Optimal Distributed Feedback Voltage Control
Under Limited Reactive Power

Guannan Qu and Na Li

Abstract—In this paper, we propose a distributed voltage control in power distribution networks through reactive power compensation. The proposed control can operate in a distributed fashion where each bus makes its decision based on local voltage measurements and communication with neighboring buses, always satisfy the reactive power capacity constraint, drive the voltage magnitude into an acceptable range, and minimize an operational cost. We also perform various numerical case studies to demonstrate the effectiveness and robustness of the controller using the nonlinear power flow model.

Index Terms—Real-time voltage control, distribution network.

NOMENCLATURE

A. Parameters

\( N, n, \mathcal{E} \) \( \mathcal{N} = \{0, 1, \ldots, n\} \) is the set of buses with 0 being the substation; \( n \) is the number of buses (except the substation); \( \mathcal{E} \) is the set of lines in the network.

\( N_i, \mathcal{P}_i \) Neighbors of bus \( i \) in the network (including \( i \) itself but excluding the substation); the set of lines in the network on the unique path from the substation to bus \( i \).

\( r_{ij}, x_{ij} \) Resistance and reactance on the transmission line between \( i, j \).

\( X, R \) Matrix in linearized branch-flow model.

\( Y \) Inverse of matrix \( X \).

\( v^{par} \) Parameter in the linearized branch flow model that means the vector of voltage when no control action is taken.

\( \tilde{q}_i, \hat{q}_i \) Upper, lower limits of reactive power injection at bus \( i \).

\( \tilde{v}_i, \underline{v}_i, \bar{v}_i, v \) Upper, lower limits of voltage at bus \( i \); vector of \( \tilde{v}_i \) and \( \bar{v}_i \) at all buses.

\( \mu, \ell \) Strongly convex and smooth parameter of the cost \( f \).

\( d \) Parameter in the cost function that balances the reactive power cost and the approximate power loss term.

\( \alpha, \beta, \gamma, c \) Step size parameters in our controller.

B. Variables and Functions

\( v_i, \mathbf{v} \) Squared voltage magnitude at bus \( i \); vector of \( v_i \) at all buses.

\( q_i, \mathbf{q} \) Reactive power injection at bus \( i \); vector of \( q_i \) at all buses.

\( p_i, \mathbf{p} \) Active power injection at bus \( i \); vector of \( p_i \) at all buses.

\( \hat{q}_i, \tilde{q}_i \) Virtual reactive power injection at bus \( i \) or the primal variable in the Lagrangian; vector of \( \hat{q}_i \) at all buses.

\( v_i(q), \mathbf{v}(q) \) Voltage at bus \( i \) as a function of the reactive power injection across the buses; the vector of \( v_i(q) \) at all buses. Not to be confused with \( v_i(t), \mathbf{v}(t) \), the voltage measurement at time \( t \).

\( p_{ij}, \ell_{ij}, \ell_{ij}, f_{ij}, f \) Active power flow, reactive power flow and the square of current magnitude on the transmission line from \( i \) to \( j \).

\( \ell, \bar{\lambda}_i, \bar{\lambda}_i \) Cost of providing reactive power at bus \( i \); the total cost.

\( \hat{\mathbf{\lambda}}, \bar{\mathbf{\lambda}}, \lambda \) Lagrangian multiplier for the voltage upper limit and voltage lower limit at bus \( i \).

\( \xi_i, \xi \) Vector of \( \xi_i \) at all buses; the vector of \( \xi_i \) at all buses.

\( \mathcal{L}(\hat{\mathbf{\xi}}, \xi, \lambda) \) Lagrangian multiplier for the reactive power capacity constraint at bus \( i \); the vector of \( \xi_i \) at all buses.

C. Notations

\([P]_{ij} \) \( \mathbb{R} \times 1 \) column vector with all ones.

\( \mathbf{x} \in \mathbb{R}^n \) \( n \)-dimensional nonnegative orthant \( \{x = [x_1, \ldots, x_n]^T : x_i \geq 0\} \).

\( x^\dagger \) The projection of \( x \) onto the nonnegative orthant.

\( x^{\ell} \) Projection of \( x \) onto the interval \([\ell, x]\), where \( x, \ell \) are scalars.

\( y \in \mathbb{R}^n \) Projection of \( y \) onto the box set \( \{z \in \mathbb{R}^n : z_i \in [y_i, \bar{y}_i]\} \), where \( y, \bar{y}, \tilde{y} \) are vectors of dimension \( n \).

\( \text{ST}^m_{\lambda_i}() \) Soft thresholding function.

The authors are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: gqu@g.harvard.edu; nali@seas.harvard.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2019.2931685
THE primary purpose of voltage control is to maintain acceptable voltages at all buses along a distribution feeder under all possible operating conditions [1]–[7]. Due to the increasing penetration of distributed energy resources (DER) such as photovoltaic and wind generation in the distribution networks, the operating conditions (supply, demand, voltages, etc) of the distribution feeder fluctuate fast and by a large amount. To overcome the challenges, it has been proposed to utilize the computing, (local) sensing, and (local) communication capabilities of the inverters in the DERs to adjust the reactive power injection in order to maintain voltage stability [8], [9].

Various voltage control methods have been proposed [10]–[15]. One popular approach is using optimization methods. Typically, an Optimal Power Flow (OPF) problem with voltage constraints is formulated and then “computationally” solved, either in a central (e.g. [10]–[13]) or distributed way (e.g. [16]–[19]). After obtaining the computational solution, the reactive power injection is adjusted. Hence, this approach is usually referred to as feedforward optimization: the algorithms assume knowledge of the disturbance (e.g. uncontrollable loads) and the system models while not using real-time measurements such as voltage magnitudes. In this paper, we mainly focus on distributed feedback voltage control algorithms, in which the controller does not know the disturbance explicitly, but takes local measurements and adjusts its reactive power output based on the local measurements and local communication with its neighboring buses.

There have been many efforts on developing feedback voltage control methods. One class is the traditional “Droop” control [14], [15], as advocated by IEEE 1547-2018 Standard [20]. It monitors the local bus voltages and adjusts the reactive power injection accordingly. However, [3] shows that droop type controllers are not able to maintain a feasible voltage profile under certain circumstances; ref. [6, Section V-A] shows droop control might experience stability and efficiency issues when the network is large. Therefore, other more sophisticated controllers have been proposed, e.g. [3]–[7], [21]–[23]. Ref. [3, Algorithm 1] and [4] propose an integration type controller that reaches a feasible voltage profile. Ref. [3, Algorithm 2] and [5] utilize a dual ascent approach that minimizes a power loss related cost while reaching a feasible voltage profile. However these methods also have their limitations. For instance, [3]–[5] ignore the hard constraints on the reactive power injection capacity; [6], [7] does not meet the hard voltage constraint. Though [21] meets both the voltage constraint and the hard reactive power constraint, there is a lack of theoretic guarantee for convergence even assuming linearized system model.

Besides the concerns on convergence and voltage/reactive power constraints, another issue is the optimality of voltage control. Since there is an acceptable range for voltage and reactive power, there is flexibility on the operating point of voltage control. Some operating points will have a lower operational cost than others. For example, for a DER with a fixed apparent power rating, it is preferable for the DER to generate less reactive power so that it can generate more active power. In other cases, it may be preferable for DERs to operate at certain power factor, which requires its reactive power injection to be close to a certain value. Though some existing methods, e.g., [3], [5]–[7], [16], [21], do optimize a particular objective, the objective can not be freely chosen by DERs and does not necessarily reflect the true cost of DERs. It will be appealing if the voltage control method not only maintains the voltage in the acceptable range, but also minimizes a cost that reflects a meaningful operation cost.

Our Contribution: To overcome these challenges, we propose a distributed feedback voltage control that unifies the above controllers in the sense that it can simultaneously (i) meet the voltage constraint asymptotically, (ii) satisfy the reactive power capacity constraint throughout, and (iii) minimize an operation cost that can be composed of a power loss related cost and reactive power operation costs. The controller takes the voltage measurements as inputs and determines the reactive power injection through local communication and computation. The communication graph is the same as the physical distribution network, meaning that each bus only needs to communicate with its 1-hop neighbors. The controller builds on the augmented Lagrangian multiplier theory [24, Sec. 3.2] and primal-dual gradient algorithms [25]–[28]. We mathematically prove the performance of the controller using linear branch flow models [29] and numerically simulate the controller on a real distribution feeder using the nonlinear power flow model. We also test the robustness of the controller against measurement errors, communication delays, modeling errors and unbalanced networks.

We also note that the use of communication in our controller is inevitable. In fact, [21] shows that for a class of communication-free controllers, there are scenarios in which those controllers can not reach a feasible operating point (that satisfies both the voltage and the capacity constraint), despite the existence of a feasible operating point. The results in this paper are consistent with the performance limit in [21] and demonstrate how to incorporate communication into controller design.

Finally, we note there are other related work [30]–[34]. For example, [31] proposes a online stochastic gradient descent based algorithm for online energy management; [32] studies using active power curtailment to mitigate the voltage rise caused by DER. Also related is the large literature on microgrid control [34]. For example, [30] uses a centralized voltage control scheme to correct voltage deviation, where a large linear equation system coupling all buses needs to be solved.

II. PRELIMINARIES: POWER FLOW MODEL AND PROBLEM FORMULATION

A. Branch Flow Model For Radial Networks and Its Linearization

In this paper, we consider a balanced distribution network with radial structure which consists of a set $\mathcal{N} = \{0, 1, \ldots, n\}$.
of buses and a set $\mathcal{E} \in \mathcal{N} \times \mathcal{N}$ of distribution lines connecting these buses.\(^2\) Bus 0 represents the substation and other buses in $\mathcal{N}$ represent branch buses. We let $\mathcal{N}_i$ denote the set of buses that are neighbors of bus $i$, including $i$ itself but excluding the substation bus. For each line $(i, j) \in \mathcal{E}$, let $z_{ij} = r_{ij} + jx_{ij}$ be the impedance on line $(i, j)$, and $s_{ij} = p_{ij} + jq_{ij}$ be the complex power flowing from buses $i$ to bus $j$. On each bus $i \in \mathcal{N}$, let $s_i = p_i + jq_i$ be the complex power injection. The branch flow model was first proposed in [1], [2] to model power flows in a radial distribution circuit [35], [36]: for each $(i, j) \in \mathcal{E}$,

$$-p_j = p_{ij} - r_{ij} \ell_{ij} - \sum_{k: (j, k) \in \mathcal{E}} p_{jk}$$

$$-q_j = q_{ij} - x_{ij} \ell_{ij} - \sum_{k: (j, k) \in \mathcal{E}} q_{jk}$$

$$v_j = v_i - 2(r_{ij}p_{ij} + x_{ij}q_{ij}) + (r_{ij}^2 + x_{ij}^2) \ell_{ij}$$

$$\ell_{ij} = \frac{p_{ij}^2 + q_{ij}^2}{v_i}$$

(1a)

(1b)

(1c)

(1d)

where $\ell_{ij}$ is the square of current magnitude on line $(i, j)$, and $v_i$ is the square of the voltage magnitude of bus $i$. As customary, we assume that the voltage magnitude $v_0$ on the substation bus is given and fixed at the nominal value (1 p.u.). Eq. (1) defines a system of equations in the variables $(p_{ij}, q_{ij}, \ell_{ij}, v_i, (i, j) \in \mathcal{E})$.\(^3\)

There are many ways to obtain linearized power flow models, e.g. [38]–[40]. In this paper, we use the Simplified Distflow model [29], which, along with similar models, has been used in distributed voltage control in the literature [5], [6], [21]–[23]. To obtain the Simplified Distflow model, we set the power loss term $\ell_{ij}$ to be 0 in the branch flow model (1) and get the following linear equations: for each $(i, j) \in \mathcal{E}$,

$$-p_j = p_{ij} - \sum_{k: (j, k) \in \mathcal{E}} p_{jk}$$

$$-q_j = q_{ij} - \sum_{k: (j, k) \in \mathcal{E}} q_{jk}$$

$$v_j = v_i - 2(r_{ij}p_{ij} + x_{ij}q_{ij})$$

(2a)

(2b)

(2c)

From (2), we can derive that the voltage vector $v = [v_1, \ldots, v_n]^T$ and power injection vector $p = [p_1, \ldots, p_n]^T$, $q = [q_1, \ldots, q_n]^T$ satisfy the following equation:

$$v = Rp + Xq + v_0 1$$

(3)

where 1 is an $n$-dimensional vector with all entries being 1, and the $(i, j)$th entry of matrix $R$, $X \in \mathbb{R}^{n \times n}$ is given as follows

$$[R]_{ij} := 2 \sum_{(h, k) \in \mathcal{P}_i \cap \mathcal{P}_j} r_{hk},$$

$$[X]_{ij} := 2 \sum_{(h, k) \in \mathcal{P}_i \cap \mathcal{P}_j} x_{hk}.$$ 

\(^2\)The edges in $\mathcal{E}$ are directed, with the natural ordering that points towards the direction farther away from the substation.

