Distributed Optimization and Market Analysis of Networked Systems

by

Ermin Wei

B.S., Computer Engineering
B.S., Mathematics
B.S., Finance
Minor in German Language, Literature, and Culture
University of Maryland, College Park (2008)
M.S., Electrical Engineering and Computer Science
Massachusetts Institute of Technology (2010)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

Author ....
Department of Electrical Engineering and Computer Science
August 29, 2014

Certified by ....
Asu Ozdaglar
Professor
Thesis Supervisor

Accepted by ....
Leslie Kolodziejski
Chairman, Department Committee on Graduate Theses

Distributed Optimization and Market Analysis of Networked Systems

by

Ermin Wei

Submitted to the Department of Electrical Engineering and Computer Science
on August 29, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

In the interconnected world of today, large-scale multi-agent networked systems are ubiquitous. This thesis studies two classes of multi-agent systems, where each agent has local information and a local objective function. In the first class of systems, the agents are collaborative and the overall objective is to optimize the sum of local objective functions. This setup represents a general family of separable problems in large-scale multi-agent convex optimization systems, which includes the LASSO (Least-Absolute Shrinkage and Selection Operator) and many other important machine learning problems. We propose fast fully distributed both synchronous and asynchronous ADMM (Alternating Direction Method of Multipliers) based methods. Both of the proposed algorithms achieve the best known rate of convergence for this class of problems, $O(1/k)$, where k is the number of iterations. This rate is the first rate of convergence guarantee for asynchronous distributed methods solving separable convex problems. For the synchronous algorithm, we also relate the rate of convergence to the underlying network topology.

The second part of the thesis focuses on the class of systems where the agents are only interested in their local objectives. In particular, we study the market interaction in the electricity market. Instead of the traditional supply-follow-demand approach, we propose and analyze a systematic multi-period market framework, where both (price-taking) consumers and generators locally respond to price. We show that this new market interaction at competitive equilibrium is efficient and the improvement in social welfare over the traditional market can be unbounded. The resulting system, however, may feature undesirable price and generation fluctuations, which imposes significant challenges in maintaining reliability of the electricity grid. We first establish that the two fluctuations are positively correlated. Then in order to reduce both fluctuations, we introduce an explicit penalty on the price fluctuation. The penalized problem is shown to be equivalent to the existing system with storage and can be implemented in a distributed way, where each agent locally responds to price. We analyze the connection between the size of storage, consumer utility function properties and generation fluctuation in two scenarios: when demand is inelastic, we can explicitly characterize the optimal storage access policy and the generation fluctuation; when demand is elastic, the relationship between concavity and generation fluctuation is studied.

Thesis Supervisor: Asu Ozdaglar
Title: Professor

Acknowledgments

The route to a PhD is not an easy one, and completing this thesis would not be possible without all the wonderful people I managed to encounter in my life. First and foremost, I'd like to thank my super supportive and caring thesis advisor, Prof. Asu Ozdaglar. I am very fortunate to have her as my advisor. The numerous meetings we had and even casual hall-way interactions all have shed lights into my life and countless times have guided me out of the dark abyss called, "I'm stuck". Her awesome mentorship, on both personal and professional fronts, has inspired me to become an academician myself, because what I learned here is too valuable to be kept to myself.

The interactions I had with my thesis committee members, Prof. Munther Dahleh, Prof. Ali Jadbabaie and Prof. Steven Low, have been all tremendously beneficial. I want to thank them for the time and dedication in helping with the development of the thesis at various stages. I also want to thank my collaborators, Prof. Azarakhsh Malekian, Shimrit Shtern for the stimulating discussions we had and ingenious works they did. I am grateful to Prof. Daron Acemoglu, Prof. Janusz Bialek, Prof. Eilyan Bitar, Dr. Ali Kakhbod, Prof. Na Li, Prof. Sean Meyn, Dr. Mardavij Roozbehani, Prof. Lang Tong, Prof. Jiankang Wang for all the fruitful discussion we had while developing my thesis. My officemates Elie Adam, Prof. Ozan Candogan, Kimon Drakopoulos, Jenny (Ji Young) Lee, Yongwhan Lim, Milashini Nambiar have all been wonderful and especially cool with my office decorations.

I would like to thank my friends at MIT for all the incredible experience we had together. The fun we had and the stories we shared will always be able to warm my heart and put a smile on my face in the future, no matter how much time passes by. A special thank goes to my family for the continuous love and support, always being there for me. The various lessons learned throughout this enriching PhD experience made me a better person and ready for the new roads awaiting.

Last but not least, I would like to thank you our funding agencies: National Science Foundation, Air Force Research Laboratory, Office of Naval Research and MIT Masdar project. Without their generous support, the work could not have been

accomplished.

Contents

1 Introduction **15**

1.1 Introduction 15

1.2 Distributed Optimization Algorithms 17

1.2.1 Motivation and Problem Formulation 17

1.2.2 Literature Review 20

1.2.3 Contributions 23

1.3 Electricity Market Analysis 24

1.3.1 Motivation 24

1.3.2 Literature Review 26

1.3.3 Contributions 28

2 Synchronous Distributed ADMM: Performance and Network Effects **31**

2.1 Preliminaries: Standard ADMM Algorithm 31

2.2 Problem Formulation 33

2.3 Distributed ADMM Algorithm 36

2.4 Convergence Analysis for Distributed ADMM Algorithm 39

2.4.1 Preliminaries 39

2.4.2 Convergence and Rate of Convergence 49

2.4.3 Effect of Network Structure on Performance 53

2.5 Summaries 60

3 Asynchronous Distributed ADMM Based Method 61

3.1 Introduction 61

3.2 Asynchronous ADMM Algorithm 63

3.2.1 Problem Formulation and Assumptions 64

3.2.2 Asynchronous Algorithm Implementation 65

3.2.3 Asynchronous ADMM Algorithm 67

3.2.4 Special Case: Distributed Multi-agent Optimization 69

3.3 Convergence Analysis for Asynchronous ADMM Algorithm 73

3.3.1 Preliminaries 75

3.3.2 Convergence and Rate of Convergence 94

3.4 Numerical Studies 108

3.5 Summaries 110

4 Market Analysis of Electricity Market: Competitive Equilibrium 113

4.1 Model 114

4.1.1 Demand side 114

4.1.2 Supply side 115

4.1.3 Competitive Equilibrium 116

4.2 Characterizing the Competitive Equilibrium 116

4.2.1 Efficiency 117

4.2.2 Efficiency Improvement Over Traditional Grid 118

4.3 Summaries 120

5 Market Analysis of Electricity Market: Generation Fluctuation Reduction 121

5.1 Price and Generation Fluctuation 121

5.2 Price Fluctuation Penalized Problem 123

5.2.1 Fluctuation Penalized Problem Formulation 123

5.2.2 Properties of the Fluctuation Penalized Problem 127

5.2.3 Distributed Implementation 146

5.3 Social Welfare and Generation Fluctuation Analysis 149

5.3.1 Social Welfare Analysis 150
5.3.2 Inelastic Demand Analysis 153
5.3.3 Elastic Demand Two Period Analysis 161
5.4 Summaries 169

6 Conclusions **171**

6.1 Summary 171
6.2 Future Directions 172

List of Figures

1-1 Reformulation of problem (1.1). 18

3-1 Asynchronous algorithm illustration: When edge $(2, 3)$ is active, only agents 2 and 3 perform update, while all the other agents stay at their previous iteration value. 67

3-2 One sample objective function value and constraint violation ($Ax = 0$) of both methods against iteration count for the network as in Figure. 3-1. The dotted black lines denote 10% interval of the optimal value and feasibility violation less than 0.1. 109

3-3 One sample objective function value and feasibility violation of both methods against iteration count for the path graph. The dotted black lines denote 10% interval of the optimal value and constraint violation ($Ax = 0$) less than 0.1. 110

List of Tables

3.1 Number of iterations needed to reach 10% of optimal function value and feasibility violation less than 0.1 for the network as in Figure. 3-1. 110

3.2 Number of iterations needed to reach 10% of optimal function value and feasibility violation less than 0.1 for the path network. 111

Chapter 1

Introduction

1.1 Introduction

In the interconnected world of today, large-scale multi-agent networked systems are ubiquitous. Some examples include sensor networks, communication networks and electricity grid. While the structure of and interaction within these systems may vary drastically, the operations of these networks share some universal characteristics, such as being large-scale, consisting of heterogeneous agents with local information and processing power, whose goals are to achieve certain optimality (either locally or globally).

Motivated by the need to enhance system performance, this thesis studies multi-agent systems in the context of two types of networks. The first type of networks consists of cooperative agents, each of which has some local information, represented as a local (convex) cost function. The agents are connected via an underlying graph, which specifies the allowed communication, i.e., only neighbors in the graph can exchange information. The system wide goal is to collectively optimize the single global objective, which is the sum of local cost functions, by performing local computation and communication only. This general framework captures many important applications such as sensor networks, compressive sensing systems, and machine learning applications. To improve performance for this type of networks, the thesis will develop fast distributed optimization algorithms, both synchronous and asynchronous, provide convergence guarantees and analyze the dependence between algorithm performance, algorithm parameters and the underlying topology. These results will help in network designs and parameter choices can be recommended.

The second type of systems is where the agents are only interested in their local objective functions. We analyze this type of networks through an example of electricity market. The agents in this system are generators and consumers. One feature that distinguishes electricity market and other markets is that supply has to equal to demand at all times due to the physics of electricity flows.[1] The traditional electricity market treats demand side as passive and inelastic. Thus the generators have to follow precisely the demand profile. However, due to the physical nature of generators, the generation level cannot change much instantaneously. In order to ensure reliability against unanticipated variability in demand, large amount of reserves is maintained in the current system. Both the demand-following nature of generation and large reserve decrease the overall system efficiency level. To improve system efficiency, the new *smart grid* paradigm has proposed to include active consumer participation, referred to as *demand response* in the literature, by passing on certain real-time price type signals to exploit the flexibility in the demand. To investigate systematically the grid with demand response integration, we propose a multi-period general market based model, where the preferences of the consumers and the cost structures of generators are reflected in their respective utility functions and the interaction between the two sides is done through pricing signals. We then analyze the competitive equilibrium and show that it is efficient and can improve significantly over the traditional electricity market (without demand response). To control undesirable price and generation fluctuation over time, we introduce a price fluctuation penalty in the social welfare maximization problem, which enables us to trade off between social welfare and price fluctuation. We show that this formulation can reduce both price and generation fluctuation. This fluctuation penalized problem can be equivalently implemented via introducing storage, whose size corresponds to the penalty parame-

[1] An imbalance of supply and demand at one place can result in black-out in a much bigger area and hence pose significant challenge to the reliability of the grid.

ters. We then analyze the properties and derive distributed implementation for this fluctuation penalized problem. The connections between fluctuation and consumer utility function for both elastic and inelastic demand are also studied.

In the following two sections, we describe the two topics in more depth and outline our main contribution for each part.

1.2 Distributed Optimization Algorithms

1.2.1 Motivation and Problem Formulation

Many networks are large-scale and consist of agents with local information and processing power. There has been a much recent interest in developing control and optimization methods for multi-agent networked systems, where processing is done locally without any central coordination [2], [3], [5], [68], [69], [70], [36], [71], [6]. These *distributed multi-agent optimization problems* are featured in many important applications including optimization, control, compressive sensing and signal processing communities. Each of the distributed multi-agent optimization problems features a set $V = \{1, \ldots, N\}$ of agents connected by M undirected edges forming an underlying network $G = (V, E)$, where E is the set of edges. Each agent has access to a privately known *local objective (or cost) function* $f_i : \mathbb{R}^n \to \mathbb{R}$, which depends on the global decision variable x in $\mathbb{R}^n$ shared among all the agents. The system goal is to collectively solve a global optimization problem.

$$\begin{aligned} \min \sum_{i=1}^{N} f_i(x) \qquad (1.1) \\ s.t. \quad x \in X. \end{aligned}$$

To illustrate the importance of problem (1.1), we consider a machine learning

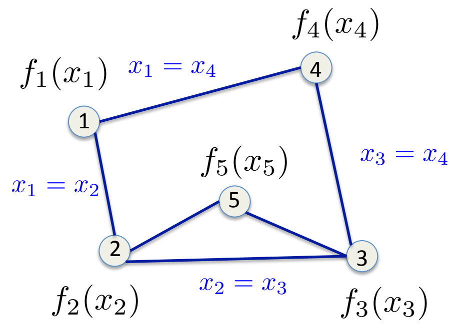

Figure 1-1: Reformulation of problem (1.1).

problem described as follows:

$$\min_x \sum_{i=1}^{N-1} l\left([W_i x - b_i]\right) + \pi \left|\left|x\right|\right|_1,$$

where W_i corresponds to the input sample data (and functions thereof), b_i represents the measured outputs, $W_i x - b_i$ indicates the prediction error and l is the loss function on the prediction error. Scalar π is nonnegative and it indicates the penalty parameter on complexity of the model. The widely used Least Absolute Deviation (LAD) formulation, the Least-Absolute Shrinkage and Selection Operator (Lasso) formulation and l_1 regularized formulations can all be represented by the above formulation by varying loss function l and penalty parameter π (see [1] for more details). The above formulation is a special case of the distributed multi-agent optimization problem (1.1), where $f_i(x) = l\left(W_i x - b_i\right)$ for $i = 1, \ldots, N-1$ and $f_N = \pi_2 \left|\left|x\right|\right|_1$. In applications where the data pairs $\left(W_i, b_i\right)$ are collected and maintained by different sensors over a network, the functions f_i are local to each agent and the need for a distributed algorithm arises naturally.

Problem (1.2) can be reformulated to facilitate distributed algorithm development by introducing a local copy x_i in $\mathbb{R}^n$ of the decision variable for each node i and imposing the constraint $x_i = x_j$ for all agents i and j with edge $(i, j) \in E$. The constraints represent the coupling across components of the decision variable imposed by the underlying information exchange network among the agents. Under the assumption that the underlying network is connected, each of the local copies are equal to each other at the optimal solution. We denote by $A \in \mathbb{R}^{Mn \times Nn}$ the *generalized edge-node incidence matrix* of network G, defined by $\tilde{A} \otimes I(n \times n)$, where $\tilde{A}$ is the standard edge-node incidence matrix in $\mathbb{R}^{M \times N}$. Each row in matrix $\tilde{A}$ represents an edge in the graph and each column represents a node. For the row corresponding to edge $e = (i, j)$, we have the $\tilde{A}_{ei} = 1$, $\tilde{A}_{ej} = -1$ and $\tilde{A}_{el} = 0$ for $l \neq i, j$. For instance, the edge-node incidence matrix of network in Fig. 1-1 is given by

$$\tilde{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix},$$

and the generalized edge-node incidence matrix is

$$A = \begin{bmatrix} 1 & -I(n \times n) & 0 & 0 & 0 \\ 0 & I(n \times n) & -I(n \times n) & 0 & 0 \\ 0 & 0 & I(n \times n) & -I(n \times n) & 0 \\ I(n \times n) & 0 & 0 & -I(n \times n) & 0 \\ 0 & I(n \times n) & 0 & 0 & -I(n \times n) \\ 0 & 0 & I(n \times n) & 0 & -I(n \times n) \end{bmatrix},$$

where each 0 is a matrix of size $n \times n$ with all zero elements.

The reformulated problem can be written compactly as

$$\begin{aligned} \min_{x_i \in X} \quad & \sum_{i=1}^{N} f_i(x_i) \\ s.t. \quad & Ax = 0, \end{aligned} \tag{1.2}$$

where x is the vector $[x_1, x_2, \ldots, x_N]'$. We will refer to this formulation as the *edge-*

based reformulation of the distributed multi-agent optimization problem. We also denote the *global objective function* given by the sum of the local objective function:

$$F(x) = \sum_{i=1}^{N} f_i(x_i). \tag{1.3}$$

This form of objective function, i.e., a sum of local functions depending only on local variables, is called *separable*.

Due to the large scale nature and the lack of centralized processing units of these problems, it is imperative that the solution we develop involves decentralized computations, meaning that each node (processor) performs calculations independently on the basis of local information available to it and only communicates this information to its neighbors according to the underlying network structure. Hence, the goal in this part of the thesis is to develop distributed optimization algorithm for problem (1.1) with provable convergence and rate of convergence guarantees for large-scale systems.

The distributed algorithms we develop will utilize the parallel processing power inherent in the system and can be used in more general parallel computation settings where the configuration of the underlying communication graph and distribution of local cost functions can be changed by the designer.[2] We will also analyze the dependence of algorithm performance on network topology. Insights from this analysis can be used to facilitate the design of a better network in the parallel computation environment.

1.2.2 Literature Review

There have been many important advances in the design of decentralized optimization algorithms in the area of multi-agent optimization, control, and learning in large scale

[2] We view parallel processing as a more general framework where the configuration of the underlying communication graph and distribution of local cost functions are choices made by the designer. Distributed algorithms, on the other hand, take the communication graph and local cost functions as given. Static sensor network with distributed data gathering, for example, falls under the category of distributed optimization.

networked systems. Most of the development builds on the seminal works [2] and [3], where distributed and parallel computations were first discussed. The standard approach in the literature involves the *consensus-based* procedures, in which the agents exchange their local estimates with neighbors with the goal of aggregating information over the entire network and reaching agreement in the limit. It has been shown that under some mild assumption on the connectivity of the graph and updating rules, both deterministic and random update rules can be used to achieve consensus (for deterministic update rules, see [4], [5], [6], [7]; for random update rules, see [8], [9], [10]). By combining the consensus procedure and parallelized first-order (sub)gradient computation, the existing literature has presented some distributed optimization methods for solving problem (1.2). The work [11] introduced a first-order primal subgradient method for solving problem (1.2) over deterministically varying networks. This method involves each agent maintaining and updating an estimate of the optimal solution by linearly combining a subgradient descent step along its local cost function with averaging of estimates obtained from his neighbors (also known as a single consensus step). Several follow-up papers considered variants of this method for problems with local and global constraints [12], [13] randomly varying networks [14], [15], [16] and random gradient errors [17], [18]. A different distributed algorithm that relies on Nesterov's dual averaging algorithm [19] for static networks has been proposed and analyzed in [20]. Such gradient methods typically have a convergence rate of $O(1/\sqrt{k})$ for general (possibly non-smooth) convex objective functions, where k is the number of iterations. [3] The more recent contribution [21] focuses on a special case of (1.2) under smoothness assumptions on the cost functions and availability of global information about some problem parameters, and provided gradient algorithms (with multiple consensus steps) which converge at the faster rate of $O(1/k^2)$. The smoothness assumption, however, is not satisfied by many important machine learning problems, the L_1 regularized Lasso for instance.

The main drawback of these (sub)gradient based existing method is the slow

[3] More precisely, for a predetermined number of steps k, the distributed gradient algorithm with constant stepsize can converge to $O(1/\sqrt{k})$ neighborhood of the optimal function value.

convergence rates (given by $O(1/\sqrt{k})$) for general convex problems, making them impractical in many large scale applications. Our goal is to provide a faster method, which preserves the distributed nature. One method that is known to perform well in a centralized setting is the *Alternating Direction Method of Multipliers (ADMM).* Alternating Direction Method of Multipliers (ADMM) was first proposed in 1970s by [22] and [23] in the context of finding zeros of the sum of two maximal monotone operators (more specifically, the Douglas-Rachford operator) and studied in the next decade by [24], [25], [26], [27]. Earlier work in this area focuses on the case $C = 2$, where C refers to the number of sequential primal updates at each iteration. Lately this method has received much attention in solving problem (1.2) (or specialized versions of it) (see [28], [29], [30], [21], [31], [32]), due to the excellent numerical performance and parallel nature. The convergence of ADMM can be established [28], [33], however the rate of convergence guarantee remained open until an important recent contribution [34]. In this work, He and Yuan (with $C = 2$) showed that the centralized ADMM implementation achieves convergence rate $O(1/k)$ in terms of objective function value. Other recent works [35], [36], [37], [38] analyzed the rate of convergence of ADMM and other related algorithms under various smoothness conditions (strongly convex objective function and/or Lipschitz gradient on the objective function), the algorithm can converge with rate $O(1/k^2)$ or even linear rate. In particular, [35], [36] studied ADMM with $C = 2$ and [37] considered a modified objective function at each primal update with $C \geq 2$. The work [38] considered the case $C > 2$ and showed that the resulting ADMM algorithm, converges under the more restrictive assumption that each f_i is strongly convex. In the recent work [39], the authors focused on the general case $C > 2$ a variant of ADMM under special assumptions on the problem structure. Using an error bound condition that estimates the distance from the dual optimal solution set, the authors established linear rate of convergence for their algorithm.

With the exception of [77] and [78], most of the algorithms provided in the literature are synchronous and assume that computations at all nodes are performed simultaneously according to a global clock. [77] provides an asynchronous subgradient method that uses gossip-type activation and communication between pairs of nodes and shows (under a compactness assumption on the iterates) that the iterates generated by this method converge almost surely to an optimal solution. The recent independent paper [78] provides an asynchronous randomized ADMM algorithm for solving problem (1.1) and establishes convergence of the iterates to an optimal solution by studying the convergence of randomized Gauss-Seidel iterations on non-expansive operators. This thesis proposes for the first time a distributed asynchronous ADMM based method with rates guarantees.

