

Quadratic Performance of Primal-Dual Methods with Application to Secondary Frequency Control of Power Systems

John W. Simpson-Porco, Bala Kameshwar Poolla, Nima Monshizadeh, and Florian Dörfler

Abstract—Primal-dual gradient methods have recently attracted interest as a set of systematic techniques for distributed and online optimization. One of the proposed applications has been optimal frequency regulation in power systems, where the primal-dual algorithm is implemented online as a dynamic controller. In this context however, the presence of external disturbances makes quantifying input/output performance important. Here we use the \mathcal{H}_2 system norm to quantify how effectively these distributed algorithms reject external disturbances. For the linear primal-dual algorithms arising from quadratic programs, we provide an explicit expression for the \mathcal{H}_2 norm, and examine the performance gain achieved by augmenting the Lagrangian. Our results suggest that the primal-dual method may perform poorly when applied to large-scale systems, and that Lagrangian augmentation can partially (or completely) alleviate these scaling issues. We illustrate our results with an application to power system frequency control by means of distributed primal-dual controllers.

I. INTRODUCTION

Primal-dual methods are a class of gradient-based algorithms for solving constrained convex optimization problems. Introduced in the early 1950's [1], [2], the methods are also frequently termed saddle-point algorithms, as they 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 (KKT) conditions, and the algorithm's internal state asymptotically converges to the global primal-dual optimizer of the optimization problem; see [3], [4] for technical convergence results. Recently, these algorithms have attracted renewed attention in the control community for solving *distributed* optimization problems, where agents cooperate through a communication network to solve an optimization problem without centralized coordination. Standard applications of distributed optimization include utility maximization [3] and congestion management in communication networks [5]. While most optimization algorithms require centralized information to compute the optimizer, primal-dual algorithms often yield distributed strategies where agents require only local information along with inter-agent communication.

Rather than solve the optimization problem only once offline, it is also desirable to run the primal-dual algorithm

online as a “controller”, so that the optimizer can be (hopefully) tracked in real-time as the problem data changes. However, like all controllers, when implemented online primal-dual methods become subject to unknown disturbances. We therefore arrive at the question of dynamic input/output performance: how well does the primal-dual method track the optimal point when subjected to disturbances?

A. Power Networks and Primal-Dual Frequency Controllers

Our particular motivation for studying primal-dual methods comes from their recent application to optimal frequency regulation in power networks. Power networks are designed to operate around a nominal frequency (e.g., 50 Hz or 60 Hz), and any steady-state deviation from this nominal value signals a global imbalance of power supply and demand. So-called primary control is a proportional control layer implemented at sources [6, Chapter 11] or loads [7] which attempts to balance supply and demand over fast time-scales, stabilizing the grid to an off-nominal frequency. Higher-level centralized control layers termed secondary and tertiary control are then tasked with regulating the grid frequency to its nominal value, meeting operational constraints, and optimizing the grid by minimizing the cost of generation. We refer to this high-level control/optimization problem as the *optimal frequency regulation* (OFR) problem.

Currently however, the rise of distributed generation is causing us to rethink how we should solve the OFR problem in future power grids. Leveraging the many new controllable power electronic devices within the grid will enable frequency control to be distributed across both producers and consumers. This decentralization could increase system resilience and allow for more localized control actions to be taken. Due to sensing and communication constraints, these devices should act based on minimal information, and ideally without detailed model information or precise knowledge of system parameters and generation/load forecasts.

Returning to primal-dual methods, in [8], [9] it was shown that the standard primary/secondary control dynamics of a power network can themselves be interpreted as a primal-dual algorithm for solving an optimal frequency regulation problem. This “reverse engineering” observation was then translated to a “forward engineering” approach to further improve the economic efficiency of the network by tweaking the desired optimization problem and reapplying the primal-dual method. The final resulting dynamics can then be interpreted as a concatenation of power system dynamics along with a real-time distributed control layer for the power grid, ostensibly replacing the secondary/tertiary centralized

J. W. Simpson-Porco is with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. Email: jwsimpson@uwaterloo.ca. B. K. Poolla and F. Dörfler are with the Automatic Control Laboratory, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland. Email: {bpoolla, dorfler}@ethz.ch. N. Monshizadeh is with the Engineering and Technology Institute, University of Groningen, The Netherlands. Email: n.monshizadeh@rug.nl.

Work supported by Waterloo, ETH, Groningen, and SNF #160573.

control layers. Extensions of this framework to load-side OFR and mixed generator-side/load-side OFR were presented in [10]–[12]; see also [13], [14] for even more recent work, and [15] for an elegant port-Hamiltonian perspective. In this power systems context, disturbances can enter the primal-dual algorithm as noisy power injections, arising from fluctuating generation and load, from measurement noise, or from generic process uncertainty. It is presently unknown how well the controller can tolerate these disturbances while it attempts to maintain the nominally optimal operating point.

