Distributed Voltage Control with Communication Delays

Sindri Magnússon, Guannan Qu, and Na Li

Abstract— **The increased penetration of volatile renewable energy into distribution networks necessities more efficient distributed (VAR) voltage control. In these control algorithms each bus regulates its voltage fluctuations based on local voltage measurements and communication to other buses. Recent studies have shown that the communication between buses is necessary to solve the general voltage control problem. However, existing literature provides only synchronous voltage control algorithms, which are clearly infeasible in large networks. In this paper, we design a distributed asynchronous feedback voltage control algorithm. The main contributions of our algorithm are: 1) it only requires local communication between neighbors in the network, 2) it enforces hard reactive power limits at every iteration, 3) it converges to hard voltage limits, 4) it converges to optimal reactive power control, 5) it is provably robust to communication delays. We prove the algorithms convergence assuming linear relationship between voltage and reactive power. We simulate the algorithm using the full nonlinear AC power flow model. Our simulations show the effectiveness of our algorithm on realistic networks with both static and fluctuating loads, even in the presence of communication delays.**

I. INTRODUCTION

One of the main goals of the smart grid is to increase the penetration of renewable energy resources in future power networks [1]. This is challenging because large scale penetration of renewable energy causes faster voltage fluctuations than today's networks can manage. This has motivated much recent research on distributed voltage (VAR) control, where the network's buses regulate their voltages by adjusting reactive power injection. In particular, each bus updates its reactive power based on feedback from local voltage measurements and possibly communication from other buses.

Many of the early studies on distributed voltage control considered local controllers where each bus updates its reactive power based only on local voltage measurements [2], [3], [4], [5]. However, even though the local control may work well in some cases (such as when some of the voltage or reactive power limits are relaxed), they can fail in providing feasible solutions as illustrated in [6]. In particular, they cannot guarantee that the voltage and reactive power limits are satisfied simultaneously. Therefore, communication is necessary to solve the general voltage control problem. This has motivated studies on distributed voltage control with communication between neighbouring buses [7], [8], [9], [10]. Many of these algorithms have the drawbacks that the reactive power is feasible only in the limit or they are not provably convergent. Moreover, all of the above algorithms are synchronous, meaning that each bus has to wait to receive information from their neighbours before they can update their control action. This is clearly unfeasible in large networks, especially if the control is suppose to respond in real-time to changes in the network.

The goal of this paper is to design an asynchronous distributed algorithm for the voltage control and prove its convergence. The main contribution of our algorithm are listed as follows.

1) The algorithm is distributed: only communication between neighbours in the network is needed.
2) The algorithm ensures that hard limits on the reactive powers on each bus are satisfied at every iteration.
3) The algorithm converges to hard voltage limits of each bus.[1]
4) The algorithm convergences to the optimal reactive power control.
5) The algorithm is provably robust to communication delays, i.e., convergence results hold even in the presence of bounded delays.

To the best of our knowledge there does not exists a voltage regulation algorithm that satisfies all of the above criterions. For example, the algorithms in [4], [7], [9], [8], [10] do not satisfy criterion 3), since in these papers there is either no constraints on the reactive power or the algorithm converges to feasible reactive power only in the limit. The algorithms in [2], [3], [5] do not satisfy criterion 2) and the algorithm in [6] does not satisfy criterion 4). Moreover, most of the algorithms lack the study of robustness to communication delays, criterion 5).

The work most closely related to ours is [11], which proves the convergence of an algorithm satisfying criterion 1)-4). However, the algorithm development is different, the algorithm in [11] is based on inexact primal-dual algorithm whereas our algorithm is based on asynchronous dual decomposition. This allows us to prove the convergence of our algorithm even in the presents of communication delays.

Our algorithm finds the optimal feasible voltage control with respect to a quadratic cost on the reactive powers. We acknowledge that such cost might sometimes be artificial. However, it allows us to use the rich literature of distributed optimization to devise our distributed control algorithm. Moreover, in some cases it can be beneficial to enforce convergence to some desirable feasible points, e.g., since

*The work was supported by NSF 1608509, NSF CAREER 1553407, AFOSR YIP, ARPA-E through the NODES program, and the Harvard Climate Change Solution Funds.

The authors are with the Harvard John A. Paulson School of Engineering and Applied Sciences (Email: sindrim@seas.harvard.edu, gqu@g.harvard.edu, and nali@seas.harvard.edu).

[1]We prove the convergence assuming a linearized relationship between voltage and reactive power. However, we illustrate the convergence of the algorithm using the full nonlinear model in the simulations.

many electronics support only limited apparent power it can be desirable to minimize the square of the reactive powers.

A. Notation

The imaginary unit is denoted by $\mathbf{i}$, i.e., $\mathbf{i} = \sqrt{-1}$. The set of real, complex, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{N}$, respectively. The set of real n vectors and $n \times m$ matrices are denoted by $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$, respectively. Otherwise, we use calligraphy letters to represent sets. The superscript $(\cdot)^{\mathrm{T}}$ stands for transpose. We let $||\cdot||$ denote the 2-norms for both vectors and matrices, i.e., $||\mathbf{A}||$ is the largest singular value of the matrix $\mathbf{A}$. For a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ we let $\texttt{dist}(i,j)$, for $i, j \in \mathcal{N}$ denote the number of edges in the shortest path between the nodes i and j.

II. System Model and Problem Formulation

A. System Model: Branch Flow for Radial Networks

Consider a radial power distribution network with $N+1$ buses represented by the set $\mathcal{N}_0 = \{0\} \cup \mathcal{N}$, where $\mathcal{N} = \{1, \ldots, N\}$. Bus 0 is a feeder bus and the buses in $\mathcal{N}$ are branch buses. Let $\mathcal{E} \subseteq \mathcal{N}_0 \times \mathcal{N}_0$ denote the set of directed flow lines, so if $(i,j) \in \mathcal{E}$ then i is the parent of j. For each i, let $p_i + \mathbf{i}q_i \in \mathbb{C}$, $V_i \in \mathbb{C}$, and $v_i \in \mathbb{R}_+$ denote the complex power injection, complex voltage, and squared voltage magnitude, respectively, at Bus i. For each $(i.j) \in \mathcal{E}$, let $P_{ij} + \mathbf{i}Q_{ij} \in \mathbb{C}$, $I_{ij} \in \mathbb{C}$, and $r_{ij} + \mathbf{i}x_{ij} \in \mathbb{C}$ denote the complex power flow, current, and impedance in the line from Bus i to Bus j. The relationship between the variables can be expressed as [13], [14],

$$-p_i = P_{\sigma_i i} - r_{\sigma_i i} l_{\sigma_i, i} - \sum_{k:(i,k)\in\mathcal{E}} P_{ik}, \qquad i \in \mathcal{N}, \quad \text{(1a)}$$

$$-q_i = Q_{\sigma_i i} - x_{\sigma_i i} l_{\sigma_i, i} - \sum_{k:(i,k)\in\mathcal{E}} Q_{ik}, \qquad i \in \mathcal{N}, \quad \text{(1b)}$$

$$v_j - v_i = -2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2) l_{ij}, \qquad (i,j) \in \mathcal{E}, \quad \text{(1c)}$$

$$l_{ij} = \frac{P_{ij}^2 + Q_{ij}^2}{v_i} \qquad (i,j) \in \mathcal{E}, \quad \text{(1d)}$$

where σ_i is the parent of bus $i \in \mathcal{N}$, i.e., the unique $\sigma_i \in \mathcal{N}_0$ with $(\sigma_i, i) \in \mathcal{E}$, and $l_{ij} = |I_{ij}|^2$.