1.2.3 Contributions

Inspired by the convergence properties of ADMM, we developed distributed versions of ADMM for problem (1.2). Our contribution can be summarized as follows:

- We develop both synchronous and asynchronous distributed fast ADMM based method for problem (1.2).

- For synchronous algorithm, we show the distributed implementation of ADMM achieves $O(1/k)$ rate of convergence, where k is the number of iteration. This means that the error to the optimal value decreases (at least) with rate $1/k$. We also analyze the dependence of the algorithm performance on the underlying network topology.

- For asynchronous algorithm, we establish $O(1/k)$ rate of convergence for both algorithms, where k is the number of iterations. This is the best known rate of convergence guarantee for this class of optimization problems and first rate of convergence analysis for asynchronous distributed methods.

- We show that the rate of convergence for the synchronous method is related to underlying network topology, the total number of edges, diameter of graph and second smallest eigenvalue of the graph Laplacian, in particular

1.3 Electricity Market Analysis

1.3.1 Motivation

The second part of the thesis focuses on the electricity market. The market structure for the electricity market is similar to any other market, where the participates can be classified as either supply side or demand side, except there is an explicit market clearing house, the *Independent System Operator (ISO)*. The ISO is a heavily regulated government created entity, which monitors, coordinates and controls the operation of the grid. Most electricity markets have two stages: day-ahead, where most of planning is done and real time (5 to 10 minutes intervals), where adjustments are made to react to the real time information. We will focus on the day-ahead market for this thesis. In the day-ahead market in a traditional grid, the ISO forecasts the electricity demand for the next day, solicits the bids from generators and solves an optimization problem, i.e. *Economic Dispatch Problem*, to minimize the cost of production while meeting the demand. The price for the electricity comes from the dual multipliers of this optimization problem. The constraint of supply equal to demand (i.e. market clearing) is a physical constraint imposed on the grid. Any violation could compromise the reliability and stability of the entire system. The demand is assumed to be passive and inelastic. Therefore the generators have to guarantee market clearing, while respecting their physical limitation in how fast they can ramp up and down. This market clearing constraint becomes harder to implement when there is high volatility in the demand, which could be a result of high intermittent renewable penetration. For instance, in a region with high wind energy integration, the traditional generator (often referred to as "residual demand" in the literature) output needs to change significantly between a windy time and another (possibly neighboring) time with no/small wind. To ensure reliability of the system, the current solution to counter-act the uncertainties imposed by the renewable energies is to increase reserve, which is generation capacity set aside and can be brought online immediately or within short time window when needed. This reserve generation capacity decreases system efficiency level and diminishes the net environmental benefit from renewables [40], [41], [42], [43]. One way to absorb variabilities in the demand, reduce reserve requirement and improve system efficiency, is by introducing *demand response* in the smart grid. The demand (consumers and/or distributors) will see real time pricing signal (either directly or indirectly through utility companies) and respond accordingly. For instance, laundry machines, dish washer, electrical car charging and other flexible loads may choose to run at night or a windy time interval when the electricity price is cheaper. The price in the smart grid will come from the market interaction between the demand and the generation.

Our work aims at providing a systematic framework to study demand response integration. To facilitate understanding of the system, we will focus on an idealistic model with no uncertainty. We develop a flexible model of market equilibrium (competitive equilibrium) in electricity markets with heterogeneous users, where both supply and demand are active market participants. Over a time horizon (24-hour in a day-ahead market example), the users respond to price and maximize their utility, which reflects their preferences over the time slots, subject to a maximal consumption level constraint. The supplier, whose cost structure characterizes the physical limitation that the generation cannot change rapidly, also responds to price and maximizes its utility (profit).[4] We consider a market where both the consumer and the supplier cannot influence price strategically, i.e., perfectly competitive market. Under some standard conditions, we can show that a market clearing price and thus a competitive equilibrium exists. We provide characterizations of equilibrium prices and quantities, which can be shown to be efficient.

The equilibrium may feature high price and load volatility, due to user preferences over time. Both of these fluctuations impose uncertainty to the market participants and therefore are undesirable. In particular, having small fluctuation in the generation levels across time is also an important feature at the market operating point. Each generator is capable of adjusting its output level within a predetermined interval, beyond which additional units need to be turned on/off and incurring a significant

[4]The inability to adjust production level, a.k.a., ramp-up and ramp-down, significantly spontaneously, can be modeled two ways, either as a hard constraint or as a high cost on the generators. The two ways are equivalent if the cost of ramp-up or ramp-down is high enough.

start-up/shut-down cost. Hence relatively steady generation output is a requirement. In addition to the short-term physical limitations, steady output level is also preferred in the long term because the generators can focus on improving efficiency and reducing environment footprint in a small output interval, which is much more feasible than improving efficiency for the entire production range. Lastly, compared with highly fluctuating generation, where generation output levels involve the entire output range, low fluctuating generation also means that the generators have more capacity to respond to sudden changes in production. Thus having a steady generation output level also increases flexibility and reliability of the system and reduces the need of reserves, which further improves the system efficiency. Price fluctuation is undesirable due to the difficulties it imposes on planning for both consumer and producer.

We present a way to reduce both price and load fluctuations by using a pricing rule different from equilibrium outcome. In particular, we systematically introduce an explicit penalty on price fluctuation and show that it can be used to reduce generation fluctuation and can be implemented via storage, whose size is directly related to the price fluctuation. We study the system performance with storage introduced and characterize the connection between storage size, social welfare, and price, load fluctuation. We analyze the relationship between price (and load fluctuation) and consumer types (with different utility preferences).

1.3.2 Literature Review

Our work is related to the rapidly growing literature in the general area of smart grid systems, especially those concerning the electricity market. The existing works in this area fall into one of the following three categories: demand management, supply management and market interaction with both supply and demand actively participating. With exogenous supply, demand side management aims at minimizing total operation cost, while meeting demand service requirement, such as total time served or total energy consumed over certain time. Works on demand management propose methods to achieve certain system goal (usually flattening the demand) by coordinating the consumers. The approaches used for demand management including direct control by the utility or aggregator, where the households give up control of their flexible loads in exchange for some incentive of participation [44], [45], [46]; or indirectly through price signals (often in the Plug-in Electric Vehicles, a.k.a. PEV, charging context). The users receiving the price signals are optimizing individual objective functions dynamically based on updated forecast of future electricity prices [47] [48], [49], [50], [51], [52]. Another approach proposed for reducing demand volatility by demand shifting is via raffle-based incentive schemes [53],[54]. The idea of raffled based incentive mechanism has been used also in other demand management applications such as decongesting transportation systems [55]. The supply management literature studies the best way to plan production and reserve under uncertainty with exogenous demand. Various approaches such as stochastic programming and robust optimization have been considered [56], [57], [58], [59].

The third category, where the supply interacts with demand and social welfare is analyzed, is most relevant for this thesis. In [60], Mansur showed that with strategic generators and numerical econometrical studies, under ramping constraints, the prices faced by consumers can be different from the true supplier marginal congestion cost. In [61], Cho and Meyn studied the competitive equilibrium in the presence of exogenous stochastic demand by modeling ramp-up/ramp-down as constraints and showed that the limited capability of generating units to meet real-time demand, due to relatively low ramping rates, does not harm social welfare, but can cause extreme price fluctuations. In another recent paper in this stream, Kizilkale and Mannor, [62] construct a dynamic game-theoretic model using specific continuous time control to study the trade-off between economic efficiency and price volatility. The supplier cost in their model depends on the overall demand, as well as the generation resources required to meet the demand. In [63], Roozbehani et al. show that under market clearing conditions, very high price volatility may result from real time pricing. In [64], Huang et al. considered the trade-off between efficiency and risk in a different dynamic oligopoly market architecture. In [65], the authors proposed to provide differentiated services to end users according to deadline to increase system efficiency.

Our work is most closely related to Xu and Tsitsiklis [66] and Nayyar et al. [67].

In [66], the authors considered a stylized multi-period game theoretic model of an electricity market with incorporation of ramping-up and down costs. Decisions are made at each period and will affect the following time periods. The cost of production depends on an exogenous state, which evolves according to a Markov chain, capturing uncertainties in the renewable generation. The consumers have different types drawn independently from a known distribution. The authors proposed a pricing mechanism, where the load needs to pay for the externality on the ramping service. They compared their pricing mechanism with marginal cost pricing (without the ramping costs), and showed that the proposed mechanism reduces the peak load and achieves asymptotic social optimal as number of loads goes to infinity. Our model considers planning for the entire time horizon as a whole and will analyze the trade-offs between welfare, price and load fluctuations. In [67], a version of demand response is considered. Each load demands a specified integer number of time slots to be served (unit amount each time) and the supply can purchase either at the day ahead market with a fixed price for all time periods or at the real time market, again with a fixed price for all time periods. The authors derive an optimal strategy for the supply to satisfy the demand while minimizing the supply side cost. The paper also includes competitive equilibrium analysis where the consumers can choose between different duration differentiated loads and optimize its utility. The major differences between [67] and our framework are that our user have a preference over time slot, which is a property fundamentally intrinsic to many loads, our users may choose any real number consumption level as opposed to discrete values in [67] and most importantly, we consider a pricing structure which is different based on the time of the day. The novelty in our work lies in proposing alternative pricing rules through introduction of explicit penalty terms on price fluctuation. In addition, we consider the change and trade-offs in social welfare, market interaction and load fluctuation.

1.3.3 Contributions

Our main findings for the electricity market analysis are summarized as follows:

- We propose a flexible multi-period market model, capturing supply and demand interaction in the electricity market.
- We show that the new model is efficient at competitive equilibrium.
- We establish that incorporating demand response can improve social welfare significantly over the traditional supply-follow-demand market.
- We introduce an explicit penalty term on the price fluctuation in the competitive equilibrium framework, which can reduce both generation and load fluctuation. We show that this penalty term is equivalent to introducing storage into the system.
- The price fluctuation penalized problem can be implemented with storage in a distributed way.
- The relationship between storage and generation fluctuation is analyzed in two cases: in the inelastic demand case, an explicit characterization of optimal storage access and generation fluctuation is given; in the elastic demand case, connection between concavity of demand utility function and generation is studied and we show some monotonicity results between concavity of consumer demand and generation fluctuation.

The rest of the thesis is organized as follows: Chapter 2 contains our development on synchronous distributed ADMM method and Chapter 3 is on asynchronous distributed ADMM method. Chapters 4 and 5 study market dynamic of electricity market. Chapter 4 proposes the competitive equilibrium framework of the electricity market and analyzes its properties. Chapter 5 addresses the issue of reducing generation fluctuation. In Chapter 6, we include summarize our findings and propose some interesting future directions.

Basic Notation and Notions:

A vector is viewed as a column vector. For a matrix A, we write $[A]_i$ to denote the i^{th} column of matrix A, and $[A]^j$ to denote the j^{th} row of matrix A. For a vector x,

x_i denotes the i^{th} component of the vector. For a vector x in $\mathbb{R}^n$, and set S a subset of $\{1, \ldots, n\}$, we denote by $[x]_S$ a vector $\mathbb{R}^n$, which places zeros for all components of x outside set S, i.e.,

$$[[x]_S]_i = \begin{cases} x_i & \text{if} \quad i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

We use x' and A' to denote the transpose of a vector x and a matrix A respectively. We use standard Euclidean norm (i.e., 2-norm) unless otherwise noted, i.e., for a vector x in $\mathbb{R}^n$, $||x|| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$. The notation $I(n)$ denotes the identity matrix of dimension n. The gradient and Hessian of function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by $\nabla f(x)$ and $\nabla^2 f(x)$ respectively (if $n = 1$ this is the same as first and second derivatives).

Chapter 2

Synchronous Distributed ADMM: Performance and Network Effects

In this chapter, we present our development on synchronous distributed ADMM method and analyze its performance and dependence on the underlying network topology. We focus on the unconstrained version of problem (1.1). This chapter is organized as follows. Section 2.1 reviews the standard ADMM algorithm. In Section 2.2, we present the problem formulation and equivalent reformulation, which enables us to develop distributed implementation in Section 2.3. Section 2.4 analyzes the convergence properties of the proposed algorithm. Section 2.5 contains our concluding remarks.

2.1 Preliminaries: Standard ADMM Algorithm

The standard ADMM algorithm solves a convex optimization problem with two primal variables. The objective function is separable and the coupling constraint is linear:[1]

$$\begin{aligned} \min_{x \in X, z \in Z} \quad & F_s(x) + G_s(z) \\ s.t. \quad & D_s x + H_s z = c, \end{aligned} \tag{2.1}$$

where $F_s : \mathbb{R}^n \to \mathbb{R}$ and $G_s : \mathbb{R}^m \to \mathbb{R}$ are convex functions, X and Z are nonempty closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, and D_s and H_s are matrices of size $w \times n$ and $w \times m$.

The augmented Lagrangian function of previous problem is given by

$$\begin{aligned} L_\beta(x, z, p) = & F_s(x) + G_s(z) - p'(D_s x + H_s z - c) \\ & + \frac{\beta}{2} \left|\left|D_s x + H_s z - c\right|\right|^2, \end{aligned} \tag{2.2}$$

where p in $\mathbb{R}^w$ is the dual variable corresponding to the constraint $D_s x + H_s z = c$ and β is a positive penalty parameter for the quadratic penalty of feasibility violation.

Starting from some initial vector (x^0, z^0, p^0), the standard ADMM method proceeds by[2]

$$x^{k+1} \in \operatorname*{argmin}_{x \in X} L_\beta(x, z^k, p^k), \tag{2.3}$$

$$z^{k+1} \in \operatorname*{argmin}_{z \in Z} L_\beta(x^{k+1}, z, p^k), \tag{2.4}$$

$$p^{k+1} = p^k - \beta(D_s x^{k+1} + H_s z^{k+1} - c). \tag{2.5}$$

The ADMM iteration approximately minimizing the augmented Lagrangian function through sequential updates of the primal variables x and z and then a gradient ascent step in the dual, using the stepsize same as the penalty parameter β (see [39] and [33]). This algorithm is particularly useful in applications where the minimization over these component functions admits simple solutions and can be implemented in a parallel or decentralized manner.

[1]Interested readers can find more details in [28] and [33].

[2]We use superscripts to denote the iteration number.

The analysis of the ADMM algorithm adopts the following standard assumption on problem (2.1).

Assumption 1. *The optimal solution set of problem (2.1) is nonempty.*

Under this assumption, the following convergence property for ADMM is known (see Section 3.2 of [28]).

Theorem 2.1.1. *Let $\{x^k, z^k, p^k\}$ be the iterates generated by the standard ADMM, then the objective function value of $F_s(x^k) + G_s(z^k)$ converges to the optimal value of problem (2.1) and the dual sequence $\{p^k\}$ converges to a dual optimal solution.*

2.2 Problem Formulation

Consider the system of N networked agents introduced in Section 1.2.1, where the underlying communication topology is represented by an undirected *connected* graph $G = (V, E)$ where V is the set of agents with $|V| = N$, and E is the set of edges with $|E| = M$. We use notation $B(i)$ to denote the set of neighbors of agent i. Each agent is endowed with a convex local objective function $f_i : \mathbb{R}^n \to \mathbb{R}$. Recall that the goal of the agents is to collectively solve the following problem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x). \tag{2.6}$$

This problem formulation arises in large-scale multi-agent (or processor) environments where problem data is distributed across N agents, i.e., each agent has access only to the component function f_i.

In this chapter, we develop an ADMM algorithm for solving problem (2.6) under the assumption that each agent perform local computations based on its own local objective function f_i and information received from its neighbors. To apply ADMM to problem (2.6), we use the same transformation as in Section 1.2.1, i.e., introducing a local copy of the global variable x for each of the agents, denoted by x_i and constraint $x_i = x_j$ for any i, j connected through an edge. Therefore problem (2.6) can be equivalently stated as

$$\min_{x_i \in \mathbb{R}^n,\ i=1,\ldots,N} \sum_{i=1}^{N} f_i(x_i) \tag{2.7}$$
$$s.t. \quad x_i = x_j \quad \text{for all } (i,j) \in E.$$

The next example illustrates using ADMM to solve problem (2.7) for a network of two agents connected through a single edge.

Example 2.2.1. *We use ADMM to solve the two agent version of problem (2.7):*

$$\min_{x_1, x_2 \in \mathbb{R}^n} f_1(x_1) + f_2(x_2)$$
$$s.t. \quad x_1 = x_2.$$

Using Eqs. (2.3)-(2.5), it can be seen that ADMM generates a primal-dual sequence $\{x_1^k, x_2^k, p^k\}$*, which at iteration* k *is updated as follows:*

$$\begin{aligned}
x_1^{k+1} &= \operatorname*{argmin}_{x_1} f_1(x_1) - (p_{12}^k)' x_1 + \frac{\beta}{2} \left|\left| x_1 - x_2^k \right|\right|_2^2 \\
x_2^{k+1} &= \operatorname*{argmin}_{x_2} f_2(x_2) + (p_{12}^k)' x_2 + \frac{\beta}{2} \left|\left| x_1^{k+1} - x_2 \right|\right|_2^2 \\
p_{12}^{k+1} &= p_{12}^k - \beta(x_1^{k+1} - x_2^{k+1}).
\end{aligned}$$

This shows that at each k*, we first update* x_1^k *and using the updated value* x_1^{k+1}*, we then update* x_2^k*.*

As the previous example highlights, direct implementation of ADMM on problem (2.7) requires an order with which the primal variables are updated. This observation was used in our recent paper [72] to develop and study an ADMM algorithm for solving problem (2.7) under the assumption that there is a globally known order on the agents (see also [30]). This algorithm cycles through the agents according to this order. In many applications however, neither the presence of such global information nor an algorithm whose runtime scales linearly with the number of agents is feasible.

To remove the ordering, we use a reformulation technique, which was used in

[3], to generate a problem,F which involves constraints separable over the primal x_i variables (in addition to a separable objective function). For each constraint $x_i = x_j$, we introduce two additional auxiliary variables z_{ij} and z_{ji} both in $\mathbb{R}^n$ and rewrite the constraint as

$$x_i = z_{ij}, \quad x_j = z_{ji}, \quad z_{ij} = z_{ji}.$$

The variables z_{ij} can be viewed as an estimate of the component x_j which is maintained and updated by agent i. To write the transformed problem compactly, we stack the vectors z_{ij} into a long vector $z = [z_{ij}]_{(i,j)\in E}$ in $\mathbb{R}^{2Mn}$. We refer to the component in vector z as either z_{ij} with two sub-indices for the component in $\mathbb{R}^n$ associated with agent i over edge (i, j) or z_p with one sub-index for any such component. Similarly, we stack x_i into a long vector x in $\mathbb{R}^{Nn}$ and we use x_i to refer to the component in $\mathbb{R}^n$ associated with agent i. We also introduce matrix D in $\mathbb{R}^{2Mn\times Nn}$ which consists of $2M$ by N $n \times n$-blocks, where each block is either all zero or $I(n)$. The (a, b) block of matrix D takes value $I(n)$ if the a^{th} component of vector z corresponds to an edge involving x_b. Hence the transformed problem (2.7) can be written compactly as

$$\begin{aligned} \min_{x\in\mathbb{R}^{Nn}, z\in Z} \quad & F(x) \\ s.t. \quad & Dx - z = 0. \end{aligned} \tag{2.8}$$

where Z is the set $\{z \in \mathbb{R}^{2Mn} \mid z_{ij} = z_{ji}, \text{ for } (i, j) \text{ in } E\}$ and

$$F(x) = \sum_{i=1}^{N} f_i(x_i). \tag{2.9}$$

We assign dual variable p in $\mathbb{R}^{2Mn}$ for the linear constraint and refer to the n-component associated with the constraint $x_i - z_{ij}$ by p_{ij}. For the rest of the chapter, we adopt the following standard assumption.

Assumption 2. *The optimal solution set of problem (2.8) is nonempty.*

In view of convexity of function F, linearity of constraints and polyhedrality of Z ensures that the dual problem of (2.8) has an optimal solution and that there is no duality gap (see [75]). A primal-dual optimal solution (x^*, z^*, p^*) is also a saddle point of the Lagrangian function,

$$L(x, z, p) = F(x) - p'(Dx - z), \tag{2.10}$$

i.e.,

$$L(x^*, z^*, p) \leq L(x^*, z^*, p^*) \leq L(x, z, p^*),$$

for all x in $\mathbb{R}^{nN}$, z in Z and p in $\mathbb{R}^{2Mn}$.

2.3 Distributed ADMM Algorithm

In this section, we apply the standard ADMM algorithm as described in Section 2.1 to problem (2.8). We then show that this algorithm admits a distributed implementation over the network of agents.

