Saddle-Flow Dynamics for Distributed Feedback-Based Optimization

Chin-Yao Chang*, Marcello Colombino*, Jorge Cortés, Emiliano Dall'Anese

Abstract—**This paper develops a distributed saddle-flow algorithm to regulate the output of a networked system – modeled as static linear map – to the solution of a constrained convex optimization problem. The algorithm is "feedback-based," in the sense that measurements of the network output are leveraged in the saddle-flow updates to avoid a complete (oracle-based) knowledge of the network map. In the distributed architecture, each actuator has access to only a subset of measurements; nevertheless, supported by a connected communication graph, a distributed protocol is implemented to achieve consensus on pertinent dual variables associated with network-level output constraints and, therefore, on the solution of the constrained problem. Using a LaSalle argument, we show that under an easily satisfiable Linear Matrix Inequality condition the proposed algorithm converges to an optimal primal-dual solution. We demonstrate the effectiveness of the proposed method in a voltage regulation problem for power systems with high penetration of renewable generation.**

Index Terms—**Distributed control, Optimization, Network analysis and control**

I. Introduction and Problem Formulation

THIS paper considers the optimal operation of a networked physical system, and addresses the design of distributed algorithmic solutions to drive the system's outputs to the solution of a constrained optimization problem [1]. A real-time implementation of this task is challenging in many realistic applications – power grids, communication systems, and transportation systems to mention a few – since saddle-flow methods [2]–[4] and discrete-time gradient-based algorithmic solutions [5] typically require an accurate knowledge of the network input-output map. Take, for example, the following linear map:

$$y = Ax + Bz, \tag{1}$$

where $x \in \mathbb{R}^N$ are the control inputs, $y \in \mathbb{R}^m$ are the outputs, $A \in \mathbb{R}^{m \times N}$ is, in general, a full (i.e., non-sparse) matrix, $B \in \mathbb{R}^{m \times E}$, and $z \in \mathbb{R}^E$ is a constant but unknown vector of disturbances or exogenous inputs. Many large-scale engineering systems with distributed sensing and decision making can be modeled as (1) to a first approximation (e.g, power, traffic and water networks). Executing saddle-flow or gradient-based algorithms to determine optimal inputs for physical systems described by (1) requires one to estimate the matrices A and B, and gather measurements of a possibly high-dimensional vector z at a central location to evaluate the map; this task might be, in fact, prohibitive in many real-world networked systems. For example, if (1) represent a linearized AC power flow model for power grids [6]–[8], z is a high-dimensional vector ($E \gg 1$) collecting the non-controllable powers at all the nodes of the grid; measuring all the entries of z and estimating the matrix B is not practical or even not feasible. This motivates the development of "feedback-based" algorithmic solutions, where measurements of the network output y are leveraged in the updates of the algorithm to avoid a complete (oracle-based) knowledge of the network map. This approach is aligned with the recent works on feedback-based gradient methods [5], [6], [9], with successful application domains ranging from communication networks [10] to power systems [6]–[8], [11], [12] to transportation [13].

To outline the problem formulation concretely, let $w := Bz$ for brevity, with w constant but unknown and assume that the vector x stacks n sub-vectors $x_i \in \mathbb{R}^{N_i}$, with $\sum_{i=1}^n N_i = N$, where each sub-vector corresponds to the inputs applied by an agent or "actuator." Assume that m sensors are deployed in the network, and let $y^{(k)}$ be a sub-vector of y collecting measurements of the k-th sensor. The optimal steady-state operation of such networks can be cast as an optimization problem of the form:

$$\begin{aligned} &\min_{x \in \mathcal{X}} \sum_{i=1}^n f_i(x_i), \\ &\text{s.t. } y^{(k)} \le b^{(k)}, \quad k = 1, \dots, m, \\ &\qquad y = Ax + w, \end{aligned} \tag{2}$$

where $f_i : \mathbb{R}^{N_i} \to \mathbb{R}$ is a convex objective function available at the actuator i, $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$, and the convex sets $\{\mathcal{X}_i \in \mathbb{R}^{N_i}\}_{i=1}^n$ represent 'hard constraints' for each actuator that cannot be violated during the execution of the algorithm. The solution of (2) depends on the unknown disturbance w. Notice that, given problem (2), even gradient-free methods (see, e.g., [14]), would require knowledge of the input-output map appearing in the constraint. Our goal here is to design a distributed algorithm that, *without* knowledge of w and in the face of the non-sparsity of the input-output map, can leverage the measurements $\{y^{(k)}\}_{k=1}^m$ and a communication network to drive x and y to the solution of the convex problem (2) as illustrated in Figure 1. This is possible because the effect of w is reflected on the output y. Based on (2), the design of the

C-Y Chang and M. Colombino are with the National Renewable Energy Laboratory (NREL), {ChinYao.Chang,Marcello.Colombino}@nrel.gov; J. Cortés is with the Department of Mechanical and Aerospace Engineering at UC San Diego, cortes@ucsd.edu; E. Dall'Anese is with the Department of Electrical, Computer, and Energy Engineering at the University of Colorado Boulder, emiliano.dallanese@colorado.edu. This work was authored in part by NREL, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by DOE Office of Electricity, Advanced Grid Modeling Program, through agreement NO. 33652, and ARPA-e NODES program.

* The first two authors contributed equally.