We develop our voltage control algorithm for the general nonlinear power flow in Equation (1). However, we prove the convergence of the algorithm by consider a linearied version of the above model. In particular, we consider the linear Distflow approximation of the above equations, which gives a good approximation in radial distribution networks [14]. The linear Distflow model is obtained by setting $l_{ij} = 0$ in which case Equation (1) can be written as

$$v = Xq + Rp + \mathbf{1}v_0, \qquad (2)$$

where $v = [v_1, \ldots, v_N]^{\mathrm{T}}$, $q = [q_1, \ldots, q_N]^{\mathrm{T}}$, $p = [p_1, \ldots, p_N]^{\mathrm{T}}$,

$$X_{ij} = 2 \sum_{(h,k)\in\mathcal{P}_i \cap \mathcal{P}_j} x_{hk}, \quad \text{and} \quad R_{ij} = 2 \sum_{(h,k)\in\mathcal{P}_i \cap \mathcal{P}_j} r_{hk},$$

where $\mathcal{P}_i \subseteq \mathcal{E}$ is the set of edges in the path from Bus 0 to Bus i.

B. Optimal Voltage Control

After the real power injection p is decided, voltage control is needed to tune the reactive power injection q with the goal of steering the voltage v to some feasible range $v \in [\underline{v}, \bar{v}]$. To that end, assume that the reactive power injection can be adjusted within some interval $q \in [\underline{q}, \bar{q}]$, e.g., by PV inverters.[2] For reactive power injection $q \in \mathbb{R}^N$, let $v(q) \in \mathbb{R}^N$ be the resulting voltage, i.e., the solution of Equation (1) for the real and reactive powers p and q. Ideally, we wish to find the optimal reactive power for the following optimization problem:

$$\begin{aligned} \underset{q_1,\ldots,q_n}{\text{minimize}} \quad & \sum_{i=1}^N C_i(q_i) := \frac{a_i}{2} q_i^2 + b_i q_i + c_i \\ \text{subject to} \quad & \underline{v} \le v(q) \le \bar{v} \\ & \underline{q} \le q \le \bar{q}. \end{aligned} \qquad (3)$$

where $C_i(q_i)$ is the quadratic cost of injecting q_i units of reactive power at bus i. Throughout the paper we assume that Problem (3) is feasible and $a_i > 0$ for all i. Moreover, set

$$a_{\min} := \min\{a_1, \ldots, a_N\}.$$

Remark 1: The cost function $C_i(q_i)$ ensures that the algorithm converges to the most desirable reactive power. For example, for PV generators there is generally an upper limit on the apparent power $\sqrt{p^2+q^2}$, so it can be desirable to minimize q^2 so we can support more real power. If there is no preference for what reactive powers are desired then a_i and b_i can be considered algorithm parameters. We leave it for future study to find good methods to tune a_i and b_i to increase the the performance of the algorithm.

The goal of this paper is to devise algorithms that solve the problem that are robust to communication delays.

C. Distributed Optimal Voltage Control

We will propose a distributed feedback control algorithm to solve (3). Ideally, we would like algorithms that use only local information. That is, each bus $i \in \mathcal{N}$ initializes its reactive power as $q_i(0) \in [\underline{q}_i, \bar{q}_i]$ and then updates it as follows, for iteration index $t \in \mathbb{N}$,

$$\textbf{Measurement:} \quad v_i(t) = v_i(q(t)) \qquad (4a)$$

$$\textbf{Control:} \quad q_i(t+1) = \texttt{Alg}_i^t(\texttt{Local\_Info}_i(t)), \qquad (4b)$$

where $\texttt{Alg}_i^t : \mathbb{R}^{2t} \to [\underline{q}_i, \bar{q}_i]$ and

$$\texttt{Local\_Info}_i(t) = \{q_i(0), \ldots, q_i(t), v_i(0), \ldots, v_i(t)\},$$

are, respectively, the local control algorithm and the local information available to bus i at iteration t. Unfortunately, there exist no local algorithm that is guaranteed to solve the optimization problem, due to the impossibility result in [15]. Therefore, it is necessary to include some communication

[2] The reactive power can be decomposed into $q = q^{\text{Adj.}} + q^{\text{Con.}}$ where $q^{\text{Adj.}}$ is adjustable reactive power and $q^{\text{Con.}}$ is the fixed reactive power consumption.

into the control law in Equation (4b). Such communication can be modeled as follows:

$$q_i(t+1) = \texttt{Alg}_i^t(\texttt{Local\_Info}_i(t), \texttt{Comm}_i(t)), \quad (5)$$

where $\texttt{Comm}_i(t)$ is information that bus i has received from other buses until iteration t.

In this paper, we consider algorithms in the form of equation (5) when the communicated information $\texttt{Comm}_i(t)$ at each iteration comes only from physical neighbours of node i. Moreover, there might be delays in the communications.

III. Algorithm and Main Results

We now illustrate the distributed algorithm for solving Problem (3). We discuss its main steps and convergence properties after the algorithm. We explain the intuition behind the algorithm in Section IV.

DIST-OPT: Distributed Optimal Voltage Control

STEP 1 **Initialization:** Set $t = 0$, $z_i(0) = \underline{\lambda}_i(0) = \bar{\lambda}_i(0) = \alpha_i(0) = \beta_i(0) = 0$ for $i \in \mathcal{N}$.

STEP 2 **Local Control:** Each bus $i \in \mathcal{N}$ injects into the grid the reactive power

$$q_i(t) = \left[\frac{1}{a_i}\left(z_i(t) - b_i\right)\right]_{\underline{q}_i}^{\bar{q}_i}.$$

STEP 3 **Local Measurement:** Each bus $i \in \mathcal{N}$ measures the voltage magnitude

$$v_i(t) = v_i(q(t))$$

and then updates

$$\underline{\lambda}_i(t+1) = \lceil \underline{\lambda}_i(t) + \gamma(\underline{v}_i - v_i(t)) \rceil_+ \quad (6a)$$

$$\bar{\lambda}_i(t+1) = \lceil \bar{\lambda}_i(t) + \gamma(v_i(t) - \underline{v}_i) \rceil_+ \quad (6b)$$

$$\lambda_i(t+1) = \underline{\lambda}_i(t+1) - \bar{\lambda}_i(t+1), \quad (6c)$$

where $\gamma > 0$ is a step-size parameter.

