

A Feedback-Based Regularized Primal-Dual Gradient Method for Time-Varying Nonconvex Optimization

Yujie Tang, Emiliano Dall’Anese, Andrey Bernstein, and Steven H. Low

Abstract—This paper considers time-varying nonconvex optimization problems, utilized to model optimal operational trajectories of systems governed by possibly nonlinear physical or logical models. Algorithms for tracking a Karush-Kuhn-Tucker point are synthesized, based on a regularized primal-dual gradient method. In particular, the paper proposes a feedback-based primal-dual gradient algorithm, where analytical models for system state or constraints are replaced with actual measurements. When cost and constraint functions are twice continuously differentiable, conditions for the proposed algorithms to have bounded tracking error are derived, and a discussion of their practical implications is provided. Illustrative numerical simulations are presented for an application in power systems.

I. INTRODUCTION

This paper considers a *time-varying* [1], [2], [3] nonconvex optimization problem of the form

$$\begin{aligned} \min_{x \in \mathcal{X}(t)} \quad & c(x; t) \\ \text{s.t.} \quad & f(x; t) \leq 0 \end{aligned} \quad (1)$$

where $t \in [0, T]$ for some T , is a temporal index; the functions $c(\cdot; t) : \mathbb{R}^n \rightarrow \mathbb{R}$ and $f(\cdot; t) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are possibly nonconvex; and the set $\mathcal{X}(t) \subset \mathbb{R}^n$ is convex and compact for all t . The optimization problem (1) can be associated with systems governed by possibly nonlinear physical or logical models. The functions c and f are time-varying, and they capture performance objectives and constraints that evolve over time. Accordingly, denoting by $x^*(t)$ an optimal solution of (1) at time t , the optimization model (1) leads to a continuous-time optimal *trajectory*. Suppose further that (some entries of) the vector-valued function $f(x; t)$ can be expressed as $f(x; t) = h(y(x; t); t)$, where

$$y(x; t) : \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^k \quad (2)$$

is a function modeling observable states or outputs of the underlying physical or logical systems, and $h(\cdot; t) : \mathbb{R}^k \rightarrow \mathbb{R}^m$ is a given function modeling constraints imposed on the physical quantities $y(x; t)$ (application-specific examples will be provided later in the paper). Given (2) and the time-varying optimization model (1), the problem addressed in this paper pertains to the development of algorithms that enable tracking of the optimal trajectory $(x^*(t))_{t \in [0, T]}$.

Y. Tang and S. H. Low are with the California Institute of Technology, Pasadena, CA. E. Dall’Anese is with the University of Colorado Boulder, Boulder, CO. A. Bernstein is with the National Renewable Energy Laboratory, Golden, CO. This work was supported by the Advanced Research Projects Agency-Energy (ARPA-E) under the Network Optimized Distributed Energy Systems (NODES) program.

For a system where the functional evaluation of the map (2) does not depend on time-varying exogenous inputs that need to be exchanged via a communication network, problem (1) might be solved in a centralized setting based on continuous-time platforms; see e.g., [4], [5], [6], [7] for convex counterparts of (2) and [1] for time-varying nonconvex settings. However, this paper focuses on discrete-time updates that allow one to account for (i) computational delays in the algorithmic updates, (ii) delays due to collection and communication of measurements and problem inputs, and (iii) the case where the update of $x(t)$ leads to control actions that are to be implemented on digital control units. To this end, consider the following sampled version of (1):

$$\begin{aligned} \min_{x \in \mathcal{X}_t} \quad & c_t(x) \\ \text{s.t.} \quad & f_t(x) \leq 0 \end{aligned} \quad (3)$$

where $t \in \mathcal{T}$ now denotes a discrete time index, and we assume that $\mathcal{T} = \{0, \dots, T\}$ for some $T \in \mathbb{N}$. Accordingly, the sampled optimal trajectory is denoted as $(x_t^*)_{t \in \mathcal{T}}$.

A *batch* solution of (3) would involve an iterative algorithm where iterations are sequentially performed until convergence [8] within each time interval $[t, t+1]$. However, a batch solution might not be appropriate in a time-varying setting, since underlying communication and computational complexity requirements may prevent the algorithm from converging within a time interval $[t, t+1]$; this is especially the case when the sampling time is chosen small enough to fully model fast-varying costs and constraints and closely approximate the continuous-time optimal trajectory. In lieu of a batch solution, this paper leverages *online* (i.e., running) primal-dual gradient methods to track the sequence of optimal points $(x_t^*)_{t \in \mathcal{T}}$ [2]. Going beyond this setting, and aligned with the works on time-varying optimization [9], [10] and subsequently [11], [12], [13], this paper further considers a *feedback-based online* implementation of primal-dual-type gradient methods where the functional evaluation of $y(x; t)$ and $h(y(x; t); t)$ is replaced with *measurements* (collected at each time $t \in \mathcal{T}$) in the implementation of the dual gradient steps. This modification leads to a closed-loop online optimization procedure [9], [10], [11], [12], [13], [14], [15], [16], [17] that allows one to: (i) bypass the need for an explicit functional form of $y(x; t)$, which may be difficult to obtain accurately in many physical systems; (ii) cope with model mismatches in the formulation of the problem constraints; and, (iii) in some applications, avoid pervasive measurements of inputs $w(t)$ that would need to be collected when $y(x; t)$ and/or $h(y(x; t); t)$ depend on $w(t)$.

The proposed technical approach leverages a regularized primal-dual method, where the regularization comes in the form of a strongly concave term in the dual variables [18]; see also [19] and [20]. The strongly concave regularization term plays a critical role in establishing Q-linear convergence of the proposed algorithm; however, as an artifact of this regularization, the iterates of the algorithm are driven to an approximate Karush-Kuhn-Tucker (KKT) point [21]. When $y(x; t)$ and $h(y(x; t); t)$ are replaced with measurements, and assuming the functions are twice continuously differentiable, we provide analytical results in terms of tracking a KKT point of the nonconvex problem (3).