Using Eqs. (2.3)-(2.5), we see that each iteration of the ADMM algorithm for problem (2.8) involves the following three steps:

a The primal variable x update

$$x^{k+1} \in \operatorname*{argmin}_x F(x) - (p^k)'Dx + \frac{\beta}{2} \left|\left|Dx - z^k\right|\right|^2. \tag{2.11}$$

b The primal variable z update

$$z^{k+1} \in \operatorname*{argmin}_{z \in Z} (p^k)'z + \frac{\beta}{2} \left|\left|Dx^{k+1} - z\right|\right|^2. \tag{2.12}$$

c The dual variable p is updated as

$$p^{k+1} = p^k - \beta(Dx^{k+1} - z^{k+1}). \tag{2.13}$$

By Assumption 2, the level sets $\{x | F(x) - (p^k)'Dx + \frac{\beta}{2} \left|\left|Dx - z^k\right|\right|^2 \leq \alpha\}$ and $\{(p^k)'z + \frac{\beta}{2} \left|\left|Dx^{k+1} - z\right|\right|^2 \geq \alpha\}$ for α in $\mathbb{R}$ are bounded and hence the search for minimum can be equivalently done over a compact level sets. Therefore, by Weierstrass theorem, the minima are obtained. Due to convexity of function F and the fact that matrix $D'D$ has full rank, the objective function in updates (2.11) and (2.12) are strictly convex. Hence, these minima are also unique.

We next show that these updates can be implementation by each agent i using its local information and estimates x_j^k communicated from its neighbors, $j \in B(i)$. Assume that each agent i maintains x_i^k and p_{ij}^k, z_{ij}^k for $j \in B(i)$ at iteration k. Using the separable nature of function F [cf. Eq. (2.9)] and structure of matrix D, the update (2.11) can be written as $x^{k+1} = \text{argmin}_x \sum_i f_i(x_i) - \sum_i \sum_{j \in B(i)} \left[(p_{ij}^k)' x_i + \frac{\beta}{2} \left|\left| x_i - z_{ij}^k \right|\right|^2 \right]$. This minimization problem can be solved by minimizing act component of the sum over x_i. Since each agent i knows $f_i(x)$, p_{ij}^k, z_{ij}^k for $j \in B(i)$, this minimization problem can be solved by agent i using local information. Each agent i then communicates their estimates x_i^{k+1} to all their neighbors $j \in N(i)$.

By a similar argument, the primal variable z and the dual variable p updates can be written as $z_{ij}^{k+1}, z_{ji}^{k+1} = \text{argmin}_{z_{ij}, z_{ji}, z_{ij} = z_{ji}} -(p_{ij}^k)'(x_i^{k+1} - z_{ij}) - (p_{ji}^k)'(x_j^{k+1} - z_{ji}) + \frac{\beta}{2} \left(\left|\left| x_i^{k+1} - z_{ij} \right|\right|^2 + \left|\left| x_j^{k+1} - z_{ji} \right|\right|^2 \right)$, $p_{ij}^{k+1} = p_{ij}^k - \beta(x_i^{k+1} - z_{ij}^{k+1}), p_{ji}^{k+1} = p_{ji}^k - \beta(x_j^{k+1} - z_{ji}^{k+1})$. The primal variable z update involves a quadratic optimization problem with linear constraints which can be solved in closed form. In particular, using first order optimality conditions, we conclude

$$z_{ij}^{k+1} = \frac{1}{\beta}(p_{ij}^k - v^{k+1}) + x_i^{k+1}, z_{ji}^{k+1} = \frac{1}{\beta}(p_{ji}^k + v^{k+1}) + x_j^{k+1},$$

where v^{k+1} is the Lagrange multiplier associated with the constraint $z_{ij} - z_{ji} = 0$ and is given by

$$v^{k+1} = \frac{1}{2}(p_{ij}^k - p_{ji}^k) + \frac{\beta}{2}(x_i^{k+1} - x_j^{k+1}).$$

The dual variable update also simplifies to

$$p_{ij}^{k+1} = v^{k+1}, \qquad p_{ji}^{k+1} = -v^{k+1}.$$

With the initialization of $p_{ij}^0 = -p_{ji}^0$, the following simplified iteration generates

$p_{ij}^{k+1} = -p_{ji}^{k+1}$ and gives the identical update sequence as above

$$v^{k+1} = p_{ij}^k + \frac{\beta}{2}(x_i^{k+1} - x_j^{k+1}),$$

$$z_{ij}^{k+1} = \frac{1}{2}(x_i^{k+1} + x_j^{k+1}), z_{ji}^{k+1} = \frac{1}{2}(x_i^{k+1} + x_j^{k+1}),$$

$$p_{ij}^{k+1} = p_{ij}^k + \frac{\beta}{2}(x_i^{k+1} - x_j^{k+1}), p_{ji}^{k+1} = p_{ji}^k + \frac{\beta}{2}(x_j^{k+1} - x_i^{k+1}).$$

Since each agent i has access to x_i^{k+1} and x_j^{k+1} for all $j \in B(i)$ (which was communicated over link (i, j)), he can perform the preceding updates using local information.

Combining the above steps leads to the following distributed ADMM algorithm.

Distributed ADMM algorithm:

A Initialization: choose some arbitrary x^0 in $\mathbb{R}^{Nn}$, z^0 in Z and p^0 in $\mathbb{R}^{2Mn}$ with $p_{ij}^0 = -p_{ji}^0$.

B At iteration k,

a Each agent i, the primal variable x_i^k is updated as $x_i^{k+1} = \operatorname{argmin}_{x_i} f_i(x_i) - \sum_{j \in B(i)} (p_{ij}^k)' x_i + \frac{\beta}{2} \sum_{j \in B(i)} \left|\left|x_i - z_{ij}^k\right|\right|^2$.

b For each pair of neighbors (i, j), the primal variables z_{ij} and z_{ji} are updated as

$$z_{ij}^{k+1} = \frac{1}{2}(x_i^{k+1} + x_j^{k+1}), z_{ji}^{k+1} = \frac{1}{2}(x_i^{k+1} + x_j^{k+1}),$$

d For each pair of neighbors (i, j), the dual variables p_{ij} and p_{ji} are updated as

$$p_{ij}^{k+1} = p_{ij}^k + \frac{\beta}{2}(x_i^{k+1} - x_j^{k+1}),$$

$$p_{ji}^{k+1} = p_{ji}^k + \frac{\beta}{2}(x_j^{k+1} - x_i^{k+1}).$$

2.4 Convergence Analysis for Distributed ADMM Algorithm

In this section, we study the convergence behavior of the distributed ADMM algorithm. We show that the primal iterates $\{x^k, z^k\}$ generated by (2.11) and (2.12) converge to an optimal solution of problem (2.8) and both the difference of the objective function values from the optimal value and the feasibility violations converge to 0 at rate $O(1/T)$. We first present some preliminary results in Section 2.4.1, which are used to establish general convergence and rate of convergence properties in Section 2.4.2. In Section 2.4.3, we analyze the algorithm performance in more depth and investigate the impact of network structure.

For notational convenience, we denote by r^{k+1} *the residual* of the form

$$r^{k+1} = Dx^{k+1} - z^{k+1}. \tag{2.14}$$

By combining the notation introduced above and update (2.13), we have

$$p^{k+1} = p^k - \beta r^{k+1}, \tag{2.15}$$

which is one of the key relations in the following analysis.

2.4.1 Preliminaries

In this section, we first provide some preliminary general results on optimality conditions (which enable us to linearize the quadratic term in the primal updates of the ADMM algorithm) and feasibility of the saddle points of the Lagrangian function of problem (2.8) (see Lemmas 2.4.1 and 2.4.2). We then use these results to rewrite the optimality conditions of (x^k, z^k) and obtain Lemma 2.4.3, based on which, we can derive bounds on two key quantities $F(x^{k+1})$ and $F(x^{k+1}) - p'r^{k+1}$ in Theorems 2.4.1 and Theorem 2.4.2 respectively. The bound on $F(x^{k+1})$ will be used to establish convergence properties of the objective function value, whereas the bound on

$F(x^{k+1}) - p'r^{k+1}$ will be used to show that feasibility violation diminishes to 0 at rate $O(1/T)$ in the next section.

The following lemma is similar to Lemma 4.1 from [3], we include here for completeness and present a different proof.

Lemma 2.4.1. *Let functions* $J_1 : \mathbb{R}^m \to \mathbb{R}$ *and* $J_2 : \mathbb{R}^m \to \mathbb{R}$ *be convex, and function* J_2 *be continuously differentiable. Let* Y *be a closed convex subset of* $\mathbb{R}^n$ *and*

$$y^* = \arg\min_{y \in Y} J_1(y) + J_2(y),$$

then

$$y^* = \arg\min_{y \in Y} J_1(y) + \nabla J_2(y^*)'y.$$

Proof. The optimality of y^* implies that there exists some subgradient $h(y^*)$ of function J_1, i.e., $h(y^*)$ in $\partial J_1(y^*)$, such that

$$(h(y^*) + \nabla J_2(y^*))' (y - y^*) \geq 0,$$

for all y in Y. Since $h(y^*)$ is the subgradient of function J_1, by definition of subgradient, we have

$$J_1(y) \geq J_1(y^*) + (y - y^*)'h(y^*).$$

By summing the above two relations, we obtain

$$J_1(y) + \nabla J_2(y^*)'y \geq J_1(y^*) + \nabla J_2(y^*)'y^*,$$

for any y in Y, and thus we establish the desired claim. □

The next lemma establishes primal feasibility (or zero residual property) of a saddle point of the Lagrangian function of problem (2.8).

Lemma 2.4.2. *Let* (x^*, z^*, p^*) *be a saddle point of the Lagrangian function defined as in Eq. (2.10) of problem (2.8). Then*

$$Dx^* - z^* = 0, . \tag{2.16}$$

Proof. We prove by contradiction. From the definition of a saddle point, we have for any multiplier p in $\mathbb{R}^{2MN}$, the following relation holds

$$F(x^*) - p'(Dx^* - z^*) \leq F(x^*) - (p^*)'(Dx^* - z^*),$$

i.e., $p'(Dx^* - z^*) \geq (p^*)'(Dx^* - z^*)$ for all p.

Assume for some i, we have $[Dx^* - z^*]_i \neq 0$, then by setting

$$\tilde{p}_j = \begin{cases} \frac{(p^*)'(Dx^* - z^*) - 1}{[Dx^* - z^*]_i} & \text{for} \quad j = i, \\ 0 & \text{for} \quad j \neq i, \end{cases}$$

we arrive at a contradiction that $\tilde{p}'(Dx^* - z^*) = (p^*)'(Dx^* - z^*) - 1 < (p^*)'(Dx^* - z^*)$. Hence we conclude that Eq. (2.16) holds. □

The next lemma uses Lemma 2.4.1 to rewrite the optimality conditions for the iterates (x^k, z^k), which is the key to later establish bounds on the two key quantities: $F(x^{k+1}) - p'r^{k+1}$ and $F(x^{k+1})$.

Lemma 2.4.3. *Let $\{x^k, z^k, p^k\}$ be the iterates generated by our distributed ADMM algorithm for problem (2.8), then the following holds for all k,*

$$F(x) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left[(r - r^{k+1}) + (z - z^{k+1})\right] - (p^{k+1})'(r - r^{k+1}) \geq 0, \tag{2.17}$$

and

$$-(p^k - p^{k+1})'(z^k - z^{k+1}) \geq 0, \tag{2.18}$$

for any x in $\mathbb{R}^{Nn}$, z in Z and residual $r = Dx - Hz$.

Proof. From update (2.11), we have x^{k+1} minimizes the function $F(x) - (p^k)'(Dx - z^k) + \frac{\beta}{2}\left|\left|Dx - z^k\right|\right|^2$ over $\mathbb{R}^{Nn}$ and thus by Lemma 2.4.1 , we have, x^{k+1} is the minimizer of the function $F(x) + [-p^k + \beta(Dx^{k+1} - z^k)]'Dx$, i.e.,

$$F(x) + [-p^k + \beta(Dx^{k+1} - z^k)]'Dx \geq F(x^{k+1}) + [-p^k + \beta(Dx^{k+1} - z^k)]'Dx^{k+1},$$

for any x in $\mathbb{R}^{Nn}$. By the definition of residual r^{k+1} [cf. Eq. (2.14)], we have that

$$\beta(Dx^{k+1} - z^k) = \beta r^{k+1} + \beta(z^{k+1} - z^k).$$

In view of Eq. (2.15), we have

$$p^k = p^{k+1} + \beta r^{k+1}.$$

By subtracting the preceding two relations, we have $-p^k + \beta(Dx^{k+1} - z^k) = -p^{k+1} + \beta(z^{k+1} - z^k)$. This implies that the inequality above can be written as

$$F(x) - F(x^{k+1}) + \left[-p^{k+1} + \beta(z^{k+1} - z^k)\right]' D(x - x^{k+1}) \geq 0.$$

Similarly, the vector z^{k+1} [c.f. Eq. (2.12)] is the minimizer of the function

$$\left[p^k - \beta(Dx^{k+1} - z^{k+1})\right]' z = \left[p^k - \beta r^{k+1}\right]' z = (p^{k+1})' z,$$

i.e.,

$$(p^{k+1})'(z - z^{k+1}) \geq 0, \tag{2.19}$$

for any z in Z. By summing the preceding two inequalities, we obtain that for any x in $\mathbb{R}^{Nn}$ and z in Z the following relation holds,

$$F(x) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' D(x - x^{k+1}) - (p^{k+1})'(r - r^{k+1}) \geq 0,$$

where we used the identities $r = Dx - z$ and Eq. (2.14).

By using the definition of residuals once again, we can also rewrite the term

$$D(x - x^{k+1}) = (r + z) - (r^{k+1} + z^{k+1}).$$

Hence, the above inequality is equivalent to

$$F(x) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left[(r - r^{k+1}) + (z - z^{k+1})\right] - (p^{k+1})'(r - r^{k+1}) \geq 0,$$

which establishes the first desired relation.

We now proceed to show Eq. (2.18).

We note that Eq. (2.19) holds for any z for each iteration k. We substitute $z = z^k$ and have

$$(p^{k+1})'(z^k - z^{k+1}) \geq 0.$$

Similarly, for iteration $k-1$ with $z = z^{k+1}$, we have

$$(p^k)'(z^{k+1} - z^k) \geq 0.$$

By summing these two inequalities, we obtain

$$-(p^k - p^{k+1})'(z^k - z^{k+1}) \geq 0,$$

hence we have shown relation (2.18) holds. □

The next theorem uses preceding lemma to establish a bound on the key quantity $F(x^{k+1})$.

Theorem 2.4.1. *Let $\{x^k, z^k, p^k\}$ be the iterates generated by our distributed ADMM algorithm and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (2.8), then the following holds for all k,*

$$F(x^*) - F(x^{k+1}) \geq \frac{\beta}{2}\left(\left|\left|(z^{k+1} - z^*)\right|\right|^2 - \left|\left|(z^k - z^*)\right|\right|^2\right) + \frac{1}{2\beta}\left(\left|\left|p^{k+1}\right|\right|^2 - \left|\left|p^k\right|\right|^2\right). \tag{2.20}$$

Proof. We derive the desired relation based on substituting $x = x^*$, $z = z^*$, $r = r^*$ into Eq. (2.17), i.e.,

$$F(x^*) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left[(r^* - r^{k+1}) + (z^* - z^{k+1})\right] - (p^{k+1})'(r^* - r^{k+1}) \geq 0. \tag{2.21}$$

By definition of residual r, we have $(r^* - r^{k+1}) + (z^* - z^{k+1}) = D(x^* - x^{k+1})$. We then rewrite the terms $\left[\beta(z^{k+1} - z^k)\right]' D(x^* - x^{k+1}) - (p^{k+1})'(r^* - r^{k+1})$. We can use Eq.

(2.14) to rewrite $r^{k+1} = Dx^{k+1} - z^{k+1}$ and based on Lemma 2.4.2, $r^* = Dx^* - z^* = 0$. Using these two observations, the term $-(p^{k+1})'(r^* - r^{k+1})$ can be written as

$$-(p^{k+1})'(r^* - r^{k+1}) = (p^{k+1})' r^{k+1},$$

and the term $D(x^* - x^{k+1})$ is the same as

$$D(x^* - x^{k+1}) = z^* - Dx^{k+1} = z^* - (r^{k+1} + z^{k+1}).$$

Eq. (2.15) suggests that $r^{k+1} = \frac{1}{\beta}(p^k - p^{k+1})$, which implies that

$$-(p^{k+1})'(r^* - r^{k+1}) = \frac{1}{\beta}(p^{k+1})'(p^k - p^{k+1}),$$

and

$$D(x^* - x^{k+1}) = -z^{k+1} + z^* - \frac{1}{\beta}(p^k - p^{k+1}).$$

By combining the preceding two relations with Eq. (2.21) yields

$$\begin{aligned} &F(x^*) - F(x^{k+1}) \\ &+ \left[\beta(z^{k+1} - z^k)\right]' \left(-z^{k+1} + z^* - \frac{1}{\beta}(p^k - p^{k+1})\right) + \frac{1}{\beta}(p^{k+1})'(p^k - p^{k+1}) \geq 0. \end{aligned}$$

Eq. (2.18) suggests that

$$\left[\beta(z^{k+1} - z^k)\right]' \left(\frac{1}{\beta}(p^k - p^{k+1})\right) \geq 0.$$

We can therefore add the preceding two relations and have

$$F(x^*) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left(-z^{k+1} + z^*\right) + \frac{1}{\beta}(p^{k+1})'(p^k - p^{k+1}) \geq 0. \quad (2.22)$$

We use the identity

$$||a + b||^2 = ||a||^2 + ||b||^2 + 2a'b,$$

for arbitrary vectors a and b to rewrite the inner product terms and have

$$\begin{aligned}&\left[\beta(z^{k+1}-z^k)\right]'\left(-z^{k+1}+z^*\right)=\\&\frac{\beta}{2}\left(\left|\left|(z^k-z^*)\right|\right|^2-\left|\left|(z^k-z^{k+1})\right|\right|^2-\left|\left|(z^{k+1}-z^*)\right|\right|^2\right),\end{aligned}$$

and

$$\frac{1}{\beta}(p^{k+1})'(p^k-p^{k+1})=\frac{1}{2\beta}\left(\left|\left|p^k\right|\right|^2-\left|\left|p^{k+1}\right|\right|^2-\left|\left|p^k-p^{k+1}\right|\right|^2\right).$$

These two equalities imply that Eq. (2.22) is equivalent to

$$\begin{aligned}F(x^*)-F(x^{k+1})\geq&\frac{\beta}{2}\left(\left|\left|(z^{k+1}-z^*)\right|\right|^2-\left|\left|(z^k-z^*)\right|\right|^2\right)+\frac{1}{2\beta}\left(\left|\left|p^{k+1}\right|\right|^2-\left|\left|p^k\right|\right|^2\right)\\&-\frac{\beta}{2}\left|\left|(z^k-z^{k+1})\right|\right|^2-\frac{1}{2\beta}\left|\left|p^k-p^{k+1}\right|\right|^2\end{aligned}$$

The last two terms are non-positive and can be dropped from right hand side and thus we establish the desired relation. □

The next lemma represents an equivalent form for some of the terms in the inequality from the preceding lemma. Theorem 2.4.2 then combines these two lemmas to establish the bound on the key quantity $F(x^{k+1})-p'r^{k+1}$.

The proof for the following lemma is based on the definition of residual r^{k+1} and algebraic manipulations, similar to those used in [3], [28] and [34].

Lemma 2.4.4. *Let $\{x^k, z^k, p^k\}$ be the iterates generated by distributed ADMM algorithm for problem (2.8). Let vectors x, z and p be arbitrary vectors in $\mathbb{R}^{Nn}$, Z and $\mathbb{R}^{2Mn}$ respectively. The following relation holds for all k,*

$$\begin{aligned}&-(p^{k+1}-p)'r^{k+1}+(r^{k+1})'\beta(z^{k+1}-z^k)-\beta(z^{k+1}-z^k)'(z-z^{k+1})\\&=\frac{1}{2\beta}\left(\left|\left|p^{k+1}-p\right|\right|^2-\left|\left|p^k-p\right|\right|^2\right)+\frac{\beta}{2}\left(\left|\left|z^{k+1}-z\right|\right|^2-\left|\left|z^k-z\right|\right|^2\right)\\&\quad+\frac{\beta}{2}\left|\left|r^{k+1}+(z^{k+1}-z^k)\right|\right|^2.\end{aligned}\tag{2.23}$$

Proof. It is more convenient to multiply both sides of Eq. (2.23) by 2 and prove

$$\begin{aligned}
&-2(p^{k+1}-p)'(r^{k+1})+2\beta(r^{k+1})'(z^{k+1}-z^k)-2\beta(z^{k+1}-z^k)'(z-z^{k+1}) \\
&=\frac{1}{\beta}\left(\left|\left|p^{k+1}-p\right|\right|^2-\left|\left|p^k-p\right|\right|^2\right)+\beta\left(\left|\left|(z^{k+1}-z)\right|\right|^2-\left|\left|(z^k-z)\right|\right|^2\right) \\
&\quad+\beta\left|\left|r^{k+1}+(z^{k+1}-z^k)\right|\right|^2.
\end{aligned} \tag{2.24}$$

Our proof will use the following two identities: Eq. (2.15), i.e.,

$$p^{k+1}=p^k-\beta r^{k+1},$$

and

$$||a+b||^2=||a||^2+||b||^2+2a'b, \tag{2.25}$$

for arbitrary vectors a and b.

