

# Data-driven Coordination of Distributed Energy Resources for Active Power Provision

Hanchen Xu, *Student Member, IEEE*, Alejandro D. Domínguez-García, *Member, IEEE* and Peter W. Sauer, *Life Fellow, IEEE*

**Abstract**—In this paper, we propose a framework for coordinating distributed energy resources (DERs) connected to a power distribution system, the model of which is not completely known, so that they collectively provide a specified amount of active power to the bulk power system, while respecting distribution line capacity limits. The proposed framework consists of (i) a linear time-varying input-output (IO) system model that represents the relation between the DER active power injections (inputs), and the total active power exchanged between the distribution and bulk power systems (output); (ii) an estimator that aims to estimate the IO model parameters, and (iii) a controller that determines the optimal DER active power injections so the power exchanged between both systems equals to the specified amount at a minimum generating cost.

We formulate the estimation problem as a box-constrained quadratic program and solve it using the projected gradient descent algorithm. To resolve the potential issue of collinearity in the measurements, we introduce random perturbations in the DER active power injections during the estimation process. Using the estimated IO model, the optimal DER coordination problem to be solved by the controller can be formulated as a convex optimization problem, which can be solved easily. The effectiveness of the framework is validated via numerical simulations using the IEEE 123-bus distribution test feeder.

**Index Terms**—data-driven, distributed energy resource, coordination, active power provision, projected gradient descent, parameter estimation.

## I. INTRODUCTION

**I**N the modernization that electric power systems are currently undergoing, one goal is to massively integrate distributed energy resources (DERs) into power distribution systems [1]. These DERs, which include distributed generation resources, energy storage, demand response resources, and typically have small capacities, may be coordinated so as to collectively provide grid support services, e.g., reactive power support for voltage control [2]–[4], and active power control for frequency regulation [5]–[7].

In this paper, we focus on the problem of coordinating the response of a set of DERs in a lossy power distribution system so that they collectively provide some amount of active power to the bulk power system, e.g., in the form of demand response. Specifically, the DERs will be requested to collectively provide in real-time a certain amount of active power at the bus where

the power distribution system is interconnected with the bulk power system. In order for the DERs to fulfill such request, it is necessary to develop appropriate schemes that explicitly take into consideration the losses incurred. One way to include the losses in the problem formulation is to utilize a power-flow-like model of the power distribution system obtained offline. However, such a model may not be available or if so, it may not be accurate [8]–[10].

As an alternative to the aforementioned model-based approach, data-driven approaches have been demonstrated to be very effective in such situations where models are not readily available [9]–[16]. The fundamental idea behind data-driven approaches is to describe the system behavior by a linear time-varying (LTV) input-output (IO) model, and estimate the parameters of this model via regression using measurements of pertinent variables [11], [12]. Many previous works have applied data-driven approaches to power system problems, both in a steady-state setting [9], [10], [13], [14], and a dynamical setting [15], [16]. For example, in [9], the authors developed a data-driven framework to estimate linear sensitivity distribution factors such as injection shifting factors [13]; they further proposed a data efficient sparse representation to estimate these sensitivities [13]. This framework was later tailored to the problem of estimating the power flow Jacobian [10]. In [14], the authors used the estimation framework proposed in [9] to solve the security constrained economic dispatch problem. Data-driven approaches have also been applied successfully to develop power systems stabilizers [15], and damping controls [16]. We refer interested readers to [17] for an overview of data-driven approaches and their applications in a variety of other areas.

Yet, due to the collinearity in the measurements [18], the regression problem may be ill-conditioned, thus resulting in large estimation errors [19]. Though numerical approaches such as locally weighted ridge regression [18] and noise-assisted ensemble regression [19] can be used to mitigate the impacts of collinearity, there is no theoretical guarantee on the estimation errors. In this paper, we pursue the data-driven approach to develop a framework for coordinating the response of a set of DERs. The proposed framework consists of three components, namely (i) a model of the system describing the relation between the variables of interest to the problem, i.e., DER active power injections and power exchanged between the distribution and bulk power systems, (ii) an estimator, which provides estimates of the parameters that populate the model in (i); and (iii) a controller that uses the model in (i) with the parameters estimated via (ii) to determine the

This work has been funded in part by the Department of Energy under a subcontract from the University of Tennessee Knoxville, the Siebel Energy Institute, and the Advanced Research Projects Agency-Energy (ARPA-E) within the NODES program, under Award DEAR0000695.

The authors are with the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. Email: {hxcu45, aledan, psauer}@illinois.edu.

active power injection set-points of the DERs. Specifically, an LTV IO model is adopted as the system model to capture the relation between the DER active power injections (inputs), and the total active power exchange (output).

The parameters in this model are estimated by the estimator via the solution of a box-constrained quadratic program, obtained by using the projected gradient descent (PGD) algorithm. Inspired by ideas in power system model identification [20], [21], we introduce random perturbations in the DER active power injections during the estimation process to resolve the potential issue of collinearity in the measurements used by the estimator. We show that the estimation algorithm converges almost surely (a.s.) under some mild conditions, i.e., the estimated parameters converge to their true values, and the total provided active power also converges to the required amount. Using the estimated IO model, the optimal DER coordination problem (ODCP) to be solved by the controller can be formulated as a convex optimization problem, which can be solved easily. The major contributions of this paper are the data-driven coordination framework, the algorithm to solve the estimation problem, and its convergence analysis.

The remainder of this paper is organized as follows. Section II describes the power distribution system model and the ODCP of interest. Section III describes the components of the data-driven DER coordination framework. A description of the algorithm used in the framework, as well as its convergence analysis, is provided in Section IV. The proposed framework is illustrated and validated via numerical simulations on the IEEE 123-bus test feeder in Section V. Concluding remarks are presented in Section VI.

## II. PRELIMINARIES

In this section, we introduce the power distribution system model adopted in this work. We then discuss the DER coordination problem of interest.

### A. Power Distribution System Model

Consider a power distribution system represented by a directed graph that consists of a set of buses indexed by the elements in the set  $\mathcal{N} = \{0, 1, \dots, N\}$ , and a set of distribution lines indexed by the elements in the set  $\mathcal{L} = \{1, 2, \dots, L\}$ . Each line  $\ell \in \mathcal{L}$  is associated with a tuple  $(i, j) \in \mathcal{N} \times \mathcal{N}$ , where  $i$  is the sending end and  $j$  is the receiving end of a line, with the direction from  $i$  to  $j$  defined to be positive. Assume bus 0 corresponds to a substation bus, which is the only connection of the power distribution system to the bulk power system. Further, assume that bus 0 is an infinite bus that maintains a constant voltage magnitude. Without loss of generality, assume there is at most one DER and/or load at each bus, except bus 0, which does not have any DER or load. Let  $\mathcal{N}^g = \{1, \dots, n\}$  denote the DER index set. Throughout this paper, DERs are assumed to be controllable. Uncontrollable DERs are modeled as negative loads.

Let  $p_i^d$  and  $q_i^d$  respectively denote the active and reactive power loads at bus  $i$ ,  $i \in \mathcal{N}$ , and define  $\mathbf{p}^d = [p_1^d, \dots, p_N^d]$ , and  $\mathbf{q}^d = [q_1^d, \dots, q_N^d]$ . Let  $p_i^g$  and  $q_i^g$  respectively denote the active and reactive power injections from DER  $i$ ,  $i \in \mathcal{N}^g$ , and

define  $\mathbf{p}^g = [p_1^g, \dots, p_n^g]^\top$ , and  $\mathbf{q}^g = [q_1^g, \dots, q_n^g]^\top$ . Let  $\underline{p}_i^g$  and  $\bar{p}_i^g$  respectively denote the minimum and maximum active power that can be provided by DER  $i$ ,  $i \in \mathcal{N}^g$ , and define  $\underline{\mathbf{p}}^g = [\underline{p}_1^g, \dots, \underline{p}_n^g]^\top$ , and  $\bar{\mathbf{p}}^g = [\bar{p}_1^g, \dots, \bar{p}_n^g]^\top$ . Similarly, let  $\underline{q}_i^g$  and  $\bar{q}_i^g$  respectively denote the minimum and maximum reactive power that can be provided by DER  $i$ ,  $i \in \mathcal{N}^g$ , and define  $\underline{\mathbf{q}}^g = [\underline{q}_1^g, \dots, \underline{q}_n^g]^\top$ , and  $\bar{\mathbf{q}}^g = [\bar{q}_1^g, \dots, \bar{q}_n^g]^\top$ . Let  $p_i$  denote the active power injection at bus  $i \in \mathcal{N}$ , and define  $\mathbf{p} = [p_1, \dots, p_N]^\top$ ; then

$$\mathbf{p} = \mathbf{C}\mathbf{p}^g - \mathbf{p}^d, \quad (1)$$

where  $\mathbf{C} \in \mathbb{R}^{N \times n}$  is the matrix that maps the DER indices to the buses. The entry at the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column of  $\mathbf{C}$  is 1 if DER  $j$  is at bus  $i$ .

Let  $\tilde{\mathbf{M}} = [M_{i\ell}] \in \mathbb{R}^{(N+1) \times L}$ , with  $M_{i\ell} = 1$  and  $M_{j\ell} = -1$  if line  $\ell$  starts from bus  $i$  and ends at bus  $j$ , and all other entries equal to zero. Let  $\mathbf{M}$  denote the  $(N \times L)$ -dimensional matrix that results from removing the first row in  $\tilde{\mathbf{M}}$ . We first make the following assumption:

**Assumption 1.** The power distribution system is radial.

Under Assumption 1,  $L = N$ , and  $\mathbf{M}$  is invertible. Let  $f_\ell$  denote the active power that flows from the sending end to the receiving end of line  $\ell \in \mathcal{L}$ , and define  $\mathbf{f} = [f_1, \dots, f_L]^\top$ . Let  $\bar{f}_\ell$  denote maximum power flows on line  $\ell \in \mathcal{L}$ , and define  $\bar{\mathbf{f}} = [\bar{f}_1, \dots, \bar{f}_L]^\top$ . Then  $\mathbf{f}$  can be approximately computed from  $\mathbf{p}$  as follows:

$$\mathbf{f} \approx \mathbf{M}^{-1}\mathbf{p} = \mathbf{M}^{-1}(\mathbf{C}\mathbf{p}^g - \mathbf{p}^d). \quad (2)$$

We would like to point out that Assumption 1 allows us to approximately compute the active power flow on the distribution lines. It is, however, not a necessary condition for the analysis to be done in the rest of the paper.