B. Contributions

As background, in Section II we review primal-dual algorithms for the relevant class of optimization problems, then recall the basic facts about the \mathcal{H}_2 norm as a measure of input/output system performance.

In Section III we consider the effect of disturbances on the primal-dual dynamics arising from linearly constrained, strictly convex quadratic optimization problems. While this is only a small subclass of the types of optimization problems primal-dual methods are applicable to, it is a particularly relevant class for the OFR problem, and should indicate a “best” case performance since it is free of both nonlinearities and hard inequality constraints. For a relevant input/output configuration, we derive an explicit expression for the primal-dual \mathcal{H}_2 norm (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. We then study the effect of augmenting the Lagrangian of the optimization problem. Under some simplifying assumptions, we derive an expression for the \mathcal{H}_2 norm, precisely quantifying how input/output performance is improved by augmentation. In Section IV we apply our results to the OFR problem.

Notation: For $A \in \mathbb{R}^{n \times n}$, A^\top is its transpose and $\text{Tr}(A) = \sum_{i=1}^n A_{ii}$. For a positive semidefinite matrix $Q \in \mathbb{R}^{n \times n}$, $Q^{\frac{1}{2}}$ is its unique square root. The $n \times n$ identity matrix is I_n , $\mathbb{0}$ is a matrix of zeros of appropriate dimension, while $\mathbb{1}_n$ (resp. $\mathbb{0}_n$) are column n -vectors of all ones and all zeros, respectively. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable, then $\nabla_x f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is its gradient.

II. REVIEW OF PRIMAL-DUAL METHODS AND \mathcal{H}_2 SYSTEM NORM

A. The Primal-Dual Method

Consider the quadratic optimization problem

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} x^\top Q x + c^\top x \\ & \text{subject to} && S x = b, \end{aligned} \quad (1)$$

where $Q = Q^\top > 0$ is positive definite, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^r$ and $S \in \mathbb{R}^{r \times n}$ where $r < n$. We make the standard assumption that S has full row rank ($\text{rank}(S) = r$), which simply means that the constraints $Sx = b$ are not redundant.*

*In some cases, it is desirable to relax this assumption to $\text{rank}(S) = \tilde{r} < r$, and to also allow for $r \geq n$. This introduces some mild technical complications, so we defer the analysis to a future work.

Under these assumptions the problem (1) has a finite optimum, the equality constraints are strictly feasible, and (1) may be equivalently studied through its Lagrange dual with zero duality gap [16]. The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}$ of the problem (1) is given by

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

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

$$\begin{aligned} \nabla_x L(x, \nu) = \mathbb{0}_n & \iff \mathbb{0}_n = Qx + S^\top \nu + c, \\ \nabla_\nu L(x, \nu) = \mathbb{0}_r & \iff \mathbb{0}_r = Sx - b, \end{aligned} \quad (3)$$

are necessary and sufficient for optimality. From these linear equations one may easily compute that the unique global primal-dual optimizer (x^*, ν^*) is

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

While (4) is the exact solution to the optimization problem (1), its evaluation requires centralized information. In many multi-agent system applications, the cost matrix Q is diagonal or block-diagonal and $x^\top Q x$ is therefore a sum of local costs, with Q_{ii} known only to the i th agent. Moreover, in certain applications the vectors b and c may change over time, and the constraints encoded in S may be quite sparse, mirroring the topology of an interaction or communication network between agents. It is therefore desirable to solve the optimization problem (1) in an online distributed fashion, where agents in the network communicate and cooperate to compute and track the global optimizer (4).

A simple continuous-time algorithm to seek this optimal point is the *primal-dual method* [3], [4], [17], [18]

$$\tau_x \dot{x} = -\nabla_x L(x, \nu), \quad \tau_\nu \dot{\nu} = \nabla_\nu L(x, \nu), \quad (5)$$

which here reduces to the linear dynamical system

$$\tau_x \dot{x} = -Qx - S^\top \nu - c, \quad (6a)$$

$$\tau_\nu \dot{\nu} = Sx - b, \quad (6b)$$

where τ_x, τ_ν are positive definite diagonal matrices of time constants. By construction, the equilibrium points of (6) are in one-to-one correspondence with the solutions of the KKT conditions (3). The following stability result shows global convergence, and can be proved using the *strict* Lyapunov function: $V(x, \nu) = (x - x^*)^\top \tau_x (x - x^*) + (\nu - \nu^*)^\top \tau_\nu (\nu - \nu^*) + \varepsilon (\nu - \nu^*)^\top S \tau_x (x - x^*)$ for $\varepsilon > 0$ sufficiently small.