We start with the first term $-2(p^{k+1}-p)'r^{k+1}$ on the left-hand side of Eq. (2.24). By adding and subtracting the term $2(p^k)'r^{k+1}$, we obtain

$$-2(p^{k+1}-p)'r^{k+1}=-2(p^{k+1}-p^k+p^k-p)'r^{k+1}=2\beta\left|\left|r^{k+1}\right|\right|^2-2(p^k-p)'r^{k+1},$$

where we used Eq. (2.15) to write $(p^{k+1}-p^k)r^{k+1}=-\beta\left|\left|r^{k+1}\right|\right|^2$. Using Eq. (2.15) once more, we can write the term $-2(p^k-p)'r^{k+1}$ as

$$\begin{aligned}
-2(p^k-p)'r^{k+1}&=\frac{2}{\beta}(p-p^k)'(p^k-p^{k+1}) \\
&=\frac{1}{\beta}\left(\left|\left|p-p^{k+1}\right|\right|^2-\left|\left|p-p^k\right|\right|^2-\left|\left|p^k-p^{k+1}\right|\right|^2\right),
\end{aligned}$$

where we applied identity Eq. (2.25) to $p-p^{k+1}=(p-p^k)+(p^k-p^{k+1})$. We also observe that Eq. (2.15) also implies

$$\beta\left|\left|r^{k+1}\right|\right|^2=\frac{1}{\beta}\left|\left|p^{k+1}-p^k\right|\right|^2.$$

We combine the above three equalities and obtain

$$-2(p^{k+1}-p)'r^{k+1} = \beta\left|\left|r^{k+1}\right|\right|^2 + \frac{1}{\beta}\left(\left|\left|p-p^{k+1}\right|\right|^2 - \left|\left|p-p^k\right|\right|^2\right). \tag{2.26}$$

We then apply Eq. (2.25) to $r^{k+1}+(z^{k+1}-z^k)$ and therefore can represent the second term on the left-hand side of Eq. (2.24), i.e., $2\beta(r^{k+1})'(z^{k+1}-z^k)$, as,

$$\begin{aligned} &2\beta(r^{k+1})'(z^{k+1}-z^k) \\ &\quad= -\beta\left|\left|r^{k+1}\right|\right|^2 - \beta\left|\left|(z^{k+1}-z^k)\right|\right|^2 + \beta\left|\left|r^{k+1}+(z^{k+1}-z^k)\right|\right|^2 \end{aligned} \tag{2.27}$$

The third term can be expressed similarly. Based on the identity $(z^k-z) = (z^k-z^{k+1})+(z^{k+1}-z)$ and Eq. (2.25), we obtain

$$\left|\left|(z^k-z)\right|\right|^2 = \left|\left|(z^k-z^{k+1})\right|\right|^2 + \left|\left|(z^{k+1}-z)\right|\right|^2 + 2(z^k-z^{k+1})'(z^{k+1}-z),$$

which implies the third term of the left-hand side of Eq. (2.24), i.e., $-2\beta(z^{k+1}-z^k)'(z-z^{k+1})$, can be written as

$$\begin{aligned} &-2\beta(z^{k+1}-z^k)'(z-z^{k+1}) \\ &\quad= \beta\left|\left|(z^k-z^{k+1})\right|\right|^2 + \beta\left|\left|(z^{k+1}-z)\right|\right|^2 - \beta\left|\left|(z^k-z)\right|\right|^2. \end{aligned} \tag{2.28}$$

By combing the equivalent representations for all three terms [cf. Eq. s (2.26), (2.27) and (2.28)], we have

$$\begin{aligned} &-2(p^{k+1}-p)'(r^{k+1}) + 2\beta(r^{k+1})(z^{k+1}-z^k) - 2\beta(z^{k+1}-z^k)'(z-z^{k+1}) \\ =&\beta\left(\left|\left|r^{k+1}\right|\right|^2 - \left|\left|r^{k+1}\right|\right|^2\right) + \beta\left(\left|\left|(z^k-z^{k+1})\right|\right|^2 - \left|\left|(z^{k+1}-z^k)\right|\right|^2\right) \\ &+\frac{1}{\beta}\left(\left|\left|p-p^{k+1}\right|\right|^2 - \left|\left|p-p^k\right|\right|^2\right) \\ &+\beta\left|\left|r^{k+1}+(z^{k+1}-z^k)\right|\right|^2 + \beta\left|\left|(z^{k+1}-z)\right|\right|^2 - \beta\left|\left|(z^k-z)\right|\right|^2. \end{aligned}$$

The terms in the first two parentheses on the right hand side cancel out, establishing the desired result. □

We can now combine results from the preceding two lemmas and obtain a bound on the key quantity $F(x^{k+1}) - p'r^{k+1}$.

Theorem 2.4.2. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the distributed ADMM algorithm and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (2.8). The following holds at each iteration k:*

$$\begin{aligned} F(x^*) - F(x^{k+1}) + p'r^{k+1} \geq \frac{1}{2\beta}\left(\left|\left|p^{k+1} - p\right|\right|^2 - \left|\left|p^k - p\right|\right|^2\right) \\ + \frac{\beta}{2}\left(\left|\left|(z^{k+1} - z^*)\right|\right|^2 - \left|\left|(z^k - z^*)\right|\right|^2\right), \end{aligned} \tag{2.29}$$

for all p in $\mathbb{R}^{2MN}$, where $r^{k+1} = Dx^{k+1} - z^{k+1}$.

Proof. We substitute $x = x^*$, $z = z^*$ and $r = r^*$ into Eq. (2.17) from Lemma 2.4.3 and obtain

$$F(x^*) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left[(r^* - r^{k+1}) + (z^* - z^{k+1})\right] - (p^{k+1})'(r^* - r^{k+1}) \geq 0.$$

By Lemma 2.4.2 and definition of r^*, we also have $r^* = Dx^* - z^* = 0$. Hence the preceding relation is equivalent to

$$F(x^*) - F(x^{k+1}) + \left[\beta(z^{k+1} - z^k)\right]' \left[(-r^{k+1}) + (z^* - z^{k+1})\right] + (p^{k+1})'(r^{k+1}) \geq 0.$$

We add and subtract a term $p'r^{k+1}$ from the left hand side of the above inequality and have

$$F(x^*) - F(x^{k+1}) + p'r^{k+1} + \left[\beta(z^{k+1} - z^k)\right]' \left[(-r^{k+1}) + (z^* - z^{k+1})\right] + (p^{k+1} - p)'(r^{k+1}) \geq 0.$$

We can now substitute $x = x^*$, $z = z^*$ into Eq. (2.23) from Lemma 2.4.4 and have

$$\begin{aligned} &-(p^{k+1} - p)'r^{k+1} + (r^{k+1})'\beta(z^{k+1} - z^k) - \beta(z^{k+1} - z^k)'(z^* - z^{k+1}) \\ &= \frac{1}{2\beta}\left(\left|\left|p^{k+1} - p\right|\right|^2 - \left|\left|p^k - p\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|(z^{k+1} - z^*)\right|\right|^2 - \left|\left|(z^k - z^*)\right|\right|^2\right) \\ &\quad + \frac{\beta}{2}\left|\left|r^{k+1} + (z^{k+1} - z^k)\right|\right|^2. \end{aligned}$$

These two relations imply

$$\begin{aligned} F(x^*) - F(x^{k+1}) + p'r^{k+1} \geq & \frac{1}{2\beta}\left(\left|\left|p^{k+1}-p\right|\right|^2 - \left|\left|p^k - p\right|\right|^2\right) \\ & + \frac{\beta}{2}\left(\left|\left|(z^{k+1}-z^*)\right|\right|^2 - \left|\left|(z^k - z^*)\right|\right|^2\right) \\ & + \frac{\beta}{2}\left|\left|r^{k+1} + (z^{k+1} - z^k)\right|\right|^2. \end{aligned}$$

By dropping the nonnegative term $\frac{\beta}{2}\left|\left|r^{k+1} + (z^{k+1} - z^k)\right|\right|^2$ from the right hand side, we establish the desired relation. □

2.4.2 Convergence and Rate of Convergence

This section studies convergence and rate of convergence properties of the algorithm. The rate will be established based on the ergodic time average of the sequences $\{x^k, z^k\}$ generated by the distributed ADMM algorithm , $\bar{x}(T)$ and $\bar{z}(T)$, defined by

$$\bar{x}(T) = \frac{1}{T}\sum_{k=1}^{T} x^k, \qquad \bar{z}(T) = \frac{1}{T}\sum_{k=1}^{T} z^k. \tag{2.30}$$

For notational convenience, we also define the ergodic time average of the residual by

$$\bar{r}(T) = D\bar{x}(T) - \bar{z}(T). \tag{2.31}$$

This lemma shows that the Lagrangian function value converges to the optimal one with rate $O(1/T)$ and will be used to show that the feasibility violation diminishes to 0 with the same rate.

Lemma 2.4.5. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the distributed ADMM algorithm and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (2.8). The following hold at each iteration T:*

$$\begin{aligned} & F(\bar{x}(T)) - p'\bar{r}(T) - F(x^*) \leq \\ & \frac{1}{2T\beta}\left|\left|p^0 - p\right|\right|^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2, \end{aligned} \tag{2.32}$$

for all p in $\mathbb{R}^{2MN}$, the dual space, where $\bar{r}(T)$ is defined as in Eq. (2.31).

Proof. Theorem 2.4.2 gives

$$F(x^*) - F(x^{k+1}) + p' r^{k+1} \geq \frac{1}{2\beta}\left(\left|\left|p^{k+1} - p\right|\right|^2 - \left|\left|p^k - p\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|z^{k+1} - z^*\right|\right|^2 - \left|\left|z^k - z^*\right|\right|^2\right),$$

for all k. We sum up this relation for $k = 0, \ldots, T-1$ and by telescoping cancellation, we have

$$TF(x^*) - \sum_{k=0}^{T-1} F(x^{k+1}) + p' \sum_{k=0}^{T-1} r^{k+1} \geq \frac{1}{2\beta}\left(\left|\left|p^T - p\right|\right|^2 - \left|\left|p^0 - p\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|z^T - z^*\right|\right|^2 - \left|\left|z^0 - z^*\right|\right|^2\right).$$

We note that by linearity of matrix vector multiplication,

$$\sum_{k=0}^{T-1} r^{k+1} = \sum_{k=0}^{T-1} Dx^{k+1} - z^{k+1} = T(D\bar{x}(T) - \bar{z}(T)) = T\bar{r}(T).$$

Convexity of function F implies $\sum_{k=0}^{T-1} F(x^{k+1}) \geq TF(\bar{x}(T))$, and thus

$$T(F(x^*) - F(\bar{x}(T)) + p'\bar{r}(T)) \geq \frac{1}{2\beta}\left(\left|\left|p^T - p\right|\right|^2 - \left|\left|p^0 - p\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|z^T - z^*\right|\right|^2 - \left|\left|z^0 - z^*\right|\right|^2\right).$$

We can drop the nonnegative terms $\frac{1}{2\beta}\left|\left|p^T - p\right|\right|^2 + \frac{\beta}{2}\left|\left|z^T - z^*\right|\right|^2$ and divide both sides by $-T$ to obtain

$$F(\bar{x}(T)) - p'\bar{r}(T) - F(x^*) \leq \frac{1}{2T\beta}\left|\left|p^0 - p\right|\right|^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2,$$

which is the desired result. □

Theorem 2.4.3. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the distributed ADMM algorithm, $\bar{x}(T)$ be the time average of the sequence $\{x^t\}$ up to time T as defined in*

Eq. (2.30) and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (2.8). The following hold at each iteration T:

$|F(\bar{x}(T)) - F(x^*)| \leq \frac{\beta}{2T} \left|\left|z^0 - z^*\right|\right|^2 + \frac{1}{2T\beta} max\{\left|\left|p^0\right|\right|^2, \left|\left|p^0 - 2p^*\right|\right|^2\}.$

Proof. We derive the bound on the absolute value by considering bounds for both term $F(\bar{x}(T)) - F(x^*)$ and $F(x^*) - F(\bar{x}(T))$ and then take the maximum of the two.

The difference $F(\bar{x}(T)) - F(x^*)$ can be bounded by substituting $p = 0$ into Eq. (2.20), summing over from $k = 0$ to T, and obtain

$$F(\bar{x}(T)) - F(x^*) \leq \frac{\beta}{2T} \left|\left|z^0 - z^*\right|\right|^2 + \frac{1}{2T\beta} \left|\left|p^0\right|\right|^2. \tag{2.33}$$

For the difference $F(x^*) - F(\bar{x}(T))$, we use the saddle point property of (x^*, z^*, p^*) and have

$$F(x^*) - F(\bar{x}(T)) + p^*\bar{r}(T) \leq 0, \tag{2.34}$$

which implies that

$$F(x^*) - F(\bar{x}(T)) \leq -p^*\bar{r}(T).$$

We next derive a bound on the right hand side term $-p^*\bar{r}(T)$. The following bound can be obtained by adding the term $p^*\bar{r}(T)$ to both sides of Eq. (2.34) and then multiply both sides by negative 1,

$$F(\bar{x}(T)) - F(x^*) - 2p^*\bar{r}(T) \geq -p^*\bar{r}(T).$$

The left hand side can be further bounded by substituting $p = 2p^*$ in Eq. (2.32), i.e., $F(\bar{x}(T)) - F(x^*) - 2p^*r(\bar{T}) \leq \frac{\beta}{2T} \left|\left|z^0 - z^*\right|\right|^2 + \frac{1}{2T\beta} \left|\left|p^0 - 2p^*\right|\right|^2$. Combining the three preceding inequalities gives

$$F(x^*) - F(\bar{x}(T)) \leq \frac{\beta}{2T} \left|\left|z^0 - z^*\right|\right|^2 + \frac{1}{2T\beta} \left|\left|p^0 - 2p^*\right|\right|^2$$

The above relation and Eq. (2.33) together yield the desired relation. □

The following lemma shows that the feasibility violation of the sequence generated by the distributed ADMM algorithm converges to 0 with rate $O(1/T)$.

Theorem 2.4.4. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the distributed ADMM algorithm (2.11)-(2.13) and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (2.8) and average residual $\bar{r}(T)$ be defined as in (2.31). The following hold at each iteration T:*

$$||\bar{r}(T)|| \leq \frac{1}{2T\beta}\left(||p^0 - p^*|| + 1\right)^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2.$$

Proof. For each T, we let $p = p^* - \frac{\bar{r}(T)}{||\bar{r}(T)||}$. We note that the product term $p'\bar{r}(T) = p^*\bar{r}(T) - ||\bar{r}(T)||$ and Eq. (2.32) becomes $F(\bar{x}(T)) - p^*\bar{r}(T) - F(x^*) + ||\bar{r}(T)|| \leq \frac{1}{2T\beta}\left|\left|p^0 - p^* - \frac{\bar{r}(T)}{||\bar{r}(T)||}\right|\right|^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2$.

We next derive an upper bound for the term $\left|\left|p^0 - p^* - \frac{\bar{r}(T)}{||\bar{r}(T)||}\right|\right|^2$ for all T. The term $\frac{\bar{r}(T)}{||\bar{r}(T)||}$ is a unit vector, and the optimal solution for the problem

$$\max_{||\alpha||\leq 1} ||b - \alpha||^2,$$

for some arbitrary vector b is given by $\alpha = -\frac{b}{||b||}$ with optimal function value of $(||b|| + 1)^2$. Therefore

$$\left|\left|p^0 - p^* - \frac{\bar{r}(T)}{||\bar{r}(T)||}\right|\right|^2 \leq \left(||p^0 - p^*|| + 1\right)^2.$$

The preceding two relations imply

$$F(\bar{x}(T)) - p^*\bar{r}(T) - F(x^*) + ||\bar{r}(T)|| \leq \frac{1}{2T\beta}\left(||p^0 - p^*|| + 1\right)^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2.$$

Since point (x^*, z^*, p^*) is a saddle point, we have $L(x^*, z^*, p^*) - L(\bar{x}(T), \bar{z}(T), p^*) = F(x^*) - F(\bar{x}(T))) + p^{*\prime}\bar{r}(T) \leq 0$, for any T.

By adding the above two relations, we have

$$||\bar{r}(T)|| \leq \frac{1}{2T\beta}\left(||p^0 - p^*|| + 1\right)^2 + \frac{\beta}{2T}\left|\left|z^0 - z^*\right|\right|^2.$$

□

Theorems 2.4.3 and 2.4.4 establish that the sequence generated by the algorithm converges to a point which is both feasible and attains optimal value at rate $O(1/T)$, and hence we conclude that the algorithm converges with rate $O(1/T)$.

2.4.3 Effect of Network Structure on Performance

In this section, we analyze the effect of network structure on the algorithm performance. In order to focus our analysis on the terms related to the network topology, we will study the bounds from Theorems 2.4.3 and 2.4.4 under the assumption that the initial point of the algorithm has all 0 for both primal and dual variables, i.e., $x^0 = 0$, $z^0 = 0$ and $p^0 = 0$. We will first bound the norm of the optimal dual variable, $||p^*||^2$, using the properties of the underlying graph topology and then use that in the bounds from the previous section.

Our bounds will be related to the *graph Laplacian matrix* $L(G)$, which is of dimension N by N, defined by

$$[L(G)]_{ij} = \begin{cases} \text{degree}(i) & i = j, \\ -1 & (i,j) \in E. \end{cases}$$

For a connected graph the eigenvalues of the graph Laplacian satisfies the following property [76].

Lemma 2.4.6. *Consider an undirected connected graph G with N nodes, let $L(G)$ denote the associated graph Laplacian. Then the matrix $L(G)$ is positive semidefinite. Scalar 0 is an eigenvalue with multiplicity 1, whose associated eigenvector is the vector of all 1 in $\mathbb{R}^N$.*

We denote by $\rho_2(L(G))$ the second smallest (smallest positive) eigenvalue of the matrix $L(G)$. This quantity is often referred to as the *algebraic connectivity* of the graph. A well connected graph has large $\rho_2 L(G)$. Our bound on $||p^*||^2$ will depend on this connectivity measure.

The following lemma establishes a key property of the optimal dual variable, which simplifies our analysis on p^* significantly.

Lemma 2.4.7. *Let (x^*, z^*, p^*) be a saddle point of problem (2.8). Then we have*

$$p_{ij}^* = -p_{ji}^*$$

for any pair of nodes i and j with edge (i, j).

Proof. Since the point (x^*, z^*, p^*) is a saddle point, we have that the primal pair (x^*, z^*) minimizes the augmented Lagrangian function given p^*, i.e.,

$$x^* \in \underset{x}{\operatorname{argmin}}\, L_\beta(x, z^*, p^*), z^* \in \underset{z \in Z}{\operatorname{argmin}}\, L_\beta(x^*, z, p^*),$$

where we recall that $L_\beta(x, z, p) = F(x) - p'(Dx - z) + \frac{\beta}{2} \left|\left|Dx - z\right|\right|^2$.

By Lemma 2.4.2, we also have $p^* = p^* - \beta(Dx^* - z^*) = p^*$.

Hence the point (x^*, z^*, p^*) is a fixed point of the ADMM iteration.

We consider an arbitrary edge (i, j). The z^* update in the algorithm is a quadratic program with linear constraint, i.e., $z_{ij}^*, z_{ji}^* = \operatorname{argmin}_{z_{ij}, z_{ji}, z_{ij} = z_{ji}} -(p_{ij}^*)'(x_i^* - z_{ij}) - (p_{ji}^*)'(x_j^* - z_{ji}) + \frac{\beta}{2} \left(\left|\left|x_i^* - z_{ij}\right|\right|^2 + \left|\left|x_j^* - z_{ji}\right|\right|^2 \right)$. The optimal solution can be written as

$$z_{ij}^* = \frac{1}{\beta}(p_{ij}^* - v^*) + x_i^*, \qquad z_{ji}^* = \frac{1}{\beta}(p_{ji}^* + v^*) + x_j^*,$$

where v^* is the Lagrange multiplier associated with the constraint $z_{ij} - z_{ji} = 0$. We use $x_i^* = z_{ij}^*$ from Lemma 2.4.2 one more time and the previous relations can be written as $p_{ij}^* = v^*, \qquad p_{ij}^* = -v^*$. We can hence conclude $p_{ij}^* = -p_{ji}^*$. □

The next theorem establishes a bound on the norm of the optimal dual variable using the underlying network properties.

Theorem 2.4.5. *There exists an optimal primal-dual solution for problem (2.8), denoted by (x^*, z^*, p^*), that satisfies*

$$\left|\left|p^*\right|\right|^2 \leq \frac{2Q^2}{\rho_2(L(G))}, \tag{2.35}$$

where Q is a bound on $||\partial F(x^*)||$ with $\partial F(x^*)$ denoting the set of subgradients of function F at point x^* and $\rho_2(L(G))$ is the second smallest positive eigenvalue of the Laplacian matrix $L(G)$ of the underlying graph.

Proof. Note that since the sub gradient of a real-valued convex function is compact, we have that $Q < \infty$. Moreover, since problem (2.8) has nonempty optimal solution set with linear constraint and a polyhedral set constraint Z, strong duality holds and there exists at least one primal-dual optimal solution.