Let  $y$  denote the active power exchanged between the distribution and bulk power systems via bus 0, defined to be positive if the flow is from the substation to the bulk power system. Conceptually,  $y$  can be represented as a function of  $\mathbf{p}^g$ ,  $\mathbf{q}^g$ ,  $\mathbf{p}^d$ ,  $\mathbf{q}^d$ . Note also  $\mathbf{q}^g$  is typically set according to some specific reactive power control rules to achieve certain objectives such as constant voltage magnitude or constant power factor [7], and thus is a function of  $\mathbf{p}^g$ ,  $\mathbf{p}^d$ ,  $\mathbf{q}^d$ . Then,  $y$  can be written as a function of  $\mathbf{p}^g$ ,  $\mathbf{p}^d$ ,  $\mathbf{q}^d$  as follows:

$$y = h(\mathbf{p}^g, \mathbf{p}^d, \mathbf{q}^d), \quad (3)$$

where  $h(\cdot)$  captures the impacts from both the physical laws as well as reactive power control. We emphasize that although the voltage control problem is not explicitly modeled in this paper, we assume certain voltage control schemes exist in the power distribution system such that the voltage profile will be maintained within an acceptable range. Indeed, voltage control schemes may have a significant impact on  $h$ . We refer interested readers to [7] for a detailed analysis on the impacts of voltage control schemes.

Note that the explicit form of  $h$  is difficult to obtain; however, we can make the following assumption on  $h$ :

**Assumption 2.** The function  $h$  is differentiable and its first order partial derivatives with respect to  $\mathbf{p}^g$  belong to  $[\underline{b}_1, \bar{b}_1]$ ,

where  $b_1, \bar{b}_1$  are some known constants. In addition,  $\frac{\partial h}{\partial \mathbf{p}^g}$  is a Lipschitz function, i.e., there exists  $b_2 > 0$  such that

$$\left\| \frac{\partial h}{\partial \mathbf{p}^g} \Big|_{\mathbf{a}} - \frac{\partial h}{\partial \mathbf{p}^g} \Big|_{\mathbf{b}} \right\| \leq b_2 \|\mathbf{a} - \mathbf{b}\|,$$

where  $\mathbf{a}, \mathbf{b} \in [\underline{\mathbf{p}}^g, \bar{\mathbf{p}}^g]$ , and  $\|\cdot\|$  denotes the  $L_2$ -norm.

Assumption 2 implies that, for fixed loads, the rate of change in  $y$  is bounded for bounded changes in the DER active power injections. In addition, the total active power provided to the bulk power system will increase when more active power is injected in the power distribution system. This assumption holds when the system is at a normal operating condition without line congestions.

### B. Optimal DER Coordination Problem

The DERs in the distribution system can collectively provide active power to the bulk power system as quantified by the power exchange between both systems at the substation bus. For example, the DERs can provide demand response services or frequency regulation services to the bulk power system; in both cases, the DERs need to be coordinated in such a way that the total active power provided to the bulk power system,  $y$ , tracks some pre-specified value, denoted by  $y^*$ . The objective of the ODCP is to determine the DER active power injections,  $\mathbf{p}^g$ , that minimize some cost function, e.g., one that reflects the cost of active power provision, while respecting to the following constraints:

- [C1.] the active power exchanged between the distribution and bulk power systems via bus 0,  $y$ , tracks some pre-specified value  $y^*$ ;
- [C2.] the active power injection from each DER  $i$ ,  $i \in \mathcal{N}^g$ , does not exceed its corresponding capacity limits, i.e.,  $\underline{\mathbf{p}}^g \leq \mathbf{p}^g \leq \bar{\mathbf{p}}^g$ ;
- [C3.] the power flow on each line  $\ell$ ,  $\ell \in \mathcal{L}$ , does not exceed its maximum capacity, i.e.,  $-\bar{\mathbf{f}} \leq \mathbf{f} \leq \bar{\mathbf{f}}$ .

Note that while constraint **C2** is a hard constraint that cannot be violated, constraint **C3** may be allowed to be violated slightly for a short period. The ODCP can be formulated as the following optimization problem:

$$\underset{\mathbf{p}^g \in [\underline{\mathbf{p}}^g, \bar{\mathbf{p}}^g]}{\text{minimize}} \quad c(\mathbf{p}^g),$$

subject to

$$h(\mathbf{p}^g, \mathbf{p}^d, \mathbf{q}^d) = y^*, \quad (4a)$$

$$-\bar{\mathbf{f}} \leq \mathbf{M}^{-1}(\mathbf{C}\mathbf{p}^g - \mathbf{p}^d) \leq \bar{\mathbf{f}}, \quad (4b)$$

where  $c(\cdot)$  denotes the cost function of the active power injections. This problem is difficult, however, when the model describing the power exchange with the bulk power system, as captured by  $h$ , is unknown. In this paper, we will resort to a data-driven approach to tackle this problem.

## III. DER COORDINATION FRAMEWORK

In this section, we describe the building blocks of the proposed framework; namely an LTV IO model, an estimator, and a controller.

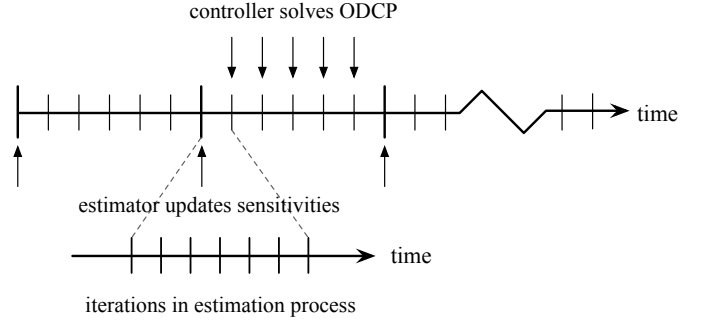


Fig. 1. Timescale separation of actions taken.

### A. Overview

The DER coordination framework consists of three components, namely (i) an LTC model of the system describing the relation between  $y$  and  $\mathbf{u}$ , (ii) an estimator that provides estimates of the parameters—the so-called sensitivity vector—that populate the model in (i); and (iii) a controller that uses the model in (i) with the parameters estimated via (ii) to solve the ODCP. This framework works on two timescales—a slow one and a fast one, as illustrated in Fig. 1. On the slow timescale, the controller determines the DER active power injection set-points by solving the ODCP periodically. The sensitivity vector is also updated by the estimator periodically. However, during each update of the sensitivity vector, the estimator needs to take actions in several iterations on a fast timescale. Since the sensitivity vector may not change significantly in a short time, it is used in the ODCP for several time instants before it is updated again. Before we proceed to presenting the detailed components in the proposed framework, we make the following assumption:

**Assumption 3.**  $\mathbf{p}^d$  and  $\mathbf{q}^d$  are constant during the estimation process; therefore, changes in  $y$  that occur across iterations in the estimation process depend only on changes in  $\mathbf{p}^g$ .

*Remark 1.* Assumption 3 allows us to determine the impacts of the DER active power injections on the output. When the load variability is significant enough so that it cannot be neglected during the estimation process, it becomes necessary to measure the loads and determine their impacts on the output as well. This is beyond the scope of the paper, and therefore we will leave it as future work.

### B. Input-Output System Model

For notational simplicity in the later development, define  $\mathbf{u} = \mathbf{p}^g$ ,  $\underline{\mathbf{u}} = \underline{\mathbf{p}}^g$ ,  $\bar{\mathbf{u}} = \bar{\mathbf{p}}^g$ , and  $\boldsymbol{\pi} = [(\mathbf{p}^d)^\top, (\mathbf{q}^d)^\top]^\top$ ; then, (3) can be written as:

$$y = h(\mathbf{u}, \boldsymbol{\pi}). \quad (5)$$

Unless otherwise noted, throughout this paper,  $x[k]$  denotes the value that some variable  $x$  takes at iteration  $k$ . It follows from (5) and Assumption 2 that  $y[k-1] = h(\mathbf{u}[k-1], \boldsymbol{\pi})$  and  $y[k] = h(\mathbf{u}[k], \boldsymbol{\pi})$ . Then, by the Mean Value Theorem, there

exists  $a_k \in [0, 1]$  and  $\tilde{\mathbf{u}}[k] = a_k \mathbf{u}[k] + (1 - a_k) \mathbf{u}[k - 1]$  such that

$$\begin{aligned} y[k] - y[k - 1] &= h(\mathbf{u}[k], \boldsymbol{\pi}) - h(\mathbf{u}[k - 1], \boldsymbol{\pi}) \\ &= \boldsymbol{\phi}[k]^\top (\mathbf{u}[k] - \mathbf{u}[k - 1]), \end{aligned}$$

where  $\boldsymbol{\phi}[k]^\top = [\phi_1[k], \dots, \phi_n[k]] = \left. \frac{\partial h}{\partial \mathbf{u}} \right|_{\bar{\mathbf{u}}[k]}$ ,<sup>1</sup> is referred to as the sensitivity vector at iteration  $k$ . It follows from Assumption 3 that  $\phi_i[k] \in [\underline{b}_i, \bar{b}_i]$ ,  $i = 1, \dots, n$ . Therefore, at any iteration  $k$ , (5) can be transformed into the following equivalent LTV IO model:

$$y[k] = y[k - 1] + \boldsymbol{\phi}[k]^\top (\mathbf{u}[k] - \mathbf{u}[k - 1]). \quad (6)$$

### C. Estimator On Fast Timescale

As illustrated in Fig. 1, the estimator updates the sensitivity vector across several iterations on the fast timescale. At iteration  $k$ , the objective of the estimator is to obtain an estimate of  $\boldsymbol{\phi}[k]$ , denoted by  $\hat{\boldsymbol{\phi}}[k]$ , using measurements collected up to iteration  $k$ , i.e.,  $y[k - 1], \mathbf{u}[k - 1], y[k - 2], \mathbf{u}[k - 2], \dots$ ; we formulate this estimation problem as follows:

$$\hat{\boldsymbol{\phi}}[k] = \arg \min_{\hat{\boldsymbol{\phi}} \in \mathcal{B}} J^e(\hat{\boldsymbol{\phi}}) = \frac{1}{2} (y[k - 1] - \hat{y}[k - 1])^2,$$

subject to

$$\hat{y}[k - 1] = y[k - 2] + \hat{\boldsymbol{\phi}}^\top (\mathbf{u}[k - 1] - \mathbf{u}[k - 2]), \quad (7)$$

where  $\mathcal{B} = [\underline{b}_1, \bar{b}_1]^n$ ,  $J^e(\cdot)$  is the cost function of the estimator, and  $\hat{y}[k - 1]$  is the value of  $y[k - 1]$  estimated by the IO model at iteration  $k$ . Essentially, (7) aims to find  $\hat{\boldsymbol{\phi}}$  that minimizes the squared error between the estimated value and the true value of  $y$ . Then,  $\hat{\boldsymbol{\phi}}[k]$  is used in the controller to determine the control for the next time instant.

During the estimation process, it is still necessary to track the output target. Therefore, at each iteration, the control is set based on the solution to the following problem:

$$\mathbf{u}[k] = \arg \min_{\mathbf{u} \in \mathcal{U}} J^c(\mathbf{u}) = \frac{1}{2} (y^* - \hat{y}[k])^2,$$

subject to

$$\hat{y}[k] = y[k - 1] + \hat{\boldsymbol{\phi}}[k]^\top (\mathbf{u} - \mathbf{u}[k - 1]), \quad (8)$$

where  $\mathcal{U} = [\underline{\mathbf{u}}, \bar{\mathbf{u}}]$ , and  $J^c(\cdot)$  is the cost function. Note that  $\hat{\boldsymbol{\phi}}[k]$  is used in (8) to predict the value of  $y[k]$  for a given  $\mathbf{u}$ . Different from the ODCP, the objective of the controller during the estimation process is that the output tracks the target and there may exist multiple solutions to this problem. This objective is chosen such that the DER active power injections behave in a way that can improve the estimation accuracy, as will be shown later in Section IV.

