

**DISTRIBUTED REAL-TIME VOLTAGE REGULATION
IN DISTRIBUTION NETWORKS**

by

XINYANG ZHOU

B.S., Zhejiang University, China, 2012

M.S., University of Colorado, Boulder, 2016

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
College of Engineering and Applied Science

2018

This thesis entitled:
Distributed Real-Time Voltage Regulation in Distribution Networks
written by Xinyang Zhou
has been approved for the College of Engineering and Applied Science

Prof. Lijun Chen

Prof. Sangtae Ha

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Zhou, Xinyang (Ph.D., Telecommunication)

Distributed Real-Time Voltage Regulation in Distribution Networks

Thesis directed by Prof. Lijun Chen

The increasing penetration of renewable and distributed energy resources (DERs) in distribution networks call for fast and efficient distributed voltage regulation algorithms. This thesis first studies the existing local Volt/VAR control and designs new local algorithms with less restrictive convergence conditions and better voltage regulation. Meanwhile, unlike the traditional assets owned and managed by utility companies, the customer-owned DERs are not necessarily subject to the control of network operators unless properly incentivized. This thesis then investigates the joint design of distributed control and incentive mechanisms for managing DERs by introducing a market-based voltage regulation framework and extending it to a real-time setting with both continuous and discrete decision variables as well as device dynamics under time-varying operating conditions. The resulting randomized distributed algorithm admits asynchronous implementation in practical systems, and its performance is analytically characterized as well as numerically evaluated.

献给我的父母.

To my parents.

Acknowledgements

Frist and foremost, I would like to express my sincere gratitude to my advisor Prof. Lijun Chen for his enormous academic and financial support during my PhD study of the past 64 months with countless invaluable advices for my research, as well as my life.

My special thanks to Dr. Emiliano Dall’Anese for two significantly productive summers at NREL, and for always being encouraging and taking care of me.

Thank the rest of my Committee, Prof. Frank Barnes, Prof. Dirk Grunwald, Prof. Sangtae Ha, and Prof. Sriram Sankaranarayanan, for your precious time and helpful suggestions to improve my work.

Thank my group members and my co-authors, Zhiyuan Liu, Fan Shen, Guohui Ding, Wanshan Yang, Masoud Farivar, Kyri Baker, Khaled Alanezi, and Andrea Simonetto, for your inspirational discussions and collaborations.

Thank my friends in and out of Colorado: Ning Gao, Yuan Sui, Ke Ma, Simeng Chen, Tong Shen, Mengyuan Wang, Qingxin Zhang, Hang Yin, Yuanzhe Zhang, Tao Jiang, Hansu Gu, Zhifu Pei, Zhenhua Wang, Wenqi Zhang, Zhanan Zou, Ibrahim Ayad, Jane Willborn, Carina Kee, Witt Keller, Zhuan Zhu, Ming Chen, Junfei Lv, Yanxin He, Qifeng Pan, Guang Lin, Yue Xu, Yuting Wang, etc. Your company has been most endearing and heartwarming to me.

Last but not least, thank my supporting family: my grandparents, and especially my parents Mr. Weiguo Zhou and Ms. Yan Zhang for raising me with their unconditional love and being extremely supportive of me following my dreams. Thank you!

Contents

Chapter

1	Introduction	1
1.1	Voltage Regulation in Distribution Networks and New Challenges	1
1.1.1	Traditional Voltage Regulation in Distribution Networks	2
1.1.2	Increasing Penetration of Renewable Energy and DERs	3
1.2	Inverter-Based Voltage Regulation	4
1.2.1	Inverter-Based Local Volt/VAR Control	5
1.3	Market-Based Distributed Voltage Regulation	7
1.4	Thesis Outline	9
2	System Model and Related Works	10
2.1	Power Flow Model for Distribution Networks	11
2.2	Inverter Model and Local Volt/VAR Control	14
2.2.1	Inverter Model	14
2.2.2	Local Volt/VAR Control	15
2.3	Related Works	17
2.3.1	Reverse Engineering	18
2.3.2	Equilibrium Characterization	18
2.3.3	Dynamics Analysis	20
2.3.4	(Sub)gradient Algorithm	22

2.4	Motivation for Improved Design	24
3	Local Volt/VAR Control Design and Analysis	25
3.1	Local Volt/VAR Control with Better Convergence	26
3.1.1	Pseudo-Gradient Algorithm Design	26
3.1.2	Dynamics Analysis	26
3.2	Local Volt/VAR Control with Better Voltage Regulation	28
3.2.1	Algorithm Design	28
3.2.2	Equilibrium Characterization	28
3.2.3	Reverse Engineering	29
3.2.4	Dynamics Analysis	30
3.3	Numerical Examples	31
3.3.1	Equilibrium	32
3.3.2	Dynamics	34
3.4	Conclusion	37
4	Distributed Voltage Regulation with Continuous Variables	38
4.1	Preliminaries and System Model	39
4.1.1	Network Model	39
4.1.2	Problem Setup	42
4.2	Incentive-Based Distributed Algorithm	44
4.2.1	Convex Reformulation	45
4.2.2	Distributed Algorithm	47
4.2.3	Performance Analysis with Nonlinear Power Flow	52
4.3	Numeric Examples	55
4.3.1	Simulation Setup	55
4.3.2	Iterative Algorithm	56
4.4	Conclusion	58

5	Distributed Voltage Regulation with Discrete Variables	59
5.1	System Modeling	60
5.1.1	Node and Device Model	60
5.1.2	Problem Formulation	62
5.2	Distributed Algorithm Design	65
5.2.1	Convex Reformulation and Signal Design	65
5.2.2	Two Timescales and Iterative Algorithm	66
5.2.3	Performance Analysis with Diminishing Stepsize	68
5.2.4	Distributed Stochastic Dual Algorithm	72
5.3	Numerical Examples	73
5.3.1	Convergence and Variance	73
5.3.2	Voltage Regulation	74
5.4	Conclusion	75
6	Online Distributed Voltage Regulation	77
6.1	System Model and Problem Formulation	78
6.1.1	Network Model	78
6.1.2	Node and Device Models	78
6.1.3	Problem Formulation	81
6.2	Distributed Stochastic Dual Algorithm	85
6.2.1	Exact Convex Relaxation	85
6.2.2	Offline Stochastic Dual Algorithm	86
6.2.3	Distributed Implementation and Receding Horizon Optimization	90
6.3	Online and Asynchronous Optimization	92
6.3.1	Online Asynchronous Algorithm	93
6.3.2	Performance Analysis	94
6.4	Numerical Examples	97

6.4.1	Simulation Setup	97
6.4.2	Online Asynchronous Distributed Algorithm	99
6.5	Conclusion	100
7	Other Works	103
7.1	Local Volt/VAR Control with Nonlinear Power Flow Model: A Game-Theoretic Perspective	105
7.1.1	Introduction	105
7.1.2	Local Volt/VAR Control with Nonlinear Power Flow Model	105
7.1.3	Voltage Control Game	108
7.1.4	Numerical Examples	110
7.1.5	Conclusion	113
7.2	Reverse Engineering and Convex Relaxation in Networks of Coupled Oscillators	114
7.2.1	Introduction	114
7.2.2	System Model	115
7.2.3	Reverse Engineering and Synchronization	118
7.2.4	Implication on Convex Relaxation	126
7.2.5	Conclusion	127
7.3	Demand Shaping in Cellular Networks	128
7.3.1	Introduction	129
7.3.2	Related Work and Issues	131
7.3.3	System Model and Problem Formulation	133
7.3.4	Offline Demand Shaping Algorithm	136
7.3.5	Online Demand Shaping Algorithm	143
7.3.6	Numerical Examples	148
7.3.7	Conclusion	154
	Bibliography	155

Tables

Table

3.1	Network Parameters of the SCE Circuit: Line impedances, peak spot load KVA, Capacitors and PV generation's nameplate ratings.	33
7.1	Traffic/Application types and examples.	131

Figures

Figure

1.1	Electricity generation, transmission, and distribution.	1
1.2	Solar PV Global Capacity and Annual Additions, 2006–2016 [101].	2
1.3	Growth in global renewable energy compared to total final energy consumption, 2004–2014 [101].	3
2.1	$\mathcal{L}_i \cap \mathcal{L}_j$ for two arbitrary buses i, j in the network and the corresponding mutual voltage-to-power-injection sensitivity factors R_{ij}, X_{ij}	13
2.2	From left to right: (left) piecewise Volt/VAR control curve based on the piecewise linear function (2.11), (middle) its inverse function, and (right) the corresponding cost function (2.15), i.e., the integral of the inverse function.	17
3.1	Circuit diagram for SCE distribution system.	32
3.2	Equilibrium voltage versus the α_i value: As α_i increases, the equilibrium voltage v_i^* deviates less from the nominal value.	33
3.3	Different possible equilibria of \mathcal{D}_4 compared with single equilibrium of \mathcal{D}_1 – \mathcal{D}_3 . Notice that voltages of 5 controlled buses are all within the 0.98 p.u.–1.02 p.u. deadband, despite that those of other buses may fall out.	34
3.4	Evolution of voltage of the dynamical system \mathcal{D}_1 with different slopes of the piecewise linear control function: Voltage does not converge when the (absolute) slope of the control function become too large (when $\alpha_i > 26$ in this example).	35

3.5	Evolution of voltage of the dynamical systems \mathcal{D}_2 and \mathcal{D}_3 with $\alpha_i = 27$: Convergence is ensured with small enough stepsizes.	35
3.6	Convergence of the dynamical system \mathcal{D}_4 to one of the equilibria with small enough stepsize γ_4 , regardless of the choice of control function.	36
3.7	Convergence of the dynamical system \mathcal{D}_2 to within a small neighborhood of the equilibrium.	37
3.8	Convergence of the dynamical system \mathcal{D}_3 to the equilibrium.	37
4.1	IEEE 37-node feeder. The boxes represent PV systems. The red nodes are the ones analyzed in the numerical example.	56
4.2	Convergence of the distributed algorithm with increasing step size ε_1 and fixed step size ε_2	57
4.3	Controlled and uncontrolled voltages at all nodes at noon.	57
5.3	Uncontrolled voltage and controlled voltage under scenario 2) with 95% confidence interval.	74
5.1	The running average of the voltage converges to the optimal voltage of the relaxed problem.	76
5.2	Different voltage variance of three different scenarios.	76
6.1	Illustration of the online asynchronous distributed algorithm.	93
6.2	Profiles of loads and power available from the PV systems. The average load profile is marked in blue.	99
6.3	Power output of one arbitrary PV inverter.	101
6.4	Temperature and power status of one arbitrary A/C.	101
6.5	SOC and power status of one arbitrary battery.	101
6.6	Temperatures of 375 households under control by A/Cs.	102
6.7	SOCs of 375 batteries under control and signals.	102
6.8	Uncontrolled and controlled voltages.	102
7.1	(Left) Voltage values of all buses as reactive power injection from Bus 26 changes; (right) voltage values of Bus 10 in both linear and non-linear models as reactive power injection from Bus 26 changes.	111

7.2	As the values of α_i increase from (a) to (d), we have slower convergence speed and finally reach a non-convergent result, when the values of α_i get too large.	111
7.3	North America smartphone web browsing activity in one day [56].	130
7.4	Strategy to calculate the gap between the equilibrium of the On-DS algorithm and that of Off-DS algorithm.	147
7.5	Base traffic: the average (blue/dotted) and a “real” trace (red/solid).	150
7.6	Repetitive experiments show that a number of 20 to 30 iterations give a satisfying result in terms of convergence.	150
7.7	Base traffic prediction error has little impact on online algorithms with updated base traffic prediction.	153
7.8	Increasing penetration level of deferrable traffic does not influence the relative gap of online algorithms.	153

Chapter 1

Introduction

Known as one of the most complex infrastructures that human beings have ever created, the electric power network is essential to the functioning of our society and daily life. The traditional electric power system mainly comprises of the generators that supply the power, the transmission system that carries the power from the generating centers to the load centers, and the distribution system that eventually delivers the power to nearby homes and industries; see Fig. 1.1.

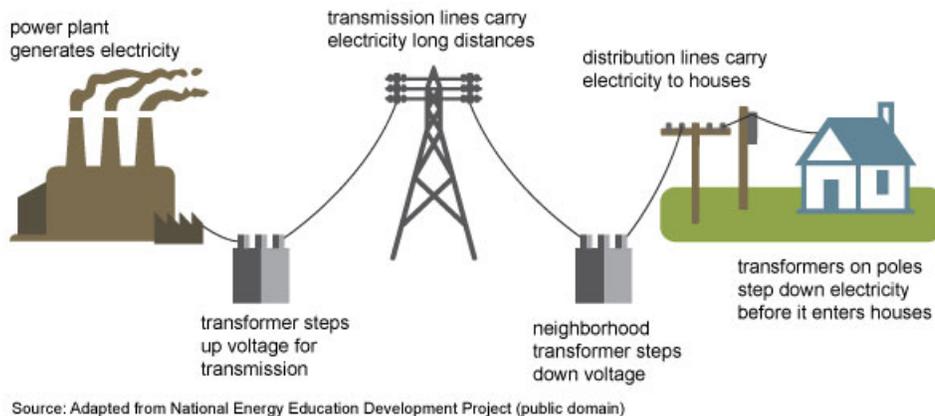


Figure 1.1: Electricity generation, transmission, and distribution.

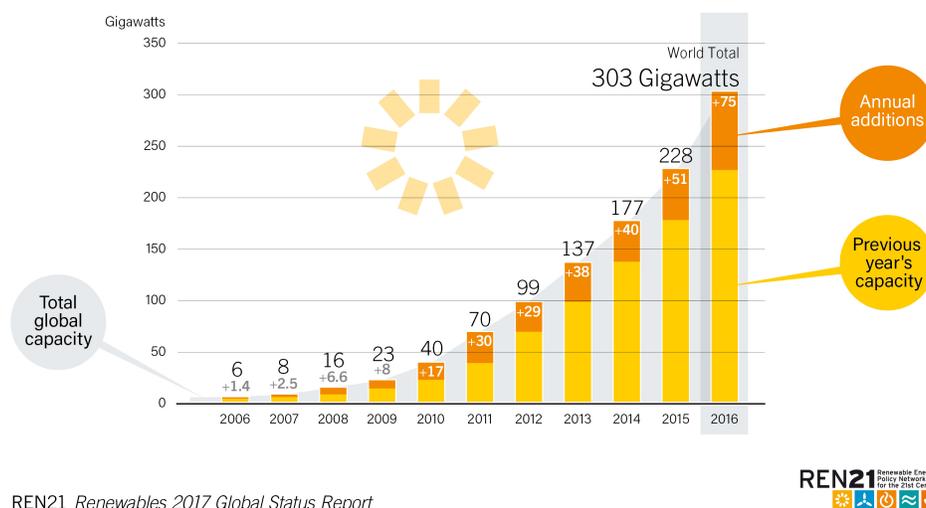
1.1 Voltage Regulation in Distribution Networks and New Challenges

Since distribution networks are directly connect to end users, keeping their voltages within specific levels is extremely important. Both overvoltages and undervoltages are problematic to end users by caus-

ing improper or inefficient operation of equipments. Moreover, overvoltages may also impair equipments through insulation damage and incur higher losses in transformers due to higher magnetizing currents, while undervoltages can overheat induction motors [108]. In US, voltage regulation in distribution networks usually follows ANSI voltage standards (ANSI C84.1-2011), e.g., maintaining the voltage magnitudes at users' meters within $\pm 5\%$ of the nominal voltage value.

1.1.1 Traditional Voltage Regulation in Distribution Networks

Voltage fluctuations are usually attributable to unbalanced power supply and demand. Traditionally, given predictable and relatively slow changes in power demand, voltage regulation is realized by tuning the outputs of the generators to match the demand, load tap changers (LTC) to change the turns ratio of transformers, and capacitor banks to inject/consume reactive power to adjust voltages [121]. Traditional voltage regulation in distribution systems is usually conducted in a centralized way a few times per day [17, 18].



REN21 *Renewables 2017 Global Status Report*

Source: IEA PVPS.

Figure 1.2: Solar PV Global Capacity and Annual Additions, 2006–2016 [101].

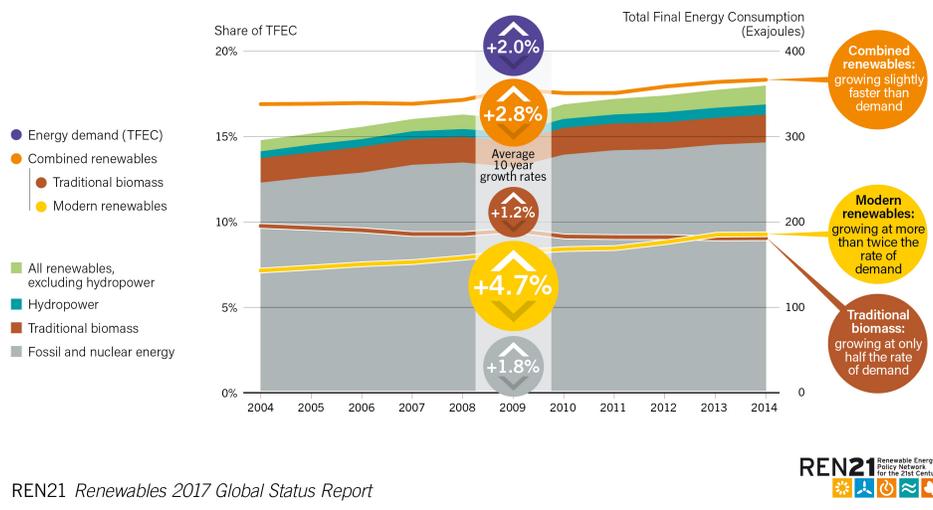


Figure 1.3: Growth in global renewable energy compared to total final energy consumption, 2004–2014 [101].

1.1.2 Increasing Penetration of Renewable Energy and DERs

Power systems have been operating under traditional control paradigms for more than a century. However, in recent year, profound evolution from two aspects is taking place in power systems, especially in distribution networks, bringing about enormous and urgent control challenges.

1.1.2.1 Renewable Energy and Real-Time Control

In pursuit of sustainable and environmentally friendly energy resources, people all over the world are implementing a large amount of renewable energy resources like photovoltaic (PV) and wind generations. The growing new investment upon renewable energy has reached a historically-high \$286 billion in 2015 [89, 101], and the cost of renewable energy has dropped significantly making it more accessible to ordinary customers [58]. As a result, the capacity and penetration of renewable energy resources have been increasing substantially for the last decade [101]; see Fig. 1.3 and Fig. 1.2.

However, the intermittent nature of the renewable energy causes frequent and rapid voltage fluctuations (at the timescale of seconds) in the distribution system, considerably beyond the reach of traditional

voltage regulation methods with capacitor banks and under load tap changers that operate at the timescale of hours. This naturally calls for a real-time control strategies that are able to react fast enough to keep up with the rapid fluctuations produced by renewable energy resources.

1.1.2.2 DERs and Market-Based Distributed Control

In the meantime, power grids are also experiencing an increasing flexibility in control on both supply and demand sides thanks to growing penetration of distributed energy resources (DERs) on the distribution level, including roof-top photovoltaic (PV) panels, electric vehicles (EV), smart batteries, thermostatically controlled loads (TCLs, e.g., water heaters, air-conditioners (A/Cs), etc.), and other responsive loads, etc. While these flexible or controllable devices can potentially provide ancillary services for the grid [38], coordinating a large number of such devices with various dynamics and constraints to achieve network-wide objectives such as voltage regulation, frequency control and economic efficiency is extremely challenging.

Moreover, unlike the traditional assets owned and managed by utility companies, the mass customer-owned devices are not necessarily subject to the control of network operators unless properly incentivized. This calls for the joint design of distributed control and incentive/market mechanisms to bring self-interested customers into the control loop so that network-wide objectives and constraints can be achieved by inducing the desired customer behaviors through proper incentives.

1.2 Inverter-Based Voltage Regulation

Even though the current IEEE Standard 1547 [111] requires distributed generation to operate at unity power factor, inverters can readily adjust real and reactive power outputs to stabilize voltages and cope with fast time-varying conditions. Indeed, the IEEE Standards group is actively exploring a new inverter-based Volt/VAR control. Unlike the capacity banks or tap changers, inverters can push and pull reactive power much faster, in a much finer granularity and with low operation costs. They will enable real-time distributed Volt/VAR control that is needed for the future power grid with a large number of renewable and DERs.

Inverter-based voltage regulation has been studied extensively in literature. Related work largely fall into the following categories:

(1) *Centralized control*: By collecting all the required information and computing a global optimal power flow (OPF) problem, a central controller determines an optimal set point for the whole network [43, 65, 71, 82, 83]. Centralized control can incorporate general objectives and operational constraints, but suffers from huge communication and computation costs when the size of the problem is large. For the same reasons, centralized control usually cannot provide fast real-time control.

(2) *Distributed control*: For OPF problems of certain structures, one can design algorithms to distribute computation with coordinative communication, which is conducted either between a central controller and agents in a hierarchical way, e.g., [13, 29, 37, 66, 77, 84, 136, 138], or among neighborhoods of individual agents without a central controller, e.g., [19, 85, 97, 107, 115, 127]. Distributed control provides a scalable way to solve a global OPF problem and can therefore provide fast real-time control.

(3) *Local control*: Based on only local information such as the voltage, local control can provide fast response without any communication within or among control hierarchies; see, e.g., [60, 102, 110, 117, 134, 142, 143].

We also refer to [92] for a recent comprehensive survey of distributed optimization and control in power systems.

1.2.1 Inverter-Based Local Volt/VAR Control

In Chapter 2–3 of this thesis, we focus on analysis and design of inverter-based local Volt/VAR control. Based on local information and reactive power control, local Volt/VAR control admits simple implementation without communication with other control nodes or central coordinator. However, characterization of its performance, especially from a global perspective as well as its dynamic properties, is challenging. In the extensive literature of local Volt/VAR control, many works such as [10, 117] focus on numerical performance evaluation only and lack analytical characterization, while some other works such as [60, 102] characterize stability analytically, but do not portray its performance from a global perspective. There are also works that provide rigorous performance analysis but are subject to special control functions; e.g., [128, 143] study linear control functions without deadband, while [110] focuses on quadratic control functions.

Given general monotone control functions, previous study [44] has reverse engineered a non-incremental local control together with the linearized power flow model as a distributed algorithm for solving a well-defined convex optimization problem that seeks a trade-off between minimizing the reactive power provisioning cost and minimizing the cost of voltage deviation from nominal value. As will be shown, in order to achieve better voltage regulation result, control functions of steeper slopes are preferred. However, convergence condition of the non-incremental control prohibits control functions of large derivatives. To break out of this dilemma, an incremental local control based on (sub)gradient algorithm for solving the same convex optimization problem is then proposed. It has been proved that given control functions of any finite derivatives, the (sub)gradient algorithm converges to its unique equilibrium point for most situations. However, the subgradient nature of this algorithm prevents it from convergence when the equilibrium point is close to the non-differentiable point in the cost function.

This motivates us to design improved local Volt/VAR control algorithms. We propose an incremental local control based on the pseudo-gradient algorithm for solving the same optimization problem as before. Moreover, we have proved that, given any control functions, this new design converges with appropriately chosen stepsize to its unique equilibrium regardless of the non-differentiable points. Following that, we design another incremental local control to assure that the stabilized voltages stay within specified bounds corresponding to the deadband in control functions given sufficient reactive power supply. These two new designs present not only better voltage regulation results, but also superior convergence properties and simpler implementation.

The convergence proofs of the new designs are challenging. For the pseudo-gradient algorithm, the Lyapunov-function-based analysis is difficult because the objective function that it solves may not be differentiable due to the deadband in control functions, and that the operator based on pseudo-gradient algorithm is generally not monotone; norm-based contraction mapping analysis does not render satisfying results either. Nevertheless, we prove the convergence by studying the Jacobian matrix of the equivalent mapping of the dynamics. We first show that the eigenvalues of the *asymmetric* Jacobian matrix are all real and non-positive. Then by Gershgorin circle theorem, we prove that given small enough stepsize, the magnitudes of the eigenvalues are all smaller than 1, and ergo the pseudo-gradient algorithm converges. We

also utilize similar approaches based on Jacobian matrix to show the convergence for the second design of strict voltage regulation.

1.3 Market-Based Distributed Voltage Regulation

Considering that the power system is not only an engineering problem, but also a market problem, we study joint distributed control and incentive/market mechanisms in Chapter 4–6.

There is a lot of effort on market-based control algorithms for tapping and coordinating the customer-owned DERs; see, e.g., [79, 86, 91, 123] for demand management, [80, 138] for voltage regulation, and [122] for frequency regulation. This thesis focuses on market-based distributed voltage regulation. The goal is to incentivize customers to provide ancillary services to the grids based on maximizing their own economic benefits and performance objectives [86, 91, 123]. For example, customers may be incentivized to adjust the output powers of DERs or the power consumption of controllable loads in real time to aid voltage regulation [80], control the aggregate network demand [79], and follow regulating signals [122].

In particular, we first consider a social welfare optimization problem that captures the operational and economic objectives of both network operator and customers as well as the voltage constraints, and design an optimization framework based on a primal-dual gradient algorithm such that the network operator and customers pursue the given operational and economic objectives while concurrently ensuring that the voltages are within the prescribed limits.

Since the framework only considers continuous decision variables, there are a few important limitations. Notably, discreteness in decision variables for certain devices such as A/Cs and batteries is mostly ignored. Discrete decision variables make the problem non-convex, for which in general no efficient algorithms exist. Considering these, we then propose a practical stochastic dual algorithm to address the problem of discrete decision variables and device dynamics, and further we develop an online implementation of the stochastic optimization approach. Specifically, we formulate a general multi-period social welfare maximization problem (minimum cost problem, in terms of minimization) with both continuous and discrete decision variables as well as device dynamics for managing DERs, and introduce a convex relaxation of the problem by replacing discrete feasible sets with their convex hulls. We then propose a distributed stochastic

algorithm that recovers discrete decision variables randomly according to convex combination coefficients from the dual gradient algorithm for the relaxed problem. The resulting algorithm is distributed, where the self-interested customers update their devices' power setpoints based on local constraints and individual cost functions for given incentive signals and the network operator updates the incentive signals based on network-wide operational constraints.

The convergence of the dual gradient method can be hard to characterize, as the dual function may be non-smooth and not strongly concave. The additional stochasticity of our algorithm makes it more challenging. In literature diminishing stepsizes are usually assumed to be necessary for convergence (see, e.g., [30, 98, 100, 140, 144]), which may not be practical in, e.g., a real-time and/or asynchronous setting. Nevertheless, we leverage recent insights in the dual method to characterize the convergence of the proposed stochastic dual algorithm with *constant* stepsizes. To the best of our knowledge, this is a first convergence characterization of its kind for the dual method applied in power systems.

Notice that, due to intermittent renewal generations and fluctuating uncontrollable loads, the operating condition of power grid may change at a fast timescale, which allows only a few iterations of the above-mentioned algorithm. Moreover, different DER devices may be featured with different timescales in control; e.g., devices with continuous decision variables can update power setpoints at relatively fast timescales, while those with discrete decision variables may only be able to update at relatively slow timescales. We therefore extend the proposed stochastic dual algorithm to a practical online realtime setting where 1) during each timeslot the algorithm can only run one or a few iterations in order to track the time-varying "optimal" operating point and 2) devices may update power setpoints asynchronously at different times. Also notice that, while we use a linearized power flow model to guide tractable algorithm design, our realtime algorithm will leverage the measured values of relevant electrical variables on the power system to account for the nonlinear power flows as well as reduce communication overhead. The resulting real-time feedback receding horizon control (RHC) algorithm provides a general online stochastic optimization algorithm for coordinating networked DERs with discrete power setpoints and dynamics to meet operational and economic objectives and constraints. We further characterize its convergence analytically and evaluate its performance numerically.

1.4 Thesis Outline

The rest of this thesis is structured as follows. Chapter 2 models the distribution system, and introduces related works including reverse engineering of the local Volt/VAR control as an optimization problem and a (sub)gradient algorithm for solving it. Chapter 3 proposes two of our new designs that improve the convergence and the voltage regulation properties with simpler implementation. Chapter 4 presents an incentive-based distributed voltage control framework by exact convex relaxation with continuous decision variables only. Chapter 5 extends the results of Chapter 4 to discrete decision variables and device dynamics and solves the problem by a stochastic dual algorithm. Chapter 6 advances the results in Chapter 5 to an online optimization and control of various DERs with mixed decision variables and device dynamics in a real-time setting. Chapter 7 introduces some other works of mine including analysis of local Volt/VAR control with nonlinear power flow model in Section 7.1, a new perspective of frequency synchronization of coupled oscillators through reverse engineering in Section 7.2, and demand shaping algorithms in cellular networks in Section 7.3.

Chapter 2

System Model and Related Works

This chapter first introduces nonlinear power flow model for distribution networks and its linearization. We then present preceding works on local Volt/VAR control including (1) a non-incremental local control together with the linearized power flow model that is reverse-engineered as a distributed algorithm for solving a well-defined convex optimization problem, and (2) an incremental local control by applying (sub)gradient algorithm to solve the same optimization problem for less restricted voltage regulation and convergence conditions. As discussed later this chapter, the presented design and analysis motivate improved local Volt/VAR control strategies for more strict voltage regulation, better convergence properties and less implementation complexities, which will be proposed next chapter. Related works of this chapter have been initially published in [44] and [45], and later improved in [141].

Main Notation For Chapter 2–3

\mathcal{N}	set of buses excluding bus 0, $\mathcal{N} := \{1, \dots, n\}$
\mathcal{L}	set of power lines
\mathcal{L}_i	set of lines from bus 0 to bus i
p_i^c, q_i^c	real, reactive power consumption at bus i
q_i^g, q_i^s	real, reactive power generation at bus i
P_{ij}, Q_{ij}	real and reactive power flow from i to j
r_{ij}, x_{ij}	resistance and reactance of line (i, j)
R, X	resistance and reactance matrices
v_i	magnitude of complex voltage at bus i
ℓ_{ij}	squared magnitude of complex current of line (i, j)
$\beta(j) \subset \mathcal{N}$	set of all descendants of bus j , $\beta(j) = \{i \mathcal{L}_j \subseteq \mathcal{L}_i\}$
\mathbb{L}	weighted Laplacian matrix
$\bar{\alpha}_i$	upper bound of the derivative of the control function at bus i
\bar{A}	$\bar{A} := \text{diag}\{\bar{\alpha}_1, \dots, \bar{\alpha}_n\}$
$[w]_{\Omega_i}$	projection of w onto set Ω_i
$[w]^+$	projection of w onto positive orthant
$\lambda_{\max}(W)$	maximum eigenvalue of matrix W
$\sigma_{\max}(W)$	maximum singular value of a matrix W

2.1 Power Flow Model for Distribution Networks

Consider a tree graph $\mathcal{G} = \{\mathcal{N} \cup \{0\}, \mathcal{L}\}$ that represents a radial distribution network consisting of $n + 1$ buses and a set \mathcal{L} of undirected lines between these buses. Bus 0 is the substation bus (slack bus) and is assumed to have a fixed voltage of 1 p.u.. Let $\mathcal{N} := \{1, \dots, n\}$. Due to the tree topology, we also have the cardinality of the line set $|\mathcal{L}| = n$. For each bus $i \in \mathcal{N}$, denote by $\mathcal{L}_i \subseteq \mathcal{L}$ the set of lines on the unique path from bus 0 to bus i , p_i^c and p_i^g the real power consumption and generation, and q_i^c and q_i^s the reactive power

consumption and generation, respectively. Let v_i be the magnitude of the complex voltage (phasor) at bus i . For each line $(i, j) \in \mathcal{L}$, denote by r_{ij} and x_{ij} its resistance and reactance, and P_{ij} and Q_{ij} the real and reactive power from bus i to bus j respectively. Let ℓ_{ij} denote the squared magnitude of the complex branch current (phasor) from bus i to bus j .

We adopt the following branch flow model introduced in [17, 18] (*DistFlow equations*) to model a *radial* distribution system:

$$P_{ij} = p_j^c - p_j^g + \sum_{k:(j,k) \in \mathcal{L}} P_{jk} + r_{ij}\ell_{ij}, \quad (2.1a)$$

$$Q_{ij} = q_j^c - q_j^g + \sum_{k:(j,k) \in \mathcal{L}} Q_{jk} + x_{ij}\ell_{ij}, \quad (2.1b)$$

$$v_j^2 = v_i^2 - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2)\ell_{ij}, \quad (2.1c)$$

$$\ell_{ij}v_i^2 = P_{ij}^2 + Q_{ij}^2. \quad (2.1d)$$

Following [16] we assume that the terms involving ℓ_{ij} are zero for all $(i, j) \in \mathcal{L}$ in (2.1). This approximation neglects the higher order real and reactive power loss terms. Since losses are typically much smaller than power flows P_{ij} and Q_{ij} , it only introduces a small relative error, typically on the order of 1%. We further assume that $v_i \approx 1$ so that we can set $v_j^2 - v_i^2 = 2(v_j - v_i)$ in equation (2.1c). This approximation introduces a small relative error of at most 0.25% if there is a 5% deviation in voltage magnitude.

With the above approximations the model (2.1) simplifies to the following linear model:

$$\begin{aligned} P_{ij} &= \sum_{k \in \beta(j)} (p_k^c - p_k^g), \\ Q_{ij} &= \sum_{k \in \beta(j)} (q_k^c - q_k^g), \\ v_i - v_j &= r_{ij}P_{ij} + x_{ij}Q_{ij}, \end{aligned}$$

where $\beta(j)$ is the set of all descendants of bus j including bus j itself, i.e., $\beta(j) = \{i | \mathcal{L}_j \subseteq \mathcal{L}_i\}$. This yields an

explicit solution for v_i in terms of v_0 (which is given and fixed):

$$\begin{aligned}
v_0 - v_i &= \sum_{(j,k) \in \mathcal{L}_i} r_{jk} P_{jk} + \sum_{(j,k) \in \mathcal{L}_i} x_{jk} Q_{jk} \\
&= \sum_{(j,k) \in \mathcal{L}_i} r_{jk} \left(\sum_{h \in \beta(k)} (p_h^c - p_h^g) \right) + \sum_{(j,k) \in \mathcal{L}_i} x_{jk} \left(\sum_{h \in \beta(k)} (q_h^c - q_h^g) \right) \\
&= \sum_{j \in \mathcal{N}} (p_j^c - p_j^g) \left(\sum_{(h,k) \in \mathcal{L}_i \cap \mathcal{L}_j} r_{hk} \right) + \sum_{j \in \mathcal{N}} (q_j^c - q_j^g) \left(\sum_{(h,k) \in \mathcal{L}_i \cap \mathcal{L}_j} x_{hk} \right) \\
&= \sum_{j \in \mathcal{N}} R_{ij} (p_j^c - p_j^g) + \sum_{j \in \mathcal{N}} X_{ij} (q_j^c - q_j^g),
\end{aligned}$$

where

$$R_{ij} := \sum_{(h,k) \in \mathcal{L}_i \cap \mathcal{L}_j} r_{hk}, \quad (2.2a)$$

$$X_{ij} := \sum_{(h,k) \in \mathcal{L}_i \cap \mathcal{L}_j} x_{hk}. \quad (2.2b)$$

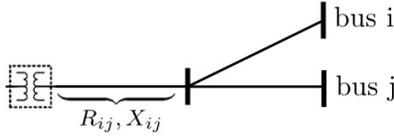


Figure 2.1: $\mathcal{L}_i \cap \mathcal{L}_j$ for two arbitrary buses i, j in the network and the corresponding mutual voltage-to-power-injection sensitivity factors R_{ij}, X_{ij} .

Fig. 2.1 gives an illustration of $\mathcal{L}_i \cap \mathcal{L}_j$ for two arbitrary buses i and j in a radial network and the corresponding R_{ij} and X_{ij} . Since

$$R_{ij} = \frac{\partial v_i}{\partial p_j^g} = -\frac{\partial v_i}{\partial p_j^c}, \quad (2.3a)$$

$$X_{ij} = \frac{\partial v_i}{\partial q_j^g} = -\frac{\partial v_i}{\partial q_j^c}, \quad (2.3b)$$

R_{ij}, X_{ij} are also referred to as the mutual voltage-to-power-injection sensitivity factors.

Define a resistance matrix $R = [R_{ij}]_{n \times n}$ and a reactance matrix $X = [X_{ij}]_{n \times n}$. Both matrices are *symmetric* and *positive*. With the matrices R and X the linearized branch flow model can be summarized compactly as:

$$v = \bar{v}_0 + R(p^g - p^c) + X(q^g - q^c), \quad (2.4)$$

where $\bar{v}_0 = [v_0, \dots, v_0]^\top$ is an n -dimensional vector. In this chapter and the next, we assume that \bar{v}^0, p^g, p^c, q^c are given constants. The only variables are (column) vectors $v := [v_1, \dots, v_n]^\top$ of squared voltage magnitudes and $q^g := (q_1^g, \dots, q_n^g)$ of generated reactive powers. Let $\tilde{v} = \bar{v}^0 + R(p^g - p^c) - Xq^c$, which is a constant vector. For notational simplicity, in the rest of this chapter and the next we will ignore the superscript in q^g and write q instead. Then the linearized branch flow model reduces to the following simple form:

$$v = Xq + \tilde{v}. \quad (2.5)$$

We have the following result, the proof of which is referred to [44].

Lemma 2.1 (Lemma 1 of [44]). *The matrices R and X are positive definite.*

2.2 Inverter Model and Local Volt/VAR Control

2.2.1 Inverter Model

We consider an inverter at bus i that can generate non-negative real power p_i and reactive power q_i that can have either sign. p_i and q_i are constrained by the apparent power capability s_i of the inverter as follows:

$$0 \leq p_i \leq s_i, \quad 0 \leq |q_i| \leq s_i, \quad p_i^2 + q_i^2 \leq s_i^2. \quad (2.6)$$

Consider power ratio $\cos \rho_i$ with $0 \leq \rho_i \leq \pi/2$ such that

$$p_i/s_i \geq \cos \rho_i. \quad (2.7)$$

Given non-controllable $p_i \leq s_i$, the feasible power set Ω_i for inverter i is cast as:

$$\Omega_i := \{q_i \mid q_i^{\min} \leq q_i \leq q_i^{\max}\}, \quad (2.8)$$

where

$$\begin{aligned} q_i^{\max} &= \min \left\{ p_i \tan \rho_i, \sqrt{s_i^2 - p_i^2} \right\}, \\ q_i^{\min} &= \max \left\{ -p_i \tan \rho_i, -\sqrt{s_i^2 - p_i^2} \right\}, \end{aligned}$$

based on (2.6)–(2.7). Here, p_i is further assumed to be sized appropriately to provide enough freedom in q_i [117]. Please also refer to [37] for more general PV inverter modeling.

For buses without inverters, we can set $s_i = p_i = q_i = 0$ and thus $\Omega_i = \{0\}$. Define $\Omega := \times_{i=1}^n \Omega_i$ for notational simplicity.

2.2.2 Local Volt/VAR Control

The goal of Volt/VAR control in a distribution network is to maintain the bus voltages v to within a tight range around their nominal values $v_i^{\text{nom}} = 1$ p.u., $i \in \mathcal{N}$ by provisioning reactive power injections $q := (q_1, \dots, q_n)$. This can be modeled as a feedback dynamical system with state $(v(t), q(t))$ at discrete time t . A general Volt/VAR control algorithm maps the current state $(v(t), q(t))$ to a new reactive power injections $q(t+1)$. The new $q(t+1)$ produces a new voltage magnitudes $v(t+1)$ according to (2.5). Usually $q(t+1)$ is determined either completely or partly by a certain Volt/VAR control function defined as follows:

Definition 2.1. A Volt/VAR control function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a collection of local control functions $f_i : \mathbb{R} \rightarrow \mathbb{R}$, each of which maps the current local voltage v_i to a local control variable u_i in reactive power at bus i :

$$u_i = f_i(v_i - v_i^{\text{nom}}), \quad \forall i \in \mathcal{N}. \quad (2.9)$$

The control functions f_i are usually decreasing but not always strictly decreasing because of a potential deadband where the control signal u_i is set to zero to prevent too frequent actuation. Assume for each bus $i \in \mathcal{N}$ a symmetric deadband $(v_i^{\text{nom}} - \delta_i/2, v_i^{\text{nom}} + \delta_i/2)$ with $\delta_i \geq 0$ around the nominal voltage v_i^{nom} . Two assumptions are made for the control functions.

Assumption 2.1. The control functions f_i are non-increasing in \mathbb{R} and strictly decreasing and differentiable in $(-\infty, -\delta_i/2)$ and $(\delta_i/2, +\infty)$.

Assumption 2.2. The derivative of the control function f_i is upper-bounded, i.e., there exist $\bar{\alpha}_i > 0$ such that $|f'_i(v_i)| \leq \bar{\alpha}_i$ for all $v_i \in (-\infty, -\delta_i/2) \cup (-\delta_i/2, \delta_i/2) \cup (\delta_i/2, +\infty)$, $\forall i \in \mathcal{N}$.

Assumption 2.2 means that an infinitesimal change in voltage should not lead to a jump in the control variable. Define $\bar{A} := \text{diag}\{\bar{\alpha}_1, \dots, \bar{\alpha}_n\} \in \mathbb{S}_{++}^N$, and let $M = \sigma_{\max}(\bar{A}X)$, the largest singular value of $\bar{A}X$. We have the following result.

Lemma 2.2 (Lipschitz continuity). *Suppose Assumptions 2.1–2.2 hold. For any $q, q' \in \Omega$, we have*

$$\|f(v(q) - v^{\text{nom}}) - f(v(q') - v^{\text{nom}})\|_2 \leq M \|q - q'\|_2. \quad (2.10)$$

Proof. Without loss of generality, assume that $v_i(q) \geq v_i(q')$. If both $v_i(q)$ and $v_i(q')$ are in $(-\infty, v_i^{\text{nom}} - \delta_i/2]$ or in $[v_i^{\text{nom}} + \delta_i/2, +\infty)$, by the mean value theorem we have $|f_i(v_i(q) - v_i^{\text{nom}}) - f_i(v_i(q') - v_i^{\text{nom}})| \leq \bar{\alpha}_i |v_i(q) - v_i(q')|$. If both are in $[v_i^{\text{nom}} - \delta_i/2, v_i^{\text{nom}} + \delta_i/2]$, $0 = |f_i(v_i(q) - v_i^{\text{nom}}) - f_i(v_i(q') - v_i^{\text{nom}})| \leq \bar{\alpha}_i |v_i(q) - v_i(q')|$. If $v_i(q) \in [v_i^{\text{nom}} + \delta_i/2, +\infty)$ and $v_i(q') \in [v_i^{\text{nom}} - \delta_i/2, v_i^{\text{nom}} + \delta_i/2]$, $|f_i(v_i(q) - v_i^{\text{nom}}) - f_i(v_i(q') - v_i^{\text{nom}})| = |f_i(v_i(q) - v_i^{\text{nom}}) - f_i(\delta_i/2)| \leq \bar{\alpha}_i |v_i(q) - (v_i^{\text{nom}} + \delta_i/2)| \leq \bar{\alpha}_i |v_i(q) - v_i(q')|$, where the first inequality follows from the mean value theorem. Similarly, we can show that $|f_i(v_i(q) - v_i^{\text{nom}}) - f_i(v_i(q') - v_i^{\text{nom}})| \leq \bar{\alpha}_i |v_i(q) - v_i(q')|$ holds under other situations too. Therefore,

$$\|f(v(q) - v^{\text{nom}}) - f(v(q') - v^{\text{nom}})\|_2 \leq \|\bar{A}(v(q) - v(q'))\|_2,$$

from which we have

$$\|f(v(q) - v^{\text{nom}}) - f(v(q') - v^{\text{nom}})\|_2 \leq \|\bar{A}X(q - q')\|_2 \leq M \|q - q'\|_2. \quad \blacksquare$$

See Fig. 2.2 (left) for an illustrative example of control function based on a piecewise linear droop control function in IEEE Standard 1547 [111]:

$$f_i(v_i) = -\alpha_i [v_i - \delta_i/2]^+ + \alpha_i [-v_i - \delta_i/2]^+ \quad (2.11)$$

with slope $-\alpha_i$ in $(-\infty, -\delta_i/2)$ and $(\delta_i/2, +\infty)$. Notice that our design and analysis in the rest of chapter (as well as the next) are not confined to the piecewise linear control functions.

Motivated by the IEEE Standard 1547, we consider a “non-incremental” control where the reactive power $q_i = u_i$, $i \in \mathcal{N}$, and obtain the following dynamical system \mathcal{D}_1 for the local Volt/VAR control:

$$\begin{cases} v(t) = Xq(t) + \tilde{v} & (2.12a) \\ q_i(t+1) = [f_i(v_i(t) - v_i^{\text{nom}})]_{\Omega_i}, \quad i \in \mathcal{N}, & (2.12b) \end{cases}$$

where $[\]_{\Omega_i}$ denotes the projection onto the set Ω_i . A fixed point (v^*, q^*) of the above dynamical system, defined as follows, represents an equilibrium operating point of the network.

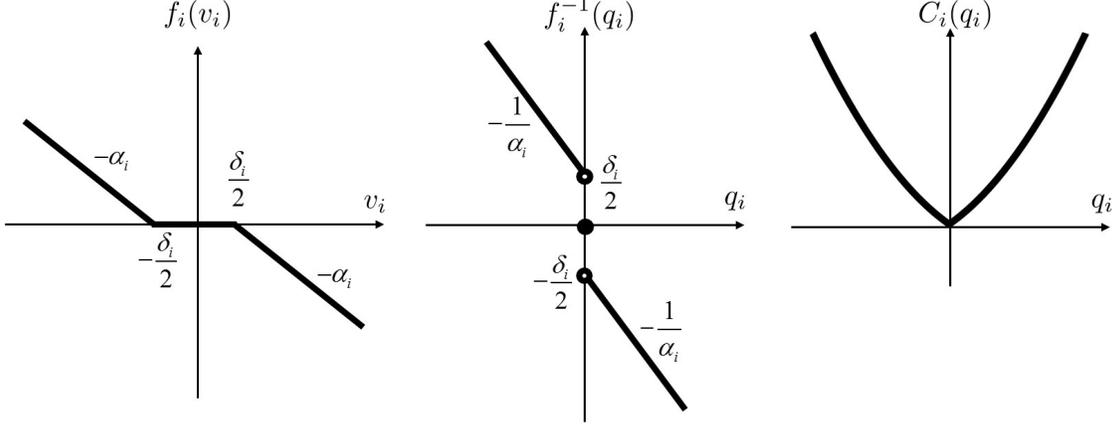


Figure 2.2: From left to right: (left) piecewise Volt/VAR control curve based on the piecewise linear function (2.11), (middle) its inverse function, and (right) the corresponding cost function (2.15), i.e., the integral of the inverse function.

Definition 2.2. (v^*, q^*) is called an equilibrium point of \mathcal{D}_1 , if it satisfies

$$v^* = Xq^* + \tilde{v}, \quad (2.13a)$$

$$q^* = [f(v^* - v^{nom})]_{\Omega}. \quad (2.13b)$$

In the next section of related works, we first characterize the equilibrium of the system \mathcal{D}_1 by showing that it is an distributed algorithm for solving a well-defined optimization problem, followed by its dynamical properties. We then introduces an incremental local Volt/VAR control based on (sub)gradient algorithm that achieves the same equilibrium as \mathcal{D}_1 while permits less restrictive convergence conditions. These works have been previously published and further improved later, especially regarding their stability analysis.

2.3 Related Works

This section first studies the equilibrium and dynamical properties of local Volt/VAR control \mathcal{D}_1 , and then designed a (sub)gradient-based control dynamics \mathcal{D}_2 for better voltage regulation results and less restricted convergence conditions.

2.3.1 Reverse Engineering

Since f_i is non-increasing, a (generalized) inverse f_i^{-1} exists. In particular, at the origin, we assign $f_i^{-1}(0) = 0$ corresponding to the deadband $[-\delta_i/2, +\delta_i/2]$ of f_i . This may introduce a discontinuity to f_i^{-1} at $q_i = 0$ if the deadband $\delta_i > 0$, with

$$f_i^{-1}(0^+) \leq -\delta_i/2 \quad \text{and} \quad f_i^{-1}(0^-) \geq \delta_i/2. \quad (2.14)$$

Define a cost function for provisioning reactive power at each bus $i \in \mathcal{N}$ as:

$$C_i(q_i) := - \int_0^{q_i} f_i^{-1}(q) dq, \quad (2.15)$$

which is convex since f_i^{-1} is decreasing. Then, given $v_i(t)$, $q_i(t+1)$ in (2.12b) is the unique solution to the following individual optimization problem:

$$q_i(t+1) = \arg \min_{q_i \in \Omega_i} C_i(q_i) + q_i (v_i(t) - v_i^{\text{nom}}), \quad (2.16)$$

i.e., (2.12b) and (2.16) are equivalent specification of $q_i(t+1)$.

Take for example the piece-wise linear control function (2.11). Its inverse is given by:

$$f_i^{-1}(q_i) := \begin{cases} -\frac{q_i}{\alpha_i} + \frac{\delta_i}{2} & \text{if } q_i < 0, \\ 0 & \text{if } q_i = 0, \\ -\frac{q_i}{\alpha_i} - \frac{\delta_i}{2} & \text{if } q_i > 0, \end{cases} \quad (2.17)$$

and the corresponding cost function is given by:

$$C_i(q_i) = \begin{cases} \frac{1}{2\alpha_i} q_i^2 - \frac{\delta_i}{2} q_i & \text{if } q_i \leq 0, \\ \frac{1}{2\alpha_i} q_i^2 + \frac{\delta_i}{2} q_i & \text{if } q_i \geq 0. \end{cases} \quad (2.18)$$

See also Fig. 2.2 (middle and right) for illustration.

2.3.2 Equilibrium Characterization

Consider the scalar valued function $F(q) : \Omega \rightarrow \mathbb{R}$:

$$F(q) := C(q) + \frac{1}{2} q^\top X q + q^\top \Delta \tilde{v}, \quad (2.19)$$

where $C(q) := \sum_{i \in \mathcal{N}} C_i(q_i)$ and $\Delta \tilde{v} := \tilde{v} - v^{\text{nom}}$, and a global optimization problem:

$$\min_{q \in \Omega} F(q). \quad (2.20)$$

Theorem 2.1 (Theorem 1 in [44]). *Suppose Assumption 2.1 holds. Then \mathcal{D}_1 has a unique equilibrium point. Moreover, a point (v^*, q^*) is an equilibrium of \mathcal{D}_1 if and only if q^* is the unique optimal solution of (2.20) and $v^* = Xq^* + \tilde{v}$.*

We refer the proof of Theorem 2.1 to [44].

With $v = Xq + \tilde{v}$, the objective can be equivalently written as:

$$F(q, v) = C(q) + \frac{1}{2}(v - v^{\text{nom}})^\top X^{-1}(v - v^{\text{nom}}) - \frac{1}{2}\Delta \tilde{v}^\top X^{-1}\Delta \tilde{v}. \quad (2.21)$$

Notice that the last term is a constant. Therefore, the local Volt/VAR control \mathcal{D}_1 can be seen as seeking an optimal trade-off between minimizing the cost of reactive power provisioning $C(q)$ and minimizing the cost of voltage deviation $\frac{1}{2}(v - v^{\text{nom}})^\top X^{-1}(v - v^{\text{nom}})$. We next provide more detailed characterization of the objective function (2.21).

2.3.2.1 Further Characterization of Equilibrium

The first term $C(q)$ of the objective (2.21) is well-defined and has the desired additive structure. It is however not clear what specific structure the second term $\frac{1}{2}(v - v^{\text{nom}})^\top X^{-1}(v - v^{\text{nom}})$ entails. We will further characterize this term in this subsection.