Based on Lemma 2.4.7, to bound the norm of the optimal dual multiplier p^*, we only need to consider half of the elements. We define a bijection mapping between $\mathbb{R}^{nM}$ and $\{p : \mathbb{R}^{2nM}, p_{ij} = -p_{ji}\}$, with

$$\lambda(p) = [p_{ij}]_{i<j}, \quad [p(\lambda)]_{ij} = \begin{cases} \lambda & i < j, \\ -\lambda & j > i. \end{cases}$$

We have

$$||p(\lambda)||^2 = 2\,||\lambda||^2\,, \tag{2.36}$$

for all λ in $\mathbb{R}^{nM}$. We will derive a bound on $\lambda(p^*)$ for some optimal dual solution p^* by using the optimality of p^* and then use the preceding equality to obtain desired relation.

For any dual optimal solution p^*, vectors $\lambda(p^*)$ and p^* are also via the node-edge incidence matrix, denoted by A in $\mathbb{R}^{nM\times nN}$, where Ax is the compact representation of the vector $[x_i - x_j]_{(i,j)\in E, i<j}$. By definition of matrix D, we have

$$(p^*)'Dx = (\lambda(p^*))'Ax \tag{2.37}$$

for any x in $\mathbb{R}^{nN}$.

The necessary and sufficient conditions for a primal-dual solution (x^*, z^*, p^*) to be optimal are $Dx^* - z^* = 0$ [cf. Lemma 2.4.2] and (x^*, z^*) minimizes the Lagrangian function $F(x) - (p^*)'(Dx - z)$, i.e.,

$$F(x^*) - (p^*)'(Dx^* - z^*) - F(x) + (p^*)'(Dx - z) \leq 0,$$

$$Dx^* - z^* = 0,$$

where x is an arbitrary vector in $\mathbb{R}^{nN}$, z is an arbitrary vector in Z. By combining Lemma 2.4.7 and the fact that $z_{ij} = z_{ji}$ for any z in Z, we have

$$(p^*)'z = \sum_{(i,j)\in E} p^*_{ij} z_{ij} - p^*_{ij} z_{ji} = 0.$$

Therefore we can rewrite the preceding inequality as

$$F(x^*) - (p^*)'Dx^* - F(x) + (p^*)'Dx \leq 0, \tag{2.38}$$

for any vector x.

We can now use Eq. (2.37) to rewrite Eq. (2.38) as

$$F(x^*) - (\lambda(p^*))'Ax^* - F(x) + (\lambda(p^*))'Ax \leq 0,$$

for any x in $\mathbb{R}^{nN}$. This optimality condition is equivalent to

$$A'\lambda(p^*) \in \partial F(x^*).$$

We let v denote that particular subgradient, i.e.,

$$A'\lambda(p^*) = v, \tag{2.39}$$

for some dual optimal solution p^*. Such v exists due to strong duality.

We observe that $(x^*, z^*, p(\lambda))$, for any λ satisfying $A'\lambda = v$ is a primal-dual optimal solution for problem (2.8). We next find the λ with the minimal norm and show that it satisfies the desired condition. This λ is the solution of the following quadratic program,

$$\min_{A'\lambda=v} \frac{1}{2} ||\lambda||^2 . \tag{QP}$$

We denote by μ^* the optimal dual multiplier associated with the constraint $A'\lambda = v$, which exists because (QP) is a feasible quadratic program. The optimal solution to (QP) given μ, denoted by λ^*, satisfies

$$\lambda^* \in \operatorname*{argmin}_{\lambda} \frac{1}{2} ||\lambda||^2 - (\mu^*)'(A'\lambda - v).$$

We can compute λ^* by taking the first order condition and have

$$\lambda^* = A\mu^*. \tag{2.40}$$

The above relation and Eq. (2.39) imply

$$A'A\mu^* = v. \tag{2.41}$$

The real symmetric matrix $A'A$ can be decomposed into $A'A = U\Lambda U'$, where U is orthonormal and Λ is diagonal with the diagonal elements equal to the eigenvalues of $A'A$. The pseudo-inverse of matrix $A'A$ denoted by $(A'A)^\dagger$ is given by $(A'A)^\dagger = U\Lambda^\dagger U'$, where $\Lambda^\dagger$ is diagonal with

$$\Lambda^\dagger_{ii} = \begin{cases} \frac{1}{\Lambda_{ii}} & \text{if} \quad \Lambda_{ii} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

The product $A'A$ is in $\mathbb{R}^{nN\times nN}$ and the (i, j) block of size $\mathbb{R}^{n\times n}$ is -1 is agents i and j are connected, the i^{th} diagonal block of size $n \times n$ is an identity matrix $I(n)$ multiplied by the degree of agent i, hence $A'A = L(G) \otimes I(n)$, where $\otimes$ denotes the Kronecker product. Hence the set of distinct eigenvalues of matrix $A'A$ are the same as those of matrix $L(G)$. This observation and Lemma 2.4.6 suggest that the null space of $A'A$ is the span of vector of all 1.

Therefore, based on Eq. (2.41), we can write $\mu^* = (A'A)^\dagger v + \alpha e$, where $(A'A)^\dagger$

is the pseudo-inverse of the matrix $A'A$, α is any scalar, and e is the vector of all 1 in $\mathbb{R}^{nN}$. We can then substitute this relation into Eq. (2.40) and obtain $\lambda^* = A\left((A'A)^{\dagger}v + \alpha e\right) = A(A'A)^{\dagger}v$, where we used the structure of matrix A to deduce that $Ae = 0$.

Hence we can now bound the norm of λ^* by $||\lambda^*||^2 = v'((A'A)^{\dagger})'A'A(A'A)^{\dagger}v = v'(A'A)^{\dagger}v \leq ||v||^2 \left||(A'A)^{\dagger}\right||$, where we used the property $B^{\dagger\prime}BB^{\dagger} = B^{\dagger}$ for B symmetric.

From definition of matrix $(A'A)^{\dagger}$ in Eq. (2.42) and the fact that $A'A$ share the same set of distinct eigenvalues as $L(G)$, we have

$$\left||(A'A)^{\dagger}\right|| \leq \frac{1}{\min_{i,\Lambda_{ii}>0} \Lambda_{ii}} = \frac{1}{\rho_2(L(G))}.$$

By definition of scalar Q, we also have $||v|| \leq Q$.

The preceding three relations imply that $||\lambda^*||^2 \leq \frac{Q^2}{\rho_2(L(G))}$.

We can now use Eq. (2.36) to relate the norm of λ^* and the dual optimal solution $p(\lambda^*)$. Thus we have established Eq. (2.35) for $p^* = p(\lambda^*)$. □

Theorem 2.4.6. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the distributed ADMM algorithm (2.11)-(2.13) with initialization $x^0 = 0$, $z^0 = 0$, $p^0 = 0$ and average residual $\bar{r}(T)$ be defined as in (2.31). Let y in $\mathbb{R}^n$ denote an optimal solution to the original problem (2.6), scalar Q be a bound on $\partial F(x^*)$ for some optimal primal variable x^* and $L(G)$ be the graph Laplacian matrix of the underlying network. The following hold at each iteration T:*

$$|F(\bar{x}(T)) - F^*| \leq \frac{\beta M ||y||^2}{T} + \frac{4Q^2}{T\beta\rho_2(L(G))}$$

$$||\bar{r}(T)|| \leq \frac{1}{2T\beta}\left(\frac{Q\sqrt{2}}{\sqrt{\rho_2(L(G))}} + 1\right)^2 + \frac{\beta M ||y||^2}{T},$$

and $\max_{i,j} \|\bar{x}_i(T) - \bar{x}_j(T)\|_1 \leq \frac{2}{T\beta}\left(\frac{Q^2}{\rho_2(L(G))} + nd(G)\right) + \frac{\beta M ||y||^2}{T}$, where M is the number of edges in the network, $d(G)$ is the diameter of the underlying graph and $\rho_2(L(G))$ is the second smallest eigenvalue (smallest positive eigenvalue) of matrix $L(G)$.

Proof. Under this particular initialization, the bounds from Theorems 2.4.3 and 2.4.4 become

$$|F(\bar{x}(T)) - F^*| \leq \frac{\beta}{2T} \left|\left|z^*\right|\right|^2 + \frac{1}{2T\beta} \left|\left|2p^*\right|\right|^2, \tag{2.43}$$

and

$$\left|\left|\bar{r}(T)\right|\right| \leq \frac{1}{2T\beta} \left(\left|\left|p^*\right|\right| + 1\right)^2 + \frac{\beta}{2T} \left|\left|z^*\right|\right|^2, \tag{2.44}$$

which hold for any primal-dual optimal solution (x^*, z^*, p^*). We use Theorem 2.4.5 and have $||p^*|| \leq \frac{Q\sqrt{2}}{\sqrt{\rho_2(L(G))}}$. Since (x^*, z^*, p^*) is a primal-dual optimal solution, we have $[z^*]_{ij} = y$ for all pair (i, j) in E and thus $||z^*|| = \sqrt{2M}\,||y||$. The term $||z^*||^2$ can be written as $||z^*||^2 = 2M\,||y||^2$. We can finally use the preceding two relations into Eqs. (2.43) and (2.44) and obtain the first two desired bounds.

To obtain the last bound, we observe that if agents i, j are connected, then $x_i - x_j$ can be obtained by $x_i - x_j = (x_i - z_{ij}) - (x_j - z_{ij})$. Moreover, for any pair of nodes i, j, we have $\|x_i - x_j\|_1 = w'(Dx - z)$, where w is a vector in $\mathbb{R}^{2nM}$ of $-1, 0, 1$ formed by traversing a path from i to j and adding pairwise difference.

For each pair of agents i, j, we substitute $p = p^* - w$ into Eq. (2.32) where w is defined to be such that $w'(D\bar{x}(T) - \bar{z}(T)) = \|\bar{x}_i(T) - \bar{x}_j(T)\|_1$. We then obtain $\|\bar{x}_i(T) - \bar{x}_j(T)\|_1 \leq F(x^*) - F(\bar{x}(T)) + p^*\bar{r}(T) + \frac{1}{2T\beta} ||p^0 - p^* + w||^2 + \frac{\beta}{2T} ||z^0 - z^*||^2$. By optimality of the primal-dual pair (x^*, z^*, p^*), we have $F(x^*) - F(\bar{x}(T)) + p^*\bar{r}(T) \leq 0$. Thus, the preceding relation implies $\|\bar{x}_i(T) - \bar{x}_j(T)\|_1 \leq \frac{1}{2T\beta} ||p^0 - p^* + w||^2 + \frac{\beta}{2T} ||z^0 - z^*||^2$. We then use the fact that $||a + b||^2 \leq 2\,||a||^2 + 2\,||b||^2$ and the fact that $p^0 = 0$ and $z^0 = 0$.

$$\|\bar{x}_i(T) - \bar{x}_j(T)\|_1 \leq \frac{1}{T\beta} \left|\left|p^*\right|\right|^2 + \frac{1}{T\beta} \left|\left|w\right|\right|^2 + \frac{\beta}{2T} \left|\left|z^*\right|\right|^2.$$

The right hand side is independent of i and j and therefore holds for the maximum pairwise difference, i.e.,

$$\max_{i,j} \|\bar{x}_i(T) - \bar{x}_j(T)\|_1 \leq \frac{1}{T\beta} \left|\left|p^*\right|\right|^2 + \frac{1}{T\beta} \left|\left|w\right|\right|^2 + \frac{\beta}{2T} \left|\left|z^*\right|\right|^2.$$

By the definition of diameter of a graph we obtain $||w||^2 \leq 2nd(G)$ and using the previously mentioned bounds on $||p^*||$ and $||z^*||$ the last bound is obtained. $\square$

Hence we observe that the number of edges, diameter of the graph and the algebraic graph connectivity all affect the algorithm performance. The ideal graph would be one with few links, small diameter and good connectivity. Expander graph, which is a well connected sparse graph, for instance, would be a candidate for a graph where the upper bounds suggest that our distributed ADMM algorithm converges fast.

2.5 Summaries

In this chapter, we present a fully distributed Alternating Direction Method of Multipliers (ADMM) based method. We analyze the convergence property of the algorithm and establish that the algorithm achieves $O(1/k)$ rate of convergence, which is the best known rate of convergence for this general class of convex optimization problems. We also show that the convergence speed is affected by the underlying graph topology through algebraic connectivity, diameter and the number of edges. Future work includes extending this network effect analysis to asynchronous implementation and time-varying network topology.

Chapter 3

Asynchronous Distributed ADMM Based Method

3.1 Introduction

In this chapter, we consider a problem more general than (1.1) and we will show that (1.1) is a special case and can be solved using the algorithm developed in this chapter. More specifically, we consider the following optimization problem with a separable objective function and linear constraints:

$$\min_{x_i \in X_i, z \in Z} \quad \sum_{i=1}^{N} f_i(x_i) \tag{3.1}$$
$$s.t. \quad Dx + Hz = 0.$$

Here each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a (possibly nonsmooth) convex function, X_i and Z are closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^W$, and D and H are matrices of dimensions $W \times nN$ and $W \times W$. The decision variable x is given by the partition $x = [x_1', \ldots, x_N']' \in \mathbb{R}^{nN}$, where the $x_i \in \mathbb{R}^n$ are components (subvectors) of x. We denote by set X the product of sets X_i, hence the constraint on x can be written compactly as $x \in X$.

As in previous chapter, problem (1.1) can be equivalently written as

$$\min_{x_i \in X} \quad \sum_{i=1}^{N} f_i(x_i) \tag{3.2}$$
$$s.t. \quad Ax = 0,$$

where $A \in \mathbb{R}^{Mn \times Nn}$ is the edge-node incidence matrix[1] of network G. Variable x is the vector $[x_1, x_2, \dots, x_N]'$. We will refer to this formulation as the *edge-based reformulation* of the multi-agent optimization problem. Note that this formulation is a special case of problem (3.1) with $D = A$, $H = 0$ and $X_i = X$ for all i.

In this chapter, we focus on the more general formulation (3.1) and propose an asynchronous decentralized algorithm based on the classical Alternating Direction Method of Multipliers (ADMM). Asynchrony is highly desirable due to the lack of central processing and coordinating unit in large scale networks. We adopt the following asynchronous implementation for our algorithm: at each iteration k, a random subset Ψ^k of the constraints are selected, which in turn selects the components of x that appear in these constraints. We refer to the selected constraints as *active constraints* and selected components as the *active components (or agents)*. We design an ADMM-type primal-dual algorithm, which at each iteration updates the primal variables using partial information about the problem data, in particular using cost functions corresponding to active components and active constraints, and updates the dual variables corresponding to the active constraints. In the context of the edge-based reformulated multi-agent optimization problem (3.2), this corresponds to a fully decentralized and asynchronous implementation in which a subset of the edges are randomly activated (for example according to local clocks associated with those edges) and the agents incident to those edges perform computations on the basis of their local objective functions followed by communication of updated values with neighbors.

[1] The edge-node incidence matrix of network G is defined as follows: Each n-row block of matrix A corresponds to an edge in the graph and each n-column block represent a node. The n rows corresponding to the edge $e = (i, j)$ has $I(n \times n)$ in the i^{th} n-column block, $-I(n \times n)$ in the j^{th} $n-$column block and 0 in the other columns, where $I(n \times n)$ is the identity matrix of dimension n.

Under the assumption that each constraint has a positive probability of being selected and the constraints have a decoupled structure (which is satisfied by reformulations of the distributed multi-agent optimization problem), our first result shows that the (primal) asynchronous iterates generated by this algorithm converge almost surely to an optimal solution. Our proof relies on relating the asynchronous iterates to *full-information* iterates that would be generated by the algorithm that uses full information about the cost functions and constraints at each iteration. In particular, we introduce a weighted norm where the weights are given by the inverse of the probabilities with which the constraints are activated and construct a Lyapunov function for the asynchronous iterates using this weighted norm. Our second result establishes a *performance guarantee of* $O(1/k)$ for this algorithm under a compactness assumption on the constraint sets X and Z, which to our knowledge is faster than the guarantees available in the literature for this problem. More specifically, we show that the expected value of the difference of the objective function value and the optimal value as well as the expected feasibility violation converges to 0 at rate $O(1/k)$.

This chapter is organized as follows: in Section 3.2, we focus on the more general formulation (3.1), present the asynchronous ADMM algorithm and apply this algorithm to solve problem (1.1) in a distributed way. Section 3.3 contains our convergence and rate of convergence analysis. We study the numerical performance of our proposed asynchronous distributed ADMM in Section 3.4. Section 3.5 concludes with closing remarks.

3.2 Asynchronous ADMM Algorithm

Extending the standard ADMM algorithm, as described in Section 2.1 from the previous chapter, we present in this section an asynchronous distributed ADMM algorithm. We present the problem formulation and assumptions in Section 3.2.1. In Section 3.2.2, we discuss the asynchronous implementation considered in the rest of this chapter that involves updating a subset of components of the decision vector at each time using partial information about problem data and without need for a global coordinator. Section 3.2.3 contains the details of the asynchronous ADMM algorithm. In Section 3.2.4, we apply the asynchronous ADMM algorithm to solve the distributed multi-agent optimization problem (1.1).

3.2.1 Problem Formulation and Assumptions

We consider the optimization problem given in (3.1), which is restated here for convenience:

$$\begin{aligned} \min_{x_i \in X_i, z \in Z} \quad & \sum_{i=1}^{N} f_i(x_i) \\ s.t. \quad & Dx + Hz = 0. \end{aligned}$$

This problem formulation arises in large-scale multi-agent (or processor) environments where problem data is distributed across N agents, i.e., each agent has access only to the component function f_i and maintains the decision variable component x_i. The constraints usually represent the coupling across components of the decision variable imposed by the underlying connectivity among the agents. Motivated by such applications, we will refer to each component function f_i as the *local objective function* and use the notation $F : \mathbb{R}^{nN} \to \mathbb{R}$ to denote the *global objective function* given by their sum:

$$F(x) = \sum_{i=1}^{N} f_i(x_i). \tag{3.3}$$

Similar to the standard ADMM formulation, we adopt the following assumption.

Assumption 3. *(Existence of a Saddle Point) The Lagrangian function of problem (3.1),*

$$L(x, z, p) = F(x) - p'(Dx + Hz), \tag{3.4}$$

has a saddle point, i.e., there exists a solution-multiplier pair (x^*, z^*, p^*) with

$$L(x^*, z^*, p) \leq L(x^*, z^*, p^*) \leq L(x, z, p^*) \tag{3.5}$$

for all x in X, z in Z and p in $\mathbb{R}^W$.

Moreover, we assume that the matrices have special structure that enables solving problem (3.1) in an asynchronous manner:

Assumption 4. *(Decoupled Constraints) Matrix H is diagonal and invertible. Each row of matrix D has exactly one nonzero element and matrix D has no columns of all zeros.*[2]

The diagonal structure of matrix H implies that each component of vector z appears in exactly one linear constraint. The conditions that each row of matrix D has only one nonzero element and matrix D has no column of zeros guarantee the columns of matrix D are linearly independent and hence matrix $D'D$ is invertible. The condition on matrix D implies that each row of the constraint $Dx + Hz = 0$ involves exactly one x_i. We will see in Section 3.2.4 that this assumption is satisfied by the distributed multi-agent optimization problem that motivates this work.

3.2.2 Asynchronous Algorithm Implementation

In the large scale multi-agent applications descried above, it is essential that the iterative solution of the problem involves computations performed by agents in a decentralized manner (with access to local information) with as little coordination as possible. This necessitates an asynchronous implementation in which some of the agents become active (randomly) in time and update the relevant components of the decision variable using partial and local information about problem data while keeping the rest of the components of the decision variable unchanged. This removes the need for a centralized coordinator or global clock, which is an unrealistic requirement in such decentralized environments.

To describe the asynchronous algorithm implementation we consider in this chapter more formally, we first introduce some notation. We call a partition of the set

[2]We assume without loss of generality that each x_i is involved at least in one of the constraints, otherwise, we could remove it from the problem and optimize it separately. Similarly, the diagonal elements of matrix H are assumed to be non-zero, otherwise, that component of variable z can be dropped from the optimization problem.

$\{1, \dots, W\}$ a *proper partition* if it has the property that if z_i and z_j are coupled in the constraint set Z, i.e., value of z_i affects the constraint on z_j for any z in set Z, then i and j belong to the same partition, i.e., $\{i, j\} \subset \psi$ for some ψ in the partition. We let Π be a proper partition of the set $\{1, \dots, W\}$, which forms a partition of the set of W rows of the linear constraint $Dx + Hz = 0$. For each ψ in Π, we define $\Phi(\psi)$ to be the set of indices i, where x_i appears in the linear constraints in set ψ. Note that $\Phi(\psi)$ is an element of the power set $2^{\{1,\dots,N\}}$.