Lemma 2.1 (Global Convergence to Optimizer): The unique equilibrium point (x^*, ν^*) given in (4) of the primal-dual dynamics (6) is globally exponentially stable.

In the remainder of the analysis we assume that we have changed variables to the error coordinates $\Delta x = x - x^*$, $\Delta \nu = \nu - \nu^*$, and with an abuse of notation we drop the Δ 's and simply refer to the error coordinates as x and ν .

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

Consider the MIMO continuous-time LTI system

$$\begin{aligned} \dot{x} &= Ax + B\eta \\ y &= Cx, \end{aligned} \quad (7)$$

with input η and output y , and where A is Hurwitz. We denote the transfer matrix from η to y by G . If (7) is input/output stable, its \mathcal{H}_2 norm $\|G\|_{\mathcal{H}_2}$ is defined as the induced norm from input signals $\eta(t) \in L_2$ to output signals $y(t) \in L_\infty$. That is, $\|G\|_{\mathcal{H}_2}$ is the worst-case L_∞ output amplification from square-integrable inputs. Other insightful interpretations exist, the most useful for our discussion being that $\|G\|_{\mathcal{H}_2}^2$ is the steady-state variance of the output

$$\|G\|_{\mathcal{H}_2}^2 = \lim_{t \rightarrow \infty} \mathbb{E}[y^\top(t)y(t)],$$

when each component of $\eta(t)$ is stochastic white noise with unit covariance (i.e., $\mathbb{E}[\eta(t)\eta^\top(\tau)] = \delta(t - \tau)I$). Thus, $\|G\|_{\mathcal{H}_2}$ measures how much the output varies in steady-state under stochastic disturbances. A convenient formula for the \mathcal{H}_2 norm is [19, Chapter 6]

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

where $X = X^\top > 0$ is the observability Gramian satisfying

$$XA + A^\top X + C^\top C = 0. \quad (9)$$

If the pair (C, A) is observable, then (9) is solvable for the unique, positive-definite observability Gramian. In what follows we will derive results on the \mathcal{H}_2 performance of primal-dual methods under disturbances by explicitly solving (9) for particular cases. Recent applications of the \mathcal{H}_2 norm to power system performance may be found in [20]–[22].

III. \mathcal{H}_2 PERFORMANCE OF PRIMAL-DUAL METHODS

After translating the equilibrium point of the system (6) to the origin, we now equip the primal-dual dynamics (6) with disturbance inputs $\eta \in \mathbb{R}^p$ and performance outputs $y \in \mathbb{R}^m$, leading to the *input-output primal-dual dynamics*

$$\begin{bmatrix} \tau_x \dot{x} \\ \tau_\nu \dot{\nu} \end{bmatrix} = \begin{bmatrix} -Q & -S^\top \\ S & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \eta, \quad (10a)$$

$$y = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix}, \quad (10b)$$

where $B_1 \in \mathbb{R}^{n \times p}$, $B_2 \in \mathbb{R}^{r \times p}$ and $C_1 \in \mathbb{R}^{m \times n}$, $C_2 \in \mathbb{R}^{m \times r}$ are the input and output matrices, respectively. With an abuse of notation, here (x, ν) are now error variables from the unique equilibrium point (4). The inputs on the primal variables (resp. dual variables) may be thought of as a disturbances to the components of the vector c (resp. the vector b), as measurement/actuation noise, or as modeling generic process uncertainty.

As the system (10) is written in error coordinates, convergence to the primal optimizer x^* from (4) is equivalent to convergence of $x(t)$ to the origin. A very natural way to measure this convergence is to use the cost matrix Q from the optimization problem (1) as a weighting matrix, and study the performance output $\|y(t)\|_2^2 = \frac{1}{2}x(t)^\top Q x(t)$,

which is obtained by choosing $m = n$, with $C_1 = \frac{1}{\sqrt{2}}Q^{\frac{1}{2}}$ and $C_2 = 0$ in (10). We arrive at our first result; all proofs are deferred to an extended publication.

Theorem 3.1 (Primal-Dual Performance): Consider the input/output primal-dual dynamics (10) with diagonal cost matrix Q . Let $C_1 = \frac{1}{\sqrt{2}}Q^{\frac{1}{2}}$ and $C_2 = 0$ so that $\|y(t)\|_2^2 = \frac{1}{2}x^\top(t)Qx(t)$. Then the squared \mathcal{H}_2 norm of the system (10) is

$$\|G\|_{\mathcal{H}_2}^2 = \frac{1}{4}\text{Tr}(B_1^\top \tau_x^{-1} B_1) + \frac{1}{4}\text{Tr}(B_2^\top \tau_\nu^{-1} B_2). \quad (11)$$

As a special case, suppose that we have

- (i) *decoupled, uniform-strength disturbances for each primal and dual channel:* $p = n + r$, $B_1 = b_1 [I_n \ 0]$ and $B_2 = b_2 [0 \ I_r]$ for constants $b_1, b_2 > 0$, and
- (ii) *uniform time constants:* $\tau_x = \tau_1 I_n, \tau_\nu = \tau_2 I_r$ for constants $\tau_1, \tau_2 > 0$.