It is worth pointing out that the development of feedback-based online optimization methods has been, so far, driven by power systems application; see, for example, [9], [10], [11], [12], [13], [14], [15], [16], [17], [22], [23] and pertinent references therein. However, the framework is generally applicable to a number of settings where the objective is to drive the operation of physical and logical systems as well as networked systems to optimal operating points in real time [24]. From a theoretical standpoint, the present paper provides a first extension of the frameworks of [9], [13], [16] from time-varying convex to time-varying nonconvex settings; it offers contributions over e.g., [11], [12], [17] by providing analytical results in terms of tracking of a time-varying KKT point; and it expands some analysis of [20] to the cases of time-varying nonconvex problems as well as feedback-based implementations. A number of works in the domain of *batch* optimization have utilized regularized primal-dual methods for nonconvex problems; see, e.g., [25], [26], [27] and references therein, with [25], [26] primarily focusing on linearly-constrained problems. The present paper departs from [25], [26], [27] by focusing on feedback-based online methods; extending some of the techniques leveraged in [25], [26], [27] to online settings is a subject of an ongoing investigation.

II. REGULARIZED PRIMAL-DUAL GRADIENT METHOD

A. Notation

For $x \in \mathbb{R}^n$, we denote the 2-norm of x by $\|x\| := \sqrt{x^T x}$. For a matrix A , its operator norm is denoted by $\|A\| := \sup_{x: \|x\|=1} \|Ax\|$. The identity matrix is denoted by I .

For a convex set $C \subseteq \mathbb{R}^n$, its normal cone at $x \in C$, defined by $\{y \in \mathbb{R}^n : y^T(z - x) \leq 0, \forall z \in C\}$, is denoted by $N_C(x)$. The projection operator onto C is denoted by \mathcal{P}_C .

For a continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, its Jacobian matrix evaluated at x is denoted by $J_f(x)$, and $f(x) \leq 0$ means each of the m components of $f(x)$ is less than or equal to zero. For a function $L(x, \lambda)$ that is twice continuously differentiable in x , we denote its Hessian with respect to x by $\nabla_{xx}^2 L(x, \lambda)$.

B. Time-Varying Optimization

Consider the time-varying optimization problem (3) and assume that the functions $c_t : \mathbb{R}^n \rightarrow \mathbb{R}$ and $f_t : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are twice continuously differentiable. For the set $\mathcal{X}_t \subset \mathbb{R}^n$,

we assume it is compact and convex and the projection onto it can be efficiently computed for each t .

We assume that the time-varying optimization problem (3) is feasible for any $t \in \mathcal{T}$, and consequently there exists a sampled trajectory $(x_t^*)_{t \in \mathcal{T}}$ of (local) optimal solutions¹. Suppose that the following constraint qualification condition holds: There is no $\lambda \in \mathbb{R}_+^m, \lambda \neq 0$ such that

$$x^T f_t(x_t^*) = 0 \text{ and } -J_{f_t}(x_t^*)^T \lambda \in N_{\mathcal{X}_t}(x_t^*).$$

Then by [28, Theorem 2A.9], there exists a Lagrange multiplier λ_t^* and we have the KKT conditions

$$(x_t^*, \lambda_t^*) \in \mathcal{X}_t \times \mathbb{R}_+^m, \quad (4a)$$

$$\nabla c_t(x_t^*) + J_{f_t}(x_t^*)^T \lambda_t^* \in -N_{\mathcal{X}_t}(x_t^*), \quad (4b)$$

$$f_t(x_t^*) \leq 0, \quad \lambda_t^{*T} f_t(x_t^*) = 0. \quad (4c)$$

We denote $z_t^* := (x_t^*, \lambda_t^*)$ and define the Lagrangian as $L_t(x, \lambda) := c_t(x) + \lambda^T f_t(x)$.

Many practical problems with time-varying systems or environments can be modeled by this formulation. In order to achieve optimal operation, we need to update and solve (3) repeatedly. As mentioned before, a batch solution might not be appropriate in a time-varying setting, and we consider online algorithms that seek suboptimal solutions.

C. Tracking Algorithm

In order to design an efficient algorithm that provides suboptimal solution to (3), let's suppose that we have obtained a suboptimal solution for time $t-1$ denoted by \hat{x}_{t-1} , and consider the following approximation of (3) at time t :

$$\begin{aligned} \min_{x \in \mathcal{X}_t} \quad & \nabla c_t(\hat{x}_{t-1})^T (x - \hat{x}_{t-1}) + \frac{\nu}{2} \|x - \hat{x}_{t-1}\|^2 \\ \text{s.t.} \quad & f_t(\hat{x}_{t-1}) + J_{f_t}(\hat{x}_{t-1})(x - \hat{x}_{t-1}) \leq 0 \end{aligned} \quad (5)$$

where ν is a positive constant. Here the objective function is replaced by a quadratic approximation, and the nonconvex constraint is replaced by its linearization around \hat{x}_{t-1} . For computational simplicity, the quadratic approximation uses a scalar matrix νI as an approximation of $\nabla_{xx}^2 L_t$; we discuss a possible generalization in Section V.

It is known [8] that a primal-dual pair is an optimal solution pair to (5) if and only if it is a solution to the following saddle point problem

$$\begin{aligned} \max_{\lambda \in \mathbb{R}_+^m} \min_{x \in \mathcal{X}_t} \quad & \nabla c_t(\hat{x}_{t-1})^T (x - \hat{x}_{t-1}) + \frac{\nu}{2} \|x - \hat{x}_{t-1}\|^2 \\ & + \lambda^T (f_t(\hat{x}_{t-1}) + J_{f_t}(\hat{x}_{t-1})(x - \hat{x}_{t-1})) \end{aligned}$$

We then add a regularization term of the Lagrange multiplier λ so that the objective function of the saddle point problem becomes strongly concave in λ :

$$\begin{aligned} \max_{\lambda \in \mathbb{R}_+^m} \min_{x \in \mathcal{X}_t} \quad & \nabla c_t(\hat{x}_{t-1})^T (x - \hat{x}_{t-1}) + \frac{\nu}{2} \|x - \hat{x}_{t-1}\|^2 \\ & + \lambda^T (f_t(\hat{x}_{t-1}) + J_{f_t}(\hat{x}_{t-1})(x - \hat{x}_{t-1})) \\ & - \frac{\epsilon}{2} \|\lambda\|^2 \end{aligned} \quad (6)$$

¹There can be multiple trajectories of local optimal solutions. In the following we only consider one of the optimal trajectories that is chosen arbitrarily.