Notice that bus 0 has a fixed voltage magnitude, which decouples different subtrees rooted at it. Therefore, without loss of generality we only consider a topology where the bus 0 is of degree 1. Denote \mathcal{T} the (sub)tree rooted at bus 1 and \mathcal{L}_T the set of links of \mathcal{T} . Define an inverse tree \mathcal{T}' that has the same sets of buses and lines as \mathcal{T} but with reciprocal line reactance $1/x_{ij}$. Let $\mathbb{L} \in \mathbb{R}^{n \times n}$ be the weighted Laplacian matrix of \mathcal{T}' defined as follows:

$$\mathbb{L}_{ij} = \begin{cases} -1/x_{ij}, & (i, j) \in \mathcal{L}_T, \\ \sum_{(i,k) \in \mathcal{L}} 1/x_{ik}, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Recall that x denotes the reactance of the line connecting buses 0 and 1, we have the following result by Liu *et al.* [81].

Theorem 2.2 (from [81]). *Given the tree graph $\mathcal{G} = \{\mathcal{N} \cup \{0\}, \mathcal{L}\}$ with bus 0 being of degree 1 and its reactance matrix X defined by (2.2b), the inverse matrix X^{-1} has the following explicit form:*

$$X^{-1} = \mathbb{L} + \begin{bmatrix} 1/x & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (2.22)$$

With the above result, the cost function (2.21) can be written as:

$$F(q, v) = C(q) + \frac{1}{2} \left(\frac{(v_1 - v^{\text{nom}})^2}{x} + \sum_{(i,j) \in \mathcal{L}_T} \frac{(v_i - v_j)^2}{x_{ij}} \right) - \frac{1}{2} \Delta \tilde{v}^\top X^{-1} \Delta \tilde{v}. \quad (2.23)$$

We see that the second term (i.e., the cost of voltage deviation) consists of two parts: the first part $(v_1 - v^{\text{nom}})^2/x$ represents the cost of voltage deviation of the bus 1 from the nominal value, and the second part $\sum_{(i,j) \in \mathcal{L}_T} (v_i - v_j)^2/x_{ij}$ gives the cost of voltage deviation among the neighboring buses. This leads to a nice leader-follower structure where the first bus (the bus 1) aims to attain the nominal voltage while each other bus tries to achieve the same voltage as that of the bus “in front of” it.

2.3.3 Dynamics Analysis

We now study the dynamic properties of the local Volt/VAR control \mathcal{D}_1 .

Theorem 2.3. *Suppose Assumptions 2.1–2.2 hold. If*

$$\sigma_{\max}(\bar{A}X) < 1, \quad (2.24)$$

then the local Volt/VAR control \mathcal{D}_1 converges to the unique equilibrium point (v^, q^*) . Moreover, it converges exponentially fast to the equilibrium.*

Proof. Write \mathcal{D}_1 equivalently as a mapping g_1 :

$$q(t+1) = g_1(q(t)) := [f(Xq(t) + \Delta \tilde{v} - v^{\text{nom}})]_\Omega. \quad (2.25)$$

By Lemma 2.2 and the non-expansiveness property of projection operator, given any feasible q, q' we have

$$\|g_1(q) - g_1(q')\|_2 \leq M\|q - q'\|_2, \quad (2.26)$$

where $M = \sigma_{\max}(\bar{A}X)$. When condition (2.24) holds, $M < 1$ and thus the mapping g_1 is a contraction, implying that $(v(t), q(t))$ converges exponentially fast to the unique equilibrium point under \mathcal{D}_1 . ■

Define the following matrix norms for some $W \in \mathbb{R}^{m \times n}$:

$$\|W\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |w_{ij}|, \quad \|W\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |w_{ij}|, \quad \|W\|_2 = \sqrt{\lambda_{\max}(W^\top W)} = \sigma_{\max}(W),$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix. By utilizing the following relationship among these matrix norms based on Hölder's inequality

$$\|W\|_2 \leq \sqrt{\|W\|_1 \cdot \|W\|_\infty}, \quad (2.27)$$

we have the following sufficient condition for convergence of \mathcal{D}_1 , which is easier to verify in practice.

Corollary 2.1. *Suppose Assumptions 2.1–2.2 hold. If*

$$\max_{i \in \mathcal{N}}(\bar{\alpha}_i) \cdot \max_{i \in \mathcal{N}} \left(\sum_{j \in \mathcal{N}} X_{ij} \right) < 1, \quad (2.28)$$

then \mathcal{D}_1 converges exponentially fast to the unique equilibrium point (v^, q^*) .*

Proof. A sufficient condition for (2.24) based on (2.27) is

$$\|\bar{A}X\|_1 < 1 \quad \text{and} \quad \|\bar{A}X\|_\infty < 1. \quad (2.29)$$

Given symmetric matrix X , (2.28) is thereafter sufficient for (2.29). ■

2.3.3.1 Limitation of the Non-Incremental Control

The voltage control (2.12b) is non-incremental, as it decides the total amount of reactive power based on the deviation of current voltage from the nominal value. Intuitively, such a control may lead to over-actuation and oscillatory behavior. In order to have converging or stable behavior, the control function

should not be too aggressive, i.e., have small derivative. This can also be seen from Theorem 2.3. In the case of the piece-wise linear control function (2.11), this implies a small α_i value.

On the other hand, seen from the equivalent objective (2.21), smaller cost functions $C_i(q_i)$ are preferred for better voltage regulation. However, a small cost function implies large derivative of the control function; see, e.g., the cost function (2.18) that becomes smaller as α_i takes larger value, as well as numerical examples shown in Section 3.3 next chapter.

Hence, there is a contention or fundamental limitation for the non-incremental control: control function with smaller derivative is preferred for convergence, while for better voltage regulation at the equilibrium control function with larger derivative is desired. This motivates us to seek new types of voltage control that are not subject to such a limitation, as will be seen in the next section.

2.3.4 (Sub)gradient Algorithm

Given an optimization problem, we may apply different algorithms to solve it. A common algorithm that often admits distributed implementation is the (sub)gradient method [31]. Applying to the problem (2.20) leads to the following voltage control:

$$q_i(t+1) = \left[q_i(t) - \gamma_2 \frac{\partial F(q(t))}{\partial q_i} \right]_{\Omega_i}, \quad i \in \mathcal{N}, \quad (2.30)$$

where $\gamma_2 > 0$ is the constant stepsize and the (sub)gradient is calculated as follows:

$$\frac{\partial F(q(t))}{\partial q_i} = \begin{cases} -f_i^{-1}(q_i(t)) + v_i(t) - v^{\text{nom}} & \text{if } q_i(t) \neq 0, \\ v_i(t) - v^{\text{nom}} & \text{if } q_i(t) = 0 \text{ and } -\delta/2 \leq v_i(t) - v^{\text{nom}} \leq \delta_i/2, \\ -\delta_i/2 + v_i(t) - v^{\text{nom}} & \text{if } q_i(t) = 0 \text{ and } v_i(t) - v^{\text{nom}} > \delta_i/2, \\ \delta_i/2 + v_i(t) - v^{\text{nom}} & \text{if } q_i(t) = 0 \text{ and } v_i(t) - v^{\text{nom}} < -\delta_i/2. \end{cases} \quad (2.31)$$

The above control is *incremental* as the change in reactive power (instead of the total reactive power) is based on the voltage deviation from the nominal value. It is also distributed, since the decision at each bus $i \in \mathcal{N}$ depends only on its current reactive power q_i and voltage v_i .

We thus obtain the following dynamical system \mathcal{D}_2 :

$$\begin{cases} v(t) = Xq(t) + \tilde{v} \\ q_i(t+1) = \left[q_i(t) - \gamma_2 \frac{\partial F(q(t))}{\partial q_i} \right]_{\Omega_i}, \quad i \in \mathcal{N}. \end{cases} \quad (2.32a)$$

$$(2.32b)$$

The following result is immediate.

Theorem 2.4. *Suppose Assumption 2.1 holds. Then there exists a unique equilibrium point for the dynamical system \mathcal{D}_2 . Moreover, a point (v^*, q^*) is an equilibrium if and only if q^* is the unique optimal solution of problem (2.20) and $v^* = Xq^* + \tilde{v}$.*

Since the feasible sets are bounded, we also have the bounded (sub)gradient of $F(q)$ with some constant $G > 0$ as

$$\|\nabla_q F(q)\|_2 \leq G, \quad \forall q \in \Omega. \quad (2.33)$$

Theorem 2.5. *Suppose Assumption 2.1 hold. The dynamical system \mathcal{D}_2 converges as*

$$\limsup_{t \rightarrow \infty} \sup_{q(t)} \sum_{\tau=1}^t \frac{F(q(\tau)) - F(q^*)}{t} = \gamma_2^2 G^2. \quad (2.34)$$

Proof. We characterize the distance between $q(t+1)$ and q^* as:

$$\begin{aligned} & \|q(t+1) - q^*\|_2^2 \\ & \leq \|q(t) - \gamma_2 \nabla_q F(q(t)) - q^*\|_2^2 \\ & = \|q(t) - q^*\|_2^2 + \gamma_2^2 \|\nabla_q F(q(t))\|_2^2 - 2\gamma_2 (q(t) - q^*)^\top \nabla_q F(q(t)) \\ & \leq \|q(t) - q^*\|_2^2 + \gamma_2^2 G^2 - (F(q(t)) - F(q^*)) \\ & \leq \|q(1) - q^*\|_2^2 + t\gamma_2^2 G^2 - \sum_{\tau=1}^t (F(q(\tau)) - F(q^*)), \end{aligned}$$

where the first inequality is due to non-expansiveness of projection operator, the second inequality is because of the definition of subgradient as well as the bounded gradient (2.33), and the last inequality is by repeating previous steps.

Because $\|q(t+1) - q^*\|_2^2 \geq 0$, it follows that:

$$\sum_{\tau=1}^t \frac{F(q(\tau)) - F(q^*)}{t} \leq \|q(1) - q^*\|_2^2 / t + \gamma_2^2 G^2. \quad (2.35)$$

When $t \rightarrow \infty$, we have (2.34). ■

2.4 Motivation for Improved Design

Notice that for any control functions f_i (that satisfies Assumption 2.1–2.2), one can always find a small enough stepsize γ_2 such that \mathcal{D}_2 converge to a neighborhood of the (v^*, q^*) of required accuracy on running average. Moreover, as proved in [45], when q^* is *not* close to the non-differentiable point, \mathcal{D}_2 converges exactly to the optima. In contrast, the convergence condition (2.24) for the non-incremental voltage control \mathcal{D}_1 does constrain the allowable control functions f_i . Therefore, \mathcal{D}_2 permits potentially better voltage regulation results than \mathcal{D}_1 ; see the discussion at the end of Section 2.3.3.1.

Nevertheless, the (sub)gradient nature of \mathcal{D}_2 may prevent it from converging to the exact optimal point. This could take place when the optimal is close to the non-differentiable point of $F(q)$ ($q_i = 0$ for some i in this case) where the value of subgradient (2.31) changes abruptly if $\delta_i \neq 0$. Moreover, the (sub)gradient-based control \mathcal{D}_2 incurs implementation complexity, since the (sub)gradient computation (2.31) requires tracking the value of v_i with respect to deadband $\pm\delta_i/2$, and takes different forms accordingly. Furthermore, it requires the calculation of the inverse of the control function f_i , which can be computationally expensive for general control functions. These limitations motivate us to design the next incremental algorithms with better convergence properties and simpler implementation in Section 3.1.

Moreover, noticing that the algorithms for solving (2.21) does not strictly assure the range of the voltages at equilibrium, we will further propose a local control to guarantee that the stabilized voltages are within bounds specified in the deadband of control functions in Section 3.2.

Chapter 3

Local Volt/VAR Control Design and Analysis

This chapter presents two incremental local Volt/VAR control strategies for better voltage regulation, improved convergence properties, and simpler implementation. In particular, we first propose a local control based on pseudo-gradient algorithm that not only achieves the same equilibrium point as the (sub)gradient algorithm and the original non-incremental control, but also assures convergence to the equilibrium point with appropriately chosen stepsize regardless of the derivatives of control functions or the non-differentiable points in the cost function. While the aforementioned incremental algorithms are able to attain smaller voltage deviation from nominal values by applying steeper control functions, they cannot strictly contain the voltages within specific ranges. In light of this limitation, we design another incremental local control that stops altering local reactive power provisioning when local voltage is within the bounds specified by the deadband in the control function. This design ensures that the stabilized voltages at control buses are strictly within specific ranges, given sufficient reactive power supply. Our designs are characterized analytically and evaluated numerically. Related works of this chapter have been presented in [134, 139, 141].

3.1 Local Volt/VAR Control with Better Convergence

3.1.1 Pseudo-Gradient Algorithm Design

The pseudo-gradient can provide a good search direction for an optimization problem without requiring the objective function to be differentiable [124]. Consider the following incremental voltage control:

$$q_i(t+1) = [(1 - \gamma_3)q_i(t) + \gamma_3 f_i(v_i(t) - v_i^{\text{nom}})]_{\Omega_i} \quad (3.1a)$$

$$= [q_i(t) - \gamma_3(q_i(t) - f_i(v_i(t) - v_i^{\text{nom}}))]_{\Omega_i}, \quad i \in \mathcal{N}, \quad (3.1b)$$

where $\gamma_3 > 0$ is a constant stepsize/weight and $q_i - f_i(v_i - v_i^{\text{nom}})$ is the pseudo-gradient direction. The above control is distributed, and is evidently simpler to implement than the (sub)gradient algorithm (2.30).

With (3.1b) we obtain the following dynamical system \mathcal{D}_3 :

$$\begin{cases} v(t) = Xq(t) + \tilde{v}, & (3.2a) \\ q_i(t+1) = [q_i(t) - \gamma_3(q_i(t) - f_i(v_i(t) - v_i^{\text{nom}}))]_{\Omega_i}, \quad i \in \mathcal{N}. & (3.2b) \end{cases}$$

Notice that \mathcal{D}_3 has the same equilibrium condition as the dynamical systems \mathcal{D}_1 and \mathcal{D}_2 . The following result is thus immediate by examining the equilibrium condition of (3.2).

Theorem 3.1. *Suppose Assumption 2.1 holds. There exists a unique equilibrium point for the dynamical system \mathcal{D}_3 . Moreover, a point (v^*, q^*) is an equilibrium if and only if q^* is the unique optimal solution of problem (2.20) and $v^* = Xq^* + \tilde{v}$.*

3.1.2 Dynamics Analysis

We now analyze the convergence of the dynamical system \mathcal{D}_3 . We shall first introduce the following useful results.

Denote by $\nabla_v f$ the diagonal matrix with each entry $(\nabla_v f)_{ii}$ representing the (sub)gradient of the control functions defined as

$$(\nabla_v f)_{ii} = \begin{cases} f'_i(v_i) & \text{if } v_i \in (-\infty, -\delta_i/2) \cup (-\delta_i/2, \delta_i/2) \cup (\delta_i/2, +\infty) \\ \frac{f'_i(v_i^-) + f'_i(v_i^+)}{2} & \text{if } v_i = -\delta_i/2 \text{ or } v_i = \delta_i/2 \end{cases}, \quad (3.3)$$

which is bounded by $-\bar{\alpha}_i \leq (\nabla_v f)_{ii} \leq 0$ based on Assumption 2.1–2.2. Denote by λ any eigenvalue of the asymmetric matrix $\nabla_v f X$. Consider $\nabla_v f X$'s similar matrix $X^{1/2} \nabla_v f X^{1/2}$, which is symmetric and negative semidefinite with real and nonpositive eigenvalues. Therefore, eigenvalues of the original asymmetric matrix $\nabla_v f X$ are also *real* and *nonpositive*. Similarly, all the eigenvalues of $\bar{A}X$ are *real* and *positive*.

Theorem 3.2. *Suppose Assumptions 2.1–2.2 hold. If the stepsize γ_3 satisfies the following condition:*

$$0 < \gamma_3 < 2/(1 + \lambda_{\max}(\bar{A}X)), \quad (3.4)$$

then the dynamical system \mathcal{D}_3 converges to its unique equilibrium point.

Proof. Write \mathcal{D}_3 equivalently as a mapping g_3 :

$$q(t+1) = g_3(q(t)) := [(1 - \gamma_3)q(t) + \gamma_3 f(v(q(t)))]_{\Omega}. \quad (3.5)$$

We study the convergence of \mathcal{D}_3 based on the Jacobian matrix of g_3 without projection operator, which is computed as

$$\nabla_q g_3 = (1 - \gamma_3)\mathcal{I} + \gamma_3 \nabla_v f X, \quad (3.6)$$

where \mathcal{I} is identity matrix. Suppose the projection operator is active for some q_i . Then the i th row of the resulting $\nabla_q g_3$ is all 0. So, by Gershgorin circle theorem [119], the magnitude of $\nabla_q g_3$'s eigenvalue without active projection has a larger bound than that with projection. Thus it is sufficient to consider g_3 without projection operator for the rest of the proof.

Denote by z the eigenvector of matrix $\nabla_v f X$ corresponding to λ , and by definition one has $\nabla_v f X z = \lambda z$. Therefore, we have

$$\nabla_q g_3 z = (1 - \gamma_3 + \gamma_3 \lambda)z, \quad (3.7)$$

that is, the corresponding eigenvalue of $\nabla_q g_3$ with respect to λ is $1 - \gamma_3 + \gamma_3 \lambda$. To ensure that g_3 is stable, one must have:

$$-1 < 1 - \gamma_3 + \gamma_3 \lambda < 1 \quad (3.8)$$

for any eigenvalue λ of $\nabla_v f X$ [46, 60]. (3.4) is sufficient for the left-hand side of (3.8) and the right-hand side always holds because λ is nonpositive. This completes the proof. ■

By Theorem 3.2, \mathcal{D}_3 converges to the optimum with properly chosen stepsize γ_3 , regardless of the non-differentiable point that potentially prevents \mathcal{D}_2 from converging to equilibrium point.

Remark 3.1. Notice that $\gamma_3 \leq 1$ in \mathcal{D}_3 gives a nice interpretation of the new decision $q_i(t+1)$ being a (positively-)weighted sum of the decision $q_i(t)$ at the previous time and the local control $u_i(t) = f_i(v_i(t) - v_i^{nom})$ in reactive power. Similar approaches in literatures can be identified as Exponentially Weighted Moving Average (EWMA) control, delayed control, etc. with $\gamma_3 \leq 1$. However, here we do not require $\gamma_3 \leq 1$ for \mathcal{D}_3 to converge, as long as condition (3.4) is satisfied.

3.2 Local Volt/VAR Control with Better Voltage Regulation

3.2.1 Algorithm Design

Consider the following voltage control scheme:

$$q_i(t+1) = \left[q_i(t) + \gamma_4 f_i(v_i(t)) \right]_{\Omega_i}, \quad i \in \mathcal{N}, \quad (3.9)$$

where each bus adjusts its reactive power provisioning *incrementally* in response to its current voltage with a certain constant stepsize $\gamma_4 > 0$, until local voltage is within deadband of the control function, i.e., when $f_i(v_i) = 0$. This control scheme is local, as each bus only needs to know its own reactive power and voltage.

With this local control scheme, the dynamical system denoted by \mathcal{D}_4 can be written as:

$$\begin{cases} v(t) = Xq(t) + \tilde{v} & (3.10a) \\ q_i(t+1) = \left[q_i(t) + \gamma_4 f_i(v_i(t)) \right]_{\Omega_i}, \quad i \in \mathcal{N}. & (3.10b) \end{cases}$$

3.2.2 Equilibrium Characterization

An equilibrium of the dynamical system \mathcal{D}_4 is defined as follows.

Definition 3.1. (v^*, q^*) is an equilibrium of dynamical system \mathcal{D}_4 if

$$v^* = Xq^* + \tilde{v}, \quad (3.11a)$$

$$\begin{cases} f_i(v_i^*) = 0, \text{ or} \\ f_i(v_i^*) < 0 \ \& \ q_i^* = q_i^{\min}, \text{ or} \\ f_i(v_i^*) > 0 \ \& \ q_i^* = q_i^{\max}. \end{cases} \quad (3.11b)$$

Theorem 3.3. *There may exist infinitely many equilibria for the dynamical system \mathcal{D}_4 .*

Proof. In order to show the possible existence of infinitely many equilibria for \mathcal{D}_4 , it is sufficient to show one such situation. For convenience, let $\bar{v}_j := v_j^{\text{nom}} + \delta_i/2$ and $\underline{v}_j := v_j^{\text{nom}} - \delta_i/2$ be the upper and lower bounds of the acceptable voltage range at bus j .

Consider a simple situation where there exists a fixed point (v^*, q^*) whose voltage value $\underline{v}_j < v_j^* < \bar{v}_j$, $\forall j \in \mathcal{N}$, and $q_j^{\text{min}} < q_j^* < q_j^{\text{max}}$, $\forall j \in \mathcal{N}$. This can be easily realized by choosing appropriate control functions and bounds on reactive powers, given the structure of the distribution network. An example is presented in Section 3.3.1.2.

Let $\epsilon > 0$ be a small enough perturbation occurring to an arbitrary bus i , so that its reactive power injection becomes $q_i^{*'} = q_i^* + \epsilon$. As a result, voltages across the networks will become $v_j^{*'} = \sum_{k \neq i} X_{jk} q_k^* + X_{ji} q_i^{*'}$, $\forall j \in \mathcal{N}$. When ϵ is small enough, we can still have $v_j^{*'} \in (\underline{v}_j, \bar{v}_j)$, $\forall j \in \mathcal{N}$, which leads to q_j^* remaining unchanged for $\forall j \neq i$. We can see that $((v_1^{*'}, \dots, v_n^{*'}), (q_1^*, \dots, q_i^{*'}, \dots, q_n^*))$ is also an equilibrium of \mathcal{D}_4 . Since ϵ can be arbitrarily small, there are infinitely many possible equilibria. ■

Remarks: Intuitively, as long as the system can provide enough reactive power so as to bring the voltages to the acceptable ranges, there are usually infinitely many equilibria for \mathcal{D}_4 . On the other hand, situations with unique equilibrium may take place when there is not enough reactive power supply.

3.2.3 Reverse Engineering

For better understanding the dynamical system \mathcal{D}_4 , we first examine if it can be reverse engineered as a distributed algorithm for solving a well-defined optimization problem. We will show that this generally cannot be done. Nevertheless, as we will show later, \mathcal{D}_4 can still be seen as a non-cooperative game where each bus minimizes its local voltage deviation from deadband.

Theorem 3.4. *Suppose A1 holds. A point (v^*, q^*) is an equilibrium of the dynamical system \mathcal{D}_4 if and only if (v^*, q^*) is a solution of the following variational inequalities (VI) problem:*

$$-f(v^*)^\top (q - q^*) \geq 0, \quad \forall q \in \Omega, \quad (3.12a)$$

$$v^* = Xq^* + \tilde{v}. \quad (3.12b)$$

Proof. This can be proved by the showing equivalence between (3.11) and (3.12). ■

We get to the next result based on the VI formulation of \mathcal{D}_4 .

Theorem 3.5. *The dynamical system \mathcal{D}_4 can be seen as a distributed algorithm for solving a convex optimization problem if and only if all control functions f_i 's are identically linear functions with $\delta_i = 0$.*

Proof. VI problem (3.12) can be seen as the optimality condition of a well-defined convex optimization problem with a scalar objective function if and only if the Jacobian matrix of $-f(v(q))$ is symmetric, $\forall q \in \Omega$ [104], that is,

$$\frac{\partial f_i(v_i(q))}{\partial q_j} = X_{ij} f'_i(v_i) = \frac{\partial f_j(v_j(q))}{\partial q_j} = X_{ji} f'_j(v_j), \quad \forall i, j \in \mathcal{N}.$$

Since $X_{ij} = X_{ji}$, it is necessary to have $f'_i(v_i) = f'_j(v_j)$, $\forall i, j$, which is true if and only if all f_i 's are identically linear with no deadband. ■

Define a cost function for bus $i \in \mathcal{N}$ as

$$h_i(v_i) := \int_0^{v_i} -f_i(v) dv, \quad (3.13)$$

penalizing local voltage deviation from deadband, and a voltage deviation minimization game \mathcal{G}_{vd} as follows.

Definition 3.2. *A non-cooperative voltage deviation minimization game is defined as a triple $\mathcal{G}_{vd} := \{\mathcal{N}, \Omega, (h_i(q_i; v_i(q)))_{i \in \mathcal{N}}\}$, where the strategic interaction among players is through the voltage $v_i(q)$, $i \in \mathcal{N}$.*

Then the VI formation of the equilibrium of \mathcal{D}_4 naturally leads to the following result [104].

Theorem 3.6. *The non-cooperative voltage deviation minimization game \mathcal{G}_{vd} is equivalent to VI problem (3.12).*

3.2.4 Dynamics Analysis

We next characterize the convergence of the dynamical system \mathcal{D}_4 .

Theorem 3.7. *Suppose Assumption 2.1–2.2 hold. If the stepsize γ_4 in \mathcal{D}_4 satisfies the following condition:*

$$0 < \gamma_4 < \frac{2}{\lambda_{\max}(\bar{A}X)}, \quad (3.14)$$

where $\bar{A} := \text{diag}(\{\bar{\alpha}_i\}_{i \in N})$, then the dynamical system \mathcal{D}_4 converges globally and asymptotically to one of its equilibrium points.

Proof. The proof is similar to that of Theorem 3.2. Rewrite \mathcal{D}_4 equivalently as the following mapping g_4 :

$$q(t+1) = g_4(q(t)) := [q(t) + \gamma_4 f(v(q(t)))]_{\Omega}. \quad (3.15)$$

We consider the Jacobian matrix of g_4 without projection operator, which will lead to a sufficient condition, computed as

$$\nabla_q g_3 = \mathcal{I} + \gamma_4 \nabla_v f X. \quad (3.16)$$

Similar to (3.7), the corresponding eigenvalue of $\nabla_q g_4$ with respect to the eigenvalue λ of $\nabla_v f X$ is $1 + \gamma_4 \lambda$. To ensure that g_4 is stable, one must have:

$$-1 < 1 + \gamma_4 \lambda < 1 \quad (3.17)$$

for any eigenvalue λ of $\nabla_v f X$ [46, 60]. Then (3.14) is sufficient for the left-hand side of (3.17) and the right-hand side always holds because λ is nonpositive as shown in Section 3.1.2. ■

Since the objective function (3.13) for each bus has an interpretation as the cost of voltage deviation from deadband, the incremental local voltage control \mathcal{D}_4 minimizes the cost of local voltage deviation. This is different from \mathcal{D}_1 – \mathcal{D}_3 which achieve a trade-off between minimizing the cost of reactive power provisioning and minimizing the cost of voltage deviation. The above reverse engineering result suggests a way to stabilize the distribution system to within desired acceptable voltage ranges by simply specifying the control function f_i 's deadband as the acceptable voltage range.

3.3 Numerical Examples

Consider a distribution feeder of South California Edison (SCE) with a high penetration of photovoltaic (PV) generation. As shown in Fig. 3.1, bus 1 is the substation (root bus) with fixed voltage v^{nom}

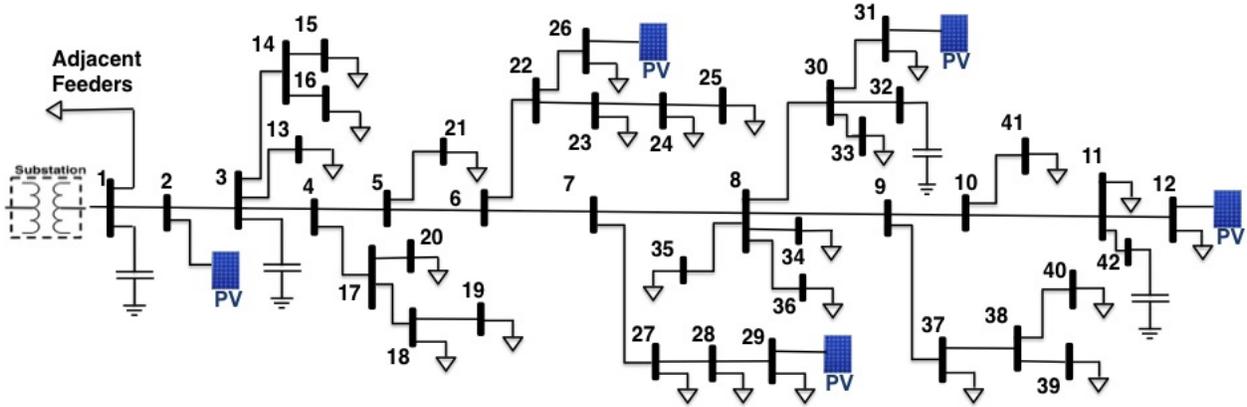


Figure 3.1: Circuit diagram for SCE distribution system.

and five PV generators are integrated at buses 2, 12, 26, 29, and 31. As we aim to study the Volt/VAR control through PV inverters, all shunt capacitors are assumed to be off. Table 3.1 contains the network data including the line impedance, the peak MVA demand of loads, and the capacity of the PV generators. It is important to note that all studies are run with a full AC power flow model with MATPOWER [145] instead of its linear approximation. As will be seen, the results we develop for the linearized model are corroborated numerically with the full power flow model.

In all numerical studies, we implement homogeneous piecewise linear droop control functions (2.11) of the IEEE 1547.8 Standard [111] for all PV inverters, with deadbands ranging from 0.98 p.u. to 1.02 p.u. and slopes α_i to be determined.

3.3.1 Equilibrium

3.3.1.1 \mathcal{D}_1 – \mathcal{D}_3 : Single Equilibrium with Tradeoff

As discussed in Section 2.3.3.1, large slopes of the control functions lead to better voltage regulation at the equilibrium. To illustrate this, we change α_i from 1 to 200 and record the corresponding equilibrium voltages v^* . As shown in Fig. 3.2, v^* gets closer to v^{nom} as α_i increases. This confirms our previous discussion that steeper control functions should be implemented for smaller voltage deviation from the nominal value.

Network Data																	
Line Data				Line Data				Line Data				Load Data		Load Data		PV Generators	
From Bus.	To Bus.	R (Ω)	X (Ω)	From Bus.	To Bus.	R (Ω)	X (Ω)	From Bus.	To Bus.	R (Ω)	X (Ω)	Bus No.	Peak MVA	Bus No.	Peak MVA	Bus No.	Capacity MW
1	2	0.259	0.808	8	34	0.244	0.046	18	19	0.198	0.046	11	0.67	28	0.27		
2	3	0.031	0.092	8	36	0.107	0.031	22	26	0.046	0.015	12	0.45	29	0.2	2	1
3	4	0.046	0.092	8	30	0.076	0.015	22	23	0.107	0.031	13	0.89	31	0.27	26	2
3	13	0.092	0.031	8	9	0.031	0.031	23	24	0.107	0.031	15	0.07	33	0.45	29	1.8
3	14	0.214	0.046	9	10	0.015	0.015	24	25	0.061	0.015	16	0.67	34	1.34	31	2.5
4	17	0.336	0.061	9	37	0.153	0.046	27	28	0.046	0.015	18	0.45	35	0.13	12	3
4	5	0.107	0.183	10	11	0.107	0.076	28	29	0.031	0	19	1.23	36	0.67		
5	21	0.061	0.015	10	41	0.229	0.122	30	31	0.076	0.015	20	0.45	37	0.13		
5	6	0.015	0.031	11	42	0.031	0.015	30	32	0.076	0.046	21	0.2	39	0.45		
6	22	0.168	0.061	11	12	0.076	0.046	30	33	0.107	0.015	23	0.13	40	0.2		
6	7	0.031	0.046	14	16	0.046	0.015	37	38	0.061	0.015	24	0.13	41	0.45		
7	27	0.076	0.015	14	15	0.107	0.015	38	39	0.061	0.015	25	0.2				
7	8	0.015	0.015	17	18	0.122	0.092	38	40	0.061	0.015	26	0.07				
8	35	0.046	0.015	17	20	0.214	0.046					27	0.13				

Table 3.1: Network Parameters of the SCE Circuit: Line impedances, peak spot load KVA, Capacitors and PV generation's nameplate ratings.

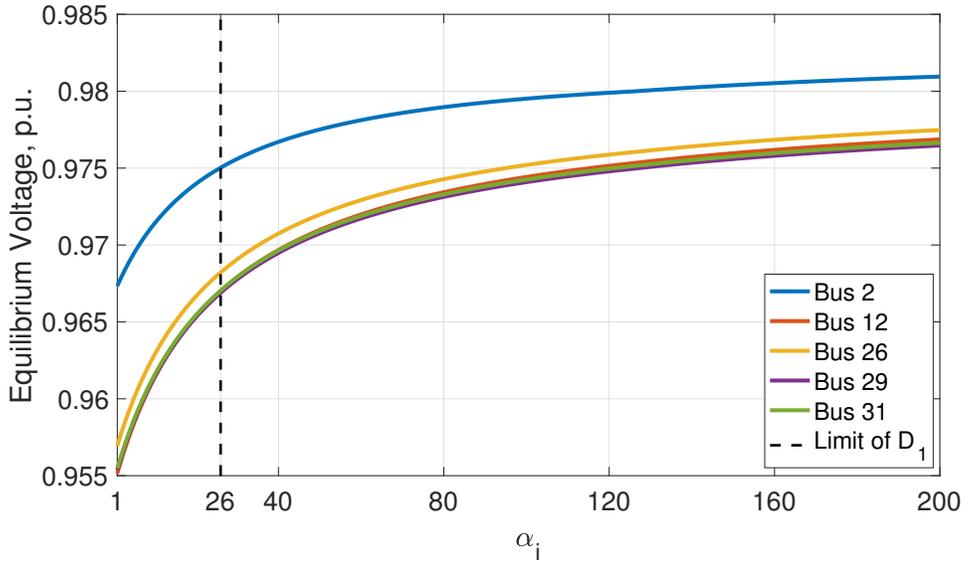


Figure 3.2: Equilibrium voltage versus the α_i value: As α_i increases, the equilibrium voltage v_i^* deviates less from the nominal value.

3.3.1.2 \mathcal{D}_4 : Multiple Equilibrium within Deadband

As proved in Theorem 3.3, \mathcal{D}_4 may possess multiple equilibrium points if we have nonzero deadband and sufficient reactive power. Here we use the same initial condition but different stepsizes to achieve various equilibrium points from \mathcal{D}_4 . The results are shown in Fig. 3.3, where the voltages of all 5 controlled buses

are confined within deadband from different equilibria, even though those of other buses may fall out of deadband.

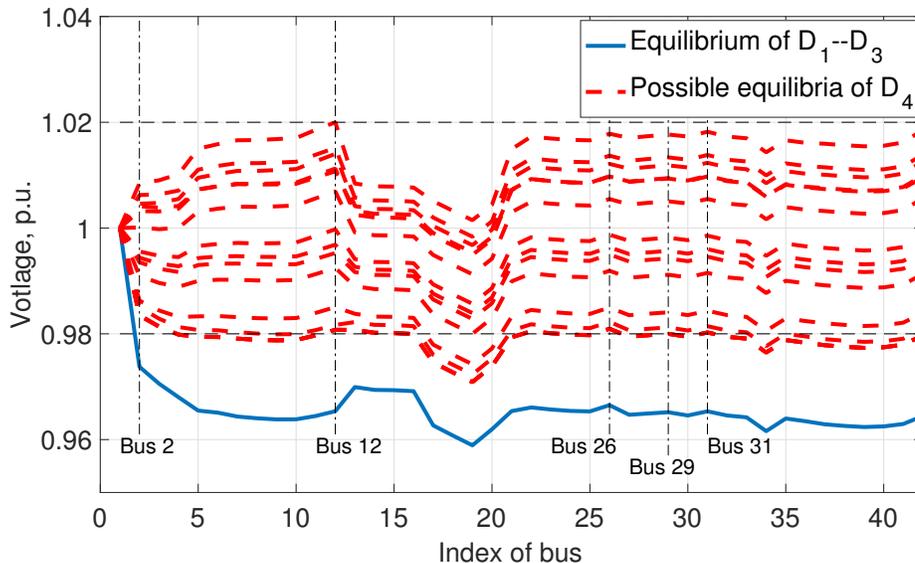


Figure 3.3: Different possible equilibria of \mathcal{D}_4 compared with single equilibrium of $\mathcal{D}_1\text{--}\mathcal{D}_3$. Notice that voltages of 5 controlled buses are all within the 0.98 p.u.–1.02 p.u. deadband, despite that those of other buses may fall out.

3.3.2 Dynamics

3.3.2.1 Convergence of Non-incremental Algorithm \mathcal{D}_1

As shown in Fig. 3.4, the dynamical system \mathcal{D}_1 displays less stable behavior as the control function become steeper with the increase of α_i value, till it ends up with oscillation when α_i becomes too large. See also the vertical dash line in Fig. 3.2, beyond which there is no convergence. As discussed in Section 2.3.3.1, a contention between convergence and equilibrium performance for the non-incremental voltage control (2.12b) exists: a smaller slope is preferred for convergence, while a larger one is selected for better voltage regulation.

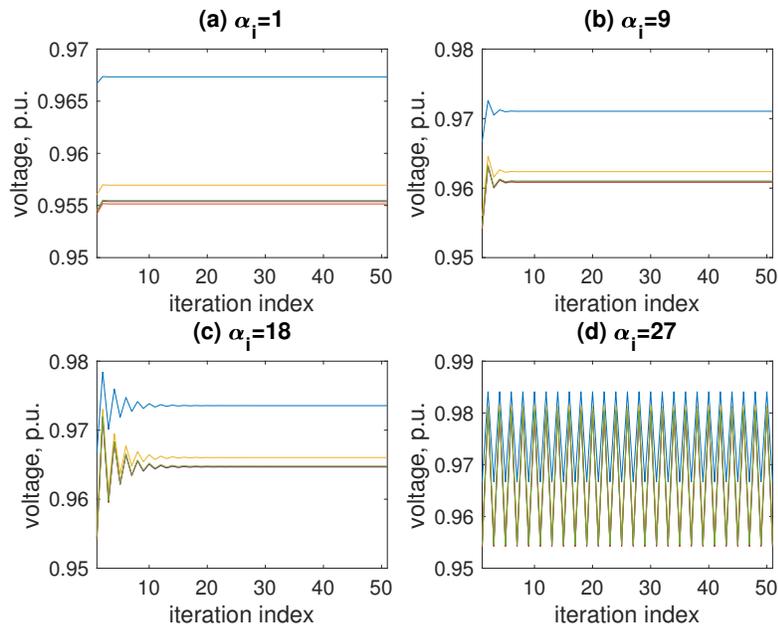


Figure 3.4: Evolution of voltage of the dynamical system \mathcal{D}_1 with different slopes of the piecewise linear control function: Voltage does not converge when the (absolute) slope of the control function become too large (when $\alpha_i > 26$ in this example).

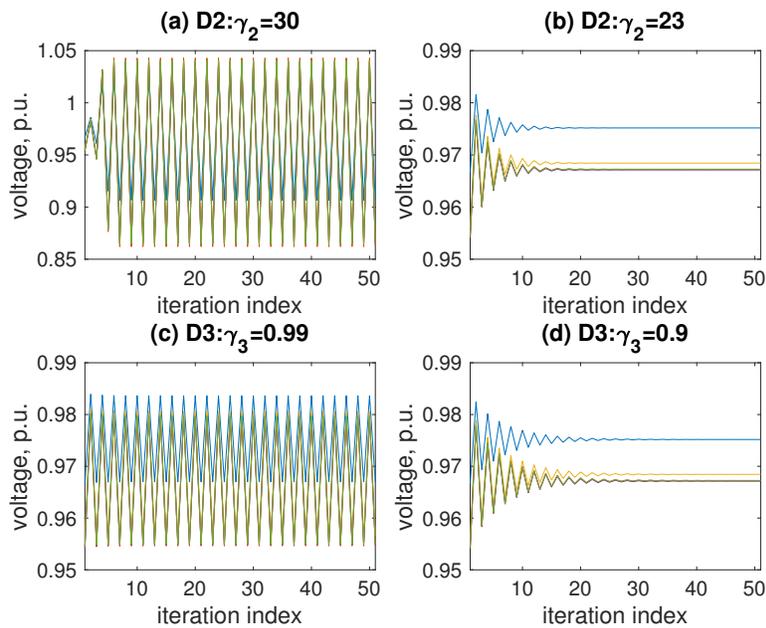


Figure 3.5: Evolution of voltage of the dynamical systems \mathcal{D}_2 and \mathcal{D}_3 with $\alpha_i = 27$: Convergence is ensured with small enough stepsizes.

3.3.2.2 Convergence of Incremental Algorithms \mathcal{D}_2 – \mathcal{D}_4

Proved in Section 2.3.4 and Section 3.1.2, given any control function, \mathcal{D}_2 and \mathcal{D}_3 converge if small enough stepsizes are chosen, and we can thus decouple the equilibrium property from the dynamical property. For instance, when $\alpha_i = 27$, the dynamical system \mathcal{D}_1 does not converge; see Fig. 3.4(b). However, when the stepsizes γ_2 and γ_3 are properly chosen, the dynamical systems \mathcal{D}_2 and \mathcal{D}_3 converge; see Fig. 3.5.

\mathcal{D}_4 also converges as long as stepsize γ_4 is chosen to be small enough regardless of the choice of the control function; see Fig. 3.6.

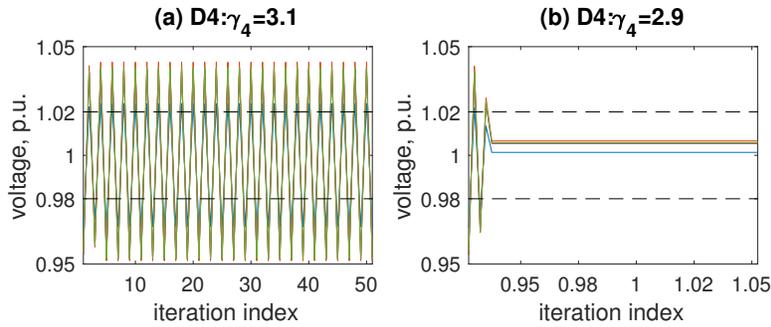


Figure 3.6: Convergence of the dynamical system \mathcal{D}_4 to one of the equilibria with small enough stepsize γ_4 , regardless of the choice of control function.

3.3.2.3 Convergence of \mathcal{D}_2 – \mathcal{D}_3 at Non-Differentiable Point

The dynamical system \mathcal{D}_2 based on subgradient algorithm can only converge to within a small neighborhood of the equilibrium if the equilibrium is near a non-differentiable point of the objective function (7.42a). We tune the parameters such that the equilibrium reactive power provisioned at certain bus — bus 2 in this case — is close to zero. As shown in Fig. 3.7, \mathcal{D}_2 eventually converges to a small region around the optimum, even with very small stepsize chosen. On the other hand, as shown in Fig. 3.8, the dynamical system \mathcal{D}_3 based on pseudo-gradient algorithm converges to the equilibrium despite the non-differentiability of the objective function at the equilibrium; see Fig. 3.7–3.8.

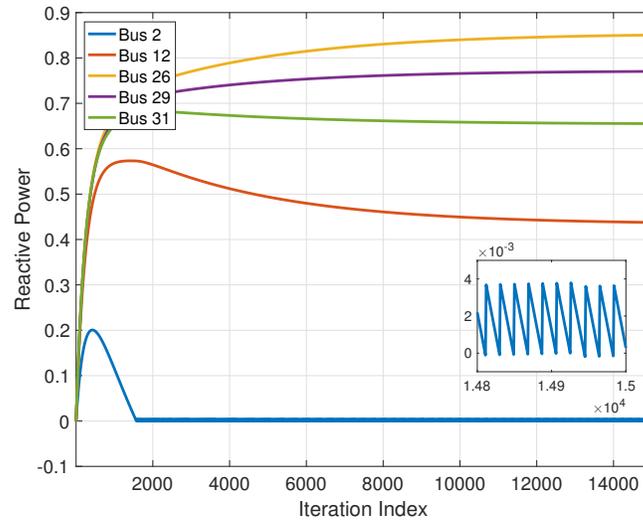


Figure 3.7: Convergence of the dynamical system \mathcal{D}_2 to within a small neighborhood of the equilibrium.

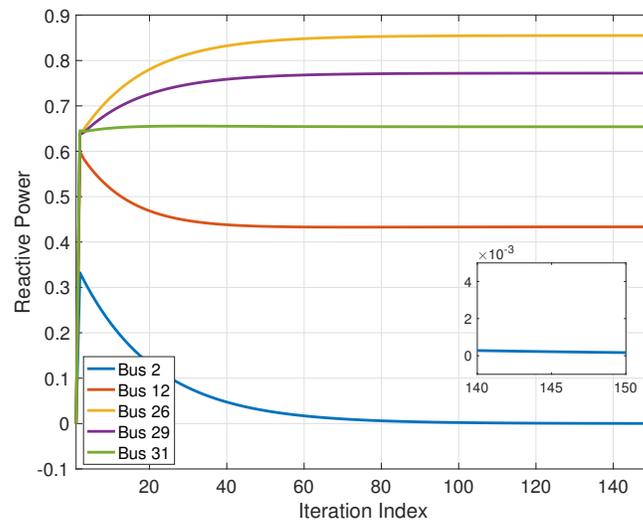


Figure 3.8: Convergence of the dynamical system \mathcal{D}_3 to the equilibrium.

3.4 Conclusion

We have proposed two incremental local Volt/VAR control algorithms that not only achieve better voltage regulation results than previous results, but also possess superior convergence properties and simpler implementation. Performance of the new designed are characterized both analytically and numerically.

Chapter 4

Distributed Voltage Regulation with Continuous Variables

As shown in previous chapters, certain voltage regulation goals can be realized by inverter-based local Volt/VAR control with carefully designed control strategies. However, in various situations, local control has limitations.

In a distribution system that has been increasingly penetrated by various controllable DERs, it is crucial to consider a control framework that is applicable to more diverse devices, with more general constraints explicitly satisfied. Moreover, customers are usually not assumed to follow certain control protocols; they are instead interested in maximizing their own utility—or optimizing a trade-off between their comfort level and electricity bill. This calls for a market-based framework that incentivizes the participation of self-interested customers into the control loop.

This chapter considers a social welfare optimization problem that captures the operational and economic objectives of both network operator and customers as well as the voltage constraints, and design an optimization framework based on a primal-dual gradient algorithm such that the network operator and customers pursue the given operational and economic objectives while concurrently ensuring that the voltages are within the prescribed limits. We consider DERs with continuous decision variables in this chapter, and will extend the results to DERs with discrete output and complex feasible sets in the next chapter. Related work of this chapter has been published in [137, 138].

Main Notation for Chapter 4–6

\mathcal{N}	Set of nodes, excluding node 0; $\mathcal{N} := \{1, \dots, N\}$
\mathcal{E}	Set of distribution lines
p_i	Net real power injected at node i
q_i	Net reactive power injected at node i
z_i	Overall power injected at node i , $z_i := [p_i, q_i]^\top$
\mathcal{Z}_i	Feasible set of real and reactive power at node i
p, q	$p := [p_1, \dots, p_N]^\top$, $q := [q_1, \dots, q_N]^\top$
z	$z := [p^\top, q^\top]^\top$
v_i	Voltage magnitude at node i
v	$v := [v_1, \dots, v_N]^\top$
A, B	Linearization parameters for real and reactive power, respectively
α_i	Signal for injected real power for node i
β_i	Signal for injected reactive power for node i
s_i	Overall signal $z_i := [\alpha_i, \beta_i]^\top$
α, β	$\alpha := [\alpha_1, \dots, \alpha_N]^\top$, $\beta := [\beta_1, \dots, \beta_N]^\top$
s	Compact signal vector $s := [\alpha^\top, \beta^\top]^\top$
$x_{i,d}$	Device state of device d at node i
y_i	System state at node i
$[\cdot]_+$	Projection of x onto the nonnegative orthant
$[\cdot]_{\mathcal{Z}}$	Projection of x onto the convex set \mathcal{Z}

4.1 Preliminaries and System Model

4.1.1 Network Model

Similar to previous Chapters, we consider a distribution network with $N + 1$ nodes collected in the set $\mathcal{N} \cup \{0\}$ with $\mathcal{N} := \{1, \dots, N\}$ and node 0 being the point of common coupling or substation, and distribution

lines collected in the set \mathcal{L} . Let $V_i \in \mathbb{C}$ denote the line-to-ground voltage at node i , and define $v_i := |V_i|$. Denote as $p_i \in \mathbb{R}$ and $q_i \in \mathbb{R}$ the (net) active and reactive power injections, respectively, of a distributed energy resource (DER) located at node $i \in \mathcal{N}$. For notational simplicity, exposition is tailored to the case where one DER is located at each node; however, the technical approach straightforwardly applies to the case where multiple DERs are connected to a node. Hereafter, \mathcal{Z}_i denotes the feasible set of active and reactive powers p_i and q_i at node $i \in \mathcal{N}$. In the following, we explain how to construct this set for some types of DERs.

Photovoltaic (PV) systems: Let $p_{i,\text{av}}$ denote the available real power from a PV system, and let η_i be the rated apparent capacity. Then, the set \mathcal{Z}_i is given by:

$$\mathcal{Z}_i = \{(p_i, q_i): 0 \leq p_i \leq p_{i,\text{av}}, p_i^2 + q_i^2 \leq \eta_i^2\}.$$

Energy storage systems: The set \mathcal{Z}_i for an energy storage system is given by:

$$\mathcal{Z}_i = \{(p_i, q_i): \underline{p}_i \leq p_i \leq \bar{p}_i, p_i^2 + q_i^2 \leq \eta_i^2\},$$

for given limits $\underline{p}_i, \bar{p}_i$ and for a given inverter capacity rating η_i . The limits $\underline{p}_i, \bar{p}_i$ are updated during the operation of the battery based on the state of charge.

Variable frequency drives: For devices such as water pumps and supply fans of commercial HVAC systems, the set \mathcal{Z}_i can be described as:

$$\mathcal{Z}_i = \{(p_i, q_i): \underline{p}_i \leq p_i \leq \bar{p}_i, q_i = 0\},$$

for given limits $\underline{p}_i, \bar{p}_i$. These limits can be fixed or updated by local controllers at a regular time intervals, based on the state of e.g., thermal loads.

The operating region of small-scale diesel generators can be modeled using constant box constraints. For DERs with discrete levels of output powers (e.g., electric vehicle chargers with discrete charging levels), \mathcal{Z}_i represents the convex envelope of the possible operating points; see e.g., [22]. Randomization techniques can then be utilized to recover a feasible setpoint. However, the development of control strategies for DERs with discrete levels of output powers is left as a future research activity.

Voltages, currents, and powers $\{p_i, q_i\}$ are related by the well-known nonlinear AC power-flow equations; assuming, for illustrative purpose, a balanced tree network, these equations read:

$$P_{ij} = -p_j + \sum_{k:(j,k) \in \mathcal{L}} P_{jk} + r_{ij} \ell_{ij}, \quad (4.1a)$$

$$Q_{ij} = -q_j + \sum_{k:(j,k) \in \mathcal{L}} Q_{jk} + x_{ij} \ell_{ij}, \quad (4.1b)$$

$$v_j^2 = v_i^2 - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2) \ell_{ij}, \quad (4.1c)$$

$$\ell_{ij} v_i^2 = P_{ij}^2 + Q_{ij}^2, \quad (4.1d)$$

where ℓ_{ij} is the squared magnitude of the current on line (i, j) , P_{ij} , Q_{ij} are real and reactive powers injected on line (i, j) , and $r_{ij} + jx_{ij}$ is the impedance on line (i, j) .

To facilitate the design and analysis of computationally-tractable algorithms, the proposed approach will employ suitable linearization approaches for (4.1). Particularly, the following approximate linear relationship between voltage magnitudes and injected powers is utilized:

$$v \approx \hat{v} = Ap + Bq + c, \quad (4.2)$$

where the parameters $A, B \in \mathbb{R}_{++}^{N \times N}$ and $a \in \mathbb{R}^N$ can be obtained using one of the two following approaches:

i) regression-based methods, based on real-time measurements of $\{v_i\}$, p , and q , e.g., the recursive least-squares method [11] can be utilized to continuously update the model parameters; and, ii) suitable linearization methods for the AC power-flow equations; see e.g., [16, 28, 36, 44, 50]. In the latter case, the model parameters A , B , and c can be time-varying too, by using current operating points as linearization points for the AC power-flow equations. Parameters A , B , and c should be re-computed every time that the system changes topology.