Then

$$\|G\|_{\mathcal{H}_2}^2 = \frac{b_1^2}{4\tau_1} n + \frac{b_2^2}{4\tau_2} r. \quad (12)$$

The most striking feature of the result (11) is that it is completely independent of both the cost matrix Q and the constraint matrix S . In other words, the constraints — be they dense or sparse — are irrelevant to the \mathcal{H}_2 performance for this performance output. The result also scales inversely with the time constants τ_1 and τ_2 , which indicates an inherent trade-off between fast convergence speed (small τ), and robustness against disturbances (large τ). In the the simplified case (12), the squared \mathcal{H}_2 norm scales linearly in the number of disturbances to the primal dynamics (here equal to n) and the number of disturbances to the dual dynamics (here equal to r). This scaling is completely independent of the constraints. While not an egregiously poor scaling, the lack of tunable controller gains other than the time constants means that convergence speed and input/output performance are always conflicting objectives, and that performance will typically degrade as the dimension of the problem grows.

A. Performance of Augmented Primal-Dual Methods

One option for improving the \mathcal{H}_2 performance of primal-dual methods is to return to the Lagrangian function (2) and instead consider the *augmented Lagrangian*

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

where we have incorporated the squared constraint $\|Sx - b\|_2^2 = 0$ into the Lagrangian with a gain $\rho \geq 0$. One way to interpret this is that we have modified the cost function $\frac{1}{2}x^\top Q x + c^\top x$ with an additional term $\frac{\rho}{2} \|Sx - b\|_2^2$ which penalizes *transient* equality constraint violation. It follows that (x, ν) is a saddle point of $L_\rho(x, \nu)$ if and only if it is a saddle point of $L(x, \nu)$, and hence L_ρ and L share the same optimizer. Applying the primal-dual method (5) to $L_\rho(x, \nu)$, we obtain the *augmented primal-dual dynamics*

$$\begin{aligned} \tau_x \dot{x} &= -(Q + \rho S^\top S)x - S^\top \nu - c + \rho S^\top b, \\ \tau_\nu \dot{\nu} &= Sx - b. \end{aligned} \quad (13)$$

One may verify that as before, the unique stable equilibrium point of (13) is given by (4). After translating the equilibrium point to the origin, we now again consider disturbance inputs η and performance outputs $y = \frac{1}{\sqrt{2}}Q^{\frac{1}{2}}x$, leading to

$$\begin{aligned} \begin{bmatrix} \tau_x \dot{x} \\ \tau_\nu \dot{\nu} \end{bmatrix} &= \begin{bmatrix} -(Q + \rho S^T S) & -S^T \\ S & \mathbb{0} \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \eta, \\ y &= \begin{bmatrix} \frac{1}{\sqrt{2}}Q^{\frac{1}{2}} & \mathbb{0} \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix}. \end{aligned} \quad (14)$$

The additional term $-\rho S^T S$ in the dynamics (14) complicates the computation of the \mathcal{H}_2 norm, and we require additional assumptions to obtain an explicit formula; these will be relaxed in a future work.

Theorem 3.2 (Augmented Primal-Dual Performance):

Consider the input/output augmented primal-dual dynamics (14) with uniform parameters $Q = qI_n$, $\tau_x = \tau_1 I_n$, $\tau_\nu = \tau_2 I_r$, $B_1 = b_1 [I_n \ \mathbb{0}]$, and $B_2 = b_2 [\mathbb{0} \ I_r]$ for scalars $q, \tau_1, \tau_2, b_1, b_2 > 0$, along with the performance output $\|y(t)\|_2^2 = \frac{1}{2}x^T(t)Qx(t) = \frac{1}{2}q\|x(t)\|_2^2$. Then the squared \mathcal{H}_2 norm of the system (14) is

$$\|G\|_{\mathcal{H}_2}^2 = \frac{b_1^2}{4\tau_1}(n-r) + \left(\frac{b_1^2}{4\tau_1} + \frac{b_2^2}{4\tau_2}\right) \sum_{i=1}^r \frac{q}{q + \rho\sigma_i^2}, \quad (15)$$

where σ_i is the i th non-zero singular value of S . In particular, in the high augmentation-gain limit $\rho \rightarrow \infty$ it holds that

$$\boxed{\lim_{\rho \rightarrow \infty} \|G\|_{\mathcal{H}_2}^2 = \frac{b_1^2}{4\tau_1}(n-r)}. \quad (16)$$