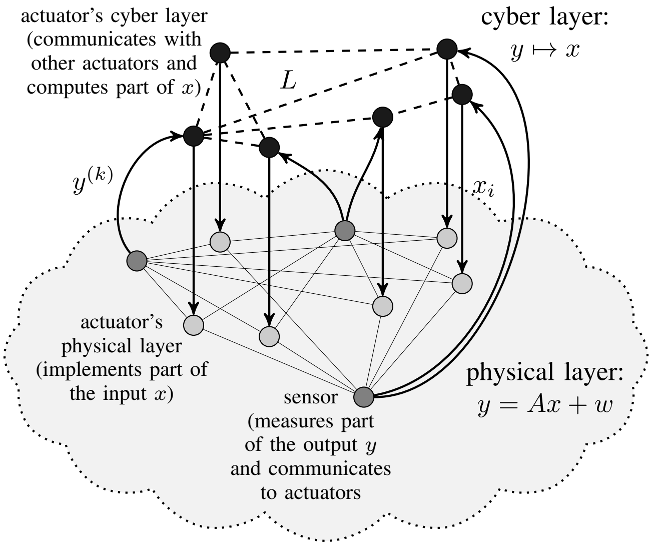

Fig. 1. Proposed distributed optimization scheme: The control framework is composed of cyber and physical layers. The cyber layer includes computation and communication that maps output measurements y to the input decisions x; the physical layer corresponds to a system that maps the inputs x and disturbances w to outputs y. The actuators communicate to each other through an undirected graph in the cyber layer and they locally update the value of their input x in real time based information received from a subset of the sensors and from other actuators. In the physical layer, the input x affects the output y which is measured by the sensors and communicated to a subset of actuators, thus closing the feedback loop.

algorithm is grounded on a saddle-flow-based approach (see e.g., [2]–[4], [15] and pertinent references therein). To address the fact that w is unknown, the standard saddle-flow dynamics [2] are appropriately modified to read the measurements $\{y^{(k)}\}_{k=1}^m$ collected by the sensors rather than using the full system model. Existing feedback-based optimization methods, however, rely either on a centralized [8], [12] or on a star (i.e., "gather and broadcast") [5] communication structures. Exceptions are, for example, the works [6], [16] (specific to voltage regulation tasks in power systems), where the communication graph matches the physical network but benefit from a reduced communication burden. In this work, we obtain a distributed architecture (Fig 1) using a consensus-enforcing term based on the vector of dual variables. Each sensor k transmits $y^{(k)}$ only to a subsets of actuators, which dynamically seek consensus on the dual variables to ensure that the constraints in (2) are satisfied. Relative to existing distributed saddle-flow methods (see e.g., [4], [15]) the proposed algorithm makes use of real-time measurements instead of the system model to distribute non-sparse constraints.

II. Preliminaries

This section presents the notation used throughout the paper and basic preliminary results. Let $\mathbb{R}$ and $\mathbb{R}_+$ denote the set of real and nonnegative reals, resp. Given a matrix A, $A^\top$ denotes its transpose, $A \succ (\prec) 0$ denotes that A is symmetric and positive (negative) definite. The matrix $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. For $x \in \mathbb{R}^n$, $\|x\|$ denotes its Euclidean norm and $\mathrm{diag}\{x\}$ a diagonal matrix with the elements of x on the main diagonal. Given vectors $x_1, \dots, x_n$, $\mathrm{diag}\{x_i\}_{i=1}^n$ denotes a block-diagonal matrix with the vectors x_i as diagonal blocks. The vector $\mathbf{1}_n$ is the vector of all 1s in $\mathbb{R}^n$. Given a set $\mathcal{X}$ and a point $x_0 \in \mathcal{X}$, we denote $\Pi_{\mathcal{X}}(x_0) = \mathrm{argmin}_{x\in\mathcal{X}} \|x - x_0\|$. Given a convex cone $\mathcal{K}$, the polar cone $\mathcal{K}^*$ is defined as $\mathcal{K}^* := \{y : y^\top x \le 0, \forall\, x \in \mathcal{K}\}$.

Lemma II.1. *Given a convex cone $\mathcal{K} \subset \mathbb{R}^n$ and a vector $a \in \mathbb{R}^n$, if $b = \Pi_{\mathcal{K}}(a)$, then $a - b \in \mathcal{K}^*$.*

Proof: We reason by contradiction. If $a - b \notin \mathcal{K}^*$, then there exists $x \in \mathcal{K}$ such that $(a-b)^\top x > 0$. By the optimality conditions of the projection operator, $(a-b)^\top(y-b) \le 0$ for all $y \in \mathcal{K}$. Since $\mathcal{K}$ is a cone, $\lambda x \in \mathcal{K}$ for $\lambda \ge 0$. Choosing $y = \lambda x$ with λ large enough leads to contradiction. ∎

Given a nonempty convex set $\mathcal{X} \subset \mathbb{R}^n$, the tangent cone of $\mathcal{X}$ at $x \in \mathcal{X}$ is a convex cone defined as

$$\mathcal{T}_{\mathcal{X}}^x := \mathrm{cl}\big(\{d \in \mathbb{R}^n : \exists\, \epsilon > 0, x + \epsilon d \in \mathcal{X}\}\big).$$

The normal cone of $\mathcal{X}$ at x is a convex cone defined as

$$\mathcal{N}_{\mathcal{X}}^x := \{d \in \mathbb{R}^n : d^\top(y - x) \le 0, \forall\, y \in \mathcal{X}\}.$$

According to [17, Theorem 6.9], $\mathcal{N}_{\mathcal{X}}^x = (\mathcal{T}_{\mathcal{X}}^x)^*$.

A differentiable function f is r-strongly convex if $(\nabla f(x) - \nabla f(y))^T(x - y) \ge r\|x - y\|_2^2$ holds for all x, y in its domain. We let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{\omega_{jk}\}_{(j,k)\in\mathcal{E}})$ denote a connected, undirected graph, without self-loops or multiple edges. Here, $\mathcal{V} = \{1, \dots, N\}$ is the set of nodes, $\mathcal{E} \subset (\mathcal{V}\times\mathcal{V}) \setminus \cup_{i\in\mathcal{V}}(i,i)$, with $|\mathcal{E}| = M$, is the set of edges, and $\{\omega_{jk}\}_{(j,k)\in\mathcal{E}}$ are the edge weights, with ω_{jk} corresponding to (j,k). The graph Laplacian $L \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ is uniquely defined by $(Lx)_i = \sum_{j:(i,j)\in\mathcal{E}} w_{ij}(x_j - x_i)$, for all $i = 1, \dots, |\mathcal{V}|$ and $x \in \mathbb{R}^{|\mathcal{V}|}$. The Laplacian L is positive semi-definite and its null-space is spanned by $\mathbf{1}_{|\mathcal{V}|}$. We refer the reader to [18] for a comprehensive review of algebraic graph theory.