STEP 4 **Communication:** Each bus $i \in \mathcal{N}$ sends the following information to its neighbours:

- **If i has a parent:** Send to parent $\sigma(i)$ the variable

$$\alpha_i(t+1) = \lambda_i(t+1) + \sum_{j \in \mathcal{C}_i} \hat{\alpha}_j(t), \quad (7)$$

where $\mathcal{C}_i$ is the set of the children of node i. The parent $\sigma(i)$ receives the possibly delayed version

$$\hat{\alpha}_i(t+1) = \alpha_i(t+1-\tau_{i\sigma(i)}(t)).$$

- **If i has a child:** Send to each child $j \in \mathcal{C}_i$ the variable[3]

$$\beta_j(t+1) = X_{ii}\left(\lambda_i(t+1) + \sum_{r \in \mathcal{C}_i \setminus \{j\}} \hat{\alpha}_r(t)\right) + \hat{\beta}_i(t). \quad (8)$$

Each child j receives the possibly delayed version

$$\hat{\beta}_j(t+1) = \beta_j(t+1-\tau_{ij}(t)).$$

STEP 5 **Local Computation:** Each bus $i \in \mathcal{N}$ updates

$$z_i(t+1) = X_{ii}\left(\lambda_i(t+1) + \sum_{j \in \mathcal{C}_i} \hat{\alpha}_j(t+1)\right) + \hat{\beta}_i(t+1). \quad (9)$$

STEP 6 **Update Iteration Index:** $t = t+1$.

In **STEP 1** each bus initializes its parameters. For simplicity of presentation, all parameters are initialized at 0. In **STEP 2** each bus i injects reactive power into the system based on the available information in $z_i(t)$. In **STEP 3** each bus i takes a local measurement of the voltage of $v_i(q(t))$, where $v(q(t))$ is the corresponding voltage for real and reactive powers p and $q(t)$. Moreover, based on these measurements bus i also updates the parameters $\underline{\lambda}_i$, $\bar{\lambda}_i$, and λ_i according to (6).[4] In **STEP 4** each bus i communicates the parameter $\alpha_i(t+1)$ to their parent bus (cf. Equation (7)) and $\beta_j(t+1)$ to each of their children buses $j \in \mathcal{C}_i$. In **STEP 5**, each bus i updates its variable $z_i(t+1)$ based on the local information $\lambda_i(t+1)$ and $\alpha_j(t+1)$ received from each of its child's $j \in \mathcal{C}_i$ and $\beta_i(t+1)$ received from its parent. Note that the information in $\alpha_j(t+1)$ and $\beta_i(t+1)$ received by bus i delayed by τ_{ji} and $\tau_{\sigma_i i}$, respectively.

We illustrate the performance of the algorithm on the full nonlinear power flow model in Section VI. However, we can also prove the algorithms convergence when the relationship between v and q is linearized according to Equation (2).

Theorem 1: Suppose that

$$v(q) = Xq + Rp + v_0 \quad (10)$$

for fixed $p \in \mathbb{R}^N$. Moreover, suppose that there exists $q \in \mathbb{R}^N$ such that (Slater's condition):

$$\underline{q} < q < \bar{q} \quad \text{and} \quad \underline{v} < v(q) < \bar{v}, \quad (11)$$

and that the communication delays are bounded by T, i.e., $\tau_{ij}(t) \leq T$ for all $i, j, t \in \mathbb{N}$. Then for any step-size γ such that

$$\gamma \in \left(0, \frac{2a_{\min}}{La_{\min} + 4((T+1)d+1)||X||^2 N}\right), \quad (12)$$

where $d = \max_{i,j \in \mathcal{N}} \texttt{dist}(i,j)$ is the diameter of the network, following holds

$$\lim_{t \to \infty} q(t) = q^\star,$$

where $q^\star$ is the optimal solution to Problem (3) (the solution is unique since the problem is strongly convex).
We prove the theorem in Section V. But first we illustrate the intuition into why the algorithm works.

[3]If node i has no parent then set $\hat{\beta}_i(t) = 0$.

[4]We show in Section IV that $\underline{\lambda}_i$, $\bar{\lambda}_i$ are the cost (or the dual variables) on violating the voltage constraint $\underline{v} \leq v(q) \leq \bar{v}$.

IV. Algorithm Intuition

We now illustrate intuitively why the algorithm should converge to the optimal solution to problem (3). In particular, we show that the algorithm is equivalent to a dual decent method where some messages are delayed. The dual problem of the optimization problem in Equation (3) is given by

$$\begin{aligned} \underset{\boldsymbol{\lambda}=(\underline{\lambda},\bar{\lambda})}{\text{maximize}} \quad & D(\boldsymbol{\lambda}) := \min_{q\in[\underline{q},\bar{q}]} \mathcal{L}(q,\boldsymbol{\lambda}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \in \mathbb{R}_+^{2N}, \end{aligned} \tag{13}$$

where $\underline{\lambda}$ and $\bar{\lambda}$ are, respectively, the dual variable associated to the lower and upper bounds and $D : \mathbb{R}^{2N} \to \mathbb{R}$ is the dual function and $\mathcal{L}(\cdot)$ is the Lagrangian function defined as

$$\mathcal{L}(q,\boldsymbol{\lambda}) = \sum_{i=1}^{n} C_i(q_i) + \underline{\lambda}^{\mathrm{T}}(\underline{v} - v(q)) + \bar{\lambda}^{\mathrm{T}}(v(q) - \bar{v}), \tag{14}$$

where $\boldsymbol{\lambda} = (\underline{\lambda}, \bar{\lambda}) \in \mathbb{R}^N \times \mathbb{R}^N$, see Chapter 5 in [16] for details. We have the following result proved in Appendix I.

Lemma 1: The dual gradient is

$$\nabla D(\boldsymbol{\lambda}) = \begin{bmatrix} \underline{v} - v(q(\boldsymbol{\lambda})) \\ v(q(\boldsymbol{\lambda})) - \bar{v} \end{bmatrix} \tag{15}$$

where

$$q(\boldsymbol{\lambda}) = \underset{q\in[\underline{q},\bar{q}]}{\text{argmin}}\ \mathcal{L}(q,\boldsymbol{\lambda}) = \left[\Lambda^{-1}X(\underline{\lambda} - \bar{\lambda}) - b\right]_{\underline{q}}^{\bar{q}}, \tag{16}$$

where $\Lambda = \mathtt{diag}(a_1, \dots, a_N)$ and $b = [b_1, \dots, b_n]^{\mathrm{T}}$. Moreover, $\nabla D(\boldsymbol{\lambda})$ is L-Lipschitz continuous where $L = 4||X||^2/a_{\min}$ where $a_{\min} = \min_{i=1,\dots,n} a_i$.

Remark 2: Note that $\Lambda^{-1}X(\underline{\lambda} - \bar{\lambda}) - b$ in Equation (16) is the unconstrained minimizer of $L(q, \boldsymbol{\lambda})$. In general, the optimal solution of a constrained optimization problem cannot be obtained by projecting the unconstrained optimizer to the constraint set, even for quadratic problems. However, this works here because of the special structures of the matrices in our problem, see Appendix I for the details.