where $\epsilon > 0$ is a constant serving as the regularization parameter. Finally, we apply a single iteration of the primal-dual gradient method [29] to the regularized saddle-point problem (6), which results in the following update rule:

$$\hat{x}_t = \mathcal{P}_{\mathcal{X}_t} \left(\hat{x}_{t-1} - \tau \nu^{-1} (\nabla c_t(\hat{x}_{t-1}) + J_{f_t}(\hat{x}_{t-1})^T \hat{\lambda}_{t-1}) \right), \quad (7a)$$

$$\hat{\lambda}_t = \mathcal{P}_{\mathbb{R}_+^m} \left((1 - \tau) \hat{\lambda}_{t-1} + \tau \epsilon^{-1} f_t(\hat{x}_{t-1}) \right), \quad (7b)$$

where $\tau \in (0, 1]$ is a constant representing the step size. Note that besides the step size τ , we also scale the primal and dual gradients by ν^{-1} and ϵ^{-1} respectively, which can be interpreted as incorporating approximate second-order information. $(\hat{x}_t, \hat{\lambda}_t)$ will then be employed as the suboptimal primal-dual pair to (3); we also denote $\hat{z}_t := (\hat{x}_t, \hat{\lambda}_t)$. The method presented in (7) for solving the time-varying optimization problem (3) will be called the *regularized primal-dual gradient method*. It is expected that, as long as the optimal pair $z_t^* = (x_t^*, \lambda_t^*)$ does not change too drastically with time and the problem (3) is sufficiently “regular”, if we start from some \hat{z}_0 that is sufficiently close to z_0^* , then the suboptimal pair \hat{z}_t should remain close to z_t^* for all $t \geq 1$. Section III will present theoretical results that make this intuition precise.

D. Feedback-based Algorithm

Let us consider a special case of the problem (3). Suppose we have a physical system whose input-output behavior is given by the time-varying map

$$y_t : \mathbb{R}^n \rightarrow \mathbb{R}^k, \quad t \in \mathcal{T} \quad (8)$$

which is twice continuously differentiable for each t ; in other words, at time t , if we set the input of the system to be $x \in \mathbb{R}^n$, then the output of the system will be $y_t(x) \in \mathbb{R}^k$. As before, we assume that the feasible set of input variables is a compact and convex set \mathcal{X}_t , and the output variables need to be constrained for operational purposes by

$$h_t(y_t(x)) \leq 0, \quad t \in \mathcal{T} \quad (9)$$

for some twice continuously differentiable $h_t : \mathbb{R}^k \rightarrow \mathbb{R}^m$ for each t . The operational cost of the system at time t is given by $c_t(x)$, as before. It can be seen that this setting can be fit into the framework of time-varying optimization (3), whereby the function f_t has the form $f_t = h_t \circ y_t$. Further, in this setting, we assume that

- (i) The function h_t is known and is *affine* in y , namely

$$h_t(y) = H_t y + \beta_t \quad (10)$$

for some $H_t \in \mathbb{R}^{m \times k}$ and $\beta_t \in \mathbb{R}^m$.

- (ii) The Jacobian matrix $J_{y_t}(x)$ is known; $J_{y_t}(x)$ represents the sensitivity of the output of the physical system to changes in the input variable x .
- (iii) The explicit functional form $y_t(x)$ is either unknown, or hard to evaluate.

To alleviate difficulty (iii) above, a *closed-loop optimization* approach is used, whereby instead of using $y_t(x)$ explicitly in

the algorithm, a measurement thereof is leveraged. Observe that

$$J_{f_t}(x) = J_{h_t}(y_t(x)) J_{y_t}(x) = H_t J_{y_t}(x).$$

Let \tilde{x}_t denote the input to the system (as computed by the algorithm below) at time t , and let \tilde{y}_t be the measurement of $y_t(\tilde{x}_{t-1})$. That is, \tilde{y}_t is the observed output of the system *right before* time t . Then, with feedback measurement, the algorithm (7) is modified as follows:

$$\begin{aligned} \tilde{x}_t &= \mathcal{P}_{\mathcal{X}_t} \left(\tilde{x}_{t-1} - \tau \nu^{-1} (\nabla c_t(\tilde{x}_{t-1}) \right. \\ &\quad \left. + [H_t J_{y_t}(\tilde{x}_{t-1})]^T \tilde{\lambda}_{t-1}) \right), \end{aligned} \quad (11a)$$

$$\tilde{\lambda}_t = \mathcal{P}_{\mathbb{R}_+^m} \left((1 - \tau) \tilde{\lambda}_{t-1} + \tau \epsilon^{-1} h_t(\tilde{y}_t) \right). \quad (11b)$$

Remark 1: For brevity, this paper focuses on the case of an affine map h_t . In this case, the feedback appears explicitly only in the dual update (11b). The analysis of the general case is left for future work.

Remark 2: In this paper, we focus on the case where the Jacobian matrix $J_{y_t}(x)$ is known. We note that this assumption is less restrictive than assuming the knowledge of the functional form of $y_t(x)$. This is particularly true in applications wherein the input-output relationship is *implicit*. For example, in the case of a power system, the input-output relationship is described by the nonlinear power flow equations of the form $x + \omega_t = g(y)$, where ω_t collects the uncontrollable inputs to the system (i.e., uncontrollable loads and generators), and is unknown. While the explicit functional form of y as a function of x is not readily available, the Jacobian matrix $J_y(x)$ can be computed by solving a system of linear equations $I = J_g(y) J_y(x)$, where $J_g(y)$ can be computed from the system parameters and the measurement of the output y . An illustrative numerical power system example is given in Section IV.