We consider projected dynamical systems of the form

$$\dot{x} = \Pi_{\mathcal{T}_{\mathcal{X}}^x}(g(x)). \tag{3}$$

Since the dynamics is discontinuous, the standard notion of solution for ordinary differential equations does not apply. Throughout the paper, $x : [0,t] \to \mathbb{R}^n$ is a (Caratheodory) solution on the interval $[0,t]$ if it is absolutely continuous on $[0,t]$ and satisfies (3) almost everywhere in $[0,t]$. For all the systems considered in this paper existence and uniqueness of Caratheodory solutions is guaranteed. We refer the reader to [19], [20] for in-depth discussions on projected dynamical systems and conditions for existence and uniqueness of different notions of solutions.

III. Distributed Optimization with Feedback

In this section we describe the proposed algorithmic solution to the problem outlined in Section I. Consider the optimization problem (2) over the setting illustrated in Fig. 1. We make the following assumptions.

Assumption 1. ***(Optimization problem).*** *The functions $f_i : \mathcal{X}_i \to \mathbb{R}$ are continuously differentiable and r-strongly convex, and the sets $\mathcal{X}_i \in \mathbb{R}^{N_i}$ are nonempty, compact, and convex. The problem* (2) *is feasible and the Slater condition is satisfied.*

Assumption 2. ***(Sensing and communication requirements).*** *For each sensor $k = 1, \dots, m$, $\mathcal{A}^{(k)} \subset \{1, \dots, n\}$, with* *cardinality $q^{(k)} \leq n$ is the set of actuators that have access to its measurement. The actuators communicate through a connected, undirected graph $\mathcal{G}$.*

Under Assumption 1, the set of primal-dual optimizer of (2) is given by those pairs $(x^\star, \lambda^\star)$ satisfying the KKT conditions

$$\begin{aligned} -\nabla_x f(x^\star) - A^\top \lambda^\star \in \mathcal{N}_{\mathcal{X}}^{x^\star}, \quad x^\star \in \mathcal{X}, \\ Ax^\star + w - b \in \mathcal{N}_{\mathbb{R}_+^m}^{\lambda^\star}, \quad \lambda^\star \in \mathbb{R}_+^m . \end{aligned} \tag{4}$$

Given the strong convexity of the cost function $f(x) := \sum_{i=1}^n f_i(x_i)$, the problem (2) has a unique optimal solution $x^\star$ satisfying (4). The dual optimizer $\lambda^\star$, however, need not be unique. We define the set of optimal dual variables

$$\Lambda = \{\lambda^\star \mid (4) \text{ are satisfied for the unique optimal } x^\star\}. \tag{5}$$

Uniqueness of the optimal dual variables is not immediate when only a subset of the constraints is dualized. We refer the reader to [21] for conditions that imply uniqueness.

For ease of presentation, our treatment begins considering a simplified instance of (2) with only one constraint (and thus one sensor). The extension to the general case is straightforward but notation-heavy and is presented later.

A. Case with one constraint and one sensor

Consider the following simplified version of problem (2):

$$\min_{x \in \mathcal{X}} \sum_{i=1}^n f_i(x_i), \qquad \text{s.t. } a^\top x + w \leq b, \tag{6}$$

where $x = [x_1^\top, \ldots, x_n^\top]^\top \in \mathbb{R}^N$, $a = [a_1^\top, \ldots, a_n^\top]^\top \in \mathbb{R}^N$ and $w, b \in \mathbb{R}$. Let $y = a^\top x + w$ be the measured data at the sensor. A projected primal-dual dynamics feedback optimization of (6) can be written as [5],

$$\dot{x}_i = \Pi_{\mathcal{T}_{\mathcal{X}_i}^x} \left(-\nabla f_i(x_i) - a_i \lambda\right), \quad \forall i = 1, \ldots, n \tag{7a}$$

$$\dot{\lambda} = \Pi_{\mathcal{T}_{\mathbb{R}_+}^\lambda} \left(y - b\right), \tag{7b}$$

where $\lambda \in \mathbb{R}_+$ is the dual variable for the constraint. Algorithm (7) cannot be directly executed under the communication restrictions of Assumption 2 (except in the special case $q = n$), because not every agent can access y to compute λ. The standard approach to implement (7) is a "gather and broadcast" strategy where sensor measurements are gathered by a central operator that performs (7b) and broadcasts λ to all actuators.

B. Distributed algorithm design

In this section, we modify the algorithm (7) to account for the restrictions imposed by the sensing and communication requirements (cf. Assumption 2). To do so, we have actuators maintain estimates $\{\hat{\lambda}_i\}_{i=1}^n$ of the dual variable λ, which are continuously updated through peer-to-peer communication. Those that have access to the sensor data update their estimate based on the output y and local communications along the graph $\mathcal{G}$, the others based on communication alone. Without loss of generality, we order the actuators so that the first q receive sensor data and the others do not. We define $\hat{\lambda} = [\hat{\lambda}_1^\top, \ldots, \hat{\lambda}_n^\top]^\top = [\hat{\lambda}_a^\top, \hat{\lambda}_b^\top]^\top \in \mathbb{R}_+^n$, where $\hat{\lambda}_a \in \mathbb{R}_+^q$ collects the estimates of λ by the actuators that have access to sensor data and $\hat{\lambda}_b \in \mathbb{R}_+^{n-q}$ collects the estimates of λ by the actuators that do not. The proposed algorithm is

$$\dot{x}_i = \Pi_{\mathcal{T}_{\mathcal{X}_i}^x} \left(-\nabla_{x_i} f_i(x_i) - a_i \hat{\lambda}_i\right), \quad i = 1, \ldots, n \tag{8a}$$