From the lemma we can derive a standard dual decent algorithm by setting

$$q_i(t) = \left[\frac{1}{a_i}\left(\sum_{j=1}^{n} X_{ij}(\lambda_i(t)) - b_i\right)\right]_{\underline{q}_i}^{\bar{q}_i}, \tag{17}$$

where $\lambda_i(t) = \underline{\lambda}_i(t) - \bar{\lambda}_i(t)$ and $\boldsymbol{\lambda} = (\underline{\lambda}_i, \bar{\lambda}_i)$ is updated according to Equation (6), which is equivalent to the following gradient update

$$\boldsymbol{\lambda}(t+1) = \lceil \boldsymbol{\lambda}(t) + \gamma \nabla D(\boldsymbol{\lambda}(t)) \rceil_+. \tag{18}$$

This algorithm is guaranteed to converge to the optimal solution (provided that $\gamma > 0$ is small enough) since it is just a gradient method for solving $D(\cdot)$ and the dual gradient $\nabla D(\cdot)$ is Lipschitz continuous. However, it has the downside that to calculate $q_i(t+1)$ in Equation (17) bus i needs information form every other bus in the network. This is countered in the **DIST-OPT** algorithm where only neighbour to neighbour communication is needed. This is obtained by replacing the sum $\sum_{j=1}^{n} X_{ij}(\lambda_i(t))$ in the update in Equation (17) by $z_i(t)$ which can be computed in distributed fashion, see Equation (9). In fact, $z_i(t)$ is just a delayed version of the $\sum_{j=1}^{n} X_{ij}(\lambda_i(t))$ as shown in the following lemma (proved in Appendix II).

Lemma 2: If $\tau_{ij}(t) = 0$ for all $i, j \in \mathcal{N}$ and $t \in \mathbb{N}$, then we have for all $i \in \mathcal{N}$ that

$$z_i(t) = \sum_{j=1}^{n} X_{ij}\lambda_j(t - d_{ij}), \tag{19}$$

where $\lambda_j(t) = 0$ for $t < 0$ and

$$d_{ij} = \begin{cases} 0 & \text{if } \mathtt{dist}(i,j) \le 1 \\ \mathtt{dist}(i,j) - 1 & \text{otherwise.} \end{cases} \tag{20}$$

If $\tau_{ij}(t) \le T$ for all $i, j \in \mathcal{N}$ and $t \in \mathbb{N}$ then

$$z_i(t) = \sum_{j=1}^{n} X_{ij}\lambda_j(t - \bar{\tau}_{ij}(t)), \tag{21}$$

where $\bar{\tau}_{ij}(t) \le (T+1)d$.

The lemma shows that the **DIST-OPT** algorithm is equivalent to

$$\lambda(t+1) = \lceil \lambda(t) + \gamma g(t) \rceil_+, \tag{22}$$

where $g(t)$ is an approximation of the dual gradient $\nabla D(\boldsymbol{\lambda}(t))$ using old λ values solve (17) (cf. Equation (19) and (21)). In particular, we have that

$$g(t) = \begin{bmatrix} \underline{v} - v(q(t)) \\ v(q(t)) - \bar{v} \end{bmatrix}. \tag{23}$$

We now use this interpretation of the **DIST-OPT** algorithm to prove Theorem 1.

V. Convergence: Proof Theorem 1

We now show that the algorithm converges to the optimal solution to Problem (3). The proof follows similar ideas as used in [17] to prove the convergence of asynchronous dual decomposition for flow control, but adjusted to our problem. In particular, the proof is based on the following two lemmas, proved in appendices III and IV, respectively.

Lemma 3: For all $t \in \mathbb{N}$ following holds:

$$||\nabla D(\boldsymbol{\lambda}(t)) - g(t)|| \le \frac{2N||X||^2}{a_{\min}} \sum_{\tau=t-t_0}^{t-1} ||\boldsymbol{\lambda}(\tau) - \boldsymbol{\lambda}(\tau+1)||,$$

where $t_0 = d(T+1)$ and $g(t)$ is the approximate dual gradient in Equation (23).

Lemma 4: For all $t \in \mathbb{N}$ following holds:

$$D(\boldsymbol{\lambda}(t+1)) \ge D(\boldsymbol{\lambda}(0)) + \left(\frac{1}{\gamma} - \frac{L}{2} - (T(d-1)+1)\frac{2N||X||^2}{a_{\min}}\right) \times \sum_{\tau=0}^{t} ||\boldsymbol{\lambda}(\tau+1) - \boldsymbol{\lambda}(\tau)||^2$$

In particular, from the Slaters condition in Equation (11) the duality gap is zero and, hence, if γ is chosen as in Equation (12) then

$$\sum_{\tau=0}^{\infty} ||\boldsymbol{\lambda}(\tau+1) - \boldsymbol{\lambda}(\tau)||^2 < \infty \text{ and } \lim_{t\to\infty} ||\boldsymbol{\lambda}(t+1) - \boldsymbol{\lambda}(t)|| = 0.$$

The two lemmas show that the approximate $g(t)$ converges to the true gradient $\nabla D(\boldsymbol{\lambda}(t))$ as t goes to infinity, i.e.,

$$\lim_{t\to\infty} ||\nabla D(\boldsymbol{\lambda}(t)) - g(t)|| = 0.$$

In particular, Lemma 3 shows that the distance between $g(t)$ and $\nabla D(\boldsymbol{\lambda}(t))$ is bounded by the finite sum

$$\sum_{\tau=t-t_0}^{t-1} ||\boldsymbol{\lambda}(\tau) - \boldsymbol{\lambda}(\tau+1)||,$$

times a constant factor. Lemma 4 the shows that the terms of the sum converge to zero when γ is chosen as in Equation (12), meaning that the sum also converges to 0. We now use these results to proof the theorem.

Proof of Theorem 1: We start by showing that every limit point of $\boldsymbol{\lambda}(t)$ is an optimal solution to the dual problem. Note that the sequence $\boldsymbol{\lambda}(t)$ is bounded since from Lemma 4

$$\boldsymbol{\lambda}(t) \in \{\boldsymbol{\lambda} \in \mathbb{R}_+^{2N} | D(\boldsymbol{\lambda}) \geq D(\boldsymbol{\lambda}(0))\} \quad \text{for all} \quad t \in \mathbb{N}$$

and every level sets is bounded [16, Proposition B.9].[5] Let $\boldsymbol{\lambda}^\star$ be some limit point of $\boldsymbol{\lambda}(t)$ and let $\boldsymbol{\lambda}(t_i)$ be a subsequence that converges to $\boldsymbol{\lambda}^\star$. Then we have from the continuity of $\nabla D(\cdot)$ that $\lim_{i\to\infty} \nabla D(\boldsymbol{\lambda}(t_i)) = \nabla D(\boldsymbol{\lambda}^\star)$. The gradient approximate sequence $g(t_i)$ also converges to $\nabla D(\boldsymbol{\lambda}^\star)$, since from the triangle inequality we get

$$\lim_{i\to\infty} ||\nabla D(\boldsymbol{\lambda}^\star) - g(t_i))|| \leq \lim_{i\to\infty} ||\nabla D(\boldsymbol{\lambda}^\star) - \nabla D(\boldsymbol{\lambda}(t_i))|| + \lim_{i\to\infty} ||\nabla D(\boldsymbol{\lambda}(t_i)) - g(t_i)|| = 0$$

where the second limit convergence to zero because of lemmas 3 and 4. Therefore, we have

$$\begin{aligned}\lceil \boldsymbol{\lambda}^\star + \gamma \nabla D(\boldsymbol{\lambda}^\star)\rceil_+ - \boldsymbol{\lambda}^\star &= \lim_{i\to\infty} \lceil \boldsymbol{\lambda}(t_i) + \gamma g(t_i)\rceil_+ - \boldsymbol{\lambda}(t_i) \\ &= \lim_{i\to\infty} \boldsymbol{\lambda}(t_i+1) - \boldsymbol{\lambda}(t_i) = 0,\end{aligned}$$

from Lemma 4. From the projection theorem [16, Proposition 2.1.3] we have that

$$\langle \nabla D(\boldsymbol{\lambda}^\star), \boldsymbol{\lambda}^\star - \boldsymbol{\lambda}\rangle \geq 0, \quad \text{for all} \quad \boldsymbol{\lambda} \in \mathbb{R}^{2N},$$

which implies that $\boldsymbol{\lambda}^\star$ is the optimal solution to the dual problem [16, Proposition 2.1.2].