First, note that (15) generalizes (12) under the assumed restrictions on parameters, since when $\rho = 0$ the expression (15) reduces to (12). Second, note that the result (15) decomposes cleanly into two terms, the first term representing the contribution to the \mathcal{H}_2 norm resulting from the unconstrained subdynamics, while the second term accounts for the constrained subdynamics. The final statement (16) emphasizes that by increasing the augmentation gain ρ , the contribution to the \mathcal{H}_2 norm from the constrained subdynamics can be made *arbitrarily small*, leaving only the contribution from the unconstrained subdynamics. While it is understood that augmented Lagrangians tend to improve convergence of optimization algorithms, the expression (15) gives an analogous input/output performance result. We conclude that augmentation of the Lagrangian may improve certain performance metrics, and may be particularly beneficial when the underlying optimization problem is heavily constrained.

As an observation, we note that even when S^T is a sparse matrix, $S^T S$ typically will not be, and hence the augmented dynamics (13) may not be implementable as a distributed algorithm. A notable exception occurs when S^T is the incidence matrix of a graph, in which case $S^T S$ is a symmetric Laplacian matrix. Incidentally, we will return to this case in the next section.

IV. APPLICATION TO DISTRIBUTED OPTIMAL FREQUENCY REGULATION

We now formulate the power system dynamics under consideration and the optimal frequency regulation optimization problem to be solved, before applying the results derived in Section III to assess the performance of primal-dual frequency controllers.

A. Power Network Model & Optimal Frequency Regulation

We model a power network as a weighted graph $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, \dots, n\}$ is the set of nodes (buses), and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges (branches) with associated edge weights $B_{ij} > 0$ for $\{i, j\} \in \mathcal{E}$. To each bus $i \in \mathcal{V}$ we associate state variables (θ_i, ω_i) corresponding to the voltage phase angle and the frequency deviation from nominal. Under the linear DC Power Flow approximation, the system evolves according to the swing dynamics

$$\begin{aligned} \dot{\theta}_i &= \omega_i, \\ M_i \dot{\omega}_i &= -D_i \omega_i + P_i^* - P_i(\theta) + p_i, \end{aligned} \quad (17)$$

where $M_i > 0$ represents inertia or inverter filter time-constants, $D_i > 0$ models damping and/or droop control, P_i^* is the constant nominal active power injection (nominal generation minus nominal load), $P_i(\theta) = \sum_{j=1}^n B_{ij}(\theta_i - \theta_j)$ is the active power injected at bus i , and p_i is the control input, corresponding to additional power generation from reserves.

When $p = \mathbb{0}_n$, the dynamics (17) converge from every initial condition to a common steady-state frequency $\omega \rightarrow \omega_{ss} \mathbb{1}_n$ which can be easily calculated to be $\omega_{ss} = (\sum_{i=1}^n P_i^*) / (\sum_{i=1}^n D_i)$. When $\omega_{ss} \neq 0$, this represents a static deviation from nominal, which we will eliminate by appropriately selecting the reserve power inputs p . To determine the the steady-state values for p_i , an *optimal frequency regulation problem* (OFRP) can be formulated as[†]

$$\text{minimize}_{p \in \mathbb{R}^n} \quad \sum_{i=1}^n \frac{1}{2} k_i p_i^2 \quad (18a)$$

$$\text{subject to} \quad 0 = \mathbb{1}_n^T (P^* + p), \quad (18b)$$

where we seek to minimize the total cost (18a) of reserve generation $p_i \in \mathbb{R}$, for some coefficients $k_i > 0$. The minimization is subject to network-wide balancing of power injections (18b). One may deduce from (17) that the constraint (18b) also ensures that $\omega = \mathbb{0}_n$ in steady-state, and thus the frequency is returned to its nominal value.

B. \mathcal{H}_2 Performance of Primal-Dual Frequency Controllers

Beginning from the OFRP (18), we roughly follow [7], [11] to derive the controller dynamics. The Lagrangian of the OFRP (18) is given by

$$L(p, \mu) = \frac{1}{2} p^T K p + \mu \mathbb{1}_n^T (P^* + p),$$

[†]A linear term could also of course be added to the cost, but we omit it here for simplicity. We assume that any inequality constraints on p are non-binding, and subsequently drop them from the problem (18), which is then similar to the *classic* economic dispatch [6, Page 405].