<sup>1</sup>We adopt the convention that the partial derivative of a scalar function with respect to a vector is a row vector.

### D. Controller On Slow Timescale

As illustrated in Fig. 1, the controller solves the ODCP to determine the least-cost active power set-points for DERs on the slow timescale. Meanwhile, it also forces the DERs to inject random active power perturbations at each iteration on the fast timescale. The ODCP to be solved by the controller is as follows:

$$\underset{\mathbf{p}^g \in [\underline{\mathbf{p}}^g, \bar{\mathbf{p}}^g]}{\text{minimize}} \quad c(\mathbf{p}^g),$$

subject to

$$y + \hat{\boldsymbol{\phi}}^\top (\mathbf{p}^g - \tilde{\mathbf{p}}^g) = y^*, \quad (9a)$$

$$-\bar{\mathbf{f}} \leq \mathbf{M}^{-1}(\mathbf{C}\mathbf{p}^g - \mathbf{p}^d) \leq \bar{\mathbf{f}}, \quad (9b)$$

where  $y$  is the output and  $\tilde{\mathbf{p}}^g$  is the DER active power injection vector at the beginning of the current time instant, and  $\hat{\boldsymbol{\phi}}$  is the up-to-date sensitivity vector. When  $c$  is a convex function, (9) is a convex problem and therefore can be solved easily with convergence guarantees. Note that this formulation is obtained by replacing (4a) in (4) by the estimated IO model.

## IV. ESTIMATION ALGORITHM AND ITS CONVERGENCE

The ODCP can be solved using existing algorithms for convex optimization and thus we do not discuss it in more details here. In this section, we focus on the problem faced by the estimator and propose a PGD based algorithm to solve it. We then provide convergence results for the proposed algorithm.

### A. Estimation Algorithm

We first describe the basic workflow of the proposed algorithm. Each iteration consists of an estimation step and a control step. At the beginning of iteration  $k$ ,  $y[k - 1]$  is available to the estimator, which uses it to update the estimate of the sensitivity vector. The updated estimate of the sensitivity vector,  $\hat{\boldsymbol{\phi}}[k]$ , is then used in the controller to determine the control,  $\mathbf{u}[k]$ . Then, the DERs are instructed to change their active power injection set-points based on  $\mathbf{u}[k]$ . At time instant  $k + 1$ , the estimation and control iterations are repeated once  $y[k]$  becomes available. The sequential process described above, which happens on the fast timescale, is illustrated as follows:

$$\dots \mathbf{u}[k - 1] \rightarrow \underbrace{\mathbf{y}[k - 1]}_{\text{estimation step}} \rightarrow \underbrace{\hat{\boldsymbol{\phi}}[k] \rightarrow \mathbf{u}[k]}_{\text{control step}} \rightarrow \mathbf{y}[k] \rightarrow \hat{\boldsymbol{\phi}}[k + 1] \dots$$

Problems (7) and (8) can be solved using the PGD method (see, e.g., [22]). Let  $\mathbf{P}_{\mathbb{V}_1 \rightarrow \mathbb{V}_2}$  denote the projection operator from a vector space  $\mathbb{V}_1$  to its (arbitrary) subspace  $\mathbb{V}_2$ , i.e.,

$$\mathbf{P}_{\mathbb{V}_1 \rightarrow \mathbb{V}_2}(\mathbf{v}_1) = \arg \min_{\mathbf{v}_2 \in \mathbb{V}_2} \|\mathbf{v}_2 - \mathbf{v}_1\|,$$

where  $\mathbf{v}_1 \in \mathbb{V}_1$ . For ease of notation, when the vector space to which  $\mathbf{v}_1$  belongs is unambiguous, we simply write  $\mathbf{P}_{\mathbb{V}_2}(\mathbf{v}_1)$  instead of  $\mathbf{P}_{\mathbb{V}_1 \rightarrow \mathbb{V}_2}(\mathbf{v}_1)$ .

Define the tracking error at iteration  $k$  as  $e[k] = y[k] - y^*$ . In addition, define  $\Delta y[k] = y[k] - y[k - 1]$  and  $\Delta \mathbf{u}[k] =$

---

**Algorithm 1:** Estimation Algorithm
 

---

**Input:**

$y$ : output  
 $y^*$ : output tracking target  
 $\delta$ : maximum allowed tracking error  
 $\hat{\phi}^0$ : initial estimate of sensitivity vector  
 $\mathbf{u}^0$ : initial DER active power injection set-point

**Output:**

$\mathbf{u}$ : DER active power injection set-point  
 $\hat{\phi}$ : estimate of sensitivity vector

**Initialization:** set  $\hat{\phi}[0] = \hat{\phi}^0$ ,  $\mathbf{u}[-1] = \mathbf{u}[0] = \mathbf{u}^0$ , obtain measurement of  $y[-1]$ , set  $k = 1$

**while**  $|e[k]| > \delta$  **do**

obtain new measurement of  $y[k-1]$   
 compute  $e[k-1] = y[k-1] - y^*$   
 compute  $\Delta y[k-1] = y[k-1] - y[k-2]$   
 compute  $\Delta \mathbf{u}[k-1] = \mathbf{u}[k-1] - \mathbf{u}[k-2]$   
 update the sensitivity vector estimate according to

$$\hat{\phi}[k] = \mathbb{P}_B \left( \hat{\phi}[k-1] - \alpha_k \Delta \mathbf{u}[k-1] \right. \\ \left. (\Delta \mathbf{u}[k-1]^\top \hat{\phi}[k-1] - \Delta y[k-1]) \right)$$

update the control vector,  $\mathbf{u}$ , according to

$$\mathbf{u}[k] = \mathbb{P}_U \left( \mathbf{u}[k-1] - \beta_k e[k-1] \mathbf{W}[k] \hat{\phi}[k] \right)$$

change DER active power injections to  $\mathbf{u}[k]$   
set  $k = k + 1$

**end**

---

$\mathbf{u}[k] - \mathbf{u}[k-1]$ . The partial derivative vector of  $J^e(\hat{\phi})$  with respect to  $\hat{\phi}$  is

$$\frac{\partial J^e(\hat{\phi})}{\partial \hat{\phi}} = \Delta \mathbf{u}[k-1] (\Delta \mathbf{u}[k-1]^\top \hat{\phi} - \Delta y[k-1]), \quad (10)$$

and that of  $J^c(\mathbf{u})$  with respect to  $\mathbf{u}$  is

$$\frac{\partial J^c(\mathbf{u})}{\partial \mathbf{u}} = \hat{\phi}[k] (\hat{\phi}[k]^\top (\mathbf{u} - \mathbf{u}[k-1]) + e[k-1]). \quad (11)$$

Instead of solving both (7) and (8) to completion, we iterate the PGD algorithm that would solve them for one step at each iteration. Specifically, at iteration  $k$ , we evaluate the new gradient at  $\hat{\phi}[k-1]$  and  $\mathbf{u}[k-1]$  and iterate once. Thus, by using (10) and (11), the update rules for  $\hat{\phi}$  and  $\mathbf{u}$ , respectively, are

$$\hat{\phi}[k] = \mathbb{P}_B \left( \hat{\phi}[k-1] - \alpha_k \Delta \mathbf{u}[k-1] \right. \\ \left. (\Delta \mathbf{u}[k-1]^\top \hat{\phi}[k-1] - \Delta y[k-1]) \right), \quad (12)$$

$$\mathbf{u}[k] = \mathbb{P}_U \left( \mathbf{u}[k-1] - \beta_k e[k-1] \hat{\phi}[k] \right), \quad (13)$$

where  $\alpha_k > 0$  and  $\beta_k > 0$  are the estimation and control step sizes at iteration  $k$ .

In order to resolve the potential issue of collinearity in the measurements used by the estimator, we introduce random perturbations during the estimation process. Define  $\mathbf{W}[k] =$

$\text{diag}(w_1[k], \dots, w_n[k])$ , where  $w_i[k]$ 's are independent random variables that follow a Bernoulli distribution with a probability parameter of 0.5. Then, the control update rule in (13) is modified, resulting in:

$$\mathbf{u}[k] = \mathbb{P}_U \left( \mathbf{u}[k-1] - \beta_k e[k-1] \mathbf{W}[k] \hat{\phi}[k] \right). \quad (14)$$

Intuitively, this means that at each iteration, the control of each DER is updated with a probability of 0.5. The random perturbation in the control is key to establish convergence of the parameter estimation process. The estimation algorithm, along with its initialization, is summarized in Algorithm 1, where  $\mathbf{u}^0$  is the vector of DER active power injections at the beginning of the time instant at which the estimation starts and  $\hat{\phi}^0$  is the up-to-date estimate of the sensitivity vector at the beginning of the same time instant.

### B. Convergence Analysis

The main convergence result for the control step during the estimation process is stated next.

**Theorem 1.** Using the estimation update rule in (12) and the control update rule in (14) with  $\beta_k \in (\frac{\epsilon}{b_1^2}, \frac{1}{nb_1^2})$ , where  $0 < \epsilon < \frac{b_1^2}{nb_1^2}$  is a given parameter, the system attains one of the following equilibria: 1)  $e[k]$  converges to 0 a.s.; 2)  $e[k]$  converges to some positive constant and  $\mathbf{u}[k]$  stays at  $\underline{\mathbf{u}}$ ; 3)  $e[k]$  converges to some negative constant and  $\mathbf{u}[k]$  stays at  $\bar{\mathbf{u}}$ . In all cases,  $\lim_{k \rightarrow \infty} \Delta \mathbf{u}[k] = \mathbf{0}_n$ , where  $\mathbf{0}_n \in \mathbb{R}^n$  is an all-zeros vector.

Theorem 1 shows something intuitive, i.e., the tracking error will be positive (negative) if the requested active power is less (more) than the minimum (maximum) amount of active power the DERs can provide; otherwise, the tracking error goes to zero a.s.

We note that  $\epsilon$  has a direct impact on the convergence rate of the control algorithm. This is more obvious in a deterministic setting, when the control update rule in (13) is used instead of the one in (14). A result on the convergence rate is given in the following corollary.

**Corollary 1.** Assume  $\mathbf{u}[k] \neq \underline{\mathbf{u}}$  and  $\mathbf{u}[k] \neq \bar{\mathbf{u}}, \forall k \in \mathbb{N}$ . Using the estimation update rule in (12) and the control update rule in (13) with  $\beta_k \in (\frac{\epsilon}{b_1^2}, \frac{1}{nb_1^2})$ , where  $\epsilon > 0$  is a given parameter,  $e[k]$  converges to 0 at a rate smaller than  $1 - \epsilon$ , i.e.,  $\left| \frac{e[k]}{e[k-1]} \right| < 1 - \epsilon$ .

We refer the readers to Appendix A for detailed proofs of the convergence results above.

