ldots, q_n)$, $S = \mathbb{1}_n^\top$, $W_b = \mathbb{1}_n^\top$, and $d := \operatorname{col}(d_1, \ldots, d_n) = b$. The interpretation of (40) is that a resource must be obtained from one of n suppliers in amount x_i, subject to a total demand satisfaction constraint. The objective function of (40) can be interpreted as the sum of the utilities $-c_i x_i$ minus the sum of the costs $q_i x_i^2 / 2$. In a multiagent context, each variable x_i is assigned to an agent, the parameters q_i, c_i, and d_i are available locally to each agent, and the agents must collectively solve the problem (40) through local exchange of information. As a concrete example, in the context of power system frequency control, the objective function models the cost of producing an auxiliary power input x_i; see our preliminary work [26] for additional details.

We will consider (40) along with several equivalent reformulations and apply our results from the previous sections to assess the input–output performance of the resulting saddle-point algorithms. External disturbances η_d will be integrated into the algorithms as $d_i \mapsto d_i + \eta_i$, where η_i models the disturbances in demand d_i. For simplicity, all time-constant matrices are assumed to be multiples of the identity matrix. To most clearly indicate which algorithms require communication of which variables, in this section, algorithms are *not* written in deviation coordinates with respect to the optimizer. In all cases, the performance output z is chosen such that $\|z(t)\|_2^2 = (x(t) - x^\star)^\top Q(x(t) - x^\star)$, where $x^\star$ is the global primal optimizer. With this choice of performance output, the transient performance of the algorithm is measured using the same relative weightings as the steady-state performance.

To begin, the augmented Lagrangian of (40) is given in vector notation by

$$
L(x, \nu) = \frac{1}{2} x^\top Q x + c^\top x + \nu \mathbb{1}_n^\top (x - d) + \frac{\rho}{2} \|\mathbb{1}_n^\top (x - d)\|_2^2
\tag{41}
$$

where $\nu \in \mathbb{R}$. Applying the saddle-point method to the Lagrangian $L(x, \nu)$ and attaching the same disturbances $\eta \in \mathbb{R}^n$ and performance outputs $z \in \mathbb{R}^n$ as before, we obtain the

centralized saddle-point dynamics

$$
\mathrm{RA}_{\mathrm{cent}}(\rho) : \begin{cases} \bar{\tau}_x \dot{x} = -Qx - c - \nu \mathbb{1}_n - \rho \mathbb{1}_n \mathbb{1}_n^\top (x - d - \eta) \\ \bar{\tau}_\nu \dot{\nu} = \mathbb{1}_n^\top (x - d - \eta) \\ z = Q^{\frac{1}{2}} (x - x^\star). \end{cases}
\tag{42}
$$

When $\rho = 0$, the algorithm (42) is of a gather-and-broadcast type [37], where all states x_i and disturbances d_i are collected and processed by a central agent with state ν. When $\rho > 0$, the additional term $\mathbb{1}_n \mathbb{1}_n^\top$ in the algorithm requires all-to-all communication of the local imbalances $x_i - d_i$.

We now consider a reformulation that results in a distributed optimization algorithm. Let $\mathcal{G} = (\{1, \ldots, n\}, \mathcal{E})$ denote a weakly connected acyclic graph over the agent set $\{1, \ldots, n\}$, and let $E \in \mathbb{R}^{n \times |\mathcal{E}|}$ denote the oriented node-edge incidence matrix of $\mathcal{G}$. The constraint $\mathbb{1}_n^\top x = \mathbb{1}_n^\top d$ in the resource allocation problem (40) is equivalent to the existence[6] of a vector $\delta \in \mathbb{R}^{|\mathcal{E}|}$ such that $E\delta = x - d$. The resource allocation problem (40) can, therefore, be equivalently rewritten as

$$
\begin{aligned}
\underset{x \in \mathbb{R}^n, \delta \in \mathbb{R}^{|\mathcal{E}|}}{\text{minimize}} \quad & \sum_{i=1}^{n} \frac{1}{2} q_i x_i^2 + c_i x_i \\
\text{subject to} \quad & E\delta = x - d
\end{aligned}
\tag{43}
$$