III. TRACKING PERFORMANCE ANALYSIS

In this section, we study the tracking performance of the proposed regularized primal-dual gradient method with feedback, measured by the quantity $\|\tilde{z}_t - z_t^*\|_{\nu, \epsilon}$, where the norm $\|z\|_{\nu, \epsilon}$ for $z = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m$ is defined as

$$\|z\|_{\nu, \epsilon} := (\|x\|^2 + \nu^{-1} \epsilon \|\lambda\|^2)^{1/2}.$$

To this end, we define the following quantities:

$$L_h = \sup_{t \in \mathcal{T}} \|H_t\|, \quad L_y(\delta) = \sup_{\substack{u: \|u\| \leq \delta \\ t \in \mathcal{T}}} \|J_{y_t}(x_t^* + u)\|, \quad (12a)$$

$$M_y(\delta) = \sup_{\substack{u: \|u\| \leq \delta \\ t \in \mathcal{T}}} \frac{\|y_t(x_t^* + u) - y_t(x_t^*) - J_{y_t}(x_t^* + u)u\|}{\|u\|^2}, \quad (12b)$$

$$M_\lambda = \sup_{t \in \mathcal{T}} \|\lambda_t^*\|. \quad (12c)$$

Roughly speaking, $L_y(\delta)$ represents the Lipschitz constant and $M_y(\delta)$ characterizes the nonlinearity of the input-output map y_t within a neighborhood of radius δ around x_t^* over all t .

We also define

$$\eta_\nu(\delta) = \sup_{\substack{u: \|u\| \leq \delta \\ t \in \mathcal{T}}} \|I - \nu^{-1} \nabla_{xx}^2 L_t(x_t^* + u, \lambda_t^*)\| \quad (13)$$

and

$$\rho_{\nu,\epsilon}(\delta, \tau) = \left[(1 - \tau + \tau \eta_\nu(\delta))^2 + \frac{\delta L_h M_y(\delta)}{\sqrt{\nu\epsilon}} (1 - \tau) \tau + \left(\frac{L_h^2 L_y^2(\delta)}{\nu\epsilon} + \frac{L_h L_y(\delta) \eta_\nu(\delta)}{\sqrt{\nu\epsilon}} \right) \tau^2 \right]^{1/2}. \quad (14)$$

We denote

$$\sigma = \sup_{t \in \mathcal{T} \setminus \{0\}} \|z_t^* - z_{t-1}^*\|_{\nu,\epsilon},$$

which characterizes the rate of change of the optimal primal-dual pair $z_t^* = (x_t^*, \lambda_t^*)$ over time. For the feedback measurement \tilde{y}_t , we assume that the measurement error is upper bounded by e_y , i.e.,

$$\|\tilde{y}_t - y_t(\tilde{x}_{t-1})\| \leq e_y, \quad \forall t \in \mathcal{T}. \quad (15)$$

The following theorem gives sufficient conditions that guarantee bounded tracking error for the regularized primal-dual gradient method with feedback (11).

Theorem 1: Suppose there exist $\delta > 0$, $\tau \in (0, 1]$, $\nu > 0$ and $\epsilon > 0$ such that

$$\rho_{\nu,\epsilon}(\delta, \tau) < 1 \quad (16a)$$

and

$$\sigma \leq (1 - \rho_{\nu,\epsilon}(\delta, \tau))\delta - \sqrt{2\nu^{-1}\epsilon}\tau(M_\lambda + \epsilon^{-1}L_h e_y). \quad (16b)$$

If the initial point $\tilde{z}_0 = (\tilde{x}_0, \tilde{\lambda}_0)$ satisfies

$$\|\tilde{z}_0 - z_0^*\|_{\nu,\epsilon} \leq \frac{\rho_{\nu,\epsilon}(\delta, \tau)\sigma + \sqrt{2\nu^{-1}\epsilon}\tau(M_\lambda + \epsilon^{-1}L_h e_y)}{1 - \rho_{\nu,\epsilon}(\delta, \tau)},$$

then the sequence $(\tilde{z}_t)_{t \in \mathcal{T}}$ generated by the regularized primal-dual gradient method with feedback (11) satisfies

$$\|\tilde{z}_t - z_t^*\|_{\nu,\epsilon} \leq \frac{\rho_{\nu,\epsilon}(\delta, \tau)\sigma + \sqrt{2\nu^{-1}\epsilon}\tau(M_\lambda + \epsilon^{-1}L_h e_y)}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} \quad (17)$$

for all $t \in \mathcal{T}$.

A sketch of the proof will be given in Appendix I.

We now discuss some implications of Theorem 1. First, we do not impose “global” convexity on the objective function c_t , and we also allow nonconvex input-output map y_t . This is especially useful for optimal power flow problems which are nonconvex in general; see the numerical example in Section IV below. On the other hand, the condition (16) implicitly requires “local” convexity of the Lagrangian: $L_t(\cdot, \lambda_t^*)$ should be strongly convex in some neighborhood of x_t^* with radius δ . This can be seen from (14) that

$$1 - \tau + \tau \eta_\nu(\delta) < \rho_{\nu,\epsilon}(\delta, \tau) < 1 \implies \eta_\nu(\delta) < 1,$$

and by the definition (13), $\nabla_{xx}^2 L_t(x, \lambda_t^*)$ should be positive definite for all x with $\|x - x_t^*\| \leq \delta$ for all t .

Next let’s look at the tracking error bound in (17), which roughly consists of three terms:

1) The first term

$$\frac{\rho_{\nu,\epsilon}(\delta, \tau)}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} \sigma$$

is proportional to σ , the rate of change of the optimal primal-dual pair over time. This term seems common for time-varying optimization algorithms [9], [12], [3].

2) The second term

$$\frac{\sqrt{2\nu^{-1}\epsilon}\tau}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} M_\lambda \quad (18)$$

is proportional to M_λ , the maximum magnitude of the optimal Lagrange multiplier λ_t^* over time. This term represents the discrepancy introduced by adding regularization in (6); similar behavior has also been observed in [18].

3) The third term

$$\frac{\sqrt{2\nu^{-1}\epsilon}\tau\epsilon^{-1}}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} L_h e_y \quad (19)$$

accounts for the measurement error introduced by feedback, and is proportional to the maximum norm of L_h and the measurement error bound e_y .

In addition, the first term has a multiplicative factor $\rho_{\nu,\epsilon}(\delta, \tau)/(1 - \rho_{\nu,\epsilon}(\delta, \tau))$, while the second and the third terms have a multiplicative factor $1/(1 - \rho_{\nu,\epsilon}(\delta, \tau))$, which are all monotonically increasing in $\rho_{\nu,\epsilon}(\delta, \tau)$. This implies that a smaller $\rho_{\nu,\epsilon}(\delta, \tau)$ will lead to better tracking performance. Here we point out two of the major factors that affect $\rho_{\nu,\epsilon}(\delta, \tau)$, as can be seen from the definition (14):

- 1) $\rho_{\nu,\epsilon}(\delta, \tau)$ is monotonically increasing in $\eta_\nu(\delta)$ when $\eta_\nu(\delta) < 1$. Then, the definition (13) indicates that, in order to achieve better tracking performance, the parameter ν should be properly chosen; the more accurately it approximates the curvature of L_t , the smaller the tracking error is.
- 2) $\rho_{\nu,\epsilon}(\delta, \tau)$ is also monotonically increasing in $M_y(\delta)$, indicating that the more nonlinear the constraints are, the more difficult it is for the algorithm (11) to track the optimal trajectory. This is in accordance with intuition.