At each iteration of the asynchronous algorithm, two random variables Φ^k and Ψ^k are realized. While the pair (Φ^k, Ψ^k) is correlated for each iteration k, these variables are assumed to be independent and identically distributed across iterations. At each iteration k, first the random variable Ψ^k is realized. The realized value, denoted by ψ^k, is an element of the proper partition Π and selects a subset of the linear constraints $Dx + Hz = 0$. The random variable Φ^k then takes the realized value $\phi^k = \Phi(\psi^k)$. We can view this process as activating a subset of the coupling constraints and the components that are involved in these constraints. If $l \in \psi^k$, we say constraint l as well as its associated dual variable p_l is *active* at iteration k. Moreover, if $i \in \Phi(\psi^k)$, we say that component i or agent i is *active* at iteration k. We use the notation $\bar{\phi}^k$ to denote the complement of set ϕ^k in set $\{1, \dots, N\}$ and similarly $\bar{\psi}^k$ to denote the complement of set ψ^k in set $\{1, \dots, W\}$.

Our goal is to design an algorithm in which at each iteration k, only active components of the decision variable and active dual variables are updated using local cost functions of active agents and active constraints. When an agent becomes active, it will update using the current locally available information (potentially outdated) and hence no central coordination or waiting is required. To that end, we define $f^k : \mathbb{R}^{nN} \to \mathbb{R}$ as the sum of the local objective functions whose indices are in the subset ϕ^k:

$$f^k(x) = \sum_{i \in \phi^k} f_i(x_i),$$

We denote by D_i the matrix in $\mathbb{R}^{W \times nN}$ that picks up the columns corresponding to x_i from matrix D and has zeros elsewhere. Similarly, we denote by H_l the diagonal

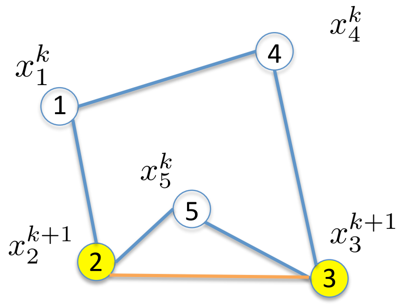

Figure 3-1: Asynchronous algorithm illustration: When edge $(2, 3)$ is active, only agents 2 and 3 perform update, while all the other agents stay at their previous iteration value.

matrix in $\mathbb{R}^{W \times W}$, which picks up the element in the l^{th} diagonal position from matrix H and has zeros elsewhere. Using this notation, we define the matrices

$$D_{\phi^k} = \sum_{i \in \phi^k} D_i, \quad \text{and} \quad H_{\psi^k} = \sum_{l \in \psi^k} H_l.$$

We impose the following condition on the asynchronous algorithm.

Assumption 5. *(Infinitely Often Update) For all k and all ψ in the proper partition Π,*

$$\mathbb{P}(\Psi^k = \psi) > 0.$$

This assumption ensures that each element of the partition Π is active infinitely often with probability 1. Since matrix D has no columns of all zeros, each of the x_i is involved in some constraints, and hence $\cup_{\psi \in \Pi} \Phi(\psi) = \{1, \ldots, N\}$. The preceding assumption therefore implies that each agent i belongs to at least one set $\Phi(\psi)$ and therefore is active infinitely often with probability 1. From definition of the partition Π, we have $\cup_{\psi \in \Pi} \psi = \{1, \ldots, W\}$. Thus, each constraint l is active infinitely often with probability 1.

3.2.3 Asynchronous ADMM Algorithm

We next describe the asynchronous ADMM algorithm for solving problem (3.1).

I. Asynchronous ADMM algorithm:

A Initialization: choose some arbitrary x^0 in X, z^0 in Z and $p^0 = 0$.

B At iteration k, random variables Φ^k and Ψ^k takes realizations ϕ^k and ψ^k. Function f^k and matrices D_{ϕ^k}, H_{ψ^k} are generated accordingly.

a The primal variable x is updated as

$$x^{k+1} \in \operatorname*{argmin}_{x \in X} f^k(x) - (p^k)' D_{\phi^k} x + \frac{\beta}{2} \left|\left| D_{\phi^k} x + H z^k \right|\right|^2. \tag{3.6}$$

with $x_i^{k+1} = x_i^k$, for i in $\bar{\phi}^k$.

b The primal variable z is updated as

$$z^{k+1} \in \operatorname*{argmin}_{z \in Z} -(p^k)' H_{\psi^k} z + \frac{\beta}{2} \left|\left| H_{\psi^k} z + D_{\phi^k} x^{k+1} \right|\right|^2. \tag{3.7}$$

with $z_i^{k+1} = z_i^k$, for i in $\bar{\psi}^k$.

c The dual variable p is updated as

$$p^{k+1} = p^k - \beta [D_{\phi^k} x^{k+1} + H_{\psi^k} z^{k+1}]_{\psi^k}. \tag{3.8}$$

We assume that the minimizers in updates (3.6) and (3.7) exist, but need not be unique.[3] The term $\frac{\beta}{2} \left|\left| D_{\phi^k} x + H z^k \right|\right|^2$ in the objective function of the minimization problem in update (3.6) can be written as

$$\frac{\beta}{2} \left|\left| D_{\phi^k} x + H z^k \right|\right|^2 = \frac{\beta}{2} \left|\left| D_{\phi^k} x \right|\right|^2 + \beta (H z^k)' D_{\phi^k} x + \frac{\beta}{2} \left|\left| H z^k \right|\right|^2,$$

where the last term is independent of the decision variable x and thus can be dropped

[3]Note that the optimization in (3.9) and (3.10) are independent of components of x not in ϕ^k and components of z not in ψ^k and thus the restriction of $x_i^{k+1} = x_i^k$, for i not in ϕ^k and $z_i^{k+1} = z_i^k$, for i not in ψ^k still preserves optimality of x^{k+1} and z^{k+1} with respect to the optimization problems in update (3.9) and (3.10).

from the objective function. Therefore, the primal x update can be written as

$$x^{k+1} \in \underset{x \in X}{\operatorname{argmin}}\, f^k(x) - (p^k - \beta H z^k)' D_{\phi^k} x + \frac{\beta}{2} \left|\left|D_{\phi^k} x\right|\right|^2. \tag{3.9}$$

Similarly, the term $\frac{\beta}{2} \left|\left|H_{\psi^k} z + D_{\phi^k} x^{k+1}\right|\right|^2$ in update (3.7) can be expressed equivalently as

$$\frac{\beta}{2} \left|\left|H_{\psi^k} z + D_{\phi^k} x^{k+1}\right|\right|^2 = \frac{\beta}{2} \left|\left|H_{\psi^k} z\right|\right|^2 + \beta (D_{\phi^k} x^{k+1})' H_{\psi^k} z + \frac{\beta}{2} \left|\left|D_{\phi^k} x^{k+1}\right|\right|^2.$$

We can drop the term $\frac{\beta}{2} \left|\left|D_{\phi^k} x^{k+1}\right|\right|^2$, which is constant in z, and write update (3.7) as

$$z^{k+1} \in \underset{z \in Z}{\operatorname{argmin}} -(p^k - \beta D_{\phi^k} x^{k+1})' H_{\psi^k} z + \frac{\beta}{2} \left|\left|H_{\psi^k} z\right|\right|^2, \tag{3.10}$$

The updates (3.9) and (3.10) make the dependence on the decision variables x and z more explicit and therefore will be used in the convergence analysis. We refer to (3.9) and (3.10) as the *primal x and z update* respectively, and (3.8) as the *dual update*.

3.2.4 Special Case: Distributed Multi-agent Optimization

We apply the asynchronous ADMM algorithm to the edge-based reformulation of the multi-agent optimization problem (3.2).[4] Note that each constraint of this problem takes the form $x_i = x_j$ for agents i and j with $(i, j) \in E$. Therefore, this formulation does not satisfy Assumption 4.

We next introduce another reformulation of this problem, used also in Example 4.4 of Section 3.4 in [3], so that each constraint only involves one component of the decision variable.[5] More specifically, we let $N(e)$ denote the agents which are the endpoints of edge e and introduce a variable $z = [z_{eq}]_{\substack{e=1,\dots,M \\ q \in N(e)}}$ of dimension $2M$, one for each endpoint of each edge. Using this variable, we can write the constraint

[4]For simplifying the exposition, we assume $n = 1$ and note that the results extend to $n > 1$.
[5]Note that this reformulation can be applied to any problem with a separable objective function and linear constraints to turn into a problem of form (3.1) that satisfies Assumption 4.

$x_i = x_j$ for each edge $e = (i, j)$ as

$$x_i = z_{ei}, \quad -x_j = z_{ej}, \quad z_{ei} + z_{ej} = 0.$$

The variables z_{ei} can be viewed as an estimate of the component x_j which is known by node i. The transformed problem can be written compactly as

$$\begin{aligned} \min_{x_i \in X, z \in Z} \quad & \sum_{i=1}^{N} f_i(x_i) \\ s.t. \quad & A_{ei} x_i = z_{ei}, \quad e = 1, \ldots, M, \, i \in \mathcal{N}(e), \end{aligned} \tag{3.11}$$

where Z is the set $\{z \in \mathbb{R}^{2M} \mid \sum_{q \in \mathcal{N}(e)} z_{eq} = 0, \; e = 1, \ldots, M\}$ and A_{ei} denotes the entry in the e^{th} row and i^{th} column of matrix A, which is either 1 or -1. This formulation is in the form of problem (3.1) with matrix $H = -I$, where I is the identity matrix of dimension $2M \times 2M$. Matrix D is of dimension $2M \times N$, where each row contains exactly one entry of 1 or -1. In view of the fact that each node is incident to at least one edge, matrix D has no column of all zeros. Hence Assumption 4 is satisfied.

One natural implementation of the asynchronous algorithm is to associate with each edge an independent Poisson clock with identical rates across the edges. At iteration k, if the clock corresponding to edge (i, j) ticks, then $\phi^k = \{i, j\}$ and ψ^k picks the rows in the constraint associated with edge (i, j), i.e., the constraints $x_i = z_{ei}$ and $-x_j = z_{ej}$.[6]

We associate a dual variable p_{ei} in $\mathbb{R}$ to each of the constraint $A_{ei} x_i = z_{ei}$, and denote the vector of dual variables by p. The primal z update and the dual update [Eqs. (3.7) and (3.8)] for this problem are given by

$$\begin{aligned} z_{ei}^{k+1}, z_{ej}^{k+1} = \operatorname*{argmin}_{z_{ei}, z_{ej}, z_{ei} + z_{ej} = 0} & -(p_{ei}^k)'(A_{ei} x_i^{k+1} - z_{ei}) - (p_{ej}^k)'(A_{ej} x_j^{k+1} - z_{ej}) \\ & + \frac{\beta}{2} \left(\left|\left| A_{ei} x_i^{k+1} - z_{ei} \right|\right|^2 + \left|\left| A_{ej} x_j^{k+1} - z_{ej} \right|\right|^2 \right), \end{aligned} \tag{3.12}$$

[6]Note that this selection is a proper partition of the constraints since the set Z couples only the variables z_{ek} for the endpoints of an edge e.

$$p_{eq}^{k+1} = p_{eq}^{k} - \beta(A_{eq}x_q^{k+1} - z_{eq}^{k+1}) \qquad \text{for } q = i, j.$$

The primal z update involves a quadratic optimization problem with linear constraints, which can be solved in closed form. In particular, using first order optimality conditions, we conclude

$$z_{ei}^{k+1} = \frac{1}{\beta}(-p_{ei}^{k} - v^{k+1}) + A_{ei}x_i^{k+1}, \qquad z_{ej}^{k+1} = \frac{1}{\beta}(-p_{ei}^{k} - v^{k+1}) + A_{ej}x_j^{k+1}, \quad (3.13)$$

where v^{k+1} is the Lagrange multiplier associated with the constraint $z_{ei} + z_{ej} = 0$ and is given by

$$v^{k+1} = \frac{1}{2}(-p_{ei}^{k} - p_{ej}^{k}) + \frac{\beta}{2}(A_{ei}x_i^{k+1} + A_{ej}x_j^{k+1}). \quad (3.14)$$

Combining these steps yields the following asynchronous algorithm for problem (3.2), which can be implemented in a decentralized manner by each node i at each iteration k having access to only his local objective function f_i, adjacency matrix entries A_{ei}, and his local variables x_i^k, z_{ei}^k, and p_{ei}^k while exchanging information with one of his neighbors.[7]

II. Asynchronous Edge Based ADMM algorithm:

A Initialization: choose some arbitrary x_i^0 in X and z^0 in Z, which are not necessarily all equal. Initialize $p_{ei}^0 = 0$ for all edges e and end points i.

B At time step k, the local clock associated with edge $e = (i, j)$ ticks,

a Agents i and j update their estimates x_i^k and x_j^k simultaneously as

$$x_q^{k+1} = \operatorname*{argmin}_{x_q \in X} f_q(x_q) - (p^k)' A_q x_q + \frac{\beta}{2} \left|\left| A_q(x_q - z^k) \right|\right|^2$$

for $q = i, j$. The updated components of x_i^{k+1} and x_j^{k+1} are exchanged

[7]The asynchronous ADMM algorithm can also be applied to a node-based reformulation of problem (1.1), where we impose the local copy of each node to be equal to the average of that of its neighbors. This leads to another asynchronous distributed algorithm with a different communication structure in which each node at each iteration broadcasts its local variables to all his neighbors, see [73] for more details.

over the edge e, while all the

b Agents i and j exchange their current dual variables p_{ei}^k and p_{ej}^k over the edge e. For $q = i, j$, agents i and j use the obtained values to compute the variable v^{k+1} as Eq. (3.14), i.e.,

$$v^{k+1} = \frac{1}{2}(-p_{ei}^k - p_{ej}^k) + \frac{\beta}{2}(A_{ei}x_i^{k+1} + A_{ej}x_i^{k+1}).$$

and update their estimates z_{ei}^k and z_{ej}^k according to Eq. (3.13), i.e.,

$$z_{eq}^{k+1} = \frac{1}{\beta}(-p_{eq}^k - v^{k+1}) + A_{eq}x_q^{k+1}.$$

c Agents i and j update the dual variables p_{ei}^{k+1} and p_{ej}^{k+1} as

$$p_{eq}^{k+1} = -v^{k+1} \qquad \text{for } q = i, j.$$

d All other agents keep the same variables as the previous time.

We note that A_q is the column of matrix A associated with agent q. Both of the terms $(p^k)'A_q x_q$ and $A_q(x_q - z^k)$ can be computed using information local to node q, i.e., the information about p_{qj} and z_{qj} for j in the neighborhood of q. Therefore, this algorithm can be implemented in a distributed way. This algorithm is also asynchronous, since agents can wake up according to their local clock and use their current available (potentially outdated) information. In the case where each time an edge becomes active at random, other than coordination between a pair of nodes on the same active edge, no other waiting or coordination is required to implement this asynchronous algorithm.

3.3 Convergence Analysis for Asynchronous ADMM Algorithm

In this section, we study the convergence behavior of the asynchronous ADMM algorithm. We show that the primal iterates $\{x^k, z^k\}$ generated by (3.9) and (3.10) converge almost surely to an optimal solution of problem (3.1). Under the additional assumption that the primal optimal solution set is compact, we further show that the corresponding objective function values converge to the optimal value in expectation at rate $O(1/k)$.

We first recall the relationship between the sets ϕ^k and ψ^k for a particular iteration k, which plays an important role in the analysis. Since the set of active components at time k, ϕ^k, represents all components of the decision variable that appear in the active constraints defined by the set ψ^k, we can write

$$[Dx]_{\psi^k} = [D_{\phi^k} x]_{\psi^k}. \tag{3.15}$$

We next consider a sequence $\{y^k, v^k, \mu^k\}$, which is formed of iterates defined by a "full information" version of the ADMM algorithm in which all constraints (and therefore all components) are active at each iteration. We will show that under the Decoupled Constraints Assumption (cf. Assumption 4), the iterates generated by the asynchronous algorithm (x^k, z^k, p^k) take the values of (y^k, v^k, μ^k) over the sets of active components and constraints and remain at their previous values otherwise. This association enables us to perform the convergence analysis using the sequence $\{y^k, v^k, \mu^k\}$ and then translate the results into bounds on the objective function value improvement along the sequence $\{x^k, z^k, p^k\}$.

More specifically, at iteration k, we define y^{k+1} by

$$y^{k+1} \in \operatorname*{argmin}_{y \in X} F(y) - (p^k - \beta H z^k)' Dy + \frac{\beta}{2} \left\| Dy \right\|^2. \tag{3.16}$$

Due to the fact that each row of matrix D has only one nonzero element [cf. Assumption 4], the norm $||Dy||^2$ can be decomposed as $\sum_{i=1}^{N} ||D_i y_i||^2$, where recall that D_i is the matrix that picks up the columns corresponding to component x_i and is equal to zero otherwise. Thus, the preceding optimization problem can be written as a separable optimization problem over the variables y_i:

$$y^{k+1} \in \sum_{i=1}^{N} \operatorname*{argmin}_{y_i \in X_i} f_i(y_i) - (p^k - \beta H z^k)' D_i y_i + \frac{\beta}{2} ||D_i y_i||^2 .$$

Since $f^k(x) = \sum_{i \in \phi^k} f_i(x_i)$, and $D_{\phi^k} = \sum_{i \in \phi^k} D_i$, the minimization problem that defines the iterate x^{k+1} [cf. Eq. (3.9)] similarly decomposes over the variables x_i for $i \in \Phi^k$. Hence, the iterates x^{k+1} and y^{k+1} are identical over the components in set ϕ^k, i.e., $[x^{k+1}]_{\phi^k} = [y^{k+1}]_{\phi^k}$. Using the definition of matrix D_{ϕ^k}, i.e., $D_{\phi^k} = \sum_{i \in \phi^k} D_i$, this implies the following relation:

$$D_{\phi^k} x^{k+1} = D_{\phi^k} y^{k+1}. \tag{3.17}$$

The rest of the components of the iterate x^{k+1} by definition remain at their previous value, i.e., $[x^{k+1}]_{\bar{\phi}^k} = [x^k]_{\bar{\phi}^k}$.

Similarly, we define vector v^{k+1} in Z by

$$v^{k+1} \in \operatorname*{argmin}_{v \in Z} -(p^k - \beta D y^{k+1})' H v + \frac{\beta}{2} ||Hv||^2 . \tag{3.18}$$

Using the diagonal structure of matrix H [cf. Assumption 4] and the fact that Π is a proper partition of the constraint set [cf. Section 3.2.2], this problem can also be decomposed in the following way:

$$v^{k+1} \in \operatorname*{argmin}_{v, [v]_\psi \in Z_\psi} \sum_{\psi \in \Pi} -(p^k - \beta D y^{k+1})' H_\psi [v]_\psi + \frac{\beta}{2} ||H_\psi [v]_\psi||^2 ,$$

where H_ψ is a diagonal matrix that contains the l^{th} diagonal element of the diagonal matrix H for l in set ψ (and has zeros elsewhere) and set Z_ψ is the projection of set Z on component $[v]_\psi$. Since the diagonal matrix H_{ψ^k} has nonzero elements only on the l^{th} element of the diagonal with $l \in \psi^k$, the update of $[v]_\psi$ is independent of the other components, hence we can express the update on the components of v^{k+1} in set ψ^k as

$$[v^{k+1}]_{\psi^k} \in \underset{v \in Z}{\operatorname{argmin}} -(p^k - \beta D_{\psi^k} x^{k+1})' H_{\psi^k} z + \frac{\beta}{2} \left|\left| H_{\psi^k} z \right|\right|^2 .$$

By the primal z update [cf. Eq. (3.10)], this shows that $[z^{k+1}]_{\psi^k} = [v^{k+1}]_{\psi^k}$. By definition, the rest of the components of z^{k+1} remain at their previous values, i.e., $[z^{k+1}]_{\bar{\psi}^k} = [z^k]_{\bar{\psi}^k}$.

Finally, we define vector μ^{k+1} in $\mathbb{R}^W$ by

$$\mu^{k+1} = p^k - \beta(Dy^{k+1} + Hv^{k+1}). \tag{3.19}$$

We relate this vector to the dual variable p^{k+1} using the dual update [cf. Eq. (3.8)]. We also have

$$[D_{\phi^k} x^{k+1}]_{\psi^k} = [D_{\phi^k} y^{k+1}]_{\psi^k} = [Dy^{k+1}]_{\psi^k},$$

where the first equality follows from Eq. (3.17) and second is derived from Eq. (3.15). Moreover, since H is diagonal, we have $[H_{\psi^k} z^{k+1}]_{\psi^k} = [Hv^{k+1}]_{\psi^k}$. Thus, we obtain $[p^{k+1}]_{\psi^k} = [\mu^{k+1}]_{\psi^k}$ and $[p^{k+1}]_{\bar{\psi}^k} = [p^k]_{\bar{\psi}^k}$.