$$\dot{\hat{\lambda}} = \begin{bmatrix} \Pi_{\mathcal{T}_{\mathbb{R}_+^q}^{\hat{\lambda}_a}} \left((y - b)\mathbf{1}_q\right) \\ 0 \end{bmatrix} - L\hat{\lambda}, \tag{8b}$$

where $\hat{\lambda}_i$ is a local copy of λ and $L \in \mathbb{R}^{n \times n}$ is the graph Laplacian of $\mathcal{G}$. Note that, thanks to the projections onto $\mathcal{T}_{\mathcal{X}_i}^x$ and $\mathcal{T}_{\mathbb{R}_+^q}^{\hat{\lambda}_a}$, together with the fact that $-L$ is a Metzler matrix, the sets $\mathcal{X}_i$ and $\mathbb{R}_+^n$ are invariant under (8a) and (8b), resp. Furthermore, both the primal update (8a) and the dual update (8b) are distributed, as those actuators that do not have access to the sensor measurement update their copy $\hat{\lambda}_i$ from communication with neighbors.

C. Convergence analysis

Here we analyze the convergence of the distributed algorithm (8). To ease the treatment, we write (8a) compactly as

$$\dot{x} = \Pi_{\mathcal{T}_{\mathcal{X}}^x} \left(-\nabla_x f(x) - \operatorname{diag}\{a_i\}_{i=1}^n \hat{\lambda}\right),$$

where $\hat{\lambda} \in \mathbb{R}^n = [\hat{\lambda}_1, \ldots \hat{\lambda}_n]^\top$. Note that $\operatorname{diag}\{a_i\}_{i=1}^n \mathbf{1}_n = a$. Next, we show that, under a mild condition on the communication graph, (8) converges to the optimal point of the optimization problem (6). We make the following simplifying assumption

Assumption 3. ***(Uniqueness of the optimal dual variables).*** *The optimal Lagrange multiplier $\lambda^\star$ is unique for problem* (6).

This assumption makes the treatment easier but it is not necessary (in fact, we drop it later in the general case).

Theorem III.1. ***(Convergence of*** (8)***).*** *Under Assumptions 1, 2, and 3, if the Linear Matrix Inequality (LMI)*

$$P = \begin{bmatrix} -rI_N & M_\xi \\ M_\xi^\top & -\frac{1}{q}L - \gamma \mathbf{1}_n \mathbf{1}_n^\top \end{bmatrix} \prec 0, \tag{9}$$

where

$$M_\xi = \frac{1}{2}\left(\operatorname{diag}\{a_i\}_{i=1}^n - \frac{1}{q}(a\mathbf{1}_n^\top)\begin{bmatrix} I_q & 0 \\ 0 & 0_{n-q} \end{bmatrix} + \xi \mathbf{1}_n^\top\right),$$

$\gamma > 0$, and $\xi \in \mathbb{R}^N$ is satisfied, then the trajectories of the online distributed algorithm (8) *converge to $[x^{\star\top}, \lambda^\star \mathbf{1}_n^\top]^\top$, where $[x^\star, \lambda^\star]^\top$ is the unique primal-dual optimum of* (6).

Proof: Let us consider the following positive-definite function

$$V(x, \hat{\lambda}) = \frac{1}{2}\left(\frac{1}{q}\|\hat{\lambda} - \hat{\lambda}^\star\|^2 + \|x - x^\star\|^2\right). \tag{10}$$

The time derivative of V along the trajectories of (8) is

$$\dot{V} = \frac{1}{q}(\hat{\lambda} - \hat{\lambda}^\star)^\top \dot{\hat{\lambda}} + (x - x^\star)^\top \dot{x}. \tag{11}$$

Using Lemma II.1 and the fact that $\mathcal{N}_{\mathcal{X}}^x = (\mathcal{T}_{\mathcal{X}}^x)^*$, we deduce

$$-\nabla_x f(x) - \operatorname{diag}\{a_i\}_{i=1}^n \hat{\lambda} - \dot{x} \in \mathcal{N}_{\mathcal{X}}^x.$$

By definition of normal cone and $x^\star \in \mathcal{X}$,

$$(x-x^\star)^\top(-\nabla_x f(x) - \mathrm{diag}\{a_i\}_{i=1}^n \hat{\lambda} - \dot{x}) \geq 0. \qquad (12)$$

With analogous reasoning we conclude that

$$\frac{1}{q}(\hat{\lambda}-\hat{\lambda}^\star)^\top \left(\begin{bmatrix} (y-b)\mathbf{1}_q \\ 0 \end{bmatrix} - L\hat{\lambda} - \dot{\hat{\lambda}} \right) \geq 0. \qquad (13)$$

By adding the non-negative quantities (12) and (13) to (11) and expanding $\dot{\hat{\lambda}}$ we obtain

$$\begin{aligned} \dot{V} \leq & \frac{1}{q}(\hat{\lambda}_a - \lambda_a^\star)^\top (y-b)\mathbf{1}_q \\ & - \frac{1}{q}(\hat{\lambda}-\hat{\lambda}^\star)^\top L(\hat{\lambda}-\hat{\lambda}^\star) \\ & - (x-x^\star)^\top \nabla_x f(x) - (x-x^\star)^\top \mathrm{diag}\{a_i\}_{i=1}^n \hat{\lambda}. \end{aligned} \qquad (14)$$

We next add the following terms to (14).