The approximate model (4.2) is utilized to facilitate the design of computationally-affordable algorithms. Section 4.2.3 will show how to leverage appropriate measurements to cope with approximation errors and systematically enforce voltage limits.

Remark 4.1. (*multiphase systems*) *For notational and exposition simplicity, the framework is outlined for a single-phase system. However, the proposed algorithmic solution is applicable to unbalanced multiphase*

networks. This can be obtained by substituting (4.2) with the linearized model recently proposed in [24] for unbalanced multiphase networks with both wye-connected and delta-connected DERs.

4.1.2 Problem Setup

The goal is to design a strategy wherein the network operator and customers pursue their own operational and economic objectives, while achieving a global coordination to enforce voltage regulation.

4.1.2.1 Customers' Optimization problem

Consider a cost function $C_i(p_i, q_i)$ that captures a well-defined performance objective for the customer(s) located at node $i \in \mathcal{N}$. Let $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}$ be incentive signals produced by the network operator (e.g., distribution system operator or aggregator) for active and reactive power injections, respectively. Given signals (α_i, β_i) , the following optimization problem is solved at each node $i \in \mathcal{N}$:

$$(\mathcal{P}_{1,i}) \quad \min_{p_i, q_i} f_i(p_i, q_i | \alpha_i, \beta_i), \quad (4.3a)$$

$$\text{s.t.} \quad (p_i, q_i) \in \mathcal{Z}_i, \quad (4.3b)$$

where the quasilinear utility function [120] is defined as

$$f_i(p_i, q_i | \alpha_i, \beta_i) := C_i(p_i, q_i) - \alpha_i p_i - \beta_i q_i, \quad (4.4)$$

portraying a trade-off between comfort level $C_i(p_i, q_i)$ and electricity bill with $\alpha_i p_i$ and $\beta_i q_i$ representing payment to/reward from the network operator. The following assumption is made.

Assumption 4.1. *Functions $C_i(p_i, q_i)$, $\forall i \in \mathcal{N}$ are continuously differentiable and strongly convex in (p_i, q_i) . Moreover, the first-order derivative of $C_i(p_i, q_i)$ is bounded in \mathcal{Z}_i .*

The assumption of bounded derivative means that an infinitesimal change in power should not lead to a jump in cost. Because (4.3a) is strictly convex in (p_i, q_i) and \mathcal{Z}_i is convex and compact, a *unique* solution (p_i^*, q_i^*) exists.

For future developments, consider the so-called best response strategy of node i , denoted as $b_i(\alpha_i, \beta_i)$, for given α_i and β_i :

$$(p_i^*, q_i^*) = b_i(\alpha_i, \beta_i) := \arg \min_{(p_i, q_i) \in \mathcal{Z}_i} f_i(p_i, q_i | \alpha_i, \beta_i). \quad (4.5)$$

4.1.2.2 Social-welfare Problem with Voltage Regulation

Consider a cost function $D(\hat{v})$ that captures network-oriented objective in voltage. For example, to minimize the voltage deviation from the nominal value v^{nom} , we can set $D(\hat{v}) = \frac{1}{2} \|\hat{v} - v^{\text{nom}}\|^2$. The following assumption is made.

Assumption 4.2. *Function $D(\hat{v})$ is continuously differentiable, convex, and with bounded first-order derivative at achievable voltage magnitude values.*

Because the feasible set of power injections (p, q) is compact and \hat{v} is a continuous function of (p, q) , the achievable values of \hat{v} are bounded. Thus, the boundedness of the first-order derivative of $D(\hat{v})$ is a reasonable assumption.

Consider the following optimization problem to be solved by the network operator, which captures both customer-oriented and network-oriented objectives of a distribution network:

$$(\mathcal{P}_2) \quad \min_{p, q, \hat{v}, \alpha, \beta} \sum_{i \in \mathcal{N}} C_i(p_i, q_i) + \gamma D(\hat{v}), \quad (4.6a)$$

$$\text{s.t.} \quad \hat{v} = Ap + Bq + c, \quad (4.6b)$$

$$\underline{v} \leq \hat{v} \leq \bar{v}, \quad (4.6c)$$

$$(p_i, q_i) = b_i(\alpha_i, \beta_i), \quad \forall i \in \mathcal{N}, \quad (4.6d)$$

where $\gamma \in \mathbb{R}_+$ is used to trade off between the customers and network-oriented objectives, and \underline{v} and \bar{v} are vectors collecting prescribed minimum and maximum voltage magnitude limits (the inequalities are component-wise). Note that the payment/rewards between network operator and customers are canceled out when we consider the social-welfare problem.

Note that (\mathcal{P}_2) is usually non-convex due to the constraint (4.6d). This is because (4.6d) is usually not affine. For better illustration of the non-convexity of (\mathcal{P}_2) , consider the following simple example with real power only. Assume a quadratic cost function $C_i(p_i) = p_i^2$ and a box feasible set $p_i \in [\underline{p}_i, \bar{p}_i]$ with upper and lower bounds for real power injections \underline{p}_i and \bar{p}_i , and we end up with a non-convex piece-wise linear function b_i :

$$p_i = b_i(\alpha_i) = \begin{cases} \underline{p}_i, & \text{if } \alpha_i/2 < \underline{p}_i \\ \alpha_i/2, & \text{if } \underline{p}_i \leq \alpha_i/2 \leq \bar{p}_i \\ \bar{p}_i, & \text{if } \alpha_i/2 > \bar{p}_i \end{cases}, \quad (4.7)$$

which becomes more complex or does not even have a closed form when we consider more complicated C_i and \mathcal{Z}_i .

4.2 Incentive-Based Distributed Algorithm

(\mathcal{P}_2) lends itself to a Stackelberg game interpretation where α and β are calculated via (\mathcal{P}_2) by the network operator (i.e., the leader) and sent to each nodes $i \in \mathcal{N}$; subsequently, each consumer (i.e., the follower) computes the power setpoints p_i^* and q_i^* from $(\mathcal{P}_{1,i})$. By design, (p^*, q^*) is an optimal point of (\mathcal{P}_2) .

However, it is challenging for the network operator to solve (\mathcal{P}_2) not only because of the non-convexity introduced by constraint (4.6d), but also because it requires knowledge of the customer's best-response function b_i (or equivalently, C_i and \mathcal{Z}_i , $\forall i \in \mathcal{N}$). To solve the problem, in Section 4.2.1 we first formulate a convex relaxation of (\mathcal{P}_2) and show that its optimum gives the optimum of (\mathcal{P}_2) , and then in Section 4.2.2 we design a distributed algorithm to solve (\mathcal{P}_2) based on the algorithm for the relaxed problem.

4.2.1 Convex Reformulation

We start by deriving a convex relaxation of the non-convex problem (\mathcal{P}_2) as well as conditions under which an optimal point of (\mathcal{P}_2) can be identified. Consider the following convex optimization problem:

$$(\mathcal{P}_3) \quad \min_{p, q, \hat{v}} \sum_{i \in \mathcal{N}} C_i(p_i, q_i) + \gamma D(\hat{v}), \quad (4.8a)$$

$$s.t. \quad \hat{v} = Ap + Bq + c, \quad (4.8b)$$

$$\underline{v} \leq \hat{v} \leq \bar{v}, \quad (4.8c)$$

$$(p_i, q_i) \in \mathcal{Z}_i, \quad \forall i \in \mathcal{N}, \quad (4.8d)$$

where we replace the non-convex constraint (4.6d) in (\mathcal{P}_2) with (4.3b), and signals α and β are to be determined later. We assume that the above problem is feasible.

Assumption 4.3 (Slater's condition). *There exists a feasible point $(\tilde{p}, \tilde{q}) \in \mathcal{Z}$, $\mathcal{Z} := \mathcal{Z}_1 \times \dots \times \mathcal{Z}_N$, such that:*

$$\underline{v} \leq A\tilde{p} + B\tilde{q} + c \leq \bar{v}. \quad (4.9)$$

Assumption 4.3 does not involve strict inequality because the constraint is linear. Given the strong convexity of the objective function (4.8a) in (p_i, q_i) and the linear relation (4.8b), a unique optimal solution exists for problem (\mathcal{P}_3) . Notice that a solution $(p_i^*, q_i^*, \hat{v}^*)$ of (\mathcal{P}_3) may not be feasible for (\mathcal{P}_2) , i.e., there does not exist a (α^*, β^*) such that $(p_i^*, q_i^*) = b_i(\alpha_i^*, \beta_i^*)$. We will, however, show next that such a (α^*, β^*) exists, and thus the solution of (\mathcal{P}_3) gives the solution of (\mathcal{P}_2) .

Denote by $\underline{\mu}$ and $\bar{\mu}$ the dual variables associated with the two inequality constraints in (4.8c). Let \hat{v}^* be the optimal voltage magnitudes produced by (\mathcal{P}_3) and $\underline{\mu}^*, \bar{\mu}^*$ the optimal dual variables. Then, we propose to design the incentive signals as follows:

$$\alpha^* = A^\top (\underline{\mu}^* - \bar{\mu}^* - \gamma \nabla_{\hat{v}} D(\hat{v}^*)), \quad (4.10a)$$

$$\beta^* = B^\top (\underline{\mu}^* - \bar{\mu}^* - \gamma \nabla_{\hat{v}} D(\hat{v}^*)), \quad (4.10b)$$

where $\nabla_{\hat{v}} D$ denotes the gradient of function D with respect to the vector \hat{v} . Note that α^* and β^* are composed of dual prices $\underline{\mu}^*, \bar{\mu}^*$ and the marginal cost of network operator $\gamma \nabla_{\hat{v}} D(\hat{v}^*)$, together with A, B characterizing

the network structure. As will be shown shortly, α^* and β^* are in fact designed based on the optimality conditions of (\mathcal{P}_2) and (\mathcal{P}_3) . The above incentive signals are bounded, which precludes the possibility of infinitely large signals.

Theorem 4.1. *Under Assumptions 4.1–4.3, the incentive signals (α^*, β^*) defined by (4.10) are bounded.*

Proof. Notice that the derivative $\nabla_{\hat{v}} D$ is bounded. To show the boundedness of (α^*, β^*) , it is enough to show that the optimal duals $(\underline{\mu}^*, \bar{\mu}^*)$ are bounded.

Consider the KKT conditions for problem (\mathcal{P}_3) :

$$\left(\nabla_p \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) + \gamma A^\top \nabla_{\hat{v}} D(\hat{v}^*) - A^\top (\underline{\mu}^* - \bar{\mu}^*) \right)^\top (p - p^*) \geq 0, \forall (p, q) \in \mathcal{Z}, \quad (4.11a)$$

$$\left(\nabla_q \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) + \gamma B^\top \nabla_{\hat{v}} D(\hat{v}^*) - B^\top (\underline{\mu}^* - \bar{\mu}^*) \right)^\top (q - q^*) \geq 0, \forall (p, q) \in \mathcal{Z}, \quad (4.11b)$$

$$\hat{v}^* = Ap^* + Bq^* + c, \quad (4.11c)$$

$$\underline{v} \leq \hat{v}^* \leq \bar{v}, \quad (4.11d)$$

$$(\hat{v}^* - \underline{v})^\top \underline{\mu}^* = 0, \quad \underline{\mu}^* \geq 0, \quad (4.11e)$$

$$(\bar{v} - \hat{v}^*)^\top \bar{\mu}^* = 0, \quad \bar{\mu}^* \geq 0. \quad (4.11f)$$

Combining (4.11a)–(4.11c) results in:

$$\begin{aligned} & \left(\nabla_p \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) + \gamma A^\top \nabla_{\hat{v}} D(\hat{v}^*) \right)^\top (p - p^*) + \left(\nabla_q \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) + \gamma B^\top \nabla_{\hat{v}} D(\hat{v}^*) \right)^\top (q - q^*) \\ & + (\bar{\mu}^* - \underline{\mu}^*)^\top (\hat{v} - \hat{v}^*) \geq 0, \forall (p, q) \in \mathcal{Z}, \forall \hat{v}, \end{aligned} \quad (4.12)$$

where the first two terms on the left of the inequality are bounded because of the bounded derivative of cost functions and the bounded set \mathcal{Z} . By the complementary slackness conditions (4.11e)–(4.11f), $\bar{\mu}_i^*$ and $\underline{\mu}_i^*$, $i \in \mathcal{N}$ cannot be nonzero at the same time. If $\bar{\mu}_i^* \rightarrow \infty$, then $\hat{v}_i^* = \bar{v}_i$ and we can choose a (p, q) and thus \hat{v}_i such that the third term on the left of (4.12) goes to $-\infty$ and (4.12) does not hold. So, $\bar{\mu}_i^*$ and thus $\bar{\mu}^*$ is bounded. Similarly, we can show that $\underline{\mu}^*$ is bounded too. The result follows. ■

By examining the optimality conditions of (\mathcal{P}_2) and (\mathcal{P}_3) , we have the following result.

Theorem 4.2. *The solutions of problem (\mathcal{P}_3) along with the signals (α^*, β^*) defined in (4.10) are global optimal solutions of problem (\mathcal{P}_2) ; i.e., problem (\mathcal{P}_3) is an exact convex relaxation of problem (\mathcal{P}_2) .*

Proof. By the signal design (4.10), (4.11a)–(4.11b) become

$$\left(\nabla_p \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) - \alpha^*\right)^\top (p - p^*) \geq 0, \quad \forall (p, q) \in \mathcal{Z}, \quad (4.13a)$$

$$\left(\nabla_q \sum_{i \in \mathcal{N}} C_i(p_i^*, q_i^*) - \beta^*\right)^\top (q - q^*) \geq 0, \quad \forall (p, q) \in \mathcal{Z}. \quad (4.13b)$$

Notice that the above variational inequalities imply that $(p_i^*, q_i^*) = b_i(\alpha_i^*, \beta_i^*)$, $i \in \mathcal{N}$. So, the solution of problem (\mathcal{P}_3) along with (α^*, β^*) defined in (4.10) is feasible for problem (\mathcal{P}_2) . The result follows, as (\mathcal{P}_3) is an exact convex relaxation of (\mathcal{P}_2) . \blacksquare

From now on, we will use the optima of (\mathcal{P}_3) and (\mathcal{P}_2) interchangeably depending on the context. Next, based on Theorem 4.2, we will develop an iterative algorithm that achieves the optimum of (\mathcal{P}_3) (and hence that of (\mathcal{P}_2)) without exposing any private information of the customers to the network operator.

Remark 4.2. *Theorem 4.2 asserts that non-convex problem (\mathcal{P}_2) can be solved through solving a convex problem (\mathcal{P}_3) . At first glance, it appears that the non-convexity of (\mathcal{P}_2) comes from a non-convex representation of the feasible set that may have a convex representation as implied by (\mathcal{P}_3) . An ongoing investigation is to identify the specific problem structure to generalize the result in Theorem 4.2 to a larger class of problems.*

4.2.2 Distributed Algorithm

For notational simplicity, let $s_i = [\alpha_i, \beta_i]^\top$ denote the overall signals for customer i and define $z_i = [p_i, q_i]^\top$. Further denote by $z := [p^\top, q^\top]^\top \in \mathbb{R}^{2N}$ the vector of stacked power injections, and by $\mu := [\underline{\mu}^\top, \bar{\mu}^\top]^\top \in \mathbb{R}_+^{2N}$ the vector of stacked dual variables. Recall that \hat{v} is a function of z and consider the following Lagrangian function associated with (\mathcal{P}_3) :

$$\mathcal{L}(z, \mu) = \sum_{i \in \mathcal{N}} C_i(z_i) + \gamma D(z) + \underline{\mu}^\top (\underline{v} - Ap - Bq - c) + \bar{\mu}^\top (Ap + Bq + c - \bar{v}), \quad (4.14)$$

which is obtained by keeping the constraints $z \in \mathcal{Z}$ and $\mu \in \mathbb{R}_+^{2N}$ implicit. Denote as (z^*, μ^*) a saddle-point of $\mathcal{L}(z, \mu)$.

To facilitate the development of provably convergent algorithms, consider the following regularized

Lagrangian function:

$$\mathcal{L}_\phi(z, \mu) := \sum_{i \in \mathcal{N}} C_i(z_i) + \gamma D(z) + \underline{\mu}^\top (\underline{v} - Ap - Bq - c) + \bar{\mu}^\top (Ap + Bq + c - \bar{v}) - \frac{\phi}{2} \|\mu\|^2, \quad (4.15)$$

where $\phi > 0$ is a predefined parameter (see e.g., [69, 109]). With the regularization term $-\frac{\phi}{2} \|\mu\|^2$, the resultant function $\mathcal{L}_\phi(z, \mu)$ is strongly concave in the dual variables. Based on (4.14), we proceed with the following minimax problem:

$$\max_{\mu \in \mathbb{R}_+^{2N}} \min_{z \in \mathcal{Z}} \mathcal{L}_\phi(z, \mu). \quad (4.16)$$

In general, the unique optimizer of (4.16), denoted by (z_ϕ^*, μ_ϕ^*) , is not a saddle-point of the Lagrangian function (4.14) because of the regularization term $-\frac{\phi}{2} \|\mu\|^2$. However, the discrepancy between the unique optimizer of (4.16) and the optimizers of (4.14) can be bounded as shown next.

Notice first that the boundedness of μ^* is shown in Theorem 4.1; μ_ϕ^* can be readily shown to be bounded too. For ease of exposition, define $f(z) := \sum_{i \in \mathcal{N}} C_i(z_i) + \gamma D(z)$ and $g(z) := \begin{bmatrix} \underline{v} - Ap - Bq - c \\ Ap + Bq + c - \bar{v} \end{bmatrix}$; this way, the Lagrangian can be re-expressed in a compact form as $\mathcal{L}(z, \mu) = f(z) + \mu^\top g(z)$ and the regularized counterpart reads $\mathcal{L}_\phi(z, \mu) = f(z) + \mu^\top g(z) - \frac{\phi}{2} \|\mu\|^2$. From Assumption 4.1–4.2, it follows that f is strongly convex in z . Equivalently, $\nabla_z f(z, \mu)$ is strongly monotone in z . Therefore, we have the following lemma:

Lemma 4.1. *There exists a scalar $m > 0$ such that $\forall z, z' \in \mathcal{Z}$,*

$$(\nabla_z f(z, \mu) - \nabla_z f(z', \mu))^\top (z - z') \geq m \|z - z'\|^2. \quad (4.17)$$

Then, the discrepancy between z^* and z_ϕ^* due to the regularization term can be bounded as follows (see also [69, Proposition 3.1]).

Theorem 4.3. *The difference between z_ϕ^* and z^* is bounded as:*

$$\|z_\phi^* - z^*\|^2 \leq \frac{\phi}{2m} (\|\mu^*\|^2 - \|\mu_\phi^*\|^2). \quad (4.18)$$

Proof. As a saddle point of (4.16), (z_ϕ^*, μ_ϕ^*) satisfies the following inequalities:

$$\mathcal{L}_\phi(z_\phi^*, \mu) \leq \mathcal{L}_\phi(z_\phi^*, \mu_\phi^*) \leq \mathcal{L}_\phi(z, \mu_\phi^*), \quad \forall z, \mu.$$

The left inequality leads to

$$(\mu_\phi^* - \mu^*)^\top g(z_\phi^*) - \frac{\phi}{2} \|\mu_\phi^*\|^2 + \frac{\phi}{2} \|\mu^*\|^2 \geq 0, \quad (4.19)$$

where we set $\mu = \mu^*$. We next characterize the term $(\mu_\phi^* - \mu^*)^\top g(z_\phi^*)$.

(i) Leveraging the definition of convex functions, $g_j(z_\phi^*)$ can be upper bounded as:

$$\begin{aligned} g_j(z_\phi^*) &\leq g_j(z^*) + \nabla_z g_j(z_\phi^*)^\top (z_\phi^* - z^*) \\ &\leq \nabla_z g_j(z_\phi^*)^\top (z_\phi^* - z^*), \end{aligned} \quad (4.20)$$

where the second inequality is due to the fact that $g_j(z^*) \leq 0$. Multiply both sides of (4.20) by $\mu_{\phi,j}^*$ (which is nonnegative) and sum up for all j to have:

$$\begin{aligned} \mu_\phi^{*\top} g(z_\phi^*) &\leq \sum_j \mu_{\phi,j}^* \cdot \nabla_z g_j(z_\phi^*)^\top (z_\phi^* - z^*) \\ &= \nabla_z \mathcal{L}_\phi(z_\phi^*, \mu_\phi^*)^\top (z_\phi^* - z^*) - \nabla_z f(z_\phi^*)^\top (z_\phi^* - z^*) \\ &\leq -\nabla_z f(z_\phi^*)^\top (z_\phi^* - z^*), \end{aligned} \quad (4.21)$$

where the second inequality is due to the first-order optimality condition $\nabla_z \mathcal{L}_\phi(z_\phi^*, \mu_\phi^*)^\top (z_\phi^* - z^*) \leq 0$.

(ii) On the other hand, one has that:

$$g_j(z_\phi^*) \geq g_j(z^*) + \nabla_z g_j(z^*)^\top (z_\phi^* - z^*). \quad (4.22)$$

Multiply both sides of (4.22) by $-\mu_j^*$ (which is nonpositive) and sum up for all j to get:

$$\begin{aligned} -\mu^{*\top} g(z_\phi^*) &\leq -\sum_j \mu_j^* g_j(z^*) - \sum_j \mu_j^* \cdot \nabla_z g_j(z^*)^\top (z_\phi^* - z^*) \\ &= \sum_j \mu_j^* \cdot \nabla_z g_j(z^*)^\top (z^* - z_\phi^*) \\ &= \nabla_z \mathcal{L}(z^*, \mu^*)^\top (z^* - z_\phi^*) - \nabla_z f(z^*)^\top (z^* - z_\phi^*) \\ &\leq \nabla_z f(z^*)^\top (z_\phi^* - z^*), \end{aligned} \quad (4.23)$$

where the first equality is due to the complimentary slackness condition and the second inequality is obtained from the first-order optimality condition.

Substitute (4.21) and (4.23) into (4.19), and use (4.17) to obtain (4.18). ■

The key advantage of utilizing the regularized Lagrangian is that the primal-dual gradient methods applied to (4.16) exhibit improved convergence properties [109] as explained next.

Hereafter, we omit the subscript ϕ from the optimization variables for notational simplicity, with the understanding that the updates of $z(k)$ and $\mu(k)$ are designed to solve the regularized saddle-point problem (4.16). Consider the following primal-dual projected gradient method, where k denotes the iteration index:

$$\begin{bmatrix} z(k+1) \\ \mu(k+1) \end{bmatrix} = \hat{T} \left(\begin{bmatrix} z(k) \\ \mu(k) \end{bmatrix} \right) := \left[\begin{bmatrix} z(k) \\ \mu(k) \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \nabla_z \mathcal{L}_\phi(z(k), \mu(k)) \\ -\varepsilon_2 \nabla_\mu \mathcal{L}_\phi(z(k), \mu(k)) \end{bmatrix} \right]_{\mathcal{Z} \times \mathbb{R}_+^{2N}}, \quad (4.24)$$

where $[\]_{\mathcal{Z} \times \mathbb{R}_+^{2N}}$ denotes the projection operation onto the set $\mathcal{Z} \times \mathbb{R}_+^{2N}$, and $\varepsilon_1, \varepsilon_2 > 0$ are prescribed step sizes for the primal and the dual updates. Notice that $\nabla_z \mathcal{L}_\phi(z, \mu)$ and $\nabla_\mu \mathcal{L}_\phi(z, \mu)$ are Lipschitz continuous and strongly monotone. Therefore by virtue of [25, Sec. 3.5, Proposition 5.4], the following result holds.

Theorem 4.4. *There exist some $\bar{\varepsilon}_1, \bar{\varepsilon}_2 > 0$ such that for any $\varepsilon_1 \in (0, \bar{\varepsilon}_1], \varepsilon_2 \in (0, \bar{\varepsilon}_2]$, \hat{T} is a contraction mapping. For $\varepsilon_1 \in (0, \bar{\varepsilon}_1], \varepsilon_2 \in (0, \bar{\varepsilon}_2]$, the sequence $\{(z(k), \mu(k))\}$ generated by (4.24) converges geometrically to the optimizer of (4.16).*

We can further provide analytical bound for such $\bar{\varepsilon}_1, \bar{\varepsilon}_2$ for completeness. We also refer to Section 4.3.2.1) for numerical characterization of step sizes as regards convergence. In the following we provide sufficient conditions on the stepsizes ε_1 and ε_2 that guarantee the operator \hat{T} in (4.24) to be a contraction.

Theorem 4.5. *If the stepsizes ε_1 and ε_2 satisfy the following conditions for any $i \in \mathcal{N}$:*

$$\varepsilon_2 < \frac{1}{2 \sum_{j \in \mathcal{N}} A_{ij}}, \quad (4.25a)$$

$$\varepsilon_1 \nabla_{p_i}^2(C_i + \gamma D) > 2\varepsilon_2 \sum_{j \in \mathcal{N}} A_{ij}, \quad (4.25b)$$

$$\varepsilon_1 \nabla_{p_i}^2(C_i + \gamma D) + 2\varepsilon_2 \sum_{j \in \mathcal{N}} A_{ij} < 2, \quad (4.25c)$$

$$\varepsilon_2 < \frac{1}{2 \sum_{j \in \mathcal{N}} B_{ij}}, \quad (4.25d)$$

$$\varepsilon_1 \nabla_{q_i}^2(C_i + \gamma D) > 2\varepsilon_2 \sum_{j \in \mathcal{N}} B_{ij}, \quad (4.25e)$$

$$\varepsilon_1 \nabla_{q_i}^2(C_i + \gamma D) + 2\varepsilon_2 \sum_{j \in \mathcal{N}} B_{ij} < 2, \quad (4.25f)$$

$$\varepsilon_1 < \frac{1}{\sum_{j \in \mathcal{N}} (A_{ij} + B_{ij})}, \quad (4.25g)$$

$$\varepsilon_1 \sum_{j \in \mathcal{N}} (A_{ij} + B_{ij}) > \varepsilon_2 \phi, \quad (4.25h)$$

$$\varepsilon_1 \sum_{j \in \mathcal{N}} (A_{ij} + B_{ij}) + \varepsilon_2 \phi < 2, \quad (4.25i)$$

then \hat{T} is a contraction.

Proof. Let $\nabla \hat{T} \in \mathbb{R}^{4N \times 4N}$ denote Jacobian matrix of \hat{T} , and let $\nabla \hat{T}_{ij}$ denote the element on row i and column j of the Jacobian matrix $\nabla \hat{T}$. To prove that \hat{T} is a contraction, it is sufficient to have the following condition:

$$\sum_j |\nabla \hat{T}_{ij}| < 1, \quad \forall i,$$

which is satisfied if the following three inequalities hold:

$$|1 - \varepsilon_1(\nabla_{p_i}^2(C_i + \gamma D))| + 2\varepsilon_2 \sum_{j \in \mathcal{N}} A_{ij} < 1, \quad (4.26a)$$

$$|1 - \varepsilon_1(\nabla_{q_i}^2(C_i + \gamma D))| + 2\varepsilon_2 \sum_{j \in \mathcal{N}} B_{ij} < 1, \quad (4.26b)$$

$$|1 - \varepsilon_2 \phi| + \varepsilon_1 \sum_{j \in \mathcal{N}} (A_{ij} + B_{ij}) < 1. \quad (4.26c)$$

Conditions (4.25) and (4.26) are equivalent. Therefore, (4.25) are sufficient for \hat{T} to be a contraction. \blacksquare

Remark 4.3. *Conditions (4.25) together with assumptions in this chapter guarantees the existence of small enough step sizes ε_1 and ε_2 to achieve convergence. This result is consistent with Theorem 5.3.*

Given Theorems 4.3–4.5, algorithm (4.24) converges to within a small neighborhood of problem (\mathcal{P}_3) (problem (\mathcal{P}_2)) whose size can be controlled by choosing a proper weight ϕ for the regularization term. Apply contraction mapping theorem, and we immediately have the following result.

Theorem 4.6. *Under the above modeling assumptions, it follows that dynamics (4.24) converges to the saddle point of (4.16), i.e.,*

$$\lim_{k \rightarrow \infty} \left\| [z(k)^\top, \mu(k)^\top]^\top - [z_\phi^{*\top}, \mu_\phi^{*\top}]^\top \right\| = 0.$$

The decomposable structure of (4.24) naturally enables the following iterative *distributed* algorithm:

$$z_i(k+1) = [z_i(k) - \varepsilon_1(\nabla_z C_i(z_i(k)) - s_i(k))]_{\mathcal{Z}_i}, \quad i \in \mathcal{N}, \quad (4.27a)$$

$$\underline{\mu}(k+1) = [\underline{\mu}(k) + \varepsilon_2(\underline{v} - \hat{v}(k) - \phi \underline{\mu}(k))]_{+}, \quad (4.27b)$$

$$\bar{\mu}(k+1) = [\bar{\mu}(k) + \varepsilon_2(\hat{v}(k) - \bar{v} - \phi \bar{\mu}(k))]_{+}, \quad (4.27c)$$

$$\alpha(k+1) = A^\top (\underline{\mu}(k+1) - \bar{\mu}(k+1) - \gamma \nabla_{\hat{v}} D(\hat{v}(k))), \quad (4.27d)$$

$$\beta(k+1) = B^\top (\underline{\mu}(k+1) - \bar{\mu}(k+1) - \gamma \nabla_{\hat{v}} D(\hat{v}(k))), \quad (4.27e)$$

$$\hat{v}(k+1) = Ap(k+1) + Bq(k+1) + c, \quad (4.27f)$$

where the power setpoints of each device are computed locally through (4.27a) and (4.27b)–(4.27f) are performed at the network operator. The resultant scheme is tabulated as Algorithm 1. Notice that each customer i *does not* share its cost function C_i or its feasible set \mathcal{Z}_i with the network operator; rather, the customer transmits to the network operator only the resultant power injections $z_i(k)$. Indeed, the results of Theorem 4.3–4.6 apply to (4.27) too.

Remark 4.4. *In (4.27), α and β are utilized by the customers to construct the primal gradient $\nabla_z \mathcal{L}_\phi$. This strategy enables a distributed implementation of the primal-dual projected gradient method (4.24) without the customers knowing information of the network.*

4.2.3 Performance Analysis with Nonlinear Power Flow

The iterative algorithm (4.27) is designed and analyzed with a linearized voltage model (4.2). However, we now analyze the stability of Algorithm 1 in the purview of the nonlinear AC power-flow model (cf.

Algorithm 1 Incentive-based iterative algorithm**repeat**[S1] Customer $i \in \mathcal{N}$ performs (4.27a) and sends $z_i(k+1)$ to network operator.

[S2] Network operator performs steps (4.27b)–(4.27f).

[S3] Network operator transmits signals $s_i(k+1)$ to customer $i \in \mathcal{N}$.**until** Stopping criterion is met

(4.1)). Particularly, the step (4.27f) is no longer executed to estimate the voltages; rather, the voltages are *computed based on a nonlinear AC power-flow model* or *directly measured* from the network.

For the rest of this section, hatted symbols (e.g., \hat{v}) refer to variables used in the linearized voltage model; on the other hand, non-hatted symbols represent electrical quantities obeying the nonlinear AC power-flow model. Accordingly, the control strategy is modified as follows:

$$z_i(k+1) = [z_i(k) - \varepsilon_1(\nabla C_i(z_i(k)) - s_i(k))]_{\mathcal{Z}_i}, \quad i \in \mathcal{N}, \quad (4.28a)$$

$$\underline{\mu}(k+1) = [\underline{\mu}(k) + \varepsilon_2(\underline{v} - v(k) - \phi \underline{\mu}(k))]_{+}, \quad (4.28b)$$

$$\bar{\mu}(k+1) = [\bar{\mu}(k) + \varepsilon_2(v(k) - \bar{v} - \phi \bar{\mu}(k))]_{+}, \quad (4.28c)$$

$$\alpha(k+1) = A^\top [\underline{\mu}(k+1) - \bar{\mu}(k+1) - \gamma \nabla_v D(v(k))], \quad (4.28d)$$

$$\beta(k+1) = B^\top [\underline{\mu}(k+1) - \bar{\mu}(k+1) - \gamma \nabla_v D(v(k))], \quad (4.28e)$$

$$v(k+1) \quad \text{obey the nonlinear model (4.1)}. \quad (4.28f)$$

The power setpoints are updated at each node $i \in \mathcal{N}$ via (4.28a) and commanded to the devices; steps (4.28b)–(4.28e) are performed by the the network operator; and, voltages are either computed based on a nonlinear AC power-flow model (e.g., using OpenDSS) or directly measured.

To establish convergence of (4.28), the following is assumed (see also [35, 142]).

Assumption 4.4. *There exists a constant $e > 0$ such that $|\hat{v}_i(z) - v_i(z)| \leq e$, $i \in \mathcal{N}$, for all $z \in \mathcal{Z}$.*

This assumption bounds the discrepancy between voltages generated based on the linearized model and the actual voltages (obtained from the nonlinear AC power-flow equations or measured). By compar-

ing (4.27) with (4.28), Assumption 4.4 naturally leads to the following bounds:

$$\begin{aligned} |\hat{\underline{\mu}}_i - \underline{\mu}_i| &\leq \varepsilon_2 e, & |\hat{\bar{\mu}}_i - \bar{\mu}_i| &\leq \varepsilon_2 e, \\ |\hat{\alpha}_i - \alpha_i| &\leq A_i^\top (\gamma \nabla^2 D(\tilde{v}) \mathbf{1}_n + \varepsilon_2) e, \\ |\hat{\beta}_i - \beta_i| &\leq B_i^\top (\gamma \nabla^2 D(\tilde{v}) \mathbf{1}_n + \varepsilon_2) e, \end{aligned}$$

for some \tilde{v} . The following bounds can be readily shown too:

$$\begin{aligned} |\hat{p}_i - p_i| &\leq \varepsilon_1 A_i^\top (\gamma \nabla^2 D(\tilde{v}) \mathbf{1}_n + \varepsilon_2) e := \delta_{1,i}, \\ |\hat{q}_i - q_i| &\leq \varepsilon_1 B_i^\top (\gamma \nabla^2 D(\tilde{v}) \mathbf{1}_n + \varepsilon_2) e := \delta_{2,i}. \end{aligned}$$

Let $\delta := [\delta_{1,1}, \dots, \delta_{1,N}, \delta_{2,1}, \dots, \delta_{2,N}] \in \mathbb{R}_+^{2N \times 1}$, and collect the primal and dual variables the vector $y := (z, \mu)$ for notational simplicity. Consequently, one has that:

$$\|\hat{T}(y) - T(y)\| \leq \|\rho\|, \quad \forall y \in \mathcal{Z} \times \mathbb{R}_+^{2N}, \quad (4.29)$$

where $\rho := [\varepsilon_2 e \cdot \mathbf{1}_{1 \times 2N}, \delta^\top]^\top$ and $T(\cdot)$ is the counterpart of $\hat{T}(\cdot)$ for the iterates (4.28). Let $\Delta \leq \bar{\Delta} < 1$ be the contraction modulus for $\hat{T}(\cdot)$ with appropriate stepsizes ε_1 and ε_2 chosen according to Theorem 4.5; by definition, we have that:

$$\|\hat{T}(y) - \hat{T}(y')\| \leq \Delta \|y - y'\|, \quad \forall y, y' \in \mathcal{Z} \times \mathbb{R}_+^{2N}, \quad (4.30)$$

and the following result can be established.

Theorem 4.7. *Iterates (4.28) converge to the unique saddle point of (4.16) within a ball of radius $\frac{\|\rho\|}{1-\Delta}$.*

Proof. Let \hat{y}_ϕ^* be the unique optimizer of (4.16) and consider bounding $\|y(k) - \hat{y}_\phi^*\|$ as follows:

$$\begin{aligned} &\|y(k) - \hat{y}_\phi^*\| \\ &= \|T(y(k-1)) - \hat{y}_\phi^*\| \\ &= \|T(y(k-1)) - \hat{T}(y(k-1)) + \hat{T}(y(k-1)) - \hat{y}_\phi^*\| \\ &\leq \|T(y(k-1)) - \hat{T}(y(k-1))\| + \|\hat{T}(y(k-1)) - \hat{y}_\phi^*\| \\ &\leq \|\rho\| + \|\hat{T}(y(k-1)) - \hat{T}(\hat{y}_\phi^*)\| \\ &\leq \|\rho\| + \Delta \|y(k-1) - \hat{y}_\phi^*\|, \end{aligned} \quad (4.31)$$

$$\leq \|\rho\| + \Delta \|y(k-1) - \hat{y}_\phi^*\|, \quad (4.32)$$

where the first inequality is due to the triangle inequality; the second inequality is due to Assumption 4.4; and the last inequality leverages the definition of contraction mapping. By repeating steps (4.31)-(4.32) recursively, one can obtain:

$$\|y(k) - \hat{y}_\phi^*\| \leq \frac{\|\rho\|(1 - \Delta^k)}{1 - \Delta} + \Delta^k \|y(0) - \hat{y}_\phi^*\|. \quad (4.33)$$

When $k \rightarrow \infty$, one has that:

$$\limsup_{k \rightarrow \infty} \|y(k) - \hat{y}_\phi^*\| = \frac{\|\rho\|}{1 - \Delta}. \quad (4.34)$$

■

4.3 Numeric Examples

4.3.1 Simulation Setup

Consider a modified version of the IEEE 37-node test feeder shown in Figure 4.1. The modified network is obtained by considering the phase ‘‘c’’ of the original system and by replacing the loads specified in the original dataset with real load data measured from feeders in Anatolia, California, during a week of August 2012 [15]. Line impedances, shunt admittances, as well as active and reactive loads are adopted from the respective data set. It is assumed that 18 PV systems are located at nodes 4, 7, 10, 13, 17, 20, 22, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35, and 36, and their generation profiles are simulated based on the real solar irradiance data available in [15]. The ratings of these inverters are 300 kVA for $i = 3$, 350 kVA for $i = 15, 16$, and 200 kVA for the remaining inverters.

The voltage limits \bar{v}_i and \underline{v}_i are set to 1.05 p.u. and 0.95 p.u. respectively, for $\forall i \in \mathcal{N}$. Various step sizes ε_1 and ε_2 are tested to provide examples of cases where the algorithm converges as well as cases where it is not convergent. The customers’ objective functions are set uniformly to $C_i(p_i, q_i) = c_p(p_{i,av} - p_i)^2 + c_q q_i^2$, in an effort to minimize the amount of real power curtailed from the available power $p_{i,av}$ based on irradiance conditions, and the amount of reactive power injected or absorbed. The coefficients are set to $c_p = 3$ and $c_q = 1$. The network-oriented objective is set to $D(v) = \frac{1}{2} \|v - v^{\text{nom}}\|_2^2$ to penalize voltage deviation from the nominal value $v^{\text{nom}} = 1$ p.u. Without loss of generality, we demonstrate our results with the trade-off

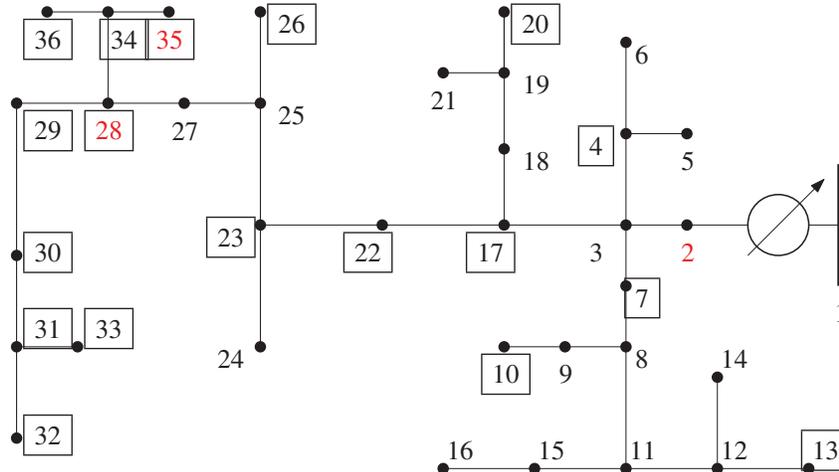


Figure 4.1: IEEE 37-node feeder. The boxes represent PV systems. The red nodes are the ones analyzed in the numerical example.

parameter γ set to either 0 or 1. For $\gamma = 1$, it is possible to trade off the customer-oriented objectives for flatness of the voltage profile. The regularization parameter ϕ is set to 10^{-4} . All the simulations are run with nonlinear AC power flow model calculated by MATPOWER [145].

4.3.2 Iterative Algorithm

We first test Algorithm 1 and show how the algorithm can address overvoltages in distribution systems [80]. To this end, we focus on a single timeslot at 12 pm.

4.3.2.1 Convergence

Let $\gamma = 0$ for simplicity. Recall from Theorem 5.3 that step sizes ε_1 and ε_2 both affect the convergence properties. For simplicity, set $\varepsilon_2 = 0.01$, and consider tuning ε_1 to achieve convergence. Similar results can be observed by fixing ε_1 and tuning ε_2 , or tuning both ε_1 and ε_2 . As shown in Figure 4.2, when ε_1 is increased from 0.01 to 0.3, we observe faster convergence. However, when we further increase ε_1 beyond 0.4, an oscillatory behavior is observed.

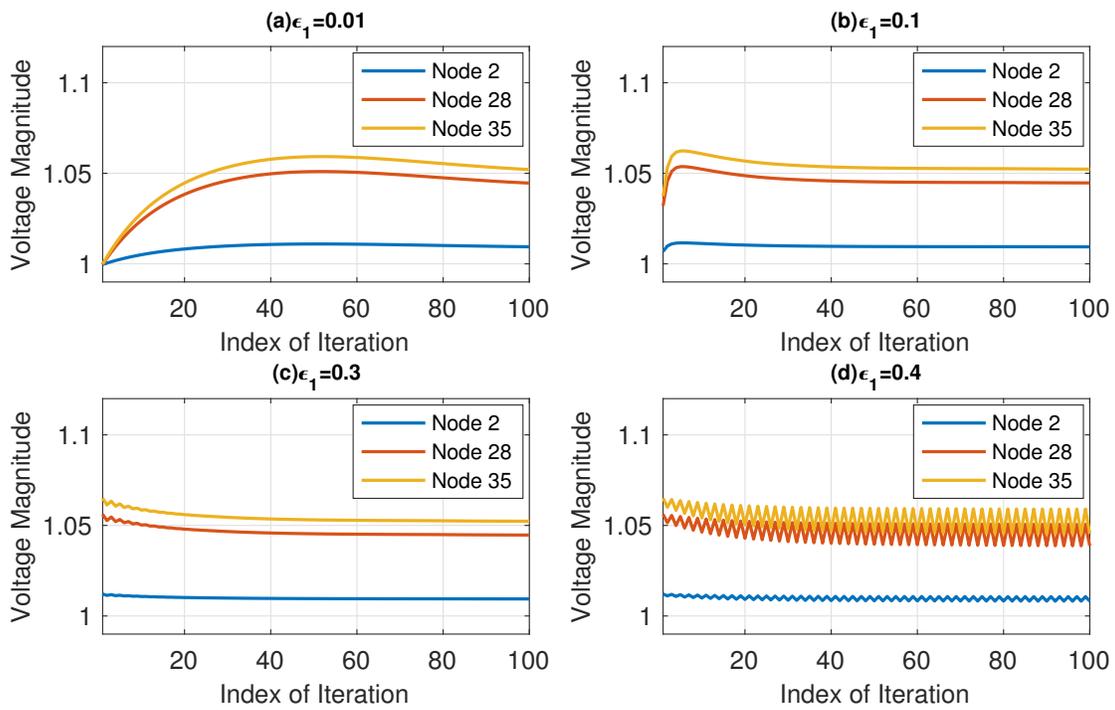


Figure 4.2: Convergence of the distributed algorithm with increasing step size ϵ_1 and fixed step size ϵ_2 .

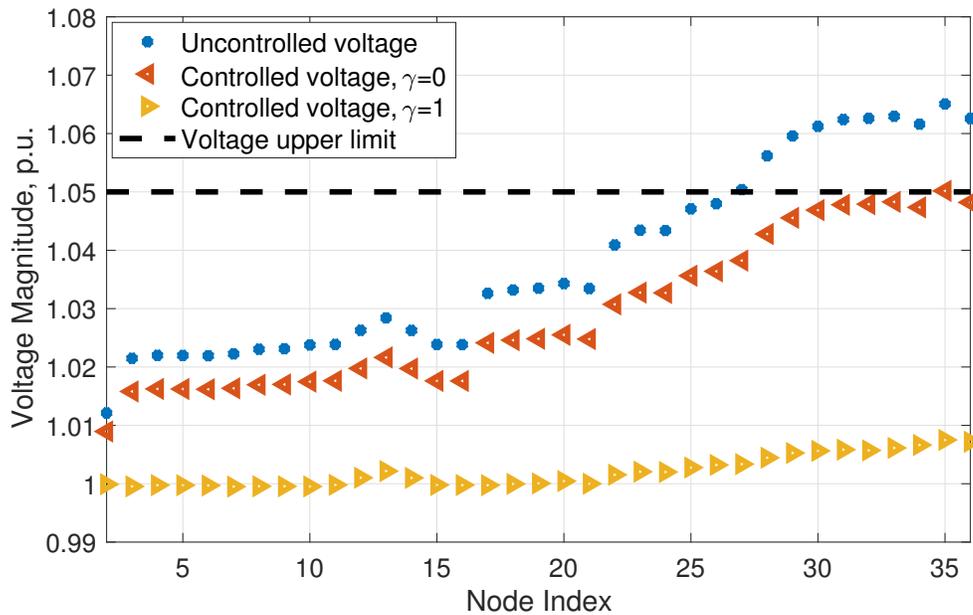


Figure 4.3: Controlled and uncontrolled voltages at all nodes at noon.

4.3.2.2 Voltage regulation

The results are plotted in Figure 4.3 corresponding to the case where $\varepsilon_1 = \varepsilon_2 = 0.01$. We show voltage profiles in three scenarios: (i) uncontrolled setting, where the PV systems operate at unity power factor and inject the maximum available power without any curtailment (blue dots), (ii) controlled voltages with $\gamma = 0$ (red dots), and (iii) controlled voltages with $\gamma = 1$ (yellow dots). It is clear that in the uncontrolled case (i) the voltage values exceed the limit of 1.05 p.u. (black dashed line) due to large reverse power flows, while the controlled scenarios (ii) and (iii) show voltage within limits. Furthermore, voltage values achieved by (iii) are closer to the nominal value than those by (ii), because (iii) also penalizes voltage deviation from 1 p.u.

4.4 Conclusion

This chapter considers a social welfare maximization problem modeling network operator and DER-owners operational objectives as well as voltage constraints. The formulated problem is non-convex; however, we propose a convex relaxation and we provide conditions under which the optimal solutions of the relaxed problem coincide with the optimal points of the non-convex social-welfare problem. We then design distributed algorithms to identify the solutions of the social welfare maximization problem without exposure of customers' privacy. Stability of the proposed schemes is analytically established and numerically corroborated.

It is worth mentioning that, the results of the exact convex relaxation and the distributed algorithm design also serve as a novel way to solve a class of Stackelberg game problem with quasilinear utility functions that has been considered difficult to solve in an efficient way.

Chapter 5

Distributed Voltage Regulation with Discrete Variables

Chapter 4 presents a market-based framework where network operator and customers pursue the given operational and economic objectives while concurrently ensuring that the voltages are within the prescribed limits. However, Chapter 4 only considers continuous decision variables (i.e., continuous DER commands), whereas in practice many appliances operate with discrete decision variables, e.g., capacity banks, thermostatically controlled loads (TCLs), and battery charging of electric vehicles (EVs), along with device dynamics as constraints.

This chapter extends the market-based framework developed in previous chapter to including discrete decision variables and device dynamics. Specifically, we first relax the discrete feasible sets to their convex hull. We then propose a distributed stochastic dual algorithm to solve the relaxed problem while recovering feasible power set points for discrete devices at every iteration, where two timescales are considered for devices of different updating frequencies. Eventually, we prove that the proposed algorithm converges to a random variable whose mean value coincides with the optimal solution of the relaxed problem. We also characterize the variance of the resultant voltage due to the stochastic process of discrete power rate recovering, and design a robust implementation for the voltage bounds accordingly. While diminishing stepsize for dual updates are utilized in this chapter, it is not practical in settings of real-time implementation and asynchronous updates. Online extension with constant stepsize will be presented next chapter. Related work of this chapter has been published in [140].

5.1 System Modeling

We use the same nonlinear AC power-flow model (4.1) and its linearization (4.2) as in last chapter. More generally, we consider a setting where each node could integrate multiple controllable devices; both continuous and discrete power outputs, as well as device dynamics, are considered for controllable devices.

5.1.1 Node and Device Model

At node $i \in \mathcal{N}$, assume that non-controllable devices and controllable devices contribute to the aggregate net power injections $p_i \in \mathcal{R}$ and $q_i \in \mathcal{R}$. Denote by $p_{i,0}$ and $q_{i,0}$ the overall active and reactive powers injected by all the non-controllable devices at node i . On the other hand, a number of controllable devices are collected in a set \mathcal{D}_i . Denote by $p_{i,d} \in \mathcal{R}$ and $q_{i,d} \in \mathcal{R}$ the active and reactive power injections of a controllable device $d \in \mathcal{D}_i$. It follows that the powers at node i can be expressed as

$$p_i = p_{i,0} + \sum_{d \in \mathcal{D}_i} p_{i,d}, \quad q_i = q_{i,0} + \sum_{d \in \mathcal{D}_i} q_{i,d}. \quad (5.1)$$

Power injections/consumption of device d is constrained as $z_{i,d} := (p_{i,d}, q_{i,d}) \in \mathcal{Z}_{i,d} \subset \mathbb{R}^2$. We consider the following two types of controllable devices.

Devices with convex sets: Devices with power injections/consumptions that can be chosen from a *convex and compact* set. These devices are assumed to be fast responding, in the sense that they regulate the power output to given commands within seconds. We collect these devices in the set $\mathcal{D}_{F_i} \subseteq \mathcal{D}_i$. For example, the feasible region of a PV inverter $d \in \mathcal{D}_{F_i}$ has the following form:

$$\mathcal{Z}_{i,d} = \{z_{i,d} \mid 0 \leq p_{i,d} \leq p_{i,d}^{\text{av}}, p_{i,d}^2 + q_{i,d}^2 \leq \eta_{i,d}^2\}, \quad (5.2)$$

where $p_{i,d}^{\text{av}}$ denotes the available active power from a PV system (based on prevailing ambient conditions), and $\eta_{i,d}$ is the rated apparent capacity.