Next, we establish the convergence of the estimation step. Define the estimation error vector at iteration  $k$  as  $\varepsilon[k] = \hat{\phi}[k] - \phi[k]$ . Since both  $\hat{\phi}[k]$  and  $\phi[k]$  are bounded,  $\varepsilon[k]$  is also bounded. Define  $\Delta \phi[k] = \phi[k] - \phi[k-1]$ . The convergence result for the estimation step is stated next.

**Theorem 2.** Using the estimation update rule in (12) and the control update rule (14), with  $\alpha_{k+1} = \frac{2}{\|\Delta \mathbf{u}[k]\|^2}$ ,  $\beta_k \in (\frac{\epsilon}{nb_1^2}, \frac{1}{nb_1^2})$ , where  $0 < \epsilon < \frac{b_1^2}{nb_1^2}$  is a given parameter, if

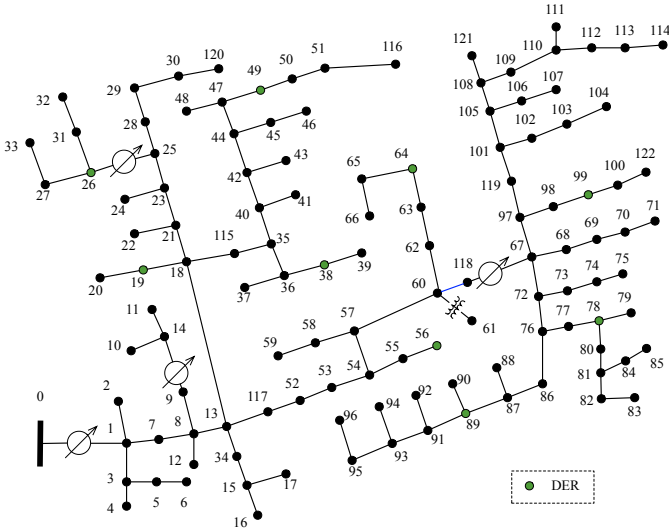


Fig. 2. IEEE 123-bus distribution test feeder.

$\mathbf{u}[k] \in (\underline{\mathbf{u}}, \bar{\mathbf{u}})$  and  $e[k] \neq 0, \forall k \in \mathbb{N}$ , then  $\|\varepsilon[k]\|$  converges to 0 a.s.

The intuition is that the estimation error goes to zero if the system can be continuously excited (guaranteed by the condition  $\mathbf{u}[k] \in (\underline{\mathbf{u}}, \bar{\mathbf{u}})$  and  $e[k] \neq 0, \forall k \in \mathbb{N}$ ).

We refer the readers to Appendix B for detailed proofs of the convergence result above.

## V. NUMERICAL SIMULATIONS

In this section, we illustrate the application of the proposed DER coordination framework and validate the theoretical results presented earlier. From a practical point of view, the timescale separation illustrated in Fig. 1 is critical for the applicability of the proposed framework. Specifically, the estimation process needs to be much faster than the timescale governing the load changes. The DERs, which are typically power electronics interfaced, can respond very quickly, on a timescale of millisecond to second [23]. In this simulation, we set the duration between two iterations to be 100 millisecond. We will show later that under this setup, the requirements on the time separation can be reasonably met.

A modified three-phase balanced IEEE 123-bus distribution test feeder from [24] (see Fig. 2 for the one-line diagram) is used for all numerical simulations. This balanced test feeder has a total active power load of 3000 kW, and a total reactive power load of 1575 kVar. DERs are added at buses 19, 26, 38, 49, 56, 64, 78, 89, 99. We assume each DER can output active power from 0 kW to 100 kW. Therefore, the maximum DER active power injections accounts for 30% of the nominal loads. To illustrate the impacts of reactive power control, we assume all DERs operate at unity power factor except DERs at buses 78 and 89, which are assumed to have enough reactive power capacity and maintain a constant voltage magnitude of 0.95 p.u. Yet, we would like to emphasize that the proposed algorithm is agnostic to the underlying reactive power control scheme and also works under other reactive power control schemes. In addition, to validate the effectiveness of the

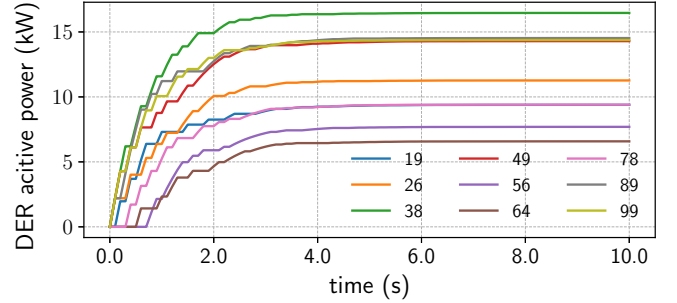


Fig. 3. DER active power injections for  $\beta_k = 0.02$  and  $y^* = -3000$  kW in Case I. (Legends indicate DER buses.)

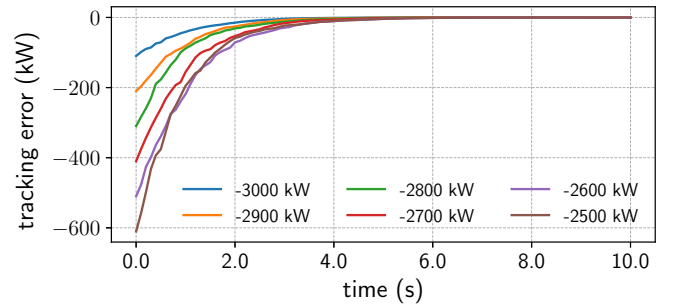


Fig. 4. Tracking error for  $\beta_k = 0.02$  under various tracking targets in Case I. (Legends indicate values of  $y^*$ .)

proposed algorithm under different operating conditions of the power distribution system, we assume there are some uncontrollable renewable energy resources in the power distribution system, which are modeled as negative loads. The underlying nonlinear power flow problem is solved using Matpower [25].

In all subsequent simulations, we set  $b_1 = 0.8$ ,  $\bar{b}_1 = 1.2$ , which are reasonable values for real power systems. Intuitively, these values indicate that the percentage of active power losses will be no larger than 20% of the total active power injections. Note that the exact value of  $b_2$  is not necessary. Under this simulation setting, as given in Theorem 1, the upper bound of the control step size is  $\frac{1}{nb_1^2} \approx 0.0694$ .

We note that comprehensive simulations including the two timescales can be done using data such as ones adopted in [26]. However, since the ODCP to be solved on the slow timescale is a standard problem, we will mainly focus on simulations for the fast timescale, where our major contributions lie.

### A. Case I

In this case, we assume the power distribution system is importing energy from the bulk power system with  $y = -3110$  kW. This corresponds to the situation where the uncontrollable renewable energy resources are not generating more active power than that needed by the loads. In addition, we set  $\hat{\phi}^0 = \mathbf{1}_n$  and  $\mathbf{u}^0 = \mathbf{0}_n$ , where  $\mathbf{1}_n \in \mathbb{R}^n$  is an all-ones vector.

1) *Tracking Performance During Estimation:* For  $y^* = -3000$  kW and a constant step size  $\beta_k = 0.02$ , the DER active

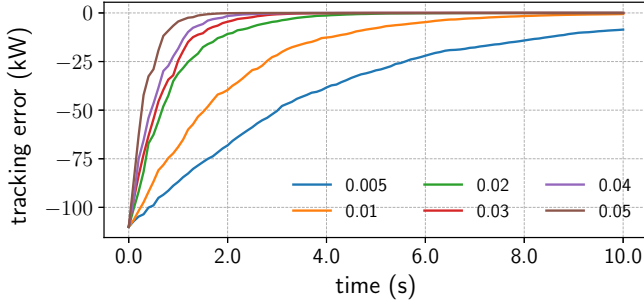


Fig. 5. Tracking error for  $y^* = -3000$  kW and various constant control step sizes in Case I. (Legends indicate values of  $\beta_k$ .)

TABLE I  
ESTIMATED SENSITIVITIES IN CASE I AFTER 60 ITERATIONS

bus	19	26	38	49	56
true	1.0394	1.0413	1.0426	1.0454	1.0467
estimate	1.0342	1.0390	1.0440	1.0468	1.0421
bus	64	78	89	99	
true	1.0702	1.0703	1.0749	1.0711	
estimate	1.0696	1.0697	1.0817	1.0702	

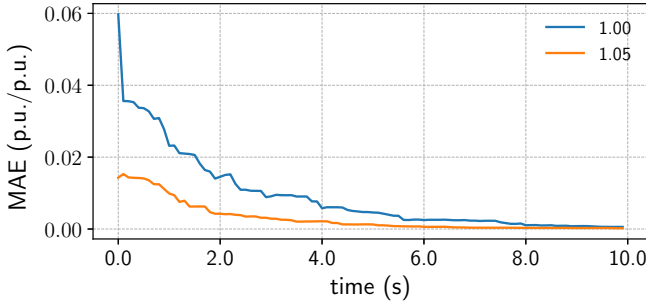


Fig. 6. MAE of estimation errors with  $\beta_k = 0.02$  in Case I. (Legends indicate the values of initial estimates.)

power injections are shown in Fig. 3. The non-smoothness in the active power profiles is caused by the random perturbation imposed on the control step. Also as shown in Fig. 4, the convergence rate of the tracking error is not affected by the tracking target, i.e., the total active power required from the bulk power system. The tracking error  $e[k]$  under various constant control step sizes is shown in Fig. 5. As expected, a larger step size will reduce the tracking error faster than a small step size.

2) *Estimation Accuracy:* With  $\beta_k = 0.02$  and  $y^* = -3000$  kW, true and estimated sensitivities are compared in Table I and the mean absolute error (MAE) of estimation errors, i.e.,  $\frac{\sum_{i=1}^n |\varepsilon_i[k]|}{N}$ , is plotted in Fig. 6. The estimated sensitivities are very close to their true values after 60 steps, which corresponds to 6 s. Note that  $\hat{\phi}^0$  has an important impact on the convergence of the sensitivity estimation algorithm. As can be seen from Fig. 6, when the initial values of the estimated sensitivities are set to 1.05, which is closer to their true values, it takes much less time to obtain a small estimation error.

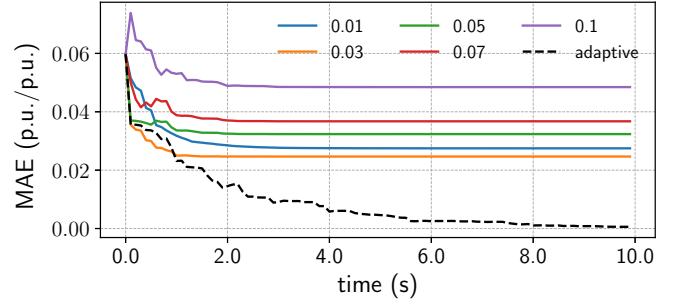


Fig. 7. MAE of estimation errors with  $\beta_k = 0.02$  under various estimation step sizes in Case I. (Legends indicate values of  $\alpha_k$ .)