with associated augmented Lagrangian

$$
\begin{aligned}
L'(x, \delta, \nu) = \ & \frac{1}{2} x^\top Q x + c^\top x + \nu^\top (E\delta - x + d) \\
& + \frac{\rho}{2} \|E\delta - x + d\|_2^2
\end{aligned}
$$

where $\nu \in \mathbb{R}^n$. This reformulation can be interpreted as a version of (1) with an expanded primal variable (x, δ) and an expanded dual variable $\nu = \operatorname{col}(\nu_1, \ldots, \nu_n)$. By applying the saddle-point method to the problem (43), we obtain the *distributed saddle-point* dynamics

$$
\mathrm{RA}_{\mathrm{dist}}(\rho) : \begin{cases} \bar{\tau}_x \dot{x} = -Qx - c + \rho(E\delta - x + d + \eta) + \nu \\ \bar{\tau}_\delta \dot{\delta} = -E^\top \nu - \rho E^\top (E\delta - x + d + \eta) \\ \bar{\tau}_\nu \dot{\nu} = E\delta - x + d + \eta \\ z = Q^{\frac{1}{2}} (x - x^\star). \end{cases}
\tag{44}
$$

When $\rho = 0$, the algorithm (44) is distributed with the topology of the graph $\mathcal{G}$, with states (x_i, ν_i) associated with each node and a state δ_i associated with each edge. When $\rho > 0$, the algorithm contains the so-called edge Laplacian matrix $E^\top E$ [38], which under our acyclic assumption is positive definite.

Our third formulation is the dual ascent algorithm (30). Substitution of the appropriate matrices into (30) leads to the *centralized dual ascent* dynamics

$$
\mathrm{RA}_{\mathrm{cent}}^{\mathrm{dual}} : \begin{cases} \bar{\tau}_\nu \dot{\nu} = -(\mathbb{1}_n^\top Q^{-1} \mathbb{1}_n)\nu - \mathbb{1}_n^\top (Q^{-1} c + d + \eta) \\ z = Q^{\frac{1}{2}} (x - x^\star) = -Q^{-\frac{1}{2}} (c + \nu \mathbb{1}_n) - Q^{\frac{1}{2}} x^\star \end{cases}
\tag{45}
$$

---

[6]This follows since $\ker(E^\top) = \operatorname{span}(\mathbb{1}_n)$, and hence, $\operatorname{Im}(E)$ is the subspace orthogonal to the vector $\mathbb{1}_n$.

TABLE I
COMPARISON OF SQUARED $\mathcal{H}_2$ NORM EXPRESSIONS

System	$\rho = 0$	$\rho \to \infty$
$\text{RA}_{\text{cent}}(\rho)$	$n/(2\overline{\tau}_\nu)$	$+\infty$
$\text{RA}_{\text{dist}}(\rho)$	$n/(2\overline{\tau}_\nu)$	$+\infty$
$\text{RA}_{\text{cent}}^{\text{dual}}$	$n/(2\overline{\tau}_\nu)$	independent of ρ
$\text{RA}_{\text{dist}}^{\text{dual}}(\rho)$	$n/(2\overline{\tau}_\nu)$	$1/(2\overline{\tau}_\nu)$

where $\nu \in \mathbb{R}$. Algorithm (45) is again centralized, with a single central agent with state ν performing all computations and broadcasting $x_i = -q_i^{-1}(c_i + \nu)$ back to each agent. For our fourth and final formulation, we apply the ADD-SP method developed in Section IV-B. For the problem (40), one quickly deduces that $\mathcal{S} = \mathcal{W}_b = I_n$, and the algorithm (33) reduces to

$$\text{RA}_{\text{dist}}^{\text{dual}}(\rho): \begin{cases} \overline{\tau}_\nu \dot{\nu} = -Q^{-1}\nu - (d + \eta) - Q^{-1}c - E\mu - \rho\text{L}\nu \\ \overline{\tau}_\mu \dot{\mu} = E^\top \nu \\ z = Q^{\frac{1}{2}}(x - x^\star) = -Q^{-\frac{1}{2}}(\nu + c) - Q^{\frac{1}{2}}x^\star \end{cases}$$
$$(46)$$

When $\rho = 0$, this algorithm is distributed with the graph $\mathcal{G}$ associated with the incidence matrix E, with states ν_i associated with nodes and states μ_{ij} associated with edges. When $\rho > 0$, the algorithm additionally contains the undirected Laplacian matrix $\text{L} = EE^\top$ of $\mathcal{G}$ and thus remains distributed.