We can now show that $q(t)$ converges to $q^\star$. The sequence $q(t)$ is bounded since it is in $[\underline{q}, \bar{q}]$. Moreover, since the function $q(\cdot)$ is continuous, see Equation (17), and from strong duality, every subsequence of $q(t)$ convergences to $q^\star = q(\boldsymbol{\lambda}^\star)$. We can conclude that $q(t)$ converges to $q^\star$.

VI. Simulations

We evaluate our algorithm **DIST-OPT** on the full nonlinear AC power flow model (1), using Matpower [19]. We do our experiments on the distribution circuit of South California Edison with a high penetration of photovoltaic (PV) generation [20].[6] In the simulation, we assume that there are Volt/Var control components at all the buses and those control components can supply or consume at most 0.4 p.u. reactive power (i.e. $\bar{q}_i = 0.4, \underline{q}_i = -0.4$). The nominal voltage magnitude is $1 p.u.$ and the acceptable range is set as $[0.95, 1.05]$ which is the plus/minus 5% of the nominate value. The parameter a_i is drawn unformly from $[1, 10]$, and b_i from $[-0.5, 0.5]$, and the delay $\tau_{ij}(t)$ is drawn uniformly from $\{0, \ldots, T\}$.

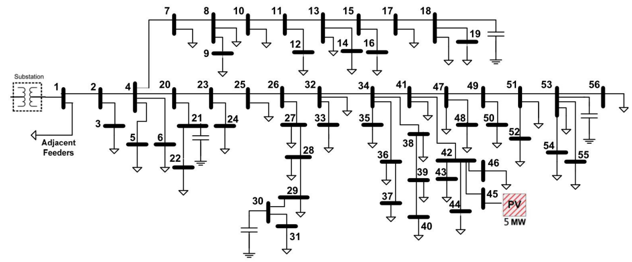

Fig. 1: Schematic diagram of two SCE distribution systems.

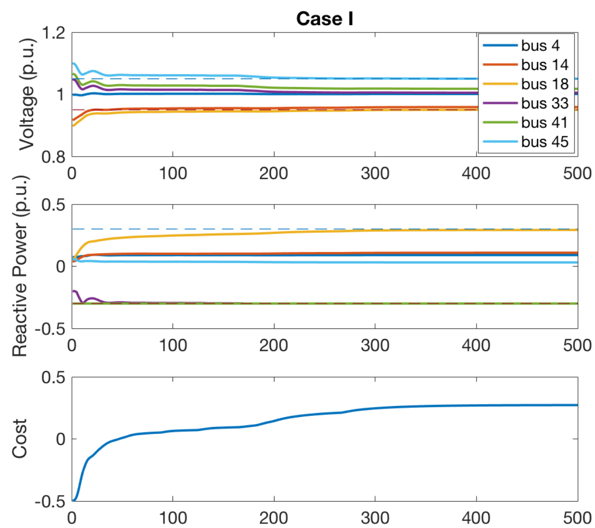

Fig. 2: Case I: zero delay, static load and PV generation.

We simulate three different cases. In all the cases, the PV generator is generating a large amount of power and some buses are heavily loaded, resulting in high voltages at some buses and low voltages at other buses. In case I, we set the maximum delay $T = 0$, i.e. there is no communication delay. In case II, we set the maximum delay $T = 10$ while other settings are the same as case I. The results of case I and case II are summarized in Figure 2 and Figure 3. It shows that our algorithm **DIST-OPT** can bring the voltage to the acceptable range and in the meanwhile not violating the capacity constraint. Also, we note that the trajectories in case II converge slower and suffer from more oscillations compared to case I, which can be attributed to the communication delays in case II.

In addition to the above two cases, we consider case III where the load and the PV generation are fluctuating

[5]Note that from Slaters condition (Equation (11)) the set of optimal solutions to the dual problem is bounded, see Lemma 1 in [18].

[6]See [20] for the network data including the line impedance, the peak MVA demand of the loads and the nameplate capacity of the shunt capacitors and the photovoltaic generation.

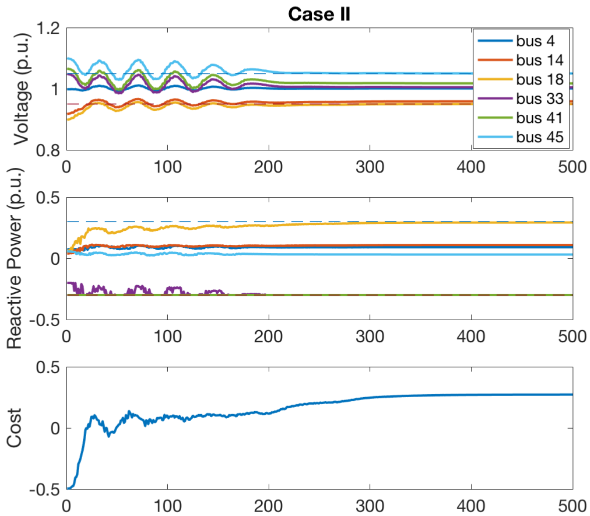

Fig. 3: Case II: positive delay, static load and PV generation.

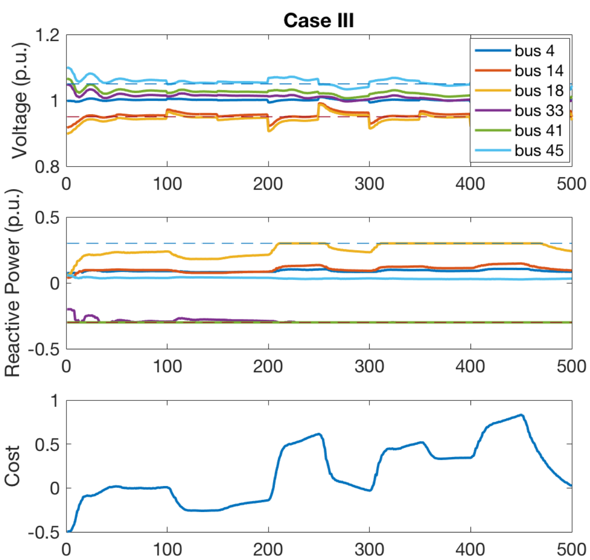

Fig. 4: Case III: positive delay, fluctuating load and PV generation.