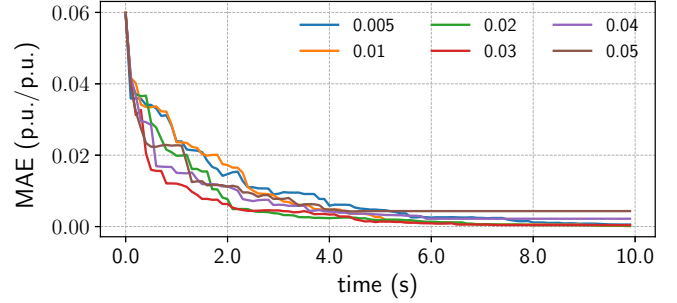


Fig. 8. MAE of estimation errors under various control step sizes in Case I. (Legends indicate values of  $\beta_k$ .)

While the estimation step size,  $\alpha_k$ , in the proposed algorithm is adaptive, we also investigate the case when  $\alpha_k$  is chosen to be constant. Figure 7 shows the MAE of estimation error under various constant estimation step sizes. As can be seen from Fig. 7, the MAE of estimation will converge to some non-zero constant under constant estimation step sizes.

The impact of the control step sizes on the estimation accuracy is also investigated. Figure 8 shows the MAE of estimation errors under various control step sizes. With a large control step size, the tracking error converges to 0 quickly, leading to a situation where the system cannot get sufficient excitation and consequently, the estimation errors cannot be further reduced.

## B. Case II

In this case, we assume the power distribution system is exporting energy to the bulk power system with  $y = 1000$  kW. This corresponds to the situation where the uncontrollable renewable energy resources are generating much more active power than that needed by the loads. We set  $\hat{\phi}^0 = \mathbf{1}_n$  and  $\mathbf{u}^0 = \mathbf{0}_n$ .

1) *Tracking Performance During Estimation:* Using a constant step size  $\beta_k = 0.02$ , the convergence rate of the tracking error under various tracking target is shown in Fig. 9. Similar to results in Case I, the convergence rate is not affected by the tracking target.

2) *Estimation Accuracy:* With  $\beta_k = 0.02$  and  $y^* = 1100$  kW, true and estimated sensitivities are compared in

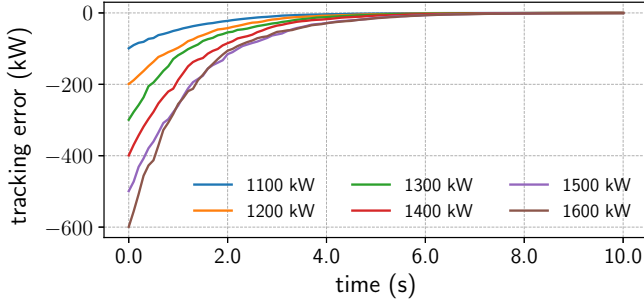


Fig. 9. Tracking error for  $\beta_k = 0.02$  under various tracking targets in Case II. (Legends indicate values of  $y^*$ .)

TABLE II  
ESTIMATED SENSITIVITIES IN CASE II AFTER 60 ITERATIONS

bus	19	26	38	49	56
true	0.9533	0.9526	0.9518	0.9509	0.9254
estimate	0.9588	0.9512	0.9497	0.9475	0.9285
bus	64	78	89	99	
true	0.8872	0.8477	0.8488	0.8700	
estimate	0.8854	0.8536	0.8396	0.8707	

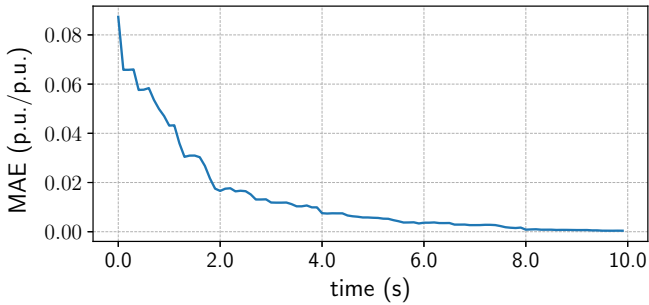


Fig. 10. MAE of estimation errors with  $\beta_k = 0.02$  in Case II.

Table II and the MAE of estimation errors is plotted in Fig. 10, respectively. Similar to the results in Case I, the estimated sensitivities are very close to their true values after 60 steps. This verifies that the proposed estimation algorithm can effectively estimate the sensitivities under different operating conditions of the power distribution system.

We note that the performance of the proposed estimation algorithm is independent of the number of DERs. To see this, we simulate a case where the DER at bus 99 gets disconnected and consequently, there are 8 DERs in the power distribution system. The sensitivities at these 8 DER buses can still be estimated effectively, as is shown in Fig. 11.

### C. Case III

In this case, we illustrate how the proposed framework handles line congestions. The setup is the same as Case II except that the tracking target is  $y^* = 1500$  kW. We set the capacity limit of line (55, 56) to 40 kW to create congestion. For simplicity, all other lines are assumed to have an infinite

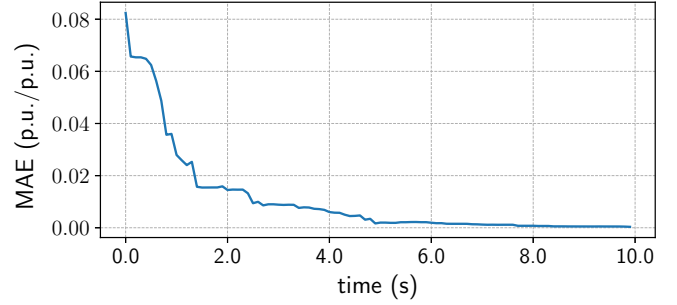


Fig. 11. MAE of estimation errors with  $\beta_k = 0.02$  in Case II with 8 DERs.

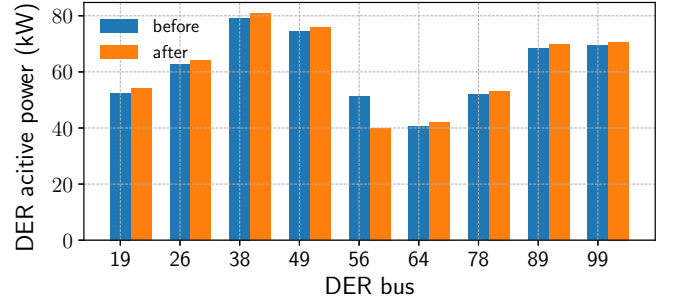


Fig. 12. DER active power set-point before and after solving the ODCP in Case III.

capacity; yet, we would like to emphasize that the proposed framework can handle multiple line congestions. The objective function in the ODCP is assumed to be  $c(\mathbf{p}^g) = \|\mathbf{p}^g - \tilde{\mathbf{p}}^g\|^2$ , where  $\tilde{\mathbf{p}}^g$  is the current DER active power injection vector as used in (9). Intuitively, this objective function will favor the solution with the least change in the DER active power injections.

The estimation algorithm is first ran to obtain an estimate of the sensitivity vector. After the estimation algorithm ends, the DER at bus 56 is generating 51.3 kW, which exceeds the capacity limit of line (55, 56). The ODCP is run afterwards to adjust the active power set-points of the DERs. Figure 12 shows the DER active power set-point before and after solving the ODCP. The DER at bus 56 is dispatched down to 40 kW, which conforms with the capacity limit of line (55, 56). All other DERs are dispatched up such that the active power exchanged between the distribution and bulk power systems still equals to 1500 kW. We note that line (55, 56) is overloaded for a short period during the estimation process but the is quickly restored to a normal loading level after the DER active power set-points are adjusted via the ODCP.

## VI. CONCLUDING REMARKS

In this paper, we proposed a data-driven coordination framework for DERs in a lossy power distribution system, the model of which is not completely known, to collectively provide some pre-specified amount of active power to a bulk power system at a minimum generating cost, while respecting distribution line capacity limits. The proposed framework consists of a LTV IO model, an estimator, and a controller.



We showed that using the estimation algorithm proposed in the framework, the estimated parameters converge to the true parameters a.s., and the total provided active power converges to the required amount during the estimation process. The data-driven nature of this framework makes it adaptive to various system operating conditions. We validated the effectiveness of the proposed framework through numerical simulations on a modified version of the IEEE 123-bus test feeder.

There are two potential directions for future work. The first one is to develop efficient estimation algorithms for scenarios in which the estimation process (with random perturbations) cannot be executed on a timescale that is much faster than the one on which the loads vary, and quantifies the impacts of load variability on the performance of the algorithm; this essentially relaxes Assumption 3. The second one is to extend the proposed estimation algorithms to scenarios where there are multiple connection buses between the distribution system and the bulk power system, and consequently, multiple outputs.

## APPENDIX

### A. Proof of Theorem 1

The convergence analysis of the control step during the estimation process relies on the following two lemmas.

**Lemma 1.** There exists  $\bar{\phi}[k]$  satisfying  $\mathbf{0}_n \leq \bar{\phi}[k] \leq \hat{\phi}[k]$ , such that (14) is equivalent to

$$\mathbf{u}[k] = \mathbf{u}[k-1] - \beta_k e[k-1] \mathbf{W}[k] \bar{\phi}[k].$$

Also,  $\bar{\phi}[k] = \mathbf{0}_n$  if and only if  $\mathbf{u}[k] = \underline{\mathbf{u}}$  or  $\mathbf{u}[k] = \bar{\mathbf{u}}$ . Furthermore, if  $\mathbf{u}[k] \neq \underline{\mathbf{u}}$  and  $\mathbf{u}[k] \neq \bar{\mathbf{u}}$ , there exists  $i \in \mathcal{N}^g$  such that  $\bar{\phi}_i[k] = \hat{\phi}_i[k] \in [\underline{b}_1, \bar{b}_1]$ .

*Proof.* If  $\mathbf{u}[k-1] - \beta_k e[k-1] \mathbf{W}[k] \hat{\phi}[k] \in \mathcal{U}$ , then we simply set  $\bar{\phi}[k] = \hat{\phi}[k]$ . Without loss of generality, first consider the case where the following holds for some  $i \in \mathcal{N}^g$ :

$$u_i[k-1] - \beta_k e[k-1] w_i[k] \hat{\phi}_i[k] > \bar{u}_i. \quad (\text{A.1})$$

Then,  $e[k-1] < 0$  and  $w_i[k] > 0$  since otherwise (A.1) cannot hold. Therefore,

$$u_i[k] = \mathbb{P}_{\mathcal{U}}(u_i[k-1] - \beta_k e[k-1] w_i[k] \hat{\phi}_i[k]) = \bar{u}_i. \quad (\text{A.2})$$

Let  $\bar{\phi}_i[k] = \frac{u_i[k-1] - \bar{u}_i}{\beta_k e[k-1] w_i[k]}$ ; by definition,  $\bar{\phi}_i[k] = 0$  if and only if  $u_i[k-1] = \bar{u}_i$ . Then, we have that:

$$u_i[k] = u_i[k-1] - \beta_k e[k-1] w_i[k] \bar{\phi}_i[k]. \quad (\text{A.3})$$

It follows from (A.1), (A.2), and (A.3) that

$$\beta_k e[k-1] \hat{\phi}_i[k] w_i[k] < \beta_k e[k-1] \bar{\phi}_i[k] w_i[k], \quad (\text{A.4})$$

which leads to  $0 \leq \bar{\phi}_i[k] < \hat{\phi}_i[k]$ . A similar argument can be used to for the case where  $u_i[k-1] - \beta_k e[k-1] w_i[k] \hat{\phi}_i[k] < \underline{u}_i$  and for some  $i \in \mathcal{N}^g$ .