For each of the four formulations above, we compute the $\mathcal{H}_2$ norm from the disturbance input η to the performance output z. For $\text{RA}_{\text{cent}}(\rho)$, $\text{RA}_{\text{cent}}^{\text{dual}}$, and $\text{RA}_{\text{dist}}^{\text{dual}}(\rho)$, this follows immediately from Theorem 3.4, Proposition 4.1, and Corollary 4.2, respectively. The algorithm $\text{RA}_{\text{dist}}(\rho)$ requires a modification of the proof of Theorem 3.4, since the objective function is no longer strongly convex in the primal variables (x, δ); we omit the details.

The first column of Table I shows the $\mathcal{H}_2$ system norms for the four formulations when $\rho = 0$, i.e., the unaugmented versions of the various saddle-point algorithms. Despite substantial differences between the algorithms in terms of information structure and number of states, all four have the same input–output performance in the $\mathcal{H}_2$ norm. This implies that a distributed implementation will perform no worse than a centralized implementation.

While these four formulations all possess identical system norms under the basic primal-dual algorithm, augmentation differentiates these methods from one another, and substantial differences between the algorithms begin to appear as ρ is increased. The limiting results are tabulated in the second column of Table I. The input–output performance of the first two formulations becomes arbitrarily bad as the augmentation gain ρ increases, while the performance of the ADD-SP algorithm improves substantially, becoming independent of the system size in the limit $\rho \to \infty$.

We illustrate the results in Table I via time-domain simulations in Figs. 2 and 3 for the system in (43) with $n = 2$ and an underlying line graph with $E = [1 \ -1]^\top$. With unit variance white noise as inputs, the unaugmented implementation in Fig. 2 for the four different algorithms results in identical steady-state output variance, numerically computed as the squared $\mathcal{H}_2$ norm in (7).

In Fig. 3, a sufficiently large augmentation factor ρ is introduced to penalize the constraint violations. It is observed that with the augmentation, the steady-state variance of the outputs for the centralized and distributed implementations in $\text{RA}_{\text{cent}}(\rho)$, $\text{RA}_{\text{dist}}(\rho)$ worsens, while that of the distributed dual implementation from $\text{RA}_{\text{dist}}^{\text{dual}}(\rho)$ improves.

Fig. 4(a) illustrates how the choice of communication graph topology influences the performance of the algorithm $\text{RA}_{\text{dist}}^{\text{dual}}(\rho)$. We consider $n = 4$ agents and implement the algorithm with line, ring, and complete communication graphs. While Corollary 4.2 and the results of Table I hold only for acyclic graphs, Fig. 4(a) shows that in all cases, the algorithm's performance improves as ρ increases. For a given value of ρ, graphs with higher connectivity show a greater improvement. This behavior is explained by noting that the algorithm (46) has the same form as the augmented saddle-point dynamics (26), and an analysis similar to that performed for Theorem 3.4 can, in fact, be performed for (46). For uniform cost function parameters and an acyclic graph, this leads to the expression

$$\|G_{\text{RA}_{\text{dist}}^{\text{dual}}(\rho)}\|_{\mathcal{H}_2}^2 = \frac{1}{2\overline{\tau}_\nu}\left(1 + \sum_{i=2}^{n}\frac{q}{q + \rho\lambda_i}\right) \quad (47)$$

where $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of the Laplacian matrix. As graph connectivity increases, so does λ_2, and therefore, performance improves. From a design perspective, note that for a fixed time constant $\overline{\tau}_\nu$ and a desired level of performance $\gamma \in (1/\sqrt{2\overline{\tau}_\nu}, \sqrt{n}/\sqrt{2\overline{\tau}_\nu})$, examination of (47) shows that a sufficient condition for $\|G_{\text{RA}_{\text{dist}}^{\text{dual}}(\rho)}\|_{\mathcal{H}_2} \leq \gamma$ is that

$$\rho \geq \frac{q}{\lambda_2}\frac{n - 2\overline{\tau}_\nu\gamma^2}{2\overline{\tau}_\nu\gamma^2 - 1}.$$

In particular, this shows that the augmentation gain should be chosen in inverse proportion to the algebraic connectivity λ_2 of the Laplacian matrix. A strongly connected graph will, therefore, require a lower augmentation gain than a weakly connected graph to achieve a desired level of $\mathcal{H}_2$ performance. Achieving an $\mathcal{H}_2$ norm lower than $1/\sqrt{2\overline{\tau}_\nu}$ requires an increase in $\overline{\tau}_\nu$.