(uniformly between 75% and 125% of their nominal values) for every other 50 controller iterations. The fluctuating speed is slower than the control algorithm updating speed, which means the controller has a certain amount of time steps to stabilize the voltages after each change in the load and the PV generation. Case III is introduced to validate the performance of **DIST-OPT** under a more realistic setting. We set $T = 4$ for case III. The simulation results are given in Figure 4. It shows that, after every change in the load and PV generation, although the voltage constraint is violated, our algorithm **DIST-OPT** can quickly bring the voltage into the acceptable range and in the mean while, not violating the capacity constraint, bringing the cost function down and maintaining it at a very low value.

References

[1] Pedro MS Carvalho, Pedro F Correia, and Luís AFM Ferreira, "Distributed reactive power generation control for voltage rise mitigation in distribution networks," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 766–772, 2008.

[2] Masoud Farivar, Lijun Chen, and Steven Low, "Equilibrium and dynamics of local voltage control in distribution systems," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 4329–4334.

[3] Pedram Jahangiri and Dionysios C Aliprantis, "Distributed volt/var control by pv inverters," *IEEE Transactions on power systems*, vol. 28, no. 3, pp. 3429–3439, 2013.

[4] N. Li, G. Qu, and M. Dahleh, "Real-time decentralized voltage control in distribution networks," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, Sept 2014, pp. 582–588.

[5] H. Zhu and H. J. Liu, "Fast local voltage control under limited reactive power: Optimality and stability analysis," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3794–3803, Sept 2016.

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

[7] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2818–2833, Nov 2013.

[8] Petr vSulc, Scott Backhaus, and Michael Chertkov, "Optimal distributed control of reactive power via the alternating direction method of multipliers," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 968–977, 2014.

[9] B. Zhang, A. Y. S. Lam, A. D. Domnguez-Garca, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1714–1726, July 2015.

[10] Sindri Magnússon, Carlo Fischione, and Na Li, "Voltage control using limited communication," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1–6, 2017.

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

[12] M. E Baran and F. F Wu, "Optimal capacitor placement on radial distribution systems," *IEEE Transactions on power Delivery*, vol. 4, no. 1, pp. 725–734, 1989.

[13] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1401–1407, Apr 1989.

[14] G. Cavraro, S. Bolognani, R. Carli, and S. Zampieri, "The value of communication in the voltage regulation problem," in *2016 54th IEEE Conference on Decision and Control (CDC)*, Des 2016.

[15] D. P. Bertsekas, *Nonlinear Programming: 2nd Edition*, Athena Scientific, 1999.

[16] S.H. Low and D.E. Lapsley, "Optimization flow control. I. basic algorithm and convergence," *Networking, IEEE/ACM Transactions on*, vol. 7, no. 6, pp. 861–874, Dec 1999.

[17] Angelia Nedic and Asuman Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.

[18] R.D. Zimmerman, C.E. Murillo-Sanchez, and R.J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *Power Systems, IEEE Transactions on*, vol. 26, no. 1, pp. 12–19, Feb 2011.

[19] Masoud Farivar, Russell Neal, Christopher Clarke, and Steven Low, "Optimal inverter var control in distribution systems with high pv penetration," in *IEEE Power and Energy Society General Meeting*, San Diego, CA, July 2012.

[20] Yurii Nesterov, *Introductory Lectures on Convex Optimization*, Springer, 2004.

Appendix I
Proof of Lemma 1

The gradient of the Lagrangian function in Equation (14) with respect to q is $\nabla\mathcal{L}(q,\boldsymbol{\lambda}) = \Lambda q + b + X(\bar{\lambda} - \underline{\lambda})$. From

Proposition 6.1.1 in [16] the dual gradient is

$$\nabla D(\lambda) = \begin{bmatrix} \underline{v} - v(q^\star(\boldsymbol{\lambda})) \\ v(q^\star(\boldsymbol{\lambda})) - \bar{v} \end{bmatrix} \tag{24}$$

where

$$q^\star(\boldsymbol{\lambda}) = \underset{q \in [\underline{q}, \bar{q}]}{\operatorname{argmin}}\ \mathcal{L}(q, \boldsymbol{\lambda}). \tag{25}$$

Therefore, to prove Equation (15) it suffices to show that $q^\star(\boldsymbol{\lambda}) = q(\boldsymbol{\lambda})$. We proof that by showing that $q(\boldsymbol{\lambda})$ is the optimal solution to the optimization problem in Equation (25) by using Proposition 2.1.2. in [16]. That is we show that $\nabla \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda})^{\mathrm{T}}(q - q(\boldsymbol{\lambda})) \geq 0$ for all $q \in [\underline{q}, \bar{q}]$. We have that

$$\nabla \mathcal{L}(q(\boldsymbol{\lambda}), \lambda)^{\mathrm{T}}(q - q(\boldsymbol{\lambda})) = \sum_{i=1}^{n} \nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \lambda)(q_i - q_i(\boldsymbol{\lambda})),$$

and hence it suffices to show that a) $\nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$ if $q_i(\boldsymbol{\lambda}) \in (\underline{q}_i, \bar{q}_i)$, b) $\nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \geq 0$ if $q_i(\boldsymbol{\lambda}) = \underline{q}_i$, and c) $\nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \leq 0$ if $q_i(\boldsymbol{\lambda}) = \bar{q}_i$. We now conclude the proof by proving a), b), and c) below.

Prove of a): Note that $q(\boldsymbol{\lambda}) = [q^\star_{\mathrm{UC}}(\boldsymbol{\lambda})]_{\underline{q}}^{\bar{q}}$ where

$$q^\star_{\mathrm{UC}}(\boldsymbol{\lambda}) := \underset{q \in \mathbb{R}^n}{\operatorname{argmin}}\ \mathcal{L}(q, \boldsymbol{\lambda}) = \Lambda^{-1} X(\underline{\lambda} - \bar{\lambda}) - b$$

is the unconstrained optimizer of $\mathcal{L}(\cdot, \boldsymbol{\lambda})$. If $q_i(\boldsymbol{\lambda}) \in (\underline{q}_i, \bar{q}_i)$ then $q_i(\boldsymbol{\lambda}) = [q^\star_{\mathrm{UC}}(\boldsymbol{\lambda})]_i$. Using that Λ is a diagonal we have

$$\begin{aligned}\nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda}) =& a_i q_i(\boldsymbol{\lambda}) + b_i + [X(\bar{\lambda} - \underline{\lambda})]_i \\ =& a_i [q^\star_{\mathrm{UC}}(\boldsymbol{\lambda})]_i + b_i + [X(\bar{\lambda} - \underline{\lambda})]_i = 0\end{aligned}$$

because $\nabla_i \mathcal{L}(q^\star_{\mathrm{UC}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$ since $q^\star_{\mathrm{UC}}(\lambda)$ is the optimizer.

Prove of b) and c): If $q_i(\boldsymbol{\lambda}) = \underline{q}_i$ then $[q^\star_{\mathrm{UC}}(\boldsymbol{\lambda})]_i \leq q(\boldsymbol{\lambda})$. Therefore, since $a_i > 0$ we have

$$\begin{aligned}0 =& \nabla_i \mathcal{L}(q^\star_{\mathrm{UC}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = a_i [q^\star_{\mathrm{UC}}(\boldsymbol{\lambda})]_i + b_i + [X(\bar{\lambda} - \underline{\lambda})]_i, \\ \leq& a_i q_i(\boldsymbol{\lambda}) + b_i + [X(\bar{\lambda} - \underline{\lambda})]_i = \nabla_i \mathcal{L}(q(\boldsymbol{\lambda}), \boldsymbol{\lambda})\end{aligned}$$

Condition c) follows from similar arguments as condition b).