$$-\frac{1}{q}(\hat{\lambda}_a - \lambda_a^\star)^\top \mathbf{1}_q (y^\star - b), \qquad (15a)$$

$$(x-x^\star)^\top (\nabla_x f(x^\star) + \mathrm{diag}\{a_i\}_{i=1}^n \hat{\lambda}^\star), \qquad (15b)$$

which are non-negative due to the KKT stationary condition. Adding (15) to the r.h.s of (14) yields

$$\begin{aligned} \dot{V} \leq & -(x-x^\star)^\top \Big(\nabla_x f(x) - \nabla_x f(x^\star)\Big) \\ & -(x-x^\star)^\top \Big(\mathrm{diag}\{a_i\}_{i=1}^n (\hat{\lambda}-\hat{\lambda}^\star)\Big) \\ & + \frac{(\hat{\lambda}_a - \lambda_a^\star)^\top \mathbf{1}_q (y-y^\star) - (\hat{\lambda}-\hat{\lambda}^\star)^\top L(\hat{\lambda}-\hat{\lambda}^\star)}{q}. \end{aligned} \qquad (16)$$

Note that

$$\begin{aligned} \frac{1}{q}\lambda_a^{\star\top}\mathbf{1}_q(y-y^\star) &= \frac{1}{q}\lambda^\star \mathbf{1}_q^\top \mathbf{1}_q a^\top (x-x^\star) \\ &= \lambda^\star a^\top (x-x^\star) \\ &= \lambda^\star \mathbf{1}_n^\top \mathrm{diag}\{a_i^\top\}_{i=1}^n (x-x^\star) \\ &= \hat{\lambda}^{\star\top} \mathrm{diag}\{a_i^\top\}_{i=1}^n (x-x^\star). \end{aligned} \qquad (17)$$

We define $\tilde{\lambda}_a := \frac{1}{q}\mathbf{1}_q^\top \hat{\lambda}_a \in \mathbb{R}_+$. Using (17) and the r-strong convexity of f, (16) becomes

$$\begin{aligned} \dot{V} \leq & -r\|x-x^\star\|^2 \\ & -(x-x^\star)^\top \Big(\mathrm{diag}\{a_i\}_{i=1}^n (\hat{\lambda} - \mathbf{1}_n \tilde{\lambda}_a)\Big) \\ & - \frac{1}{q}(\hat{\lambda}-\hat{\lambda}^\star)^\top L(\hat{\lambda}-\hat{\lambda}^\star). \end{aligned} \qquad (18)$$

Let us rewrite $\hat{\lambda} = \hat{\lambda}_\parallel + \hat{\lambda}_\perp$, where $\mathbf{1}_n^\top \hat{\lambda}_\perp = 0$ and $\hat{\lambda}_\perp^\top \hat{\lambda}_\parallel = 0$. Next, we denote

$$\lambda_{a,\parallel} = \begin{bmatrix} I_q & 0 \\ 0 & 0_{n-q} \end{bmatrix} \hat{\lambda}_\parallel, \quad \lambda_{a,\perp} = \begin{bmatrix} I_q & 0 \\ 0 & 0_{n-q} \end{bmatrix} \hat{\lambda}_\perp.$$

Since all components of $\hat{\lambda}_\parallel$ are equal, we derive

$$\hat{\lambda}_\parallel = \frac{1}{q}\mathbf{1}_n \mathbf{1}_q^\top \lambda_{a,\parallel}. \qquad (19)$$

Using (19) and the definition of $\tilde{\lambda}_a$, we conclude that

$$\begin{aligned} & \mathrm{diag}\{a_i\}_{i=1}^n (\hat{\lambda} - \mathbf{1}_n \tilde{\lambda}_a) \\ = & \ \mathrm{diag}\{a_i\}_{i=1}^n \Big(\hat{\lambda}_\perp + \hat{\lambda}_\parallel - \frac{1}{q}\mathbf{1}_n \mathbf{1}_q^\top (\lambda_{a,\perp} + \lambda_{a,\parallel})\Big) \\ = & \left(\mathrm{diag}\{a_i\}_{i=1}^n - \frac{1}{q}(a\mathbf{1}_n^\top) \begin{bmatrix} I_q & 0 \\ 0 & 0_{n-q} \end{bmatrix} + \xi \mathbf{1}_n^\top \right) \hat{\lambda}_\perp, \end{aligned} \qquad (20)$$

for any $\xi \in \mathbb{R}^N$. Finally, we note that

$$\frac{1}{q}(\hat{\lambda}-\hat{\lambda}^\star)^\top L(\hat{\lambda}-\hat{\lambda}^\star) = \frac{1}{q}\hat{\lambda}_\perp^\top L \hat{\lambda}_\perp. \qquad (21)$$

Using (20) and (21) into (18), we conclude that

$$\begin{aligned} \dot{V} \leq & -r\|x-x^\star\|^2 - 2(x-x^\star)^\top M_\xi \hat{\lambda}_\perp \\ & - \frac{1}{q}\hat{\lambda}_\perp^\top (L - \gamma \mathbf{1}_n \mathbf{1}_n^\top)\hat{\lambda}_\perp \leq 0, \end{aligned} \qquad (22)$$

for any $\gamma \in \mathbb{R}$ and $\xi \in \mathbb{R}^N$. We therefore have $\dot{V} \leq z^\top P z \leq 0$, where $z = [(x-x^\star)^\top, \hat{\lambda}_\perp^\top]^\top$ and the last inequality follows from (9). The level sets of V are compact and positively invariant. The application of the LaSalle's invariance principle [22] ensures that the trajectories of (8) converge to the largest invariant set $M \subset \Omega$, where $\Omega = \{[x^\top, \hat{\lambda}^\top]^\top | \dot{V} = 0\}$. If $\dot{V} = 0$, then $z^\top P z = 0$, which is equivalent to $x = x^\star$, $\hat{\lambda}_\perp = 0$. Since $\hat{\lambda} = \lambda^\star \mathbf{1}$ is a necessary condition for invariance of M (since $\dot{x} \neq 0$ otherwise), the result is proven. ■

The following result shows that the LMI condition (9) in Theorem III.1 is always satisfied when the coupling in the communication graph is strong enough.