Finally, Fig. 4(b) plots the system norm as a function of ρ for the three augmented algorithms, for a test case with $n = 4$ agents. The norm is not a monotonic function of the augmentation factor ρ for the implementations in $\text{RA}_{\text{cent}}(\rho)$ and $\text{RA}_{\text{dist}}(\rho)$, but is monotonic for $\text{RA}_{\text{dist}}^{\text{dual}}(\rho)$ applied to resource allocation problems, in agreement with the result for the parametrically uniform case in (47).
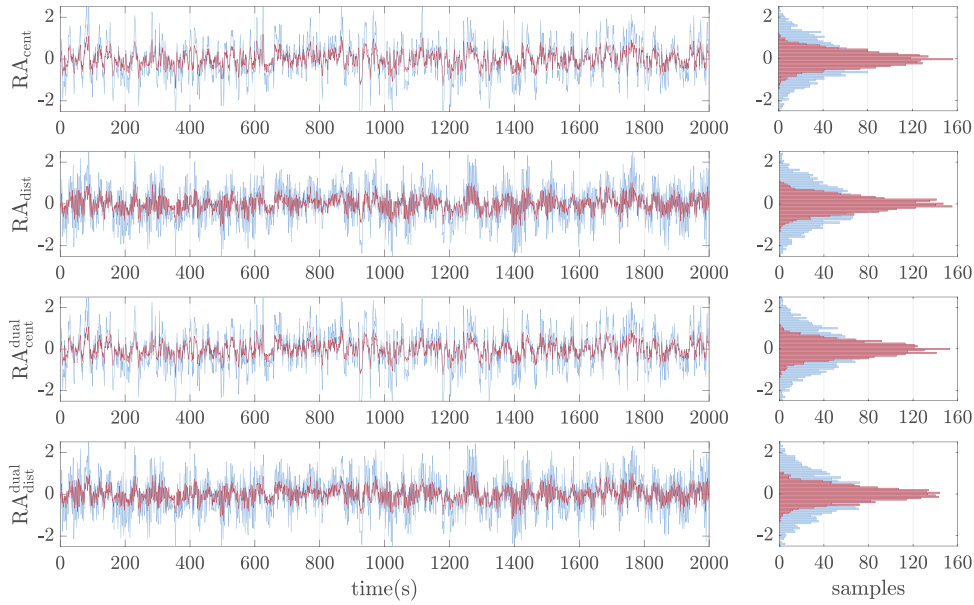
Fig. 2. Steady-state variance for the four different implementations for the unaugmented case, for parameters $n = 2$, $Q = \mathrm{diag}(4, 25)$, $\tau_x = \tau_\delta = \tau_\nu = \tau_\mu = 1$, $E = [1 \;\; -1]^\top$; the remaining parameters do not influence the results.
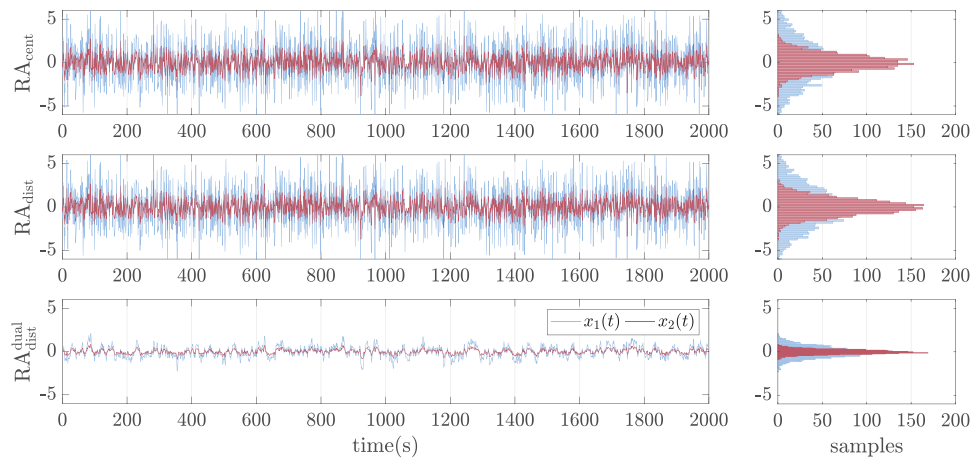