Finally we show that $\nabla D(\boldsymbol{\lambda})$ is L-Lipschitz continues with $L = 4||X||^2/a_{\min}$. Take $\boldsymbol{\lambda}_1 = (\underline{\lambda}_1, \bar{\lambda}_1), \boldsymbol{\lambda}_2 = (\underline{\lambda}_2, \bar{\lambda}_2) \in \mathbb{R}^{2n}_+$, then from Equations (15) and (16)

$$\begin{aligned}||\nabla D(\boldsymbol{\lambda}_1) - \nabla D(\boldsymbol{\lambda}_2)|| \leq& 2||v(q(\boldsymbol{\lambda}_1)) - v(q(\boldsymbol{\lambda}_2))|| \\ \leq& 2||X||\ ||q(\boldsymbol{\lambda}_1) - q(\boldsymbol{\lambda}_2)|| \\ \leq& 2||X||^2 ||\Lambda^{-1}||\ ||\underline{\lambda}_1 - \underline{\lambda}_2 + \bar{\lambda}_2 - \bar{\lambda}_1|| \\ \leq& 4\frac{||X||^2}{a_{\min}}\ ||\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2||,\end{aligned}$$

where we have used the triangle inequality in the first and last inequality and the fact that $||\Lambda^{-1}|| = 1/a_{\min}$.

Appendix II
Proof of Lemma 2

We need the following definition.

Definition 1: Consider a rooted tree $\mathcal{T} = (\mathcal{N}, \mathcal{E})$.

- For each node $i \in \mathcal{N}$ we define the set of **r-th descendants** of i as follows

$$\mathcal{C}_i^r = \{j \in \mathcal{N} : \sigma^r(j) = i\}.$$

  Moreover, define the set of i and all of its descendants as follows

$$\mathcal{D}(i) = \bigcup_{k=0}^{\infty} \mathcal{C}_i^r$$

- We define the **height** of a node $i \in \mathcal{N}$ as follows

$$\texttt{Height}(i) = \max\{n \in \mathbb{N} : C_i^n \neq \emptyset\}.$$

- We define the **depth** of node $i \in \mathcal{N}$ as the distance from i to the root node $R \in \mathcal{R}$, i.e.,

$$\texttt{Depth}(i) = \texttt{dist}(i, R).$$

- We define the **most recent common ancestor** of nodes $i, j \in \mathcal{N}$ as follows

$$\texttt{MRCA}(i, j) = \underset{k \in \mathcal{A}_i \cap \mathcal{A}_j}{\operatorname{argmax}}\ \texttt{Depth}(k),$$

  where $\mathcal{A}_i = \{k \in \mathcal{N} : \sigma^r(i) = k \text{ for some } r \in \mathbb{N}\}$ is the set of ancestors of node i.

Using the notation from the definition, we have the following claims (proved in the sequel):

- **Claim 1:** For $\alpha_i(t)$ defined in Equation (7) we have

$$\lambda_i(t) + \sum_{j \in \mathcal{C}_i} \alpha_j(t) = \sum_{j \in \mathcal{D}(i)} \lambda_j(t - d_{ij}),$$

  where we set $\lambda_j(t) = 0$ for $t < 0$.
- **Claim 2:** For $\beta_i(t)$ defined in Equation (8) we have

$$\beta_i(t) = \sum_{k=1}^{\texttt{Depth}(i)} \chi_{\sigma^k(i)} \sum_{j \in \mathcal{D}(i,k)} \lambda(t - d_{ij}),$$

  where $\chi = [X_{11}, \dots, X_{NN}]$ and $\mathcal{D}(i, k) = \mathcal{D}(\sigma^k(i)) \setminus \mathcal{D}(\sigma^{k-1}(i))$ and $\lambda_j(t) = 0$ for $t < 0$.
- **Claim 3:** We have $X_{ij} = X_{kk}$ where $k = \texttt{MRCA}(i, j)$, i.e., $X_{ij} = \chi_{\texttt{MRCA}(i,j)}$.

Plug in the equations from the three claims into Equation (9) proves equations (19) and (21). We now prove the claims.

Prove of **Claim 1**: The equation follows from the following equation (proved in the sequel)

$$\begin{aligned}\alpha_i(t) =& \lambda_i(t) + \sum_{r=1}^{\texttt{Height}(i)} \sum_{j \in \mathcal{C}_i^r} \lambda_j(t - r), \\ =& \sum_{j \in \mathcal{D}(i)} \lambda_j(t - \texttt{dist}(i, j)).\end{aligned}$$

We proof the result by induction over `Height(i)`. Suppose first that `Height(i)` $= 0$, i.e., node i is a leave. Then the result follows from Equation (7). Suppose now that the equation holds for `Height(i)` $= r$. Then from Equation (7)

$$\begin{aligned}\alpha_i(t) =& \lambda_i(t) + \sum_{j \in \mathcal{C}_i} \alpha_j(t) \\ =& \lambda_i(t) + \sum_{j \in \mathcal{C}_i} \lambda_j(t - 1) + \sum_{r=2}^{\texttt{Height}(i)} \sum_{j \in \mathcal{C}_i^r} \lambda_j(t - r) \\ =& \lambda_i(t) + \sum_{r=1}^{\texttt{Height}(i)} \sum_{j \in \mathcal{C}_i^r} \lambda_j(t - r)\end{aligned}$$

where we use the induction premises in the second equality.

Prove of **Claim 2**: Writing out the recursion in Equation (8) and using that $\beta_j(t) = 0$, for all t, if i if i has no parent (i.e., if j is the root) then we get

$$\beta_i(t) = \sum_{k=1}^{\mathtt{Depth}(i)} \chi_{\sigma^k(i)} \left(\lambda_{\sigma^k(i)}(t+1-k) + \sum_{r \in \mathcal{C}_{\sigma^k(i)} \setminus \{\sigma^{k-1}(i)\}} \alpha_j(t-k) \right)$$
$$= \sum_{k=1}^{\mathtt{Depth}(i)} \chi_{\sigma^k(i)} \sum_{j \in \mathcal{D}(i,k)} \lambda_{\sigma^k(i)}(t+1-(k+\overbrace{\mathtt{dist}(\sigma^k(i),j)}^{=\mathtt{dist}(i,j)}))$$

Prove of **Claim 3**: Follows from that $X_{ij} = 2\sum_{(h,k)\in\mathcal{P}_i\cap\mathcal{P}_j} x_{hk}$ and that $\mathtt{MRCA}(i,j)$ is the end point of the intersection of the two paths $\mathcal{P}_i$ and $\mathcal{P}_j$.