Finally, we briefly discuss the effect of the parameters τ and ϵ . Using smaller τ and ϵ can help reduce (18), i.e., the discrepancy that results from regularization. On the other hand, a smaller ϵ will increase $\rho_{\nu,\epsilon}(\delta, \tau)$ by (14), which may degrade the tracking performance; it may also increase (19) and make the algorithm more sensitive to measurement noise. The relation between the quantities $\rho_{\nu,\epsilon}(\delta, \tau)$ and τ, ϵ is complicated. These observations suggest that there are complicated trade-offs in choosing τ and ϵ in order to achieve good tracking performance. In practice, they can be chosen heuristically by experiments or simulations.

IV. ILLUSTRATIVE NUMERICAL EXAMPLE

In this section we present a numerical example of the proposed algorithm applied to a power system test case.

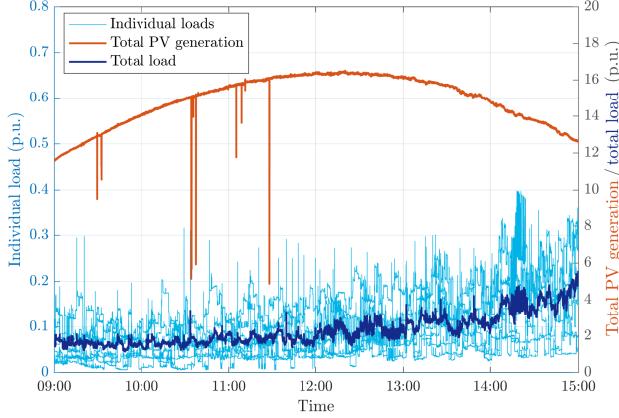


Fig. 1. Profiles of individual loads, total load and total PV generation.

The test case is based on a single-phase version of the IEEE 37 node test feeder with high penetration of photovoltaic (PV) systems. We replace the loads in the original test case with real load data measured from feeders in Anatolia, CA during the week of August 2012 [30]. A total of 18 PV systems are installed at node 4, 7, 10, 13, 17, 20, 22, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35 and 36. The rating of these inverters are 100 kVA at node 10, 350 kVA at node 36, and 200 kVA for the remaining ones. The generation profiles are simulated based on the real solar irradiance data available in [30]. The load and generation profiles are shown in Figure 1. The decision variables are the real power injection p_i and reactive power injection q_i for each inverter $i = 1, \dots, 18$. The objective function in (3) is

$$c_t(p, q) = \sum_{i=1}^{18} c_p (p_i - P_{i,t})^2 + c_q q_i^2,$$

where $P_{i,t}$ is the maximum real power available at PV system i at time t . The cost coefficients are $c_p = 3$ and $c_q = 1$. We set

$$\mathcal{X}_t = \{(p, q) : p_i^2 + q_i^2 \leq S_{i,\max}^2, 0 \leq p_i \leq P_{i,t}, \forall i\},$$

where $S_{i,\max}$ is the rated apparent power for inverter i . The function y_t is the mapping from power injections to the voltage magnitudes at each node, and is an implicit function derived from power flow equations². We require the voltage magnitudes at each node to be within $[0.95, 1.05]$ p.u., which constitutes the nonconvex constraints (9). The proposed algorithm updates the setpoints of the inverters every 0.5 s, and the parameters are $\tau = 0.04$, $\nu = 2$ and $\epsilon = 3.2 \times 10^{-5}$. We also add noise with $c_y = 8 \times 10^{-3}$ p.u. on the voltage magnitudes in the dual step to simulate the measurement error.

Figure 2 shows the tracking error $\|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}$ obtained from the proposed algorithm (11). It can be seen that apart

²It has been empirically observed that this implicit function exists given a “reasonable” set of injections in most situations. Recent work [31], [32] has identified conditions under which there is a unique power flow solution within a certain domain.

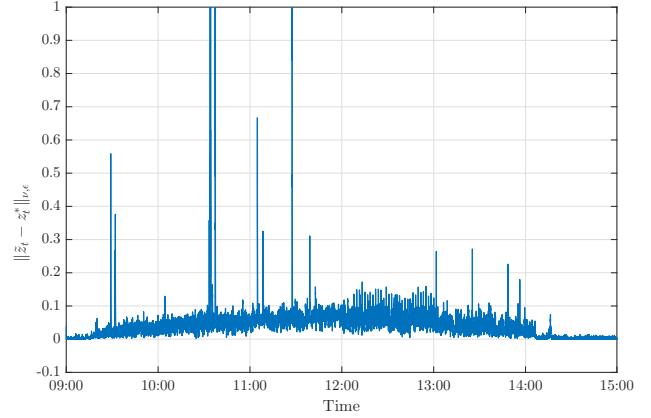


Fig. 2. The tracking error $\|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}$.

from some spikes, the tracking error is bounded below 0.2 for all t . We also have the following statistics:

$$\begin{aligned} \frac{1}{T} \sum_t \|\tilde{z}_t - z_t^*\|_{\nu,\epsilon} &= 4.3 \times 10^{-2}, \\ \frac{1}{T} \sum_t \frac{\|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}}{\|z_t^*\|_{\nu,\epsilon}} &= 1.2 \times 10^{-2}. \end{aligned}$$

By comparing Figure 2 with Figure 1, we can see that most of the spikes occur where there are spikes in the load or generation profiles. Moreover, all the spikes are quickly detected and suppressed by the proposed algorithm.