\(^3\)Given these variables, the phase angles of voltages and currents can be uniquely determined for radial networks [33], [37].

Here $\mathcal{P}_i \subset \mathcal{E}$ is the set of lines on the unique path from bus 0 to bus $i$. The detailed derivation is given in [41]. In [5], [41], it has been shown that when the resistances and reactances of the lines in the network are all positive, the following proposition holds.

**Proposition 1:** If $x_{ij} > 0$ for all edges $(i, j)$, then $X$ is positive definite; further define $Y := X^{-1}$, then

$$[Y]_{ij} = \begin{cases} \frac{1}{x_{ik}} & \text{if } i = j \\ -\frac{1}{x_{ij}} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

here $i \sim j$ means $i \neq j$ and $i$ is a neighbor of $j$ in the network.

**B. Problem Formulation**

We separate the reactive power $q$ into two parts, $q = q^e + q^p$, where $q^e$ denotes the control action, i.e. the reactive power injection governed by the control components and $q^p$ denotes any other reactive power injection. Let $v^{par} := Rp + Xq^e + v_0 1$. Then, given control action $q^e$, the voltage profile by (3) is

$$v(q^e) = Xq^e + v^{par}.$$  

(4)

Here we intentionally write voltage profile $v(q^e)$ as a function of $q^e$ to emphasize the input-output relationship between the control action and the voltage profile. We comment that in (4), $v^{par}$ carries the loading conditions of the network that are unknown and not controllable.

**Feedback Control Loop:** We consider the following feedback control setting. At time $t$, let the reactive power injections of the nodes be $q(t)$, then it determines the voltage profile $v(t) := v(q(t))$ through (4).\(^4\) Then given the voltage profile $v(t)$ and other available information, the controller determines a new reactive power injection $q(t + 1)$. Mathematically, the voltage control problem is formulated as the following closed loop dynamical system,

$$v(t) = v(q(t)) = Xq(t) + v^{par}$$

where the information available to $i$ at $t$ may include the local voltage measurement, information received from neighbors, and locally stored internal states (that may have their own update equations).

We also illustrate the feedback structure of the problem in Fig. 1. We make the following remarks regarding the feedback nature of our problem.

- The network is treated as an input-output system, and the controller has no knowledge of the internal details of the model (e.g. the $v^{par}$ in (4)), except that it can measure the output $v$ of the model. For now we assume the model is the Linearized Distflow (4), because the particular structure of (4) will facilitate our controller design and

\(^4\)Without causing any confusion, we abuse the notation $v(\cdot)$ to denote both the network model (4) mapping $q$ to the voltage profile $v(q)$, and the voltage profile at a certain time step $v(t)$. 
theoretic performance analysis. We emphasize that the controller in this paper can also be applied to nonlinear power flow model (1). In our numerical case studies, we test the performance of our controller on the nonlinear model.

- To facilitate theoretic analysis, we assume the \( v^{\text{bar}} \) (which captures the uncontrollable active and reactive generations and demands) in (4) is fixed during the control process. This is not required when applying our feedback controller.

In the numerical case study, we test our controller under time varying loading conditions to show how our controller automatically adapt to the changing conditions.

**Voltage Control Problem:** The objective of voltage control is to design a controller that meets the following four requirements.

**Requirement 1. Information:** The controller at \( i \) shall only use information that is accessible locally and from neighboring buses in the network. This includes local decision variable \( q_i(t) \), local voltage measurement \( v_i(t) \), other local auxiliary variables and variables communicated from neighboring buses.

**Requirement 2. Asymptotic voltage constraint:** The voltage profile reaches the acceptable lower and upper limits \( \underline{v}_i, \bar{v}_i \), i.e. \( \forall i, v_i(t) \) converges to a point inside the interval \( [\underline{v}_i, \bar{v}_i] \).

**Requirement 3. Hard capacity constraint:** For each \( i \), we introduce scalar \( \bar{q}_i \) and \( q_i \), the lower and upper reactive power capacity limit for the device at node \( i \). We require that, for any \( t \), this capacity constraint shall not be violated, i.e. \( q_i \leq q_i(t) \leq \bar{q}_i \).

**Requirement 4. Optimality:** We introduce \( f_i : \mathbb{R} \rightarrow \mathbb{R} \), the operating cost of control action for each individual bus \( i \). We require that, under any system condition \( v^{\text{bar}} \), the controller drives the distribution system to the optimal point of the following optimization problem.

\[
\begin{align*}
\min_{q_i} f(q) & \triangleq \sum_{i=1}^{n} f_i(q_i) + \frac{d}{2} q^T X q \\
\text{s.t.} & \quad \underline{v}_i \leq v_i(q) \leq \bar{v}_i \\
& \quad q_i \leq q_i \leq \bar{q}_i
\end{align*}
\]

In the optimization problem, the cost function (5a) is composed of the sum of the operating costs \( f_i \), as well as a network level cost \( \frac{d}{2} q^T X q \), with a weighting parameter \( d \geq 0 \) balancing the two costs. The cost \( \frac{d}{2} q^T X q \) is an approximation network loss term (up to a multiplicative factor and an additive term that does not depend on \( q \), under the assumption that the \( R/X \) ratio of the network is constant [5, Lemma 2]. This cost has appeared in a few related work [3], [21]. If the cost \( \frac{d}{2} q^T X q \) is not an important factor for the system operator, then \( d \) can be set as 0. We make the following regularity assumption on problem (5).

**Assumption 1:** (i) The cost function \( f \) is differentiable, and is \( \mu \)-strongly convex and \( l \)-smooth, i.e. \( \forall q, q' \in \mathbb{R}^n \),

\[
\mu \|q - q'\|^2 \leq \langle q - q', \nabla f(q) - \nabla f(q') \rangle \leq l \|q - q'\|^2.
\]

(ii) There exists a feasible solution \( \hat{q}_i^0 \) for problem (5) that meets the voltage constraint (5b) with strict inequality. In other words, \( \hat{q}_i^0 \) satisfies \( v_i < v_i(\hat{q}_i^0) < \bar{v}_i \) and \( q_i \leq \hat{q}_i^0 \leq \bar{q}_i \), \( \forall i \).

**Remark 1:** Assumption 1(i) is a standard assumption in the optimization literature. For the particular cost function in (5a) to meet Assumption 1(i), we need \( f_i \) to be convex with Lipschitz gradients. For example, any linear or quadratic \( f_i \), or zero cost function \( f_i = 0 \) meets Assumption 1(i). Assumption 1(ii) ensures that the problem we are considering is feasible.

**Remark 2:** The reactive power provider sets the cost function \( f_i \). One example of \( f_i \) is the loss of opportunity cost for providing active power service. If the apparent power limit for the device is \( s_i^\text{max} \), and the provider can sell the active power at price \( \eta_i \), then the provider can earn (at most) \( \eta_i \sqrt{(s_i^\text{max})^2 - q_i^2} \) when generating \( q_i \) amount of reactive power. If the provider does not generate any reactive power, the provider can earn \( \eta s_i^\text{max} \).

Therefore, the lost of opportunity cost is

\[
\eta s_i^\text{max} - \eta \sqrt{(s_i^\text{max})^2 - q_i^2} = \frac{\eta^2 q_i^2}{\eta s_i^\text{max} + \eta \sqrt{(s_i^\text{max})^2 - q_i^2}} \approx \frac{\eta^2 q_i^2}{\eta s_i^\text{max}}
\]

which can be set as the cost function \( f_i(q_i) \). Intuitively, the above cost function can also be understood as a penalization on large reactive power injection in order to provide more flexibility in supplying active power or maintain a power factor closer to 1.

Finally, we would like to emphasize that the fact our algorithm can guarantee optimality is an add-on functionality on top of the more basic requirements of meeting the voltage constraint with limited reactive power and local communication. If the users have no cost in supplying \( q_i, f_i \) can be set to 0.

**Remark 3:** For easy exposition and without loss of generality, we assume there is a control component at each bus \( i \). In Section IV-C, we discuss the scenario when some buses do not have control components.

## III. DISTRIBUTED VOLTAGE CONTROLLER

In this section, we formally introduce our controller, Optimal Distributed Feedback Voltage Control (OptDIST-VC). For each bus \( i \), we introduce auxiliary variables, \( \hat{q}_i, \xi_i, \hat{\lambda}_i, \hat{\zeta}_i \). At each iteration \( t \), node \( i \) measures the local voltage \( v_i(t) \), then computes variables \( \hat{q}_i(t+1), q_i(t+1), \xi_i(t+1), \hat{\lambda}_i(t+1), \hat{\zeta}_i(t+1) \) and injects the reactive power \( q_i(t+1) \), and lastly passes certain variables to its neighboring buses in the network. The detailed implementation of OptDIST-VC is given as follows.

**OPTDIST-VC:**

At time \( t \), each bus \( i \) follows the following 4 steps.

**Step 1 (Measuring):** Receive voltage measurement \( v_i(t) \).
and certain “correction directions”, are physical quantities (reactive power $q_v$). Calculate $\eta$ as the soft-thresholding function defined in (6d).

Step 2 (Calculating): Calculate $\xi(t+1), \lambda_i(t+1), \lambda_j(t+1)$ as follows.

$$
\xi(t+1) = \xi_t(t) + \beta \left( \lambda_i(t) + \gamma (v_i(t) - \hat{v}_i) \right)$$

$$
\lambda_i(t+1) = \lambda_i(t) + \gamma (v_i(t) - \hat{v}_i) + \frac{\eta_i(t+1)}{c}$$

where $\frac{\cdot}{\cdot}$ means projection onto the nonnegative orthant; quantity $\alpha, \beta, \gamma$ and $c$ are positive scalar parameters. For any $b_1 < b_2$, function $\text{ST}_{b_2}^y(\cdot)$ is the soft-thresholding function defined as, $\text{ST}_{b_2}^y(y) = \max(\min(y-b_1,0), y-b_2)$ (see Fig. 2 for an illustration).

Step 3 (Injecting Reactive Power): Set reactive power injection at time $t+1$ as

$$
q_i(t+1) = \left[ \xi_i(t+1) \right]_{q_i(t+1)}^{\eta_i(t+1)}$$

where $\left[\cdot\right]_{q_i}^{\eta_i}$ means projection onto the set $[\hat{q}_i, q_i]$.

Step 4 (Communicating): Send values $f_i'(\hat{q}_i(t+1)) + \text{ST}_{\hat{q}_i(t+1)}^b(\xi_i(t+1) + \gamma q_i(t+1))$ to neighbors.

Comment: Fig. 3 shows the information exchange between different buses and between the cyber layer (controller) and physical layer (network model) under OptDist-VC. As Fig. 3 shows, the cyber layer interacts with the physical layer only through voltage measurement $v_i(t)$ (Step 1) and reactive power injection $q_i(t)$ (Step 3), while all other activities of OptDist-VC are conducted entirely inside the cyber layer, including calculation (Step 2) and communication (Step 4). A few comments on OptDist-VC are in place.