A key term in our analysis will be the *residual* defined at a given primal vector (y, v) by

$$r = Dy + Hv. \tag{3.20}$$

The residual term is important since its value at the primal vector (y^{k+1}, v^{k+1}) specifies the update direction for the dual vector μ^{k+1} [cf. Eq. (3.19)]. We will denote the residual at the primal vector (y^{k+1}, v^{k+1}) by

$$r^{k+1} = Dy^{k+1} + Hv^{k+1}. \tag{3.21}$$

3.3.1 Preliminaries

We proceed to the convergence analysis of the asynchronous algorithm. We first provide some preliminary general results on optimality conditions (which enable us to linearize the quadratic term in the primal updates of the ADMM algorithm) and feasibility of the saddle points of the Lagrangian function of problem (3.1) (using Lemma 2.4.1 from previous chapter and Lemma and 3.3.1). We then use these results to provide bounds on the difference of the objective function value of the vector y^k from the optimal value, the distance between μ^k and an optimal dual solution and distance between v^k and an optimal solution z^*, which will be used later to establish convergence properties of asynchronous algorithm (see Theorem 3.3.1). We also provide a set of sufficient conditions for a limit point of the sequence $\{x^k, z^k, p^k\}$ to be a saddle point of the Lagrangian function, which is later used to establish almost sure convergence (see Lemma 3.3.5). The results of this section are independent of the probability distributions of the random variables Φ^k and Ψ^k.

The next lemma establishes primal feasibility (or zero residual property) of a saddle point of the Lagrangian function of problem (3.1).

Lemma 3.3.1. *Let (x^*, z^*, p^*) be a saddle point of the Lagrangian function defined as in Eq. (3.4) of problem (3.1). Then*

$$Dx^* + Hz^* = 0, . \tag{3.22}$$

Proof. We prove by contradiction. From the definition of a saddle point [cf. Eq. (3.5)], we have for any multiplier p in $\mathbb{R}^W$, the following relation holds

$$F(x^*) - p'(Dx^* + Hz^*) \leq F(x^*) - (p^*)'(Dx^* + Hz^*),$$

i.e., $p'(Dx^* + Hz^*) \geq (p^*)'(Dx^* + Hz^*)$ for all p.

Assume for some i, we have $[Dx^* + Hz^*]_i \neq 0$, then by setting

$$\tilde{p}_j = \begin{cases} \frac{(p^*)'(Dx^*+Hz^*)-1}{[Dx^*+Hz^*]_i} & \text{for} \quad j = i, \\ 0 & \text{for} \quad j \neq i, \end{cases}$$

we arrive at a contradiction that $\tilde{p}'(Dx^* + Hz^*) = (p^*)'(Dx^* + Hz^*) - 1 < (p^*)'(Dx^* + Hz^*)$. Hence we conclude that Eq. (3.22) holds. □

The next lemma uses Lemma 2.4.1 to rewrite the optimality conditions for the iterates (x^k, z^k) and (y^k, v^k), which will be used later in Theorem 3.3.1 to establish bounds on the two key quantities: $F(y^{k+1}) - \mu' r^{k+1}$ and $\frac{1}{2\beta}\left|\left|\mu^{k+1} - p^*\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^*)\right|\right|^2$.

Lemma 3.3.2. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). Let $\{y^k, v^k, \mu^k\}$ be the sequence defined in Eqs. (3.16)-(3.19) and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (3.1). The following hold at each iteration k:*

(a) For all $y \in X$, we have

$$f^k(y)+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]' D_{\phi^k} y \geq f^k(x^{k+1})+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]' D_{\phi^k} x^{k+1}. \tag{3.23}$$

(b) For all $v \in Z$, we have

$$-(p^{k+1})' H_{\psi^k} v \geq -(p^{k+1})' H_{\psi^k} z^{k+1}. \tag{3.24}$$

(c) For all $\mu \in \mathbb{R}^W$, we have

$$\begin{aligned} &F(x^*) - F(y^{k+1}) + \mu' r^{k+1} \\ &\quad - [\mu - \mu^{k+1}]'(r^{k+1}) - \beta[H(z^k - v^{k+1})]' r^{k+1} - \beta[H(z^k - v^{k+1})]' H(z^* - v^{k+1}) \geq 0. \end{aligned} \tag{3.25}$$

(d) We have

$$(r^{k+1})' H(z^k - v^{k+1}) \geq 0. \tag{3.26}$$

Proof. (a) By Lemma 2.4.1 and definition of y^{k+1} in Eq. (3.16), it follows that y^{k+1} is the minimizer of the function $F(y) + [-p^k + \beta(Dy^{k+1} + Hz^k)]' Dy$, i.e.,

$$F(y) + [-p^k + \beta(Dy^{k+1} + Hz^k)]' Dy \geq F(y^{k+1}) + [-p^k + \beta(Dy^{k+1} + Hz^k)]' Dy^{k+1},$$

for any y in X.

Recall the definition of μ^{k+1} in Eq. (3.19):

$$\mu^{k+1} = p^k - \beta(Dy^{k+1} + Hv^{k+1}),$$

which implies

$$-p^k + \beta(Dy^{k+1} + Hz^k) = -p^k + \beta(Dy^{k+1} + Hv^{k+1}) + \beta H(z^k - v^{k+1}) = -\mu^{k+1} + \beta H(z^k - v^{k+1}).$$

We now apply the equivalent representation of the term $-p^k + \beta(Dy^{k+1} + Hz^k)$ to the preceding inequality and obtain,

$$F(y) + [-\mu^{k+1} + \beta H(z^k - v^{k+1})]' Dy \geq F(y^{k+1}) + [-\mu^{k+1} + \beta H(z^k - v^{k+1})]' Dy^{k+1}.$$

By setting $y_j = y_j^{k+1}$ for all $j \notin \phi^k$ and canceling the terms that appear on both sides, we have only the components related to f^k and D_{ϕ^k} remain, i.e.,

$$f^k(y) + [-\mu^{k+1} + \beta H(z^k - v^{k+1})]' D_{\phi^k} y \geq f^k(y^{k+1}) + [-\mu^{k+1} + \beta H(z^k - v^{k+1})]' D_{\phi^k} y^{k+1}.$$

The above relation combined with the fact that $[x^{k+1}]_{\phi^k} = [y^{k+1}]_{\phi^k}$ and Eq. (3.17), i.e., $D_{\phi^k} x^{k+1} = D_{\phi^k} y^{k+1}$, proves part (a).

(b) Lemma 2.4.1 implies that the vector v^{k+1} is the minimizer of the function

$$\left[-p^k + \beta(Dy^{k+1} + Hv^{k+1})\right]' Hv = -(\mu^{k+1})' Hv,$$

where we used the definition of μ^{k+1} [cf. Eq. (3.19)]. This yields for all $v \in Z$,

$$-(\mu^{k+1})' Hv \geq -(\mu^{k+1})' Hv^{k+1}. \tag{3.27}$$

Setting $v_j = v_j^{k+1}$ for all $j \notin \psi^k$ and canceling the terms that appear on both sides, we have only the components corresponding to ψ^k remain in the preceding inequality, i.e., $-(\mu^{k+1})' H[v]_{\psi^k} \geq -(\mu^{k+1})' H[v^{k+1}]_{\psi^k}$. By using the diagonal structure of H, and the fact that $[p^{k+1}]_{\psi^k} = [\mu^{k+1}]_{\psi^k}$, $[z^{k+1}]_{\psi^k} = [v^{k+1}]_{\psi^k}$, this shows Eq. (3.24) for all v

in Z.

(c) We apply Eq. (3.21) to rewrite the terms Dy^{k+1} and Dy in Eq. (3.23) and obtain

$$F(y)+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'(r-Hv) \geq F(y^{k+1})+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'(r^{k+1}-Hv^{k+1}).$$

By adding the above two inequalities together and rearranging the terms, we obtain

$$F(y) - F(y^{k+1}) + \beta[H(z^k - v^{k+1}) - \mu^{k+1}]'(r - r^{k+1}) - \beta[H(z^k - v^{k+1})]'H(v - v^{k+1}) \geq 0$$

for all y in X and v in Z. We then set $y = x^*$, $v = z^*$ in the preceding relation and by Lemma 3.3.1, we have $r = Hz^* + Dx^* = 0$. Thus, we have

$$F(x^*) - F(y^{k+1}) + \beta[H(z^k - v^{k+1}) - \mu^{k+1}]'(-r^{k+1}) - \beta[H(z^k - v^{k+1})]'H(z^* - v^{k+1}) \geq 0.$$

By adding and subtracting the term $\mu' r^{k+1}$ from the left-hand side of the above inequality, and regrouping of the terms, we obtain

$$\begin{aligned} F(x^*) - F(y^{k+1}) + \mu' r^{k+1} - [\mu - \mu^{k+1}]'(r^{k+1}) - \beta[H(z^k - v^{k+1})]' r^{k+1} \\ - \beta[H(z^k - v^{k+1})]'H(z^* - v^{k+1}) \geq 0. \end{aligned}$$

(d) We next establish Eq. (3.26) by using the following equality

$$H = \sum_{\psi \in \Pi} H_\psi,$$

where we used the fact that the values of sets ψ form a partition of the constraints. Fix any k. For any set ψ, we denote by τ the largest iteration number where $\psi^\tau = \psi$ and $\tau < k$. Then the dual variable p and primal variable z corresponding to the constraints in set ψ was last updated at time τ, i.e.,

$$p_\psi^{\tau+1} = \ldots = p_\psi^k, \quad \text{and} \quad z_\psi^{\tau+1} = \ldots = z_\psi^k. \tag{3.28}$$

By Eq. (3.24), we have

$$-(p^{\tau+1})' H_\psi v \geq -(p^{\tau+1})' H_\psi z^{\tau+1},$$

for all v in Z, where we used the fact that $\psi^\tau = \psi$. Since multiplication by H_ψ picks components of a vector in set ψ, we can use Eq. (3.28) and replace $p^{\tau+1}$ with p^k, $z^{\tau+1}$ with z^k, set $v = v^{k+1}$ and have

$$-(p^k)' H_\psi (v^{k+1} - z^k) \geq 0.$$

The preceding relation holds for all ψ, thus we can sum over all sets ψ and obtain

$$\sum_{\psi \in \Pi} -(p^k)' H_\psi (v^{k+1} - z^k) = -(p^k)' H (v^{k+1} - z^k) \geq 0,$$

We then substitute $v = z^k$ into Eq. (3.27), which yields

$$-(\mu^{k+1})' H (z^k - v^{k+1}) \geq 0.$$

The above two inequalities together with Eq. (3.19) imply the desired relation

$$\frac{1}{\beta}(p^k - \mu^{k+1})' H (z^k - v^{k+1}) = (r^{k+1})' H (z^k - v^{k+1}) \geq 0.$$

□

The following lemma provides an equivalent representation of some terms in Eq. (3.25), which will be used to derive the bounds on the two key quantities $F(y^{k+1}) - \mu' r^{k+1}$ and $\frac{1}{2\beta} \left|\left|\mu^{k+1} - p^*\right|\right|^2 + \frac{\beta}{2} \left|\left|H(v^{k+1} - z^*)\right|\right|^2$. The proof is based on the definition of residual r^{k+1} and algebraic manipulations, similar to those used in [28] and [34].

Lemma 3.3.3. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). Let $\{y^k, v^k, \mu^k\}$ be the sequence defined in Eqs. (3.16)-(3.19). The following holds at each iteration k:*

$$\begin{aligned} &-(\mu^{k+1}-\mu)'(r^{k+1}) - \beta(r^{k+1})'H(v^{k+1}-z^k) - \beta(v^{k+1}-z^k)'H'H(v-v^{k+1}) \\ &\qquad = \frac{1}{2\beta}\left(\left|\left|\mu^{k+1}-\mu\right|\right|^2 - \left|\left|p^k-\mu\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1}-v)\right|\right|^2 - \left|\left|H(z^k-v)\right|\right|^2\right) \\ &\qquad + \frac{\beta}{2}\left|\left|r^{k+1}-H(v^{k+1}-z^k)\right|\right|^2, \end{aligned} \tag{3.29}$$

for all y in X, v in Z and μ in $\mathbb{R}^W$.

Proof. It is more convenient to multiply both sides of Eq. (3.29) by 2 and prove

$$\begin{aligned} &-2(\mu^{k+1}-\mu)'(r^{k+1}) - 2\beta(r^{k+1})'H(v^{k+1}-z^k) - 2\beta(v^{k+1}-z^k)'H'H(v-v^{k+1}) \\ &= \frac{1}{\beta}\left(\left|\left|\mu^{k+1}-\mu\right|\right|^2 - \left|\left|p^k-\mu\right|\right|^2\right) + \beta\left(\left|\left|H(v^{k+1}-v)\right|\right|^2 - \left|\left|H(z^k-v)\right|\right|^2\right) \\ &\quad + \beta\left|\left|r^{k+1}-H(v^{k+1}-z^k)\right|\right|^2. \end{aligned} \tag{3.30}$$

Our proof will use the following two identities:

$$\mu^{k+1} = p^k - \beta r^{k+1}, \tag{3.31}$$

which follows from Eqs. (3.19) and the definiton of the residual r^{k+1}, and

$$||a+b||^2 = ||a||^2 + ||b||^2 + 2a'b, \tag{3.32}$$

for arbitrary vectors a and b.

We start with the first term $-2(\mu^{k+1}-\mu)'r^{k+1}$ on the left-hand side of Eq. (3.30). By adding and subtracting the term $2(p^k)'r^{k+1}$, we obtain

$$-2(\mu^{k+1}-\mu)'r^{k+1} = -2(\mu^{k+1}-p^k+p^k-\mu)'r^{k+1} = 2\beta\left|\left|r^{k+1}\right|\right|^2 - 2(p^k-\mu)'r^{k+1},$$

where we used Eq. (3.31) to write $(\mu^{k+1} - p^k) r^{k+1} = -\beta \left|\left|r^{k+1}\right|\right|^2$. Using Eq. (3.31) once more, we can write the second term on the right-hand side as

$$-2(p^k - \mu)' r^{k+1} = -\frac{2}{\beta}(p^k - \mu)'(p^k - \mu^{k+1}) = \frac{1}{\beta}\left(\left|\left|\mu - \mu^{k+1}\right|\right|^2 - \left|\left|\mu - p^k\right|\right|^2 - \left|\left|p^k - \mu^{k+1}\right|\right|^2\right),$$

where we applied identity Eq. (3.32) to $\mu - \mu^{k+1} = (\mu - p^k) + (p^k - \mu^{k+1})$. We also observe that Eq. (3.31) also implies

$$\beta \left|\left|r^{k+1}\right|\right|^2 = \frac{1}{\beta}\left|\left|\mu^{k+1} - p^k\right|\right|^2.$$

We combine the above three equalities and obtain

$$-2(\mu^{k+1} - \mu)' r^{k+1} = \beta \left|\left|r^{k+1}\right|\right|^2 + \frac{1}{\beta}\left(\left|\left|\mu - \mu^{k+1}\right|\right|^2 - \left|\left|\mu - p^k\right|\right|^2\right). \tag{3.33}$$

We apply Eq. (3.32) to $r^{k+1} - H(v^{k+1} - z^k) = r^{k+1} + (-H(v^{k+1} - z^k))$ and therefore can represent the second term on the left-hand side of Eq. (3.30), i.e., $-2\beta(r^{k+1})' H(v^{k+1} - z^k)$, as,

$$-2\beta(r^{k+1})' H(v^{k+1} - z^k) = -\beta \left|\left|r^{k+1}\right|\right|^2 - \beta \left|\left|H(v^{k+1} - z^k)\right|\right|^2 + \beta \left|\left|r^{k+1} - H(v^{k+1} - z^k)\right|\right|^2 \tag{3.34}$$

The third term can be expressed similarly. Based on the identity $H(z^k - v) = H(z^k - v^{k+1}) + H(v^{k+1} - v)$ and Eq. (3.32), we obtain

$$\left|\left|H(z^k - v)\right|\right|^2 = \left|\left|H(z^k - v^{k+1})\right|\right|^2 + \left|\left|H(v^{k+1} - v)\right|\right|^2 + 2(z^k - v^{k+1})' H' H(v^{k+1} - v),$$

which implies the third term of the left-hand side of Eq. (3.30), i.e., $-2\beta(v^{k+1} - z^k)' H' H(v - v^{k+1})$, can be written as

$$-2\beta(v^{k+1} - z^k)' H' H(v - v^{k+1}) = \beta \left|\left|H(z^k - v^{k+1})\right|\right|^2 + \beta \left|\left|H(v^{k+1} - v)\right|\right|^2 - \beta \left|\left|H(z^k - v)\right|\right|^2. \tag{3.35}$$

By combing the equivalence relations for all three terms [cf. Eq. s (3.33), (3.34) and (3.35)], we have

$$\begin{aligned}
&-2(\mu^{k+1}-\mu)'(r^{k+1})-2\beta(r^{k+1})'H(v^{k+1}-z^k)-2\beta(v^{k+1}-z^k)'H'H(v-v^{k+1})\\
=&\beta\left(\left|\left|r^{k+1}\right|\right|^2-\left|\left|r^{k+1}\right|\right|^2\right)+\beta\left(\left|\left|H(z^k-v^{k+1})\right|\right|^2-\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right)\\
&+\frac{1}{\beta}\left(\left|\left|\mu-\mu^{k+1}\right|\right|^2-\left|\left|\mu-p^k\right|\right|^2\right)+\beta\left|\left|r^{k+1}-H(v^{k+1}-z^k)\right|\right|^2\\
&+\beta\left|\left|H(v^{k+1}-v)\right|\right|^2-\beta\left|\left|H(z^k-v)\right|\right|^2.
\end{aligned}$$

The terms in the first two parentheses cancel out, establishing the desired result. □

The next theorem combines the preceding results and provides bounds on two key quantities, $F(y^{k+1})-\mu' r^{k+1}$ and $\frac{1}{2\beta}\left|\left|\mu^{k+1}-p^*\right|\right|^2+\frac{\beta}{2}\left|\left|H(v^{k+1}-z^*)\right|\right|^2$. These quantities will be related to the iterates generated by the asynchronous ADMM algorithm via a weighted norm and a weighted Lagrangian function in Section 3.3.2. The weighted version of the quantity $\frac{1}{2\beta}\left|\left|\mu^{k+1}-p^*\right|\right|^2+\frac{\beta}{2}\left|\left|H(v^{k+1}-z^*)\right|\right|^2$ is used to show almost sure convergence of the algorithm and the quantity $F(y^{k+1})-\mu' r^{k+1}$ is used in the convergence rate analysis.

Theorem 3.3.1. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). Let $\{y^k, v^k, \mu^k\}$ be the sequence defined in Eqs. (3.16)-(3.19) and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (3.1). The following hold at each iteration k:*

$$\begin{aligned}
F(x^*)-F(y^{k+1})+\mu' r^{k+1}\geq&\frac{1}{2\beta}\left(\left|\left|\mu^{k+1}-\mu\right|\right|^2-\left|\left|p^k-\mu\right|\right|^2\right) \qquad (3.36)\\
&+\frac{\beta}{2}\left(\left|\left|H(v^{k+1}-z^*)\right|\right|^2-\left|\left|H(z^k-z^*)\right|\right|^2\right)+\frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2+\frac{\beta}{2}\left|\left|H(v^{k+1}-z^k)\right|\right|^2,
\end{aligned}$$

for all μ *in* $\mathbb{R}^W$, *and*

$$0 \geq \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - p^*\right|\right|^2 - \left|\left|p^k - p^*\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right)$$
$$+ \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2. \quad (3.37)$$

Proof. By Lemma 3.3.0(c), we have

$$F(x^*) - F(y^{k+1}) + \mu' r^{k+1}$$
$$- [\mu - \mu^{k+1}]'(r^{k+1}) - \beta[H(z^k - v^{k+1})]' r^{k+1} - \beta[H(z^k - v^{k+1})]' H(z^* - v^{k+1}) \geq 0,$$

Using Eq. (3.29) with $v = z^*$ (cf. Lemma 3.3.3), we can express the last three terms on the left-hand side of this inequality as

$$(\mu - \mu^{k+1})'(r^{k+1}) + \beta(r^{k+1})' H(z^k - v^{k+1}) + \beta(z^k + v^{k+1})' H' H(z^* - v^{k+1})$$
$$= \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - \mu\right|\right|^2 - \left|\left|p^k - \mu\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right)$$
$$+ \frac{\beta}{2}\left|\left|r^{k+1} - H(v^{k+1} - z^k)\right|\right|^2,$$

Lemma 3.3.0(d) implies,

$$\left|\left|r^{k+1} - H(v^{k+1} - z^k)\right|\right|^2 = \left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1} - z^k)\right|\right|^2 + 2(r^{k+1})' H(z^k - v^{k+1})$$
$$\geq \left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1} - z^k)\right|\right|^2.$$

Combining with the preceding relation, we obtain

$$F(x^*) - F(y^{k+1}) + \mu' r^{k+1} \geq \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - \mu\right|\right|^2 - \left|\left|p^k - \mu\right|\right|^2\right)$$
$$+ \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) + \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2$$

showing Eq. (3.36).

To show Eq. (3.37), we let $\mu = p^*$ in the preceding inequality:

$$F(x^*) - F(y^{k+1}) + (p^*)'r^{k+1} \geq \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - p^*\right|\right|^2 - \left|\left|p^k - p^*\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) + \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2.$$

By Assumption 3, we have

$$F(y^{k+1}) - p^* r^{k+1} \geq F(x^*) - p^*(Dx^* + Hz^*) = F(x^*),$$

where the equality follows from result $Dx^* + Hz^* = 0$ in Lemma 3.3.1. We then combine the above two inequalities and obtain

$$0 \geq \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - p^*\right|\right|^2 - \left|\left|p^k - p^*\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) + \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2,$$

which establishes Eq. (3.37). □

The next lemma provides a set of sufficient conditions for a vector to be a saddle point of the Lagrangian function. It will be used to establish Lemma 3.3.5, which analyzes the limiting properties of the sequence $\{x^k, z^k, p^k\}$.