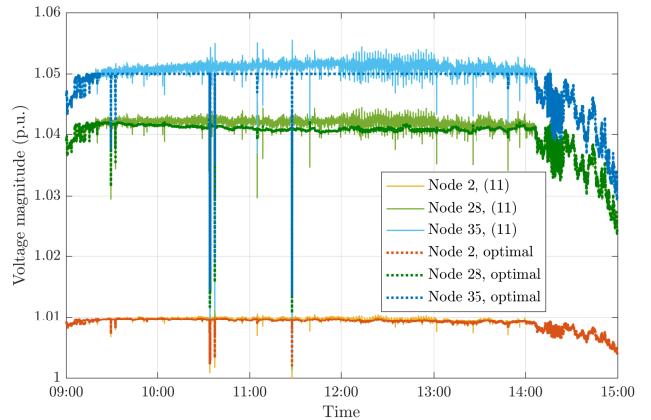


Fig. 3. Voltage magnitudes of Node 2, 28 and 35.

Figure 3 compares the voltage magnitude profiles at three nodes obtained from the proposed algorithm (11) and the optimal solutions. It can be seen that the voltage magnitudes at node 35 are slightly above 1.05 p.u. for certain time instants, indicating that constraints (9) are violated. The violation is very small though, as simulation gives

$$\frac{1}{T} \sum_t \| [f_t(\tilde{x}_t)]_+ \| = 8.7 \times 10^{-4},$$

where $[f_t(\tilde{x}_t)]_+$ denotes the vector formed by the positive part of each entry of $f_t(\tilde{x}_t)$. The violation is partly a con-

sequence of introducing the regularization $-\epsilon\|\lambda\|^2/2$, which drives the Lagrange multiplier towards zero and leads to an underestimation of the optimal dual variable. Fortunately, the constraint violation can possibly be upper bounded as the tracking error $\|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}$ can be upper bounded theoretically. Furthermore, our ongoing work suggests that the constraint violation can be alleviated by employing a more complicated regularization on the Lagrange multiplier, which will be discussed in our future work.

V. CONCLUSION

In this paper, we propose a feedback-based regularized primal-dual gradient method for tracking a trajectory of KKT points of a time-varying nonconvex system, where analytical models for the system are replaced with actual measurements. We develop analytical results that guarantee bounded tracking error, and present detailed discussion on their implications. We also give simulation results on a power system test case.

Some potential generalizations are as follows:

- 1) In (7) the parameters ν and ϵ are uniform for each entry of the primal variable and the dual variable respectively. It is straightforward to make the generalization that we assign different ν and ϵ to each entry of the primal and dual variables.
- 2) In (6) we employ a simple regularization $-\epsilon\|\lambda\|^2/2$ that drives λ towards zero. We have found that by employing a more sophisticated regularization term, the discrepancy (18) can be reduced and better tracking performance can be achieved.
- 3) The feedback formulation assumes that h_t is an affine map for all t for simplicity. We are working on the general case where h_t is a nonlinear map.
- 4) In this paper we assume that J_{y_t} can be accurately and efficiently constructed, but it is not always the case in practice. To facilitate the computation of J_{y_t} , we can utilize feedback measurement data and/or employ good approximation, which will be left for future work where general feedback and approximate gradient are considered.
- 5) In this paper we only discuss briefly on the relation between the parameters of the proposed algorithm and the tracking performance. A thorough analysis of how to choose the parameters will be both theoretically important and practically relevant.

APPENDIX I PROOF SKETCH OF THEOREM 1

The proof is by induction. Suppose that (17) holds for time $t-1$. We can see from the conditions (16) that

$$\begin{aligned} \|\tilde{z}_{t-1} - z_t^*\| &\leq \|\tilde{z}_{t-1} - z_{t-1}^*\| + \|z_{t-1}^* - z_t^*\| \\ &\leq \frac{\rho_{\nu,\epsilon}(\delta, \tau)}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} \sigma + \frac{\sqrt{2\nu^{-1}}\epsilon\tau(M_\lambda + \epsilon^{-1}L_h e_y)}{1 - \rho_{\nu,\epsilon}(\delta, \tau)} + \sigma \\ &= \frac{\sigma + \sqrt{2\nu^{-1}}\epsilon\tau(M_\lambda + \epsilon^{-1}L_h e_y)}{1 - \rho_{\nu,\epsilon}(\delta, \zeta, \tau)}, \end{aligned} \quad (20)$$

which can further be shown to be less than or equal to δ . The KKT conditions (4) imply that

$$x_t^* = \mathcal{P}_{\mathcal{X}_t}(x_t^* - \tau\nu^{-1}(\nabla c_t(x_t^*) + [H_t J_{y_t}(x_t^*)]^T \lambda_t^*)), \quad (21a)$$

$$\lambda_t^* = \mathcal{P}_{\mathbb{R}_+^m}(\lambda_t^* + \tau\epsilon^{-1}h_t(y_t(x_t^*))). \quad (21b)$$

By comparing them with (11) and using the fact that $\mathcal{P}_{\mathcal{X}_t}$ and $\mathcal{P}_{\mathbb{R}_+^m}$ are non-expansive, it can be shown that

$$\begin{aligned} \|\tilde{x}_t - x_t^*\|^2 &\leq \left\| \left[\int_0^1 (I - \tau\nu^{-1}\Delta_t(\theta)) d\theta \right] (\tilde{x}_{t-1} - x_t^*) \right. \\ &\quad \left. - \tau\nu^{-1}[H_t J_{y_t}(\tilde{x}_{t-1})]^T (\tilde{\lambda}_{t-1} - \lambda_t^*) \right\|^2, \\ \|\tilde{\lambda}_t - \lambda_t^*\|^2 &\leq \left\| (1 - \tau)(\tilde{\lambda}_{t-1} - \lambda_t^*) \right. \\ &\quad \left. + \tau\epsilon^{-1}H_t(y_t(\tilde{x}_{t-1}) - y_t(x_t^*)) \right. \\ &\quad \left. + \tau(\epsilon^{-1}H_t(\tilde{y}_t - y_t(\tilde{x}_{t-1})) - \lambda_t^*) \right\|^2, \end{aligned} \quad (22)$$

where

$$\Delta_t(\theta) = \nabla_{xx}^2 L_t(x_t^* + \theta(\tilde{x}_{t-1} - x_t^*), \lambda_t^*).$$