• $q_i(t)$ and $v_i(t)$ are physical quantities (reactive power injection and voltage), while $(\hat{q}_i(t), \xi_i(t), \lambda_i(t), \lambda_j(t))$ are “digital” variables stored in the memory of the controller.

• Variable $\hat{q}_i(t)$ is the desired amount of reactive power to inject at time $t$. However it might violate the capacity constraint. Therefore, we simply set $\hat{q}_i(t)$ as the projection of $\hat{q}_i(t)$ onto the capacity constraint. As will be shown later, $\hat{q}_i(t)$ will eventually become very close to $q_i(t)$.

• The update equation (6a) for the desired reactive power injection $\hat{q}_i(t)$ is to drive $\hat{q}_i(t)$ towards the superposition of the gradient of $f$ and certain “correction directions”, related to $\lambda_j(t) - \lambda_j(t)$ and $\xi_j(t)$, which pulls $\hat{q}_i(t)$ into the constraints set. Because of the superposition of the two directions, $\hat{q}_i(t)$ will be driven to minimize $f$ and in the meanwhile avoid violating the constraints.

• The variable $\xi(t), \lambda_i(t), \lambda_j(t)$ are Lagrangian multipliers that reflect the level of constraint violation of $\hat{q}_i(t)$.

It can be readily seen that OptDist-VC meets Requirement 1 since it only needs local and neighboring information. Moreover, (7) implies that $q_i(t)$ always lies within $[\underline{q}_i, \bar{q}_i]$ and hence Requirement 3 is met. At last, the following Theorem 2 shows that $q(t)$ will converge to the optimal solution of (5) and hence Requirement 2 and 4 are satisfied. In conclusion, OptDist-VC meets all the four design requirements.

**Theorem 2:** In OptDist-VC, for any $c > 0$, when $\alpha, \beta, \gamma$ are small enough and satisfy mild conditions, $q(t)$ will converge to the unique optimizer of (5).

OptDist-VC is based on a primal-dual gradient algorithm [25]–[28] for an augmented Lagrangian [24], in which $\hat{q}_i(t)$ is the primal variable, $\xi_i(t), \lambda_i(t), \lambda_j(t)$ are the dual variables. We will elaborate on this in Section IV.

We here provide a comparison of OptDist-VC with related methods in [3]–[6], [21]–[23]. In terms of underlying methodology, [3, Algorithm 1], [6] are based on (projected) primal gradient descent, [3, Algorithm 2], [5], [21], [23] are based on dual gradient ascent, while [22] and our method are based on (partial) primal-dual gradient algorithm. Compared to our method, [3], [5] do not consider reactive power capacity limit; [6], [22], [23] treat the voltage constraint as a soft penalty in the cost function in the form of (weighted) voltage deviation, whereas our work treats the voltage constraint as hard constraint; [21] does not have theoretic guarantee for convergence when the reactive power constraint is enforced at every time step.6 We provide a summary of the comparison of these methods in Table I.

**Remark 4:** The values of $\alpha, \beta, \gamma$ will affect convergence, and hence controller performance. One one hand, these values should be relatively small to make the algorithm converge. On the other hand, very small $\alpha, \beta, \gamma$ will make the convergence very slow. We provide a theoretic upper bound on $\alpha, \beta, \gamma$ in Remark 6. However, the theoretic upper bound is conservative, which is due

---

5For the detailed conditions on the step sizes, please see Remark 6 in Appendix-B.

6Ref. [21] does have theoretic guarantee for a variant of the algorithm therein, but that variant may violate the capacity constraint before it converges to an equilibrium point.
to all known step size bounds that guarantee the “contraction” property of the primal-dual algorithm, a property that is crucial in the analysis, are known to be conservative. For this reason, in the simulations we use trial and error to obtain the step sizes. Roughly speaking, the step size values would depend on the network structure and parameters, as well as the condition number of $f$. In practice, to choose good parameter values would require some modeling and numerical testing before physical implementation, which is common for many (control) algorithms. Finally, we comment that our current results assume the network topology is fixed. It remains our future work to address the issue of robustness against system reconfiguration.

Remark 5: Under the IEEE 1547-2018 standard [20], our algorithm fits into the voltage-reactive power operation mode, where the reactive power injection responds to the voltage measurement.

IV. RATIONALE BEHIND CONTROLLER DESIGN

In this section, we will describe in detail the rationale behind our controller design, whereas the detailed mathematical proof is deferred to the appendix.

A. OPTDIST-VC as a Primal-Dual Algorithm

We introduce Lagrangian multipliers $\lambda = [\lambda^T, \bar{\lambda}^T]^T \in \mathbb{R}^{2n}$ for optimization problem (5), in which $\lambda = [\bar{\lambda}_1, \ldots, \bar{\lambda}_n]^T$ and $\bar{\lambda}_i$ corresponds to the lower limit in the voltage constraint (5b); $\bar{\lambda}^T = [\bar{\lambda}_1, \ldots, \bar{\lambda}_n]^T$ and $\bar{\lambda}_i$ corresponds to the upper limit in (5b). We also introduce multiplier $\xi = [\xi_1, \ldots, \xi_n]^T$ for the reactive power constraint (5c), and $\xi_i$ is for both the upper limit and the upper limit in the capacity constraint (5c). Next, we introduce the augmented Lagrangian for the optimization problem. Throughout this paper, we reserve letter $q$ for the physical reactive power injection, and use $\hat{q}$ to represent variables in the Lagrangian $\mathcal{L}$.

$$
\mathcal{L}(\hat{q}, \xi, \lambda) = f(\hat{q}) + \lambda^T (\mathbf{y} - \mathbf{v}(\hat{q})) + \bar{\lambda}^T (\mathbf{v}(\hat{q}) - \bar{\mathbf{v}}) + K(\xi, \hat{q})
$$

(8)

Here $K(\xi, \hat{q}) = \sum_{i=1}^{n} K_i(\xi_i, \hat{q}_i)$, and $K_i(\xi_i, \hat{q}_i)$ is a quadratic penalty function defined to be,

$$
K_i(\xi_i, \hat{q}_i) = \begin{cases} 
\xi_i(\hat{q}_i - q) + \frac{\xi_i^2}{2} & \hat{q}_i + \frac{\xi_i}{c} < q_i \\
\frac{\xi_i^2}{2} & q_i \leq \hat{q}_i + \frac{\xi_i}{c} \leq q_i \\
\xi_i(\hat{q}_i - q) + \frac{\xi_i^2}{2}(\hat{q}_i - q)^2 & \hat{q}_i + \frac{\xi_i}{c} > q_i
\end{cases}
$$

and we note that the partial derivatives of $K_i(\cdot, \cdot)$ are given as,

$$
\frac{\partial K_i(\xi_i, \hat{q}_i)}{\partial \hat{q}_i} = \text{ST}_{c q_i} \left( \xi_i + c\hat{q}_i \right)
$$

$$
\frac{\partial K_i(\xi_i, \hat{q}_i)}{\partial \xi_i} = \frac{1}{c} \left[ \text{ST}_{c q_i} (\xi_i(t) + c\hat{q}_i) - \xi_i \right].
$$

In (8), $\lambda^T (\mathbf{y} - \mathbf{v}(\hat{q})) + \bar{\lambda}^T (\mathbf{v}(\hat{q}) - \bar{\mathbf{v}})$ is the standard term in Lagrangian multiplier theory that penalizes violation of the voltage constraint, while the $K_i(\xi_i, \hat{q}_i)$ term is a special quadratic penalty function that penalizes violation of both the upper limit and the lower limit of constraint (5c). For details of such quadratic penalty functions, we refer the readers to [24, Section 3.2], [28, Appendix-G]. In short, with the penalty functions, the max-min problem

$$
\max_{\lambda \in \mathbb{R}^{2n}_+} \min_{\xi \in \mathbb{R}^n} \min_{\hat{q} \in \mathbb{R}^n} \mathcal{L}(\hat{q}, \xi, \lambda)
$$

(9)

is equivalent to the original optimization problem (5) (cf. Lemma 3). The reason we use the augmented Lagrangian instead of the standard Lagrangian is that the primal-dual gradient algorithm associated with the augmented Lagrangian avoids projection and has better convergence properties [28].

We then write down the standard primal-dual gradient algorithm [25]–[28] for solving the max-min problem (9),

$$
\hat{q}_i(t + 1) = \hat{q}_i(t) - \alpha \frac{\partial \mathcal{L}(\hat{q}(t), \xi(t), \lambda(t))}{\partial \hat{q}_i}
$$

$$
= \hat{q}_i(t) - \alpha \left[ f_i(\hat{q}_i(t)) + \text{ST}_{c q_i} (\xi_i(t) + c\hat{q}_i(t)) + \sum_{j=1}^{n} [X]_{ij} (\lambda_j(t) - \bar{\lambda}_j(t) + \delta \bar{\lambda}_j(t)) \right]
$$

(10a)

$$
\xi_i(t + 1) = \xi_i(t) + \beta \frac{\partial \mathcal{L}(\hat{q}(t), \xi(t), \lambda(t))}{\partial \xi_i}
$$

$$
= \xi_i(t) + \beta \frac{\text{ST}_{c q_i} (\xi_i(t) + c\hat{q}_i(t)) - \xi_i}{c}
$$

(10b)

$$
\lambda_i(t + 1) = \lambda_i(t) + \gamma \frac{\partial \mathcal{L}(\hat{q}(t), \xi(t), \lambda(t))}{\partial \lambda_i}
$$

$$
= \lambda_i(t) + \gamma \left( v_i(\hat{q}(t)) - \bar{v}_i \right) + \gamma \left( \hat{q}_i(t) - \bar{q}_i \right)
$$

(10c)

$$
\bar{\lambda}_i(t + 1) = \bar{\lambda}_i(t) + \gamma \left( \frac{\partial \mathcal{L}(\hat{q}(t), \xi(t), \lambda(t))}{\partial \bar{\lambda}_i} \right)
$$

$$
= \bar{\lambda}_i(t) + \gamma \left( \hat{q}_i(t) - \bar{q}_i \right)
$$

(10d)

TABLE I

| Algorithm          | Underlying Methodology | Voltage Constraint | Reactive Power Constraint | Cost       | Communication |
|--------------------|------------------------|--------------------|---------------------------|------------|---------------|
| [3, Algo. 1]      | Integral control       | Satisfied          | Not considered            | N/A        | Not needed    |
| [3, Algo. 2]      | Dual ascent            | Satisfied          | Not considered            | $\frac{1}{2}q^TXq$ | Not needed    |
| [4]               | Integral control       | Penalization in cost | Not considered            | $v$ deviation | Not needed    |
| [5]               | Dual ascent            | Penalization in cost | Not considered            | Power loss | Local comm.   |
| [6]               | Gradient Descent       | Penalization in cost | Considered                | Weighted $v$ deviation + $q$ cost | Not needed    |
| [21]              | Dual ascent            | Penalization in cost | Considered                | $\frac{1}{2}q^TXq$ | Local comm.   |
| [22]              | Primal-Dual            | Penalization in cost | Considered                | Weighted $v$ deviation + $v$ deviation | Local comm.   |
| [23]              | Accelerated dual ascent | Penalization in cost | Considered                | $v$ deviation | Local comm.   |
| Our work          | Primal-Dual            | Satisfied          | Considered                | $\frac{1}{2}q^TXq + q$ cost | Local comm.   |
Eq. (10) is the standard primal-dual gradient algorithm (also known as saddle point algorithm). At every time step, $\hat{q}_i(t)$ conducts a gradient descent step along the gradient of $\mathcal{L}$ w.r.t. $\hat{q}_i$, since $\hat{q}_i(t)$ seeks to minimize $\mathcal{L}$ (cf. the max-min problem (9)), while $\xi(t)$ and $\lambda(t)$ conducts a gradient ascent step along the gradient of $\mathcal{L}$ w.r.t. $\xi$ and $\lambda$, since $\xi(t)$ and $\lambda(t)$ seek to maximize $\mathcal{L}$. Though literature has shown the convergence of primal-dual gradient algorithms with properly chosen step sizes and under some conditions [25–28], the algorithm in (10) does not meet our design requirements. Firstly, step (10a) involves a summation from 1 to $n$ and requires information across the network to implement, violating Requirement 1. Secondly, the $\hat{q}_i(t)$ in step (10a) might violate the capacity constraint, violating Requirement 3.