Fig. 3. Steady-state variance for the three different implementations for the augmented case, for parameters $n = 2$, $\rho = 100$, $Q = \mathrm{diag}(4, 25)$, $\tau_x = \tau_\delta = \tau_\nu = \tau_\mu = 1$, and $E = [1 - 1]^\top$; the remaining parameters do not influence the results.
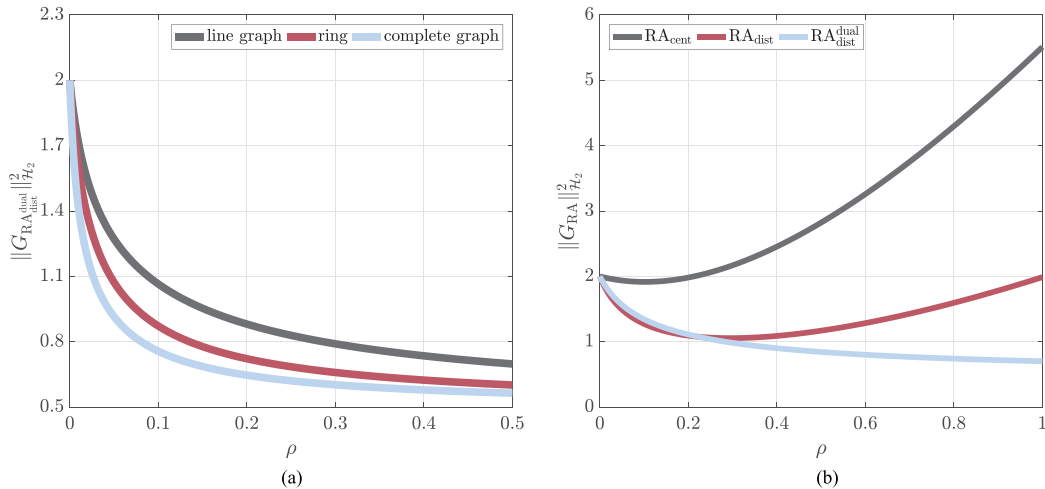

Fig. 4. Squared $\mathcal{H}_2$ norm as a function of ρ, for parameters $n = 4$, $\tau_x = \tau_\delta = \tau_\nu = \tau_\mu = 1$, and unweighted graphs. (a) $\mathrm{RA}_{\mathrm{dist}}^{\mathrm{dual}}(\rho)$ for $Q = \mathrm{diag}(4, 25, 16, 49)$. (b) $\mathrm{RA}_{\mathrm{cent}}$, $\mathrm{RA}_{\mathrm{dist}}$, $\mathrm{RA}_{\mathrm{dist}}^{\mathrm{dual}}$ for $Q = \mathrm{diag}(4, 4, 4, 9)$ and line graph.

VI. Conclusion

We have studied the input–output performance of continuous-time saddle-point methods for solving linearly constrained convex quadratic programs, providing an explicit formula for the $\mathcal{H}_2$ norm under a relevant input–output configuration. We then studied the effects of Lagrangian regularization and augmentation on this norm and derived a distributed dual version of the augmented algorithm, which overcomes some of the limitations of naive augmentation. We then applied the results to compare several implementations of the saddle-point method to resource allocation problems.

Open directions for future research include input–output performance metrics for problems involving inequality constraints, for distributed implementations where communication delays occur between agents, and for other classes of distributed optimization algorithms. An analogous study in an $\mathcal{H}_\infty$ performance framework has not been completed. Another interesting question is how to further improve the $\mathcal{H}_2$ performance of saddle-point methods by designing auxiliary feedback controllers; augmentation is but one approach. Finally, extending these results beyond the case of quadratic cost functions with linear constraints could be pursued via nonlinear and robust control approaches, as in [21]–[25]; see [39] for recent work in this direction, in the context of robust $\mathcal{H}_2$ analysis of gradient methods.