By the definitions (12) and (13) and the assumption (15), we have

$$\begin{aligned} &\left\| \int_0^1 (I - \tau\nu^{-1}\Delta_t(\theta)) d\theta \right\| \\ &= \left\| (1 - \tau)I + \tau \int_0^1 (I - \nu^{-1}\Delta_t(\theta)) d\theta \right\| \\ &\leq 1 - \tau + \tau\eta_\nu(\delta), \end{aligned}$$

$$\|H_t J_{y_t}(\tilde{x}_{t-1})^T (\tilde{\lambda}_{t-1} - \lambda_t^*)\| \leq L_h L_y(\delta) \|\tilde{x}_{t-1} - x_t^*\|,$$

$$\|H_t(y_t(\tilde{x}_{t-1}) - y_t(x_t^*))\| \leq L_h L_y(\delta) \|\tilde{x}_{t-1} - x_t^*\|,$$

$$\|\epsilon^{-1}H_t(\tilde{y}_t - y_t(\tilde{x}_{t-1})) - \lambda_t^*\| \leq \epsilon^{-1}L_h e_y + M_\lambda,$$

and

$$\begin{aligned} &\left\| (1 - \tau)(y_t(\tilde{x}_{t-1}) - y_t(x_t^*)) \right. \\ &\quad \left. - J_{y_t}(\tilde{x}_{t-1}) \left[\int_0^1 (I - \tau\nu^{-1}\Delta_t(\theta)) d\theta \right] (\tilde{x}_{t-1} - x_t^*) \right\| \\ &= \left\| (1 - \tau)(y_t(\tilde{x}_{t-1}) - y_t(x_t^*)) - J_{y_t}(\tilde{x}_{t-1})(\tilde{x}_{t-1} - x_t^*) \right. \\ &\quad \left. - \tau J_{y_t}(\tilde{x}_{t-1}) \left[\int_0^1 (I - \nu^{-1}\Delta_t(\theta)) d\theta \right] (\tilde{x}_{t-1} - x_t^*) \right\| \\ &\leq ((1 - \tau)\delta M_y(\delta) + \tau L_y(\delta)\eta_\nu(\delta)) \|\tilde{x}_{t-1} - x_t^*\|. \end{aligned}$$

By combining these bounds with (22), it can be shown that

$$\begin{aligned} \|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}^2 &= \|\tilde{x}_t - x_t^*\|^2 + \nu^{-1}\epsilon\|\tilde{\lambda}_t - \lambda_t^*\|^2 \\ &\leq \left[(1 - \tau + \tau\eta_\nu(\delta))^2 + \tau^2 \frac{L_h^2 L_y^2(\delta)}{\nu\epsilon} \right] \|\tilde{x}_{t-1} - x_t^*\|^2 \\ &\quad + \nu^{-1}\epsilon \left[(1 - \tau)^2 + \tau^2 \frac{L_h^2 L_y^2(\delta)}{\nu\epsilon} \right] \|\tilde{\lambda}_{t-1} - \lambda_t^*\|^2 \end{aligned}$$

$$\begin{aligned}
& + 2\tau\nu^{-1} [(1-\tau)\delta L_h M_y(\delta) + \tau L_h L_y(\delta)\eta_\nu(\delta)] \\
& \cdot \|\tilde{x}_{t-1} - x_t^*\| \|\tilde{\lambda}_{t-1} - \lambda_t^*\| \\
& + 2\tau\sqrt{\nu^{-1}\epsilon} (M_\lambda + \epsilon^{-1}L_h e_y) \\
& \cdot \left[\frac{\tau L_h L_y(\delta)}{\sqrt{\nu\epsilon}} \|\tilde{x}_{t-1} - x_t^*\| + (1-\tau)\sqrt{\frac{\epsilon}{\nu}} \|\tilde{\lambda}_{t-1} - \lambda_t^*\| \right] \\
& + \tau^2\nu^{-1}\epsilon (M_\lambda + \epsilon^{-1}L_h e_y)^2. \tag{23}
\end{aligned}$$

By Young's inequality, we have

$$\begin{aligned}
& \|\tilde{x}_{t-1} - x_t^*\| \|\tilde{\lambda}_{t-1} - \lambda_t^*\| \\
& \leq \frac{1}{2} \sqrt{\frac{\nu}{\epsilon}} \left(\|\tilde{x}_{t-1} - x_t^*\|^2 + \nu^{-1}\epsilon \|\tilde{\lambda}_{t-1} - \lambda_t^*\|^2 \right)
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{\sqrt{2}} \left[\frac{\tau L_h L_y(\delta)}{\sqrt{\nu\epsilon}} \|\tilde{x}_{t-1} - x_t^*\| + (1-\tau)\sqrt{\frac{\epsilon}{\nu}} \|\tilde{\lambda}_{t-1} - \lambda_t^*\| \right] \\
& \leq \left[\frac{\tau^2 L_h^2 L_y^2(\delta)}{\nu\epsilon} \|\tilde{x}_{t-1} - x_t^*\|^2 + \frac{(1-\tau)^2\epsilon}{\nu} \|\tilde{\lambda}_{t-1} - \lambda_t^*\|^2 \right]^{\frac{1}{2}}
\end{aligned}$$

Then by combining them with (23) and the definition of $\rho_{\nu,\epsilon}(\delta, \tau)$ (14), it can be shown that

$$\begin{aligned}
& \|\tilde{z}_t - z_t^*\|_{\nu,\epsilon}^2 \\
& \leq \rho_{\nu,\epsilon}^2(\delta, \tau) \|\tilde{z}_{t-1} - z_t^*\|_{\nu,\epsilon}^2 + 2\tau^2\nu^{-1}\epsilon (M_\lambda + \epsilon^{-1}L_h e_y)^2 \\
& \quad + 2\tau\sqrt{2\nu^{-1}\epsilon} (M_\lambda + \epsilon^{-1}L_h e_y) \rho_{\nu,\epsilon}(\delta, \tau) \|\tilde{z}_{t-1} - z_t^*\|_{\nu,\epsilon}
\end{aligned}$$

or

$$\begin{aligned}
& \|\tilde{z}_t - z_t^*\|_{\nu,\epsilon} \\
& \leq \rho_{\nu,\epsilon}(\delta, \tau) \|\tilde{z}_{t-1} - z_t^*\|_{\nu,\epsilon} + \sqrt{2\nu^{-1}\epsilon} \tau (M_\lambda + \epsilon^{-1}L_h e_y).
\end{aligned}$$

Finally, by plugging in (20), it's straightforward to check that (17) also holds for time t , and by induction we complete the proof.