We now propose two modifications to (10) to meet the design requirements and the two modifications together change (10) into OptDist-VC.

Modification (a): We now modify (10a) such that each bus only needs local and neighbor’s information to update. Eq. (10a) is a gradient update for the $\hat{q}_i$ coordinates of $\mathcal{L}$, and the gradient is given by,

$$
\nabla_\hat{q}\mathcal{L}(\hat{q}_i, \xi(t), \lambda(t)) = \nabla f(\hat{q}_i) + X(\hat{\lambda}(t) - \hat{\lambda}(t)) + \nabla_\hat{q}K(\xi(t), \hat{q}_i).
$$

(11)

Because of the sparse structure of $Y := X^{-1}$ (cf. Proposition 1), the scaled gradient $Y \nabla_\hat{q}\mathcal{L}(\hat{q}_i, \xi(t), \lambda(t))$ is given by,

$$
[Y \nabla_\hat{q}\mathcal{L}(\hat{q}_i, \xi(t), \lambda(t))]_i = \left\{ \hat{\lambda}_i(t) - \lambda_i(t) + d\hat{q}_i,
\sum_{j \in N_i} [Y]_{ij} \left[ f'_j(\hat{q}_j) + ST^{\hat{q}_j}_{\phi_{\hat{q}_j}}(\xi_j(t) + c\hat{q}_j) \right] \right\}.
$$

To calculate the $i$’th element of the scaled gradient $Y \nabla_\hat{q}\mathcal{L}(\hat{q}_i, \xi(t), \lambda(t))$, bus $i$ only needs local information ($\hat{\lambda}_i(t), \lambda_i(t)$) and information from neighbors ($f'_j(\hat{q}_j), ST^{\hat{q}_j}_{\phi_{\hat{q}_j}}(\xi_j(t) + c\hat{q}_j)$) where $j \in N_i$. Moreover, since $Y$ is positive definite, $Y \nabla_\hat{q}\mathcal{L}(\hat{q}_i, \xi(t), \lambda(t))$ is still a descent direction for $\mathcal{L}$ (in the $\hat{q}_i$ coordinates) and hence using the scaled gradient in the primal dual algorithm still has convergence guarantee [26], [27]. Therefore, we change (10a) into the following “scaled” gradient update, which gives rise to step (6a) in OptDist-VC.

$$
\hat{q}_i(t + 1) = \hat{q}_i(t) - \alpha[Y \nabla_\hat{q}\mathcal{L}(\hat{q}_i(t), \xi(t), \lambda(t))]_i.
$$

(12)

Modification (b): To fix the problem that $\hat{q}_i(t)$ may violate the capacity constraint, we do not actually implement $\hat{q}_i(t)$, but instead implement $q_i(t) = [\hat{q}_i(t)]_{\text{sad}}$, the projection of $\hat{q}_i(t)$ onto the capacity constraint. This gives rise to (7) in our controller. Another issue is that update (10c) (10d) uses $v_i(\hat{q}_i(t))$, which is not the measured voltage since the implemented reactive power is not $\hat{q}_i(t)$. Therefore, we replace the $v_i(\hat{q}_i(t))$ in (10c) (10d) with $v_i(q(t))$ and get,

$$
\tilde{\lambda}_i(t + 1) = \tilde{\lambda}_i(t) + \frac{\partial \mathcal{L}(q(t), \xi(t), \lambda(t))}{\partial \lambda_i} + \gamma(v_i(q(t)) - \tilde{v}_i) + \gamma(q_i(q(t)) - \tilde{v}_i) + \gamma(q_i(q(t)) - \tilde{v}_i(t)) + \gamma(q_i(q(t)) - \tilde{v}_i(t))
$$

which uses the measured voltage $v_i(q(t)) = v_i(t)$ and gives rise to (6c) (6d). We emphasize that after the change, the update for $\tilde{\lambda}_i(t), \hat{\lambda}_i(t)$ does not use the true gradient of $\mathcal{L}$, $\nabla_\lambda\mathcal{L}(\hat{q}_i(t), \xi(t), \lambda(t))$ any more; but uses a gradient that is evaluated at a different point, $\nabla_\lambda\mathcal{L}(q(t), \xi(t), \lambda(t))$. We will show in Section IV-B that despite of the inconsistency, we will have $\mathcal{L}(q(t) - \hat{q}_i(t) \rightarrow 0$ and the controller still converges.

B. Algorithm Analysis

We first analyze the Lagrangian (8), and show that the original problem (5) is indeed equivalent to the max-min problem (9). The proof of Lemma 3 is in Appendix-A.

Lemma 3: $\mathcal{L}(\hat{q}_i, \xi, \lambda)$ is convex in $\hat{q}_i$, concave in $\xi, \lambda$ and has a saddle point $(\hat{q}_i^\text{sad}, \xi^\text{sad}, \lambda^\text{sad})$ satisfying $\mathcal{L}(\hat{q}_i^\text{sad}, \xi^\text{sad}, \lambda^\text{sad}) = \max_{\tilde{\xi} \in \mathbb{R}^n, \tilde{\lambda} \in \mathbb{R}^m} \mathcal{L}(\hat{q}_i^\text{sad}, \xi, \lambda) = \min_{\xi \in \mathbb{R}^n, \lambda \in \mathbb{R}^m} \mathcal{L}(\hat{q}_i, \xi^\text{sad}, \lambda^\text{sad})$. Moreover, for any saddle point $(\hat{q}_i^\text{sad}, \xi^\text{sad}, \lambda^\text{sad})$, $q^\text{sad}$ must be the unique solution of the optimization problem (5).

As discussed before, our algorithm is essentially the primal-dual gradient algorithm with scaled gradient, except that in the update for $\lambda(t)$, the gradient is evaluated at $(\hat{q}(t), \xi(t), \lambda(t))$ instead of $(\hat{q}_i(t), \xi(t), \lambda(t))$. Our proof essentially shows that $q(t) - \hat{q}_i(t) \rightarrow 0$, and therefore, our algorithm is approximately the true primal-dual gradient algorithm (with scaled gradient) and it converges to a saddle point of $\mathcal{L}$. The rigorous proof of Theorem 2 can be found in Appendix-B.

C. Implementation on a Subset of Nodes

Based on the derivations in Section IV-A, we can develop a variation of the algorithm OptDist-VC that adapts to the case where only a subset of the buses are equipped with control components (referred to as controllable buses hereafter). Let the set of controllable buses be $C$, and the set of uncontrollable buses be $\bar{C}$. Then, after rearranging the labels of the buses, we can write $X$ matrix as,

$$
X = \begin{bmatrix}
X_C & G \\
G^T & X_{\bar{C}}
\end{bmatrix}
$$
and the voltage at the controllable buses $v_C$ is given by

$$v_C(q_C) = X_C q_C + G q_d + v_d^{par}$$

We can formulate an analogous optimization problem as that in (5),\(^7\)

$$\min_{q_C} f(q_C) \triangleq \sum_{i \in C} f_i(q_i) \tag{13a}$$

subject to

$$v_i \leq v_i(q_C) \leq \bar{v}_i, \forall i \in C \tag{13b}$$

$$q_i \leq \bar{q}_i, \forall i \in C \tag{13c}$$

where we note that the decision variables are now only $q_C$, the reactive power injection at the controllable buses, and the voltage constraint and the capacity constraint are also only placed on the controllable buses. Following the same modified primal-dual approach as in Section IV-A, we have the following resulting algorithm that operates on the controllable buses. For any $i \in C$,

$$\dot{q}_i(t + 1) = \hat{q}_i(t) - \alpha \left[ \lambda_i(t) - \tilde{\lambda}_i(t) + \sum_{j \in C} [X_C^{-1}]_{ij} \left[ f'_j(\hat{q}_j(t)) + ST \frac{c_{ij}}{2} (\xi_j(t) + \bar{c}_j \hat{q}_j(t)) \right] \right] \tag{14a}$$

$$\dot{\xi}_i(t + 1) = \xi_i(t) + \frac{ST \frac{c_{ij}}{2} (\xi_i(t) + \bar{c}_j \hat{q}_j(t))}{c} \tag{14b}$$

$$\tilde{\lambda}_i(t + 1) = [\lambda_i(t) + \gamma(v_i(t) - \bar{v}_i)]^+ \tag{14c}$$

$$\bar{\lambda}_i(t + 1) = [\bar{\lambda}_i(t) + \gamma(\bar{v}_i - v_i(t))]^+ \tag{14d}$$

$$q_i(t + 1) = [\hat{q}_i(t + 1) - \bar{q}_i]^+ \tag{14e}$$

The same theoretic guarantee holds for the above algorithm (14). The only substantial difference is that the matrix $X_C^{-1}$ has a sparsity pattern different from that of the physical network topology. In detail, $i$ and $j$ can communicate if and only if the path between $i$ and $j$ in the physical network does not pass any other bus in $C$. An illustration of this communication topology

\(^7\)Similar to the $\frac{1}{2} q^T X q$ term in (5), we can also add a term $\frac{1}{2} q_C^T X_C q_C$ to (13a) without changing too much of the discussion in this subsection.
Fig. 6. Simulation results for the case with static load and PV generation. The left plot and the middle plot are the reactive power injection (control action) and the voltage profile for a selected subset of buses. The right plot is the cost function.

Fig. 7. Aggregate load and PV generation profile.

Impact of Measurement Error and Delay: We test the robustness of OptDIST-VC against communication delay. In particular, when node $j$ sends variables to node $i$, we let the communication be delayed for a random amount of time up to 10 iterations (60 seconds). This means that the controller at bus $i$ implements step (6a) in the following way

$$
\hat{q}_i(t + 1) = \hat{q}_i(t) - \alpha \left\{ \delta_i(t) - d\hat{q}_i(t) + \sum_{j \in N_i} [Y]_{ij} f'_j(\hat{q}_j(t - \Delta t_{ij})) + ST_c\hat{q}_j(t - \Delta t_{ij}) + c\hat{q}_j(t - \Delta t_{ij}) \right\}
$$

where the information received from $j$ is delayed by $\Delta t_{ij}$ steps, and $\Delta t_{ij}$ is a random integer between 0 and 10. We implement OptDIST-VC under the above setting, and plot the results in Fig. 9b. It can be seen that in the delayed communication case OptDIST-VC can still guarantee the voltage stays inside the acceptable range for most of the time. Compared with Fig. 8, the voltage trajectory under communication delay shown in Fig. 9b violates the voltage constraint by larger amounts.

Impact of Modeling Error: Recall that the implementation of OptDIST-VC requires knowledge of $[Y]_{ij}$, which is essentially the reactance of the power lines (cf. Proposition 1). We test OptDIST-VC with incorrect values of $[Y]_{ij}$ that are within plus/minus 20% of the true value. The results are shown in Fig. 9c. Fig. 9c shows that under inaccurate model OptDIST-VC performs similarly as the case with accurate model (Fig. 8).