Devices with discrete sets: Devices that admit a *discrete set* of possible setpoints. Control actions of these devices are usually implemented at a slower timescale. Collect these devices in the set $\mathcal{D}_{S_i} \subseteq \mathcal{D}_i$, and denote

by $\mathbb{P}_{i,d}$ the set of discrete power setpoints of a device $d \in \mathcal{D}_{S_i}$; i.e.,

$$z_{i,d} \in \mathbb{P}_{i,d}. \quad (5.3)$$

The devices may also feature states with given dynamics. For a device d , let $x_{i,d}$ denote the states of the device. We postulate an affine relationship between $x_{i,d}$ and $z_{i,d}$; i.e.,

$$x_{i,d} = f_{i,d}(z_{i,d}). \quad (5.4)$$

On the other hand, constraints on the state are modeled via a function $\psi_{i,d}$:

$$\psi_{i,d}(x_{i,d}) \leq 0. \quad (5.5)$$

By substituting (5.4) into (5.5), one obtains the following constraint:

$$\psi_{i,d}(f_{i,d}(z_{i,d})) \leq 0. \quad (5.6)$$

For a device $d \in \mathcal{D}_{S_i}$, its feasible power set is then written as

$$\mathcal{Z}_{i,d} = \{z_{i,d} | (5.3), (5.6)\}. \quad (5.7)$$

The set $\mathcal{Z}_{i,d}$ is then convex if $d \in \mathcal{D}_{F_i}$; on the other hand, it is *discrete* and *non-convex* if $d \in \mathcal{D}_{S_i}$. In the following, we provide two examples.

a) For a heating, ventilation, and air conditioning (HVAC) system $d \in \mathcal{D}_{S_i}$, let $x_{i,d}$ be the room temperature (i.e., the state) to be controlled. The dynamics (5.4) and constraints (5.5) are, in this case, in the form:

$$x_{i,d} = T_{i,d}^{\text{in}} + \theta_1(T_{i,d}^{\text{out}} - T_{i,d}^{\text{in}}) + \theta_2 p_{i,d}, \quad (5.8)$$

$$\underline{T}_{i,d}^{\text{in}} \leq x_{i,d} \leq \overline{T}_{i,d}^{\text{in}}, \quad (5.9)$$

where $T_{i,d}^{\text{in}}$ and $T_{i,d}^{\text{out}}$ are current room temperature and ambient temperature, θ_1, θ_2 are parameters specifying thermal characteristics of the HVAC and its operating environment, and $\underline{T}_{i,d}^{\text{in}}$ and $\overline{T}_{i,d}^{\text{in}}$ are customized temperature bounds; see, e.g., [75]. We assume that the HVAC system can be turned on/off; that is,

$$p_{i,d} \in \mathbb{P}_{i,d} = \{p_{i,d}^{\text{on}}, 0\}, \quad (5.10)$$

with working power rate $p_{i,d}^{\text{on}} < 0$. Then, the feasible set is $\mathcal{Z}_{i,d} = \{z_{i,d} | (5.8)–(5.10)\}$.

b) For an electric vehicle (EV) at charging station $d \in \mathcal{D}_{S_i}$, $x_{i,d}$ represents its state of charge (SOC) ; then, (5.4) represents the evolution of SOC for $m = 1, \dots, w$; i.e.,

$$x_{i,d} = E_{i,d} + \xi p_{i,d}, \quad (5.11)$$

where $E_{i,d}$ is the current SOC and ξ is the charging efficiency. The constraint (5.5) in this case takes the following form:

$$\underline{E}_{i,d} \leq x_{i,d} \leq \bar{E}_{i,d}, \quad (5.12)$$

with an acceptable SOC range of $[\underline{E}_{i,d}, \bar{E}_{i,d}]$. Further, a discrete set of feasible charging rates can be written as:

$$p_{i,d} \in \mathbb{P}_{i,d} = \{p_{i,d}^{\text{cha},1}, p_{i,d}^{\text{cha},2}, \dots, p_{i,d}^{\text{cha},N}, 0\}, \quad (5.13)$$

with feasible charging power rates $p_{i,d}^{\text{cha},1}, p_{i,d}^{\text{cha},2}, \dots, p_{i,d}^{\text{cha},N} < 0$. The EV battery's feasible set is then $\mathcal{Z}_{i,d} = \{z_{i,d} | (5.11)–(5.13)\}$.

5.1.2 Problem Formulation

We aim to design an algorithmic strategy where network operator and the DER-owners pursue their own operational goals and economic objectives, while concurrently achieving global coordination to enforce engineering constraints. For simplicity of exposition, consider the following notation

$$\mathcal{Z}_i := \prod_{d \in \mathcal{D}_i} \mathcal{Z}_{i,d}, \quad \mathcal{Z} := \prod_{i \in \mathcal{N}} \mathcal{Z}_i,$$

$$z_i = \{z_{i,d}\}_{d \in \mathcal{D}_i} \in \mathcal{Z}_i, \quad z = \{z_i\}_{i \in \mathcal{N}} \in \mathcal{Z}.$$

5.1.2.1 Customer's problem

The discrete nature of the control actions associated with devices in the sets \mathcal{D}_{S_i} would render pertinent optimization problems nonconvex. Consider then utilizing the convex hull, defined as follows:

$$\text{conv}(\mathcal{Z}_{i,d}) := \{z_{i,d} | (5.6), z_{i,d} \in \text{conv}(\mathbb{P}_{i,d})\}. \quad (5.14)$$

The relaxed set $\text{conv}(\mathcal{Z}_{i,d})$ is *convex and compact*. Accordingly, consider the following convex sets:

$$\text{conv}(\mathcal{Z}_i) := \left(\prod_{d \in \mathcal{D}_{F_i}} \mathcal{Z}_{i,d} \right) \times \left(\prod_{d \in \mathcal{D}_{S_i}} \text{conv}(\mathcal{Z}_{i,d}) \right),$$

and

$$\text{conv}(\mathcal{Z}) := \prod_{i \in \mathcal{N}} \text{conv}(\mathcal{Z}_i). \quad (5.15)$$

While a convex hull is utilized for the algorithmic design, a randomization strategy will be leveraged to recover feasible control actions from $\text{conv}(\mathcal{Z}_{i,d})$.

Consider a cost function $C_i(x_i, z_i)$ that captures well-defined performance objectives of all devices at node i . Since x_i is in fact a function of z_i , we will henceforth write the cost as $C_i(z_i)$.

Assumption 5.1. *Functions $C_i(z_i)$, $\forall i \in \mathcal{N}$ are strongly convex in $z_i \in \text{conv}(\mathcal{Z}_i)$.*

Let $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}$ be the pricing/reward signals sent by the network operator (e.g., distribution system operator or aggregator) to customer i for real and reactive power injections, respectively. Customers are assumed to be rational and self-interested to minimize their own cost, by solving the following problem ($\mathcal{P}'_{1,i}$) given signals (α_i, β_i) :

$$\min_{z_i} C_i(z_i) - \sum_{d \in \mathcal{D}_i} (\alpha_i p_{i,d} + \beta_i q_{i,d}), \quad (5.16a)$$

$$\text{s.t. } z_i \in \text{conv}(\mathcal{Z}_i), \quad (5.16b)$$

where $\alpha_i p_{i,d}$ and $\beta_i q_{i,d}$ represent payment to/reward from the network operator owing to all controllable device $d \in \mathcal{D}_i$.

Because (5.16a) is strongly convex in z_i , and \mathcal{Z}_i is convex and compact, a unique solution z_i^* exists. For future developments, consider a “best-response” strategy b_i of customer i as the following function of

(α_i, β_i) :

$$z_i^* = b_i(\alpha_i, \beta_i) := \arg \min_{z_i \in \text{cov}(\mathcal{Z}_i)} C_i(z_i) - \sum_{d \in \mathcal{D}_i} (\alpha_i p_{i,d} + \beta_i q_{i,d}).$$

5.1.2.2 Recover feasible power rates

Given $p_{i,d}^*$ solved from the relaxed feasible set for slow device d , we randomly select a feasible power rate $p_{i,d}$ based on the probability distribution such that $E[p_{i,d}] = p_{i,d}^*$, where $E[\cdot]$ denotes the expectation.

While there are multiple ways to determine the probability distribution of feasible setpoints based on $p_{i,d}^*$, we exemplify the procedure with two-point distribution for illustrative purpose. We select two feasible power rates and denote them as $\underline{p}_{i,d}$ and $\bar{p}_{i,d}$, such that $\underline{p}_{i,d} \leq p_{i,d}^* \leq \bar{p}_{i,d}$. Then the related probability of the corresponding two-point distribution is calculated as:

$$\begin{cases} \mathbb{P}(p_{i,d} = \underline{p}_{i,d}) &= (\bar{p}_{i,d} - p_{i,d}^*) / (\bar{p}_{i,d} - \underline{p}_{i,d}), \\ \mathbb{P}(p_{i,d} = \bar{p}_{i,d}) &= (p_{i,d}^* - \underline{p}_{i,d}) / (\bar{p}_{i,d} - \underline{p}_{i,d}), \end{cases} \quad (5.17)$$

according to which $p_{i,d}$ is randomly chosen.

5.1.2.3 Social-welfare problem

Similar to (\mathcal{P}_2) in last Chapter, consider now the following optimization problem (\mathcal{P}'_2) capturing both social welfare and voltage constraints of a distribution feeder:

$$\min_{z, \hat{v}, \alpha, \beta} \sum_{i \in \mathcal{N}} C_i(z_i), \quad (5.18a)$$

$$\text{s.t. } p_i = p_i^0 + \sum_{d \in \mathcal{D}_i} p_{i,d}, \quad q_i = q_i^0 + \sum_{d \in \mathcal{D}_i} q_{i,d}, \quad (5.18b)$$

$$\hat{v} = Ap + Bq + c, \quad (5.18c)$$

$$\underline{v} \leq \hat{v} \leq \bar{v}, \quad (5.18d)$$

$$z_i = b_i(\alpha_i, \beta_i), \quad i \in \mathcal{N}. \quad (5.18e)$$

Notice that the problem formulation naturally extends to the case where multiple customers are located at node i , but we outline the problem in this way to limit the complexity of the notation.

Since (\mathcal{P}'_2) is structured based on the relaxed sets of discrete devices, convex reformulation as well as the result of exact convex relaxation introduced in the Section 4.2.1 holds, as will be presented in the following section. Then, based on the stochastic dual algorithm, we propose a distributed algorithm that prevents any exposure of private information from the customers while solving a convex optimization problem where devices admits discrete power levels.

5.2 Distributed Algorithm Design

5.2.1 Convex Reformulation and Signal Design

Consider the following convex optimization problem (\mathcal{P}'_3) :

$$\min_{z, \hat{v}} \sum_{i \in \mathcal{N}} C_i(z_i), \quad (5.19a)$$

$$\text{s.t. } p_i = p_i^0 + \sum_{d \in \mathcal{D}_i} p_{i,d}, \quad q_i = q_i^0 + \sum_{d \in \mathcal{D}_i} q_{i,d}, \quad (5.19b)$$

$$\hat{v} = Ap + Bq + c, \quad (5.19c)$$

$$\underline{v} \leq \hat{v} \leq \bar{v}, \quad (5.19d)$$

$$z_i \in \mathcal{Z}_i, \forall i \in \mathcal{N}, \quad (5.19e)$$

with non-convex constraint (5.18e) replaced with (5.19e), and α, β to be designed later. We assume (\mathcal{P}'_3) is feasible:

Assumption 5.2. *Slater's condition holds for (\mathcal{P}'_3) .*

Given the strong convexity of the objective function (5.19a) in z , a unique optimal solution exists for problem (\mathcal{P}'_3) . Notice that a solution (z^*, \hat{v}^*) of (\mathcal{P}'_3) might not be feasible for (\mathcal{P}'_2) , because there might not exist a (α^*, β^*) such that $z_i^* = b_i(\alpha_i^*, \beta_i^*)$. We will, however, show next that such (α^*, β^*) exists; thus, the solution of (\mathcal{P}'_3) gives the solution of (\mathcal{P}'_2) .

Substitute (5.19c) into (5.19d), and denote by $\underline{\mu}$ and $\bar{\mu}$ the dual variables associated with the constraints (5.19d). Let \hat{v}^* be the optimal voltage magnitudes produced by (\mathcal{P}'_3) , and denote the optimal dual

variables associated with (5.19d) as $\underline{\mu}^*, \bar{\mu}^*$. Then, we design the signals as:

$$\alpha^* = A^\top [\underline{\mu}^* - \bar{\mu}^*], \quad (5.20a)$$

$$\beta^* = B^\top [\underline{\mu}^* - \bar{\mu}^*]. \quad (5.20b)$$

Note that α^* and β^* are composed of dual prices $\underline{\mu}^*$ and $\bar{\mu}^*$ with A, B characterizing the network structure.

We can prove that the following results similarly as Theorem 4.1–4.2 in Section 4.2.1.

Theorem 5.1. *Under Assumptions 5.1–5.2, the signals (α^*, β^*) defined by (5.20) are bounded.*

By examining the optimality conditions of (\mathcal{P}'_2) and (\mathcal{P}'_3) , we have the following result.

Theorem 5.2. *The solution of problem (\mathcal{P}'_3) along with the signals (α^*, β^*) defined in (5.20) is a global optimal solution of problem (\mathcal{P}'_2) ; i.e., problem (\mathcal{P}'_3) is an exact convex relaxation of problem (\mathcal{P}'_2) .*

Next, based on Theorem 5.2, we will develop an iterative algorithm that achieves the optimum of (\mathcal{P}'_3) (and hence that of (\mathcal{P}'_2)) without exposing any private information of the customers to the network operator.

5.2.2 Two Timescales and Iterative Algorithm

In this part, we design an iterative algorithm to solve (\mathcal{P}'_3) . As mentioned, we have two types of devices with two different update frequencies. Assume that slow devices update M times slower than fast devices with integer $M \geq 1$. We index by $k \in \mathbb{Z}_{++}$ the iterations when fast devices update. Then slow devices updates when $k = tM$ with index $t \in \mathbb{Z}_{++}$. We put the two timescales update in Algorithm 2 for easy reference later. Based on this strategy, we next propose a stochastic dual algorithm to solve (\mathcal{P}'_3) while recovering feasible power rates for slow devices.

Denote by $\mu := [\underline{\mu}^\top, \bar{\mu}^\top]^\top \in \mathbb{R}_+^{2N}$ the vector of stacked dual variables, and denote $g(z) = \begin{bmatrix} \underline{v} - Ap - Bq - c \\ Ap + Bq + c - \bar{v} \end{bmatrix}$.

For notational simplicity, we here use $z_s := \{z_{i,d}\}_{d \in \mathcal{D}_S}^{i \in \mathcal{N}}$ and $z_f := \{z_{i,d}\}_{d \in \mathcal{D}_F}^{i \in \mathcal{N}}$. We now can write the Lagrangian of (\mathcal{P}'_3) as:

$$\mathcal{L}(z, \mu) = \mathcal{L}(z_s, z_f, \mu) = \sum_{i \in \mathcal{N}} C_i(z_i) + \mu^\top g(z), \quad (5.21)$$

Algorithm 2 Two-timescale update

if iteration $k = tM$ **then**

Customer i solves $z_i^*(k+1) = \arg \min_{z_i \in \mathcal{Z}_i} C_i(z_i) - \sum_{d \in \mathcal{D}_i} (\alpha_i(k)p_{i,d} + \beta_i(k)q_{i,d})$, recovers $z_{i,d}(k+1)$ with

$E[z_{i,d}(k+1)] = z_{i,d}^*(k+1)$ for $d \in \mathcal{D}_{S_i}$, and sets $z_{i,d}(k+1) = z_{i,d}^*(k+1)$ for $d \in \mathcal{D}_{F_i}$.

else iteration $k = tM + m$, $m = 1, \dots, M-1$

Customer i keeps $z_{i,d}(k+1) = z_{i,d}(k)$ for $d \in \mathcal{D}_{S_i}$, and gets $z_{i,d}(k+1)$ for $d \in \mathcal{D}_{F_i}$ by solving:

$$\arg \min_{z_{F_i} \in \mathcal{Z}_{F_i}} C_i(z_{F_i} | z_{S_i}) - \sum_{d \in \mathcal{D}_{F_i}} (\alpha_i(k)p_{i,d} + \beta_i(k)q_{i,d}),$$

end if

which is obtained by keeping the constraints $z \in \mathcal{Z}$ and $\mu \in \mathbb{R}_+^{2N}$ implicit. Fix the value of z_S , and define the resultant form as a “reduced” Lagrangian $\mathcal{L}_F(z_F, \mu | z_S)$.

We will implement a dual algorithm with two timescales to solve the following minimax problems based on $\mathcal{L}(z, \mu)$ and $\mathcal{L}_F(z_F, \mu | z_S)$:

$$\max_{\mu \in \mathbb{R}_+^{2N}} \min_{z \in \text{conv}(\mathcal{Z})} \mathcal{L}(z, \mu), \quad (5.22a)$$

$$\max_{\mu \in \mathbb{R}_+^{2N}} \min_{z_F \in \mathcal{Z}_F} \mathcal{L}_F(z_F, \mu | z_S). \quad (5.22b)$$

To this end, we define two concave dual functions for \mathcal{L} and \mathcal{L}_F , respectively:

$$h(\mu) := \min_{z \in \mathcal{Z}} \mathcal{L}(z, \mu), \quad (5.23a)$$

$$h_F(\mu | z_S) := \min_{z_F \in \mathcal{Z}_F} \mathcal{L}_F(z_F, \mu | z_S), \quad (5.23b)$$

with corresponding dual problems:

$$\max_{\mu \in \mathbb{R}_+^{2N}} h(\mu), \quad (5.24a)$$

$$\max_{\mu \in \mathbb{R}_+^{2N}} h_F(\mu | z_S). \quad (5.24b)$$

Considering the dual algorithm to solve (5.22) while recovering implementable feasible power rates for discrete devices at each iteration, we have the following stochastic dual algorithm:

$$z(k+1) \quad \text{set by Algorithm 2}, \quad (5.25a)$$

$$\mu(k+1) = [\mu(k) + \varepsilon_k g(z(k+1))]_+, \quad (5.25b)$$

where $[\]_+$ is a projection operator onto the nonnegative orthant, and the stepsize ε_k to be selected. Also notice that $g(z(k+1))$ is subgradient of $h(\mu(k))$ when $k = tM$ and that of $h_F(\mu(k))$ when $k = tM + m$, $m = 1, \dots, M-1$.

Because \mathcal{Z} is compact and $g(z)$ is linear in z , there exists some constant $G > 0$ such that $E[|g(z(k))|] \leq G$ for all k . We next show the stability of dynamics (5.25) with diminishing stepsize in the following subsection.

5.2.3 Performance Analysis with Diminishing Stepsize

In this part, we choose stepsize ε_k to be square summable but not summable, i.e.:

$$\sum_{k=1}^{\infty} \varepsilon_k^2 < \infty, \quad \sum_{k=1}^{\infty} \varepsilon_k = \infty, \quad (5.26)$$

e.g., $\varepsilon_k = 1/t$ at iteration $k = tM + m$, $m = 0, \dots, M-1$. With such diminishing stepsize, we will prove the convergence of the sequence $\{\mu_k\}$ generated by (5.25) to a random vector. Moreover, we characterize the variance of voltage due to the randomness, and we propose a robust design.

5.2.3.1 Convergence

To show the convergence of dynamics (5.25), we utilize the next lemma [98].

Lemma 5.1. *Consider a sequence of random variables $\omega(1), \dots, \omega(k) \geq 0$, $E[\omega(1)] < \infty$ and*

$$E[\omega(k+1)|\omega(1), \dots, \omega(k)] \leq (1 + x_k)\omega(k) + y_k,$$

with

$$\sum_{k=1}^{\infty} x_k < \infty, \quad \sum_{k=1}^{\infty} y_k < \infty, \quad x_k \geq 0, \quad \text{and} \quad y_k \geq 0.$$

Then

$$\omega(k) \rightarrow \omega(\infty)$$

almost surely, where $\omega(\infty) \geq 0$ is some random variable.

We accordingly have the following convergence result.

Theorem 5.3. *If the stepsize ε_k is chosen as in (5.26), the sequence $\{\mu(k)\}$ generated by (5.25) converges to certain random vector $\mu(\infty)$ almost surely.*

Proof. The convergence of sequence $\{\mu(k)\}$ is equivalent to that of sequence $\{\|\mu(k) - \mu^*\|^2\}$, where μ^* is an optimal dual of (\mathcal{P}'_3) . Then, at iteration $k = tM$ with some $t > 0$:

$$\begin{aligned}
& E[\|\mu(tM + 1) - \mu^*\|^2 | \mu(1), \dots, \mu((t-1)M + 1)] \\
& \leq E[\|\mu(tM) + \varepsilon_{tM} g(z(tM)) - \mu^*\|^2 | \mu((t-1)M + 1)] \\
& \leq E[\|\mu(tM) - \mu^*\|^2 | \mu((t-1)M + 1)] + \varepsilon_{tM}^2 G^2 + 2\varepsilon_{tM} (\mu(tM) - \mu^*)^\top g(z^*(tM)) \\
& \leq \|\mu((t-1)M + 1) - \mu^*\|^2 + \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2 + \sum_{m=0}^{M-1} 2\varepsilon_{tM-m} (\mu(tM-m) - \mu^*)^\top g(z^*(tM-m)) \\
& \leq \|\mu((t-1)M + 1) - \mu^*\|^2 + \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2 + 2\varepsilon_{tM} (h(\mu(tM)) - h(\mu^*)) + \sum_{m=1}^{M-1} 2\varepsilon_{tM-m} (h_F(\mu(tM-m)) - h_F(\mu^*)) \\
& \leq \|\mu((t-1)M + 1) - \mu^*\|^2 + \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2,
\end{aligned}$$

where the first inequality comes from the non-expansiveness of projection operator, the third from repeating previous steps, the fourth from the definition of the subgradient, and the last from the definition of optimality of concave functions h and h_F .

Because $\sum_{t=1}^{\infty} \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2 < \infty$, by Lemma 5.1 the sequence $\{\|\mu(k) - \mu^*\|^2\}$ converges to some random variable $\{\|\mu(\infty) - \mu^*\|^2\}$ almost surely, and therefore the sequence $\{\mu(k)\}$ converges to some random vector $\mu(\infty)$ almost surely. \blacksquare

Denote $\tilde{\mu}(k) := E[\mu(k)]$. By Theorem 5.3, $\lim_{k \rightarrow \infty} \tilde{\mu}(k) = \tilde{\mu}(\infty)$, where $\tilde{\mu}(\infty)$ is the mean value of random variable $\mu(\infty)$ to which $\{\mu(k)\}$ converges. We next show that $\tilde{\mu}(\infty) = \mu^*$.

Theorem 5.4. *Select the stepsize ε_k as in (5.26). The sequence $\{\tilde{\mu}(k)\}$ generated by (5.25) converges to μ^* . Meanwhile, the running average of $h(\mu(k))$ approaches $h(\mu^*)$ as $k \rightarrow \infty$:*

$$\lim_{k \rightarrow \infty} h(\mu^*) - \sum_{\kappa=1}^k h(\mu(\kappa)) / k = 0. \tag{5.27}$$

Proof. Similar to the proof of Theorem 5.3, we have:

$$\begin{aligned}
& E[\|\mu(tM + 1) - \mu^*\|^2] \\
\leq & E[\|\mu((t-1)M + 1) - \mu^*\|^2] + \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2 + 2\varepsilon_{tM} (h(\mu(tM)) - h(\mu^*)) + \sum_{m=1}^{M-1} 2\varepsilon_{tM-m} (h_F(\mu(tM-m)) - h_F(\mu^*)) \\
\leq & E[\|\mu((t-1)M + 1) - \mu^*\|^2] + \sum_{k=(t-1)M+1}^{tM} \varepsilon_k^2 G^2 + 2\varepsilon_{tM} (h(\mu(tM)) - h(\mu^*)).
\end{aligned}$$

Apply the above steps recursively to obtain:

$$E[\|\mu(tM + 1) - \mu^*\|^2] \leq E[\|\mu(1) - \mu^*\|^2] + \sum_{k=1}^{tM} \varepsilon_k^2 G^2 + \sum_{\tau=1}^t 2\varepsilon_{\tau M} E[h(\mu(\tau M)) - h(\mu^*)].$$

Because $E[\|\mu(tM + 1) - \mu^*\|^2] \geq 0$, the following holds:

$$\sum_{\tau=1}^t 2\varepsilon_{\tau M} E[h(\mu^*) - h(\mu(\tau M))] \leq E[\|\mu(1) - \mu^*\|^2] + \sum_{k=1}^{tM} \varepsilon_k^2 G^2. \quad (5.28)$$

By Jensen's inequality:

$$E[h(\mu^*) - h(\mu(\tau M))] \geq h(\mu^*) - h(\tilde{\mu}(\tau M)). \quad (5.29)$$

Therefore, by considering $\sum_{k=1}^{\infty} \varepsilon_k^2 < \infty$ from (5.26), we have from (5.28) and (5.29) that:

$$\lim_{t \rightarrow \infty} \sum_{\tau=1}^t 2\varepsilon_{\tau M} (h(\mu^*) - h(\tilde{\mu}(\tau M))) < \infty. \quad (5.30)$$

We next show $h(\mu^*) = h(\tilde{\mu}(\infty))$ by contradiction. Recalling that $h(\mu^*) \geq h(\mu)$ for any feasible μ , assume there exists some $e > 0$ such that $h(\mu^*) - h(\tilde{\mu}(\infty)) \geq e$. Because $\sum_{\tau=1}^{\infty} \varepsilon_{\tau M} = \infty$, we must have $\lim_{\tau \rightarrow \infty} \sum_{\tau=1}^t 2\varepsilon_{\tau M} (h(\mu^*) - h(\tilde{\mu}(\tau M))) = \infty$, which contradicts (5.30). Hence, $h(\mu^*) = h(\tilde{\mu}(\infty))$.

Further, because μ is statistically stationary, its ensemble average and time average are identical. (5.27) follows. ■

Remark 5.1. With strongly convex cost functions C_i , $\{\mu(k)\}$ generated by (5.25) usually converges to a random vector even with constant stepsize, where all the properties we obtain in Theorem 5.3–5.4 hold. This will be shown with numerical examples in Section 5.3, and proved next setion.

5.2.3.2 Variance and Robust Design

The randomness in the power rate selection for discrete devices of \mathcal{D}_S leads to volatility of voltages. Let D_S be the number of all discrete devices. We next characterize the variance of the voltage (the result is tailored to the variable (5.17), but can be straightforwardly generalized).

Proposition 1. *Using the randomized selection strategy (5.17), the voltage variance $\text{Var}(v_i)$ at node $i \in \mathcal{N}$ is bounded as:*

$$\text{Var}(\hat{v}_i) \leq D_S / 4 \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot \max_{j,d} (\bar{p}_{j,d} - \underline{p}_{j,d})^2. \quad (5.31)$$

Proof. The variance of \hat{v}_i can be written as:

$$\begin{aligned} \text{Var}(\hat{v}_i) &= E[|\hat{v}_i - \hat{v}_i^*|^2] = E\left[\left|\sum_{j \in \mathcal{N}} A_{ij}(p_j - p_j^*)\right|^2\right] \\ &\leq \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot E\left[\sum_{j \in \mathcal{N}} (p_j - p_j^*)^2\right] \\ &\leq \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot \sum_{j \in \mathcal{N}} \sum_{d \in \mathcal{D}_{S_j}} E[(p_{j,d} - p_{j,d}^*)^2] \\ &= \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot \sum_{j \in \mathcal{N}} \sum_{d \in \mathcal{D}_{S_j}} (p_{j,d}^* - \underline{p}_{j,d})(\bar{p}_{j,d} - p_{j,d}^*) \\ &\leq \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot \sum_{j \in \mathcal{N}} \sum_{d \in \mathcal{D}_{S_j}} (\bar{p}_{j,d} - \underline{p}_{j,d})^2 / 4, \\ &\leq D_S / 4 \sum_{j \in \mathcal{N}} A_{ij}^2 \cdot \max_{j,d} (\bar{p}_{j,d} - \underline{p}_{j,d})^2, \end{aligned}$$

where we apply Cauchy-Schwarz inequality in the first inequality, Jensen's inequality in the second, and the second equality is based on the probability distribution (5.17). \blacksquare

The result of Proposition 1 motivates us to propose the following robust design. We choose tighter voltage bounds $[\underline{v}', \bar{v}']$ with $\underline{v} < \underline{v}' < v^{\text{nom}} < \bar{v}' < \bar{v}$, and replace the original bounds with the tighter bounds in the algorithm so that the resultant voltage falls within the original bounds with required probability. We design the tighter bounds based on Chebyshev's inequality [112] as shown next.

Proposition 2 (robust implementation). *Given $\delta > 0$, if the voltage upper and lower bounds are set as:*

$\bar{v}'_i \leq \bar{v}_i - \delta$ and $\underline{v}'_i \geq \underline{v}_i + \delta$, then:

$$\mathbb{P}[\hat{v}_i \geq \bar{v}'_i] \leq \text{Var}(\hat{v}_i) / 2\delta^2, \text{ and } \mathbb{P}[\hat{v}_i \leq \underline{v}'_i] \leq \text{Var}(\hat{v}_i) / 2\delta^2.$$

Proof. Let $\tilde{v}_i = E[\hat{v}_i]$. By Chebyshev's inequality [112], given $\delta > 0$, we have

$$\mathbb{P}[|\hat{v}_i - \tilde{v}_i| \geq \delta] \leq \frac{\text{Var}(\hat{v}_i)}{\delta^2}.$$

Consider the upper bound first. Design the robust bound as $\bar{v}'_i \leq \bar{v}_i - \delta$, so that we must have $\tilde{v}_i \leq \bar{v}'_i$ by Assumption 6.3 and Theorem 5.4. And we write the probability of voltage violation as follows:

$$\begin{aligned} \mathbb{P}[\hat{v}_i \geq \bar{v}_i] &= \mathbb{P}[\hat{v}_i - \tilde{v}_i \geq \bar{v}_i - \tilde{v}_i] \\ &\leq \mathbb{P}[\hat{v}_i - \tilde{v}_i \geq \bar{v}_i - \bar{v}'_i] \leq \mathbb{P}[\hat{v}_i - \tilde{v}_i \geq \delta] \\ &= \frac{1}{2} \cdot \mathbb{P}[|\hat{v}_i - \tilde{v}_i| \geq \delta] \leq \frac{\text{Var}(\hat{v}_i)}{2\delta^2}. \end{aligned}$$

Similar process applies to $\mathbb{P}[\hat{v}_i \leq \underline{v}_i]$. ■

Remark 5.2. *These bounds are admittedly conservative; however, reasonable values can be obtained in realistic settings. For example, scenario 2) in Section 5.3.1 with variance estimated by upper-bound (5.31) leads to robust bounds $\bar{v}'_i = 1.035$ p.u. and $\underline{v}'_i = 0.965$ p.u., with $\text{Var}(\hat{v}_i)/2\delta^2 \leq 5\%$ and $\bar{v}_i = 1.05$ p.u. and $\underline{v}_i = 0.95$ p.u. Tighter bounds can be obtained empirically.*

5.2.4 Distributed Stochastic Dual Algorithm

The decomposable structure of (5.25) naturally enables the following iterative *distributed* algorithm:

$$z(k+1) \quad \text{set by Algorithm 2,} \tag{5.32a}$$

$$\hat{v}(k+1) = Ap(k+1) + Bq(k+1) + c \tag{5.32b}$$

$$\underline{\mu}(k+1) = [\underline{\mu}(k) + \varepsilon_k(\underline{v} - \hat{v}(k+1))]_+, \tag{5.32c}$$

$$\bar{\mu}(k+1) = [\bar{\mu}(k) + \varepsilon_k(\hat{v}(k+1) - \bar{v})]_+, \tag{5.32d}$$

$$\alpha(k+1) = A^\top [\underline{\mu}(k+1) - \bar{\mu}(k+1)], \tag{5.32e}$$

$$\beta(k+1) = B^\top [\underline{\mu}(k+1) - \bar{\mu}(k+1)], \tag{5.32f}$$

where the power set points of the devices are computed and implemented locally through (5.32a), and (5.32b)–(5.32f) are performed at the network operator. Notice that each customer i *does not* share its cost function

C_i or its feasible set \mathcal{Z}_i with the network operator; rather, the customer transmits to the network operator only the resultant power injections $z_i(k)$. Indeed, (5.25) and (5.32) are equivalent and the results of Theorem 5.3–5.4 apply to (5.32).

5.3 Numerical Examples

Based on the same simulation setup as in Section 4.3.1, we further add the following devices with discrete power rates and device dynamics. We install 15 identical A/Cs at each of the following 25 nodes: 2, 5, 6, 7, 9, 10, 11, 13, 14, 16, 18, 19, 20, 21, 22, 24, 26, 27, 28, 29, 30, 32, 33, 35, and 36, totaling 375 A/Cs that comprise the set \mathcal{D}_S . We set a uniform cost function for all A/Cs as $C_{i,d}(T_{i,d}^{\text{in}+}) = 20(T_{i,d}^{\text{in}+} - T_{i,d}^{\text{nom}})^2$, where $T_{i,d}^{\text{nom}}$ is a preferred room temperature set at 75°F, and the room temperature 15 minutes later is modeled as $T_{i,d}^{\text{in}+} = T_{i,d}^{\text{in}} + 0.1(T_{i,d}^{\text{out}} - T_{i,d}^{\text{in}}) - 0.001p_{i,d}$. Also, $T_{i,d}^{\text{in}+}$ should be within [70°F, 80°F]. For each A/C, there are two possible power rates: 0 and -4 kW. The cost function of customer i sums the cost functions of all its devices $C_i(z_i) = \sum_{d \in \mathcal{D}_i} C_{i,d}(z_{i,d})$.

The voltage limits are $\bar{v}_i = 1.05$ p.u. and $\underline{v}_i = 0.95$ p.u., and robust voltage limits are set to $\bar{v}'_i = 1.04$ p.u. and $\underline{v}'_i = 0.96$ p.u. for $\forall i \in \mathcal{N}$. Stepsize $\varepsilon = 0.1$ is constant, which enables us to achieve all the results proved under diminishing stepsize.

5.3.1 Convergence and Variance

We update PVs every iteration, and A/Cs every 60 iterations, with the following scenarios:

- 1) All 15 A/Cs at each node are combined to be controlled together with two power rates of 0 kW and -60 kW.
- 2) 15 A/Cs at each node are controlled independently.
- 3) All 15 A/Cs at each node are combined to be controlled together with 16 power rates: 0 kW, -4 kW, ..., -60 kW.

Without loss of generality, we use scenario 2) to show convergence. The results are plotted in Fig. 5.1. Though the resultant voltage (red line) is changing randomly, it fluctuates around the solution of the

relaxed problem (blue line). Also, the running average of the fluctuating voltage (green line) approaches the solution of the relaxed problem as iteration number increases, verifying Theorem 5.3–5.4 even with constant stepsize. Moreover, in between updates of A/Cs, their random consequence is absorbed by the PV’s faster updates.

Next, we compare the variance of the resultant voltages among scenarios 1)–3). Based on design, these three scenarios have the same optimality on average. By Proposition 1, the voltage variance stems from the number of random variables and the control granularity. We compare scenario 1) and 3) to illustrate that finer granularity generates less variance, and scenario 2) and 3) to show that less random variables results in less variance. The results are presented in Fig. 5.2, where three scenarios are marked with different colors.

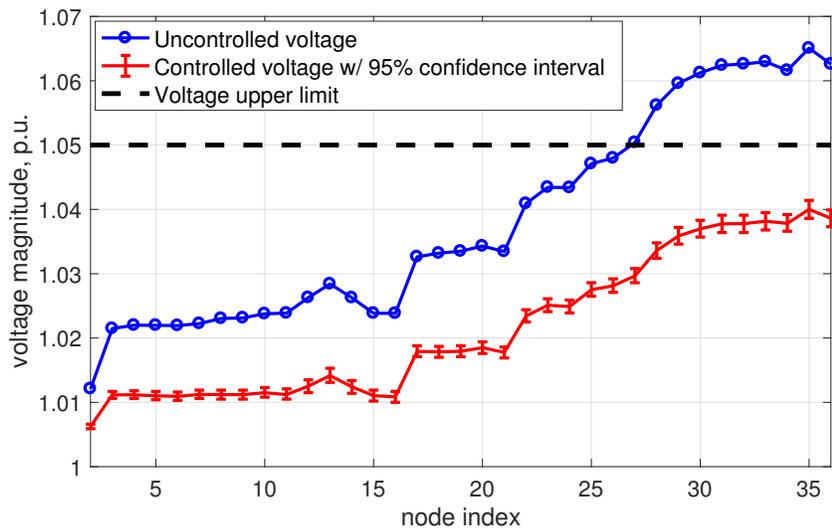


Figure 5.3: Uncontrolled voltage and controlled voltage under scenario 2) with 95% confidence interval.

5.3.2 Voltage Regulation

In this part, we use scenario 2) to compare the resultant voltages at all nodes to those without any voltage regulation. We record 25,000 random processes since convergence, and we plot the mean values of the voltages together with their 95% confidence intervals in Fig. 5.3. Because of the robust bounds as well as the small variance of the resultant voltages, the controlled voltages are all less than the original voltage

upper-limit of 1.05 p.u.

5.4 Conclusion

Based on the market-based framework developed in Chapter 4, we have proposed an iterative distributed stochastic dual algorithm that allows the distribution network operator and the customers to coordinate with private information preserved to optimize the social welfare while concurrently recovering feasible power rates for discrete devices and ensuring that the voltage magnitudes are within the prescribed limits. We prove its convergence and analyze its performance. Numerical examples are provided to support the theoretical results.

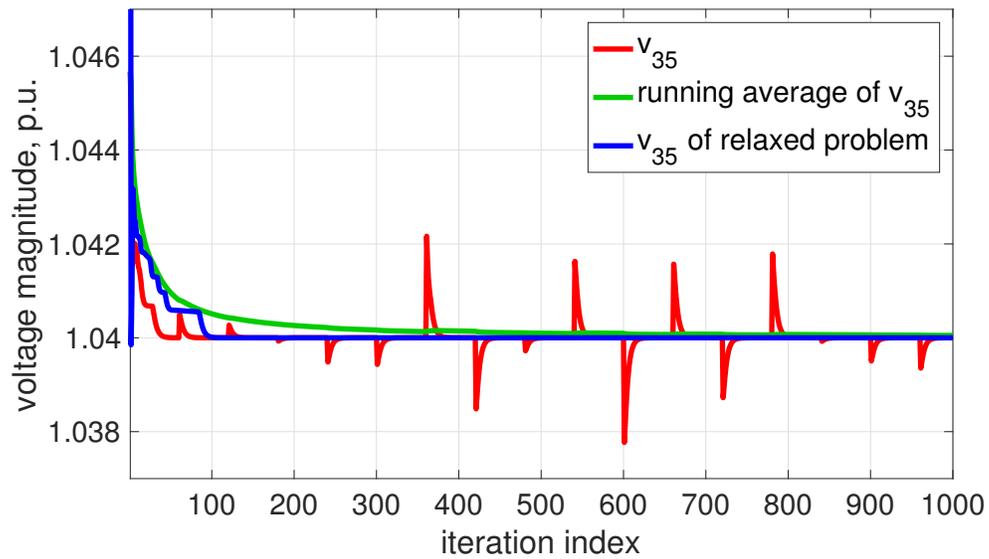


Figure 5.1: The running average of the voltage converges to the optimal voltage of the relaxed problem.

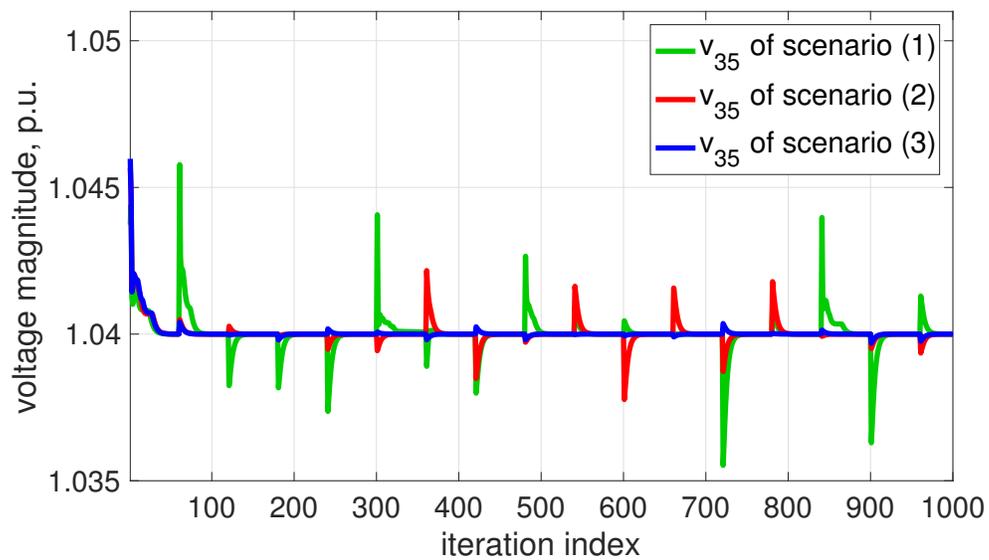


Figure 5.2: Different voltage variance of three different scenarios.

Chapter 6

Online Distributed Voltage Regulation

Based on results of joint distributed control and incentive/market mechanisms to coordinate DERs in distribution grids from Chapter 4–5, this chapter formulates a multi-period social welfare maximization problem, and an online setting with time-varying operating conditions, asynchronous updates by devices, and feedback being leveraged to account for nonlinear power flows as well as reduce communication overhead. The resulting algorithm provides a general online stochastic optimization algorithm for coordinating networked DERs with mixed continuous/discrete power setpoints and dynamics to meet operational and economic objectives and constraints.

In literature diminishing stepsizes are usually assumed to be necessary for convergence of the dual method. Nevertheless, we leverage recent insights in dual method to characterize the convergence of the proposed stochastic dual algorithm with constant stepsizes. To the best of our knowledge, this is a first convergence characterization of its kind for the dual method applied in power systems. Performance of the online algorithm are then analytically characterized and numerically corroborated. Related work of this chapter has been presented in [136].

6.1 System Model and Problem Formulation

6.1.1 Network Model

Since this chapter studies a time-varying setting, we recast the system modeling accordingly. Consider a distribution network with $N + 1$ nodes collected in the set $\mathcal{N} \cup \{0\}$, $\mathcal{N} := \{1, \dots, N\}$, and distribution lines represented by the set \mathcal{E} . Let $p_i^t \in \mathbb{R}$ and $q_i^t \in \mathbb{R}$ denote the aggregate net active and reactive power injections, respectively, at node $i \in \mathcal{N}$ at time t . Further, let y^t be a vector collecting certain electrical quantities of interest; for example, voltage magnitudes at some selected nodes, current on some lines, or power flows at the substation. The electrical quantities collected in y^t are related to p_i^t and q_i^t via a nonlinear relationship that follows from Ohm's and Kirchhoff's Laws. In this chapter, we utilize the following approximate linear relationship:

$$y^t \approx \hat{y}^t = Ap^t + Bq^t + c, \quad (6.1)$$

where A , B and c are linearization parameters that can be computed as shown in, e.g., [23, 36] and pertinent references therein. It is worth pointing out that the linearized model (6.1) is utilized to facilitate the design of computationally-light algorithms that admit a real-time implementation. In Section 6.3, we will show how to leverage appropriate measurements from the distribution grid and DERs to cope with the inaccuracies in the representation of the AC power flows and we will establish appropriate convergence claims. As shown in [23, 36], the linear model (6.1) can be built in the way to account for an unbalanced system operation and for both wye and delta connections.

6.1.2 Node and Device Models

At time t , denote by $p_{i,0}^t$ and $q_{i,0}^t$ the overall active and reactive powers injected by all the *non-controllable* devices at node i . On the other hand, a number of controllable devices are collected in a set \mathcal{D}_i . Denote by $p_{i,d}^t \in \mathbb{R}$ and $q_{i,d}^t \in \mathbb{R}$ the active and reactive power injections of a controllable device $d \in \mathcal{D}_i$ at time t . It follows that the powers at node i can be expressed as:

$$p_i^t = p_{i,0}^t + \sum_{d \in \mathcal{D}_i} p_{i,d}^t, \quad q_i^t = q_{i,0}^t + \sum_{d \in \mathcal{D}_i} q_{i,d}^t. \quad (6.2)$$

Power injections/consumption of device d at time t is constrained as $z_{i,d}^t := (p_{i,d}^t, q_{i,d}^t) \in \mathcal{Z}_{i,d}^t \subset \mathbb{R}^2$. Also, denote by $W_t^w = \{t, t+1, \dots, t+w\}$ the time window from time t up to time $t+w$, and $\mathbf{z}_{i,d}^t := [(z_{i,d}^t)^\top, \dots, (z_{i,d}^{t+w})^\top]^\top$ the power trajectory and $\mathcal{Z}_{i,d}^t = \times_{\tau=t}^{t+w} \mathcal{Z}_{i,d}^\tau$ the feasible set of device d within this time window. We consider the following two types of controllable devices within the time window W_t^w .

Devices with convex sets: Devices with power injections/consumptions that can be chosen from a *convex and compact* set. These devices are assumed to be fast responding, in the sense that they regulate the power output to given commands within seconds. We collect these devices in the set $\mathcal{D}_{F_i} \subseteq \mathcal{D}_i$. For example, the feasible region of a PV inverter $d \in \mathcal{D}_{F_i}$ has the following form for time t :

$$\mathcal{Z}_{i,d}^t = \{z_{i,d}^t | 0 \leq p_{i,d}^t \leq p_{i,d}^{\text{av},t}, p_{i,d}^{t2} + q_{i,d}^{t2} \leq \eta_{i,d}^2\}, \quad (6.3)$$

where $p_{i,d}^{\text{av},t}$ denotes the available active power from a PV system at time t (based on prevailing ambient conditions), and $\eta_{i,d}$ is the rated apparent capacity.

We use bold set symbol $\mathcal{Z}_{i,d}^t$ to represent the feasible set of device d within the window of W_t^w . For a device without temporarily correlated constraint like a PV inverter, $\mathcal{Z}_{i,d}^t$ can be the Cartesian product of the time-varying feasible sets within the time window W_t^w , e.g., $\mathcal{Z}_{i,d}^t = \times_{\tau=t}^w \mathcal{Z}_{i,d}^\tau$. This is not usually the case for devices with dynamics, as to be introduced next.

Devices with discrete sets: Devices that admit a *discrete set* of possible setpoints. Control actions of these devices are usually implemented at a slower timescale. Collect these devices in the set $\mathcal{D}_{S_i} \subseteq \mathcal{D}_i$, and denote by $\mathbb{P}_{i,d}$ the set of power setpoints of a device $d \in \mathcal{D}_{S_i}$; i.e.,

$$z_{i,d}^t \in \mathbb{P}_{i,d}. \quad (6.4)$$

The devices may also feature states with given dynamics. For a device d , let $\mathbf{x}_{i,d}^t := [x_{i,d}^t, \dots, x_{i,d}^{t+w}]^\top$ collect the states of the device from the current time t up to time $t+w$ (i.e., within the time window W_t^w). We postulate an affine relationship between $\mathbf{x}_{i,d}^t$ and $\mathbf{z}_{i,d}^t$; i.e.,

$$\mathbf{x}_{i,d}^t = \mathbf{f}_{i,d}^t(\mathbf{z}_{i,d}^t) \quad (6.5)$$

where $\mathbf{f}_{i,d}^t(\mathbf{z}_{i,d}^t) := [f_{i,d}^t(\mathbf{z}_{i,d}^t), \dots, f_{i,d}^{t+w}(\mathbf{z}_{i,d}^t)]^\top$. On the other hand, constraints on the state are modeled via a

general *convex* vector-valued function $\psi_{i,d}^t$:

$$\psi_{i,d}^t(\mathbf{x}_{i,d}^t) \leq 0. \quad (6.6)$$

By substituting (6.5) into (6.6), one obtains the following constraint:

$$\psi_{i,d}^t(f_{i,d}^t(\mathbf{z}_{i,d}^t)) \leq 0. \quad (6.7)$$

The set $\mathcal{Z}_{i,d}^t$ is then convex if $d \in \mathcal{D}_{F_i}$; on the other hand, it is *discrete* and *non-convex* if $d \in \mathcal{D}_{S_i}$.

In the following, we provide two examples.

a) For a heating, ventilation, and air conditioning (HVAC) system $d \in \mathcal{D}_{S_i}$, let $x_{i,d}^t$ be the room temperature (i.e., the state) to be controlled at time t . The dynamics (6.5) are, in this case, in the form:

$$x_{i,d}^{t+m} = T_{0,i,d}^{t+m} + \sum_{\tau=0}^{m-1} (1 - \theta_1)^{m-1-\tau} \theta_2 p_{i,d}^{t+\tau}. \quad (6.8)$$

Here, $T_{0,i,d}^{t+m}$ is a constant characterized as

$$T_{0,i,d}^{t+m} = (1 - \theta_1)^m x_{i,d}^t + \sum_{\tau=0}^{m-1} (1 - \theta_1)^{m-1-\tau} \theta_1 T_{out,i,d}^{t+\tau}, \quad (6.9)$$

where $T_{out,i,d}^{t+\tau}$ is the ambient temperature, and θ_1, θ_2 are parameters specifying thermal characteristics of the HVAC and its operating environment; see, e.g., [75]. The constraint (6.6) can be

$$\underline{\mathbf{T}}_{i,d}^t \leq \mathbf{x}_{i,d}^t \leq \overline{\mathbf{T}}_{i,d}^t, \quad (6.10)$$

which is to confine the room temperatures within a given comfort range $[\underline{\mathbf{T}}_{i,d}^t, \overline{\mathbf{T}}_{i,d}^t]$. We assume that the HVAC system can be turned on/off; that is,

$$p_{i,d}^t \in \{p_{i,d}^{\text{on}}, 0\}, \quad \forall t, \quad (6.11)$$

with working power rate $p_{i,d}^{\text{on}} < 0$. Then, the feasible set within W_i^w is $\mathcal{Z}_{i,d}^t = \{\mathbf{z}_{i,d}^t | (6.8) - (6.11)\}$.

b) For an electric vehicle (EV) at charging station $d \in \mathcal{D}_{S_i}$, $x_{i,d}^t$ represents its state of charge (SOC) at time t ; then, (6.5) represents the evolution of SOC for $m = 1, \dots, w$; i.e.,

$$x_{i,d}^{t+m} = x_{i,d}^t + \sum_{\tau=0}^{m-1} \xi p_{i,d}^{t+\tau}, \quad (6.12)$$

where ξ is the charging efficiency. The constraint (6.6) in this case takes the following form:

$$\underline{E}_{i,d}^t \leq x_{i,d}^t \leq \bar{E}_{i,d}^t, \quad (6.13)$$

with an acceptable SOC range of $[\underline{E}_{i,d}^t, \bar{E}_{i,d}^t]$. Further, a discrete set of feasible charging rates can be written as:

$$p_{i,d}^t \in \{0, p_{i,d}^{\text{cha},1}, p_{i,d}^{\text{cha},2}, \dots, p_{i,d}^{\text{cha},N}\}, \quad \forall t. \quad (6.14)$$

The EV battery's feasible set from within W_i^w is then $\mathcal{Z}_{i,d}^t = \{z_{i,d}^t | (6.12)-(6.14)\}$.

6.1.3 Problem Formulation

We aim to design an algorithmic strategy where network operator and the DER-owners pursue their own operational goals and economic objectives, while concurrently achieving global coordination to enforce engineering constraints. For simplicity of exposition, consider the following notation:

$$\mathcal{Z}_i^t := \prod_{d \in \mathcal{D}_i} \mathcal{Z}_{i,d}^t, \quad \mathcal{Z}^t := \prod_{i \in \mathcal{N}} \mathcal{Z}_i^t,$$

$$z_i^t = \{z_{i,d}^t\}_{d \in \mathcal{D}_i} \in \mathcal{Z}_i^t, \quad z^t = \{z_i^t\}_{i \in \mathcal{N}} \in \mathcal{Z}^t.$$

We now describe pertinent optimization problems that model operational goals and constraints of network operator and the DER-owners. For simplicity of exposition, again we outline the problem formulation under the presumption that one customer/DER-owner is located at a node of the electrical network; however, the proposed methodology is applicable to the case where multiple customers are located at a node (this is the case, for example, where multiple houses are connected to the same distribution transformer).

6.1.3.1 Customer-level problem

Let $\alpha_i^t = [\alpha_i^t, \dots, \alpha_i^{t+w}] \in \mathbb{R}^{w+1}$ and $\beta_i^t = [\beta_i^t, \dots, \beta_i^{t+w}] \in \mathbb{R}^{w+1}$ be vectors collecting ‘‘incentive signals’’ that are sent by the network operator to customer i within the time window W_i^w . Let $x_i^t := \{x_{i,d}^t\}_{d \in \mathcal{D}_i}$ be a vector collecting states of all devices of node i at time t , and consider a cost function $C_i^t(z_i^t, x_i^t)$ that

captures well-defined performance objectives of all devices at node i within the time window W_i^w . Since \mathbf{x}_i^t is in fact a function of \mathbf{z}_i^t , we will henceforth write the cost as $C_i^t(\mathbf{z}_i^t)$. Notice further that the cost function captures objectives over the time window W_i^w , and it can be expanded as, but not confined to, an additive form, e.g., $C_i^t(\mathbf{z}_i^t) = \sum_{\tau=t}^{t+w} \sum_{d \in \mathcal{D}_i} \text{cost}_{i,d}^\tau(\mathbf{z}_{i,d}^\tau)$, where $\text{cost}_{i,d}^\tau(\mathbf{z}_{i,d}^\tau)$ denotes the cost function for device d at time τ .