Appendix III
Proof of Lemma 3

From equations (15) and (23), we have

$$\begin{aligned}||\nabla D(\boldsymbol{\lambda}(t)) - g(t)|| &= \sqrt{2}||v(q(\boldsymbol{\lambda}(t))) - v(q(t))|| \\ &\le \frac{2N||X||^2}{a_{\min}} \sum_{\tau=t-t_0}^{t-1} ||\boldsymbol{\lambda}(\tau) - \boldsymbol{\lambda}(\tau+1)||\end{aligned}$$

where the $\sqrt{2}$ factor in the first equation comes from the duplication of $v(\cdot)$ in equations (15) and (23) and the second equation comes from the following two inequalities:

$$||v(q(\boldsymbol{\lambda}(t))) - v(q(t))|| \le ||X||||q(\boldsymbol{\lambda}(t)) - q(t)||, \quad (26)$$

$$||q(\boldsymbol{\lambda}(t)) - q(t)|| \le \frac{N\sqrt{2}||X||}{a_{min}} \sum_{\tau=t-t_0}^{t-1} ||\boldsymbol{\lambda}(\tau) - \boldsymbol{\lambda}(\tau+1)||. \quad (27)$$

Equation (26) follows from the definition of $v(\cdot)$ in Equation (10) and the fact that $||X||$ is the induced matrix norm. To prove Equation (27), we recall that from Equation (16) and Lemma 2 we have

$$q_i(t) = \left[\frac{1}{a_i} \sum_{j=1}^N X_{ij} \lambda_j(t - \bar{\tau}_{ji}(t)) - b_i \right]_{\underline{q}_i}^{\bar{q}_i},$$

where $\bar{\tau}_{ij}(t) \le t_0 = d(T+1)$. Therefore, focussing on component i of the vector $q(\boldsymbol{\lambda}(t)) - q(t)$ and using the non-expansion property of the projection we get

$$\begin{aligned}|q_i(\boldsymbol{\lambda}(t)) - q_i(t)| &\le \frac{||X||}{a_{\min}} \sum_{j=1}^N |\lambda_j(t) - \lambda_j(t - \bar{\tau}_{ji}(t))| \\ &\le \frac{||X||}{a_{\min}} \sum_{j=1}^N \sum_{k=t-\bar{\tau}_{ji}}^{t} |\lambda_j(k+1) - \lambda_j(k)| \\ &\le \frac{||X||\sqrt{N}}{a_{\min}} \sum_{k=t-t_0}^{t-1} ||\lambda(k+1) - \lambda(k)|| \\ &\le \frac{||X||\sqrt{2N}}{a_{\min}} \sum_{k=t-t_0}^{t-1} ||\boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)||\end{aligned}$$

where the first inequality comes from the Cauchy-Schwarz, the triangle inequality, the fact that $||\cdot|| \le ||\cdot||_1$, and the fact that $1/a_i \le 1/a_{\min}$ for all i. The second inequality comes by using the triangle inequality. The third inequality comes by adding extra terms to the inner sum (every term is positive) so it runs from $k = t - t_0$ to t, swapping the sums, and using the equivalence of norms, i.e., $||\cdot||_1 \le \sqrt{N}||\cdot||$. The final inequality is obtained by noting that $\lambda(k) = \underline{\lambda}(k) - \bar{\lambda}(k)$ so

$$\begin{aligned}||\lambda(k+1) - \lambda(k)||^2 &\le 2(||\underline{\lambda}(k+1) - \underline{\lambda}(k)||^2 + ||\bar{\lambda}(k+1) - \bar{\lambda}(k)||^2) \\ &= 2||\boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)||^2.\end{aligned}$$

Equation (27) can now be obtained by using the equivalence of the $||\cdot||_\infty$ and $||\cdot||$ norms as in the prove of Equation (26).

Appendix IV
Proof of Lemma 4

Set $\Delta(k) = \boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)$. From the convexity of $-D(\cdot)$ we have [**?**, Theorem 2.1.5]

$$\begin{aligned}-D(\boldsymbol{\lambda}(t+1)) &\le -D(\boldsymbol{\lambda}(t)) - \langle \nabla D(\boldsymbol{\lambda}(t)), \Delta(t)\rangle + \frac{L}{2}||\Delta(t)||^2 \\ &= -D(\boldsymbol{\lambda}(t)) - \langle \nabla D(\boldsymbol{\lambda}(t)) - g(t), \Delta(t)\rangle \\ &\quad - \langle g(t), \Delta(t)\rangle + \frac{L}{2}||\Delta(t)||^2 \\ &\le -D(\boldsymbol{\lambda}(t)) + ||\nabla D(\boldsymbol{\lambda}(t)) - g(t)||\,||\Delta(t)|| \\ &\quad - \left(\frac{1}{\gamma} - \frac{L}{2}\right)||\Delta(t)||^2,\end{aligned}$$

where in the last inequality we have used $\frac{1}{\gamma}||\Delta(t)||^2 \le \langle g(t), \Delta(t)\rangle$, which is obtained by noting that $\boldsymbol{\lambda}(t+1) = \lceil \boldsymbol{\lambda}(t) + \gamma g(t) \rceil_+$ (Equation (22)) and hence from the projection theorem in [16, Lemma 2.1.3 (b)] we have

$$\begin{aligned}0 &\ge \langle \boldsymbol{\lambda}(t) + \gamma g(t) - \boldsymbol{\lambda}(t+1), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}(t+1)\rangle \\ &= -\gamma\langle g(t), \Delta(t)\rangle + ||\Delta(t)||^2.\end{aligned}$$

Expanding further by using Lemma 3 we get

$$\begin{aligned}-D(\boldsymbol{\lambda}(t+1)) &\le -D(\boldsymbol{\lambda}(t)) - \left(\frac{1}{\gamma} - \frac{L}{2}\right)||\Delta(t)||^2 \\ &\quad + \frac{2N||X||^2}{a_{\min}} \sum_{k=t-t_0}^{t-1} ||\Delta(k)||\,||\Delta(t)|| \\ &\le -D(\boldsymbol{\lambda}(t)) - \left(\frac{1}{\gamma} - \frac{L}{2}\right)||\Delta(t)||^2 \\ &\quad + \frac{2N||X||^2}{a_{\min}} \sum_{\tau=t-t_0}^{t} ||\Delta(k)||^2,\end{aligned}$$

where the final inequality is obtained by using the fact that for any $s_{t-t_0}, \dots, s_t \in \mathbb{R}_+$ it holds that $\sum_{k=t-t_0}^{t-1} s_k s_t \le \sum_{k=t-t_0}^{t} s_k^2$. If we sum over t we get

$$\begin{aligned}-D(\boldsymbol{\lambda}(t+1)) &\le -D(\boldsymbol{\lambda}(0)) - \left(\frac{1}{\gamma} - \frac{L}{2}\right)\sum_{k=0}^{t}||\Delta(k)||^2 \\ &\quad + \frac{2N||X||^2}{a_{\min}} \sum_{k_1=0}^{t} \sum_{k_2=\tau_1-t_0}^{k_1} ||\Delta(k)||^2, \\ &\le -D(\boldsymbol{\lambda}(0)) - \left(\frac{1}{\gamma} - \frac{L}{2} - (t_0+1)\frac{2N||X||^2}{a_{\min}}\right) \\ &\quad + \sum_{\tau=0}^{t} ||\boldsymbol{\lambda}(\tau+1) - \boldsymbol{\lambda}(\tau)||^2,\end{aligned}$$

which concludes the proof.