B. Three-Phase Unbalanced Network

We evaluate OptDIST-VC on a three-phase unbalanced network. We use the network with 126 multiphase nodes, the PV generation and the load profile in [43]. See [43, Section V] for high fidelity three phase models that are suited for optimization and control algorithm design are still being actively developed, our future work includes rerun these tests using newly developed models and softwares.
a detailed description of the network structure and parameters. The control components are located at parts of the buses of the network, and for the buses that have control components, the control components are located at parts of the phases. We run OptDist-VC on the network, and show the voltage profile of phase a of the buses that have control components in Fig. 10 (left). The voltage profile of the other phases is similar. We also show the voltage profile when no controller is used in Fig. 10 (right). Fig. 10 shows that OptDist-VC can maintain the voltage inside the acceptable limit despite the network is unbalanced.

VI. CONCLUSION

This paper proposes a distributed feedback voltage controller OptDist-VC that can (i) meet the voltage constraint asymptotically, (ii) satisfy the reactive power capacity constraint
throughout, and (iii) minimize a cost that is composed of a power loss related cost and reactive power operation costs. Future work includes extending the approach to jointly control active and reactive power.

APPENDIX

A. Proof of Lemma 3

For notational simplicity, we define $\tilde{X} = [-X, X]^T$, $v_b = [-\nabla^T + (\nu^{par})^T, \tilde{\nu}^T - (\nu^{par})^T]^T$ and rewrite Lagrangian (8) as,

$$L(\tilde{q}, \xi, \lambda) = f(\tilde{q}) + \lambda^T(\tilde{X}\tilde{q} - v_b) + K(\xi, \tilde{q}).$$

(15)

Recall that $(\tilde{q}, \xi, \lambda)$ is a saddle point if and only if $L(\tilde{q}, \xi, \lambda) = \min_{\tilde{q} \in \mathbb{R}^n} \max_{\xi, \lambda} L(\tilde{q}, \xi, \lambda) = \max_{\xi, \lambda} \min_{\tilde{q} \in \mathbb{R}^n} L(\tilde{q}, \xi, \lambda)$. Notice that $L(\tilde{q}, \xi, \lambda)$ is convex and lower bounded in $\xi$, concave and upper bounded in $\xi$, and linear in $\lambda$, so $(\tilde{q}, \xi, \lambda)$ is a saddle point if and only if for all $i$,

$$\frac{\partial}{\partial q_i} L(\tilde{q}, \xi, \lambda) = 0$$

(16)

$$\frac{\partial}{\partial \xi_i} L(\tilde{q}, \xi, \lambda) = 0$$

(17)

$$\lambda_i \frac{\partial}{\partial \lambda_i} L(\tilde{q}, \xi, \lambda) = \lambda_i \frac{\partial}{\partial \lambda_i} L(\tilde{q}, \xi, \lambda) = 0$$

(18)

$$\lambda_i \geq 0, \tilde{\lambda}_i \geq 0$$

(19)

$$v_i \leq v_i(\tilde{q}) \leq \tilde{v}_i.$$  

(20)

We claim the above equations are equivalent to the KKT condition of optimization problem (5). First notice that $\frac{\partial}{\partial q_i} L(\tilde{q}, \xi, \lambda) = 0$ is

$$ST_{\alpha i}^q(\xi_i + c\tilde{q}_i) = \xi_i, \forall i.$$  

Recall that $ST_{\alpha i}^q(\cdot)$ is the soft thresholding function, and we can check that the above equation is equivalent to

$$\tilde{q}_i \geq q_i$$

(21)

$$\tilde{q}_i \leq q_i$$

(22)

$$(\tilde{q}_i - q_i) \max(\xi_i, 0) = 0$$

(23)

$$(\tilde{q}_i - q_i) \min(\xi_i, 0) = 0$$

(24)

Next, notice that (18) is equivalent to,

$$\lambda_i(v_i(\tilde{q}) - v_i) = \tilde{\lambda}_i(v_i(\tilde{q}) - \bar{v}_i) = 0.$$  

(25)

Further, $\nabla \tilde{q} L(\tilde{q}, \xi, \lambda) = 0$ can be rewritten as

$$0 = \nabla f(\tilde{q}) + \tilde{X}\tilde{q} + \sum_{i=1}^{n} ST_{\alpha i}^q(\xi_i + c\tilde{q}_i) e_i$$

$$= \nabla f(\tilde{q}) + \tilde{X}\tilde{q} + \sum_{i=1}^{n} \xi_i e_i$$

$$= \nabla f(\tilde{q}) + \tilde{X}\tilde{q} + \sum_{i=1}^{n} \max(\xi_i, 0) e_i$$

$$+ \sum_{i=1}^{n} (- \min(\xi_i, 0))(-e_i)$$

(26)

where $e_i \in \mathbb{R}^n$ is a vector where the $i$’th entry is 1 and other entries are 0.

Now we replace $\max(\xi_i, 0)$ with $\tilde{\xi}_i$ and replace $-\min(\xi_i, 0)$ with $\bar{\xi}_i$ and impose constraint $\tilde{\xi}_i \geq 0, \bar{\xi}_i \geq 0$. Then (19), (20), (21), (22), (23), (24), (25), (26), together with $\tilde{\xi}_i \geq 0, \bar{\xi}_i \geq 0$, are exactly the KKT condition of optimization problem (5).

Since the Slater condition of problem (5) holds, we have the KKT condition of problem (5) has a solution, and hence $L$ has a saddle point $(\tilde{q}_{\text{sad}}, \xi_{\text{sad}}, \lambda_{\text{sad}})$. Moreover, since the objective of (5) is strongly convex, the primal solution is unique. Therefore, for any saddle point $(\tilde{q}_{\text{sad}}, \xi_{\text{sad}}, \lambda_{\text{sad}})$ of $L(\cdot, \cdot, \cdot)$, $\tilde{q}_{\text{sad}}$ must be the unique solution of problem (5).}

B. Proof of Theorem 2

As mentioned in Section IV, OptDist-VC is not the standard primal dual gradient algorithm because in update (6c) (6d), the gradient of $L$ is not evaluated at $\tilde{q}(t)$, but at $q(t)$ instead. In what follows, we define

$$S(\lambda) = \max_{\tilde{q}} \min_{\xi} L(\tilde{q}, \xi, \lambda)$$

(27)

i.e. $S(\lambda)$ is the solution of the inner layer of the max-min problem (9) given fixed $\lambda$. Our analysis treats the update of $\tilde{q}(t)$ as the gradient projection algorithm for $S(\lambda)$ where the gradient contains some “error”. To fully understand the property of $S$, we prove the following Lemma on the max-min problem of $L(\cdot, \cdot, \cdot)$ given fixed $\lambda$. The proof of Lemma 4 is given in Appendix-C.

**Lemma 4:** The following statements hold.

a) For every $\lambda$, function $L(\cdot, \cdot, \lambda)$ has a unique saddle point $(\tilde{q}^*(\lambda), \xi^*(\lambda))$ satisfying $L(\tilde{q}^*(\lambda), \xi^*(\lambda), \lambda) = \min_{\tilde{q} \in \mathbb{R}^n} L(\tilde{q}, \xi^*(\lambda), \lambda) = \max_{\xi \in \mathbb{R}^n} L(\tilde{q}^*(\lambda), \xi, \lambda).$ Moreover, $\tilde{q}^*(\lambda) \in [q_l, q_u]$.

b) $\xi^*(\lambda)$ and $\tilde{q}^*(\lambda)$ is Lipschitz in $\lambda$.

$$||\tilde{q}^*(\lambda) - \tilde{q}^*(\lambda')|| \leq \frac{||\tilde{X}||}{\mu} ||\lambda - \lambda'||$$

(28)

$$||\xi^*(\lambda) - \xi^*(\lambda')|| \leq 2\frac{1}{\mu} ||\tilde{X}|| ||\lambda - \lambda'||$$

(29)

c) $S(\lambda)$ is a concave function, and $\nabla S(\lambda) = \tilde{X}\tilde{q}^*(\lambda) - v_b$, and $S$ is $\frac{||\tilde{X}||^2}{\mu}$-smooth.

d) $S(\lambda)$ is upper bounded over $\lambda \in \mathbb{R}^2$ and moreover, for any real number $a \in \mathbb{R}$, level set $Level_a = \{\lambda \in \mathbb{R}^n : S(\lambda) \geq a\}$ is bounded.

e) Let $\lambda^*$ be any solution of $\max_{\lambda \in \mathbb{R}^2} S(\lambda)$, then $\tilde{q}^*(\lambda^*)$ is the unique solution of optimization problem (5).

Now we show that the update of $\lambda(t)$ can be written as the gradient projection algorithm for $S(\lambda)$ with noisy gradient. For notational simplicity, we abuse the notation $\tilde{q}^*(\cdot), \xi^*(\cdot)$ in Lemma 4 and define $\tilde{q}_+(t) := \tilde{q}(\lambda(t)), \xi(\cdot) := \xi^*(\lambda(t))$. Then, the gradient of $S$ at $\lambda(t)$ is (by Lemma 4(e))

$$\nabla S(\lambda(t)) = \tilde{X}\tilde{q}_+(t) - v_b.$$  

(30)

It is easy to check such replacement is one-to-one and onto.

(26) is the stationarity condition, (20), (21), (22) are the primal feasibility condition, (19) and $\tilde{\xi}_i \geq 0, \bar{\xi}_i \geq 0$ are the dual feasibility condition, (23), (24), (25), (26) are the complimentary slackness condition.
In the meanwhile, notice the update for \( \lambda(t) \) (6c) (6d) can be written as,
\[
\lambda(t + 1) = [\lambda(t) + \gamma(\hat{X}q(t) - v_b)]^+.
\]
So the update for \( \lambda(t) \) is “inexact” projected gradient method for function \( S \), and the inexact gradient is \( \hat{X}q(t) - v_b \). Therefore, the gradient error is
\[
\epsilon(t) = \|\hat{X}(q(t) - \hat{q}^I(t))\|.
\]
We show in Lemma 5 the gradient error is bounded, whose proof is deferred to Appendix-D.

**Lemma 5:** The gradient error \( \epsilon(t) = \|\hat{X}(q(t) - \hat{q}^I(t))\| \) is bounded by
\[
\epsilon(t) \leq \|\hat{X}\|\|q(t) - \hat{q}^I(t)\| \leq \|\hat{X}\|\|\hat{q}(t) - \hat{q}^I(t)\|
\]
for constants \( \rho, C_1 \) and \( C_2 \) defined as follows. First, define \( \mu' = \mu\hat{X}(Y), l' = \tilde{l}\hat{a}(Y) \). Further define constants,
\[
\alpha = 20l' \left[ \max \left( \frac{\hat{a}\hat{X}(Y)}{\mu'}, \frac{\tilde{l'}}{\mu'} \right) \right]^2 \left[ \max \left( \frac{\beta}{\alpha'}, \frac{l'}{\mu'} \right) \right]^2 \kappa(Y)
\]
\[
\tau = \frac{\beta}{2\alpha} \hat{a}(Y)
\]
\[
\nu = \max \left( \sqrt{2(l' + \hat{a}(Y)c)^2 + 2\frac{\beta^2}{\alpha^2}\hat{a}(Y)}, \sqrt{2\hat{a}(Y) + 8\frac{\beta^2}{\alpha^2}c^2} \right)
\]
\[
P_0 = \begin{bmatrix}
\begin{array}{rr}
\frac{\beta}{\alpha}I & \frac{\beta}{\alpha}Y^{1/2} \\
\frac{\beta}{\alpha}Y^{1/2} & aI \\
\end{array}
\end{bmatrix}
\]
\[
P = \begin{bmatrix}
\frac{\beta}{\alpha}Y^{-1} & \frac{\beta}{\alpha}I \\
\frac{\beta}{\alpha}I & aI \\
\end{bmatrix}
\]
Then, we can define \( \rho, C_1, \) and \( C_2 \) as follows,
\[
\rho = \left( e^{-\frac{\alpha^2}{2}} + \frac{\alpha^2l^2\kappa(P_0)}{2} \right)
\]
\[
C_1 = \|\hat{X}\|\sqrt{\kappa(P)}\|z(0) - z'(0)\|
\]
\[
C_2 = \|\hat{X}\|^2\sqrt{\kappa(P)}\sqrt{1 + \frac{4l^2}{\mu^2}}.
\]
Further, we have when fixing \( \beta/\alpha, \rho < 1 \) when \( \alpha \) is small enough.