The discrete nature of the control actions associated with devices in the sets $\{\mathcal{D}_{S_i}\}$ would render pertinent optimization problems nonconvex. As in last chapter, we then consider utilizing the convex hull, defined as follows:

$$\text{conv}(\mathcal{Z}_{i,d}^t) := \{\mathbf{z}_{i,d}^t | (6.7), \mathbf{z}_{i,d}^\tau \in \text{conv}(\mathbb{P}_{i,d}), \forall \tau = t, \dots, t+w\}. \quad (6.15)$$

The relaxed set $\text{conv}(\mathcal{Z}_{i,d}^t)$ is *convex and compact*. Accordingly, consider the following convex sets:

$$\text{conv}(\mathcal{Z}_i^t) := \left(\prod_{d \in \mathcal{D}_{F_i}} \mathcal{Z}_{i,d}^t \right) \times \left(\prod_{d \in \mathcal{D}_{S_i}} \text{conv}(\mathcal{Z}_{i,d}^t) \right),$$

and

$$\text{conv}(\mathcal{Z}^t) := \prod_{i \in \mathcal{N}} \text{conv}(\mathcal{Z}_i^t). \quad (6.16)$$

While a convex hull is utilized for the algorithmic design, a randomization strategy will be leveraged to recover feasible control actions from $\text{conv}(\mathcal{Z}_{i,d}^t)$.

For given the vectors $(\boldsymbol{\alpha}_i^t, \boldsymbol{\beta}_i^t)$, and for a given optimization horizon of $w + 1$ time slots, the following problem is solved at node i at time t :

$$(\mathcal{P}_{1,i}^t) \min_{\mathbf{z}_i^t} C_i^t(\mathbf{z}_i^t) - \sum_{\tau=t}^{t+w} \sum_{d \in \mathcal{D}_i} (\alpha_i^\tau p_{i,d}^\tau + \beta_i^\tau q_{i,d}^\tau), \quad (6.17a)$$

$$\text{s.t.} \quad \mathbf{z}_i^t \in \text{conv}(\mathcal{Z}_i^t), \quad (6.17b)$$

where the terms $\alpha_i^\tau p_{i,d}^\tau$ and $\beta_i^\tau q_{i,d}^\tau$ are utilized to enable responsiveness to the network-level signals $(\boldsymbol{\alpha}_i^t, \boldsymbol{\beta}_i^t)$ (which can be interpreted as incentives or prices), and device-specific states are naturally accounted for in (6.17b). The following assumption is imposed.

Assumption 6.1. Functions $C_i^t(\mathbf{z}_i^t)$, $\forall i \in \mathcal{N}$ are strongly convex in $\mathbf{z}_i^t \in \text{conv}(\mathcal{Z}_i^t)$; i.e., there exists some $\sigma_c > 0$ such that $\nabla_{\mathbf{z}_i^t}^2 C_i^t(\mathbf{z}_i^t) \geq \sigma_c I$. Moreover, the gradient of $C_i^t(\mathbf{z}_i^t)$ with respect to \mathbf{z}_i^t is Lipschitz continuous for any $\mathbf{z}_i^t \in \text{conv}(\mathcal{Z}_i^t)$.

Since (6.17a) is strongly convex and that the constraints are convex, a unique solution \mathbf{z}_i^{t*} exists for given α_i^t and β_i^t , represented as the following mapping:

$$\mathbf{z}_i^{t*} = \mathbf{b}_i^t(\alpha_i^t, \beta_i^t). \quad (6.18)$$

6.1.3.2 Social-welfare problem

Let $g^t(\hat{\mathbf{y}}^t)$ be an affine function of the electrical quantities of interest $\hat{\mathbf{y}}^t$, and let the following constraints capture network-level operational constraints at time t :

$$g^t(\hat{\mathbf{y}}^t) \leq 0. \quad (6.19)$$

By equations (6.1)–(6.2), $\hat{\mathbf{y}}^t$ is affine in \mathbf{z}^t . For future development, define the vector-valued function $\mathbf{g}^t(\mathbf{z}^t) := [g^t(\hat{\mathbf{y}}^t(\mathbf{z}^t))^\top, \dots, g^{t+w}(\hat{\mathbf{y}}^{t+w}(\mathbf{z}^{t+w}))^\top]^\top$ within the time window W_t^w , which is affine in \mathbf{z}^t . Notice that the operational constraints are usually “independent” in the sense that none of them will subsume any other. We therefore have the following assumption.

Assumption 6.2. Function \mathbf{g}^t is an affine function of \mathbf{z}^t with full row rank.

With Assumption 6.2 and the boundedness of the set $\text{conv}(\mathcal{Z}^t)$, the Jacobian of $\mathbf{g}^t(\mathbf{z}^t)$ is bounded over $\text{conv}(\mathcal{Z}^t)$; i.e., there exists some constant $\sigma_g > 0$ such that

$$\|\nabla_{\mathbf{z}} \mathbf{g}^t(\mathbf{z}^t)\|_F \leq \sigma_g, \quad \forall \mathbf{z}^t \in \text{conv}(\mathcal{Z}^t), \quad (6.20)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

With these definitions in place, the following optimization problem is formulated to minimize the

aggregate cost incurred by the customers, subject to network constraints:

$$(\mathcal{P}_2^t) \quad \min \quad \sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t) \quad (6.21a)$$

$$\text{over } \mathbf{z}^t, \hat{\mathbf{y}}^t, \boldsymbol{\alpha}^t, \boldsymbol{\beta}^t \quad (6.21b)$$

$$\text{s.t. } \mathbf{p}_i^t = \mathbf{p}_{i,0}^t + \sum_{d \in \mathcal{D}_i} \mathbf{p}_{i,d}^t, \quad i \in \mathcal{N}, \quad (6.21c)$$

$$\mathbf{q}_i^t = \mathbf{q}_{i,0}^t + \sum_{d \in \mathcal{D}_i} \mathbf{q}_{i,d}^t, \quad i \in \mathcal{N}, \quad (6.21d)$$

$$\hat{\mathbf{y}}^\tau = A\mathbf{p}^\tau + B\mathbf{q}^\tau + \mathbf{c}, \quad \tau = t, \dots, t+w, \quad (6.21e)$$

$$\mathbf{g}^t(\hat{\mathbf{y}}^t) \leq 0, \quad (6.21f)$$

$$\mathbf{z}_i^t = \mathbf{b}_i^t(\boldsymbol{\alpha}_i^t, \boldsymbol{\beta}_i^t), \quad \forall i \in \mathcal{N}, \quad (6.21g)$$

where (6.21g) explicitly accounts for the solution of the customer-level problem $(\mathcal{P}_{1,i}^t)$. Notice that $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ are optimization variables, and they are designed in a way that customer responses (6.21g) do not lead to violation of network constraints.

Notice that even though (6.21g) renders problem (\mathcal{P}_2^t) nonconvex as it is usually not affine, convex reformulation and exact convex relaxation utilized in Section 4.2.1 still apply to (\mathcal{P}_2^t) . Before outlining the solution approach to address nonconvexity brought in by (6.21g) (see Section 6.2.1), we briefly explain how to recover discrete control actions for the devices \mathcal{D}_{S_i} in this scenario next.

6.1.3.3 Recovering feasible power setpoints

For a device $d \in \mathcal{D}_{S_i}$ with discrete power commands, the solution $\mathbf{p}_{i,d}^{t*}$ of the problem $(\mathcal{P}_{1,i}^t)$ may not be implementable, since it is computed based on the convex hull of the discrete set $\mathbb{P}_{i,d}^t$. To recover a feasible command implementable at time t , notice that $\mathbf{p}_{i,d}^{t*} \in \text{conv}(\mathbb{P}_{i,d}^t)$ can be written as a convex combination $\sum_m \theta_m \mathbf{p}_{i,d}^{t,m}$ of certain points $\mathbf{p}_{i,d}^{t,m}$ in $\mathbb{P}_{i,d}^t$, where $\theta_m \geq 0$ are such that $\sum_m \theta_m = 1$. Then, a feasible power command $\mathbf{p}_{i,d}^{t,m} \in \mathbb{P}_{i,d}^t$ can be randomly selected with probability θ_m .

As an illustrative example, assume that a device has two feasible power commands denoted by $\mathbf{p}_{i,d}^{t,1}$ and $\mathbf{p}_{i,d}^{t,2}$. It follows that $\mathbf{p}_{i,d}^{t,1} \leq \mathbf{p}_{i,d}^{t*} \leq \mathbf{p}_{i,d}^{t,2}$. Then, one can select $\mathbf{p}_{i,d}^{t,1}$ with probability $(\mathbf{p}_{i,d}^{t,2} - \mathbf{p}_{i,d}^{t*}) / (\mathbf{p}_{i,d}^{t,2} - \mathbf{p}_{i,d}^{t,1})$, and $\mathbf{p}_{i,d}^{t,2}$ with probability $(\mathbf{p}_{i,d}^{t*} - \mathbf{p}_{i,d}^{t,1}) / (\mathbf{p}_{i,d}^{t,2} - \mathbf{p}_{i,d}^{t,1})$.

6.2 Distributed Stochastic Dual Algorithm

At time t , problem (\mathcal{P}_2^t) naturally leads to a Stackelberg game where: (i) α^t and β^t are calculated via (\mathcal{P}_2^t) by the network operator and broadcasted to all nodes $i \in \mathcal{N}$; and, (ii) each consumer computes the power set points z_i^{t*} from $(\mathcal{P}_{1,i}^t)$. By design, z^{t*} is in fact an optimal point for (\mathcal{P}_2^t) .

In the following, we first recall a result presented in Section 4.2.1 to obtain an exact convex relaxation of (\mathcal{P}_2^t) regarding non-convex constraint (6.21g). Then, Section 6.2.2 and Section 6.2.3 will present a stochastic dual algorithm to identify optimal points of (\mathcal{P}_2^t) and to recover feasible power commands for devices in $\{\mathcal{D}_{S_i}\}$.

6.2.1 Exact Convex Relaxation

Consider the following convex optimization problem:

$$\mathcal{P}_3^t) \quad \min \quad \sum_{i \in \mathcal{N}} C_i^t(z_i^t) \quad (6.22a)$$

$$\text{over } z^t, \hat{y}^t \quad (6.22b)$$

$$\text{s.t. } p_i^t = p_{i,0}^t + \sum_{d \in \mathcal{D}_i} p_{i,d}^t, \quad i \in \mathcal{N}, \quad (6.22c)$$

$$q_i^t = q_{i,0}^t + \sum_{d \in \mathcal{D}_i} q_{i,d}^t, \quad i \in \mathcal{N}, \quad (6.22d)$$

$$\hat{y}^\tau = A p^\tau + B q^\tau + c, \quad \tau = t, \dots, t+w, \quad (6.22e)$$

$$g^t(\hat{y}^t) \leq 0, \quad (6.22f)$$

$$z_i^t \in \text{conv}(\mathcal{Z}_i^t), \quad i \in \mathcal{N}, \quad (6.22g)$$

which is obtained from (\mathcal{P}_2^t) by dropping the nonconvex constraint (6.21g) and instead adding the device-specific constraints (6.22g). The following is assumed.

Assumption 6.3. *Problem (\mathcal{P}_3^t) is strictly feasible, i.e., it satisfies Slater's condition at each time t .*

Given the strong convexity of the cost function (6.22a) in z^t , a unique optimal solution exists for problem (\mathcal{P}_3^t) . Leveraging the results of Section 4.2.1, we can show that (\mathcal{P}_3^t) together with a particular choice of α^t and β^t lead to an exact convex relaxation of (\mathcal{P}_2^t) ; i.e., the solution of (\mathcal{P}_3^t) coincides with an

optimal solution of (\mathcal{P}_2^t) . Specifically, substitute (6.22c)–(6.22e) into (6.22f), and denote by μ^τ the dual variable associated with the constraints (6.22f) for a given τ . Let \hat{y}^{t*} be the optimal solution of (\mathcal{P}_3^t) , and denote the optimal dual variables associated with (6.22f) as $\mu^{\tau*}$ for time τ . Consider then the following choice for α^t and β^t :

$$\alpha^{\tau*} = -A^\top \nabla_{\hat{y}} g^\tau(\hat{y}^{\tau*}) \mu^{\tau*}, \quad (6.23a)$$

$$\beta^{\tau*} = -B^\top \nabla_{\hat{y}} g^\tau(\hat{y}^{\tau*}) \mu^{\tau*} \quad (6.23b)$$

for all $\tau = t, \dots, t + w$. Then, Theorems 4.1–4.2 can be extended to obtain the following result.

Theorem 6.1. *Under Assumptions 6.1–6.3, it follows that $(\alpha^{t*}, \beta^{t*})$ defined by (6.23) are bounded. Moreover, the solution of problem (\mathcal{P}_3^t) along with $(\alpha^{t*}, \beta^{t*})$ defined in (6.23) is a globally optimal solution of the nonconvex problem (\mathcal{P}_2^t) .*

Hereafter, we will refer to the globally optimal solutions of (\mathcal{P}_3^t) and (\mathcal{P}_2^t) interchangeably, depending on the context. In the next section, we will design a stochastic dual algorithm for solving (\mathcal{P}_3^t) in an offline or batch setting, and show how to recover feasible commands for devices with discrete power levels. Subsequently, we will develop an online stochastic dual algorithm.

Remark 6.1. *From a practical standpoint, α^t and β^t can be interpreted as “incentive signals” or “prices” that the network operator communicates to customers to ensure that engineering constraints in the network are satisfied. For example, when y_t collects voltage magnitudes and the function g is designed to ensure voltage regulation, α^t and β^t can be understood as prices of voltage violation.*

6.2.2 Offline Stochastic Dual Algorithm

Consider the Lagrangian function associated with (\mathcal{P}_3^t) :

$$\mathcal{L}^t(\mathbf{z}^t, \boldsymbol{\mu}^t) = \sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t) + \boldsymbol{\mu}^{t\top} \mathbf{g}^t(\mathbf{z}^t), \quad (6.24)$$

where $\boldsymbol{\mu}^t$ is the vector of dual variables associated with the network constraints (6.22f). The Lagrangian (6.24) is obtained by substituting (6.22c)–(6.22e) into (6.22f) and by keeping the constraints (6.22g) implicit. De-

find the primal optimal solution at time t as for a given dual $\boldsymbol{\mu}^t$:

$$\mathbf{z}^{t*} := \arg \min_{\mathbf{z}^t} \mathcal{L}^t(\mathbf{z}^t, \boldsymbol{\mu}^t), \quad (6.25a)$$

$$\text{s.t. (6.22g),} \quad (6.25b)$$

and consider the following dual problem

$$\max_{\boldsymbol{\mu}^t \geq 0} h^t(\boldsymbol{\mu}^t) \quad (6.26)$$

where the dual function is defined as:

$$h^t(\boldsymbol{\mu}^t) = \mathcal{L}^t(\mathbf{z}^{t*}, \boldsymbol{\mu}^t). \quad (6.27)$$

We then have the following result.

Lemma 6.1 (Theorem 2.1 of [93]). *Under Assumptions 6.1–6.2, the gradient of the dual function is given by $\nabla_{\boldsymbol{\mu}} h^t(\boldsymbol{\mu}^t) = \mathbf{g}^t(\mathbf{z}^{t*})$. Furthermore, the gradient of the dual function is Lipschitz continuous with constant $\frac{\sigma_g^2}{\sigma_c}$, i.e.,*

$$\|\nabla_{\boldsymbol{\mu}} h^t(\boldsymbol{\mu}) - \nabla_{\boldsymbol{\mu}} h^t(\tilde{\boldsymbol{\mu}})\| \leq \frac{\sigma_g^2}{\sigma_c} \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\| \quad (6.28)$$

for any feasible $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\mu}}$.

Using the results of Theorem 2.1 of [93], we augment a dual gradient method for solving (6.26) with the a randomization strategy to recover discrete power commands at each iteration. The resultant stochastic dual gradient algorithm involves a sequential execution of the following steps:

$$[\text{S1}] \mathbf{z}^{t*}(k+1) = \arg \min_{\mathbf{z}^t} \mathcal{L}^t(\mathbf{z}^t, \boldsymbol{\mu}^t(k)), \text{ s.t. (6.22g),} \quad (6.29a)$$

[S2] For $d \in \mathcal{D}_{S_i}$, pick $z_{i,d}^t(k+1) \in \mathcal{Z}_{i,d}^t$ randomly based on

$$z_{i,d}^{t*}(k+1) \text{ using scheme described in Section 6.1.3.3,} \quad (6.29b)$$

[S3] Set $z_{i,d}^{\tau}(k+1) = z_{i,d}^{t*}(k+1), \forall d \in \mathcal{D}_{S_i}, \forall \tau \neq t$,

$$\text{and } z_{i,d}^t(k+1) = z_{i,d}^{t*}(k+1), \forall d \in \mathcal{D}_{F_i}, \quad (6.29c)$$

$$[\text{S4}] \boldsymbol{\mu}^t(k+1) = [\boldsymbol{\mu}^t(k) + \varepsilon \mathbf{g}^t(\mathbf{z}^t(k+1))]_{+}, \quad (6.29d)$$

where $[\]_+$ denotes the projection onto the nonnegative orthant, and $\varepsilon > 0$ is a given constant stepsize. Notice that (6.29c) recovers feasible power commands only for the current time t (and not for $\tau = t + 1, \dots, t + w$). Further, in step (6.29d), $\mathbf{g}^t(\mathbf{z}^t(k+1))$ is stochastic gradient of $h(\boldsymbol{\mu}^t(k))$ with $E[h(\boldsymbol{\mu}^t(k))] = \mathbf{g}^t(\mathbf{z}^{t*}(k+1))$ (the gradient).

Lemma 6.2. *Under Assumptions 6.1-6.2, the dual function $h^t(\boldsymbol{\mu}^t)$ is strongly concave.*

Proof. By definition, the dual function is given by

$$\begin{aligned} h^t(\boldsymbol{\mu}^t) &= \min_{\mathbf{z}^t} \sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t) + \boldsymbol{\mu}^{t\top} \mathbf{g}^t(\mathbf{z}^t) \\ &= -\max_{\mathbf{z}^t} \sum_{i \in \mathcal{N}} -C_i^t(\mathbf{z}_i^t) - \boldsymbol{\mu}^{t\top} \mathbf{g}^t(\mathbf{z}^t). \end{aligned} \quad (6.30)$$

On the other hand, the conjugate function of $\sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t)$ is defined as

$$\tilde{C}^t(\boldsymbol{\phi}^t) = \sum_{i \in \mathcal{N}} \tilde{C}_i^t(\boldsymbol{\phi}_i^t) := \max_{\mathbf{z}^t} \boldsymbol{\phi}^{t\top} \mathbf{z}^t - \sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t), \quad (6.31)$$

which further gives:

$$\tilde{C}^t(-\nabla_{\mathbf{z}} \mathbf{g}^t \cdot \boldsymbol{\mu}^t) = \max_{\mathbf{z}^t} -(\nabla_{\mathbf{z}} \mathbf{g}^t \cdot \boldsymbol{\mu}^t)^\top \mathbf{z}^t - \sum_{i \in \mathcal{N}} C_i^t(\mathbf{z}_i^t).$$

Under Assumption 6.2, \mathbf{g}^t is affine in \mathbf{z}^t . Therefore, $h^t(\boldsymbol{\mu}^t) = -\tilde{C}^t(-\nabla_{\mathbf{z}} \mathbf{g}^t \boldsymbol{\mu}^t)$. Moreover, under Assumption 6.1, it follows from [103, Proposition 12.60] that the conjugate function $\tilde{C}_i^t(\boldsymbol{\phi}_i^t)$ is strongly convex. Thus $h^t(\boldsymbol{\mu}^t)$ is strongly concave. \blacksquare

Let $\sigma_h > 0$ be the strong concavity coefficient of $h^t(\boldsymbol{\mu}^t)$; that is, for any feasible $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\mu}}$, it holds that:

$$(\nabla_{\boldsymbol{\mu}} h^t(\boldsymbol{\mu}) - \nabla_{\boldsymbol{\mu}} h^t(\tilde{\boldsymbol{\mu}}))^\top (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) \leq -\sigma_h \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|^2. \quad (6.32)$$

Furthermore, let $\Delta > 0$ be a given scalar such that the following holds uniformly in time:

$$\text{Var}(\mathbf{g}^t(\mathbf{z}^t)) := E[\|\mathbf{g}^t(\mathbf{z}^t)\|^2] - \|\mathbf{g}^t(\mathbf{z}^{t*})\|^2 \leq \Delta, \quad (6.33)$$

where the variance is taken with respect to the randomization step (6.29c). In other words, $\text{Var}(\mathbf{g}^t(\mathbf{z}^t))$ represent the variance of the discrepancy between the constraint function evaluated at the relaxed solution and at the randomized solution.

The following convergence result for the dual variables can be stated.

Theorem 6.2. *Under Assumptions 6.1–6.3, and with a stepsize ε chosen such that*

$$0 < 1 + \varepsilon^2 \sigma_g^4 / \sigma_c^2 - 2\varepsilon \sigma_h < 1, \quad (6.34)$$

the stochastic dual algorithm (6.29) converges as

$$\lim_{k \rightarrow \infty} E[\|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2] = \frac{\varepsilon \Delta}{2\sigma_h - \varepsilon \sigma_g^4 / \sigma_c^2}. \quad (6.35)$$

□

Proof. Let $\boldsymbol{\mu}^{t*}$ be the vector of unique optimal dual variables. Then,

$$\begin{aligned} & E[\|\boldsymbol{\mu}^t(k+1) - \boldsymbol{\mu}^{t*}\|^2 | \boldsymbol{\mu}^t(k)] \\ & \leq E[\|\boldsymbol{\mu}^t(k) + \varepsilon \mathbf{g}^t(\mathbf{z}^t(k+1)) - \boldsymbol{\mu}^{t*} - \varepsilon \mathbf{g}^t(\mathbf{z}^{t*})\|^2 | \boldsymbol{\mu}^t(k)] \\ & = \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2 + \varepsilon^2 E[\|\mathbf{g}^t(\mathbf{z}^t(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})\|^2 | \boldsymbol{\mu}^t(k)] + 2\varepsilon (\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*})^\top (\mathbf{g}^t(\mathbf{z}^{t*}(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})) \\ & = \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2 + \varepsilon^2 E[\|\mathbf{g}^t(\mathbf{z}^t(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})\|^2 | \boldsymbol{\mu}^t(k)] + 2\varepsilon (\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*})^\top (\nabla h^t(\boldsymbol{\mu}^t(k)) - \nabla h^t(\boldsymbol{\mu}^{t*})) \\ & \leq \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2 + \varepsilon^2 E[\|\mathbf{g}^t(\mathbf{z}^t(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})\|^2 | \boldsymbol{\mu}^t(k)] - 2\varepsilon \sigma_h \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2, \end{aligned}$$

where the first inequality follows from the non-expansiveness property of the projection operator; the first equality is due to the fact that \mathbf{g}^t is linear and it accounts for the recovery of discrete solutions; the second equality follows from Lemma 6.1; and the last inequality is due to the strong concavity of $h^t(\boldsymbol{\mu}^t)$.

Notice that

$$\begin{aligned} & E[\|\mathbf{g}^t(\mathbf{z}^t(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})\|^2 | \boldsymbol{\mu}^t(k)] \\ & = \text{Var}(\mathbf{g}^t(\mathbf{z}^t(k+1)) | \boldsymbol{\mu}^t(k)) + \|\mathbf{g}^t(\mathbf{z}^{t*}(k+1)) - \mathbf{g}^t(\mathbf{z}^{t*})\|^2 \\ & \leq \Delta + \|\nabla h^t(\boldsymbol{\mu}^t(k)) - \nabla h^t(\boldsymbol{\mu}^{t*})\|^2 \\ & \leq \Delta + \sigma_g^4 / \sigma_c^2 \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2, \end{aligned}$$

where the last inequality is due to the Lipschitz continuity of $\nabla h^t(\boldsymbol{\mu}^t)$. We then obtain the following inequality:

$$E[\|\boldsymbol{\mu}^t(k+1) - \boldsymbol{\mu}^{t*}\|^2 | \boldsymbol{\mu}^t(k)] \leq (1 + \varepsilon^2 \sigma_g^4 / \sigma_c^2 - 2\varepsilon \sigma_h) \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2 + \varepsilon^2 \Delta.$$

By taking the total expectation on both sides and by recursively computing the steps above, one obtains:

$$\begin{aligned} & E[\|\boldsymbol{\mu}^t(k+1) - \boldsymbol{\mu}^{t*}\|^2] \\ & \leq (1 + \varepsilon^2 \sigma_g^4 / \sigma_c^2 - 2\varepsilon \sigma_h)^k E[\|\boldsymbol{\mu}^t(1) - \boldsymbol{\mu}^{t*}\|^2] + \varepsilon^2 \Delta (1 - (1 + \varepsilon^2 \sigma_g^4 / \sigma_c^2 - 2\varepsilon \sigma_h)^k) / (2\varepsilon \sigma_h - \varepsilon^2 \sigma_g^4 / \sigma_c^2). \end{aligned}$$

With ε chosen as in (6.34), the result (6.35) follows. ■

Notice that (6.34) allows to select a stepsize $\varepsilon > 0$ that is “small enough” to satisfy the converge requirement. It is also worth pointing out that when $\Delta = 0$ (i.e., no devices with discrete power commands are present), the right-hand-side of (6.35) goes to zero, corresponding to (6.29) reducing to the standard dual gradient algorithm when no discrete variables are present.

Corollary 6.1. *When $\mathcal{D}_{S_i} = \emptyset$ for all i (i.e., there are no discrete optimization variables), and under Assumptions 6.1–6.3, if the stepsize ε satisfies (6.34), then (6.29) converges to the exact optimal solution of (\mathcal{P}_3^t) asymptotically; i.e.,*

$$\lim_{k \rightarrow \infty} \|\boldsymbol{\mu}^t(k) - \boldsymbol{\mu}^{t*}\|^2 = 0 \tag{6.36a}$$

$$\lim_{k \rightarrow \infty} \|\mathbf{z}^t(k) - \mathbf{z}^{t*}\|^2 = 0. \tag{6.36b}$$

We next present a distributed implementation of the proposed method, along with a receding horizon optimization strategy.

6.2.3 Distributed Implementation and Receding Horizon Optimization

The stochastic dual algorithm (6.29) can be implemented in a distributed fashion by leveraging the decomposability of the Lagrangian function. Specifically, step (6.29a) is decomposable on a per-node basis, where each customer/node i can update the power commands of the devices \mathcal{D}_i once the vectors $(\boldsymbol{\alpha}^t, \boldsymbol{\beta}^t)$ are received. On the other hand, the dual step is performed by a network operator. The overall distributed algorithm is tabulated as Algorithm 3.

Problem (\mathcal{P}_3^t) is a multi-period problem. To optimize the operation of both network and devices, problem (\mathcal{P}_3^t) can be embedded into a receding horizon control (RHC) strategy [88] where:

- (i) At time t , the temporal window of w time slots $\mathcal{T}_k := \{t_k, t_{k+1}, \dots, t_{k+w}\}$ is considered, and the offline Algorithm 3 is utilized to solve (\mathcal{P}_3^t) to convergence;
- (ii) The solution $\{z_{i,d}^{t*}\}$ corresponding the first slot t is implemented;
- (iii) The temporal window is shifted of one time slot $\mathcal{T}_k \rightarrow \mathcal{T}_{k+1}$; and,
- (iv) Once the window is shifted, point (i) is repeated.

This strategy is consistent with traditional RHC methods. However, the premises here are that:

- Each time slot is “long enough” to allow the offline Algorithm 3 to converge to the solution of (\mathcal{P}_3^t) ;
- Within each time slot, the problem inputs are not changing; i.e., prevailing ambient and operational conditions are invariant over a time slot.

In the following section, we will present an *online* algorithm that can cope with cases where prevailing ambient and operational conditions vary fast, and they lead to problem inputs that change even within an iteration (or a few iterations) of the Algorithm 3. The resultant algorithm can be interpreted as a real-time RHC strategy where the temporal window is shifted every iteration of the online algorithm (with the shift being determined by the computational time of the algorithmic steps).

Remark 6.2. *Since this chapter primarily focuses on the design and analysis of Algorithm 3 and its online counterpart presented in Section 6.3, errors in the forecasting of problem inputs (such as maximum available PV generation, uncontrollable loads, etc; e.g., [61,64]) are not considered in the chapter. On the other hand, the proposed framework could be equally applicable to robust or chance-constrained counterparts of (\mathcal{P}_3^t) , so long as the optimization problem is convex.*

Algorithm 3 Offline Distributed Algorithm

At time t ,

repeat

[S1] Node i measure the state x_i^t .

[S2] Given $(\alpha_i^t(k), \beta_i^t(k))$, node i computes $z^{t*}(k+1)$ by solving $(\mathcal{P}_{1,i}^t)$, and recovers feasible power set points $z^t(k+1)$ for $d \in \mathcal{D}_{S_i}$ via randomization.

[S3] Node i updates $p_i^t(k+1)$ and $q_i^t(k+1)$ based on (6.2) and sends the results to the network operator.

[S4] Network operator calculates system states for $\tau = t, \dots, t+w$ as:

$$\hat{y}^\tau(k+1) = A p^\tau(k+1) + B q^\tau(k+1) + c.$$

[S5] Network operator updates dual variables as:

$$\mu^t(k+1) = [\mu^t(k) + \varepsilon g^t(\hat{y}^t(k+1))]_{+}.$$

[S6] Network operator computes the following quantities for $\tau = t, \dots, t+w$:

$$\alpha^\tau(k+1) = -A^\top \nabla_{\hat{y}} g^\tau(\hat{y}^\tau(k+1)) \mu^\tau(k+1),$$

$$\beta^\tau(k+1) = -B^\top \nabla_{\hat{y}} g^\tau(\hat{y}^\tau(k+1)) \mu^\tau(k+1),$$

and sends them to the nodes.

until meeting stopping criterion

6.3 Online and Asynchronous Optimization

In this section, we address the case where computational and communication constraints may render infeasible the distributed solution of (\mathcal{P}_3^t) — and, hence, (\mathcal{P}_2^t) — to convergence (i.e., offline solution) at a timescale that is consistent with the variability of the underlying ambient and operational conditions. We consider a situation with:

(i) Real-time implementation: power setpoints are implemented whenever they are updated numerically.

(ii) Asynchronous computation: power setpoints for devices in the set \mathcal{D}_{F_i} (with continuous decision

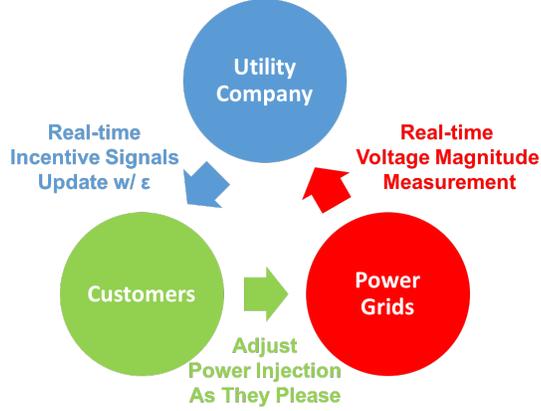


Figure 6.1: Illustration of the online asynchronous distributed algorithm.

variables) are updated every iteration of the algorithm (e.g., every second or a few seconds); on the other hand, the setpoints of device \mathcal{D}_{S_i} (with discrete decision variables) are updated at a slower time scale (depending on, e.g., the availability of the device). This leads to an *asynchronous* control algorithm, where different devices are controlled at different timescales.

(iii) The temporal window of the multi-period problem is advanced at each iteration (or every few iterations, depending on particular implementations) of the algorithm.

(iv) The resultant online algorithm will leverage *feedback* (see also [39, 138]), to substitute the linearized model (6.1) with measurements of \hat{y}^t for the first timeslot of the optimization window. This accounts for the nonlinear power flows and also helps reduce communication and computation overhead.

6.3.1 Online Asynchronous Algorithm

Let $\mathcal{T}_{i,d} := \{t_{i,d}^1, t_{i,d}^2, \dots\}$ denote the set of time indexes where the device d of node i updates the local variables. Let $|\mathcal{T}_{i,d}|$ denote the cardinality of $\mathcal{T}_{i,d}$. Let $\mathcal{D}_i^t \subset \mathcal{D}_i$ be the subset of devices that update power setpoints at time t , and let $\mathcal{D}^t = \bigcup_{i \in \mathcal{N}} \mathcal{D}_i^t$. The quantities $\{z_{i,d}^t\}_{d \in \mathcal{D}^t}$ are treated as constants at time t ; thus, define a “reduced” Lagrangian as

$$\mathcal{L}_r^t(\{z_{i,d}^t\}_{d \in \mathcal{D}^t}, \mu_r^t) := \mathcal{L}^t(\{z_{i,d}^t\}_{d \in \mathcal{D}^t}, \mu^t | \{z_{i,d}^t\}_{d \notin \mathcal{D}^t}), \quad (6.37)$$

and define its corresponding “reduced” dual function as:

$$h_r^t(\boldsymbol{\mu}_r^t) := \min_{\{z_{i,d}^t\}_{d \in \mathcal{D}^t}} \mathcal{L}_r^t(\{z_{i,d}^t\}_{d \in \mathcal{D}^t}, \boldsymbol{\mu}_r^t), \quad (6.38)$$

whose unique optimal solution is:

$$\boldsymbol{\mu}_r^{t*} := \arg \max_{\boldsymbol{\mu}_r^t \geq 0} h_r^t(\boldsymbol{\mu}_r^t). \quad (6.39)$$

Based on the definitions above, we propose the online asynchronous algorithm tabulated as Algorithm 4. Notice that we have included an additional subscript to denote the time when decision variables are calculated; e.g., $\boldsymbol{\mu}_{r,t}^{t+1}$ is the dual vector for the reduced Lagrangian from $t + 1$ to $t + w + 1$ calculated at time t . We also recall that in Algorithm 4, only *one* iteration is carried out at time t , and then the temporal window is shifted from t to $t + 1$.

6.3.2 Performance Analysis

To analyze the performance of Algorithm 4, the following assumptions are made.

Assumption 6.4. *There exists some constant $e > 0$ such that the variation of the optimal $\boldsymbol{\mu}_r^*$ of (\mathcal{P}_3^t) over any two consecutive timeslots is bounded as*

$$\|\boldsymbol{\mu}_r^{t+1*} - \boldsymbol{\mu}_r^{t*}\|^2 \leq e, \quad \forall t. \quad (6.40)$$

This is a standard assumption in the domain of time-varying optimization [39, 109] to characterize the variability of optimal solutions from the current timeslot to the next (and, hence, the variability of problem inputs).

Let $\hat{\mathbf{y}}^t := [\hat{y}^{t\top}, \hat{y}^{t+1\top}, \dots, \hat{y}^{t+w\top}]^\top$, and $\mathbf{y}^t := [y^{t\top}, \hat{y}^{t+1\top}, \dots, \hat{y}^{t+w\top}]^\top$, where the first element of the vector $\hat{\mathbf{y}}^t$ is measured from the grid. We next assume bounded error due to the linearization of power flow equations.

Assumption 6.5. *Given any feasible power injection z^t , there exists a constant $\rho > 0$ such that*

$$\|\mathbf{g}^t(\mathbf{y}^t(z^t)) - \mathbf{g}^t(\hat{\mathbf{y}}^t(z^t))\|^2 \leq \rho. \quad (6.41)$$

Algorithm 4 Online Asynchronous Algorithm

At time t ,

[S1] Node i updates measures $x_{i,t}^t$.

[S2] Given $(\alpha_{i,t-1}^t, \beta_{i,t-1}^t)$, Node i sets $z_{i,d,t}^t = z_{i,d,t-1}^{t-1}$, $\forall d \notin \mathcal{D}_i^t$, and solves $(\mathcal{P}_{\mathbf{1},i}^t)$ over $\{z_{i,d,t}^t\}_{d \in \mathcal{D}_i^t}$ to get the solution $\{z_{i,d,t}^{t*}\}_{d \in \mathcal{D}_i^t}$. Recover and implement discrete power set points for \mathcal{D}_{S_i} via randomization, with $E[z_{i,d,t}^t] = z_{i,d,t}^{t*}$.

[S3] Node i calculates future aggregated power $(z_{i,t}^{t+1*}, \dots, z_{i,t}^{t+w*})$ based on (6.2), and sends the result to network operator.

[S4] Network operator measures the current states y_t^t , and estimates the future states \hat{y}_t^τ for $\tau = t+1, \dots, t+w$ by

$$\hat{y}_t^\tau = Ap_t^{\tau*} + Bq_t^{\tau*} + c.$$

[S5] Network operator updates the dual variables for $t+1$ with *measured* states by

$$\mu_{r,t}^{t+1} = [\mu_{r,t-1}^t + \varepsilon g^t(y_t^t)]_+.$$

and updates the dual variables for $\tau = t+1, \dots, t+w$ with *predicted* states by

$$\mu_{r,t}^{\tau+1} = [\mu_{r,t-1}^\tau + \varepsilon g^\tau(\hat{y}_t^\tau)]_+.$$

[S6] Network operator updates signals for $t+1$ by

$$\alpha_i^{t+1} = -A^\top \nabla_{\hat{y}} g^t(y_t^t) \mu_{r,t}^{t+1}, \beta_t^{t+1} = -B^\top \nabla_{\hat{y}} g^t(y_t^t) \mu_{r,t}^{t+1},$$

and those for $\tau = t+1, \dots, t+w$ by

$$\alpha_i^{\tau+1} = -A^\top \nabla_{\hat{y}} g^\tau(\hat{y}_t^\tau) \mu_{r,t}^{\tau+1}, \beta_t^{\tau+1} = -B^\top \nabla_{\hat{y}} g^\tau(\hat{y}_t^\tau) \mu_{r,t}^{\tau+1},$$

and sends the results $(\alpha_{i,t}^{t+1}, \beta_{i,t}^{t+1})$ to node i .

[S7] Shift temporal window from t to $t+1$, and go to [S1].

Then, the following convergence result can be stated.

Theorem 6.3. *Under Assumptions 6.1–6.5, and given a stepsize ε chosen to satisfy (6.34), Algorithm 4 converges as*

$$\limsup_{t \rightarrow \infty} E[\|\boldsymbol{\mu}_{r,t}^{t+1} - \boldsymbol{\mu}_r^{t+1*}\|^2] = \frac{\varepsilon^2 \Delta + \varepsilon^2 \rho + e}{2\varepsilon\sigma_h - \varepsilon^2\sigma_g^4/\sigma_c^2}. \quad (6.42)$$

Proof. Denote by $\boldsymbol{\mu}_{r,t}^{t+1}$ the dual produced by Algorithm 4 at time t and $\boldsymbol{\mu}_r^{t+1*}$ the optimal dual solution. We have

$$\begin{aligned} & E[\|\boldsymbol{\mu}_{r,t}^{t+1} - \boldsymbol{\mu}_r^{t+1*}\|^2] \\ & \leq E[\|\boldsymbol{\mu}_{r,t-1}^t + \varepsilon \mathbf{g}^t(\hat{\mathbf{y}}_{t-1}^t) + \varepsilon \mathbf{g}^t(\mathbf{y}_{t-1}^t) - \varepsilon \mathbf{g}^t(\hat{\mathbf{y}}_{t-1}^t) - \boldsymbol{\mu}_r^{t*} + \boldsymbol{\mu}_r^{t*} - \boldsymbol{\mu}_r^{t+1*}\|^2] \\ & \leq E[\|\boldsymbol{\mu}_{r,t-1}^t + \varepsilon \mathbf{g}^t(\hat{\mathbf{y}}_{t-1}^t) - \boldsymbol{\mu}_r^{t*}\|^2 + \|\boldsymbol{\mu}_r^{t*} - \boldsymbol{\mu}_r^{t+1*}\|^2 + \|\varepsilon \mathbf{g}^t(\mathbf{y}_{t-1}^t(\mathbf{z}_{t-1}^t)) - \varepsilon \mathbf{g}^t(\hat{\mathbf{y}}_{t-1}^t(\mathbf{z}_{t-1}^t))\|^2] \\ & \leq E[\|\boldsymbol{\mu}_{r,t-1}^t + \varepsilon \mathbf{g}^t(\mathbf{z}_{t-1}^t) - \boldsymbol{\mu}_r^{t*} - \varepsilon \mathbf{g}^t(\mathbf{z}^*)\|^2] + e + \varepsilon^2 \rho, \end{aligned}$$

where the last inequality is due to (6.40)–(6.41). Using Lemma 6.1 and Lemma 6.2 (which still hold for the reduced dual function), and recalling that Δ is the upper bound on the variance $\text{Var}(\mathbf{g}^t(\mathbf{z}^t))$, one has that:

$$\begin{aligned} & E[\|\boldsymbol{\mu}_{r,t}^{t+1} - \boldsymbol{\mu}_r^{t+1*}\|^2] \\ & \leq (1 + \varepsilon^2\sigma_g^4/\sigma_c^2 - 2\varepsilon\sigma_h)E[\|\boldsymbol{\mu}_{r,t-1}^t - \boldsymbol{\mu}_r^{t*}\|^2] + \varepsilon^2\Delta + \varepsilon^2\rho + e \\ & \leq (1 + \varepsilon^2\sigma_g^4/\sigma_c^2 - 2\varepsilon\sigma_h)^t E[\|\boldsymbol{\mu}_{r,0}^1 - \boldsymbol{\mu}_r^{1*}\|^2] + (\varepsilon^2\Delta + \varepsilon^2\rho + e) \frac{1 + (1 + \varepsilon^2\sigma_g^4/\sigma_c^2 - 2\varepsilon\sigma_h)^t}{2\varepsilon\sigma_h - \varepsilon^2\sigma_g^4/\sigma_c^2}. \end{aligned}$$

Selecting the stepsize ε in a way to satisfy (6.34), and letting $t \rightarrow \infty$, the main result (6.42) follows. \blacksquare

The bound (6.42) provides a characterization of the discrepancy between the optimal dual variable and the dual variable generated by the online algorithm. The asymptotic bound depends on the underlying dynamics of the optimization problem through e and on the measurement errors and linearization errors through ρ . The result (6.42) can also be interpreted as input-to-state stability, where the trajectory of the optimal dual variables is taken as a reference.

The result (6.42) also suggests ways to improve the performance of Algorithm 4 in terms of tracking. For example, more frequent updates leads to smaller e ; a finer control granularity for discrete devices leads to a smaller Δ .

6.4 Numerical Examples

6.4.1 Simulation Setup

Consider a modified version of the IEEE 37-node test feeder shown in Fig. 4.1. The modified network is obtained by considering the phase “c” of the original system and by replacing the loads specified in the original dataset with real load data measured from feeders in Anatolia, California during a week of August 2012 [15]. Particularly, the data have a granularity of 1 second, and represent the loading of secondary transformers. Line impedances, shunt admittances as well as active and reactive loads are adopted from the respective data set. It is assumed that 18 PV systems are located at nodes 4, 7, 10, 13, 17, 20, 22, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35 and 36, and their generation profiles are simulated based on the real solar irradiance data available in [15]. The ratings of these inverters are 300 kVA for $i = 3$, 350 kVA for $i = 15, 16$, and 200 kVA for the remaining inverters. Loads and the power available from a PV system with capacity of 50 kW are reported in Fig. 6.2 for illustrative purpose. We then install 15 A/Cs and 15 batteries at each of the following 25 nodes: 2, 5, 6, 7, 9, 10, 11, 13, 14, 16, 18, 19, 20, 21, 22, 24, 26, 27, 28, 29, 30, 32, 33, 35 and 36, totaling 375 A/Cs and 375 batteries. The detailed simulation modeling of PV inverters, A/Cs, and batteries are described as follows.

PV inverters: The PV inverters’ objective functions are set uniformly to

$$C_{i,d}^t(p_{i,d}^t, q_{i,d}^t) = c_p (p_{i,d}^{\text{av},t} - p_{i,d}^t)^2 + c_q q_{i,d}^{t2}, \quad (6.43)$$

with positive constant c_p and c_q , in an effort to minimize the amount of real power curtailed from the available power $p_{i,d}^{\text{av},t}$ based on irradiance conditions at time t , and the amount of reactive power injected or absorbed. PV inverters are set to be updated every second within a convex feasible set as (6.3).

A/Cs: We set a uniform cost function for all A/Cs as

$$C_{i,d}(T_{i,d}^t) = c_t \|T_{i,d}^t - \mathbf{1} \cdot T_{i,d}^{\text{nom}}\|^2, \quad (6.44)$$

where c_t is positive constant, $T_{i,d}^{\text{nom}}$ is a preferred room temperature set at 75°F, and the future room temperature and constraints are modeled according to (6.8)–(6.10) with θ_1 uniformly set to 0.1, and θ_2 randomly

picked from $[-0.0009, -0.0011]$. The acceptable ranges of room temperatures are set to $[70^\circ\text{F}, 80^\circ\text{F}]$. Each A/C updates every 15 minutes with two possible power status: 4 kW (on) and 0 (off).

Batteries: We set a uniform cost function for batteries as

$$C_{i,d}(\mathbf{S}_{i,d}^t) = c_b \|\mathbf{S}_{i,d}^t - \mathbf{1} \cdot S_{i,d}^{\text{nom}}\|^2, \quad (6.45)$$

where c_b is a positive constant, $S_{i,d}^{\text{nom}}$ is a preferred battery state of charge set to 0.5 for all batteries for illustrative purpose, and the battery dynamics and constraints are modeled as (6.12)–(6.13), with charging/discharging efficiency ξ set to 1 and SOC bounds to $[0.2, 0.8]$ uniformly for simplicity.¹ For each battery, we set a uniform charging rate of 4 kW, and a fixed discharging rate randomly picked between 3.6 kW and 4.4 kW. Batteries update power status of charging/off/discharging every 15 minutes.

Operational constraints $g^t(\hat{y}^t) \leq 0$ are set to voltage regulation:

$$\underline{v}_i \leq \hat{v}_i^t \leq \bar{v}_i, \quad (6.46)$$

with voltage upper and lower bounds \bar{v}_i and \underline{v}_i set to 1.04 p.u. and 0.96 p.u. respectively for $\forall i \in \mathcal{N}$.

For the rest of this section, we will focus on illustrating performance of online asynchronous algorithm, i.e., Algorithm 4. We refer to numerical results in Chapter 5 for numerical examples of stability analysis for offline algorithm.

¹ Here, this battery model is simplified for better illustrating battery's response towards signals. In practice, one can apply objective function to avoid frequent switch between charging and discharging, deep charge/discharge that can damage battery, etc., and the constraints can involve charging deadline, e.g., charge the EV battery with at least 90% SOC by 8 am.

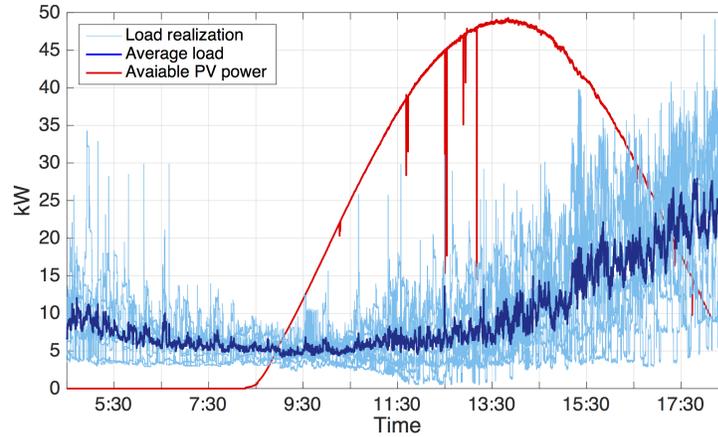


Figure 6.2: Profiles of loads and power available from the PV systems. The average load profile is marked in blue.

6.4.2 Online Asynchronous Distributed Algorithm

We implement Algorithm 4 for one day to coordinate network operator and customers to achieve operational and economic goals, and examine individual and aggregate behaviors of the controllable devices. The asynchrony is simplified as follows: at the beginning of every minute, one-fifth of A/Cs and batteries update their setpoints while the rest maintain previous decisions. In practice, more asynchronous updates are expected, which usually leads to better and smoother results.

6.4.2.1 Dynamics of Individual Device

We first zoom in to examine individual controllable devices.

PV inverter: We select an arbitrary PV inverter and plot its power output in Fig. 6.3. Positive real power curtailment and negative reactive power injection can be observed in response to the signal that incentivizes negative power injection.

A/C: We select an arbitrary house from an arbitrary node and plot its inside temperature. As shown in Fig. 6.4, the room temperature is controlled within the acceptable range.

Battery: We select an arbitrary battery and plot its SOC along with its power setpoint in Fig. 6.5. We observe a SOC near 0.5 but deviating due to the response to incentive signal, which we will further illustrate

later.

6.4.2.2 Aggregated Behavior & Voltage Regulation

We zoom out to examine the aggregated behavior of hundreds of discrete devices.

As shown in Fig. 6.6, the temperatures of all 375 houses with A/Cs are controlled within the acceptable temperature range, with an average temperature around the preferred value of 75°F. Tighter temperature bounds can be achieved by finer control granularity and more frequent control.

Fig. 6.7 shows the SOCs of all 375 batteries, together with exemplifying incentive signals for real power at 3 of nodes (see equation (6.23)). One can observe that batteries are incentivized to charge more during the middle of the day with average SOC higher than the preferred one.

Lastly, we plot the voltage in Fig. 6.8. Due to coordinated efforts of all controllable devices, the voltage magnitude is brought within the upper bound, compared with uncontrolled one.

6.5 Conclusion

We have proposed a distributed stochastic dual algorithm for managing DERs with both continuous and discrete decision variables as well as device dynamics, and extended it to the practical realtime setting with time-varying operating conditions, asynchronous updates by devices, and feedback being leveraged to account for nonlinear power flows as well as reduce communication overhead. The resulting algorithm provides a general online stochastic optimization algorithm for coordinating networked DERs to meet operational and economic objectives and constraints. We characterize the convergence of the algorithm analytically and evaluate its performance numerically. Particularly, the convergence characterization of the proposed stochastic dual algorithm is the first of its kind in power system to our best knowledge.

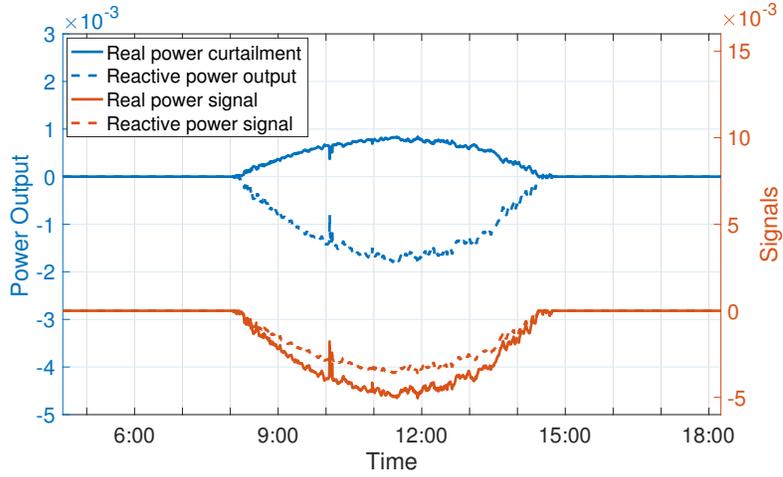


Figure 6.3: Power output of one arbitrary PV inverter.

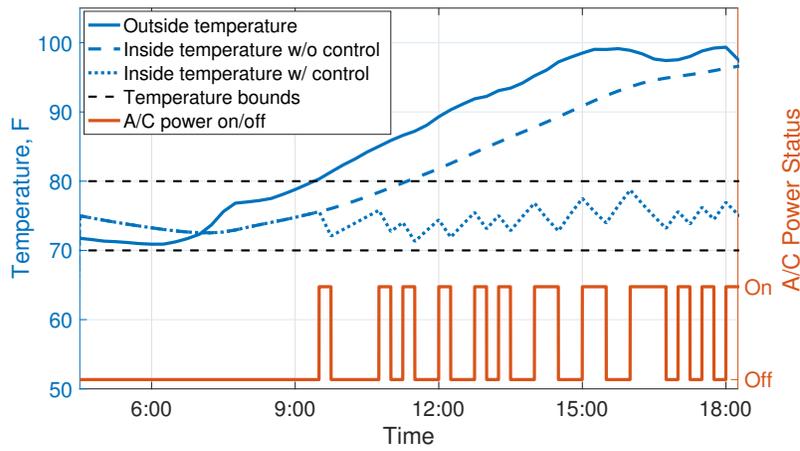


Figure 6.4: Temperature and power status of one arbitrary A/C.