where $K = \text{diag}(k_i)$ and $\mu \in \mathbb{R}$ is a multiplier. Computing $\text{argmin}_{p \in \mathbb{R}^n} L(p, \mu)$, one finds that $p = -\mu K^{-1} \mathbf{1}_n$ is the unique minimizer. From this one quickly calculates the dual function $\Phi(\mu) = \inf_{p \in \mathbb{R}^n} L(p, \mu)$, and the dual OFRP is

$$\underset{\mu \in \mathbb{R}}{\text{maximize}} \quad \Phi(\mu) = \sum_{i=1}^n \mu P_i^* - \frac{1}{2k_i} \mu^2, \quad (19)$$

where we seek to maximize $\Phi(\mu)$ over the common variable $\mu \in \mathbb{R}$. To distribute (19), we introduce local variables $\mu_i \in \mathbb{R}$ for each bus, and consider the equivalent constrained optimization problem

$$\begin{aligned} & \underset{\mu \in \mathbb{R}^n}{\text{maximize}} \quad \sum_{i=1}^n \mu_i P_i^* - \frac{1}{2k_i} \mu_i^2 \\ & \text{subject to} \quad 0 = \mu_i - \mu_j, \quad \{i, j\} \in \mathcal{E}_c, \end{aligned} \quad (20)$$

where \mathcal{E}_c is the edge set of a connected, undirected, and acyclic[‡] communication graph $(\mathcal{V}, \mathcal{E}_c)$ between the buses. The additional constraints $\mu_i - \mu_j = 0$ force the local variables μ_i to agree at optimality. Letting $E_c \in \mathbb{R}^{n \times |\mathcal{E}_c|}$ denote the incidence matrix of the communication graph, the dual OFRP (20) is written in vector notation as

$$\begin{aligned} & \underset{\mu \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} \mu^\top K^{-1} \mu - (P^*)^\top \mu \\ & \text{subject to} \quad \mathbb{0}_{|\mathcal{E}_c|} = E_c^\top \mu, \end{aligned} \quad (21)$$

where now $\mu = (\mu_1, \dots, \mu_n)$. The problem (21) is a linearly constrained, strictly convex quadratic program of the form (1) from Section I, with $Q = K^{-1}$, $c = -P^*$, $S = E_c^\top$ and $b = \mathbb{0}_{|\mathcal{E}_c|}$. The corresponding Lagrangian is $L(\mu, \nu) = \frac{1}{2} \mu^\top K^{-1} \mu - (P^*)^\top \mu + \nu^\top E_c^\top \mu$, where $\nu \in \mathbb{R}^{|\mathcal{E}_c|}$ is a vector of multipliers, and the primal-dual algorithm with control output p then becomes

$$\begin{aligned} \tau_\mu \dot{\mu} &= -K^{-1} \mu + P^* - E_c \nu, \\ \tau_\nu \dot{\nu} &= E_c^\top \mu, \\ p &= -K^{-1} \mu, \end{aligned} \quad (22)$$

where as before τ_μ and τ_ν are positive diagonal matrices of controller gains. Since the graph $(\mathcal{V}, \mathcal{E}_c)$ is acyclic, the incidence matrix E_c has full column rank. Therefore by Lemma 2.1, the controller (22) converges exponentially to the global optimizer (μ^*, ν^*) of the problem (21). The output $p(t)$ of (22) is the input to the swing dynamics (17); the interconnection is a cascade. Since $p(t)$ an exponentially converging input to the exponentially stable linear system (17), the cascade is exponentially stable; we omit the details. It follows from the cascade structure that the map from P^* to p given by (22) is the same as the map from P^* to p after the systems are interconnected.

To evaluate the input/output performance of the primal-dual controller (22), we consider the case where P^* is subject to an additive disturbance, modeling fluctuating generation/load, noise, or other uncertainty. As in Section II, we shift the undisturbed equilibrium point of (22) to the origin, and following Section III we define the performance output

[‡]The acyclic assumption is made for consistency with our assumption that $\text{rank}(S) = r$ from Section II.

$y(t) = \frac{1}{\sqrt{2}} K^{\frac{1}{2}} p(t)$, such that $\|y(t)\|_2^2 = \frac{1}{2} p(t)^\top K p(t)$. Since $p(t) = -K^{-1} \mu(t)$, the performance output becomes $y(t) = -\frac{1}{\sqrt{2}} K^{-\frac{1}{2}} \mu$ or $\|y(t)\|_2^2 = \frac{1}{2} \mu(t)^\top K^{-1} \mu(t)$. We now apply Theorem 3.1 to obtain the following result.