Proposition III.2. *Let μ be the second smallest eigenvalue of L. If $\mu \geq \frac{q}{r}\min_\xi \|M_\xi\|^2$, then the LMI (9) is satisfied.*

Proof: Note that for any $y \in \mathbb{R}^n$, for $\gamma \geq \mu/(qn)$ $y^\top L y + q\gamma y^\top \mathbf{1}_n \mathbf{1}_n^\top y \geq \mu \|y\|^2$. To see this, it suffices to write $y = (I - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)y + \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top y$ and follow simple algebraic manipulations noting that $L\mathbf{1}_n = 0$. Let us fix $\xi \in \mathbb{R}^N$. Using the Schur complement, (9) is satisfied if and only if $-L - q\gamma\mathbf{1}_n\mathbf{1}_n^\top + \frac{q}{r}M_\xi^\top M_\xi$ is negative definite, which is true when $-\mu + \frac{q}{r}\|M_\xi\|^2 < 0$, completing the proof. ■

D. General case: multiple constraints, multiple sensors and non-unique multipliers

In this section, we extend the results above on problem (2) for the case of one constraint and one sensor to the general case with multiple constraints and multiple sensors. For each $k = 1, \ldots, m$, we let $a^{(k)}$ denote the corresponding column of $A^\top$ and $B^{(k)} := \sum_{i \in \mathcal{A}^{(k)}} e_i e_i^\top \in \mathbb{R}^{n\times n}$, where e_i is the i^{th} canonical basis vector. Consider the dynamics

$$\dot{x} = \Pi_{\mathcal{T}_\mathcal{X}^x}\left(-\nabla_x f(x) - \sum_{k=1}^m \mathrm{diag}\{a_i^{(k)}\}_{i=1}^n \hat{\lambda}^{(k)}\right), \qquad (23a)$$

$$\dot{\hat{\lambda}}^{(k)} = B^{(k)} \Pi_{\mathcal{T}_{\mathbb{R}_+^n}^{\hat{\lambda}^{(k)}}}\Big((y^{(k)} - b^{(k)})\mathbf{1}_n\Big) - L^{(k)}\hat{\lambda}^{(k)}, \qquad (23b)$$

where $\mathcal{A}^{(k)}$ is defined in Assumption 2, $\hat{\lambda}^{(k)} \in \mathbb{R}_+^n$ is the vector of estimates of the multiplier for the k^{th} constraint and $L^{(k)}$ is the Laplacian of the graph associated to the communication network used by the actuator to reach consensus on $\hat{\lambda}^{(k)}$ (in the simplest case all $L^{(k)}$ are identical, but this is not necessary). Since the term

$$\Pi_{\mathcal{T}_{\mathbb{R}_+^n}^{\hat{\lambda}^{(k)}}}\Big((y^{(k)} - b^{(k)})\mathbf{1}_n\Big)$$

in (23b) is pre-multiplied by $B^{(k)}$, it needs to be computed only by those actuators that have access to the k^{th} sensor. This means that the algorithm is fully distributed, i.e., only relies on neighbor to neighbor communication. In the following we prove the convergence of (23) to the set of primal-dual optimal solutions of (2).

Theorem III.3. ***(Convergence of*** (23)***).*** *Under Assumptions 1 and 2, let*

$$P := \begin{bmatrix} -rI_N & M^{(1)}_{\xi^{(1)}} & \cdots & M^{(m)}_{\xi^{(m)}} \\ M^{(1)\top}_{\xi^{(1)}} & -\frac{1}{q^{(1)}}L^{(1)} - \gamma^{(1)}\mathbf{1}_n\mathbf{1}_n^\top & & \\ \vdots & & \ddots & \\ M^{(m)\top}_{\xi^{(m)}} & & & -\frac{1}{q^{(m)}}L^{(m)} - \gamma^{(m)}\mathbf{1}_n\mathbf{1}_n^\top \end{bmatrix},$$

with $M^{(k)}_{\xi^{(k)}}$ defined as

$$M^{(k)}_{\xi^{(k)}} := \frac{1}{2}\left(\mathrm{diag}\{a_i^{(k)}\}_{i=1}^n - \frac{1}{q^{(k)}}(a^{(k)}\mathbf{1}_n^\top)B^{(k)} + \xi^{(k)}\mathbf{1}_n^\top\right),$$

and $\gamma^{(k)} > 0$, and $\xi^{(k)} \in \mathbb{R}^N$ arbitrary. If $P \prec 0$, then the trajectories of the online distributed algorithm (23) *converge to $[x^{\star\top}, \lambda^{\star(1)}\mathbf{1}_n^\top, \ldots, \lambda^{\star(m)}\mathbf{1}_n^\top]^\top$, where $x^\star$ is the unique solution of* (2) *and $[\lambda^{\star(1)}, \ldots, \lambda^{\star(m)}]^\top \in \Lambda$ is an optimal dual variable.*

Proof: We first note that the primal optimum is unique under Assumption 1, but there could be multiple dual optima. Let us define the set $\hat{\Lambda} := \{[\lambda^{\star(1)}\mathbf{1}_n^\top, \ldots, \lambda^{\star(m)}\mathbf{1}_n^\top : [\lambda^{\star(1)}, \ldots, \lambda^{\star(m)}]^\top \in \Lambda\}$ and $\hat{\lambda}^{\star(k)}(\hat{\lambda}^{(k)}) := \Pi_{\hat{\Lambda}}(\hat{\lambda}^{(k)})$. Consider the following positive-definite function

$$V(x, \hat{\lambda}) = \frac{1}{2}\left(\|x - x^\star\|^2 + \sum_{k=1}^m \frac{\|\hat{\lambda}^{(k)} - \hat{\lambda}^{\star(k)}(\hat{\lambda}^{(k)})\|^2}{q^{(k)}}\right). \quad (24)$$

By Danskin's Theorem [23], the time derivative of (24) along the trajectories of (23) is given by

$$\dot{V} = \sum_{k=1}^m \frac{\left(\hat{\lambda}^{(k)} - \hat{\lambda}^{\star(k)}(\hat{\lambda}^{(k)})\right)\dot{\hat{\lambda}}^{(k)}}{q^{(k)}} + (x - x^\star)^\top \dot{x} \quad (25)$$