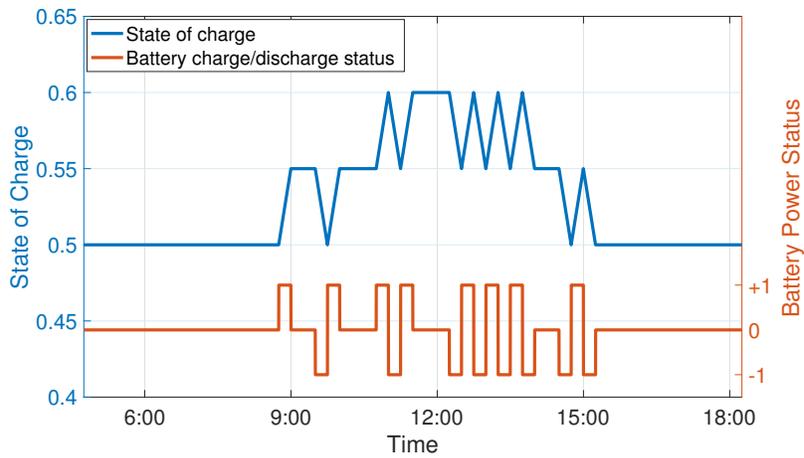


Figure 6.5: SOC and power status of one arbitrary battery.

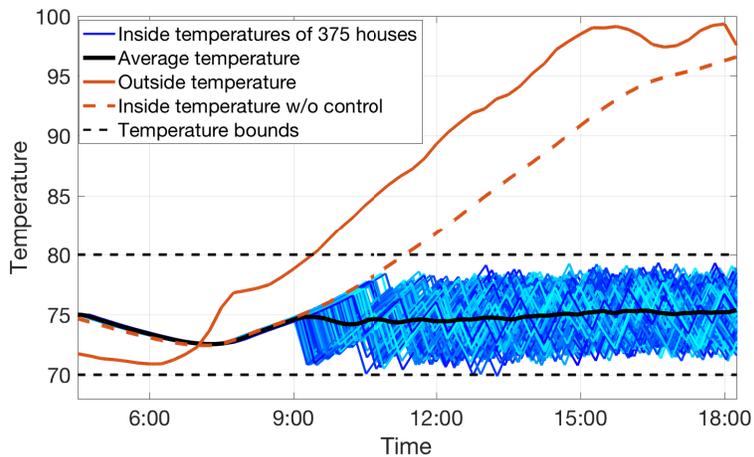


Figure 6.6: Temperatures of 375 households under control by A/Cs.

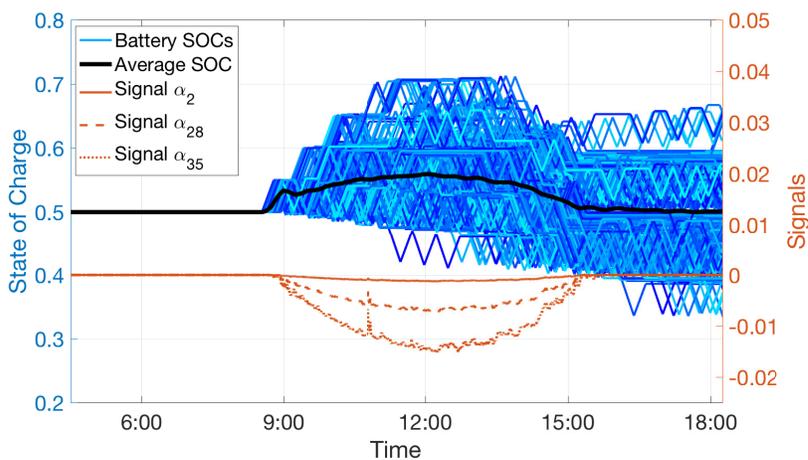


Figure 6.7: SOCs of 375 batteries under control and signals.

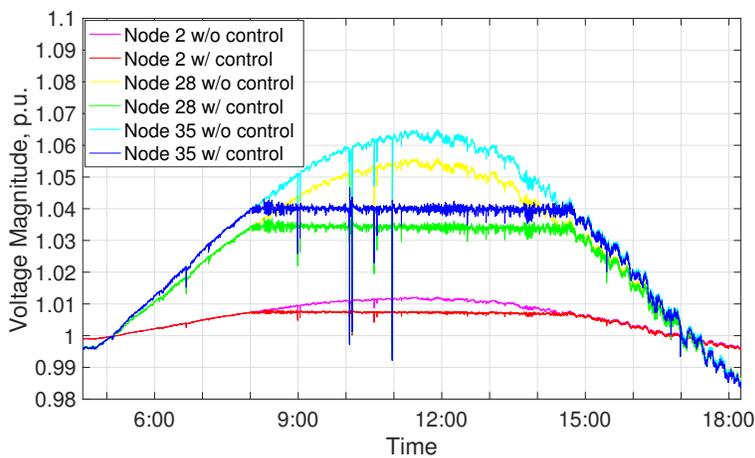


Figure 6.8: Uncontrolled and controlled voltages.

Chapter 7

Other Works

This chapter consists of some other works of mine, organized as follows.

Section 7.1 seeks contribution in the domain of reactive power compensation by establishing stability of local Volt/VAR controllers. In lieu of the approximate linear surrogate used in the existing work, the paper establishes existence and uniqueness of an equilibrium point using nonlinear AC power flow model. Key to this end is to consider a nonlinear dynamical system with non-incremental local Volt/VAR control, cast the Volt/VAR dynamics as a game, and leverage the fixed-point theorem as well as pertinent contraction mapping argument. Numerical examples are provided to complement the analytical results.

Section 7.2 takes a new approach to investigate synchronization in networks of coupled oscillators. We show that the coupled oscillator system when restricted to a proper region is a distributed partial primal-dual gradient algorithm for solving a well-defined convex optimization problem and its dual. We characterize conditions for synchronization solution of the KKT system of the optimization problem, based on which we derive conditions for synchronization equilibrium of the coupled oscillator network. This new approach reduces the hard problem of synchronization of coupled oscillators to a simple problem of verifying synchronization solution of a system of linear equations, and leads to a complete characterization of synchronization condition for the coupled oscillator network in an interesting and practically important region. Our synchronization condition is stated elegantly as the existence of solution for a system of linear equations, of which one best existing synchronization condition is a special sufficient condition case. In addition, we formulate a non-convex optimization problem with the force balance constraint for which the afore convex optimization problem is relaxation, and show that the coupled oscillator system is also a

distributed algorithm for solving this non-convex problem. This has interesting implication on exact convex relaxation, and confirms the insight that a physical system usually solves a convex problem even though it may have a non-convex representation.

Section 7.3 proposes a promising way to mitigate the wireless cellular capacity shortfall in the presence of ever-increasing wireless data demand. We formulate demand shaping as an optimization problem that minimizes the variation in aggregate traffic. We design a distributed and randomized offline demand shaping algorithm under complete traffic information and prove its almost surely convergence. We further consider a more realistic setting where the traffic information is incomplete but the future traffic can be predicted to a certain degree of accuracy. We design an online demand shaping algorithm that updates the schedules of deferrable applications (DAs) each time when new information is available, based on solving at each timeslot an optimization problem over a shrinking horizon from the current time to the end of the day. We compare the performance of the online algorithm against the optimal offline algorithm, and provide numerical examples to complement the theoretical analysis.

Related works have been published in [132, 133, 135, 142].

7.1 Local Volt/VAR Control with Nonlinear Power Flow Model: A Game-Theoretic Perspective

7.1.1 Introduction

Previously, we have studied the Volt/VAR control grounded on a linearized AC power flow model for mathematical tractability. With a linearized model, voltage values can be approximated as a linear function of reactive power injections. In contrast, this section seeks an analytical characterization of Volt/VAR control using exact nonlinear AC power flow models. To this end, the section utilizes a reverse-engineering approach to cast the nonlinear dynamical system with non-incremental Volt/VAR control as a game, where each node acts as a “self-interested player” who uses its local control function as a best-response strategy to minimize its own cost function. Using this approach, we show that the equilibrium of the Volt/VAR control dynamics is equivalent to the equilibrium of the resulting game. We further prove the existence and uniqueness of the equilibrium by leveraging the fixed-point theorem as well as contraction mapping argument.

7.1.2 Local Volt/VAR Control with Nonlinear Power Flow Model

The power flow equations (2.1) can be represented in the following compact form:

$$F(P, Q, \ell, v, q) = 0, \quad (7.1)$$

where F is twice continuously differentiable with respect to q and $y := (P, Q, \ell, v)$. For prevailing ambient conditions, given the reactive powers q , y is uniquely determined for distribution networks setups where $v_0 \approx 1$ and r_{ij}, x_{ij} are sufficiently small [35]. It has also been shown in [35] (Proposition 4-1) that, under the setup of $v_0 \approx 1$ and small r_{ij}, x_{ij} , the Jacobian matrix $\partial_y F(y, q)$ is nonsingular for a topology of one main feeder with its direct laterals. This result can be straightforwardly extended to general radial networks where laterals may have laterals and so on, because the corresponding expanded Jacobian matrix keeps the same crucial properties for both its diagonal and off-diagonal blocks. We therefore focus on the setup of $v_0 \approx 1$ and small r_{ij}, x_{ij} in this part. Then, by the implicit function theorem, it follows that equation (7.1) (i.e., equations (2.1)) defines implicitly a twice continuously differentiable function $y = y(q)$. Since Ω is

compact, $\partial_q v(q)$ is bounded uniformly on Ω , i.e., the first-order derivatives of P, Q, ℓ, v with respect to q are all bounded on Ω .

Specifically, we represent voltage magnitude as a function of q :

$$v = v(q), \quad (7.2)$$

with bounded $\frac{\partial v_i}{\partial q_j}$, $\forall i, j \in \mathcal{N}$. Define the reactance matrix $X = [X_{ij}]_{n \times n}$ with entries

$$X_{ij} := \sum_{(h,k) \in \mathcal{L}_i \cap \mathcal{L}_j} 2 \cdot x_{hk} > 0.$$

It is straightforward to check that X is symmetric. In the following, we will particularly relate the derivative of v with respect to q to the matrix X , which will be useful for the analysis in Section 7.1.3.2.

Lemma 7.1. *In a radial distribution system, for $\forall i, j \in \mathcal{N}$, we have*

$$\left| \frac{\partial v_i}{\partial q_j} \right| \leq (1 + \eta) X_{ij}, \quad (7.3)$$

for some $\eta > 0$.

Proof. In Chapter 2, we have shown that for the linearized and lossless power flow models (i.e., terms involving ℓ_{lk} , $\forall (l, k) \in \mathcal{L}$, are all set to zero), the approximated voltage value, denoted \hat{v} , satisfies:

$$\hat{v} = Xq + \tilde{v}, \quad (7.4)$$

where \tilde{v} is a constant determined by the system setup.

Considering the nonlinear model, we add up (4.1c) from any node $i \in \mathcal{N}$ all the way back to node 0 to obtain:

$$\begin{aligned} v_i &= v_0 + \sum_{(l,k) \in \mathcal{L}_i} \left(2(r_{lk}P_{lk} + x_{lk}Q_{lk}) - (r_{lk}^2 + x_{lk}^2)\ell_{lk} \right) \\ &= \hat{v}_i + 2 \sum_{(l,k) \in \mathcal{L}_i} \left(r_{lk} \sum_{(r,s) \in \tilde{\mathcal{L}}_k} r_{rs} \ell_{rs} \right) + 2 \sum_{(l,k) \in \mathcal{L}_i} \left(x_{lk} \sum_{(r,s) \in \tilde{\mathcal{L}}_k} x_{rs} \ell_{rs} \right) - \sum_{(l,k) \in \mathcal{L}_i} (r_{lk}^2 + x_{lk}^2) \ell_{lk}. \end{aligned}$$

Let $G_i(\ell)$ denote the three summation terms on the right-hand side of (7.5), and substitute \hat{v} with (7.4). Then equation (7.5) takes a simpler form of

$$v_i = \sum_{m \in \mathcal{N}} X_{im} q_m + \tilde{v}_i + G_i(\ell). \quad (7.5)$$

Notice that $G_i(\ell)$ is a function of single variable ℓ with an order of one. Take derivative of (7.5) on both sides with respect to q_j from any bus $j \in \mathcal{N}$, and we have

$$\frac{\partial v_i}{\partial q_j} = X_{ij} + \frac{\partial G_i(\ell)}{\partial q_j}, \quad (7.6)$$

and thus

$$\left| \frac{\partial v_i}{\partial q_j} \right| \leq X_{ij} + \left| \frac{\partial G_i(\ell)}{\partial q_j} \right|.$$

Since $\left| \frac{\partial \ell_{lk}}{\partial q_j} \right|$ is bounded and X_{ij} is nonzero,¹ there exists an $\eta_{ij} > 0$ such that

$$\left| \frac{\partial G_i(\ell)}{\partial q_j} \right| \leq \eta_{ij} \cdot X_{ij}.$$

Take $\eta = \max_{i,j \in \mathcal{N}} \eta_{ij}$, and the inequality (7.3) follows. ■

Remarks: An accurate characterization of η is challenging. In the numerical experiments, we have found that η is usually a small number. For example, $\eta \leq 0.2$ with the setup of the 42-bus distribution network used in Section 7.1.4. This is due to the fact that $\left| \frac{\partial \ell_{lk}}{\partial q_j} \right|$ is bounded, and that r_{lk}, x_{lk} are small.

Equation (7.2) together with the control function (2.12b) yields the following nonlinear dynamical system for Volt/VAR control:

$$v(t) = v(q(t)), \quad (7.7a)$$

$$q_i(t+1) = \left[f(v_i(t) - v_i^{\text{nom}}) \right]_{\Omega_i}, \quad (7.7b)$$

with locally measured $v(t)$ as the only control input, and $q(t)$ the only control variables.

Definition 7.1. A point (q^*, v^*) is called an equilibrium, if it is a fixed point of the dynamical system (7.7), i.e.,

$$v^* = v(q^*), \quad (7.8a)$$

$$q^* = \left[f(v^* - v^{\text{nom}}) \right]_{\Omega}. \quad (7.8b)$$

¹ Here we have assumed that, without loss of generality, the bus 0 has only one child node. Notice that, when the bus 0 has multiple child nodes, as the squared voltage magnitude v_0 is a constant, the branches of different child nodes are independent.

7.1.3 Voltage Control Game

It has been shown in [44] that, given voltage $v_i(t)$, the reactive power $q_i(t+1)$ in (7.7b) is the unique solution of the following optimization problem:

$$q_i(t+1) = \arg \min_{q_i \in \Omega_i} u_i(q_i; v_i(t)), \quad (7.9)$$

where

$$u_i(q_i; v_i) := C_i(q_i) + q_i v_i \quad (7.10)$$

with $C_i(q_i) := -\int_0^{q_i} f_i^{-1}(q) dq$ is a convex function since f_i^{-1} is non-increasing under Assumption 2.1. This result motivates us to cast the dynamics (7.7) as a game as shown next.

7.1.3.1 A Game-Theoretic Perspective

We view each node $i \in \mathcal{N}$ as a player with strategy space Ω_i and a cost function $u_i(q_i; v_i(q))$ defined by (7.10). Recalling that $\Omega := \prod_{i \in \mathcal{N}} \Omega_i$, the voltage control game is defined next.

Definition 7.2. A non-cooperative voltage control game is defined as a triple $\mathcal{G}_{vc} := \{\mathcal{N}, \Omega, (u_i(q_i; v_i(q)))_{i \in \mathcal{N}}\}$, where the strategic interaction among players is through the voltage $v_i(q)$, $i \in \mathcal{N}$.

In consistence with the introduction of the cost function u_i , we need to extend the concept of usual Nash equilibrium.

Definition 7.3. An equilibrium of the voltage control game \mathcal{G}_{vc} is a tuple $(q^*; v^*)$ such that $\forall i \in \mathcal{N}$, q_i^* is the best response to $v_i(q^*)$, i.e., for $\forall i \in \mathcal{N}$,

$$u_i(q_i^*; v_i^*) \leq u_i(q_i; v_i^*), \quad \forall q_i \in \Omega_i, \quad (7.11a)$$

$$v^* = v(q^*). \quad (7.11b)$$

In the above definition of equilibrium, the buses respond directly to given voltages. Similar to the price-taking behavior of the agents in a competitive market, we call such a behavior here *signal-taking*, i.e.,

when a bus makes decisions, it takes the voltage v_i as given but does not take into consideration the impact of its own decision upon the voltages.²

Recall that $v_i(q)$ is the implicit function from the implicit equation (7.17) (i.e., the power flow equations (2.1)), the following result is immediate.

Theorem 7.1. *The dynamical system (7.7) can be viewed as the best response algorithm for the voltage control game \mathcal{G}_{vc} . Moreover, a point (q^*, v^*) is an equilibrium of (7.7) if and only if (q^*, v^*) is an equilibrium of \mathcal{G}_{vc} .*

We further show the existence of the equilibrium of \mathcal{G}_{vc} .

Theorem 7.2. *Given the continuous Volt/VAR control functions $f(\cdot)$, there exists an equilibrium for the voltage control game \mathcal{G}_{vc} .*

Proof. Recalling from Section 7.1.2 that $v(q)$ is continuously differentiable, we know that the best response algorithm of the game \mathcal{G}_{vc} , i.e., the dynamical system (7.7), is a continuous mapping from Ω to itself. Since Ω is compact, by Brouwer's fixed-point theorem there exists an equilibrium for the game \mathcal{G}_{vc} . ■

7.1.3.2 Convergence of Dynamics and Uniqueness of Equilibrium

In this subsection, we establish a sufficient condition for convergence of the dynamical system (7.7) by leveraging the pertinent contraction mapping. Existence and uniqueness of its equilibrium will follow.

To this end, consider rewriting the control dynamics using the following mapping $g : \Omega \rightarrow \Omega$ as

$$q(t+1) = [f(v(q(t)))]_{\Omega} := g(q(t)).$$

Lemma 7.2. *If condition*

$$\alpha_i(1 + \eta) \sum_j X_{ij} < 1, \quad \forall i \in \mathcal{N} \tag{7.12}$$

holds, then the mapping g is a contraction mapping.

² When a bus takes into consideration the impact of its own decision upon the voltages, we say that this behavior is *signal-anticipating* [81]. With signal-anticipating buses, we can define the usual Nash equilibrium for the voltage control game, which we will investigate in another paper.

Proof. Define $\tilde{X} = [\tilde{X}_{ij}]_{n \times n}$ with its entries $\tilde{X}_{ij} = \frac{\partial v_i}{\partial q_j}$, $\forall i, j \in \mathcal{N}$. With induced matrix norm $\|\cdot\|_\infty$ as the maximum row sum, we have

$$\begin{aligned} \|\nabla_q g\|_\infty &\leq \max_i \left(|f'_i(v_i)| \sum_j |\tilde{X}_{ij}| \right) \\ &\leq \max_i \left(\alpha_i \sum_j |\tilde{X}_{ij}| \right) \\ &\leq \max_i \left(\alpha_i (1 + \eta) \sum_j X_{ij} \right) \\ &< 1, \end{aligned}$$

where the four inequalities respectively come from 1) the possibility of q 's being projected onto the boundary of Ω , making the corresponding derivative of g equal to zero (as well as a very small chance for \tilde{X}_{ij} to be negative), 2) assumption A2, 3) Lemma 7.1, and 4) condition (7.12). Hence, given $\forall q_x, q_y \in \Omega$, we have

$$\|g(q_x) - g(q_y)\|_\infty \leq \left\| \frac{\partial g}{\partial q} \right\|_\infty \cdot \|q_x - q_y\|_\infty < \|q_x - q_y\|_\infty,$$

i.e., g is a contraction mapping. ■

Then, using the contraction mapping theorem [25], the following result can be demonstrated.

Theorem 7.3. *Under the condition (7.12), the dynamics (7.7) converges to the unique equilibrium point.*

Remarks: Notice that when $\eta = 0$, i.e., when we ignore line loss, (7.12) coincides with the convergence condition for the same control strategy with linearized model for \mathcal{D}_1 in Section 2.3.3. While it is challenging to characterize η , as will be shown by numerical examples in Section 7.1.4.2, by setting $\eta = 0$, we usually still have a practical sufficient convergence condition for non-linear model, because (7.12) is a conservative condition, and that η itself is normally small.

7.1.4 Numerical Examples

We now provide numerical examples to complement the theoretical analysis in previous sections.

7.1.4.1 Simulation Setup

We use the same network setup of simulation in Chapter 3.

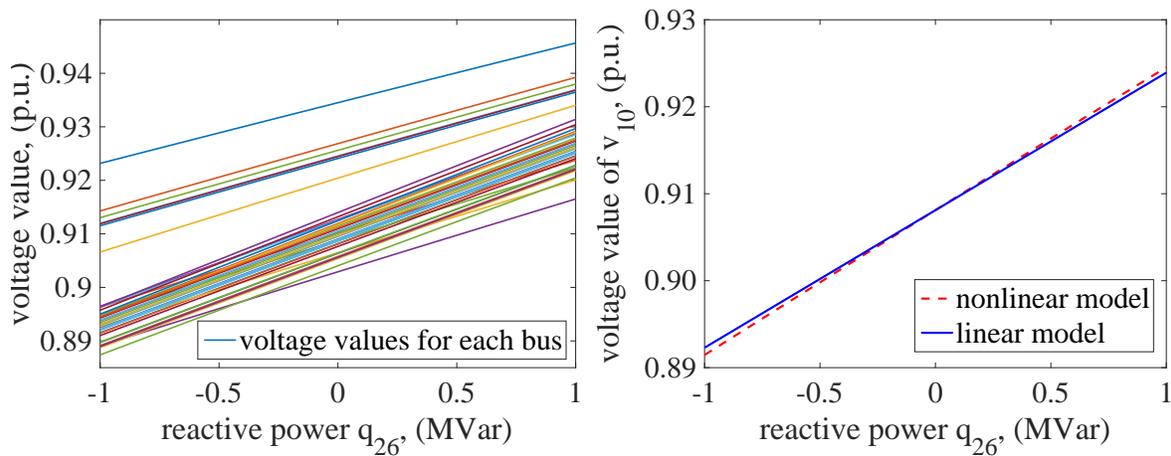


Figure 7.1: (Left) Voltage values of all buses as reactive power injection from Bus 26 changes; (right) voltage values of Bus 10 in both linear and non-linear models as reactive power injection from Bus 26 changes.

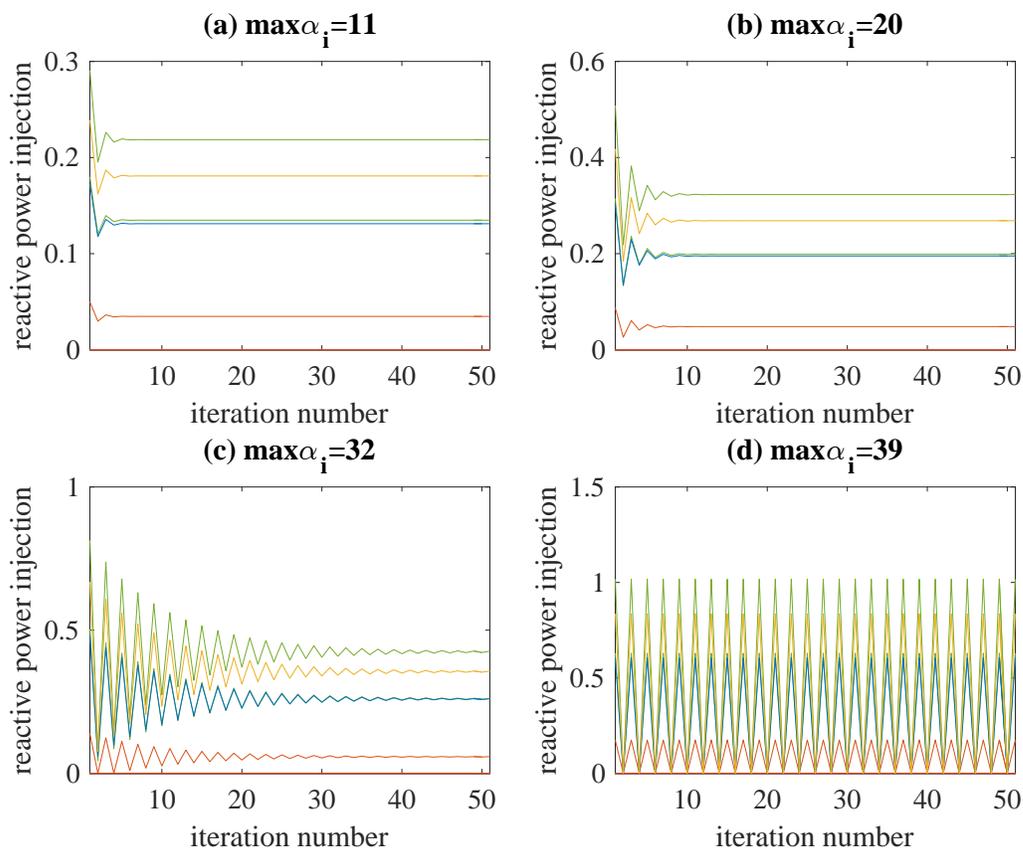


Figure 7.2: As the values of α_i increase from (a) to (d), we have slower convergence speed and finally reach a non-convergent result, when the values of α_i get too large.

We use the piecewise linear droop control functions (2.11) with their slopes α_i to be determined and analyzed. We assume that all the control functions have identical acceptable voltage range $[0.98^{p.u.}, 1.02^{p.u.}]$, i.e., $\delta_i = 0.04$ p.u., $\forall i \in \mathcal{N}$.

7.1.4.2 Effects of Reactive Power Injections upon Voltage Values

In this part, we examine how voltage values change with different reactive injections. We fix the reactive power injections of all inverters as 0 except that at Bus 26. We sweep the reactive power injections at Bus 26 from -1 MW to 1 MW with granularity of 0.1 MW, and record the consequent voltage changes at all buses. Similar results are observed by engaging any other inverters.

According to (7.6), we do not preclude the possibility of negative $\frac{\partial v_i}{\partial q_j}$, but it rarely takes place since the second term in (7.6) is usually much smaller than X_{ij} , almost always resulting in positive $\frac{\partial v_i}{\partial q_j}$, $\forall i, j \in \mathcal{N}$, as we can see from the illustration in Fig. 7.1 (left).

We then arbitrarily pick Bus 10 to compare its voltage changes against reactive power injections in both nonlinear and linear models. As illustrated in Fig. 7.1 (right), the slopes exhibiting $\frac{\partial v_{10}}{\partial q_{26}}$ in two models are very close. We can find a small parameter η , such that the slope in nonlinear model is upper-bounded by that in linear model multiplied by $(1 + \eta)$. In this case, η can be set as 0.2. Similar results are observed from voltage values of any other buses and reactive power injections from any other inverters, though the slope in nonlinear model is not necessarily greater than that in linearized model due to the possibility of $\frac{\partial \ell_{ik}}{\partial q_j}$ being negative.

7.1.4.3 Convergence

In this part, we set different α_i values in the piecewise linear droop control functions to see how they affect the convergence of our Volt/VAR control dynamics (7.7). As illustrated in Fig. 7.2, when we increase simultaneously the values of slopes for all five droop control functions from Fig. 7.2(a) to Fig. 7.2(d), we observe decreasing convergence speeds, until convergence is no longer available with too large α_i , where oscillation occurs (Fig. 7.2(d)).

From simulation results, we also observe that, the maximum allowed slope value $\max_i \alpha_i \approx 33$, is much

larger than the $\alpha < 4.6$ upper bound calculated by convergence condition (7.12) with $\eta = 0$. With larger η , we get even smaller sufficient upper bound. This is because the sufficient condition (7.12) is conservative estimation.

7.1.5 Conclusion

In order to analytically characterize the equilibrium and convergence of the local Volt/VAR control dynamics with nonlinear power flow model, we reverse-engineer the dynamical system with non-incremental control as a voltage control game. We then establish the existence, uniqueness, and convergence of the equilibrium by the fixed-point theorem and pertinent contraction mapping argument. We also extend the results to the incremental Volt/VAR controls. Numerical examples are provided to complement the analytical results.

7.2 Reverse Engineering and Convex Relaxation in Networks of Coupled Oscillators

7.2.1 Introduction

The network of coupled oscillators and its synchronization is one of the most investigated network dynamical systems and behaviors. It has broad applications in various disciplines from biology and medicine to chemistry and physics and to engineering and social sciences; see, e.g., [27, 33, 40, 41, 53, 59, 68, 78, 87, 94, 113, 114, 118, 125, 126, 130]. Despite its broad applications, a complete or tight characterization of the condition for synchronization of coupled oscillators is mostly an open question.

In this part, we consider a general coupled oscillator model that is partly motivated by the frequency dynamics and control in power networks: some of the oscillators are subject to the second-order Newtonian dynamics while the others are subject to the first-order kinematic dynamics, and they are sinusoidally coupled; see, e.g., [33] and [41]. This coupled oscillator model and its various special cases have been studied extensively; see, e.g., the above cited references and particularly [41] for a brief review. In particular, [41] presents an elegant closed-form condition for synchronization that significantly improves upon the existing conditions and is provably exact for various interesting network topologies and parameters.

Motivated by our prior work on the reverse engineering of the frequency control in the power network ([33]), we take a new approach to investigate synchronization of coupled oscillators. Specifically, we show that the coupled oscillator system when restricted to a proper region is a distributed partial primal-dual gradient algorithm for solving a well-defined convex optimization problem and its dual. We characterize conditions for synchronization solution of the KKT system of the optimization problem, based on which we derive conditions for synchronization equilibrium of the coupled oscillator network. This new approach reduces the hard problem of synchronization of coupled oscillators to a simple problem of verifying synchronization solution of a system of linear equations, and leads to a complete characterization of synchronization condition for the coupled oscillator network in an interesting and practically important region. Our synchronization condition is stated elegantly as the existence of solution for a system of linear equations, of which one synchronization condition of [41] is a special sufficient condition case.

We then formulate a non-convex optimization problem with the force balance constraint for which the

above-mentioned convex optimization problem is a relaxation. We show that the coupled oscillator system is also a distributed algorithm for solving this non-convex problem. This has an interesting implication on exact convex relaxation: a non-convex problem may be solved through solving its convex relaxation using a carefully chosen algorithm. This kind of exact convex relaxation is a bit different from the conventional one where the optimum of the convex problem is always a feasible point of the original non-convex problem, and confirms the insight that a physical system usually solves a convex problem even though it may have a non-convex representation.

7.2.2 System Model

Consider a network modeled by a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with a set \mathcal{N} of nodes and a set \mathcal{E} of undirected links connecting the nodes. Each node $i \in \mathcal{N}$ denotes an oscillator with phase $\theta_i \in \mathbb{R}$ and frequency $\omega_i = \dot{\theta}_i \in \mathbb{R}$, and each link $(i, j) \in \mathcal{E}$ (or $l \in \mathcal{E}$)³ is associated with a weight or coupling constant $b_{ij} > 0$ (or $b_l > 0$). The node set is partitioned into two disjoint sets $\mathcal{N} = \mathcal{N}^1 \cup \mathcal{N}^2$. Consider the following coupled oscillator system:

$$M_i \dot{\omega}_i + F_i(\omega_i) = f_i - \sum_{\{j:(i,j) \in \mathcal{E}\}} b_{ij} \sin(\theta_i - \theta_j), \quad i \in \mathcal{N}^1, \quad (7.13)$$

$$F_i(\omega_i) = f_i - \sum_{\{j:(i,j) \in \mathcal{E}\}} b_{ij} \sin(\theta_i - \theta_j), \quad i \in \mathcal{N}^2, \quad (7.14)$$

where each oscillator $i \in \mathcal{N}^1$ follows the second-order Newtonian dynamics with an inertia constant $M_i > 0$ and each oscillator $i \in \mathcal{N}^2$ follows the first-order kinematic dynamics. Each oscillator $i \in \mathcal{N}$ is subject to a constant force of $f_i \in \mathbb{R}$ and a frequency-dependent damping of $F_i(\omega_i)$. The function $F_i(\cdot)$ is assumed to be Lipschitz continuous and strictly increasing.

The above coupled oscillator model (7.13)–(7.14) is partly motivated by the frequency dynamics and control in the power network, and a huge literature exists on the synchronization of this general system and its various special cases; see, e.g., [41] and [33] and references therein. For instance, for the frequency dynamics of the power network, the set \mathcal{N}^1 is the set mechanical generators and \mathcal{N}^2 the set of load buses; f_i

³ We use (i, j) and l interchangeably to denote a link in \mathcal{E} . Note that in this section (i, j) is an un-ordered pair, i.e., $(i, j) = (j, i)$. But from the next section on, $l \in \mathcal{E}$ is directed and $(i, j) \neq (j, i)$.

is the power inject or draw, $F_i(\omega_i) = D_i\omega_i$ with damping coefficient $D_i > 0$,⁴ and M_i the generator inertia; and $b_{ij} = \frac{v_i v_j}{x_{ij}}$ with v_i the voltage magnitude at bus i and x_{ij} the reactance of power line (i, j) ; see, e.g., [20] and [33].

We aim to characterize conditions under which the network of coupled oscillators has a synchronization equilibrium and its stability.

Definition 7.4. (*Synchronization equilibrium*) A synchronization equilibrium $(\omega, \theta = \{\theta_i; i \in \mathcal{N}\}, \theta^0 = \{\theta_i^0; i \in \mathcal{N}\})$ of the coupled oscillator system (7.13)–(7.14) is defined by the following relations:

$$\omega_i = \omega, \quad i \in \mathcal{N}, \quad (7.15a)$$

$$\theta_i(t) = \theta_i^0 + \omega t, \quad i \in \mathcal{N}, \quad (7.15b)$$

$$F_i(\omega) = f_i - \sum_{\{j:(i,j) \in \mathcal{E}\}} b_{ij} \sin(\theta_i - \theta_j), \quad i \in \mathcal{N}, \quad (7.15c)$$

where $\theta_i^0 \in [0, 2\pi)$, $i \in \mathcal{N}$.

Motivated by the application in the power network where a security constraint $|\theta_i - \theta_j| < \frac{\pi}{2}$, $(i, j) \in \mathcal{E}$ is usually imposed (see, e.g., [20, 21], and [42]), we are particularly interested in the synchronization equilibrium with $|\theta_i^0 - \theta_j^0| < \frac{\pi}{2}$, $(i, j) \in \mathcal{E}$.

Definition 7.5. (*Phase Cohesiveness* ([41])) Given $\gamma \in [0, \frac{\pi}{2})$, a synchronization equilibrium $(\omega, \theta, \theta^0)$ is γ phase cohesive if $|\theta_i^0 - \theta_j^0| \leq \gamma$, $(i, j) \in \mathcal{E}$.

7.2.2.1 Reverse Engineering of Network Dynamics with Linearized Coupling

Assume that the system is initially at a synchronization equilibrium with a “nominal” frequency ω^n and phases θ_i^n , $i \in \mathcal{N}$ such that $|(\theta_i - \theta_j) - (\theta_i^n - \theta_j^n)| \ll 1$, $(i, j) \in \mathcal{E}$. Let $\tilde{b}_{ij} = b_{ij} \cos(\theta_i^n - \theta_j^n)$, and consider

⁴ Note that this damping term can result from frequency-sensitive load or frequency-based load or generation control. We can include more than one of such terms at each node as in [33], which will not change the structure of the problem and the results of this section.

the following system with linearized coupling between oscillators:

$$M_i \dot{\omega}_i + F_i(\omega_i) = f_i - \sum_{\{j:(i,j) \in \mathcal{E}\}} p_{ij}, \quad i \in \mathcal{N}^1, \quad (7.16a)$$

$$F_i(\omega_i) = f_i - \sum_{\{j:(i,j) \in \mathcal{E}\}} p_{ij}, \quad i \in \mathcal{N}^2, \quad (7.16b)$$

$$\dot{p}_{ij} = \tilde{b}_{ij}(\omega_i - \omega_j), \quad (i, j) \in \mathcal{E}. \quad (7.16c)$$

In the power network application, $b_{ij} \sin(\theta_i - \theta_j)$ is the nonlinear power flow from bus i to bus j , and the above linearization corresponds to the assumption of small phase angle deviation; see, e.g., [20].

Let $d_i = F_i(\omega_i)$, and $F_i^{-1}(d_i)$ is well-defined because of F_i being strictly monotone. As in [33], we introduce a cost function corresponding to each damping term:

$$C_i(d_i) = \int F_i^{-1}(d_i) dd_i, \quad i \in \mathcal{N}, \quad (7.17)$$

which is a strictly convex function by the assumption on the function F_i , and a convex optimization problem:

$$\min_{\mathbf{d}, \mathbf{p}} \sum_{i \in \mathcal{N}} C_i(d_i) \quad (7.18a)$$

$$\text{s.t.} \quad f_i = d_i + \sum_{\{j:(i,j) \in \mathcal{E}\}} p_{ij}, \quad i \in \mathcal{N}, \quad (7.18b)$$

where $\mathbf{d} = \{d_i; i \in \mathcal{N}\}$ and $\mathbf{p} = \{p_{ij}; (i, j) \in \mathcal{E}\}$. The cost function $C_i(d_i)$ and problem (7.18a)–(7.18b) can have different interpretations, depending on specific applications. For instance, in the power network, $d_i = F_i(\omega_i)$ can be the primary frequency control and $C_i(d_i)$ is then the cost associated with the generation control, and problem (7.18a)–(7.18b) is a DC optimal power flow problem ([33]). Notice that there may be “operational” constraints on d_i . For instance, in the power network, there is a limited capacity for generation. These operational constraints can be incorporated implicitly through carefully defining the domain of function F_i or explicitly through adding to the optimization problem (7.18a)–(7.18b).

It has been shown that the system dynamics (7.16a)–(7.16c) can be seen as a distributed algorithm for solving the problem (7.18a)–(7.18b) and its dual; see, e.g., [33]:

Theorem 7.4. *(Theorem 1 in [33] tailored to system (7.16a)–(7.16c)) The set of saddle points of the Lagrangian of problem (7.18a)–(7.18b) is the set of synchronization equilibria of dynamical system (7.16a)–*

(7.16c). Moreover, the dynamics (7.16a)–(7.16c) is a partial primal-dual gradient algorithm for solving the problem (7.18a)–(7.18b) and its dual.

We have applied the above reverse engineering result to guide the design of new frequency control algorithms for the power system to not only recover nominal frequency but also achieve economic efficiency; see, e.g., [33] and [78]. However, the above linearized model and reverse engineering result applies to the system with small phase angle derivation from an initial synchronization equilibrium, which is limited in applicability. An important question is if the above reverse engineering result can extend to the coupled oscillator system (7.13)–(7.14) with nonlinear coupling. In the next sections, we give a positive answer to this question, and use it to characterize the condition for synchronization in the network of coupled oscillators, as well as discuss its implication on convex relaxation.

7.2.3 Reverse Engineering and Synchronization

We first introduce a few notations to simplify the presentation of the system and its analysis. Assigning an arbitrary direction to each link $l \in \mathcal{E}$, we define a $|\mathcal{N}| \times |\mathcal{E}|$ incidence matrix A with entry

$$A_{li} = \begin{cases} 1, & \text{if node } i \text{ is the source node of link } l \\ -1, & \text{if node } i \text{ is the sink node of link } l \\ 0, & \text{otherwise} \end{cases}$$

Since \mathcal{G} is connected, we have $\text{Rank}(A) = |\mathcal{N}| - 1$ and $\text{Ker}(A^\top) = \text{span}(\mathbf{1}_{|\mathcal{N}|})$; see, e.g., [26]. With the incident matrix, we can rewrite the problem (7.18a)–(7.18b) as:

$$\min_{\mathbf{d}, \mathbf{p}} \sum_{i \in \mathcal{N}} C_i(d_i) \tag{7.19a}$$

$$\text{s.t. } \mathbf{f} = \mathbf{d} + A\mathbf{p}, \tag{7.19b}$$

where $\mathbf{f} = \{f_i; i \in \mathcal{N}\}$. Notice that the coupling $b_{ij} \sin(\theta_i - \theta_j)$ between nodes i and j is bounded by $\pm b_{ij}$.

This implies that an additional constraint should be imposed on the problem (7.19a)–(7.19b):

$$-\mathbf{b} \leq \mathbf{p} \leq \mathbf{b}, \tag{7.20}$$

where $\mathbf{b} = \{b_l; l \in \mathcal{E}\}$. The problem (7.19a)–(7.20) is convex, so all its optima are global optima. We further assume that the problem (7.19)–(7.20) is strictly feasible.

Lemma 7.3. *The problem (7.19)–(7.20) may have multiple (global) optima in \mathbf{p} , but has an unique optimum in \mathbf{d} .*

Proof. The objective function is not strictly convex in all the decision variable, so the problem (7.19)–(7.20) may have multiple global optima. Suppose that \mathbf{d} can take two values $\bar{\mathbf{d}}$ and $\hat{\mathbf{d}}$ at optima. Since the objective function is strictly convex in \mathbf{d} , we have $\sum_{i \in \mathcal{N}} C_i(d_i) < \alpha \sum_{i \in \mathcal{N}} C_i(\bar{d}_i) + (1 - \alpha) \sum_{i \in \mathcal{N}} C_i(\hat{d}_i)$ for any $\mathbf{d} = \alpha \bar{\mathbf{d}} + (1 - \alpha) \hat{\mathbf{d}}$, $0 < \alpha < 1$. This contradicts the fact that $\bar{\mathbf{d}}$ and $\hat{\mathbf{d}}$ are optima. So, the problem (7.19)–(7.20) has an unique optimum in \mathbf{d} . ■

7.2.3.1 Synchronization Solution of the KKT System

Introduce Lagrangian multiplier λ_i for each constraint in (7.19b), and write down the KKT condition of the problem (7.19)–(7.20) (see, e.g., [31]):

$$f_i = d_i + \sum_{l \in \mathcal{E}} A_{il} p_l, \quad i \in \mathcal{N}, \quad (7.21a)$$

$$d_i = F_i(\lambda_i), \quad i \in \mathcal{N}, \quad (7.21b)$$

$$p_l = b_l, \quad \text{if } \lambda_{s_l} > \lambda_{d_l}, \quad l \in \mathcal{E}, \quad (7.21c)$$

$$p_l = -b_l, \quad \text{if } \lambda_{s_l} < \lambda_{d_l}, \quad l \in \mathcal{E}, \quad (7.21d)$$

$$p_l \in [-b_l, b_l], \quad \text{if } \lambda_{s_l} = \lambda_{d_l}, \quad l \in \mathcal{E}, \quad (7.21e)$$

where s_l and d_l denote the source and sink nodes of link l respectively. For the reason that will become clear later, we focus on those “synchronization” solutions to the above KKT system.

Definition 7.6. *A solution to the KKT system (7.21a)–(7.21e) is said to be a synchronization solution if $\lambda_i = \lambda$ for all $i \in \mathcal{N}$.*

It is obvious that, if there is a synchronization solution, λ is uniquely determined by $\sum_{i \in \mathcal{N}} f_i =$

$\sum_{i \in \mathcal{N}} F_i(\lambda)$. Let $\bar{f}_i = f_i - F_i(\lambda)$. Then $\sum_{i \in \mathcal{N}} \bar{f}_i = 0$, and at a synchronization solution

$$\bar{\mathbf{f}} = A\mathbf{p}, \quad (7.22)$$

$$-\mathbf{b} \leq \mathbf{p} \leq \mathbf{b}, \quad (7.23)$$

where $\bar{\mathbf{f}} = \{\bar{f}_i; i \in \mathcal{N}\}$. From equation (7.22) we have

$$\mathbf{p} = A^T(AA^T)^\dagger \bar{\mathbf{f}} + \bar{\mathbf{p}}, \quad (7.24)$$

$$\bar{\mathbf{p}} \in \text{Ker}(A), \quad (7.25)$$

where ‘ \dagger ’ denotes Moore-Penrose pseudo inverse, and AA^T and its pseudo inverse satisfies $AA^T(AA^T)^\dagger = (AA^T)^\dagger AA^T = I_{|\mathcal{N}|} - \frac{1}{|\mathcal{N}|} \mathbf{1}_{|\mathcal{N}| \times |\mathcal{N}|}$. The space $\text{Ker}(A)$ is related to the cycles in the network; see, e.g., [26].

Theorem 7.5. *The following three statements are equivalent:*

- (1) *There exists at least one $\bar{\mathbf{p}} \in \text{Ker}(A)$ such that $-\mathbf{b} \leq A^T(AA^T)^\dagger \bar{\mathbf{f}} + \bar{\mathbf{p}} \leq \mathbf{b}$.*
- (2) *The KKT system (7.21a)–(7.21e) has a synchronization solution.*
- (3) *All the solutions of the KKT system (7.21a)–(7.21e) are synchronization solutions.*

Proof. The equivalence of statements (1) and (2) is already shown in the above. The statement (3) obviously implies statement (2). Now, suppose that (2) holds but (3) does not, i.e., there exists another solution that is not a synchronization solution. Thus, the problem (7.19)–(7.20) has two different solutions in terms of \mathbf{d} by equation (7.21b), which contradicts Lemma 7.3. So, (2) implies (3), and thus statements (2) and (3) are equivalent. ■

The statement (1) of Theorem 7.5 gives a sufficient and necessary condition for the synchronization solution of the KKT system (7.21a)–(7.21e). To verify this condition is a linear programming (LP) problem, for which efficient algorithms exist; see, e.g., [31].

Corollary 7.1. *A sufficient condition for the existence of the synchronization solution of the KKT system (7.21a)–(7.21e) is given by*

$$\|(\text{diag}(\mathbf{b}))^{-1} A^T(AA^T)^\dagger \bar{\mathbf{f}}\|_\infty \leq 1, \quad (7.26)$$

where $\|\cdot\|_\infty$ denotes infinity norm.

Proof. Let $\bar{\mathbf{p}} = \mathbf{0}$. If condition (7.26) holds, then $-\mathbf{b} \leq \mathbf{p} = A^\top(AA^\top)^\dagger \bar{\mathbf{f}} \leq \mathbf{b}$. So, \mathbf{p} is a synchronization solution to the KKT system (7.21a)–(7.21e). ■

The condition (7.26) is easier to verify. It is also necessary for synchronization solution in certain networks with special structure, e.g., for the tree network where $\text{Ker}(A) = \{\mathbf{0}\}$. An interesting question is how tight the condition (7.26) is for general networks, compared with the sufficient and necessary condition in the statement (1) of Theorem 7.5. We will investigate related issues in future work.

7.2.3.2 Primal-Dual Gradient Algorithm

Let $\boldsymbol{\lambda} = \{\lambda_i; i \in \mathcal{N}\}$, and consider the Lagrangian for the problem (7.19)–(7.20):

$$L(\mathbf{d}, \mathbf{p}; \boldsymbol{\lambda}) = \sum_{i \in \mathcal{N}} C_i(d_i) + \boldsymbol{\lambda}^\top (\mathbf{f} - \mathbf{d} - A\mathbf{p}). \quad (7.27)$$

A saddle point of L is a primal-dual optimum of the problem (7.19)–(7.20) and its dual (see, e.g., [31]), and moreover, the saddle point is unique in $\boldsymbol{\lambda}$ by Lemma 7.3 and the strict monotonicity of the functions F_i .

Define a reduced Lagrangian:

$$\bar{L}(\mathbf{p}; \boldsymbol{\lambda}^1) = \max_{\boldsymbol{\lambda}^2} \min_{\mathbf{d}} L(\mathbf{d}, \mathbf{p}; \boldsymbol{\lambda}), \quad (7.28)$$

where $\boldsymbol{\lambda}^1 = \{\lambda_i; i \in \mathcal{N}^1\}$ and $\boldsymbol{\lambda}^2 = \{\lambda_i; i \in \mathcal{N}^2\}$. From the inner minimization in (7.28) we have

$$d_i = F_i(\lambda_i), \quad i \in \mathcal{N}. \quad (7.29)$$

The function $\min_{\mathbf{d}} L(\mathbf{d}, \mathbf{p}; \boldsymbol{\lambda})$ is strictly concave and continuously differentiable in $\boldsymbol{\lambda}$ by the assumption on the functions F_i . From the outer maximization we have

$$F_i(\lambda_i) = f_i - \sum_{l \in \mathcal{E}} A_{il} p_l, \quad i \in \mathcal{N}^2. \quad (7.30)$$

Since $\min_{\mathbf{d}} L(\mathbf{d}, \mathbf{p}; \boldsymbol{\lambda})$ is strictly concave in $\boldsymbol{\lambda}$, the reduced Lagrangian \bar{L} is strictly concave in $\boldsymbol{\lambda}^1$.

Applying the continuous-time primal-dual gradient algorithm (aka, saddle point dynamics) to the reduced Lagrangian, we have

$$\dot{p}_l = -\epsilon_l \frac{\partial \bar{L}}{\partial p_l} = \sqrt{b_l^2 - p_l^2} (\lambda_{s_l} - \lambda_{d_l}), l \in \mathcal{E}, \quad (7.31a)$$

$$\dot{\lambda}_i = \gamma_i \frac{\partial \bar{L}}{\partial \lambda_i} = \frac{1}{M_i} (f_i - F_i(\lambda_i) - \sum_{l \in \mathcal{E}} A_{il} p_l), i \in \mathcal{N}^1, \quad (7.31b)$$

where we have chosen specific scaling factors $\epsilon_l = \sqrt{b_l^2 - p_l^2}$ and $\gamma_i = \frac{1}{M_i}$. Notice that in equation (7.31a) the choice of the scaling factor ensures that the constraint (7.20) is satisfied. As $d p_l / \sqrt{b_l^2 - p_l^2} = d \arcsin(p_l/b_l)$, if we identify λ_i with ω_i , the algorithm (7.29)–(7.31) is equivalent to the dynamical system (7.13)–(7.14) with the phases being restricted to $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$. We thus have the following result.

Theorem 7.6. *If identifying λ_i with ω_i for all $i \in \mathcal{N}$, the network dynamics (7.13)–(7.14) in the region defined by $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$ is a distributed partial primal-dual gradient algorithm for solving the following problem and its dual:*

$$\min_{\mathbf{d}, \mathbf{p}} \sum_{i \in \mathcal{N}} C_i(d_i) \quad (7.32a)$$

$$s.t. \quad \mathbf{f} = \mathbf{d} + \mathbf{A}\mathbf{p}, \quad (7.32b)$$

$$-\mathbf{b} < \mathbf{p} < \mathbf{b}. \quad (7.32c)$$

Moreover, the set of synchronization equilibria of the dynamical system (7.13)–(7.14) in the region defined by $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$ is a subset of the set of saddle points of the Lagrangian L .

We will study the synchronization equilibrium and its stability of the network of coupled oscillators (7.13)–(7.14) from the perspective that it is a primal-dual gradient algorithm for solving the problem (7.32a)–(7.32c) and its dual, i.e., we will study the network dynamics (7.13)–(7.14) through studying the algorithm (7.29)–(7.31b). For this purpose, in the rest of this section we will assume that there exists at least one $\bar{\mathbf{p}} \in \text{Ker}(A)$ such that $-\mathbf{b} < A^\top(AA^\top)^\dagger \bar{\mathbf{f}} + \bar{\mathbf{p}} < \mathbf{b}$, under which all the primal-dual optima of the problem (7.32a)–(7.32c) and its dual are synchronization solutions by Theorem 7.5. We will also use ω_i and λ_i interchangeably from now on.

7.2.3.3 Synchronization Equilibrium and Its Stability

We first study the convergence of the primal-dual gradient algorithm (7.29)–(7.31b) in the region defined by $-\mathbf{b} < \mathbf{p} < \mathbf{b}$.