Theorem 4.1 (Primal-Dual OFRP Performance): For the primal-dual OFRP dynamics (22), consider the corresponding shifted, input/output dynamics

$$\begin{aligned} \begin{bmatrix} \tau_\mu \dot{\mu} \\ \tau_\nu \dot{\nu} \end{bmatrix} &= \begin{bmatrix} -K^{-1} & -E_c \\ E_c^\top & \mathbb{0} \end{bmatrix} \begin{bmatrix} \mu \\ \nu \end{bmatrix} + \begin{bmatrix} B_1 \\ \mathbb{0} \end{bmatrix} \eta, \\ y &= -\frac{1}{\sqrt{2}} K^{-\frac{1}{2}} \mu, \end{aligned} \quad (23)$$

with disturbances η and performance outputs y . Then the squared \mathcal{H}_2 norm of (23) is

$$\|G\|_{\mathcal{H}_2}^2 = \text{Tr}(B_1^\top \tau_\mu^{-1} B_1) / 4. \quad (24)$$

Moreover, assuming that $\tau_\mu = \tau I_n$ and $B_1 = b I_n$ for some $\tau, b > 0$, we have that

$$\|G\|_{\mathcal{H}_2}^2 = \frac{b^2}{4\tau} n. \quad (25)$$

The result indicates that the input/output performance of the primal-dual frequency controller (22) is completely independent of the cost coefficients k_i and the incidence matrix E_c used to implement the distributed control; it depends *only* on the controller time-constants τ_μ , the disturbance strength B_1 , and the number of buses subject to disturbances.

Finally, we can consider an augmented Lagrangian leading to the augmented primal-dual OFRP dynamics

$$\begin{aligned} \tau_\mu \dot{\mu} &= -K^{-1} \mu + P^* - E_c \nu - \rho E_c E_c^\top \mu, \\ \tau_\nu \dot{\nu} &= E_c^\top \mu, \\ p &= -K^{-1} \mu, \end{aligned} \quad (26)$$

The matrix $\rho E_c E_c^\top$ is in fact a Laplacian matrix for the graph $(\mathcal{V}, \mathcal{E}_c)$. The additional term arising from the augmentation is therefore a distributed proportional consensus-type term on the μ variables, complementing the integral consensus-type term $-E_c \nu$. Applying Theorem 3.2, we obtain the following.

Theorem 4.2 (Augmented OFRP Performance): For the primal-dual OFRP dynamics (26) with the uniform parameters $K = k I_n$, $\tau_\mu = \tau_1 I_n$ and $\tau_\nu = \tau_2 I_{|\mathcal{E}_c|}$ for constants $k, \tau_1, \tau_2 > 0$, consider the corresponding shifted, input-output dynamics

$$\begin{aligned} \begin{bmatrix} \tau_1 \dot{\mu} \\ \tau_2 \dot{\nu} \end{bmatrix} &= \begin{bmatrix} -(\frac{1}{k} I_n + \rho E_c E_c^\top) & -E_c \\ E_c^\top & \mathbb{0} \end{bmatrix} \begin{bmatrix} \mu \\ \nu \end{bmatrix} + \begin{bmatrix} b I_n \\ \mathbb{0} \end{bmatrix} \eta, \\ y &= -\frac{1}{\sqrt{2k}} \mu, \end{aligned} \quad (27)$$

with disturbance inputs η and performance outputs y . Then the squared \mathcal{H}_2 norm of (27) is

$$\|G\|_{\mathcal{H}_2}^2 = \frac{b^2}{4\tau_1} + \frac{b^2}{4\tau_1} \sum_{i=1}^{n-1} \frac{1}{1 + \rho k \sigma_i^2}, \quad (28)$$

where σ_i is the i th non-zero singular value of E_c . Moreover, in the high augmentation gain limit $\rho \rightarrow \infty$, we have that

$$\lim_{\rho \rightarrow \infty} \|G\|_{\mathcal{H}_2}^2 = \frac{b^2}{4\tau_1}.$$

Proof: The proof is immediate by applying Theorem 3.2 and noting that $r = n - 1$ for the constraint $E_c^T \mu = \mathbf{0}_r$, since the graph $(\mathcal{V}, \mathcal{E}_c)$ is acyclic. \square

Remarkably, we find that by designing the frequency controller based on the augmented Lagrangian and making the gain ρ sufficiently high, the performance of the primal-dual OFR controller (26) becomes *independent* of network size, converging to a constant $b^2/4\tau_1$. Moreover, due to the special structure of graph incidence matrices, the algorithm remains distributed. Theorem 4.2 demonstrates that the performance characteristics of unaugmented primal-dual frequency controllers (Theorem 4.1) do not represent a fundamental performance limit for the approach.

V. CONCLUSIONS

We have quantified the input/output performance of primal-dual methods for online optimization by providing an explicit formula for the system's \mathcal{H}_2 norm from disturbances to a relevant performance output. Under some parametric restrictions, we extended our results to augmented Lagrangian primal-dual methods, and we then applied the results to quantify the performance of distributed secondary frequency controllers for power systems. Our results indicate that performance of the primal-dual method will degrade as the size of the network of cooperating agents grows, and that this scaling issue can be partially — and sometimes completely — compensated by using controllers derived from augmented Lagrangians.