Let us rewrite $\hat{\lambda}^{(k)} = \hat{\lambda}^{(k)}_\parallel + \hat{\lambda}^{(k)}_\perp$, where $\mathbf{1}_n^\top \hat{\lambda}^{(k)}_\perp = 0$ and $\hat{\lambda}^{(k)\top}_\perp \hat{\lambda}^{(k)}_\parallel = 0$. Using the same arguments that lead to (20) in the proof of Theorem III.1, we conclude that

$$\begin{aligned} \dot{V} \le{}& -r\|x - x^\star\|^2 - 2(x - x^\star)^\top \sum_{k=1}^m M^{(k)}_{\xi^{(k)}}\hat{\lambda}^{(k)}_\perp \\ & - \sum_{k=1}^m \hat{\lambda}^{(k)\top}_\perp \left(\frac{1}{q^{(k)}}L^{(k)} - \gamma\mathbf{1}_n\mathbf{1}_n^\top\right)\hat{\lambda}^{(k)}_\perp = z^\top P z. \end{aligned} \quad (26)$$

where $z = [(x - x^\star)^\top, \hat{\lambda}^{(1)\top}_\perp, \ldots, \hat{\lambda}^{(m)\top}_\perp]^\top$. The level sets of V are compact and positively invariant. Therefore, by LaSalle's invariance principle [22], the trajectories of (23) converge to the largest invariant set $M \subset \Omega$, where $\Omega = \{[x^\top, \hat{\lambda}^{(1)\top}, \ldots, \hat{\lambda}^{(m)\top}]^\top \,|\, \dot{V} = 0\}$. If $\dot{V} = 0$ then $z^\top P z = 0$, and the latter is equivalent to $x = x^\star$ and $[\hat{\lambda}^{(1)\top}_\perp, \ldots, \hat{\lambda}^{(m)\top}_\perp]^\top = 0$. Since $\hat{\lambda}^{(k)} = \lambda^{\star(k)}\mathbf{1}_n$ with $\lambda^\star = [\lambda^{\star(1)}, \ldots, \lambda^{\star(m)}]^\top \in \Lambda$ is necessary for invariance of M ($\dot{x} \neq 0$ otherwise) the result is proven. ■

The next result shows that the matrix P in the statement of Theorem III.3 can always be made negative definite by strengthening the coupling in the communication graph.

Proposition III.4. *Let $\mu^{(k)}$ be the second smallest eigenvalue of $L^{(k)}$. If $\min_k \mu^{(k)} \ge \frac{\max_k q^{(k)}}{r}\min_{\xi^{(k)}} \|[M^{(1)}_{\xi^{(1)}}, \ldots, M^{(k)}_{\xi^{(k)}}]\|^2$, the matrix P in the statement of Theorem III.3 is negative definite.*

The proof is analogous to that of Proposition III.2 and we omit it for space reasons. Theorem III.3 and Proposition III.4 show that the distributed algorithm (23) converges to a primal-dual optimal point of (2) for any connected communication graph with strong enough coupling.

Remark III.5. Nonlinear notwork maps of the form $y = g(x, w)$ would make problem (2) nonconvex; we do not cover this case formally in this work. In practice, the method proposed here can be employed using local linear approximations. Preliminary results for power systems [7], [12] and the simulation results below suggest that the use of feedback in the algorithm (23) through sensor measurements improves robustness to model mismatch and the quality of the solution.

IV. Numerical Example in Power Systems

In this section we illustrate how the distributed feedback-based optimization algorithm (23) can be used as a distributed controller to optimally manage the operation of a distribution feeder with high renewable penetration. We consider the modified IEEE 37-node feeder with high penetration of Photo-Voltaic (PV) generation illustrated in Figure 2. We simulate the feeder using real data from Anatolia, CA, USA, for solar irradiance and load consumption for 10 hours with granularity of one second. We assume that the feeder is divided in three areas, each of which has access to the voltage measurements within the area and can decide to curtail the PV active power production to avoid over-voltage.

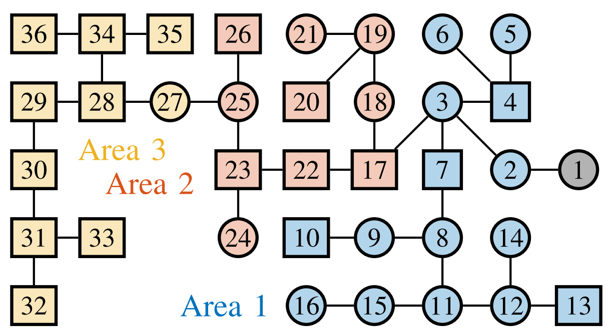

Fig. 2. IEEE 37-node feeder. Node 1 is the Point of Common Coupling (PCC). All other nodes are connected to a load and a voltage sensor. The square nodes are equipped with PV systems. The grid has been divided into three "areas" each of which contains an "actuator". Each actuator is responsible for computing the necessary curtailment of the PV systems in the pertinent area based on local (within the area) voltage measurements and communication to the other two actuators over the ring graph.

Following [7], we approximate the voltage magnitude by linearizing the AC power-flow equations around the zero-load operating point and we obtain a linear model of the form $y \approx Ax + w$, where y is a vector collecting the voltage magnitude at each bus, x is a vector collecting the decision variables (PV curtailment and reactive power injections) and w is a (time-varying) term that depends on the non-controllable loads (active and reactive) and the solar irradiance. We formulate a problem of the form (2) with upper and lower bounds on the voltages y, local constraint sets $\mathcal{X}_i$ on the decision variables and quadratic functions $x_i \mapsto f_i(x_i)$ that penalize active power curtailment and reactive power injections at the PVs. With one actuator per area, we use the ring graph with uniform unit weights for communication (instead, the LMI in Theorem III.3, which is conservative, would require a gain of 17.25). To demonstrate the robustness of the proposed feedback-based optimization algorithm, cf. Remark III.5, we use Matpower as a nonlinear AC power-flow solver (and not the linearized model) to simulate the feeder and obtain the data. Figure 3 shows the effectiveness of the distributed feedback-based method in avoiding over-voltages. Figure 4 shows a close-up of the Lagrange multipliers at node 35. We note that consensus is maintained during execution despite the time-varying nature of the problem.