References

[1] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica*, vol. 24, no. 1, pp. 59–70, Jan. 1956.

[2] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, CA, USA: Stanford Univ. Press, 2006.

[3] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, Dec. 2010.

[4] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proc. IEEE Conf. Decis. Control/Eur. Control Conf.*, Orlando, FL, USA, Dec. 2011, pp. 3800–3805.

[5] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, Jan. 2016.

[6] K. C. Kosaraju, S. Mohan, and R. Pasumarthy, "On the primal-dual dynamics of support vector machines," 2018, *arXiv:1805.00699*.

[7] B. Gharesifard and J. Cortes, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.

[8] S. H. Low and D. E. Lapsey, "Optimization flow control. I: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–74, Dec. 1999.

[9] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Breaking the hierarchy: Distributed control & economic optimality in microgrids," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 241–253, Sep. 2016.

[10] N. Li, L. Chen, C. Zhao, and S. H. Low, "Connecting automatic generation control and economic dispatch from an optimization view," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 735–740.

[11] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, Dec. 2016.

[12] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Regulation of dynamical systems to optimal solutions of semidefinite programs: Algorithms and applications to ac optimal power flow," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, Jul. 2015, pp. 2087–2092.

[13] T. Stegink, C. D. Persis, and A. van der Schaft, "A unifying energy-based approach to stability of power grids with market dynamics," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2612–2622, Jun. 2017.

[14] E. Mallada, C. Zhao, and S. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6294–6309, Dec. 2017.

[15] J. T. Wen and M. Arcak, "A unifying passivity framework for network flow control," *IEEE Trans. Autom. Control*, vol. 49, no. 2, pp. 162–174, Feb. 2004.

[16] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Oct. 2010, pp. 557–561.

[17] G. Droge, H. Kawashima, and M. B. Egerstedt, "Continuous-time proportional-integral distributed optimisation for networked systems," *J. Control Decis.*, vol. 1, no. 3, pp. 191–213, Aug. 2014.

[18] G. Droge and M. Egerstedt, "Proportional integral distributed optimization for dynamic network topologies," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 3621–3626.

[19] T. Hatanaka, N. Chopra, T. Ishizaki, and N. Li, "Passivity-based distributed optimization with communication delays using pi consensus algorithm," 2017. [Online]. Available: https://arxiv.org/abs/1609.04666

[20] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J. Slotine, "Contraction and robustness of continuous time primal-dual dynamics," *IEEE Control Syst. Lett.*, vol. 2, no. 4, pp. 755–760, Oct. 2018.

[21] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of strong convexity-concavity in the convergence and robustness of the saddle-point dynamics," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2016, pp. 504–510.

[22] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of convexity on saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2449–2464, Aug. 2018.

[23] J. W. Simpson-Porco, "Input/output analysis of primal-dual gradient algorithms," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2016, pp. 219–224.

[24] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, Jan. 2016.

[25] B. Hu and L. Lessard, "Control interpretations for first-order optimization methods," in *Proc. Amer. Control Conf.*, Seattle, WA, USA, May 2017, pp. 3114–3119.

[26] J. W. Simpson-Porco, B. K. Poolla, N. Monshizadeh, and F. Dörfler, "Quadratic performance of primal-dual methods with application to secondary frequency control of power systems," in *Proc. IEEE Conf. Decis. Control*, Las Vegas, NV, USA, Dec. 2016, pp. 1840–1845.

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

[28] M. Benzi, G. H. Golub, and J. Liesen, "Numerical solution of saddle point problems," *Acta Numer.*, vol. 14, pp. 1–137, May 2005.

[29] A. Nedic and A. Ozdaglar, "Subgradient methods for saddle-point problems," *J. Optim. Theory Appl*, vol. 142, no. 1, pp. 205–228, Jul. 2009.

[30] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[31] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory* (ser. Texts in Applied Mathematics), vol. 36. New York, NY, USA: Springer, 2000.