Theorem 7.7. *The primal-dual gradient algorithm (7.29)–(7.31) converges locally to a primal-dual optimum of the problem (7.32a)–(7.32c) and its dual.*

Proof. Let $(\mathbf{p}^*, \boldsymbol{\lambda}^*)$ be a primal-dual optimum⁵ of the problem (7.32a)–(7.32c) and its dual. Consider the Lyapunov function:

$$U(\mathbf{p}, \boldsymbol{\lambda}^1; \mathbf{p}^*, \boldsymbol{\lambda}^{1*}) = \sum_{l \in \mathcal{E}} \int_{p_l^*}^{p_l} \frac{q_l - p_l^*}{\sqrt{b_l^2 - q_l^2}} dq_l + \sum_{i \in \mathcal{N}^1} \frac{M_i}{2} (\lambda_i - \lambda_i^*)^2, \quad (7.33)$$

which is strictly convex if $-\mathbf{b} < \mathbf{p} < \mathbf{b}$. Consider its Lie-derivative under the algorithm (7.29)–(7.31):

$$\begin{aligned} \dot{U}(\mathbf{p}, \boldsymbol{\lambda}^1; \mathbf{p}^*, \boldsymbol{\lambda}^{1*}) &= -(\mathbf{p} - \mathbf{p}^*)^\top \nabla_{\mathbf{p}} \bar{L} + (\boldsymbol{\lambda}^1 - \boldsymbol{\lambda}^{1*})^\top \nabla_{\boldsymbol{\lambda}^1} \bar{L} \\ &\leq \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^1) - \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^1) + \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^1) - \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^{1*}) \end{aligned} \quad (7.34)$$

$$\begin{aligned} &= \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^1) - \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^{1*}) \\ &= \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^1) - \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^{1*}) + \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^{1*}) - \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^{1*}) \\ &\leq 0, \end{aligned} \quad (7.35)$$

where inequality (7.34) follows from the fact that \bar{L} is convex in \mathbf{p} and concave in $\boldsymbol{\lambda}^1$, and inequality (7.35) from the fact that $(\mathbf{p}^*; \boldsymbol{\lambda}^{1*})$ is a saddle point of \bar{L} . Notice that if $\dot{U}(\mathbf{p}, \boldsymbol{\lambda}^1; \mathbf{p}^*, \boldsymbol{\lambda}^{1*}) = 0$, then all the inequalities become equality, and $\bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^1) = \bar{L}(\mathbf{p}^*; \boldsymbol{\lambda}^{1*})$ and $\bar{L}(\mathbf{p}; \boldsymbol{\lambda}^{1*}) = \bar{L}(\mathbf{p}; \boldsymbol{\lambda}^{1*})$. From LaSalle's invariance principle ([67]), the trajectory of the algorithm (7.29)–(7.31b) will be eventually contained in a compact subset of the invariant set

$$\mathcal{I} = \{(\mathbf{p}, \boldsymbol{\lambda}) : \dot{U}(\mathbf{p}, \boldsymbol{\lambda}^1; \mathbf{p}^*, \boldsymbol{\lambda}^{1*}) = 0\}. \quad (7.36)$$

Since $\bar{L}(\mathbf{p}; \boldsymbol{\lambda}^1)$ is strictly concave in $\boldsymbol{\lambda}^1$, by Proposition 11 in [33] the invariant set \mathcal{I} is a subset of the primal-dual optima of the problem (7.32a)–(7.32c) and its dual, and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ for all $(\mathbf{p}, \boldsymbol{\lambda}) \in \mathcal{I}$. When the network is a

⁵ Notice that $\boldsymbol{\lambda}^* = \boldsymbol{\lambda} \mathbf{1}_{|\mathcal{N}^1|}$.

tree, the set \mathcal{I} is a singleton, and obviously the algorithm (7.29)–(7.31) converges to the unique primal-dual optimum of the problem (7.32a)–(7.32c) and its dual. In general, for any networks, since the algorithm converges to the compact set \mathcal{I} as $t \rightarrow \infty$, there exists a convergence subsequence $\{(\mathbf{p}(t_k), \boldsymbol{\lambda}(t_k))\}_{k=1,2,\dots}$ with $0 \leq t_1 < t_2 < \dots$ and $\lim_{k \rightarrow \infty} t_k \rightarrow \infty$, such that $\lim_{k \rightarrow \infty} \mathbf{p}(t_k) = \mathbf{p}^\infty$ and $\lim_{k \rightarrow \infty} \boldsymbol{\lambda}(t_k) = \boldsymbol{\lambda}^*$ for some $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*) \in \mathcal{I}$. Since the Lyapunov function can be defined in terms of any primal-dual optimum, we choose the Lyapunov function to be $U(\mathbf{p}, \boldsymbol{\lambda}^1; \mathbf{p}^\infty, \boldsymbol{\lambda}^{1*})$. Notice that $U \geq 0$ with $U = 0$ only if $\mathbf{p} = \mathbf{p}^\infty$, and $\dot{U} \leq 0$ along the trajectory $(\mathbf{p}(t), \boldsymbol{\lambda}(t))$ of the algorithm (7.29)–(7.31). By the continuity of U , we have

$$\begin{aligned} & \lim_{t \rightarrow \infty} U(\mathbf{p}(t), \boldsymbol{\lambda}^1(t); \mathbf{p}^\infty, \boldsymbol{\lambda}^{1*}) \\ &= \lim_{k \rightarrow \infty} U(\mathbf{p}(t_k), \boldsymbol{\lambda}^1(t_k); \mathbf{p}^\infty, \boldsymbol{\lambda}^{1*}) \\ &= U(\mathbf{p}^\infty, \boldsymbol{\lambda}^{1*}; \mathbf{p}^\infty, \boldsymbol{\lambda}^{1*}) = 0. \end{aligned}$$

This implies that $(\mathbf{p}(t), \boldsymbol{\lambda}(t))$ converges to $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$, which is a primal-dual optimum of the problem (7.32a)–(7.32c) and its dual. ■

Theorem 7.7 does not implies *global* convergence of the algorithm (7.29)–(7.31), as its proof requires that the trajectory of the algorithm is contained in the region defined by $-\mathbf{b} < \mathbf{p} < \mathbf{b}$. Moreover, the convergence is *trajectory-wise* and does not necessarily imply the local stability of the primal-dual optimum $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$. We will however show that the convergence point $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$ is unique, i.e., independent of the specific trajectories, and is indeed locally stable.

Theorem 7.8. *The primal-dual gradient algorithm (7.29)–(7.31) converges to a unique and locally stable primal-dual optimum of the problem (7.32a)–(7.32c) and its dual.*

Proof. By Theorem 7.6 or equation (7.31a), at a convergence point $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$, there exist phases $\boldsymbol{\theta}$ with $|\theta_{s_l} - \theta_{d_l}| < \pi/2$, $l \in \mathcal{E}$ such that

$$\bar{f}_i = \sum_{l \in \mathcal{E}} A_{il} p_l = \sum_{l \in \mathcal{E}} A_{il} b_l \sin(\theta_{s_l} - \theta_{d_l}), \quad i \in \mathcal{N}. \quad (7.37)$$

Notice that the above mapping from $\boldsymbol{\theta}$ to $\bar{\mathbf{f}}$ is one to one in the domain defined by $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$; see, e.g., [12]. So, \mathbf{p}^∞ and thus $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$ are uniquely determined, independent of specific trajectories of the algorithm (7.29)–(7.31b). This further implies that the convergence point $(\mathbf{p}^\infty, \boldsymbol{\lambda}^*)$ is locally stable. ■

Combining Theorems 7.6–7.8, we have the following result.

Theorem 7.9. *The following two statements are equivalent:*

- (1) *There exists at least one $\bar{\mathbf{p}} \in \text{Ker}(A)$ such that $-\mathbf{b} < A^\top (AA^\top)^\dagger \bar{\mathbf{f}} + \bar{\mathbf{p}} < \mathbf{b}$.*
- (2) *The network of coupled oscillators (7.13)–(7.14) has a unique and locally stable synchronization equilibrium with cohesive phases $|\theta_i^0 - \theta_j^0| < \pi/2$, $(i, j) \in \mathcal{E}$.*

Theorem 7.9 states the synchronization condition elegantly as the existence of solution for a system of linear equations, and is a complete characterization of the condition for synchronization of coupled oscillator network in the region defined by $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$.

Similar to Corollary 7.1, we have the following sufficient condition for synchronization equilibrium if choosing $\bar{\mathbf{p}} = 0$.

Corollary 7.2. *The network of coupled oscillators (7.13)–(7.14) has a unique and locally stable synchronization equilibrium with cohesive phases $|\theta_i^0 - \theta_j^0| < \pi/2$, $(i, j) \in \mathcal{E}$, if*

$$\|(\text{diag}(\mathbf{b}))^{-1} A^\top (AA^\top)^\dagger \bar{\mathbf{f}}\|_\infty < 1. \quad (7.38)$$

This sufficient condition is exactly one condition given in [41].

Notice that the synchronization equilibrium of the coupled oscillator network is locally stable. An important question is to characterize its region of attraction, which has important implication in applications to, e.g., the power network. We will explore the Lyapunov function (7.33) and its convexity to investigate this question in future work.

To recapture, the above conditions for synchronization of the coupled oscillator network (7.13)–(7.14) are carried over from the conditions for synchronization solution of the KKT system for the problem (7.19)–(7.20). We have reduced the hard problem of synchronization of coupled oscillators to a simple problem of verifying solution of a system of linear equations, by identifying the network system dynamics as a distributed partial primal-dual gradient algorithm for solving a well-defined convex optimization problem and its dual.

7.2.4 Implication on Convex Relaxation

Consider the following optimization problem:

$$\min_{\mathbf{d}, \boldsymbol{\theta}} \sum_{i \in \mathcal{N}} C_i(d_i) \quad (7.39a)$$

$$\text{s.t. } f_i = d_i + \sum_{l \in \mathcal{E}} A_{il} b_l \sin(\theta_{s_l} - \theta_{d_l}), \quad i \in \mathcal{N}, \quad (7.39b)$$

where the “physical” constraint (7.39b) captures the force balance that should hold at an equilibrium. The problem (7.39a)–(7.39b) looks a more natural problem to study than the problem (7.19)–(7.20), as it captures directly nonlinear coupling between the oscillators. In the power network application, for instance, the problem (7.39a)–(7.39b) corresponds to an optimal power flow problem with nonlinear branch flows, i.e., without assuming small phase deviation as in usual DC power flow approximation.

The problem (7.39a)–(7.39b) is nonconvex, even if the phases are constrained to $|\theta_{s_l} - \theta_{d_l}| \leq \pi/2$, $l \in \mathcal{E}$. Notice that the problem (7.19)–(7.20) is a convex relaxation of the problem (7.39a)–(7.39b), and at its optimum the constraint (7.39b) is satisfied when solved using the algorithm (7.29)–(7.31b).

Theorem 7.10. *The network dynamics (7.13)–(7.14) in the region defined by $|\theta_i - \theta_j| < \pi/2$, $(i, j) \in \mathcal{E}$ is a distributed algorithm for solving the problem (7.39a)–(7.39b).*

Notice that an optimum of the problem (7.19)–(7.20) may not be an optimum of the problem (7.39a)–(7.39b). Theorem 7.10 thus has an interesting implication: a non-convex problem may be solved through solving its convex relaxation using a *carefully chosen algorithm*. This kind of exact convex relaxation is a bit different from the “conventional” exact relaxation where the optimum of the convex problem is always a feasible point of the original non-convex problem. Physically, this confirms an insight that a physical system usually solves a convex problem (e.g., the problem (7.19)–(7.20)) even though it may have a non-convex representation (e.g., the problem (7.39a)–(7.39b)). Even though the above implication on exact convex relaxation is based on the result when the phases are restricted to $|\theta_{s_l} - \theta_{d_l}| < \pi/2$, $l \in \mathcal{E}$, we expect that it holds generally and will further investigate it in future work.

7.2.5 Conclusion

We have taken a new approach to investigate synchronization in the coupled oscillator network, by identifying the network system dynamics as a distributed primal-dual gradient algorithm for solving a well-defined convex optimization problem and its dual. This new approach reduces the hard problem of synchronization of coupled oscillators to a simple problem of verifying synchronization solution of a system of linear equations, and leads to a complete characterization of synchronization condition for the coupled oscillator network in an interesting and practically important region. Our synchronization condition is stated elegantly as the existence of solution for a system of linear equations, of which one best existing synchronization condition is a special case of sufficient condition. We have also formulated a non-convex optimization problem with the force balance constraints for which the afore convex optimization problem is relaxation, and showed that the coupled oscillator system is also a distributed algorithm for solving this non-convex problem. This has interesting implication on exact convex relaxation, and confirms the insight that a physical system usually solves a convex problem even though it may have a non-convex representation.

7.3 Demand Shaping in Cellular Networks

Main Notation for Section 7.3

t	time index, $t \in \mathcal{T} := \{1, \dots, T\}$
n	DA index, $n \in \mathcal{N} := \{1, \dots, N\}$
\mathcal{N}'	set of N' continuous DAs
\mathcal{N}''	set of $N'' = N - N'$ discrete DAs
$\hat{\mathcal{N}}_t''$	set of discrete DAs started earlier
$\tilde{\mathcal{N}}_t$	set of DAs adjustable at time t
b	base traffic profile, $b = \{b(t); t \in \mathcal{T}\}$
p_n	data rate profile of DA n , $p_n = \{p_n(t); t \in \mathcal{T}\}$
$\bar{p}_n(t)$	upper bounds of DA n on the data rate at time t
r_n	constant bit rate for DA $n \in \mathcal{N}''$
l_n	number of timeslots to finish transmission for DA $n \in \mathcal{N}''$
q	virtual deferrable traffic profile
d	average traffic profile
\hat{d}	average traffic profile of online ODS
\hat{d}^*	average traffic profile of online relaxed ODS
d^*	average traffic profile of offline relaxed ODS
P_n	total traffic required from DA n , $P_n = \sum_{t \in \mathcal{T}} p_n(t)$
$P_n(t)$	remaining traffic to be served for DA $n \in \mathcal{N}'_t$
x_n^k	change in traffic profile of DA n , $x_n^k = p_n^{k+1} - p_n^k$
t_n^a	arrival time of DA n
t_n^d	deadline of DA n
A_n	number of feasible profiles of DA $n \in \mathcal{N}''$
$f_{n,a}$	a -th feasible profile of DA $n \in \mathcal{N}''$
$u_{n,a}$	probability corresponding to $f_{n,a}$
\mathcal{F}_n	set of all feasible traffic profiles for discrete DAs, $\mathcal{F}_n = \{f_{n,a}; 1 \leq a \leq A_n\}$
$V(d)$	objective value: (time) variance of d

7.3.1 Introduction

We have witnessed in recent years rapid increase in demand for wireless data, driven by the proliferation of smart mobile devices. The global mobile traffic in 2016 has nearly reached 84 exabytes, more than 80 times greater than the entire global Internet traffic in 2000; yet, this number is expected to be increasing at a compound annual growth rate (CAGR) of 47% in the coming five years, i.e., a seven-fold growth from 2016 to 2021 [55]. However, despite frequent upgrades of cellular networks technology from 2G to 4G LTE and beyond, wireless service providers fall short of keeping up with this increasing wireless data demand, leading to congestion in the network, especially in areas of dense population. As a result, users' data rates have to be throttled to ease congestions [2, 6, 9], at the cost of the degraded quality of service (QoS).

Admittedly, the capacity shortfall of cellular networks can be mitigated by allocating more wireless spectrum and deploying more wireless infrastructures including more and smaller cells and WiFi networks offloading, etc. However, spectrum allocation and infrastructure upgrading are not only costly but also time-consuming, while WiFi networks may not always be available and secure. A promising alternative, inspired by the similar problem of demand response in power networks, is to improve spectrum and infrastructure efficiency through managing wireless data traffic (i.e., demand). Notice that wireless traffic or demand usually fluctuates with a large peak-to-valley ratio throughout a day; see Fig. 7.3 for a trace of smartphone web browsing activity over a day. However, wireless capacity needs to be provisioned to meet the peak demand rather than the average. This means that the cellular network is usually stressed in peak hours while largely underutilized at other times. If the demand profile can be shaped to reduce the peak and smooth the time variation, not only can more traffic be accommodated under limited existing capacity constraints, but also additional spectrum allocation and infrastructure upgrades can be slowed down, which together greatly improve wireless network efficiency and QoS, and yield huge savings for service providers.

In this section, we focus on designing demand shaping algorithms for cellular networks. We divide wireless traffic into two categories: non-deferrable traffic and deferrable traffic. Non-deferrable traffic refers to the traffic of those applications such as online gaming that have no or low delay tolerance, and constitutes the base traffic whose profile cannot be shaped. Deferrable traffic refers to the traffic of those applications

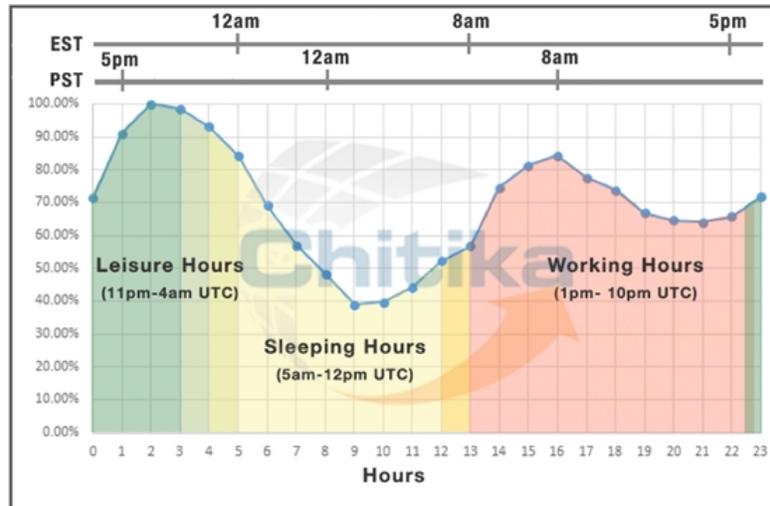


Figure 7.3: North America smartphone web browsing activity in one day [56].

such as file uploading/downloading that are flexible in time and only require being served by a designated deadline, e.g., finishing photo backup on cellphone by 12 am. Deferrable applications (DAs) are further divided into two major types: (1) continuous-rate interruptible applications such as photos backup and applications update that allow any data rates—e.g., the *delayed offloading* in [73, 90], and (2) discrete-rate non-interruptible applications such as online movie streaming and video conference that usually requires certain constant data rate [3, 4] and should not be interrupted once started, e.g., one can schedule movie watching or video conference to the “valley” time to enjoy better graphic quality and incur less data cost if he/she has the time flexibility. See Table 7.1 for a summary of traffic types and examples. We seek to schedule the deferrable traffic to flatten the aggregate traffic profile over a day.

Specifically, we formulate the cellular traffic demand shaping as an optimization problem that minimizes the (time) variation in the aggregate traffic profile subject to the time and rate specification on each DA. We first assume complete traffic information and design an offline demand shaping algorithm. There are two challenging issues in the offline algorithm design. First, the optimization problem is non-convex because of discrete-rate non-interruptible applications. We instead solve its convex relaxation and design a randomized scheme based on the solution to the relaxed problem. Second, demand shaping involves potentially a huge number of applications and users. A centralized algorithm is not scalable. We instead design an

Traffic/Application Type	Examples
Non-deferrable application	Online gaming, web browsing
Discrete-rate non-interruptible DA	Movie streaming, video conference
Continuous-rate interruptible DA	Applications update, photos backup

Table 7.1: Traffic/Application types and examples.

iterative and distributed algorithm based on the descent method. We establish the almost surely convergence for the algorithm based on supermartingale theory.

We then consider a more realistic setting with incomplete information where we can only predict future traffic to a certain degree of accuracy, and design an online and distributed demand shaping algorithm that updates the schedules of DAs each timeslot when new information and updated prediction are available, based on the offline algorithm for an optimization problem over a shrinking horizon from the current time to the end of the day. We compare the performance of the online algorithm against the optimal offline algorithm, and provide numerical examples to complement the theoretical analysis.

The rest of the section is organized as follows. Section 7.3.2 briefly reviews some related work and discusses some related issues. Section 7.3.3 describes the system model and problem formulation. Section 7.3.4 presents an offline distributed algorithm for demand shaping under the assumption of complete traffic information and characterizes its performance. Section 7.3.5 considers a realistic setting of incomplete traffic information, and presents an online algorithm for demand shaping. Section 7.3.6 provides numerical examples to complement theoretical analysis, and Section 7.3.7 concludes the section.

7.3.2 Related Work and Issues

Demand shaping in cellular networks is similar to demand response in power networks, in terms of design objectives, problem formulation, and the associated algorithmic challenges. Indeed, we borrow insights from demand response in power networks; see, e.g., [32, 47, 48, 76]. In particular, our online demand shaping algorithm is motivated by the solution approach for online control of continuous load in reference [48], and mathematically can be seen as its extension to incorporate discrete decision variables considered in reference [47]. However, our model captures realistic cellular traffic settings, as it includes

both continuous and discrete decision variables. Moreover, the integration of discrete decision variables into the online algorithm makes the performance analysis of the algorithm more challenging, compared to that in [48]. Related work also includes Zhao *et al* [131] that designs a centralized online EV charging algorithm to minimize the peak procurement from the grid under uncertain prediction of future demand and renewable energy supply, and Parise *et al* [95] that proposes a decentralized charging control for EVs to flatten the aggregate power demand profile. They all consider only continuous decision variables.

To ease the stress from high demand in cellular networks, various demand-shaping-based methodologies as well as traffic offloading strategies have been studied in existing literatures. Tadrous *et al* in [116] propose a paradigm to proactively serve peak-hour requests during the off-peak time based on prediction to smoothen the traffic demand over time without changing customers' activity pattern. However, such strategy is limited to routine behaviors only. In [54] Hajiesmaili *et al* introduce an online procurement auction framework to incentivize mobile devices to participate in device-to-device load balancing to offload traffic from one heavy-loaded base station to adjacent idle ones. Besides, WiFi and femtocell offloading of cellular data is another major approach to easing the congestion of cellular networks; see [14, 34, 57, 72, 73, 90] for related works.

In this section we have focused on designing demand shaping algorithms based on a general and simplified system model. We do not investigate the important practical issues such as the timescale and granularity at which we schedule and reschedule the DAs. We plan to develop a platform to enable automatic demand shaping in the future, and will investigate various practical issues then. Also, demand shaping involves not only the design of control algorithms but also the design of right mechanisms to incentivize the users to move out of their "comfortable zone" in wireless applications and data usage. Incentive design for demand shaping is currently an active research area; see, e.g., the smart data pricing in wireless networks [51, 106, 129]; pricing design in general network service to remove congestions [62, 96], pricing/reward signals in power distribution system [74, 138], and the references therein.

Some discussion on the practicality of demand shaping is also in place. People tend to use mobile data services whenever they want, regardless of whether it is at peak time or valley time for the cellular network. However, a survey [52] conducted in India and USA in 2012 shows that, given proper monetary incentive,

many people are willing to postpone their mobile data usage, with acceptable postponement varying from minutes to hours, depending on different types of services and different individual preferences [51]. For example, wireless service providers can motivate the users to shift their demand by implementing the time-dependent pricing (TDP) strategy. TDP is now applied as a simple two-period plan by many wireless service providers around the world, in voice services and data services; e.g., Verizon [8] and Sprint [5] in the US have “happy hours” in the night and weekend for voice service, TelCom [7] in South Africa has “Night Surfer” plans giving free data from 11pm to 5am, and Airtel [1] in India provides unlimited data in the night. More refined TDP strategies can be applied to maximize benefits for both wireless service providers and users, by dynamically adjusting prices according to the data usage of the current time and predicted future. For instance, Ha *et al* [51] have worked on a TDP-based application named TUBE. Trials in cooperation with a local wireless service provider shows its effectiveness in shaping the traffic profile [63]. Also refer to [105] for a review of pricing strategies.

7.3.3 System Model and Problem Formulation

Consider a cellular network that serves users for different applications such as web browsing, file sharing, real-time entertainment, etc. The applications can be broadly divided into two categories: deferrable applications (DAs) and non-deferrable applications (non-DAs). DAs refer to those applications that are flexible in the starting time and/or data rate, while the non-DAs refer to those that should be served immediately and often have stringent data rate requirement. Please refer to the third paragraph of Section 7.3.1 and TABLE 7.1 for more detailed description and examples of DAs and non-DAs.

This work aims to schedule the traffic of DAs so as to flatten the aggregate traffic profile over a day, subject to the time constraints and rate constraints of each application. We use a discrete-time model where one day is divided equally into T timeslots, indexed by $t \in \mathcal{T} = \{1, 2, \dots, T\}$. The duration of a timeslot can be, e.g., 30 minutes or 1 hour [51], depending on the time resolution of scheduling decisions.

7.3.3.1 Non-Deferrable Applications

Non-DAs include web browsing, online gaming, and real-time chatting with multimedia, etc. The latency tolerated by these applications usually varies from hundreds of milliseconds to seconds. Since these applications should be served immediately upon request, their traffic is inelastic and constitutes the *base traffic* whose profile cannot be shaped. Denote the base traffic profile by $b = \{b(t); t \in \mathcal{T}\}$. As we can only predict the base traffic to a certain accuracy, we model it as a random vector with mean $\bar{b} = \{\bar{b}(t); t \in \mathcal{T}\}$ and random derivation $\delta b = \{\delta b(t); t \in \mathcal{T}\}$ from the mean, i.e., $b = \bar{b} + \delta b$. We assume that $\delta b(t)$ has a mean of 0 and variance of $\delta^2(t)$, and may be temporally correlated. We further assume that we can make better prediction for the timeslots that are closer to current time, modeled by a time-dependent deviation from the mean, i.e., the base traffic at some future time $\tau \in \mathcal{T}$ is predicted at current time t by

$$b_t(\tau) = \bar{b}(\tau) + \delta b_t(\tau), \quad (7.40)$$

where the subscript t represents the timeslot when the prediction is made, and $\delta b_t(\tau)$ has a decreasing variance $\delta_t^2(\tau)$ as t approaches τ . More concrete model for prediction will be introduced in Section 7.3.6. The parameters \bar{b} and δ_t will be specified exogenously, and can be estimated from the historical traffic records.

7.3.3.2 Deferrable Applications

Assume that there are N DAs in the network, indexed by $n \in \mathcal{N} = \{1, \dots, N\}$. Each DA n is characterized by an arrival time t_n^a when it is requested or after which it can be started, a deadline t_n^d by which its transmission must be done, and certain requirement or constraint on data rate $p_n = \{p_n(t); t \in \mathcal{T}\}$. Let P_n denote the total traffic required by DA n , i.e., $\sum_{t \in \mathcal{T}} p_n(t) = P_n$. We can classify DAs into two main categories: *continuous-rate interruptible* DAs (or continuous DAs for simplicity) that allow any data rates between certain upper and lower bounds and can be interrupted and resumed at any time before the deadline, and *discrete-rate non-interruptible* DAs (or discrete DAs for simplicity) that require certain (roughly) constant data rate and cannot be interrupted once they are started. For example, system backup is usually interruptible and allows any continuous data rates, while video conference is usually preferred to be non-interruptible and runs at a constant (thus discrete) data rate once it is started.

Among the total N DAs, we assume there are N' continuous DAs, indexed by $n \in \mathcal{N}' = \{1, \dots, N'\}$. For each continuous DA, denote by $\underline{p}_n(t)$ and $\bar{p}_n(t)$ the lower and upper bounds on its data rate at time $t \in \mathcal{T}$, i.e.,

$$\underline{p}_n(t) \leq p_n(t) \leq \bar{p}_n(t), \quad t \in \mathcal{T}. \quad (7.41)$$

Naturally, $0 \leq \underline{p}_n(t) \leq \bar{p}_n(t)$. The lower bounds $\underline{p}_n(t)$ are usually zero, and the upper bounds $\bar{p}_n(t)$ can be set according to, e.g., the available bandwidth. The arrival time t_n^a and the deadline t_n^d can be integrated into the rate constraints (7.41) by setting $\bar{p}_n(t) = 0$ for $t < t_n^a$ and $t > t_n^d$, i.e., no traffic is transmitted before arrival time or after deadline.

Index the rest $N'' = N - N'$ discrete DAs by $n \in \mathcal{N}'' = \{N' + 1, \dots, N\}$. For a discrete DA such as a streaming application, a constant bit rate r_n corresponds to a certain graphic quality, e.g., $r_n = 3$ Mbps for a SD quality movie on Netflix [4], and $r_n = 1.2$ Mbps for a HD video call on Skype [3]. As the graphic quality usually (preferably) does not change during those applications, this seemingly over-simplified assumption of a single discrete rate is reasonable.

For each DA $n \in \mathcal{N}''$ with its total traffic P_n and the rate r_n , it takes $l_n = P_n/r_n$ consecutive timeslots (or equivalently the other way around, i.e., we calculate $P_n = l_n * r_n$ based on l_n and r_n). Therefore, the number of its feasible traffic profiles is $A_n = t_n^d - t_n^a - l_n + 1$, wherein the a -th feasible profile is denoted as

$$f_{n,a} = \left\{ p_n \middle| p_n(t) = \begin{cases} r_n, & \text{if } t_n^a + a - 1 \leq t \leq t_n^a + a + l_n \\ 0, & \text{otherwise} \end{cases} \right\}.$$

We denote the set of all feasible traffic profiles of DA $n \in \mathcal{N}''$ by $\mathcal{F}_n = \{f_{n,a} : 1 \leq a \leq A_n\}$, i.e., $p_n \in \mathcal{F}_n$, $\forall n \in \mathcal{N}''$.

Remark 7.1. *All the modeled traffic parameters can be reasonably accessed or estimated in practice. For example, information regarding total required traffic P_n and video streaming rate r_n is available from meta-data of traffic to be transmitted, parameters like t_n^a and t_n^d are specified by the users in advance (and \mathcal{F}_n can then be calculated accordingly), whereas data rate bounds $\underline{p}_n(t)$ and $\bar{p}_n(t)$ can be either determined by available bandwidth or designated by the users. See, e.g., [51] for an example system involving similar information requirement and implemented with real users and service provider.*

7.3.3.3 Problem Formulation

We aim to schedule the traffic of DAs, so as to flatten the aggregate traffic profile as much as possible. Denote the “average” traffic profile by $d = \{d(t); t \in \mathcal{T}\} := \frac{1}{N}(b + \sum_{n \in \mathcal{N}} p_n)$. Traffic flattening can be achieved by minimizing the time variance of d , formulated as the following optimal demand shaping (**ODS**) problem:

ODS:

$$\min_{p,d} \quad V(d) = \frac{1}{T} \sum_{t \in \mathcal{T}} (d(t) - \frac{1}{T} \sum_{\tau \in \mathcal{T}} d(\tau))^2 \quad (7.42a)$$

$$\text{s.t.} \quad d(t) = \frac{1}{N}(b(t) + \sum_{n \in \mathcal{N}} p_n(t)), \quad t \in \mathcal{T}, \quad (7.42b)$$

$$\underline{p}_n(t) \leq p_n(t) \leq \bar{p}_n(t), \quad t \in \mathcal{T}, \quad n \in \mathcal{N}', \quad (7.42c)$$

$$\sum_{t \in \mathcal{T}} p_n(t) = P_n, \quad n \in \mathcal{N}', \quad (7.42d)$$

$$p_n \in \mathcal{F}_n, \quad n \in \mathcal{N}''. \quad (7.42e)$$

Notice that the constraints (7.42e) for discrete DAs are non-convex. In next section, we will investigate an *offline* algorithm together with a randomized scheme for solving the ODS problem under the assumption of complete information on the base traffic and DAs. Then in Section 7.3.5, we will study an *online* algorithm for demand shaping under a more realistic setting of incomplete information where we can only predict the future traffic to a certain degree of accuracy. The offline ODS problem and algorithm will later serve as a benchmark to characterize the performance of the online algorithm.

7.3.4 Offline Demand Shaping Algorithm

In this section, we assume complete traffic information, i.e., the base traffic and arrival of DAs are accurately known, and study how to solve the resulting **offline ODS** problem. The offline problem and algorithm will provide insights into the online algorithm design for realistic settings of incomplete information that will be considered in Section 7.3.5.

7.3.4.1 Convex Relaxation and Randomized Scheme

The offline ODS problem is non-convex, as each discrete DA has to pick a traffic profile from a discrete set; see constraint (7.42e). Consider the convex hull of \mathcal{F}_n , defined as

$$\text{conv}(\mathcal{F}_n) := \left\{ p_n \mid p_n = \sum_{a=1}^{A_n} u_{n,a} \cdot f_{n,a}, u_{n,a} \geq 0 \text{ and } \sum_{a=1}^{A_n} u_{n,a} = 1 \right\}, \quad (7.43)$$

where $u_n := \{u_{n,1}, \dots, u_{n,A_n}\}$ is the convex combination coefficients, and will be interpreted as probability distribution in the randomized algorithm to be introduced soon. We will instead solve the convex relaxation of the ODS problem by replacing (7.42e) with the following constraint:

$$p_n \in \text{conv}(\mathcal{F}_n), n \in \mathcal{N}''. \quad (7.44)$$

We call the relaxed problem (7.42a)–(7.42d)(7.44) the **R-ODS** problem. However, a solution $p_n^* \in \text{conv}(\mathcal{F}_n)$, $n \in \mathcal{N}''$ to the R-ODS problem might not be feasible for original ODS, i.e., $p_n^* \notin \mathcal{F}_n$. But since by definition (7.43) a solution p_n^* can always be written as the convex combination $\sum_{a=1}^{A_n} u_{n,a} f_{n,a}$ we will randomly pick a traffic profile $p_n = f_{n,a} \in \mathcal{F}_n$ with corresponding probability $u_{n,a}$. That said, we will design a randomized algorithm for the offline ODS problem, based on the solution to the R-ODS problem. We will integrate it into a distributed algorithm next.

7.3.4.2 Distributed Algorithm

Solving the R-ODS problem (and the offline ODS problem) directly in a centralized way requires collecting information on all DAs, which may incur too much communication overhead and is impractical in the real network. Moreover, the users may not be willing to reveal information on DAs due to privacy concern. Therefore, we seek to solve it in a distributed way. Noticing that R-ODS problem has decoupled constraints, we attempt to design an *iterative and distributed* algorithm based on the decent method [?].

Before deriving the algorithm, we establish the following useful results. At k -th iteration, let $p^k = \{p_n^k; n \in \mathcal{N}\}$ be the traffic profiles of all DAs, $d^k = \frac{1}{N}(b + \sum_{n \in \mathcal{N}} p_n^k)$ the average traffic profile, and $x_n^k = p_n^{k+1} - p_n^k$, $n \in \mathcal{N}$ the change in traffic profile of DA n from iteration k to $k + 1$. We have:

$$E\left[\left\| \sum_{n \in \mathcal{N}} x_n^k \right\|_2^2\right] = \sum_{n \in \mathcal{N}} \text{Var}(x_n^k) + \left\| \sum_{n \in \mathcal{N}} E[x_n^k] \right\|_2^2, \quad (7.45)$$

where the variance $Var(x_n^k) := E[\|x_n^k\|_2^2] - \|E[x_n^k]\|_2^2$, and $E[\cdot]$ denotes the average.⁶ By Jensen's inequality,

$$\left\| \sum_{n \in \mathcal{N}} E[x_n^k] \right\|_2^2 \leq N \sum_{n \in \mathcal{N}} \|E[x_n^k]\|_2^2. \quad (7.46)$$

Therefore, one has

$$E\left[\left\| \sum_{n \in \mathcal{N}} x_n^k \right\|_2^2\right] \leq \sum_{n \in \mathcal{N}} Var(x_n^k) + N \sum_{n \in \mathcal{N}} \|E[x_n^k]\|_2^2. \quad (7.47)$$

And it follows that

$$\begin{aligned} & TN^2(E[V(d^{k+1})|p^k] - V(d^k)) \\ &= E\left[\left\| \sum_{n \in \mathcal{N}} x_n^k \right\|_2^2 + 2\langle Nd^k, \sum_{n \in \mathcal{N}} x_n^k \rangle\right] \\ &\leq \sum_{n \in \mathcal{N}} Var(x_n^k) + N \sum_{n \in \mathcal{N}} \|E[x_n^k]\|_2^2 + 2 \sum_{n \in \mathcal{N}} E[\langle Nd^k, x_n^k \rangle] \\ &= \sum_{n \in \mathcal{N}'} (2\langle Nd^k, x_n^k \rangle + N\|x_n^k\|_2^2) + \sum_{n \in \mathcal{N}''} (2\langle Nd^k, E[x_n^k] \rangle + N\|E[x_n^k]\|_2^2 + Var(x_n^k)). \end{aligned} \quad (7.48)$$

Denote by W_1 the first term in (7.48) and W_2 the second. For $n \in \mathcal{N}'$, we choose p_n^{k+1} so as to minimize W_1 , i.e., to solve

$$\min_{p_n} \quad 2\langle d^k, p_n - p_n^k \rangle + \|p_n - p_n^k\|_2^2 \quad (7.49a)$$

$$\text{s.t.} \quad (7.42c)-(7.42d). \quad (7.49b)$$

On the other hand, after some mathematical manipulations, we have

$$W_2 = \sum_{n \in \mathcal{N}''} \left(2N\langle d^k - p_n^k, E[p_n^{k+1}] \rangle + (N-1)\|E[p_n^{k+1}]\|_2^2 \right) + \Pi^k,$$

where Π^k is a constant given p_n^k . For $n \in \mathcal{N}''$, we choose p_n^{*k+1} so as to minimize W_2 , i.e., to solve

$$\min_{p_n \in \text{conv}(\mathcal{F}_n)} \quad 2\langle d^k - p_n^k, p_n \rangle + \frac{N-1}{N} \|p_n\|_2^2. \quad (7.50)$$

In essence, what we have done is to *maximize the expected incremental decrease in the objective value* $V(d)$ at each iteration (i.e., steepest descent). This motivates a distributed demand shaping algorithm with the collaboration of a coordinator; see Algorithm 5. The wireless service provider can implement a logical coordinator at the base station.

⁶ Notice that we consider a randomized scheme only for discrete DAs. That said, for continuous DAs there is no randomness and their variance is zero.

Algorithm 5 Offline Demand Shaping (Off-DS) Algorithm

At k -th iteration:

(1) Upon gathering traffic profiles p_n^k from DAs, the coordinator calculates the average traffic profile

$$d^k = \frac{1}{N}(b + \sum_{n \in \mathcal{N}} p_n^k) \text{ and announces it to DAs (or the end users) over a signaling or control channel.}$$

(2) Upon receiving the average traffic profile d^k ,

- DA $n \in \mathcal{N}'$ updates its traffic profile by

$$p_n^{k+1} = \arg \min_{p_n} \|p_n - p_n^k + d^k\|_2^2$$

s.t. (7.42c)–(7.42d),

and submits it to the coordinator.

- DA $n \in \mathcal{N}''$ calculates the average traffic profile by

$$p_n^{*k+1} = \arg \min_{p_n \in \text{conv}(\mathcal{F}_n)} \left\| p_n - \frac{N}{N-1}(p_n^k - d^k) \right\|_2^2,$$

which is $p_n^{*k+1} = \sum_{a=1}^{A_n} u_{n,a}^{k+1} f_{n,a}$, and then randomly chooses a traffic profile $p_n^{k+1} = f_{n,a}$ with probability $u_{n,a}^{k+1}$ and submits it to the coordinator.

The Off-DS algorithm is a distributed algorithm wherein each DA solves its own simple optimization problem based on its previous decision, the average traffic profile d^k , and local constraints, while the coordinator collects the proposed traffic profiles and updates the average traffic profile. Therefore, this algorithm is not only preserving privacy of the users, but also scalable and thus capable of quick response, which is crucial especially in real-time implementation in Section 7.3.5.

The computational complexity of the Off-DS algorithm is estimated as follows for completeness. Given certain accuracy requirement $\epsilon > 0$ in the objective function value, the descent method requires $O(\log(1/\epsilon))$ iterations [31]. At each iteration, DAs solves an easy quadratic programming with a polynomial complexity of $O(T^{O(1)})$ [99]. On the other hand, the coordinator calculates the average traffic profile which requires $O(N)$ complexity each iteration. As a result, the Off-DS algorithm requires overall computational complexity of $O((N + T^{O(1)}) \log(1/\epsilon))$.

Remark 7.2. For simpler expression, we use p_n as the decision variable for DA $n \in \mathcal{N}''$ in algorithm design and analysis, while in real implementation, it is more convenient to use probability distribution u_n as the equivalent decision variable. Also notice that, if there is no continuous DA, Algorithm 5 reduces to the stochastic algorithm in [47]. We expect that the solution approach—randomized algorithm based on the “steepest” descent method for the convex relaxed problem—that we lay out in Sections 7.3.4.1 and 7.3.4.2 will find broad application in designing efficient algorithms for optimization problems that involve both continuous and discrete decision variables.

7.3.4.3 Convergence

Before showing the convergence of the Off-DS algorithm, we first establish two useful relations. For each DA $n \in \mathcal{N}'$, since p_n^{k+1} solves the problem (7.49), we have the first-order optimality condition

$$\langle p_n^{k+1} - p_n^k + d^k, p_n - p_n^{k+1} \rangle \geq 0 \quad (7.51)$$

for any feasible p_n . Set $p_n = p_n^k$ to obtain

$$\langle d^k, p_n^{k+1} - p_n^k \rangle \leq -\|p_n^{k+1} - p_n^k\|_2^2. \quad (7.52)$$

For each DA $n \in \mathcal{N}''$, recalling that $p_n^{*k+1} = E[p_n^{k+1}]$, by the first-order optimality condition, we have

$$\left\langle \frac{N}{N-1}(d^k - p_n^k) + p_n^{*k+1}, p_n - p_n^{*k+1} \right\rangle \geq 0 \quad (7.53)$$

for any feasible p_n . Set $p_n = p_n^k$ to get

$$\langle Nd^k, p_n^{*k+1} - p_n^k \rangle \leq -(N-1)\|p_n^{*k+1} - p_n^k\|_2^2 + \langle p_n^k, p_n^{*k+1} - p_n^k \rangle. \quad (7.54)$$

Now, construct a filtration Σ^* of the probability space $\{\Omega, \Sigma, \mathcal{P}\}$, where the sample space Ω is the feasible set specified by the constraints (7.42c)–(7.42e), the σ -algebra $\Sigma_k = \Omega$, $k \geq 0$, and $\mathcal{P}(\Sigma_k) = \{\delta(p_n - p_n^k), n \in \mathcal{N}'; u_{n,a}^k, 1 \leq a \leq A_n, n \in \mathcal{N}''\}$, i.e., determined by the k -th iteration of the Off-DS algorithm.

Theorem 7.11. The pair $(V(d), \Sigma^*)$ is a supermartingale. □

Proof. First, notice that $V(d)$ is bounded from below. So, $E[-\min\{0, V(d)\}] < \infty$. Second, applying relations (7.52)–(7.54) to equation (7.48), we obtain

$$\begin{aligned}
& TN^2(E[V(d^{k+1})|p^k] - V(d^k)) \\
& \leq \sum_{n \in \mathcal{N}'} -N\|x_n^k\|_2^2 + \sum_{n \in \mathcal{N}''} (\text{Var}(x_n^k) + (-N + 2)\|E[x_n^k]\|_2^2 + 2\langle p_n^k, p_n^{*k+1} - p_n^k \rangle) \\
& = \sum_{n \in \mathcal{N}'} -N\|x_n^k\|_2^2 + \sum_{n \in \mathcal{N}''} (-N + 1)\|E[x_n^k]\|_2^2 \\
& \leq 0,
\end{aligned}$$

i.e., $E[V(d^{k+1})|p^k] \leq V(d^k)$. By definition, $(V(d), \Sigma^*)$ is a supermartingale [49]. \blacksquare

Notice that $(V(d), \Sigma^*)$ is a nonnegative supermartingale. By the martingale convergence theorem [49], the following result is immediate.

Corollary 7.3. $V(d^\infty) = \lim_{k \rightarrow \infty} V(d^k)$ exists almost surely, where $V(d^\infty)$ is some random variable. \square

Theorem 7.12. Denote by \mathcal{P}^∞ an “equilibrium” distribution over traffic profiles that $(V(d), \Sigma^*)$ converges to. The support of \mathcal{P}^∞ is a singleton. \square

Proof. When $(V(d), \Sigma^*)$ converges, $E[V(d^{k+1})|p^k] = V(d^k)$. This requires $E[x_n^k] = E[x_{n'}^k]$, $n, n' \in \mathcal{N}$, $p_n^{k+1} = p_n^k$, $n \in \mathcal{N}'$, and $p_n^{*k+1} = p_n^k$, $n \in \mathcal{N}''$ for (7.47), (7.52), and (7.54) to hold with equality. Notice that $p_n^{*k+1} = p_n^k$ implies $p_n^{k+1} = p_n^k$, as different feasible traffic profiles of DA $n \in \mathcal{N}''$ are linearly independent. Thus, $p_n^{k+1} = p_n^k$, $n \in \mathcal{N}$. So, the support of \mathcal{P}^∞ contains only one point. \blacksquare

Denote by p^∞ an “equilibrium” traffic profile of the Off-DS algorithm, i.e., if $p^k = p^\infty$, then $p^{k+1} = p^\infty$. Obviously the set of equilibrium profiles is not empty, as an optimum of the offline ODS problem is an equilibrium. The following result follows immediately from Theorem 7.12 and Corollary 7.3.

Theorem 7.13. The Off-DS algorithm converges almost surely to an equilibrium traffic profile. \square

By equations (7.51)–(7.53), we have the following optimality conditions at equilibrium p^∞ : for any feasible p_n ,

$$\langle b + \sum_{m \in \mathcal{N}} p_m^\infty, p_n - p_n^\infty \rangle \geq 0, \quad n \in \mathcal{N}', \quad (7.55a)$$

$$\langle b + \sum_{m \neq n} p_m^\infty, p_n - p_n^\infty \rangle \geq 0, \quad n \in \mathcal{N}''. \quad (7.55b)$$

7.3.4.4 Performance Analysis of the Offline Algorithm

We now characterize the performance of Off-DS algorithm with respect to the relaxed problem R-ODS that at optimum may attain a lower objective value than the ODS problem. Specifically, denote by p^* the solution of R-ODS, we bound the gap between the equilibrium of the Off-DS algorithm and the solution of the R-ODS problem as: $G^{\text{off}} := V(d^\infty) - V(d^*)$, where $d^\infty = (b + \sum_{n \in \mathcal{N}} p_n^\infty)/N$ and $d^* = (b + \sum_{n \in \mathcal{N}} p_n^*)/N$. Denote by $G_r^{\text{off}} := (V(d^\infty) - V(d^*))/V(d^*)$ the relative gap achieved by the Off-DS algorithm.

Theorem 7.14. *For the Off-DS algorithm, the gap G^{off} is bounded as follows:*

$$G^{\text{off}} \leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|p_n^\infty\|_2^2. \quad (7.56)$$

Moreover, the relative gap diminishes as the number N'' of discrete DAs increases, i.e.,

$$\lim_{N'' \rightarrow \infty} G_r^{\text{off}} = 0. \quad (7.57)$$

Proof. For notational simplicity, let $c_d := \sum_{t \in \mathcal{T}} d(t)/T$, which is a constant given the total amount of traffic.

The objective value can be written as

$$\begin{aligned} V(d) &= \frac{1}{T} \|d - c_d \cdot \mathbf{1}\|_2^2 \\ &= \frac{1}{T} (\|d\|_2^2 + c_d^2 \|\mathbf{1}\|_2^2 - 2\langle d, \mathbf{1} \rangle) \\ &= \frac{1}{T} (\|d\|_2^2 + T \cdot c_d^2 - 2T \cdot c_d), \end{aligned}$$

where only the part $\|d\|_2^2$ contains decision variables. We can thus write the gap G^{off} as

$$\begin{aligned} G^{\text{off}} &= V(d^\infty) - V(d^*) = \frac{1}{T} (\|d^\infty\|_2^2 - \|d^*\|_2^2) \\ &= \frac{1}{T} (-\|d^\infty - d^*\|_2^2 + \langle 2d^\infty, d^\infty - d^* \rangle) \\ &\leq \frac{1}{T} \langle 2d^\infty, d^\infty - d^* \rangle \\ &= \frac{2}{N^2} \left(\sum_{n \in \mathcal{N}'} \langle Nd^\infty, p_n^\infty - p_n^* \rangle + \sum_{n \in \mathcal{N}''} \langle Nd^\infty, p_n^\infty - p_n^* \rangle \right) \\ &\leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \langle p_n^\infty, p_n^\infty - p_n^* \rangle \\ &\leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|p_n^\infty\|_2^2, \end{aligned}$$

where the second inequality follows from (7.55). Note that $\|p_n^\infty\|_2^2$ is a constant for $n \in \mathcal{N}''$. Then the relative gap G_r^{off} can be bounded as

$$\begin{aligned} G_r^{\text{off}} &\leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|p_n^\infty\|_2^2 / V(d^*) \\ &= \frac{\sum_{n \in \mathcal{N}''} \|p_n^\infty\|_2^2}{\|b + \sum_{n \in \mathcal{N}} p_n^*\|_2^2 + N^2(T \cdot c_d^2 - 2T \cdot c_d)}, \end{aligned} \quad (7.58)$$

whose numerator increases linearly with N'' and denominator increases linearly with the square of N'' . Equation (7.57) follows. \blacksquare

Remark 7.3. *We use the relaxed problem R-ODS for comparison instead of the ODS problem for two reasons. First, it is difficult to characterize the optimum of the non-convex ODS problem, and thus evaluating the gap between the equilibrium of the Off-DS algorithm and the optimum of ODS problem is mathematically hard. Second, R-ODS achieves an optimal objective value that is not greater than ODS, resulted from convex relaxation for the discrete decision variables. Therefore, G^{off} provides an upper bound for the “actual” sub-optimality, i.e., the gap between the equilibrium of Off-DS and the optimum of ODS.*

7.3.5 Online Demand Shaping Algorithm

In this section, we consider a realistic setting with incomplete information where we can only predict future traffic to a certain degree of accuracy, and study online demand shaping that makes decisions based on the prediction of future traffic and updates the decision as new information becomes available.

A typical algorithm used in this setting is the receding horizon control; see, e.g., [70]. However, as the objective function (7.42a) does not have a nice additive structure, receding horizon control algorithm does not admit an easy analysis. We will instead extend a shrinking horizon control algorithm, which is used in [48] that studies mathematically the same problem with only continuous DAs, to include discrete DAs, and apply it to our online demand shaping (**online DS**) problem.

7.3.5.1 Online Algorithm

We assume that the number m_t of DAs arriving at time t is randomly distributed with a mean λ_t and variance $(\delta\lambda_t)^2$, and the total amount of traffic of each DA is randomly distributed with a mean P and

variance $(\delta P)^2$. Denote by $\mathcal{N}'_t = \{1, \dots, N'_t\}$ the set of continuous DAs and $\mathcal{N}''_t = \{N'_t + 1, \dots, N_t\}$ the set of discrete DAs that have arrived by time $t \in \mathcal{T}$, and let $\mathcal{N}_t = \mathcal{N}'_t \cup \mathcal{N}''_t$ and $N''_t = N_t - N'_t$. Notice that we cannot reschedule the remaining traffic of a discrete DA that has already started. Denote by $\tilde{\mathcal{N}}''_t \subseteq \mathcal{N}''_t$ the set of discrete DAs that have not been started by time t . For DA $n \in \tilde{\mathcal{N}}''_t$, denote by $\mathcal{F}_n(t) = \{f_{n,a}; 1 \leq a \leq A_n(t)\}$ the set of feasible traffic profiles at time t . Let $\tilde{\mathcal{N}}_t = \mathcal{N}'_t \cup \tilde{\mathcal{N}}''_t$ be the set of DAs whose profiles are still adjustable at time t (i.e., all the continuous DAs and the discrete DAs that have not started by time t).