If  $\mathbf{u}[k] \neq \underline{\mathbf{u}}$  and  $\mathbf{u}[k] \neq \bar{\mathbf{u}}$ , then there exists  $i \in \mathcal{N}^g$  such that  $\underline{u}_i < u_i[k] < \bar{u}_i$ , which implies

$$u_i[k] = u_i[k-1] - \beta_k e[k-1] w_i[k] \hat{\phi}_i[k]. \quad (\text{A.5})$$

Therefore,  $\bar{\phi}_i[k] = \hat{\phi}_i[k]$ . Consequently,  $\bar{\phi}_i[k] = \hat{\phi}_i[k] \in [\underline{b}_1, \bar{b}_1]$ . It can be easily seen that if  $\mathcal{U}$  is sufficiently large and no DER hits its capacity limits, then  $\bar{\phi}[k] = \hat{\phi}[k]$ .  $\square$

**Lemma 2.** Let  $X_k, k = 1, 2, \dots$ , be independently identically distributed (i.i.d.) random variables. Assume  $X_k > 0$  and  $\mathbb{E}[X_k] \in (0, 1)$ , where  $\mathbb{E}$  denotes expectation. Let  $Y_k = \prod_{i=1}^k X_i$ . Then,  $\lim_{k \rightarrow \infty} Y_k = 0$  a.s.

*Proof.* Note that  $Y_k = \exp\left\{\sum_{i=1}^k \log X_i\right\}$ . By the Strong Law of Large Numbers (see Proposition 2.15 in [27]), we have that

$$\lim_{k \rightarrow \infty} \sum_{i=1}^k \frac{1}{k} \log X_i = \mathbb{E}[\log X_1], \text{ a.s.} \quad (\text{A.6})$$

By Jensen's inequality (see Theorem 2.18 in [27]), we have that

$$\mathbb{E}[\log X_1] \leq \log \mathbb{E}[X_1] < 0. \quad (\text{A.7})$$

Therefore,

$$\lim_{k \rightarrow \infty} \sum_{i=1}^k k \frac{1}{k} \log X_i = -\infty, \text{ a.s.}, \quad (\text{A.8})$$

which leads to

$$\begin{aligned} \lim_{k \rightarrow \infty} Y_k &= \lim_{k \rightarrow \infty} \exp\left\{\sum_{i=1}^k k \frac{1}{k} \log X_i\right\} \\ &= \exp\left\{\lim_{k \rightarrow \infty} \sum_{i=1}^k k \frac{1}{k} \log X_i\right\} \\ &= 0, \text{ a.s.}; \end{aligned} \quad (\text{A.9})$$

this completes the proof.  $\square$

Using Lemma 1 and Lemma 2, we can prove the Theorem 1 as follows:

*Proof.* By (6), we have that

$$e[k] - e[k-1] = \phi[k]^\top \Delta \mathbf{u}[k]. \quad (\text{A.10})$$

By Lemma 1, we have that

$$\Delta \mathbf{u}[k] = -\beta_k e[k-1] \mathbf{W}[k] \bar{\phi}[k], \quad (\text{A.11})$$

where  $\mathbf{0}_n \leq \bar{\phi}[k] \leq \hat{\phi}[k]$ . Substituting (A.11) into (A.10) leads to

$$e[k] = (1 - \beta_k \phi[k]^\top \mathbf{W}[k] \bar{\phi}[k]) e[k-1]. \quad (\text{A.12})$$

Define  $\rho_k = 1 - \beta_k \phi[k]^\top \mathbf{W}[k] \bar{\phi}[k]$ , then

$$e[k] = e[0] \prod_{i=1}^k \rho_i. \quad (\text{A.13})$$

By Assumption 3,  $0 < \underline{b}_1 \leq \phi_i[k] \leq \bar{b}_1$ . In addition, it follows from Lemma 1 that  $0 \leq \bar{\phi}_i[k] \leq \hat{\phi}_i[k] \leq \bar{b}_1$ . Therefore,  $\phi[k]^\top \mathbf{W}[k] \bar{\phi}[k]$  can be bounded as follows:

$$0 \leq \phi[k]^\top \mathbf{W}[k] \bar{\phi}[k] = \sum_{i=1}^n w_i[k] \phi_i[k] \bar{\phi}_i[k] \leq n \bar{b}_1^2. \quad (\text{A.14})$$

Since  $\beta_k < \frac{1}{nb_1^2}$ , then all  $e[k]$  has the same sign for all  $k$  (positive if  $e[0] > 0$ , and negative otherwise). As a result, the entries of  $\Delta \mathbf{u}[k]$  always have the same sign by (A.11).

(a) If  $e[k] = 0$  for some  $k \in \mathbb{N}$ , then  $\Delta \mathbf{u}[k+1] = \mathbf{0}_n$ . The control and estimation algorithms will stop updating according to (12) and (14). In this case,  $\mathbf{u}[k]$  may equal to  $\underline{\mathbf{u}}$  or  $\bar{\mathbf{u}}$  or neither, and the system attains an equilibrium.

(b) Now suppose  $e[k] \neq 0, \forall k \in \mathbb{N}$ . Since the increments of  $\mathbf{u}$  always have the same sign, the entries of  $\mathbf{u}$  cannot hit their bounds in different directions, i.e., some hit their lower bounds while others hit their upper bounds.

(b.1) If  $\mathbf{u}[k] = \underline{\mathbf{u}}$  for some iteration  $k$ , then  $e[k] > 0, \forall k \in \mathbb{N}$ . By (14), we have that

$$\mathbf{u}[k+1] = \mathbf{P}_U(\underline{\mathbf{u}} - \beta_k e[k] \mathbf{W}[k+1] \hat{\phi}[k+1]) = \underline{\mathbf{u}}. \quad (\text{A.15})$$

Thus,  $\Delta \mathbf{u}[k+1] = \mathbf{0}_n$ , which leads to  $e[k+1] = e[k]$  by (A.10). Therefore,  $\mathbf{u}$  will equal to  $\underline{\mathbf{u}}$  and  $e[k'] = e[k] > 0$  for all  $k' > k$ .

Similarly, when  $\mathbf{u}[k] = \bar{\mathbf{u}}$ ,  $\mathbf{u}$  will be equal to  $\bar{\mathbf{u}}$ , and  $e$  will be equal to  $e[k] < 0$  in all future time intervals. The system attains an equilibrium in both cases.

(b.2) If  $\mathbf{u}[k] \neq \underline{\mathbf{u}}$  and  $\mathbf{u}[k] \neq \bar{\mathbf{u}}$ ,  $\forall k \in \mathbb{N}$ , by Lemma 1, there exists  $i \in \mathcal{N}^g$  such that  $\bar{\phi}_i[k] \in [b_1, \bar{b}_1]$ . Then,

$$\phi[k]^\top \mathbf{W}[k] \bar{\phi}[k] = \sum_{i=1}^n \phi_i[k] \bar{\phi}_i[k] w_i[k] \geq \bar{b}_1^2 w_i[k]. \quad (\text{A.16})$$

Thus, by using (A.14) and (A.16), it follows that  $\rho_k \in [1 - \beta_k n \bar{b}_1^2, 1 - \beta_k \bar{b}_1^2 w_i[k]]$ . Define  $\bar{\rho}_k = 1 - \epsilon w_i[k]$ , then  $\bar{\rho}_k$  equals to  $1 - \epsilon$  or 1, each with probability 0.5, and  $\mathbb{E}[\bar{\rho}_k] = 1 - \frac{\epsilon}{2} \in (0, 1)$ . Note that  $0 < \epsilon < \frac{b_1^2}{nb_1^2}$  implies  $\bar{\rho}_k > 0$ . By Lemma 2,

$$\lim_{k \rightarrow \infty} \prod_{i=1}^k \bar{\rho}_i = 0, \text{ a.s.} \quad (\text{A.17})$$

When  $\beta_k \in (\frac{\epsilon}{b_1^2}, \frac{1}{nb_1^2})$ ,  $0 \leq \rho_k \leq \bar{\rho}_k$ . Then, in an a.s. sense,

$$\lim_{k \rightarrow \infty} |e[k]| = |e[0]| \lim_{k \rightarrow \infty} \prod_{i=1}^k \rho_i \leq |e[0]| \lim_{k \rightarrow \infty} \prod_{i=1}^k \bar{\rho}_i = 0. \quad (\text{A.18})$$

Since  $|e[k]| \geq 0$ ,  $\lim_{k \rightarrow \infty} |e[k]| = 0$  a.s. In addition, by (A.11),  $\lim_{k \rightarrow \infty} \Delta \mathbf{u}[k] = \mathbf{0}_n$  a.s.  $\square$

*Remark 2.* If  $\mathcal{U}$  is sufficiently large and no DER hits the capacity limits, then  $\bar{\phi}[k] = \hat{\phi}[k]$  and  $\phi[k]^\top \mathbf{W}[k] \bar{\phi}[k] \geq \bar{b}_1^2 \sum_{i=1}^n w_i[k]$ . Following a similar argument as in part (b.2) in the proof of Theorem 1, we can show  $e[k]$  converges to 0 a.s. when  $\beta_k \in (\frac{\epsilon}{nb_1^2}, \frac{1}{nb_1^2})$ , where  $0 < \epsilon < \frac{b_1^2}{b_1^2}$ .

Following a similar argument, Corollary 1 can be proved as follows:

*Proof.* When the control update rule in (13) is used instead of the one in (14),

$$e[k] = (1 - \beta_k \phi[k]^\top \bar{\phi}[k]) e[k-1]. \quad (\text{A.19})$$

If  $\mathbf{u}[k] \neq \underline{\mathbf{u}}$  and  $\mathbf{u}[k] \neq \bar{\mathbf{u}}$ ,  $\phi[k]^\top \bar{\phi}[k] = \phi_i[k] \bar{\phi}_i[k] \geq \bar{b}_1^2$ . Define  $\rho_k = 1 - \beta_k \phi[k]^\top \bar{\phi}[k]$ . When  $\beta_k \in (\frac{\epsilon}{b_1^2}, \frac{1}{nb_1^2})$ ,  $\rho_k < 1 - \epsilon$ . Therefore,  $\left| \frac{e[k]}{e[k-1]} \right| = \rho_k < 1 - \epsilon$ .  $\square$

## B. Proof of Theorem 2

The convergence analysis of the estimation step uses some convergence results for  $\Delta \phi[k]$ , which are presented next.

**Lemma 3.** Let  $X_k, k = 1, 2, \dots$ , be i.i.d. random variables that take value 1 with probability 0.5, or some constant  $x \in (0, 1)$ , also with probability 0.5. Let  $Y_k = \prod_{i=1}^k X_i$  and  $Z = \sum_{i=1}^{\infty} Y_i$ . Then,  $Z$  is bounded a.s.