From the perspective of distributed optimization, an open direction is to extend the system norm calculations considered herein to allow for more general strictly convex objective functions and for inequality constraints; the approach in [23] may be fruitful. Another important question is how one could further improve the \mathcal{H}_2 performance of primal-dual methods via feedback while maintaining the same distributed architecture of (6). Here we have considered a specific performance output; other outputs may be important to examine as well depending on the application of interest.

On the power systems side, the framework introduced here for measuring performance is by no means all-encompassing, and other performance metrics such as tracking errors under time-varying disturbances are of interest. An important extension of the present work would be to the case of “partial” primal-dual methods, where only a subset of buses participate in control. It will also be important to compare the results here to analogous calculations for both centralized optimal frequency controllers and distributed optimal frequency controllers based on consensus [24]; some of these extensions will be pursued in an extended publication to follow.

ACKNOWLEDGMENTS

The authors thank D. Gross and S. Bolognani for their valuable comments.

REFERENCES

- [1] T. Kose, “Solutions of saddle value problems by differential equations,” *Econometrica*, vol. 24, no. 1, pp. 59–70, 1956.
- [2] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*. Stanford University 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, 2010.
- [4] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in *IEEE Conf. on Decision and Control and European Control Conference*, Orlando, FL, USA, Dec. 2011, pp. 3800–3805.
- [5] S. H. Low and D. E. Lapsley, “Optimization flow control I: Basic algorithm and convergence,” *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–74, 1999.
- [6] A. R. Bergen and V. Vittal, *Power Systems Analysis*, 2nd ed. Prentice Hall, 2000.
- [7] C. Zhao, U. Topcu, N. Li, and S. Low, “Design and stability of load-side primary frequency control in power systems,” *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1177–1189, 2014.
- [8] C. Zhao, U. Topcu, and S. Low, “Swing dynamics as primal-dual algorithm for optimal load control,” in *IEEE Int. Conf. on Smart Grid Communications*, Tainan, Taiwan, Nov 2012, pp. 570–575.
- [9] N. Li, L. Chen, C. Zhao, and S. H. Low, “Connecting automatic generation control and economic dispatch from an optimization view,” in *American Control Conference*, Portland, OR, USA, Jun. 2014, pp. 735–740.
- [10] E. Mallada, C. Zhao, and S. H. Low, “Optimal load-side control for frequency regulation in smart grids,” in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sep. 2014, pp. 731–738.
- [11] E. Mallada and S. H. Low, “Distributed frequency-preserving optimal load control,” in *IFAC World Congress*, Cape Town, South Africa, Aug. 2014, pp. 5411–5418.
- [12] C. Zhao, E. Mallada, and S. Low, “Distributed generator and load-side secondary frequency control in power networks,” in *Conference on Information Sciences and Systems*, Baltimore, MD, USA, Mar. 2015, pp. 1–6.
- [13] S. You and L. Chen, “Reverse and forward engineering of frequency control in power networks,” in *IEEE Conf. on Decision and Control*, Los Angeles, CA, USA, Dec. 2014, pp. 191–198.
- [14] X. Zhang and A. Papachristodoulou, “A real-time control framework for smart power networks: Design methodology and stability,” *Automatica*, vol. 58, pp. 43 – 50, 2015.
- [15] T. Stegink, C. D. Persis, and A. van der Schaft, “A unifying energy-based approach to optimal frequency and market regulation in power grids,” 2015, submitted.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] M. Benzi, G. H. Golub, and J. Liesen, “Numerical solution of saddle point problems,” *Acta Numerica*, vol. 14, pp. 1–137, 2005.
- [18] A. Nedic and A. Ozdaglar, “Subgradient methods for saddle-point problems,” *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 205–228, 2009.
- [19] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory*, ser. Texts in Applied Mathematics. Springer, 2000, no. 36.
- [20] E. Tegling, B. Bamieh, and D. F. Gayme, “The price of synchrony: Evaluating the resistive losses in synchronizing power networks,” *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 254–266, 2015.
- [21] E. Tegling, M. Andreasson, J. W. Simpson-Porco, and H. Sandberg, “Improving performance of droop-controlled microgrids through distributed PI-control,” in *American Control Conference*, Boston, MA, USA, Jul. 2016, pp. 2321–2327.
- [22] B. K. Poolla, S. Bolognani, and F. Dörfler, “Placing Rotational Inertia in Power Grids,” in *American Control Conference*, Boston, MA, USA, Jul. 2016, pp. 2314–2320.
- [23] L. Lessard, B. Recht, and A. Packard, “Analysis and design of optimization algorithms via integral quadratic constraints,” *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.
- [24] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, “Breaking the hierarchy: Distributed control & economic optimality in microgrids,” *IEEE Transactions on Control of Network Systems*, 2016, to appear.