[32] M. B. Khuzani and N. Li, "Distributed regularized primal-dual method: Convergence analysis and trade-offs," 2016, *arXiv:1609.08262*.

[33] A. Simonetto and G. Leus, "Double smoothing for time-varying distributed multiuser optimization," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 852–856.

[34] E. Dall'Anese and A. Simonetto, "Optimal power flow pursuit," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 942–952, Mar. 2018.

[35] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 2008.

[36] F. Bullo, *Lectures on Network Systems*, version 0.85, with contributions by J. Cortés, F. Dörfler, and S. Martínez, May 2016. [Online]. Available: http://motion.me.ucsb.edu/book-lns

[37] F. Dörfler and S. Grammatico, "Gather-and-broadcast frequency control in power systems," *Automatica*, vol. 79, pp. 296–305, May 2017.

[38] D. Zelazo and M. Mesbahi, "Edge agreement: Graph-theoretic performance bounds and passivity analysis," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 544–555, Mar. 2011.

[39] N. S. Aybat, A. Fallah, M. Gürbüzbalaban, and A. Ozdaglar, "Robust accelerated gradient methods for smooth strongly convex functions," 2018. [Online]. Available: https://arxiv.org/abs/1805.10579

John W. Simpson-Porco (S'11–M'16) received the B.Sc. degree in engineering physics from Queen's University, Kingston, ON, Canada, in 2010, and the Ph.D. degree in mechanical engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2015.

He is currently an Assistant Professor of Electrical and Computer Engineering with the University of Waterloo, Waterloo, ON. He was previously a Visiting Scientist with the Automatic Control Lab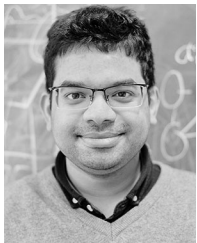oratory, ETH Zürich, Zürich, Switzerland. His research interests include feedback control theory and applications of control in modernized power grids.

Prof. Simpson-Porco is a recipient of the 2012–2014 IFAC Automatica Prize and the Center for Control, Dynamical Systems and Computation Best Thesis Award and Outstanding Scholar Fellowship.

Nima Monshizadeh received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2007, the M.Sc. degree in control engineering from the K.N. Toosi University of Technology, Tehran, in 2009, and the Ph.D. (Hons.) degree (*cum laude*) from the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands, in December 2013.

He then became a Postdoctoral Researcher with the Engineering and Technology Institute, University of Groningen. He was a Research Associate with the Control group of the University of Cambridge, before taking a tenure track position with the University of Groningen in January 2018. His research interests include power networks, model reduction, optimization, and control of complex networks.

Florian Dörfler (S'09–M'13) received the Diploma degree in engineering cybernetics from the University of Stuttgart, Stuttgart, Germany, in 2008, and the Ph.D. degree in mechanical engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2013.

He is currently an Assistant Professor with the Automatic Control Laboratory, ETH Zürich, Zürich, Switzerland. From 2013 to 2014, he was an Assistant Professor with the University of California, Los Angeles. His primary research interests include distributed control, complex networks, and cyber-physical systems currently with applications in energy systems and smart grids.

Dr. Dörfler received the 2010 ACC Student Best Paper Award, the 2011 O. Hugo Schuck Best Paper Award, and the 2012–2014 *Automatica* Best Paper Award, and the 2016 IEEE Circuits and Systems Guillemin–Cauer Best Paper Award. He is a recipient of the 2009 Regents Special International Fellowship, the 2011 Peter J. Frenkel Foundation Fellowship, and the 2015 UCSB ME Best Ph.D. Award. His students were winners or finalists for Best Student Paper Awards at the European Control Conference (2013), the American Control Conference (2016), and the PES PowerTech Conference (2017).

Bala Kameshwar Poolla (S'15) received the B.Tech. degree in electrical engineering and the M.Tech. degree in control systems engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2013. He is currently working toward the Ph.D. degree with the Automatic Control Laboratory, ETH Zürich, Zürich, Switzerland.

His research interests include applications of control theory to power grids, renewable energy systems, and energy markets.

Mr. Poolla was a finalist for the ACC 2016 Best Student Paper Award.