At time t , we make a prediction $b_t(t : T)$ of base traffic for the rest timeslots of the day, and we also have the information on DA $n \in \mathcal{N}_t$ and the expected total future deferrable traffic $\sum_{\tau=t+1}^T P\lambda_\tau$. Following [48], we introduce a *virtual* deferrable traffic profile $q(t : T) = \{q(\tau); t \leq \tau \leq T\}$ with $q(t) = 0$ and $\sum_{\tau=t}^T q(\tau) = \sum_{\tau=t+1}^T P\lambda_\tau$, to emulate the impact of the future deferrable traffic upon the current demand shaping decision. With the afore setup, we aim to schedule and reschedule the DAs, so as to solve the following problem at each timeslot $t \in \mathcal{T}$.

ODS_t:

$$\min V(d) = \frac{1}{T-t+1} \sum_{\tau=t}^T \left(d(\tau) - \frac{\sum_{s=\tau}^T d(s)}{T-t+1} \right)^2 \quad (7.59a)$$

over $p(t : T)$, $d(t : T)$, $q(t : T)$

$$\text{s.t. } d(\tau) = \frac{b_t(\tau) + q(\tau) + \sum_{n \in \mathcal{N}_t} p_n(\tau)}{N_t}, \quad \tau \geq t, \quad (7.59b)$$

$$\underline{p}_n(\tau) \leq p_n(\tau) \leq \bar{p}_n(\tau), \quad \tau \geq t, \quad n \in \mathcal{N}'_t, \quad (7.59c)$$

$$\sum_{\tau=t}^T p_n(\tau) = P_n(t), \quad n \in \mathcal{N}'_t, \quad (7.59d)$$

$$p_n \in \mathcal{F}_n(t), \quad n \in \tilde{\mathcal{N}}''_t, \quad (7.59e)$$

$$\sum_{\tau=t}^T q(\tau) = \sum_{\tau=t+1}^T P\lambda_\tau, \quad (7.59f)$$

where $p(t : T) = \{p_n(\tau); t \leq \tau \leq T, n \in \tilde{\mathcal{N}}_t\}$, $d(t : T) = \{d(\tau); t \leq \tau \leq T\}$, and $P_n(t) = P_n - \sum_{\tau=1}^{t-1} p_n(\tau)$, $n \in \mathcal{N}'_t$ is the amount of traffic to be served at or after time t .

We can solve the ODS_t problem at each timeslot the same way as we solve the offline ODS problem (7.42), constituting an online demand shaping algorithm; see Algorithm 6, wherein the convergence (and computational complexity) of Step 2) can be established (and analyzed) in the same way as Algorithm 5.

Algorithm 6 Online Demand Shaping (On-DS) Algorithm

At each timeslot $t \in \mathcal{T}$:

(1) Denote by $p_n^{(t-1)}$, $n \in \mathcal{N}_{t-1}$ the schedules determined by time $t-1$, and by $\hat{\mathcal{N}}_t'' \subseteq \mathcal{N}_t''$ the set of discrete DAs that has been started before time t . For each DA $n \in \hat{\mathcal{N}}_t''$, set its schedule $p_n(t; T) = \{p_n(\tau); t \leq \tau \leq T\}$ as $p_n(\tau) = p_n^{(t-1)}(\tau)$, $t \leq \tau \leq T$.

(2) Solve the ODS_t problem iteratively: at k -th iteration,

(a) Upon gathering traffic profiles $p_n^k(t : T) = \{p_n^k(\tau); t \leq \tau \leq T\}$ from DAs $n \in \tilde{\mathcal{N}}_t$, the coordinator solves

$$\begin{aligned} \min_{q(t+1:T)} \sum_{\tau=t+1}^T & \left(b_t(\tau) + q(\tau) + \sum_{n \in \hat{\mathcal{N}}_t''} p_n(\tau) + \sum_{n \in \tilde{\mathcal{N}}_t} p_n^k(\tau) \right)^2 \\ \text{s.t.} \quad & (7.59\text{f}), \end{aligned}$$

to obtain a virtual deferrable traffic $\{q^k(\tau); t+1 \leq \tau \leq T\}$, and then calculates the average traffic $d^k(\tau) = \frac{1}{N_t}(b_t(\tau) + q^k(\tau) + \sum_{n \in \hat{\mathcal{N}}_t''} p_n(\tau) + \sum_{n \in \tilde{\mathcal{N}}_t} p_n^k(\tau))$ for $\tau \geq t$ and announces it to DA $n \in \tilde{\mathcal{N}}_t$ over a signaling or control channel.

(b) Upon receiving the average traffic profile d^k ,

- DA $n \in \mathcal{N}'_t$ obtains $p_n^{k+1}(t : T)$ by

$$\begin{aligned} \min_{p_n(t:T)} \quad & \left\| p_n(t : T) - p_n^k(t : T) + d^k(t : T) \right\|_2^2 \\ \text{s.t.} \quad & (7.59\text{c})\text{--}(7.59\text{d}), \end{aligned}$$

and submits the updated profile to the coordinator.

- DA $n \in \tilde{\mathcal{N}}_t''$ calculates $p_n^{*k+1}(t : T)$ by

$$\begin{aligned} \min_{p_n(t:T)} \quad & \left\| p_n(t : T) - \frac{N_t}{N_t - 1} (p_n^k(t : T) - d^k(t : T)) \right\|_2^2 \\ \text{s.t.} \quad & p_n(t : T) \in \text{conv}(\mathcal{F}_n(t)), n \in \tilde{\mathcal{N}}_t'', \end{aligned}$$

represents it as a convex combination $p_n^{*k+1} = \sum_{a=1}^{A_n(t)} u_{n,a}^{k+1} f_{n,a}$, and randomly chooses a traffic profile $p_n^{k+1} = f_{n,a}$ with probability $u_{n,a}^{k+1}$ and submits it to the coordinator.

7.3.5.2 Performance Analysis of the Online Algorithm

We now characterize the performance of On-DS algorithm with respect to the result of Off-DS algorithm which serves as a benchmark. We will make the following assumptions to simplify the analysis and obtain insights into how uncertainties affect the performance of On-DS algorithm.

Assumption 7.1. *The amount of deferrable traffic is large and flexible enough so that a valley-filling schedule exists at every time $t = 1, \dots, T$, i.e., there exists some constant $C(t) \geq b_t(\tau), \forall \tau = t, \dots, T$ such that*

$$Nd(t) = C(t) = \frac{1}{T-t+1} \left(\sum_{\tau=t}^T b_t(\tau) + \sum_{\tau=t+1}^T P\lambda_\tau + \sum_{n=1}^{N_t} P_n(t) \right). \quad (7.60)$$

Remark 7.4. *Assumption 7.1 looks a strong assumption, and we do not have empirical evidence to support it as demand shaping has not being widely adopted in current cellular networks. However, with increasing penetration of deferrable traffics and users, this assumption expects to hold. One purpose of algorithm design as in this section and incentive design as in [51] is to facilitate and incentivize wide adoption of demand shaping. On the other hand, valley-filling represents the scenario where demand shaping is most useful and presents a benchmark for the potential of demand shaping. Mathematically, it is very difficult to analyze the performance of the online algorithm under more general assumption than Assumption 7.1. However, notice that in numerical examples in Section 7.3.6, we do not impose Assumption 7.1 while the results still fall into the bound specified in Theorem 7.15.*

Assumption 7.2. *The base traffic prediction at t is modeled as the following causal filter*

$$b_t(\tau) = \bar{b}(\tau) + \sum_{s=1}^T e(s)f(\tau-s), \quad \tau = 1, \dots, T, \quad (7.61)$$

where $e = \{e(s)\}_{s=1}^T$ is an uncorrelated sequence of independent and identically distributed random variables with mean 0 and variance δ^2 , and $f = \{f(\tau)\}_{\tau=-\infty}^{\infty}$ is the impulse response with $f(0) = 1$. Let $F(t) := \sum_{s=0}^t f(s)$.

We denote by G^{on} the gap defined as the expected difference between the results of On-DS algorithm and Off-DS algorithm, i.e., $G^{\text{on}} = E[V(\hat{d}) - V(d^{\infty})]$, where E denotes the expectation, and \hat{d} and d^{∞} denote the average traffic profiles achieved by the On-DS algorithm and the offline-DS algorithm respectively. It

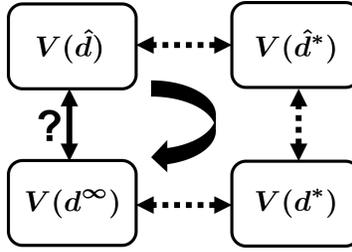


Figure 7.4: Strategy to calculate the gap between the equilibrium of the On-DS algorithm and that of Off-DS algorithm.

turns out that direct calculation of this gap is difficult. We therefore utilize two intermediate variables: d^* , the average traffic profile achieved by the R-ODS problem, and \hat{d}^* , the average traffic profile achieved by the relaxed online DS, i.e., the counterpart of R-ODS problem in the online scenario. Similar notations are applied to individual traffic profile p_n . With the relation shown in Fig. 7.4, we can write online gap as

$$\begin{aligned} G^{\text{on}} &= E[V(\hat{d}) - V(\hat{d}^*) + V(\hat{d}^*) - V(d^*) + V(d^*) - V(d^\infty)] \\ &= E[V(\hat{d}) - V(\hat{d}^*)] + E[V(\hat{d}^*) - V(d^*)] + E[V(d^*) - V(d^\infty)]. \end{aligned} \quad (7.62)$$

Theorem 7.15. *The gap, i.e., the expected difference between the results of On-DS algorithm and Off-DS algorithm is bounded as follows:*

$$G^{\text{on}} = E[V(\hat{d}) - V(d^\infty)] \leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|\hat{p}_n\|_2^2 + \frac{(\delta\lambda)^2}{T} \sum_{t=2}^T \frac{1}{t} + \frac{\delta^2}{T^2} \sum_{t=0}^{T-1} F^2(t) \frac{T-t-1}{t+1}. \quad (7.63)$$

Proof. Applying the approach and results from Theorem 7.14, we have

$$0 \leq E[V(\hat{d}) - V(\hat{d}^*)] \leq \frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|\hat{p}_n\|_2^2, \quad (7.64a)$$

$$-\frac{2}{TN^2} \sum_{n \in \mathcal{N}''} \|\hat{p}_n\|_2^2 \leq E[V(d^*) - V(d^\infty)] \leq 0. \quad (7.64b)$$

For the second term of (7.62), under Assumptions 7.1–7.2, following [48], we get

$$E[V(\hat{d}^*) - V(d^*)] = \frac{(\delta\lambda)^2}{T} \sum_{t=2}^T \frac{1}{t} + \frac{\delta^2}{T^2} \sum_{t=0}^{T-1} F^2(t) \frac{T-t-1}{t+1}. \quad (7.65)$$

Combine (7.64a)–(7.65) to obtain (7.63). ■

Theorem 7.15 indicates that, the size of the gap between online and offline algorithms changes monotonically with prediction error of both base traffic and future arrival of deferrable traffic. Accordingly we

can improve the result of On-DS algorithm by implementing better prediction mechanism, e.g., On-DS algorithm which updates its prediction to keep the value of prediction error small. Also, if the impulse response f is chosen to fade quickly enough, then as we have finer time granularity, we have $T \rightarrow \infty$, and $G^{\text{on}} \rightarrow 0$, which intuitively indicates that, with infinitely small timeslot, we can update our decisions frequently enough to mitigate prediction errors, and therefore have a negligible performance gap.

Lastly, similar to Theorem 7.14, define a relative gap $G_r^{\text{on}} := G^{\text{on}}/V(d^\infty)$. The following result is immediate.

Theorem 7.16. *The relative gap G_r^{on} diminishes as the number of discrete DAs N'' increases, i.e.,*

$$\lim_{N'' \rightarrow \infty} G_r^{\text{on}} = 0. \quad (7.66)$$

Remark 7.5. *It is worth noting that equation (7.66) does not necessarily imply a monotone decreasing of G_r^{on} with respect to N'' . This can be seen from Fig. 7.8 in Section 7.3.6 that does not show a decreasing G_r^{on} as N'' increases.*

By equations (7.64a) and (7.65), it is straightforward to obtain the following result.

Corollary 7.4. *The expected difference between the On-DS algorithm and the optimum of the R-ODS problem is bounded as follows:*

$$E[V(\hat{d}) - V(d^*)] \leq \frac{2}{TN^2} \sum_{n \in N''} \|\hat{p}_n\|_2^2 + \frac{(\delta\lambda)^2}{T} \sum_{t=2}^T \frac{1}{t} + \frac{\delta^2}{T^2} \sum_{t=0}^{T-1} F^2(t) \frac{T-t-1}{t+1}.$$

7.3.6 Numerical Examples

In this section, we provide numerical examples to evaluate the performance of the On-DS algorithm. We use certain composite traffic traces to drive simulations to show the impact of base traffic prediction errors, deferrable traffic prediction errors, and deferrable traffic penetration levels. We expect the conclusions obtained to hold for real traffic.

7.3.6.1 Experimental Setup

Consider a 48-hour period of time starting from 4:00 pm to 3:59 pm two days later. We divide the 48 hours equally into 96 timeslots, each 30 minutes long. We consider scheduling traffic that arrives within the

first 24 hours only, which may be allocated to the second 24 hours.

Non-deferrable traffic: The “real” trace we use for non-deferrable traffic, or base traffic, is shown in Fig. 7.5 (red line). It is constructed by random fluctuation around the average base traffic trace (blue line) composed based on North American mobile web browsing activity by time of day in 2013 [56], shown in Fig. 7.3. As modeled in Section 7.3.3.1, the prediction of base traffic follows (7.40), consisting of average base traffic $\bar{b}(\tau)$ and random deviation $\delta b_t(\tau)$ from the average value. Following [48], at time t , $\delta b_t(\tau)$ is modeled as

$$\delta b_t(\tau) = \sum_{s=t+1}^{\tau} \omega_s(\tau), \quad t < \tau \leq T, \quad (7.67)$$

where $\omega_s(\tau)$ are random variables of Gaussian distribution with 0 mean and variances

$$E[\omega_s^2(\tau)] = \frac{\sigma^2}{\tau - s + 1}, \quad 1 \leq s \leq \tau \leq T. \quad (7.68)$$

In this way, $\delta b_t(\tau)$ has decreasing variance as t approaches τ , simulating a gradually improving prediction for some future timeslot τ as one gets closer to it. In simulation, we take the values of σ^2 in (7.68) from 0 to 100 with increment of 10, corresponding to a root-mean-square prediction error (RMSE) ranging from 0% to 32%, looking 48 timeslots (24 hours) ahead.

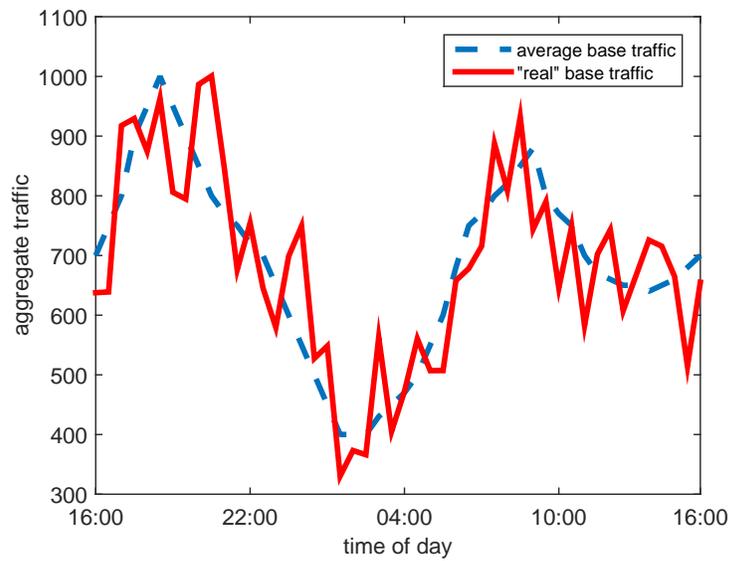


Figure 7.5: Base traffic: the average (blue/dotted) and a “real” trace (red/solid).

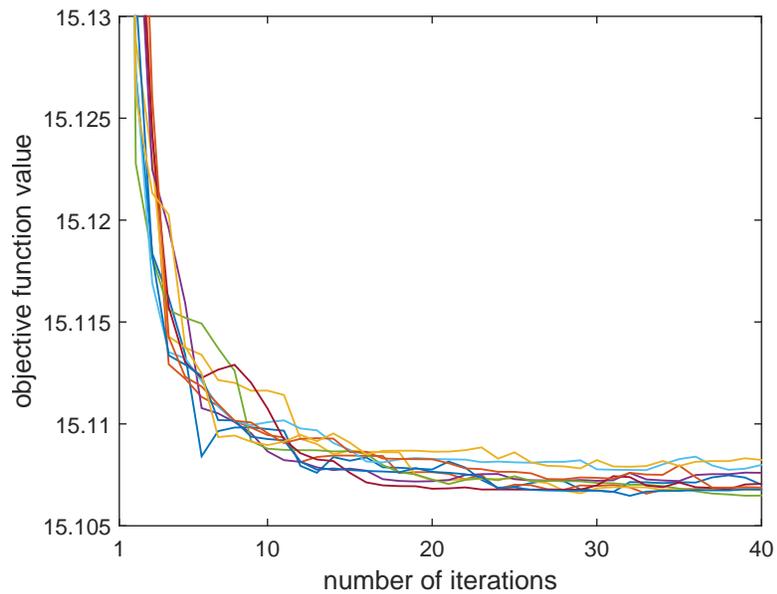


Figure 7.6: Repetitive experiments show that a number of 20 to 30 iterations give a satisfying result in terms of convergence.

Deferrable traffic:

We assume that the number of DAs arriving at each timeslot follows a “shifted” Poisson process

$m + \text{poissrnd}(\lambda_p)$, with $m \geq 0$ and $\text{poissrnd}(\lambda_p)$ denoting a Poisson process with rate λ_p . Here, we set $\lambda_p = 4$, while each DA has a 50-50 chance to require continuous- or discrete-rate traffic. The total traffic P_n of each DA is uniformly distributed in $[\underline{P}, \bar{P}]$ where we set $\underline{P} = 12$ and $\bar{P} = 24$. The deadline for DA n is uniformly distributed in $[t_n^a + l_n + \underline{D}, t_n^a + l_n + \bar{D}]$, where $l_n = \lceil P_n / \bar{p}_n \rceil$ is the minimum number of timeslots required by the DA calculated by ceiling function $\lceil \cdot \rceil$. We set $\underline{D} = 6$, $\bar{D} = 14$, and a universal bit rate upper bound $\bar{p}_n = 3$.

Benchmarks for comparison: We compare the performance of the On-DS algorithm with a few typical benchmarks to evaluate the impact of base traffic prediction error, the benefit of updating the prediction in real time, and the impact of deferrable traffic's penetration level. We thus consider the followings five cases in our experiments:

- (0) *Offline demand shaping w/ Off-DS algorithm.* We use “real” trace for future base traffic and use arrival information recorded from case (1) below for DAs. Applied with Off-DS algorithm, this case gives the optimal performance used as benchmark to characterize the gap of other cases.
- (1) *Online demand shaping w/ On-DS algorithm.* We make prediction for both DAs' arrival and base traffic in the future. Prediction is updated at each timeslot. We run On-DS algorithm to schedule traffic.
- (2) *Online demand shaping w/ exact information for base traffic and w/o exact information for DAs.* We use “real” trace for base traffic and prediction for DAs. We apply On-DS algorithm. Comparison of case (2) with case (1) shows the impact of uncertainty in base traffic.
- (3) *Demand shaping w/ updating prediction of base traffic and w/ exact information for DAs.* We use DAs arrival information recorded from case (1). Instead of applying virtual deferrable traffic, we schedule traffic profiles for all the future deferrable traffic. Since the exact base traffic information is not available, we updated base traffic prediction at each timeslot. Comparison of case (3) and (1) shows the impact of uncertainty in DAs arrival prediction.
- (4) *Demand shaping w/o updating prediction of base traffic and w/ exact information for DAs.* We

use prediction of the base traffic at the beginning ($t = 1$) without further updating, and use arrival information recorded from case (1) for DAs. This case shows how the online algorithm benefits from updating prediction at each timeslot.

We use the metric of relative gap $G_r(d) = (V(d) - V(d^0))/V(d^0)$ to evaluate the performance, where d^0 is the results obtained from case (0). Also notice that when d is calculated based on case (1), $G_r(d)$ becomes G_r^{on} in Theorem 7.16.

7.3.6.2 Experiment Results

Considering randomness in DAs' arrivals, base traffic prediction, and deciding traffic profiles for discrete DAs, we run simulation for 10 times, and take the average as the final result to present.

Convergence Speed: We first run a case of randomly generated 143 continuous DAs and 150 discrete DAs by Off-DS algorithm with different numbers of iterations ranging from 1 to 40 for 10 times. Because of the random process in choosing traffic profiles for discrete DAs, we observe oscillation in objective function values for each individual run. However, the oscillation has a trend of diminishing as the more iterations are implemented, with satisfying enough results generated from running 20 to 30 iterations. See Fig.7.6 for the results. We will implement a number of 30 iterations for each decision to be made in the rest of simulation.

Impact of Base Traffic Prediction Error: As described in Section 7.3.6.1, we can tune the variance σ to emulate situations with different prediction errors in base traffic. As Fig. 7.7 shows, with updated prediction, case (1)'s performance is barely affected by the increasing prediction error, keeping its relative gap under 5%. This is almost as good as that of case (2) with perfect base traffic information. We can also see from the performance of case (3) the pure impact from prediction errors, while case (4) gives an example showing what happens if there is no updated prediction.

Impact of Penetration Level of Discrete DAs: In this case, we fix the prediction error in base traffic at $\sigma^2 = 40$ and the average number of DAs' arrival at each timeslot at $\lambda_p = 4$. We then tune the penetration level of discrete DAs from 25% to 75% with granularity of 5%. As shown in Fig. 7.8, the relative gap maintain relatively unaffected by the changes of discrete DAs whose penetration has increased by three

times. Here, we do not observe a decreasing relative gap mainly because the gap is *not* monotonically decreasing with number of N'' .

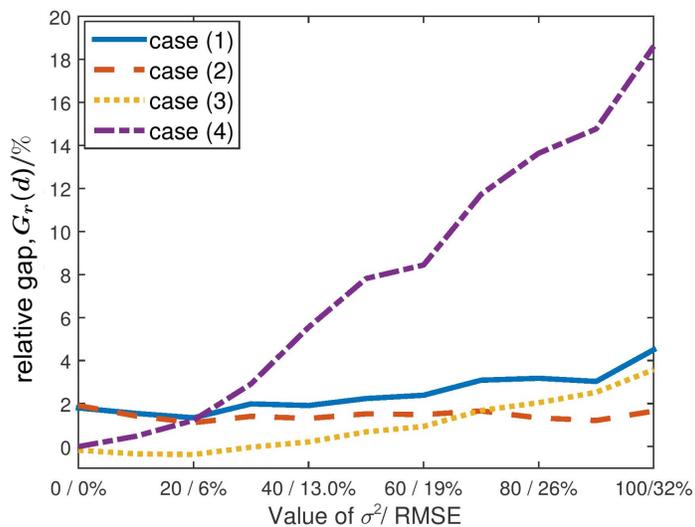


Figure 7.7: Base traffic prediction error has little impact on online algorithms with updated base traffic prediction.

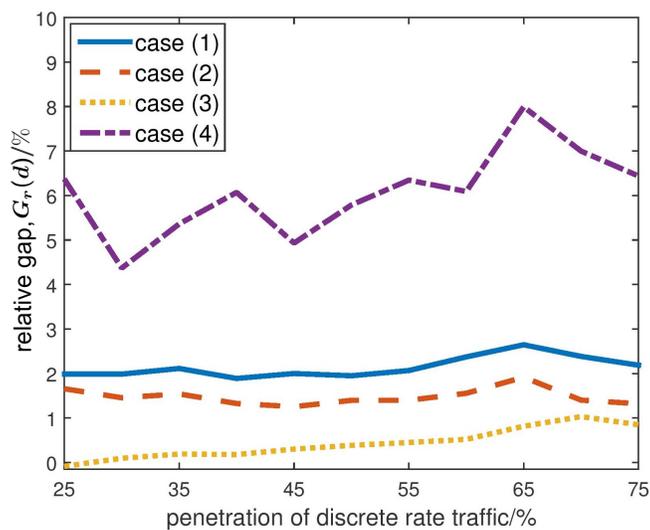


Figure 7.8: Increasing penetration level of deferrable traffic does not influence the relative gap of online algorithms.

7.3.7 Conclusion

We have formulated demand shaping in cellular networks as an optimization problem that minimizes the time variation in aggregate traffic subject to the rate and time requirements of the applications. We design a distributed and randomized offline demand shaping algorithm under complete traffic information and prove its almost surely convergence. We then consider a realistic setting with incomplete information where we can only predict future traffic to a certain degree of accuracy, and design an online demand shaping algorithm that updates the schedules of deferrable applications each time new information is available, based on solving at each timeslot an optimization problem over a shrinking horizon from the current time to the end of the day. We compare the performance of the online algorithm against the optimal offline algorithm analytically and numerically. As future work, we are investigating to integrate the incentive mechanisms such as the smart data pricing into the demand shaping algorithm design. We also plan to develop a platform to enable automatic demand shaping in cellular networks and investigate the related practical issues.

Bibliography

- [1] Airtel launches unlimited-usage night plans for calls, internet. <http://businesstoday.intoday.in/story/airtel-night-plans-unlimited-usage-for-calls-internet/1/205272.html>.
- [2] AT&T still throttles “unlimited data”—even when network not congested. <https://arstechnica.com/information-technology/2014/12/att-still-throttles-unlimited-data-even-when-network-not-congested/>.
- [3] How much bandwidth does Skype need? <https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need>.
- [4] Netflix internet connection speed recommendations. <https://help.netflix.com/en/node/306>.
- [5] Sprint night and weekend minutes. http://shop2.sprint.com/en/stores/popups/voice_nights_weekends_7pm_popup.shtml.
- [6] T-mobile now throttling mobile hotspots when network is congested. <https://arstechnica.com/information-technology/2016/10/t-mobile-now-throttling-mobile-hotspots-when-network-is-congested/>.
- [7] Telkom night surfer plan. <http://www.telkonomie.co.za/plans/prepaid-data/60gbpromo/>.
- [8] Verizon nationwide for business plans. <http://business.verizonwireless.com/content/b2b/en/shop-business-products/business-plans/nationwide-for-business.html>.
- [9] Verizon wireless to slow down users with unlimited 4G LTE plans. <https://arstechnica.com/information-technology/2014/07/verizon-wireless-to-slow-down-users-with-unlimited-4g-lte-plans/>.
- [10] Filip Andr n, Benoit Bletterie, Serdar Kadam, Panos Kotsampopoulos, and Christof Bucher. On the stability of local voltage control in distribution networks with a high penetration of inverter-based generation. *IEEE Trans. on Industrial Electronics*, 62(4):2519–2529, 2015.
- [11] Daniele Angelosante, Juan Andr s Bazerque, and Georgios B Giannakis. Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1 -norm. *IEEE Trans. on Signal Processing*, 58(7):3436–3447, Jul. 2010.
- [12] A Araposthatis, S Sastry, and P Varaiya. Analysis of power-flow equation. *International Journal of Electrical Power & Energy Systems*, 3(3):115–126, 1981.

- [13] Kyri Baker, Andrey Bernstein, Emiliano Dall’Anese, and Changhong Zhao. Network-cognizant voltage droop control for distribution grids. *IEEE Trans. on Power Systems*, 33(2):2098–2108, 2018.
- [14] Aruna Balasubramanian, Ratul Mahajan, and Arun Venkataramani. Augmenting mobile 3G using WiFi. *Proc. of International Conference on Mobile Systems, Applications, and Services*, pages 209–222, 2010.
- [15] Jason Bank and Joshua Hambrick. Development of a high resolution, real time, distribution-level metering system and associated visualization modeling, and data analysis functions. National Renewable Energy Laboratory, Tech. Rep. NREL/TP-5500-56610, May 2013.
- [16] Mesut E Baran and Felix F Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans. on Power Delivery*, 4(2):1401–1407, 1989.
- [17] Mesut E Baran and Felix F Wu. Optimal capacitor placement on radial distribution systems. *IEEE Trans. on Power Delivery*, 4(1):725–734, 1989.
- [18] Mesut E Baran and Felix F Wu. Optimal sizing of capacitors placed on a radial distribution system. *IEEE Trans. on Power Delivery*, 4(1):735–743, 1989.
- [19] Mohammadhafez Bazrafshan and Nikolaos Gatsis. Decentralized stochastic optimal power flow in radial networks with distributed generation. *IEEE Trans. on Smart Grid*, 8(2):787–801, 2017.
- [20] Arthur R Bergen. *Power Systems Analysis*. Pearson Education India, 2009.
- [21] Arthur R Bergen and David J Hill. A structure preserving model for power system stability analysis. *IEEE Trans. on Power Apparatus and Systems*, (1):25–35, 1981.
- [22] Andrey Bernstein, Niek J Bouman, and Jean-Yves Le Boudec. Design of resource agents with guaranteed tracking properties for real-time control of electrical grids. arXiv preprint arXiv: 1511.08628.
- [23] Andrey Bernstein and Emiliano Dall’Anese. Linear power-flow models in multiphase distribution networks. *Proc. of 2017 IEEE PES ISGT-Europe*, pages 1–6, 2017.
- [24] Andrey Bernstein, Cong Wang, Emiliano Dall’Anese, Jean-Yves Le Boudec, and Changhong Zhao. Load-flow in multiphase distribution networks: Existence, uniqueness, and linear models. *arXiv preprint arXiv:1702.03310*, 2017.
- [25] Dimitri P Bertsekas and John N Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*, volume 23. Prentice hall Englewood Cliffs, NJ, 1989.
- [26] Norman Biggs. *Algebraic Graph Theory*. Cambridge university press, 1993.
- [27] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175 – 308, 2006.
- [28] Saverio Bolognani and Florian Dörfler. Fast power system analysis via implicit linearization of the power flow manifold. *Proc. of IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 402–409, 2015.
- [29] Saverio Bolognani and Sandro Zampieri. A distributed control strategy for reactive power compensation in smart microgrids. *IEEE Trans. on Automatic Control*, 58(11):2818–2833, 2013.

- [30] Stephen Boyd and Almir Mutapcic. Subgradient methods. *Lecture notes of EE364b, Stanford University, Winter Quarter*, 2007, 2006.
- [31] Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2004.
- [32] Lijun Chen, Na Li, Libin Jiang, and Steven H Low. Optimal demand response: Problem formulation and deterministic case. *Control and Optimization Theory for Electric Smart Grids*, 2012.
- [33] Lijun Chen and Seungil You. Reverse and forward engineering of frequency control in power networks. *IEEE Trans. on Automatic Control*, 62(9):4631–4638, 2017.
- [34] Man Hon Cheung, Fen Hou, Jianwei Huang, and Richard Southwell. Congestion-aware dns for integrated cellular and Wi-Fi networks. *IEEE Journal on Selected Areas in Communications*, 35(6):1269–1281, 2017.
- [35] H-D Chiang and Mesut E Baran. On the existence and uniqueness of load flow solution for radial distribution power networks. *IEEE Trans. on Circuits and Systems*, 37(3):410–416, 1990.
- [36] Konstantina Christakou, Jean-Yves LeBoudec, Mario Paolone, and Dan-Cristian Tomozei. Efficient computation of sensitivity coefficients of node voltages and line currents in unbalanced radial electrical distribution networks. *IEEE Trans. on Smart Grid*, 4(2):741–750, 2013.
- [37] Emiliano Dall’Anese, Sairaj V Dhople, and Georgios B Giannakis. Optimal dispatch of photovoltaic inverters in residential distribution systems. *IEEE Trans. on Sustainable Energy*, 5(2):487–497, 2014.
- [38] Emiliano Dall’Anese, Pierluigi Mancarella, and Antonello Monti. Unlocking flexibility: Integrated optimization and control of multienergy systems. *IEEE Power and Energy Magazine*, 15(1):43–52, 2017.
- [39] Emiliano Dall’Anese and Andrea Simonetto. Optimal power flow pursuit. *IEEE Trans. on Smart Grid*, 9(2):942–952, 2018.
- [40] Florian Dörfler and Francesco Bullo. On the critical coupling for kuramoto oscillators. *SIAM Journal on Applied Dynamical Systems*, 10(3):1070–1099, 2011.
- [41] Florian Dörfler, Michael Chertkov, and Francesco Bullo. Synchronization in complex oscillator networks and smart grids. *Proc. of the National Academy of Sciences*, 110(6):2005–2010, 2013.
- [42] Florian Dörfler, John W Simpson-Porco, and Francesco Bullo. Breaking the hierarchy: Distributed control and economic optimality in microgrids. *IEEE Trans. on Control of Network Systems*, 3(3):241–253, 2016.
- [43] M. Farivar, C. R. Clarke, S. H. Low, and K. M. Chandy. Inverter var control for distribution systems with renewables. *Proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 457–462, 2011.
- [44] Masoud Farivar, Lijun Chen, and Steven Low. Equilibrium and dynamics of local voltage control in distribution systems. *Proc. of IEEE Conference in Decision and Control (CDC)*, pages 4329–4334, December 2013.
- [45] Masoud Farivar, Xinyang Zhou, and Lijun Chen. Local voltage control in distribution systems: An incremental control algorithm. *Proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 732–737, 2015.

- [46] Oded Galor. *Discrete Dynamical Systems*. Springer Science & Business Media, 2007.
- [47] Lingwen Gan, Ufuk Topcu, and Steven H Low. Stochastic distributed protocol for electric vehicle charging with discrete charging rate. *Proc. of Power and Energy Society General Meeting*, pages 1–8, 2012.
- [48] Lingwen Gan, Adam Wierman, Ufuk Topcu, Niangjun Chen, and Steven H Low. Real-time deferrable load control: Handling the uncertainties of renewable generation. *Proc. of International Conference on Future Energy Systems*, pages 113–124, 2013.
- [49] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford university press, 2001.
- [50] Swaroop S Guggilam, Emiliano Dall’Anese, Yu Christine Chen, Sairaj V Dhople, and Georgios B Giannakis. Scalable optimization methods for distribution networks with high PV integration. *IEEE Trans. on Smart Grid*, 7(4):2061–2070, July 2016.
- [51] Sangtae Ha, Soumya Sen, Carlee Joe-Wong, Youngbin Im, and Mung Chiang. Tube: Time-dependent pricing for mobile data. *ACM SIGCOMM Computer Communication Review*, 42(4):247–258, 2012.
- [52] Sangtae Ha, Soumya Sen, Carlee Joe-Wong, Youngbin Im, and Muung Chiang. Tube survey questions and demographics. http://www.princeton.edu/~cjoe/TUBE_Survey.pdf, Jan 2012.
- [53] Seung-Yeal Ha, Eunhee Jeong, and Moon-Jin Kang. Emergent behaviour of a generalized viscek-type flocking model. *Nonlinearity*, 23(12):3139, 2010.
- [54] Mohammad H Hajiesmaili, Lei Deng, Minghua Chen, and Zongpeng Li. Incentivizing device-to-device load balancing for cellular networks: An online auction design. *IEEE Journal on Selected Areas in Communications*, 35(2):265–279, 2017.
- [55] Cisco Visual Networking Index. Global mobile data traffic forecast update, 2016–2021. 2017.
- [56] Chitika Insights. Hour-by-hour examination: Smartphone, tablet, and desktop usage rates, 2013.
- [57] George Iosifidis, Lin Gao, Jianwei Huang, and Leandros Tassioulas. A double-auction mechanism for mobile data-offloading markets. *IEEE/ACM Trans. on Networking (TON)*, 23(5):1634–1647, 2015.
- [58] IRENA. Renewable power generation costs in 2017. 2017.
- [59] Ali Jadbabaie, Nader Motee, and Mauricio Barahona. On the stability of the kuramoto model of coupled nonlinear oscillators. *Proc. of American Control Conference (ACC)*, 5:4296–4301, June 2004.
- [60] Pedram Jahangiri and Dionysios C Aliprantis. Distributed Volt/VAr control by PV inverters. *IEEE Trans. on Power Systems*, 28(3):3429–3439, 2013.
- [61] Huaiguang Jiang, Yingchen Zhang, Eduard Muljadi, Jun Zhang, and Wenzhong Gao. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. *IEEE Trans. on Smart Grid*, 2016.
- [62] Libin Jiang, Shyam Parekh, and Jean Walrand. Time-dependent network pricing and bandwidth trading. *Proc. of IEEE Network Operations and Management Symposium Workshops*, pages 193–200, 2008.

- [63] Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. Time-dependent broadband pricing: Feasibility and benefits. *Proc. of International Conference on Distributed Computing Systems (ICDCS)*, pages 288–298, 2011.
- [64] Hristiyan Kanchev, Di Lu, Frederic Colas, Vladimir Lazarov, and Bruno Francois. Energy management and operational planning of a microgrid with a PV-based active generator for smart grid applications. *IEEE Trans. on Industrial Electronics*, 58(10):4583–4592, 2011.
- [65] Vassilis Kekatos, Gang Wang, Antonio J Conejo, and Georgios B Giannakis. Stochastic reactive power management in microgrids with renewables. *IEEE Trans. on Power Systems*, 30(6):3386–3395, 2015.
- [66] Vassilis Kekatos, Liang Zhang, Georgios B Giannakis, and Ross Baldick. Voltage regulation algorithms for multiphase power distribution grids. *IEEE Trans. on Power Systems*, 31(5):3913–3923, 2016.
- [67] Hassan K Khalil. *Nonlinear Systems*. Prentice-Hall, New Jersey, 2002.
- [68] Istvan Z. Kiss, Yumei Zhai, and John L. Hudson. Emerging Coherence in a Population of Chemical Oscillators. *Science*, 296(5573):1676–1678, May 2002.
- [69] Jayash Koshal, Angelia Nedić, and Uday V Shanbhag. Multiuser optimization: Distributed algorithms and error analysis. *SIAM J. on Optimization*, 21(3):1046–1081, 2011.
- [70] W. H. Kwon and A.E. Pearson. A modified quadratic cost problem and feedback stabilization of a linear system. *IEEE Trans. on Automatic Control*, 22(5):838–842, Oct 1977.
- [71] Javad Lavaei and Steven H Low. Zero duality gap in optimal power flow problem. *IEEE Trans. on Power Systems*, 27(1):92–107, 2012.
- [72] Joohyun Lee, Yung Yi, Song Chong, and Youngmi Jin. Economics of WiFi offloading: Trading delay for cellular capacity. *IEEE Trans. on Wireless Communications*, 13(3):1540–1554, 2014.
- [73] Kyunghan Lee, Joohyun Lee, Yung Yi, Injong Rhee, and Song Chong. Mobile data offloading: How much can WiFi deliver? *IEEE/ACM Trans. on Networking*, 21(2):536–550, 2013.
- [74] Na Li. A market mechanism for electric distribution networks. *Proc. of IEEE Annual Conference on Decision and Control (CDC)*, pages 2276–2282, 2015.
- [75] Na Li, Lijun Chen, and Steven H Low. Optimal demand response based on utility maximization in power networks. *Proc. of IEEE Power and Energy Society General Meeting*, pages 1–8, 2011.
- [76] Na Li, Lijun Chen, and Steven H Low. Optimal demand response based on utility maximization in power networks. *Proc. of IEEE Power Engineering Society General Meeting*, July 2011.
- [77] Na Li, Guannan Qu, and Munther Dahleh. Real-time decentralized voltage control in distribution networks. *Proc. of IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 582–588, 2014.
- [78] Na Li, Changhong Zhao, and Lijun Chen. Connecting automatic generation control and economic dispatch from an optimization view. *IEEE Trans. on Control of Network Systems*, 3(3):254–264, 2016.

- [79] Sen Li, Wei Zhang, Jianming Lian, and Karanjit Kalsi. Market-based coordination of thermostatically controlled loads-Part I: A mechanism design formulation. *IEEE Trans. on Power Systems*, 31(2):1170–1178, 2016.
- [80] Y. Liu, J. Bebic, B. Kroposki, J. de Bedout, and W. Ren. Distribution system voltage performance analysis for high-penetration PV. *Proc. of IEEE Energy 2030 Conference*, Nov. 2008.
- [81] Zhiyuan Liu, Jeries Shihadeh, Seungil You, Guohui Ding, Xinyang Zhou, and Lijun Chen. Signal-anticipating in local voltage control in distribution systems. *arXiv*, 2018.
- [82] Steven H Low. Convex relaxation of optimal power flow—Part I: Formulations and equivalence. *IEEE Trans. on Control of Network Systems*, 1(1):15–27, 2014.
- [83] Steven H Low. Convex relaxation of optimal power flow—Part II: Exactness. *IEEE Trans. on Control of Network Systems*, 1(2):177–189, 2014.
- [84] Sindri Magnússon, Carlo Fischione, and Na Li. Voltage control using limited communication. *IFAC-PapersOnLine*, 50(1):1–6, 2017.
- [85] Sindri Magnússon, Pradeep Chathuranga Weeraddana, and Carlo Fischione. A distributed approach for the optimal power-flow problem based on ADMM and sequential convex approximations. *IEEE Trans. on Control of Network Systems*, 2(3):238–253, 2015.
- [86] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Basar. Dependable demand response management in the smart grid: A Stackelberg game approach. *IEEE Trans. on Smart Grid*, 4(1):120–132, 2013.
- [87] Enrique Mallada and Steven H Low. Distributed frequency-preserving optimal load control. *IFAC Proceedings Volumes*, 47(3):5411–5418, 2014.
- [88] Jacob Mattingley, Yang Wang, and Stephen Boyd. Receding horizon control. *IEEE Control Systems*, 31(3):52–65, 2011.
- [89] Angus McCrone, Ulf Moslener, Francoise d’Estais, Eric Usher, and Christine Grüning. Global trends in renewable energy investment 2016. *Frankfurt School UNEP Collaborating Centre for Climate and Sustainable Energy Finance*, 2016.
- [90] Fidan Mehmeti and Thrasyvoulos Spyropoulos. Is it worth to be patient? Analysis and optimization of delayed mobile data offloading. *Proc. of IEEE INFOCOM*, pages 2364–2372, 2014.
- [91] Amir-Hamed Mohsenian-Rad, Vincent WS Wong, Juri Jatskevich, Robert Schober, and Alberto Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans. on Smart Grid*, 1(3):320–331, 2010.
- [92] Daniel K Molzahn, Florian Dörfler, Henrik Sandberg, Steven H Low, Sambuddha Chakrabarti, Ross Baldick, and Javad Lavaei. A survey of distributed optimization and control algorithms for electric power systems. *IEEE Trans. on Smart Grid*, 2017.
- [93] Ion Necoara and Valentin Nedelcu. Rate analysis of inexact dual first-order methods application to dual decomposition. *IEEE Trans. on Automatic Control*, 59(5):1232–1243, 2014.
- [94] Derek A Paley, Naomi Ehrich Leonard, Rodolphe Sepulchre, Daniel Grunbaum, and Julia K Parrish. Oscillator models and collective motion. *IEEE Control Systems*, 27(4):89–105, Aug 2007.

- [95] Francesca Parise, Marcello Colombino, Sergio Grammatico, and John Lygeros. Mean field constrained charging policy for large populations of plug-in electric vehicles. *Proc. of IEEE Annual Conference on Decision and Control (CDC)*, pages 5101–5106, 2014.
- [96] I Ch Paschalidis and John N Tsitsiklis. Congestion-dependent pricing of network services. *IEEE/ACM Trans. on Networking*, 8(2):171–184, 2000.
- [97] Qiuyu Peng and Steven H Low. Distributed optimal power flow algorithm for radial networks, I: Balanced single phase case. *IEEE Trans. on Smart Grid*, 2016.
- [98] Boris T Polyak. *Introduction to Optimization*. Optimization Software, Inc, New York, 1987.
- [99] Florian A Potra and Stephen J Wright. Interior-point methods. *Journal of Computational and Applied Mathematics*, 124(1):281–302, 2000.
- [100] S Sundhar Ram, A Nedić, and Venugopal V Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, 2009.
- [101] REN21. Renewables 2017 global status report. 2017.
- [102] Brett A Robbins, Christoforos N Hadjicostis, and Alejandro D Domínguez-García. A two-stage distributed architecture for voltage control in power distribution systems. *IEEE Trans. on Power Systems*, 28(2):1470–1482, 2013.
- [103] R Tyrrell Rockafellar and Roger J-B Wets. Variational analysis. *Springer Science & Business Media*, 317, 2009.
- [104] Gesualdo Scutari, Daniel P Palomar, Francisco Facchinei, and Jong-shi Pang. Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine*, 27(3):35–49, 2010.
- [105] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. Pricing data: A look at past proposals, current plans, and future trends. *arXiv preprint arXiv:1201.4197*, 2012.
- [106] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. Smart data pricing (SDP): Economic solutions to network congestion. *SIGCOMM eBook on Recent Advances in Networking*, 2013.
- [107] Wenbo Shi, Xiaorong Xie, Chi Cheng Chu, and Rajit Gadh. Distributed optimal energy management in microgrids. *IEEE Trans. on Smart Grid*, 6(3):1137–1146, 2015.
- [108] Thomas Allen Short. *Electric Power Distribution Handbook*. CRC press, 2014.
- [109] Andrea Simonetto and Geert Leus. Double smoothing for time-varying distributed multiuser optimization. *Proc. of IEEE Global Conference on Signal and Information Processing*, pages 852–856, Dec. 2014.
- [110] John W Simpson-Porco, Florian Dörfler, and Francesco Bullo. Voltage stabilization in microgrids via quadratic droop control. *IEEE Trans. on Automatic Control*, 62(3):1239–1253, 2017.
- [111] Standards Coordinating Committee 21 of Institute of Electrical and Electronics Engineers, Inc., IEEE Standard, P1547.8TM/D8. Recommended practice for establishing methods and procedures that provide supplemental support for implementation strategies for expanded use of IEEE Standard 1547. *IEEE ballot document*, Aug. 2014.

- [112] Henry Stark and John W Woods. *Probability and Random Processes with Applications to Signal Processing: International Edition*. Pearson Higher Ed, 2014.
- [113] Steven H. Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, 2000.
- [114] Steven Henry Strogatz. *Sync: the Emerging Science of Spontaneous Order*. Theia, New York, 2003.
- [115] Petr Šulc, Scott Backhaus, and Michael Chertkov. Optimal distributed control of reactive power via the alternating direction method of multipliers. *IEEE Trans. on Energy Conversion*, 29(4):968–977, 2014.
- [116] John Tadrous, Atilla Eryilmaz, and Hesham El Gamal. Proactive resource allocation: Harnessing the diversity and multicast gains. *IEEE Trans. on Information Theory*, 59(8):4833–4854, 2013.
- [117] Konstantin Turitsyn, Petr Šulc, Scott Backhaus, and Michael Chertkov. Options for control of reactive power by distributed photovoltaic generators. *Proc. of the IEEE*, 99(6):1063–1073, 2011.
- [118] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nat Rev Neurosci*, 2(4):229–239, Apr 2001.
- [119] Richard S Varga. *Matrix Iterative Analysis*, volume 27. Springer Science & Business Media, 2009.
- [120] Hal R Varian. *Microeconomic analysis*. WW Norton, 1992.
- [121] Alexandra Von Meier. *Electric Power Systems: A Conceptual Introduction*. John Wiley & Sons, 2006.
- [122] Evangelos Vrettos, Frauke Oldewurtel, and Göran Andersson. Robust energy-constrained frequency reserves from aggregations of commercial buildings. *IEEE Trans. on Power Systems*, 31(6):4272–4285, 2016.
- [123] Y. Wang, X. Lin, and M. Pedram. A Stackelberg game-based optimization framework of the smart grid with distributed PV power generations and data centers. *IEEE Trans. on Energy Conversion*, 29(4):978–987, Dec. 2014.
- [124] J.Y. Wen, Q.H. Wu, D.R. Turner, S.J. Cheng, and J. Fitch. Optimal coordinated voltage control for power system voltage stability. *IEEE Trans. on Power Systems*, 19(2):1115–1122, 2004.
- [125] Kurt Wiesenfeld, Pere Colet, and Steven H. Strogatz. Frequency locking in josephson arrays: Connection with the kuramoto model. *Phys. Rev. E*, 57:1563–1569, Feb 1998.
- [126] Arthur T Winfree. *The Geometry of Biological Time*, volume 12 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2001.
- [127] Chenye Wu, Gabriela Hug, and Soumya Kar. Distributed voltage regulation in distribution power grids: Utilizing the photovoltaics inverters. *Proc. of American Control Conference (ACC)*, pages 2725–2731, 2017.
- [128] Baosen Zhang, Alejandro D Dominguez-Garcia, and David Tse. A local control approach to voltage regulation in distribution networks. *Proc. of IEEE North American Power Symposium (NAPS)*, 2013, pages 1–6, 2013.

- [129] Liang Zhang. *Smart Data Pricing in Wireless Data Networks: An Economic Solution to Congestion*. PhD thesis, The Hong Kong Polytechnic University, 2016.
- [130] Changhong Zhao, Ufuk Topcu, Na Li, and Steven Low. Design and stability of load-side primary frequency control in power systems. *IEEE Trans. on Automatic Control*, 59(5):1177–1189, 2014.
- [131] Shizhen Zhao, Xiaojun Lin, and Minghua Chen. Peak-minimizing online EV charging: Price-of-uncertainty and algorithm robustification. *Proc. of IEEE Conference on Computer Communications (INFOCOM)*, pages 2335–2343, 2015.
- [132] Xinyang Zhou and Lijun Chen. Demand shaping in cellular networks. *Proc. of IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 621–628, 2014.
- [133] Xinyang Zhou and Lijun Chen. A new perspective to synchronization in networks of coupled oscillators: Reverse engineering and convex relaxation. *IFAC-PapersOnLine*, 48(22):40–45, 2015.
- [134] Xinyang Zhou and Lijun Chen. An incremental local algorithm for better voltage control in distribution networks. *Proc. of IEEE Conference on Decision and Control (CDC)*, pages 2396–2402, 2016.
- [135] Xinyang Zhou and Lijun Chen. Demand shaping in cellular networks. *IEEE Trans. on Control of Network Systems*, 2018.
- [136] Xinyang Zhou, Emiliano Dall’Anese, and Lijun Chen. Online stochastic control of discrete loads in distribution grids. *arXiv preprint arXiv:1711.09953*, 2017.
- [137] Xinyang Zhou, Emiliano Dall’Anese, Lijun Chen, and Kyri Baker. Incentive-based voltage regulation in distribution networks. *Proc. of American Control Conference (ACC)*, pages 2732–2738, 2017.
- [138] Xinyang Zhou, Emiliano Dall’Anese, Lijun Chen, and Andrea Simonetto. An incentive-based online optimization framework for distribution grids. *IEEE Trans. on Automatic Control*, 2017.
- [139] Xinyang Zhou, Masoud Farivar, and Lijun Chen. Pseudo-gradient based local voltage control in distribution networks. *Proc. of IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 173–180, 2015.
- [140] Xinyang Zhou, Zhiyuan Liu, Emiliano Dall’Anese, and Lijun Chen. Stochastic dual algorithm for voltage regulation in distribution networks with discrete loads. *Proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2017.
- [141] Xinyang Zhou, Zhiyuan Liu, Masoud Farivar, Lijun Chen, and Steven Low. Reverse and forward engineering of local voltage control in distribution networks. *arXiv preprint arXiv:1801.02015*, 2018.
- [142] Xinyang Zhou, Jie Tian, Lijun Chen, and Emiliano Dall’Anese. Local voltage control in distribution networks: A game-theoretic perspective. *Proc. of North American Power Symposium (NAPS)*, pages 1–6, 2016.
- [143] Hao Zhu and Hao Jan Liu. Fast local voltage control under limited reactive power: Optimality and stability analysis. *IEEE Trans. on Power Systems*, 31(5):3794–3803, 2016.
- [144] Minghui Zhu and Sonia Martínez. On distributed convex optimization under inequality and equality constraints. *IEEE Trans. on Automatic Control*, 57(1):151–164, 2012.

- [145] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. on Power Systems*, 26(1):12–19, 2011.