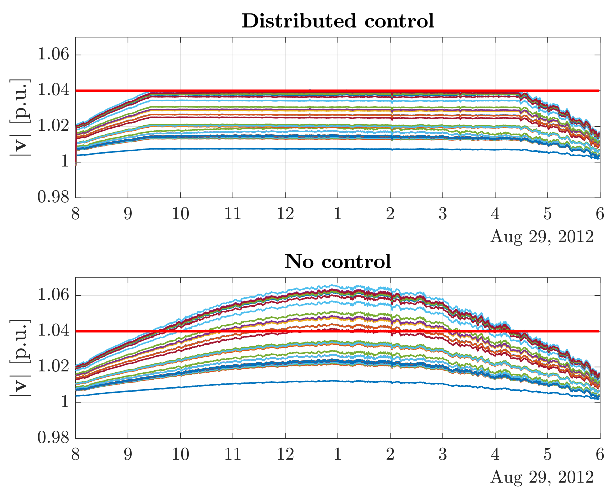

Fig. 3. Voltage magnitudes, the PVs are curtailed using the distributed feedback optimization scheme (23) is compared with the voltage magnitude without any control. We observe a significant improvement in avoiding over-voltage situations.

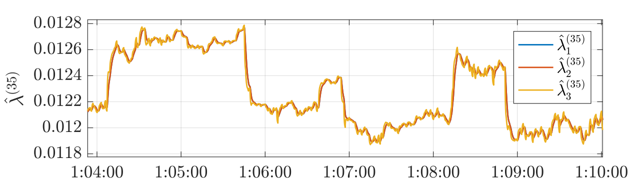

Fig. 4. A close-up of the time evolution of the Lagrange multipliers estimates $\{\hat{\lambda}_i^{(35)}\}_{i=1}^3$. Note that only the actuator in Area 3 has access to the voltage measurement at bus 35 to compute $\hat{\lambda}_3^{(35)}$. The other estimates are maintained using consensus alone.

V. Conclusions and Outlook

We have proposed a feedback-based, distributed, saddle-flow algorithm that, relying on sparse measurements and peer-to-peer communication, provably steers the output of a static linear system to the solution of a convex optimization problem. We have demonstrated its efficacy on a distributed voltage regulation problem. Interesting future research directions include tightening the stability conditions, extending the current results to time-varying and non-convex optimization problems, exploring the effect on algorithm performance of non-negligible dynamics in the physical system, and studying the robustness of the feedback-based algorithm to model mismatches and time-varying disturbances.

References

[1] A. Jokić, M. Lazar, and P. Van den Bosch, "On constrained steady-state regulation: Dynamic KKT controllers," *IEEE Trans. Auto. Contr.*, vol. 54, no. 9, pp. 2250–2254, Sep. 2009.

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

[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 *Proc. of 50th IEEE Conf. on Decision and Control*, Orlando, FL, Dec. 2011.

[5] A. Bernstein, E. Dall'Anese, and A. Simonetto, "Online primal-dual methods with measurement feedback for time-varying convex optimization," *IEEE Transactions on Signal Processing*, 2019, to appear; Arxiv preprint: https://arxiv.org/abs/1804.05159.

[6] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Trans. on Automatic Control*, vol. 60, no. 4, pp. 966–981, Apr. 2015.

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

[8] Y. Tang, K. Dvijotham, and S. Low, "Real-time optimal power flow," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2963–2973, 2017.

[9] L. S. Lawrence, J. W. Simpson-Porco, and E. Mallada, "The optimal steady-state control problem," *arXiv preprint arXiv:1810.12892*, 2018.

[10] 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. 861–874, 1999.

[11] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 3, pp. 254–264, 2016.

[12] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Projected gradient descent on riemannian manifolds with applications to online power system optimization," in *Communication, Control, and Computing, 2016 54th Annual Allerton Conference on*. IEEE, 2016, pp. 225–232.

[13] M. Vaquero and J. Cortés, "Distributed augmentation-regularization for robust online convex optimization," in *IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Groningen, The Netherlands, 2018, pp. 230–235.

[14] S. Michalowsky, B. Gharesifard, and C. Ebenbauer, "On the Lie bracket approximation approach to distributed optimization: Extensions and limitations," in *European Control Conference*, 2018, pp. 119–124.

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

[16] G. Qu and N. Li, "An optimal and distributed feedback voltage control under limited reactive power," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–7.

[17] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.

[18] F. Bullo, J. Cortés, and S. Martinez, *Distributed Control of Robotic Networks*, ser. Applied Mathematics Series. Princeton University Press, 2009.

[19] A. Nagurney and D. Zhang, *Projected dynamical systems and variational inequalities with applications*. Springer Science & Business Media, 2012, vol. 2.

[20] J. Cortés, "Discontinuous dynamical systems," *IEEE control Systems*, vol. 28, no. 3, 2008.

[21] S. Deng, "On uniqueness of lagrange multipliers in composite optimization," *Journal of mathematical analysis and applications*, vol. 201, no. 3, pp. 689–696, 1996.

[22] J. P. LaSalle, "Some extensions of Liapunov's second method," *Circuit Theory, IRE Transactions on*, vol. 7, no. 4, pp. 520–527, 1960.

[23] P. Bernhard and A. Rapaport, "On a theorem of Danskin with an application to a theorem of Von Neumann-Sion," *Nonlinear analysis*, vol. 24, no. 8, pp. 1163–1182, 1995.