Inexact gradient method with this type of gradient error has already been shown to have convergence guarantee, according to some related studies (e.g. [45, Sec. IV-D]). For completeness we give a result of convergence for projected inexact gradient method in the following Lemma 6, whose proof is deferred to Appendix-E.

**Lemma 6:** Recall that by Lemma 4, \( S \) is a concave and \( \frac{\|\hat{X}\|^2}{\mu'} \)-smooth function and \( S \) is upper bounded. Consider the following algorithm
\[
\lambda(t + 1) = [\lambda(t) + q(t)]^+
\]
where \([\cdot]^+\) is the projection onto the nonnegative orthant \( \mathbb{R}^n_0 \), and \( g(t) \) is an inexact gradient that satisfies
\[
\|g(t) - \nabla S(\lambda(t))\| \leq C_1\rho^t + C_2\sum_{k=0}^{t-1}\rho^{t-k-1}\|\lambda(k + 1) - \lambda(k)\|
\]
for \( C_1, C_2 > 0 \) and \( \rho \in (0, 1) \). Then, when \( \gamma < \min\left( \frac{\mu}{2\|X\|^2}, \frac{\|\hat{X}\|^21-\rho^2}{4\mu\rho^2} \right) \), we will have (i), \( \lambda(t + 1) - \lambda(t) \to 0 \), (ii) \( S(\lambda(t)) \) is lower bounded.

The above lemma shows when \( \gamma \) is small enough, \( \lambda(t + 1) - \lambda(t) \to 0 \). By Lemma 5, this further implies \( \hat{q}(t) - \hat{q}^I(t) \to 0 \) and \( q(t) - \hat{q}^I(t) \to 0 \). Since \( S(\lambda(t)) \) is lower bounded, we have \( \lambda(t) \) is a bounded sequence by Lemma 4(d). Then, sequence \( (q(t), \lambda(t)) \) is bounded and hence has a limit point \( (q_{lim}, \lambda_{lim}) \), with subsequence \( (q_{tk}, \lambda_{tk}) \) converging to \( (q_{lim}, \lambda_{lim}) \). Also since \( \hat{q}^I(t_k) = \hat{q}^I(\lambda_{lim}) \), and such \( q_{lim} = \hat{q}^I(\lambda_{lim}) \). We next show that \( \lambda_{lim} \) must be a maximizer of \( S(\lambda) \) over \( \mathbb{R}^n_0 \).

\[
\lim_{k \to \infty} \lambda(t_k + 1) = \lim_{k \to \infty} [\lambda(t_k) + \gamma(\hat{X}q(t_k) - v_b)]^+ = [\lambda_{lim} + \gamma(\hat{X}q_{lim} - v_b)]^+
\]
\[
= [\lambda_{lim} + \gamma(\hat{X}\hat{q}^I(\lambda_{lim}) - v_b)]^+ = [\lambda_{lim} + \gamma\nabla S(\lambda_{lim})]^+.
\]

Also since \( \lambda(t_k + 1) - \lambda(t_k) \to 0 \), we have \( \lambda(t_k + 1) \to \lambda_{lim} \), and hence,
\[
\lambda_{lim} = [\lambda_{lim} + \gamma\nabla S(\lambda_{lim})]^+.
\]

This implies that, for any \( \lambda \in \mathbb{R}^n_0 \),
\[
\nabla S(\lambda_{lim})^T(\lambda - \lambda_{lim}) = \frac{1}{\gamma}(\lambda_{lim} + \gamma\nabla S(\lambda_{lim}) - [\lambda_{lim} + \gamma\nabla S(\lambda_{lim})]^+)^T(\lambda - [\lambda_{lim} + \gamma\nabla S(\lambda_{lim})]^+)
\]
\[
\leq 0
\]
where we have used the projection theorem. Since \( S \) is a concave function, we have \( \lambda_{lim} \) is indeed a maximizer of \( S \) over \( \mathbb{R}^n_0 \). By Lemma 4(e), we have \( q_{lim} = \hat{q}^I(\lambda_{lim}) \) is the unique solution of the original optimization problem (5). This also shows the accumulation point of the bounded sequence \( q(t) \) is unique. So \( q(t) \) must converge to the unique solution of optimization problem (5).

**Remark 6:** We here summarize what are the theoretic step size requirements for OptDist-VC to converge. First we are given \( l, \mu, c, Y \) as part of the problem parameters. Then, fix the ratio \( \beta/\alpha \) to be any positive real number \( \eta \) based on which the constants \( a, \tau, \nu, C_2 \) and matrix \( P_0, P \) in Lemma 5 can be determined. Henceforth \( \rho = e^{-\frac{\alpha^2}{2}} + \frac{\alpha^2l^2\kappa(P_0)}{2} \) will only depend on \( \alpha \). Notice that \( \rho \) as a function of \( \alpha \) equals 1 when
\[ R - \bar{R}^0 \text{ so to guarantee } \langle \alpha \rangle < \tilde{\alpha} - \tilde{l} \geq \xi \nabla \tilde{\alpha} \leq \xi \xi R \in \tilde{R}(31b) \] 

\[ \text{We can show when fixing } \langle \rho \rangle < \tilde{\rho}^0 \text{ and } (\langle \xi \rangle, \tilde{\xi}\langle \lambda \rangle, \tilde{\xi}^0, \lambda) = (\tilde{X} \tilde{\alpha} \langle \lambda \rangle) - v_b. \] 

In (34), we have shown the first two terms of the RHS of (34) are nonnegative and hence \( \langle X^T (\lambda - \lambda') \rangle, \tilde{\xi} \langle \lambda \rangle - \tilde{\xi}^0 \langle \lambda \rangle \rangle \leq 0. \) Therefore, 

\[ \langle \nabla S(\lambda) - \nabla S(\lambda'), \lambda - \lambda' \rangle = \langle \tilde{X} (\tilde{\xi} \langle \lambda \rangle - \tilde{\xi}^0 \langle \lambda \rangle), \lambda - \lambda' \rangle \] 

\[ = \langle \tilde{X}^T (\lambda - \lambda'), \tilde{\xi} \langle \lambda \rangle - \tilde{\xi}^0 \langle \lambda \rangle \rangle \leq 0. \]

Therefore, \( S \) is a concave function. To show the smoothness of \( S \), we have 

\[ \| \nabla S(\lambda) - \nabla S(\lambda') \| \leq \| \tilde{X} ||\langle \lambda \rangle - \tilde{\xi}^0 \langle \lambda \rangle \rangle \| \] 

\[ \leq \| \tilde{X} \|^2 \mu ||\langle \lambda \rangle - \lambda' ||. \]

\[ \text{Proof of (d): Let } \tilde{q}^0 \in \mathbb{R}^n \text{ be a feasible solution to (5) and meets (5b) with strict inequality. Then, for } \lambda \in \mathbb{R}^{2n}_0, \text{ noticing } \lambda_i(v_i(\tilde{q}^0) - \bar{v}_i) \leq 0 \text{ and } \lambda_i(\bar{v}_i - v_i(\tilde{q}^0)) \leq 0, \text{ we have,} \]

\[ \min_{q\in\mathbb{R}^n} \mathcal{L}(\tilde{q}^0, \xi, \lambda) \leq \mathcal{L}(\tilde{q}^0, \xi, \lambda) \]

\[ = f(\tilde{q}^0) + \tilde{X}^T (v(\tilde{q}^0) - \bar{v}) + \lambda^T (v - v(\tilde{q}^0)) + K(\xi, \tilde{q}^0) \leq f(\tilde{q}^0) + K(\xi, \tilde{q}^0). \]

Since \( \tilde{q}^0 \in [\tilde{q}_i, \tilde{q}_i], \) we have \( K_i(0, \tilde{q}^0) = 0 \) and \( \frac{\partial K}{\partial \xi}(0, \tilde{q}^0) = 0. \) This, together with the fact that \( K_i(\xi, \tilde{q}^0) \) is concave (in \( \xi_i \)), we have \( K_i(\xi_i, \tilde{q}^0) \leq 0, \forall \xi_i \in \mathbb{R}. \) Therefore, \( \min_{q\in\mathbb{R}^n} \mathcal{L}(\tilde{q}^0, \xi, \lambda) \leq f(\tilde{q}^0), \forall \xi \in \mathbb{R}^n \) and hence, \( S(\lambda) \leq f(\tilde{q}^0) \) (for all \( \lambda \in \mathbb{R}^{2n}_0 ). \) Therefore \( S(\lambda) \) is bounded over \( \mathbb{R}^{2n}_0 \).

Further, assuming \( \lambda \in Level_\alpha, \) then still using the fact \( K(\xi, \tilde{q}^0) \leq 0, \)

\[ a \leq S(\lambda) = \max_{\xi \in \mathbb{R}} \min_{q\in\mathbb{R}^n} \mathcal{L}(\tilde{q}^0, \xi, \lambda) \leq \max_{\xi \in \mathbb{R}^n} \mathcal{L}(\tilde{q}^0, \xi, \lambda) \]

\[ \leq f(\tilde{q}^0) + \tilde{X}^T (v(\tilde{q}^0) - \bar{v}) + \lambda^T (v - v(\tilde{q}^0)). \]

Define \( \epsilon = -\max_{i=1, \ldots, n} \max_{v_i(\tilde{q}^0) - \bar{v}_i, \bar{v}_i - v_i(\tilde{q}^0)}. \) We have \( \epsilon > 0 \) since \( \tilde{q}^0 \) satisfies the voltage constraint with strict
inequality. Hence, \( a \leq f(q^0) - \epsilon \sum_{i=1}^{n} (\lambda_i + \bar{\lambda}_i) \) and
\[
\sum_{i=1}^{n} (\bar{\lambda}_i + \lambda_i) \leq \frac{1}{\epsilon} (f(q^0) - a).
\]

This together with the fact that \( \lambda \in \mathbb{R}^{2n}_+ \) shows that the level set Levela is bounded.

**Proof of (e):** By definition, \((q^*(\lambda^*), \xi^*(\lambda^*))\) is a saddle point of \(L(\cdot, \cdot, \cdot)\). Then by Lemma 3 we have \(\tilde{q}^*(\lambda^*)\) is the unique optimizer of problem (5).

**D. Bounded Gradient Error, Proof of Lemma 5**

Notice that by Lemma 4(a), \(\tilde{q}^*(\lambda)\) lies within the capacity constraint. Moreover, since \(\tilde{q}(t)\) is the projection of \(\tilde{q}(t)\) onto the capacity constraint (cf. (7)), by projection theorem we have \(\|\tilde{q}(t) - \tilde{q}(\lambda)\| \leq \|	ilde{q}(t) - \tilde{q}(\lambda)\|\). Hence the gradient error satisfies
\[
\epsilon(t) \leq \|\tilde{X}\|\|\tilde{q}(t) - \tilde{q}(\lambda)\|.
\]
(35)

By (35), to bound the gradient error we only need to bound \(\|\tilde{q}(t) - \tilde{q}(\lambda)\|\). To this end, we now analyze what happens if we conduct one step of update for \(q(t), \xi(t)\) (Eq. (6a) and (6b)). If we think of \(\lambda(t)\) as fixed, then the one step update for \(q(t), \xi(t)\) is simply a primal-dual gradient update step for function \(L(\cdot, \cdot, \cdot, \cdot, \lambda(t))\). By our recent work on the geometric convergence rate of primal-dual gradient algorithm [28], after one step of update for \(q(t), \xi(t)\) (Eq. (6a) and (6b)), the distance between \((\tilde{q}(t+1), \xi(t+1))\) and the unique saddle point of \(L(\cdot, \cdot, \cdot, \lambda(t))\), \((q^*(\lambda(t)), \xi^*(\lambda(t)))\) will shrink at least by a fixed ratio compared to the distance between \((\tilde{q}(t), \xi(t))\) and \((q^*(\lambda(t)), \xi^*(\lambda(t)))\). Here the distance is measured by a weighted norm. Formally, we stack \(q(t)\) into one large vector \(z(t) = [\tilde{q}(t)^T, \xi(t)^T]^T\), and similarly define \(z^*(t) = [\tilde{q}^*(\lambda(t))^T, \xi^*(\lambda(t))^T]^T\). Then, we have the following lemma, which is a consequence of the results in [28]. We will provide the detailed derivations in Appendix-F.