*Proof.* Let  $K$  denote the maximum number of 1's that appears continuously in the sequence  $\{X_k\}$ ; then, the sequence  $\{Y_k\}$  will have a new (smaller) value at most after  $K+1$  steps. We claim  $Z$  is unbounded only if  $K$  is infinite. Suppose  $X_j = x$ , and  $X_k = 1$  for  $k = j+1, \dots, j+m$ , then  $Y_j = Y_{j+1} = \dots = Y_{j+m}$  and  $\sum_{i=j}^{j+m} Y_i = (m+1)Y_j \leq (K+1)Y_j$ . Therefore,

$$Z = \sum_{i=1}^{\infty} Y_i \leq (K+1) \sum_{i=0}^{\infty} x^i = \frac{K+1}{1-x}. \quad (\text{A.20})$$

It follows from (A.20) that  $Z$  is unbounded only if  $K$  is infinite. However,  $\mathbb{P}\{M = \infty\} \leq \mathbb{P}\{X_{i+1} = \dots = X_{i+K} = 1, \text{ for some } i\} = \frac{1}{2^\infty} = 0$ , where  $\mathbb{P}$  denotes probability. Thus,  $Z$  is bounded a.s.  $\square$

**Lemma 4.** Using estimation update rule (12) and control update rule (14), with  $\beta_k \in (\frac{\epsilon}{nb_1^2}, \frac{1}{nb_1^2})$ , where  $0 < \epsilon < \frac{b_1^2}{b_1^2}$  is a given parameter, then

$$\lim_{k \rightarrow \infty} \|\Delta \phi[k]\| = 0, \text{ a.s.} \quad (\text{A.21})$$

and

$$\sum_{k=1}^{\infty} \|\Delta \phi[k]\| < \infty, \text{ a.s.} \quad (\text{A.22})$$

*Proof.* It follows from the proof of Theorem 1 that the entries of  $\Delta \mathbf{u}[k]$  always have the same sign. First consider the case where  $\Delta \mathbf{u}[k] \geq \mathbf{0}_n$  for all  $k \in \mathbb{N}$ . Note that  $\phi[k]^\top = \frac{\partial h}{\partial \mathbf{u}} \Big|_{\tilde{\mathbf{u}}[k]}$ , where  $\tilde{\mathbf{u}}[k] = a_k \mathbf{u}[k] + (1 - a_k) \mathbf{u}[k-1]$  with  $a_k \in [0, 1]$ , i.e.,  $\mathbf{u}[k-1] \leq \tilde{\mathbf{u}}[k] \leq \mathbf{u}[k]$ . Similarly,  $\phi[k-1]^\top = \frac{\partial h}{\partial \mathbf{u}} \Big|_{\tilde{\mathbf{u}}[k-1]}$ , where  $\mathbf{u}[k-2] \leq \tilde{\mathbf{u}}[k-1] \leq \mathbf{u}[k-1]$ . Thus, by Assumption 3, we have that

$$\begin{aligned} \|\Delta \phi[k]\| &\leq b_2 \|\tilde{\mathbf{u}}[k] - \tilde{\mathbf{u}}[k-1]\| \\ &\leq b_2 \|\mathbf{u}[k] - \mathbf{u}[k-2]\| \\ &= b_2 \|\Delta \mathbf{u}[k] + \Delta \mathbf{u}[k-1]\| \\ &\leq b_2 (\|\Delta \mathbf{u}[k]\| + \|\Delta \mathbf{u}[k-1]\|). \end{aligned} \quad (\text{A.23})$$

Since  $\lim_{k \rightarrow \infty} \|\Delta \mathbf{u}[k]\| = 0$  a.s. by Theorem 1, as a result,  $\lim_{k \rightarrow \infty} (\|\Delta \mathbf{u}[k]\| + \|\Delta \mathbf{u}[k-1]\|) = 0$  a.s., which gives

$$\lim_{k \rightarrow \infty} \|\Delta \phi[k]\| = 0, \text{ a.s.} \quad (\text{A.24})$$

Assume  $\mathbf{u}[k] = \mathbf{0}_n$  for all  $k < 0$ , then we have that

$$\begin{aligned}
\sum_{k=1}^{\infty} \|\Delta\phi[k]\| &\leq \sum_{k=1}^{\infty} b_2(\|\Delta\mathbf{u}[k]\| + \|\Delta\mathbf{u}[k-1]\|) \\
&\leq 2b_2 \sum_{k=0}^{\infty} \|\Delta\mathbf{u}[k]\| \\
&\leq 2b_2 \sum_{k=0}^{\infty} \|\beta_k \mathbf{W}[k] \hat{\phi}[k] e[k-1]\| \\
&\leq \frac{2b_2}{nb_1} \sqrt{nb_1} \sum_{k=0}^{\infty} |e[k-1]| \\
&= \frac{2b_2}{\sqrt{nb_1}} \sum_{k=-1}^{\infty} |e[k]|. \tag{A.25}
\end{aligned}$$

Recall that  $\bar{\rho}_k$  equals to  $1 - \epsilon$  or  $1$ , each with probability  $0.5$ , where  $\bar{\rho}_k$  is defined in the proof of Theorem 1. Therefore, by Lemma 3,  $\sum_{k=1}^{\infty} \prod_{i=1}^k \bar{\rho}_i$  is bounded a.s. When  $\beta_k \in (\frac{\epsilon}{b_1^2}, \frac{1}{nb_1})$ ,  $0 \leq \rho_k \leq \bar{\rho}_k$ , and

$$\sum_{k=0}^{\infty} |e[k]| = |e[0]|(1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \rho_i) \leq |e[0]|(1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \bar{\rho}_i). \tag{A.26}$$

As a result,  $\sum_{k=-1}^{\infty} |e[k]|$  is bounded a.s. since  $|e[-1]|$  is also bounded. The case where  $\Delta\mathbf{u}[k] \leq \mathbf{0}_n$  for all  $k \in \mathbb{N}$  can be proved similarly.  $\square$

The convergence analysis of the estimation step also relies on the following lemma (see Theorem 1 in [28]).

**Lemma 5.** Let  $X_k, Y_k, Z_k$ ,  $k = 1, 2, \dots$ , be non-negative variables in  $\mathbb{R}$  such that  $\sum_{k=0}^{\infty} Y_k < \infty$ , and  $X_{k+1} \leq X_k + Y_k - Z_k$ , then  $X_k$  converges and  $\sum_{k=0}^{\infty} Z_k < \infty$ .

Using Lemma 4 and Lemma 5, Theorem 2 can then be proved as follows:

*Proof.* Consider an arbitrary sample path. Without loss of generality, assume  $e[k] < 0$ , it follows from Theorem 1 that  $e[k] < 0, \forall k \in \mathbb{N}$ . Since  $\mathbf{u}[k] \in (\underline{\mathbf{u}}, \bar{\mathbf{u}}), \forall k \in \mathbb{N}$ , (14) becomes

$$\Delta\mathbf{u}[k] = -\beta_k e[k-1] \mathbf{W}[k] \hat{\phi}[k]. \tag{A.27}$$

It follows from (6) and (12) that

$$\hat{\phi}[k+1] = \mathcal{P}_B(\hat{\phi}[k] - \alpha_{k+1} \Delta\mathbf{u}[k] \Delta\mathbf{u}[k]^\top \boldsymbol{\varepsilon}[k]). \tag{A.28}$$

By definition, the estimation error at iteration  $k$  is

$$\boldsymbol{\varepsilon}[k+1] = \mathcal{P}_B(\hat{\phi}[k] - \alpha_{k+1} \Delta\mathbf{u}[k] \Delta\mathbf{u}[k]^\top \boldsymbol{\varepsilon}[k]) - \phi[k+1]. \tag{A.29}$$

Since  $\phi[k+1] = \mathcal{P}_B(\phi[k+1])$ , by the non-expansiveness of the projection operation (see Proposition 1.1.9 in [29]), then

$$\begin{aligned}
\|\boldsymbol{\varepsilon}[k+1]\| &\leq \|\boldsymbol{\varepsilon}[k] - \alpha_{k+1} \Delta\mathbf{u}[k] \Delta\mathbf{u}[k]^\top \boldsymbol{\varepsilon}[k] - \Delta\phi[k+1]\| \\
&\leq \|\boldsymbol{\varepsilon}[k] - \alpha_{k+1} \Delta\mathbf{u}[k] \Delta\mathbf{u}[k]^\top \boldsymbol{\varepsilon}[k]\| + \|\Delta\phi[k+1]\|. \tag{A.30}
\end{aligned}$$

Let  $g(\alpha_{k+1}) = \|\boldsymbol{\varepsilon}[k] - \alpha_{k+1} \Delta\mathbf{u}[k] \Delta\mathbf{u}[k]^\top \boldsymbol{\varepsilon}[k]\|^2$ ; then,  $g$  attains its minimum at  $\alpha_{k+1} = \frac{1}{\|\Delta\mathbf{u}[k]\|^2}$ , which is

$$\|\boldsymbol{\varepsilon}[k]\|^2 - (\boldsymbol{\varepsilon}[k]^\top \frac{\Delta\mathbf{u}[k]}{\|\Delta\mathbf{u}[k]\|})^2 = \|\boldsymbol{\varepsilon}[k]\|^2 - (\boldsymbol{\varepsilon}[k]^\top \frac{\mathbf{W}[k] \hat{\phi}[k]}{\|\mathbf{W}[k] \hat{\phi}[k]\|})^2. \tag{A.31}$$

Define  $\cos \theta_k = \frac{\boldsymbol{\varepsilon}[k]^\top \mathbf{W}[k] \hat{\phi}[k]}{\|\boldsymbol{\varepsilon}[k]\| \|\mathbf{W}[k] \hat{\phi}[k]\|}$ . Consequently,  $g(\alpha_{k+1}) = (1 - \sin^2 \theta_k) \|\boldsymbol{\varepsilon}[k]\|^2$ , and

$$\|\boldsymbol{\varepsilon}[k+1]\| \leq |\sin \theta_k| \|\boldsymbol{\varepsilon}[k]\| + \|\Delta\phi[k+1]\|. \tag{A.32}$$

Let  $X_k = \|\boldsymbol{\varepsilon}[k]\|$ ,  $Y_k = \|\Delta\phi[k+1]\|$ , and  $Z_k = (1 - |\sin \theta_k|) \|\boldsymbol{\varepsilon}[k]\|$ . Then,  $X_{k+1} \leq X_k + Y_k - Z_k$ . Also,  $\sum_{k=0}^{\infty} Y_k = \sum_{k=1}^{\infty} \|\Delta\phi[k]\| < \infty$  by Lemma 4. Therefore, by Lemma 5,  $\|\boldsymbol{\varepsilon}[k]\|$  converges, and  $\sum_{k=1}^{\infty} (1 - |\sin \theta_k|) \|\boldsymbol{\varepsilon}[k]\| < \infty$ , which further implies  $\lim_{k \rightarrow \infty} (1 - |\sin \theta_k|) \|\boldsymbol{\varepsilon}[k]\| = 0$ . Let  $\varepsilon^*$  denote the limit of  $\|\boldsymbol{\varepsilon}[k]\|$ ; then,

$$\begin{aligned}
\lim_{k \rightarrow \infty} |\sin \theta_k| \|\boldsymbol{\varepsilon}[k]\| &= \lim_{k \rightarrow \infty} (|\sin \theta_k| - 1) \|\boldsymbol{\varepsilon}[k]\| + \lim_{k \rightarrow \infty} \|\boldsymbol{\varepsilon}[k]\| \\
&= \varepsilon^*. \tag{A.33}
\end{aligned}$$