REFERENCES

- [1] A. L. Dontchev, M. I. Krastanov, R. T. Rockafellar, and V. M. Veliov, "An Euler–Newton continuation method for tracking solution trajectories of parametric variational inequalities," *SIAM J. Control Optim.*, vol. 51, no. 3, pp. 1823–1840, 2013.
- [2] A. Simonetto and G. Leus, "Double smoothing for time-varying distributed multiuser optimization," in *IEEE Global Conf. on Signal and Information Processing*, pp. 852–856, Dec. 2014.
- [3] A. Simonetto, "Time-varying convex optimization via time-varying averaged operators." [Online] Available at: <https://arxiv.org/abs/1704.07338>.
- [4] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Nonlinear Programming*. Stanford University Press, 1958.
- [5] A. Cherukuri, B. Gharesifard, and J. Cortés, "Saddle-point dynamics: Conditions for asymptotic stability of saddle points," *SIAM J. Control Optim.*, vol. 55, no. 1, pp. 486–511, 2017.
- [6] S. Rahili and W. Ren, "Distributed continuous-time convex optimization with time-varying cost functions," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1590–1605, 2017.
- [7] M. Fazlyab, C. Nowzari, G. J. Pappas, A. Ribeiro, and V. M. Preciado, "Self-triggered time-varying convex optimization," in *Proceedings of the 55th IEEE Conference on Decision and Control*, (Las Vegas, NV, US), pp. 3090–3097, Dec. 2016.
- [8] D. P. Bertsekas, *Nonlinear Programming*. Belmont, Massachusetts: Athena Scientific, 1999.
- [9] E. Dall'Anese and A. Simonetto, "Optimal power flow pursuit," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 942–952, 2018.
- [10] A. Bernstein, L. Reyes-Chamorro, J.-Y. Le Boudec, and M. Paolone, "A composable method for real-time control of active distribution networks with explicit power setpoints. Part I: Framework," *Electric Power Systems Research*, vol. 125, pp. 254–264, Aug. 2015.
- [11] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Projected gradient descent on Riemannian manifolds with applications to online power system optimization," in *Proceedings of the 54th Annual Allerton Conference on Communications, Control and Computing*, (Monticello, IL, USA), pp. 225–232, 2016.
- [12] Y. Tang, K. Dvijotham, and S. Low, "Real-time optimal power flow," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2963–2973, 2017.
- [13] A. Bernstein and E. Dall'Anese, "Real-time feedback-based optimization of distribution grids: A unified approach," [Online] Available at: <https://arxiv.org/abs/1711.01627>.
- [14] Y. Tang and S. Low, "Distributed algorithm for time-varying optimal power flow," in *Proceedings of the 56th IEEE Conference on Decision and Control*, (Melbourne, Australia), pp. 3264–3270, 2017.
- [15] S. Bolognani, R. Carli, G. Cavarro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 966–981, 2015.
- [16] Y. Zhang, E. Dall'Anese, and M. Hong, "Dynamic ADMM for real-time optimal power flow," in *IEEE Global Conf. on Signal and Information Processing*, pp. 1085–1089, Nov. 2017.
- [17] A. Hauswirth, A. Zanardi, S. Bolognani, F. Dörfler, and G. Hug, "Online optimization in closed loop on the power flow manifold," in *Proceedings of IEEE PES PowerTech*, pp. 1–6, Jun. 2017.
- [18] J. Koshal, A. Nedić, and U. V. Shanbhag, "Multiuser optimization: Distributed algorithms and error analysis," *SIAM J. Optimization*, vol. 21, no. 3, pp. 1046–1081, 2011.
- [19] R. T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 97–116, 1976.
- [20] M. B. Khuzani and N. Li, "Distributed regularized primal-dual method: Convergence analysis and trade-offs." [Online] Available at: <https://arxiv.org/abs/1609.08262>.
- [21] R. Andreani, G. Haeser, and J. M. Martínez, "On sequential optimality conditions for smooth constrained optimization," *Optimization*, vol. 60, no. 5, pp. 627–641, 2011.
- [22] L. Gan and S. H. Low, "An online gradient algorithm for optimal power flow on radial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 625–638, 2016.
- [23] H. Zhu and H. J. Liu, "Fast local voltage control under limited reactive power: Optimality and stability analysis," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3794–3803, 2016.
- [24] F. Bullo, *Lectures on Network Systems*. 2018. With contributions by J. Cortés, F. Dörfler, and S. Martínez. [Online] Available at: <http://motion.me.ucsb.edu/book-lns>.
- [25] M. Hong, "Decomposing linearly constrained nonconvex problems by a proximal primal dual approach: Algorithms, convergence, and applications," [Online] Available at: <https://arxiv.org/abs/1604.00543>.
- [26] M. Hong, J. D. Lee, and M. Razaviyayn, "Gradient primal-dual algorithm converges to second-order stationary solutions for nonconvex distributed optimization," [Online] Available at: <https://arxiv.org/pdf/1802.08941.pdf>.
- [27] T. Sun, R. Barrio, L. Cheng, and H. Jiang, "Precompact convergence of the nonconvex primal-dual hybrid gradient algorithm," *J. of Computational and Applied Mathematics*, vol. 330, pp. 15–27, Mar. 2018.
- [28] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*. New York, USA: Springer, 2nd ed., 2014.
- [29] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *J. Optim. Theory Appl.*, vol. 142, no. 1, pp. 205–228, 2009.
- [30] J. Bank and J. Hambrick, "Development of a high resolution, real time, distribution-level metering system and associated visualization modeling, and data analysis functions," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-5500-56610, May 2013.
- [31] K. Dvijotham, E. Mallada, and J. W. Simpson-Porco, "High-voltage solution in radial power networks: Existence, properties, and equivalent algorithms," *IEEE Control Syst. Lett.*, vol. 1, no. 2, pp. 322–327, 2017.
- [32] C. Wang, A. Bernstein, J.-Y. Le Boudec, and M. Paolone, "Explicit conditions on existence and uniqueness of load-flow solutions in distribution networks," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 953–962, 2018.