**Lemma 7:** Define \(\mu' = \mu \sigma(Y)^{1/2}, \mu'' = \mu \sigma(Y)^{-1/2}\). Further define constants,

\[
a = 20\left[ \max \left( \frac{c\sigma(Y)}{\mu}, \frac{\mu''}{\mu'} \right) \right]^2 \left[ \max \left( \frac{\beta}{\alpha c''}, \frac{\mu'}{\mu''} \right) \right]^2 \kappa(Y),
\]
\[
\tau = \frac{\beta \sigma(Y)}{2 \alpha},
\]
\[
\nu = \max \left( \sqrt{2(l' + \sigma(Y))^2 + 2 \frac{\beta}{\alpha c''} \sigma(Y)}, \sqrt{2 \sigma(Y) + 8 \frac{\beta}{\alpha c''} \sigma(Y)} \right).
\]

Then we define matrix,
\[
P_0 = \begin{bmatrix} \frac{\beta}{\alpha} I & \frac{\beta}{\alpha} Y^{1/2} \\ \frac{\beta}{\alpha} a Y^{-1/2} & a I \end{bmatrix},
\]
and constant,
\[
\rho = \left( e^{-\frac{\alpha^2 \nu^2 \kappa(P_0)}{2}} \right).
\]

Then, we have,
\[
\|z(t + 1) - z^*(t)\|_p \leq \rho \|z(t) - z^*(t)\|_p
\]
(36)

where norm \(\|\cdot\|_p\) is defined as \(\|z\|_p = \sqrt{z^T P z}\). Further, fixing \(\beta / \alpha\), we have \(\rho < 1\) when \(\alpha\) is small enough.

Lemma 7, specifically eq. (36), says that that \(z(t + 1)\) gets closer to the equilibrium point \(z^*(t)\) by at least a fixed ratio \(\rho\) compared to \(z(t)\), where the distance is measured in a specially constructed norm \(\|\cdot\|_p\). From (36) we can easily get,
\[
\|z(t) - z^*(t)\|_p \leq \rho \|z(t - 1) - z^*(t - 1)\|_p + \|z(t) - z^*(t)\|_p
\]
\[
\rho \rho^k \|z(0) - z^*(0)\|_p + \sum_{k=0}^{t-1} \rho^{t-k} \|z^*(k + 1) - z^*(k)\|_p
\]
(37)

We now convert (37) into a bound on \(\|\tilde{q}(t) - \tilde{q}^*(t)\|\). First we have
\[
\|z(t) - z^*(t)\|_p \geq \sqrt{\sigma(P)} \|\tilde{q}(t) - \tilde{q}^*(t)\|.
\]
(38)

Recall that \(z^*(t) = [\tilde{q}^*(\lambda(t))^T, \xi^*(\lambda(t))^T]^T\) is the unique saddle point of \(L(\cdot, \cdot, \cdot, \cdot, \lambda(t))\). Using Lemma 4(b) (also using \(\|\tilde{X}\| = \sqrt{2} \|X\|\)),
\[
\|z^*(t + 1) - z^*(t)\|_p \leq \sigma(P) \left( \|\tilde{q}(t+1) - \tilde{q}^*(t)\|^2 + \|\xi(t+1) - \xi^*(t)\|^2 \right)
\]
\[
: = C_1
\]
\[
\leq \sigma(P) \left( \frac{\|\tilde{X}\|^2}{\mu} + 8 \left( \frac{1}{\mu} \right)^2 \|X\|^2 \right) \|\lambda(t+1) - \lambda(t)\|^2.
\]
(39)

Using (38) and (39) in (37), we have,
\[
\epsilon(t) \leq \|\tilde{X}\|\|\tilde{q}(t) - \tilde{q}^*(t)\|
\]
\[
\leq \|\tilde{X}\| \frac{1}{\sqrt{\sigma(P)}} \|z(t) - z^*(t)\|_p
\]
\[
\leq \|\tilde{X}\| \sqrt{\kappa(P)} \|z(0) - z^*(0)\|_p \rho^t
\]
\[
+ \|\tilde{X}\| \sqrt{\kappa(P)} C_1 \sum_{k=0}^{t-1} \rho^{t-1-k} \|\lambda(k + 1) - \lambda(k)\|
\]
(40)

which concludes the proof of Lemma 5.

**E. Inexact Gradient Method, Proof of Lemma 6**

**Proof:** Let \(l'' = \frac{\|X\|^2}{\mu}\). Then \(S\) is \(l''\)-smooth. By concavity and smoothness,
\[
S(\lambda(t+1)) - S(\lambda(t)) \geq \langle \nabla S(\lambda(t)), \lambda(t+1) - \lambda(t) \rangle - \frac{l''}{2} \|\lambda(t+1) - \lambda(t)\|^2
\]
\[
= \langle g(t), \lambda(t+1) - \lambda(t) \rangle
\]
By projection property, we have,
\[ (\lambda(t) - \lambda(t) + \lambda(t) + \gamma g(t) - \lambda(t + 1) \leq 0 \]
which implies
\[ (g(t), \lambda(t + 1) - \lambda(t)) \geq \frac{1}{\gamma} \|\lambda(t + 1) - \lambda(t)\|^2. \quad (42) \]
Also notice that
\[ \langle \nabla S(\lambda(t)) - g(t), \lambda(t) - \lambda(t) \rangle \]
\[ \geq -\frac{l''}{2} \|\lambda(t) - \lambda(t)\|^2 - \frac{1}{2l''} \|\nabla S(\lambda(t)) - g(t)\|^2. \]
Plugging the above and (42) into (41), we have
\[ S(\lambda(t + 1)) - S(\lambda(t)) \]
\[ \geq \left( \frac{1}{\gamma} - l'' \right) \|\lambda(t + 1) - \lambda(t)\|^2 - \frac{1}{2l''} \|\nabla S(\lambda(t)) - g(t)\|^2. \]
Summing up the above from \( t = 0, \ldots, \tau \), we get,
\[ \sum_{t=0}^{\tau} \|\nabla S(\lambda(t)) - g(t)\|^2 \]
\[ \leq C_1 \rho + \frac{1}{2} \sum_{t=0}^{\tau} \|\lambda(t) - \lambda(t)\|^2. \quad (43) \]
Now we bound \( \sum_{t=0}^{\tau} \|\nabla S(\lambda(t)) - g(t)\|^2. \) By the condition in this lemma, we have
\[ \|\nabla S(\lambda(t)) - g(t)\| \]
\[ \leq C_1 \rho + C_2 \sum_{t=0}^{\tau} \rho^{j-1} \|\lambda(k + 1) - \lambda(k)\| = \langle \chi, \nu_t \rangle \]
where we define vector \( \chi = [C_1, C_2, \|\lambda(1) - \lambda(0)\|, \ldots, C_2, \|\lambda(\tau - 1)\|] \in \mathbb{R}^{\tau + 1} \), and vector \( \nu_t = [\rho^0, \rho^{j-1}, \ldots, \rho, 0, 0, \ldots, 0] \in \mathbb{R}^{\tau + 1} \) for \( 0 \leq t \leq \tau \). Then
\[ \sum_{t=0}^{\tau} \|\nabla S(\lambda(t)) - g(t)\|^2 \leq \sum_{t=0}^{\tau} \chi^T (\nu_t \nu_t^T) \chi = \chi^T V \chi \]
where \( V = \sum_{t=0}^{\tau} \nu_t \nu_t^T \in \mathbb{R}^{(\tau + 1) \times (\tau + 1)} \). Obviously \( V \) is symmetric and positive semi-definite, with each entry being nonnegative. Now, for \( 1 \leq i \leq j \leq \tau + 1 \), the \( (i, j) \)’th entry of \( V \) is \( V_{ij} = \sum_{t=j-i}^{\tau} \rho^{j-i} \rho^{t-j} < \rho^{j-i} \frac{1}{\rho} < \rho^{j-i} \frac{1}{\rho} \). Therefore, fixing \( i, j \),
\[ \sum_{t=j-i}^{\tau} |V_{ij}| = \sum_{t=1}^{\tau} |V_{ij}| < \sum_{t=1}^{\tau} \rho^{t-j} \frac{1}{\rho} < \frac{1}{\rho} \tau \rho \],
\[ \sum_{j=i}^{\tau} |V_{ij}| = \sum_{j=1}^{\tau} |V_{ij}| = \sum_{j=1}^{\tau} \rho^j \frac{1}{\rho} < \frac{1}{(1-\rho)^2}. \]
So,
\[ \|V\|_1 \leq \sum_{j=1}^{\tau} |V_{ij}| \leq \frac{2}{(1-\rho)^2}. \]
This implies that
\[ \sum_{t=0}^{\tau} \|\nabla S(\lambda(t)) - g(t)\|^2 \]
\[ \leq \frac{2}{(1-\rho)^2} \|\lambda\|^2 \]
\[ \leq \frac{2}{(1-\rho)^2} C_1^2 + \frac{2}{(1-\rho)^2} C_2^2 \sum_{t=0}^{\tau-1} \|\lambda(t) - \lambda(t)\|^2. \]
Plugging this into (43), we have
\[ S(\lambda(\tau + 1)) - S(\lambda(0)) \]
\[ \geq \left( \frac{1}{\gamma} - l'' - \frac{C_2^2}{l''(1-\rho)^2} \right) \sum_{t=0}^{\tau} \|\lambda(t) - \lambda(t)\|^2 \]
\[ - \frac{C_2^2}{l''(1-\rho)^2}. \]
So if we choose \( \gamma < \min(\frac{1}{(1-\rho)^2}, \frac{l''(1-\rho)^2}{2l''}) \), we have \( \frac{1}{\gamma} - l'' - \frac{C_2^2}{l''(1-\rho)^2} > 0 \). Since \( S \) is upper bounded, we have \( \sum_{t=0}^{\infty} \|\lambda(t) - \lambda(t)\|^2 < \infty \), which implies \( \|\lambda(t) - \lambda(t)\|^2 \to 0 \). The above inequality also implies that \( S(\lambda(\tau + 1)) \) is lower bounded regardless of \( \tau \). \hfill \square

F. Proof of Lemma 7

In this section, since we only consider one step update, we drop the dependence on \( t \) in the notations. Specifically, we write \( \lambda(t) \) as \( \lambda \), \( \hat{q}(t) \) and \( \xi(t) \) as \( \hat{q} \) and \( \xi \), \( \hat{q}(t + 1) \) and \( \xi(t + 1) \) as \( \hat{q}^+ \) and \( \xi^+ \), and at last, \( z^+ \) and \( z \). We now do the following change of variable from \( q \) to \( y \), \( y = Y^{-1/2} q \) (\( y^+ \), \( y^+ \) are defined similarly), while \( \xi \) stays unchanged. Correspondingly, vector \( z \) becomes
\[ w = \begin{bmatrix} Y^{-1/2} & 0 \\ 0 & I \end{bmatrix} z = \begin{bmatrix} Y^{-1/2} \hat{q} \\ \xi \end{bmatrix} \]
and \( w^+ \), \( w^+ \) are defined correspondingly. We then rewrite the Lagrangian in the new variables, \( \hat{L}(y, \xi) = L(Y^{-1/2} y, \xi, \lambda) \) (we drop the dependence of \( \hat{L} \) on \( \lambda \) since throughout this section, \( \lambda \) is fixed). Then, the saddle point of \( \hat{L} \) is simply \( w^+ \). Recall the update equation from \( z \) to \( z^+ \) are
\[ \hat{q}^+ = \hat{q} - \alpha Y^2 q \hat{L}(\hat{q}, \xi, \lambda) \]
\[ \hat{q}^+ = \xi + \beta \hat{L}(\hat{q}, \xi, \lambda). \]
Now we rewrite the above update equation in variable \( w \) and get,
\[ y^+ = y + \alpha Y^{-1/2} \hat{L}(\hat{q}, \xi, \lambda) \]
\[ = y - \alpha \hat{L}(y, \xi) \]
\[ = y + \beta \hat{L}(\hat{q}, \xi, \lambda) \]
\[ = \xi + \beta \hat{L}(\hat{q}, \xi, \lambda). \]

Next, notice that $\tilde{L}(y, \xi) = f(Y^{1/2}y) + \lambda^T (\tilde{X}Y^{1/2}y - v_b) + K(\xi, Y^{1/2}y)$ is precisely the augmented Lagrangian for the following optimization problem,

$$\min_y f(Y^{1/2}y) + \lambda^T (\tilde{X}Y^{1/2}y - v_b)$$

s.t. $q \leq Y^{1/2}y \leq q.$ \hfill (46)

The objective in (46) is $\mu'$-strongly convex and $l'$-smooth, where $\mu' = \mu g(Y), l' = l\sigma(Y).$ Then, by [28, Thm. 3, Lem. 4, Lem. 5], we have the update (44), (45) is a contraction, which is formally stated in the following Proposition.