Next, we show  $\varepsilon^* = 0$  by contradiction. Assume  $\varepsilon^* > 0$ . Since both  $\|\boldsymbol{\varepsilon}[k]\|$  and  $|\sin \theta_k| \|\boldsymbol{\varepsilon}[k]\|$  converges to  $\varepsilon^*$ ,

$$\lim_{k \rightarrow \infty} |\sin \theta_k| = \frac{\lim_{k \rightarrow \infty} |\sin \theta_k| \|\boldsymbol{\varepsilon}[k]\|}{\lim_{k \rightarrow \infty} \|\boldsymbol{\varepsilon}[k]\|} = 1, \tag{A.34}$$

which implies  $|\cos \theta_k|$  converges to  $0$ . Since  $\|\boldsymbol{\varepsilon}[k]\|$  and  $\|\mathbf{W}[k] \hat{\phi}[k]\|$  are bounded, then  $|\boldsymbol{\varepsilon}[k]^\top \mathbf{W}[k] \hat{\phi}[k]|$  converges to  $0$ . Define  $\mathbf{E}_i[k] = \{w_j[k] = 1 \text{ if } j = i, w_j[k] = 0 \text{ otherwise}\}$ ; then  $\mathbb{P}\{\mathbf{E}_i[k]\} = \frac{1}{2^n}$ . Consequently,  $\sum_{k=1}^{\infty} \mathbb{P}\{\mathbf{E}_i[k]\} = \infty$ . Also note that  $\mathbf{E}_i[k]$ ,  $k \in \mathbb{N}$ , are independent. By the Borel-Cantelli Lemma (see Lemma 1.3 in [27]),  $\mathbb{P}\{\mathbf{E}_i[k] \text{ infinitely often}\} = 1$ ; therefore, there are infinitely many time instants that  $w_i[k] = 1$  and  $w_j[k] = 0$  for all  $j \neq i$ . Let  $\mathcal{K}_i$  denote the set of such time instants. Then  $|\boldsymbol{\varepsilon}[k]^\top \mathbf{W}[k] \hat{\phi}[k]| = |\varepsilon_i[k] \hat{\phi}_i[k]|$  for  $k \in \mathcal{K}_i$ . The sequence  $\{|\varepsilon_i[k] \hat{\phi}_i[k]|, k \in \mathcal{K}_i\}$  is a subsequence of  $\{|\boldsymbol{\varepsilon}[k]^\top \mathbf{W}[k] \hat{\phi}[k]|\}$ ; therefore, it also converges to  $0$ . Note that  $\hat{\phi}[k] > 0$ ; thus,  $\varepsilon_i[k]$  converges to  $0$ . Since  $i$  is arbitrary, we conclude that  $\|\boldsymbol{\varepsilon}[k]\|$  converges to  $0$ , which implies  $\varepsilon^* = 0$ , contradiction. Since this result holds for all sample paths, then we conclude that  $\|\boldsymbol{\varepsilon}[k]\|$  converges to  $0$  a.s.  $\square$

## REFERENCES

- [1] J. Driesen and F. Katiraei, "Design for distributed energy resources," *IEEE Power Energy Mag.*, vol. 6, no. 3, pp. 30–40, May 2008.
- [2] A. Kulmala, S. Repo, and P. Järventausta, "Coordinated voltage control in distribution networks including several distributed energy resources," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 2010–2020, July 2014.
- [3] H. Xu, A. D. Domínguez-García, and P. W. Sauer, "A data-driven voltage control framework for power distribution systems," in *Proc. of IEEE PES General Meeting*, Portland, OR, Aug. 2018, pp. 1–5.
- [4] K. Zhang, W. Shi, H. Zhu, E. Dall'Anese, and T. Basar, "Dynamic power distribution system management with a locally connected communication network," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 673–687, Aug 2018.
- [5] A. D. Domínguez-García and C. N. Hadjicostis, "Coordination and control of distributed energy resources for provision of ancillary services," in *Proc. of IEEE International Conference on Smart Grid Communications*, Oct. 2010, pp. 537–542.
- [6] E. Dall'Anese, S. Guggilam, A. Simonetto, Y. C. Chen, and S. V. Dhople, "Optimal regulation of virtual power plants," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1868–1881, March 2018.

- [7] H. Xu, A. D. Domínguez-García, and P. W. Sauer, "Adaptive coordination of distributed energy resources in lossy power distribution systems," in *Proc. of IEEE PES General Meeting*, Portland, OR, Aug. 2018, pp. 1–5.
- [8] O. Ardakanian, V. W. Wong, R. Dobbe, S. H. Low, A. von Meier, C. Tomlin, and Y. Yuan, "On identification of distribution grids," *arXiv preprint arXiv:1711.01526*, 2017.
- [9] Y. C. Chen, A. D. Domínguez-García, and P. W. Sauer, "Measurement-based estimation of linear sensitivity distribution factors and applications," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1372–1382, May 2014.
- [10] Y. C. Chen, J. Wang, A. D. Domínguez-García, and P. W. Sauer, "Measurement-based estimation of the power flow jacobian matrix," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2507–2515, Sept. 2016.
- [11] Z. Hou and S. Jin, "Data-driven model-free adaptive control for a class of mimo nonlinear discrete-time systems," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2173–2188, Dec. 2011.
- [12] J. Zhang, X. Zheng, Z. Wang, L. Guan, and C. Y. Chung, "Power system sensitivity identification – inherent system properties and data quality," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2756–2766, July 2017.
- [13] Y. C. Chen, A. D. Domínguez-García, and P. W. Sauer, "A sparse representation approach to online estimation of power system distribution factors," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1727–1738, July 2015.
- [14] K. E. V. Horn, A. D. Domínguez-García, and P. W. Sauer, "Measurement-based real-time security-constrained economic dispatch," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3548–3560, Sept. 2016.
- [15] C. Lu, Y. Zhao, K. Men, L. Tu, and Y. Han, "Wide-area power system stabiliser based on model-free adaptive control," *IET Control Theory Appl.*, vol. 9, no. 13, pp. 1996–2007, 2015.
- [16] J. Zhang, C. Y. Chung, and Y. Han, "Online damping ratio prediction using locally weighted linear regression," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1954–1962, May 2016.
- [17] Z. Hou, R. Chi, and H. Gao, "An overview of dynamic-linearization-based data-driven control and applications," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4076–4090, May 2017.
- [18] J. Zhang, Z. Wang, X. Zheng, L. Guan, and C. Y. Chung, "Locally weighted ridge regression for power system online sensitivity identification considering data collinearity," *IEEE Trans. Power Syst.*, vol. PP, no. 99, pp. 1–1, 2017.
- [19] J. Zhang, C. Y. Chung, and L. Guan, "Noise effect and noise-assisted ensemble regression in power system online sensitivity identification," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2302–2310, Oct. 2017.
- [20] J. Zhang and H. Xu, "Microperturbation method for power system online model identification," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1055–1063, June 2016.
- [21] J. Zhang and H. Xu, "Online identification of power system equivalent inertia constant," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8098–8107, Oct 2017.
- [22] A. Iusem, "On the convergence properties of the projected gradient method for convex optimization," *Computational & Applied Mathematics*, vol. 22, no. 1, pp. 37–52, 2003.
- [23] O. Ajala, M. Almeida, P. W. I. Celanovic, Sauer, and A. D. Domínguez-García, "A hierarchy of models for microgrids with grid-feeding inverters," in *Proc. of the IREP Bulk Power System Dynamics and Control Symposium*, Aug. 2017.
- [24] IEEE distribution test feeders. [Online]. Available: <https://ewh.ieee.org/soc/pes/dsacom/testfeeders/>
- [25] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
- [26] B. A. Robbins and A. D. Domínguez-García, "Optimal reactive power dispatch for voltage regulation in unbalanced distribution systems," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2903–2913, 2016.
- [27] B. Hajek, *Random processes for engineers*. Cambridge University Press, 2015.
- [28] H. Robbins and D. Siegmund, "A convergence theorem for non negative almost supermartingales and some applications," in *Herbert Robbins Selected Papers*. Springer, 1985, pp. 111–135.
- [29] D. P. Bertsekas, *Convex optimization theory*. Athena Scientific Belmont, 2009.

**Hanchen Xu** (S'12) received the B.Eng. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2012 and 2014, respectively, and the M.S. degree in applied mathematics from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2017, where he is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering. His current research interests include control, optimization, reinforcement learning, with applications to power systems and electricity markets.

**Alejandro D. Domínguez-García** (S'02, M'07) received the degree of electrical engineering from the University of Oviedo (Spain) in 2001 and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 2007.

He is Professor with the Department of Electrical and Computer Engineering (ECE), and Research Professor with the Coordinated Science Laboratory and the Information Trust Institute, all at the University of Illinois at Urbana-Champaign. He is affiliated with the ECE Power and Energy Systems area, and has been a Grainger Associate since August 2011.

His research interests are in the areas of system reliability theory and control, and their applications to electric power systems, power electronics, and embedded electronic systems for safety-critical/fault-tolerant aircraft, aerospace, and automotive applications.

Dr. Domínguez-García received the NSF CAREER Award in 2010, and the Young Engineer Award from the IEEE Power and Energy Society in 2012. In 2014, he was invited by the National Academy of Engineering to attend the US Frontiers of Engineering Symposium, and was selected by the University of Illinois at Urbana-Champaign Provost to receive a Distinguished Promotion Award. In 2015, he received the U of I College of Engineering Dean's Award for Excellence in Research.

He is currently an editor for the IEEE Transactions on Control of Network Systems; he also served as an editor of the IEEE Transactions on Power Systems and IEEE Power Engineering Letters from 2011 to 2017.

**Peter W. Sauer** (S'73, M'77, SM'82, F'93, LF'12) obtained his Bachelor of Science degree in electrical engineering from the University of Missouri at Rolla in 1969. From 1969 to 1973, he was the electrical engineer on a design assistance team for the Tactical Air Command at Langley Air Force Base, Virginia. He obtained the Master of Science and Ph.D. degrees in Electrical Engineering from Purdue University in 1974 and 1977 respectively. From August 1991 to August 1992 he served as the Program Director for Power Systems in the Electrical and Communication Systems Division of the National Science Foundation in Washington D.C. He is a cofounder of the Power Systems Engineering Research Center (PSERC) and the PowerWorld Corporation. He is a registered Professional Engineer in Virginia and Illinois, a Fellow of the IEEE, and a member of the U.S. National Academy of Engineering. He is currently the Grainger Chair Professor of Electrical Engineering at Illinois. Additional information can be found at: <https://ece.illinois.edu/directory/profile/psauer>.