**Proposition 8:** ([28, Thm. 3, Lem. 4, Lem. 5]) Given $\mu', l', c$ and matrix $Y$, define constants,

$$a = 20l' \left[ \max \left( \frac{c(\sigma(Y)}{\mu'} \right) l' \right] \left[ \max \left( \frac{\beta}{\alpha l', c} \right) l' \right] \kappa(Y).$$

$$\tau = \frac{\beta}{2\alpha} \sigma(Y) a$$

$$\nu = \max \left( \sqrt{2(\gamma + \sigma(Y))^2 + \frac{2\beta^2}{\alpha^2} \sigma(Y)}, \sqrt{2\sigma(Y) + \frac{8\beta^2}{\alpha^2 c^2}} \right).$$

Then we define matrix,

$$P_0 = \begin{bmatrix} \frac{\beta}{\alpha} a I & \frac{\beta}{\alpha} Y^{1/2} \\ \frac{\beta}{\alpha} Y^{1/2} & a I \end{bmatrix},$$

and constant,

$$\rho = \left( e^{-\frac{\tau}{2\alpha}} + \alpha^2 \nu^2 \kappa(P_0) \right).$$

Then we have,

$$\|w^+ - w^*\|_P \leq \rho \|w - w^*\|_P$$

and further, fixing the ratio $\beta/\alpha$, we have $\rho < 1$ if $\alpha$ is small enough.

Since $w, w^+$ and $w^*$ are linear transforms of $z, z^+$ and $z^*$, we have

$$\|z^+ - z\|_P \leq \rho \|z - z^*\|_P$$

where

$$P = \begin{bmatrix} Y^{-1/2} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \frac{\beta}{\alpha} a I & \frac{\beta}{\alpha} Y^{1/2} \\ \frac{\beta}{\alpha} Y^{1/2} & a I \end{bmatrix} \begin{bmatrix} Y^{-1/2} & 0 \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\beta}{\alpha} a Y^{-1} & \frac{\beta}{\alpha} I \\ \frac{\beta}{\alpha} I & a I \end{bmatrix}.$$

This concludes the proof of Lemma 7.

**REFERENCES**

[1] M. E. Baran and F. F. Wu, “Optimal capacitor placement on radial distribution systems,” *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 725–734, Jan. 1989.

[2] M. E. Baran and F. F. Wu, “Optimal sizing of capacitors placed on a radial distribution system,” *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 735–743, Jan. 1989.

[3] N. Li, G. Qu, and M. Dahleh, “Real-time decentralized voltage control in distribution networks,” in *Proc. 52nd Ann. Allerton Conf. Commun., Control Comput.*, 2014, pp. 582–588.

[4] B. Zhang, A. D. Domínguez-García, and D. Tse, “A local control approach to voltage regulation in distribution networks,” in *Proc. North Amer. Power Symp.*, 2013, pp. 1–6.

[5] 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, Apr. 2015.

[6] H. Zhu and J. H. Liu, “Fast local voltage control under limited reactive power optimality and stability analysis,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3794–3803, Sep. 2016.

[7] V. Kekatos, L. Zhang, G. B. Giannakis, and R. Baldick, “Fast localized voltage regulation in single-phase distribution grids,” in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2015, pp. 725–730.

[8] J. Smith, W. Sunderman, D. Ragan, and B. Seal, “Smart inverter Volt/VAR control functions for high penetration of PV on distribution systems,” in *Proc. IEEE Power Syst. Conf. Expo.*, 2011, pp. 1–6.

[9] K. Tursitsyn, P. Sulc, S. Backhaus, and M. Chertkov, “Options for control of reactive power by distributed photovoltaic generators,” *IEEE Trans.*, vol. 99, no. 6, pp. 1063–1073, Jun. 2011.

[10] M. Farivar, R. Neal, C. Clarke, and S. Low, “Optimal inverter VAR control in distribution systems with high PV penetration,” in *Proc. IEEE Power Energy Soc. General Meeting*, 2012, pp. 1–7.

[11] N. Li, L. Chen, and S. H. Low, “Exact convex relaxation of OPF for radial networks using branch flow model,” in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun.*, 2012, pp. 7–12.

[12] H.-G. Yeh, D. F. Gayme, and S. H. Low, “Adaptive VAR control for distribution circuits with photovoltaic generators,” *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1656–1663, Aug. 2012.

[13] S. Deshmukh, B. Natarajan, and A. Pahwa, “Voltage/VAR control in distribution networks via reactive power injection through distributed generators,” *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1226–1234, Sep. 2012.

[14] M. Farivar, L. Chen, and S. Low, “Equilibrium and dynamics of local voltage control in distribution systems,” in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 4329–4334.

[15] P. Jahangiri and D. C. Aliprantis, “Distributed Volt/VAR control by PV inverters,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3429–3439, Aug. 2013.

[16] B. Zhang, A. Y. Lam, A. D. Domínguez-García, and D. Tse, “An optimal and distributed method for voltage regulation in power distribution systems,” *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1714–1726, Jul. 2015.

[17] M. Kraning et al., “Dynamic network energy management via proximal message passing,” *Found. Trends Optim.*, vol. 1, no. 2, pp. 73–126, 2014.

[18] E. Dall’Anese, H. Zhu, and G. B. Giannakis, “Distributed optimal power flow for smart microgrids,” *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1464–1475, Sep. 2013.

[19] P. Sulc, S. Backhaus, and M. Chertkov, “Optimal distributed control of reactive power via the alternating direction method of multipliers,” *IEEE Trans. Energy Convers.*, vol. 29, no. 4, pp. 968–977, Dec. 2014.

[20] “IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces,” IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)-Redline, Apr. 2018, pp. 1–227.

[21] G. Cavaro, S. Bolognani, R. Carli, and S. Zampieri, “The value of communication in the voltage regulation problem,” in *Proc. IEEE Conf. Decis. Control*, 2016, pp. 5781–5786.

[22] H. J. Liu, W. Shi, and J. Zhu, “Hybrid voltage control in distribution networks under limited communication rates,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2416–2427, May 2019.

[23] Z. Tang, D. J. Hill, and T. Liu, “Fast distributed reactive power control for voltage regulation in distribution networks,” *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 802–805, Jan. 2019.

[24] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.

[25] A. Nedić and A. Ozdaglar, “Subgradient methods for saddle-point problems,” *J. Optim. Theory Appl.*, vol. 142, no. 1, pp. 205–228, 2009.

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

[27] D. Feijer and F. Paganini, “Stability of primal–dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
[28] G. Qu and N. Li, “On the exponential stability of primal-dual gradient dynamics,” 2018, arXiv:1803.01825.

[29] M. Baran and F. Wu, “Network reconfiguration in distribution systems for loss reduction and load balancing,” IEEE Trans. Power Del., vol. 4, no. 2, pp. 1401–1407, Apr. 1989.

[30] G. Massa, G. Gross, V. Galdi, and A. Piccolo, “Dispersed voltage control in microgrids,” IEEE Trans. Power Syst., vol. 31, no. 5, pp. 3950–3960, Sep. 2016.

[31] B. A. Robbins, C. N. Hadjicostis, and A. D. Domínguez-García, “A two-stage distributed architecture for voltage control in power distribution systems,” IEEE Trans. Power Syst., vol. 28, no. 2, pp. 1470–1482, May 2013.

[32] P. Carvalho, L. Ferreira, and M. Ilic, “Distributed energy resources integration challenges in low-voltage networks: Voltage control limitations and risk of cascading,” IEEE Trans. Sustain. Energy, vol. 4, no. 1, pp. 82–88, Jan. 2013.

[33] A. Kam and J. Simonelli, “Stability of distributed, asynchronous VAR-based closed-loop voltage control systems,” in Proc. IEEE PES General Meeting Conf. Expo., Jul. 2014, pp. 1–5.

[34] A. Bidram and A. Davoudi, “Hierarchical structure of microgrids control system,” IEEE Trans. Smart Grid, vol. 3, no. 4, pp. 1963–1976, Dec. 2012.

[35] M. Farivar and S. H. Low, “Branch flow model: Relaxations and convexification—Part I,” IEEE Trans. Power Syst., vol. 28, no. 3, pp. 2554–2564, Aug. 2013.

[36] L. Gan, N. Li, U. Topcu, and S. Low, “Branch flow model for radial networks: Convex relaxation,” in Proc. 51st IEEE Conf. Decis. Control, 2012, pp. 1–8.

[37] S. H. Low, “Convex relaxation of optimal power flow—Part I: Formulations and equivalence,” IEEE Trans. Control Netw. Syst., vol. 1, no. 1, pp. 15–27, Mar. 2014.

[38] A. C. Rueda-Medina, J. F. Franco, M. J. Rider, A. Padiha-Feltrin, and R. Romero, “A mixed-integer linear programming approach for optimal type, size and allocation of distributed generation in radial distribution systems,” Electric Power Syst. Res., vol. 97, pp. 133–143, 2013.

[39] A. Garces, “A linear three-phase load flow for power distribution systems,” IEEE Trans. Power Del., vol. 5, no. 1, pp. 827–828, Jan. 2016.

[40] R. Cespedes, “New method for the analysis of distribution networks,” IEEE Trans. Power Del., vol. 33, no. 6, pp. 5832–5843, Nov. 2018.

[41] G. Qu and N. Li, “Harnessing smoothness to accelerate distributed optimization,” IEEE Trans. Control Netw. Syst., vol. 5, no. 3, pp. 1245–1260, Sep. 2018.

Guannan Qu received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2014. Since 2014, he has been a graduate student with the School of Engineering and Applied Sciences, Harvard University. His research interests include network control and optimization.

Na Li received the B.S. degree in mathematics and applied mathematics from Zhejiang University, Hangzhou, China, in 2007 and the Ph.D. degree in control and dynamical systems from the California Institute of Technology, Pasadena, CA, USA, in 2013. She is an Associate Professor with the School of Engineering and Applied Sciences, Harvard University. She was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. Her research interests include the design, analysis, optimization, and control of distributed network systems, with particular applications to cyber-physical network systems. She received National Science Foundation CAREER Award in 2016, AFOSR Young Investigator Award (2017), Office of Naval Research Young Investigator Award in 2019, Donald P. Eckman Award in 2019 among others.