Lemma 3.3.4. *A point (x^*, z^*, p^*) is a saddle point of the Lagrangian function $L(x, z, p)$ of problem (3.1) if it satisfies the following conditions:*

$$[Dx^* + Hz^*]_l = 0 \qquad \text{for all } l = 1, \ldots, W, \tag{3.38}$$

$$f_i(x_i^*) - (p^*)'D_i x^* \leq f_i(x_i) - (p^*)'D_i x \qquad \text{for all } i = 1, \ldots N,\ x \in X, \tag{3.39}$$

$$-(p^*)' \sum_{\psi \in \Pi} H_\psi z^* \leq -(p^*)' \sum_{\psi \in \Pi} H_\psi z \qquad \text{for all } z \in Z. \tag{3.40}$$

Proof. From the definition of a saddle point [cf. Assumption 3], a point (x^*, z^*, p^*) is a saddle point of the Lagrangian function $L(x, z, p) = F(x) - p'(Dx + Hz)$ if

$$L(x^*, z^*, p) \leq L(x^*, z^*, p^*) \leq L(x, z, p^*),$$

for any x in X, z in Z and p in $\mathbb{R}^W$. We derive the sufficient condition (3.38) from the first inequality and (3.39)-(3.40) from the second inequality.

The first inequality $L(x^*, z^*, p) \leq L(x^*, z^*, p^*)$, by using definition of the Lagrangian function can be written as

$$-p'(Dx^* + Hz^*) \leq -(p^*)'(Dx^* + Hz^*).$$

This holds for all p in $\mathbb{R}^W$ if Eq. (3.38), i.e., $[Dx^*+Hz^*]_l = 0$, holds for all $l = 1, \ldots, W$.

The second inequality $L(x^*, z^*, p^*) \leq L(x, z, p^*)$, by the definition of the Lagrangian function can be written as

$$F(x^*) - (p^*)'(Dx^* + Hz^*) \leq F(x) - (p^*)'(Dx^* + Hz^*).$$

We now substitute equalities $F(x) = \sum_{i=1}^N f_i(x_i)$, $D = \sum_{i=1}^N D_i$ and $H = \sum_{\psi \in \Pi} H_\psi$ and obtain

$$\sum_{i=1}^N [f_i(x_i^*) - (p^*)'D_i x^*] - (p^*)' \sum_{\psi \in \Pi} H_\psi z^* \leq \sum_{i=1}^N [f_i(x_i) - (p^*)'D_i x] - (p^*)' \sum_{\psi \in \Pi} H_\psi z.$$

Since the vector $D_i x$ only involves x_i, we have that the above inequality is separable in x_i. The above inequality can further be decomposed into terms related to variable x and terms related to variable z. Hence the above inequality holds for all x in X, z in Z if the following two inequalities are satisfied,

$$f_i(x_i^*) - (p^*)'D_i x^* \leq f_i(x_i) - (p^*)'D_i x,$$

$$-(p^*)' \sum_{\psi \in \Pi} H_\psi z^* \leq -(p^*)' \sum_{\psi \in \Pi} H_\psi z,$$

which are relations (3.39) and (3.40). □

The following lemma analyzes the limiting properties of the sequence $\{x^k, z^k, p^k\}$. The results will later be used in Lemma 3.3.7, which provides a set of sufficient conditions for a limit point of the sequence $\{x^k, z^k, p^k\}$ to be a saddle point. This is later used to establish almost sure convergence.

Lemma 3.3.5. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). Let $\{y^k, v^k, \mu^k\}$ be the sequence defined in Eqs. (3.16)-(3.19). Suppose the sequence $\left\{\left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1} - z^k)\right|\right|^2\right\}$ converges to 0 and the sequence $\{z^k, p^k\}$ is bounded, where each vector r^k is the residual defined as in Eq. (3.21). Then, the sequence $\{x^k, y^k, z^k\}$ has a limit point, which is a saddle point of the Lagrangian function of problem (3.1).*

Proof. We will establish the lemma in three steps:

(a) The scalar sequence $\left\{\left|\left|Dx^{k+1} + Hz^{k+1}\right|\right|\right\}$ converges to 0.

(b) The sequence $\{x^k, z^k, p^k\}$ is bounded, and therefore has a limit point.

(c) A limit point $\{\tilde{x}, \tilde{z}, \tilde{p}\}$ of the sequence $\{x^k, z^k, p^k\}$ is a saddle point of the Lagrangian function of problem (3.1).

We first explicitly write down the components of the vectors $Dx^{k+1} + Hz^{k+1}$ and p^{k+1} in terms of the iterates at time k and vectors $y^{k+1}, v^{k+1}, \mu^{k+1}$ (in particular the residual vector r^{k+1}), for each k and each realization of random variables Φ^k and Ψ^k. The properties of vector $Dx^{k+1} + Hz^{k+1}$ will be used in proving part (a) and those of vector p^{k+1} will be used to establish part (c).

By Assumption 4, each row l of matrix D has exactly one nonzero element. We denote this element by $i(l)$. By the definition of random variable $\phi^k = \Phi(\psi^k)$ in our asynchronous implementation [cf. Section 3.2.2], if constraint l is active, then component $i(l)$ is active, i.e., if $l \in \psi^k$, then $i(l) \in \phi^k$. However, note that even though constraint l is not active, i.e., $l \notin \psi^k$, we could have component $i(l)$ active, i.e., $i(l) \in \phi^k$ (since that component may appear in another row of matrix D, which may become active at iteration k). Thus the index l falls into one of the following three cases and we investigate the change in the components $[Dx^{k+1} + Hz^{k+1}]_l$ and p_l^{k+1} in each scenario.

(I) When $i(l) \notin \phi^k$ and $l \notin \psi^k$, we have the $x_{i(l)}$, z_l and p_l associated with the l^{th} row of the constraint stay at their values form the previous iteration, hence,

$$[Dx^{k+1} + Hz^{k+1}]_l = [Dx^k + Hz^k]_l,$$

and

$$p_l^{k+1} = p_l^k.$$

(II) When $i(l) \in \phi^k$ and $l \in \psi^k$, the variables $x_{i(l)}$, z_l and p_l all update at iteration k, and we have

$$[Dx^{k+1} + Hz^{k+1}]_l = [Dy^{k+1} + Hv^{k+1}]_l = [r^{k+1}]_l,$$

and

$$p_l^{k+1} = \mu_l^{k+1}.$$

(III) When $i(l) \in \phi^k$ and $l \notin \psi^k$, variable $x_{i(l)}$ updates, while variables z_l and p_l remain at the same value as the previous iteration. We have $[Dx^{k+1}]_l = [Dy^{k+1}]_l$ and $[Hz^{k+1}]_l = [Hz^k]_l$, therefore

$$[Dx^{k+1} + Hz^{k+1}]_l = [Dy^{k+1} + H(v^{k+1} - v^{k+1} + z^k)]_l = [r^{k+1}]_l - [H_l(v^{k+1} - z^k)]_l,$$

and

$$p_l^{k+1} = p_l^k = \mu_l^{k+1} + \beta r_l^{k+1},$$

where the last equality follows from Eq. (3.19).

Let $l \in \{1, \dots, W\}$. We now proceed to prove part (a) by showing that $\lim_{k\to\infty}[Dx^{k+1} + Hz^{k+1}]_l = 0$. Since the scalar sequence $\left\{ \left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1} - z^k)\right|\right|^2 \right\}$ converges to 0 and both of the terms are nonnegative, the scalar sequences $\left|\left|H(v^{k+1} - z^k)\right|\right|^2$ and

$\left|\left|r^{k+1}\right|\right|^2$ both converge to 0, which implies that

$$\lim_{k\to\infty} H_l(v^{k+1} - z^k) = 0 \tag{3.41}$$

and

$$\lim_{k\to\infty} [r^{k+1}]_l = 0. \tag{3.42}$$

Let k be any iteration count larger than the first time when all ψ in Π have been active for at least once. Because each set ψ appears infinitely often in the asynchronous algorithm [cf. Assumption 5], such finite index k exists with probability 1. Define $\tau(k,i)$ to be the last time x_i is updated up to and include k, i.e., $\tau(k,i) \leq k$, and $\tau(\tau(k,i),i) = \tau(k,i)$. Recall the three scenarios, we have

$$[Dx^{k+1} + Hz^{k+1}]_l = \begin{cases} [Dx^k + Hz^k]_l & \text{for } l \text{ in case I, and } \tau(k,i) < k, \\ [r^{k+1}]_l & \text{for } l \text{ in case II, and } \tau(k,i) = k, \\ [r^{k+1}]_l - [H_l(v^{k+1} - z^k)]_l & \text{for } l \text{ in case III, and } \tau(k,i) = k. \end{cases} \tag{3.43}$$

In case I, we have

$$[Dx^{k+1} + Hz^{k+1}]_l = \ldots = [Dx^{\tau(k,i)+1} + Hz^{\tau(k,i)+1}]_l.$$

Since $\tau(\tau(k,i),i) = \tau(k,i)$, from Eq. (3.43), we have

$$[Dx^{\tau(k,i)+1} + Hz^{\tau(k,i)+1}]_l = [r^{\tau(k,i)+1}]_l \quad \text{or} \quad [r^{\tau(k,i)+1}]_l - [H_l(v^{\tau(k,i)+1} - z^{\tau(k,i)})]_l.$$

Thus we can conclude that for each k and l

$$[Dx^{k+1} + Hz^{k+1}]_l = [r^{\tau(k,i)+1}]_l \quad \text{or} \quad [r^{\tau(k,i)+1}]_l - [H_l(v^{\tau(k,i)+1} - z^{\tau(k,i)})]_l.$$

By Eqs. (3.41) and (3.42), both of the sequences $[r^{k+1}]_l$ and $[r^{k+1}]_l - [H_l(v^{k+1} - z^k)]_l$ converges to 0 as k grows large, which implies both sequences $[r^{\tau(k,i)+1}]_l$ and $[r^{\tau(k,i)+1}]_l - [H_l(v^{\tau(k,i)+1} - z^{\tau(k,i)})]_l$ converges to 0 as k grows large, since the function

$\tau(k,i)$ is monotone in k and is unbounded. Therefore, we have the sequence $\{[Dx^{k+1} + Hz^{k+1}]_l\}$ is convergent and its limit is given by

$$\lim_{k\to\infty}[Dx^{k+1} + Hz^{k+1}]_l = 0,$$

for all l. This shows statement (a).

We now proceed to statement (b). Since the sequence $\{z^k, p^k\}$ is assumed be to bounded, we only need to establish that the sequence $\{x^k\}$ is bounded to guarantee the sequence $\{x^k, z^k, p^k\}$ is bounded. We note by triangle inequality that

$$\limsup_{k\to\infty}\left|\left|Dx^k\right|\right| \leq \limsup_{k\to\infty}\left|\left|Dx^k + Hz^k\right|\right| + \left|\left|Hz^k\right|\right|.$$

In view of the fact that the sequence $\{z^k\}$ is bounded, the sequence $\{\left|\left|Hz^k\right|\right|\}$ is also bounded. Statement (a) implies sequence $\{\left|\left|Dx^k + Hz^k\right|\right|\}$ is bounded. Hence the right-hand of the above inequality is bounded. Therefore, the scalar sequence $\{\left|\left|Dx^k\right|\right|\}$ is bounded. Since the matrix $D'D$ is positive definite (which follows from Assumption 4), this shows that the sequence $\{x^k\}$ is bounded.

To establish the last statement (c), we analyze the properties of the limit point to establish that it satisfies the sufficient conditions to be a saddle point of the Lagrangian function given in Lemma 3.3.4. We denote by $(\tilde{x}, \tilde{z}, \tilde{p})$ a limit point of the sequence (x^k, z^k, p^k), i.e., the limit of sequence (x^k, z^k, p^k) along a subsequence k in κ. By statement (a) and the definition of a limit point, we have

$$[D\tilde{x} + H\tilde{z}]_l = \lim_{k\in\kappa, k\to\infty}[Dx^k + Hz^k]_l = 0. \tag{3.44}$$

By Eq. (3.24) from Lemma 3.3.0(b), we have for all k and $z \in Z$,

$$-(p^{k+1})' H_{\psi^k} z^{k+1} \leq -(p^{k+1})' H_{\psi^k} z. \tag{3.45}$$

We will first show that

$$-(p^{k+1})' \sum_{\psi \in \Pi} H_\psi z^{k+1} \leq -(p^{k+1})' \sum_{\psi \in \Pi} H_\psi z,$$

for all j. By restricting to the convergent subsequence $\{p^k, z^k\}$ for k in κ and using a continuity argument, this proves condition (3.40). Since we are only interested in the properties of the sequence for k large, without loss of generality we can restrict our attention to sufficiently large indices k, where all constraints have been active for at least once.[8]

Since the partition Π is proper and the matrix H is diagonal, the update of variables z and p are independent of components in $\bar{\psi}^k$, i.e., when $\psi^k \neq \psi$, $[z^{k+1}]_\psi = [z^k]_\psi$ and $[p^{k+1}]_\psi = [p^k]_\psi$. Thus, for every $\psi \neq \psi^k$, we have

$$-(p^{k+1})' H_\psi z^{k+1} = -(p^{\tau+1})' H_\psi z^{\tau+1} \leq -(p^{\tau+1})' H_\psi z = -(p^{k+1})' H_\psi z,$$

where $\tau < k$ is the largest index where $\psi^\tau = \psi$ and the inequality follows from Lemma 3.3.0(b).

By summing the above relation with Eq. (3.45), we have

$$-(p^{k+1})' \sum_{\psi \in \Pi} H_\psi z^{k+1} \leq -(p^{k+1})' \sum_{\psi \in \Pi} H_\psi z,$$

for all z in Z. The preceding relation holds for all k and therefore for k in κ.

Since the subsequence (z^k, p^k) is convergent and multiplication by H_ψ is continuous, we can now take limit in k and obtain

$$\lim_{k \in \kappa, k \to \infty} -(p^k)' \sum_{\psi \in \Pi} H_\psi z^k \leq \lim_{k \in \kappa, k \to \infty} -(p^k)' \sum_{\psi \in \Pi} H_\psi z,$$

i.e.,

$$-\tilde{p}' \sum_{\psi \in \Pi} H_\psi \tilde{z} \leq -\tilde{p}' \sum_{\psi \in \Pi} H_\psi z.$$

[8]Due to Assumption 5, the time index for when all constraints are active for at least once is finite almost surely.

What remains is to show Eq. (3.39). By Lemma 3.3.0(a), we have

$$f^k(y)+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'D_{\phi^k}y \geq f^k(x^{k+1})+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'D_{\phi^k}x^{k+1},$$

for all k. Since the above relation holds for all y in X, we can isolate each x_i by setting $y_j = x_j^{k+1}$ for $j \neq i$ and have for i in ϕ^k

$$f_i(y_i)+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'D_i y \geq f_i(x_i^{k+1})+[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'D_i x^{k+1}.$$

Since agent i is active, the matrix D_i has nonzero elements only in rows corresponding to agent i in ϕ^k, i.e., $[[-\mu^{k+1}+\beta H(z^k-v^{k+1})]'D_i]_l = 0$ for l in case I. If l falls in case II, we have

$$[-\mu^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i = [-p^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i,$$

whereas in case III, we have

$$[-\mu^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i = [-p^{k+1}-\beta r^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i.$$

We can write the above inequality as

$$\begin{aligned} f_i(y_i) &+ \sum_{l\in II}[-p^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i y + \sum_{l\in III}[-p^{k+1}-\beta r^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i y \\ &\geq f_i(x_i^{k+1}) + \sum_{l\in II}[-p^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i x^{k+1} \\ &+ \sum_{l\in III}[-p^{k+1}-\beta r^{k+1}+\beta H(z^k-v^{k+1})]_l'D_i x^{k+1}, \end{aligned} \tag{3.46}$$

where we use the notation $l \in II$ to indicate the set of indices l in case II and similarly for case III. This relation holds for all i in ϕ^k. For i not in ϕ^k, let $\tau(k,i)$ be the last time x_i has been updated up to time k. We have

$$\begin{aligned}
f_i(y_i)+&\sum_{l\in II}[-p^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i y\\
&+\sum_{l\in III}[-p^{\tau(k,i)+1}-\beta r^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i y\\
\geq& f_i(x_i^{\tau(k,i)+1})+\sum_{l\in II}[-p^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i x^{\tau(k,i)+1}\\
&+\sum_{l\in III}[-p^{\tau(k,i)+1}-\beta r^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i x^{\tau(k,i)+1}.
\end{aligned}$$

Since the values of x_i^{k+1} and p^{k+1} related to the constraints involving x_i, i.e., p_l^{k+1} for l in II and III, have the same value as $x_i^{\tau(k,i)+1}$ and $p^{\tau(k,i)+1}$, the above relation can be written as

$$\begin{aligned}
f_i(y_i)+&\sum_{l\in II}[-p^{k+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i y\\
&+\sum_{l\in III}[-p^{k+1}-\beta r^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i y\\
\geq& f_i(x_i^{k+1})+\sum_{l\in II}[-p^{k+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i x^{k+1}\\
&+\sum_{l\in III}[-p^{k+1}-\beta r^{\tau(k,i)+1}+\beta H(z^{\tau(k,i)}-v^{\tau(k,i)+1})]_l' D_i x^{k+1}.
\end{aligned}$$

iteration count where i is last updated up to and including iteration k. This τ index is a function of i and iteration count k. Since each agent is active infinitely often with probability 1 by Assumption 5, we have as k goes to infinity, the index $\tau(k,i)$ also approaches infinity for all i.

We now consider the subsequence $\{x^k, z^k, p^k\}$ for k in κ. By relations (3.42) and (3.41), we have for all i, the terms involving $H(z^{\tau(k,i)}-v^{\tau(k,i)+1})$ and $r^{\tau(k,i)+1}$ diminish as $k\to\infty$ for k in κ. Function f is continuous (since it is convex over $\mathbb{R}^n$) and therefore we can take limit of the preceding relation and Eq. (3.46). Both of them satisfy

$$\lim_{k\in\kappa,k\to\infty} f_i(y_i) + [-p^k]'D_i y \geq \lim_{k\in\kappa,k\to\infty} f_i(x_i^k) + [-p^k]'D_i x^k,$$

where we combined the $[p^t]_l$ term for both $l \in II$ and $l \in III$. Since the subsequence for k in κ converges to the point $(\tilde{x}, \tilde{z}, \tilde{p})$, the limit satisfies

$$f_i(y_i) - [\tilde{p}]'D_i y \geq f_i(\tilde{x}_i) - [\tilde{p}]'D_i\tilde{x},$$

for all y in X and for all i. Thus condition (3.39) is satisfied and by Lemma 3.3.4 the limit point $(\tilde{x}, \tilde{z}, \tilde{p})$ is a saddle point of the Lagrangian function, which establishes the desired result. □

3.3.2 Convergence and Rate of Convergence

The results of the previous section did not rely on the probability distributions of random variables Φ^k and Ψ^k. In this section, we will introduce a weighted norm and weighted Lagrangian function where the weights are defined in terms of the probability distributions of random variables Ψ^k and Φ^k representing the active constraints and components. We will use the weighted norm to construct a nonnegative supermartingale along the sequence $\{x^k, z^k, p^k\}$ generated by the asynchronous ADMM algorithm and use it to establish the almost sure convergence of this sequence to a saddle point of the Lagrangian function of problem (3.1). By relating the iterates generated by the asynchronous ADMM algorithm to the variables (y^k, v^k, μ^k) through taking expectations of the weighted Lagrangian function and using results from Theorem 3.3.1(which provides a bound on the difference of the objective function value of the vector y^k from the optimal value), we will show that under a compactness assumption on the constraint sets X and Z, the asynchronous ADMM algorithm converges with rate $O(1/k)$ in expectation in terms of both objective function value and constraint violation.

We use the notation α_i to denote the probability that component x_i is active at one iteration, i.e.,

$$\alpha_i = \mathbb{P}(i \in \Phi^k), \tag{3.47}$$

and the notation λ_l to denote the probability that constraint l is active at one iteration, i.e.,

$$\lambda_l = \mathbb{P}(l \in \Psi^k). \tag{3.48}$$

Note that, since the random variables Φ^k (and Ψ^k) are independent and identically distributed for all k, these probabilities are the same across all iterations. We define a diagonal matrix Λ in $\mathbb{R}^{W\times W}$ with elements λ_l on the diagonal, i.e.,

$$\Lambda_{ll} = \lambda_l \qquad \text{for each } l \in \{1, \dots, W\}.$$

Since each constraint is assumed to be active with strictly positive probability [cf. Assumption 5], matrix Λ is positive definite. We write $\bar{\Lambda}$ to indicate the inverse of matrix Λ. Matrix $\bar{\Lambda}$ induces a *weighted vector norm* for p in $\mathbb{R}^W$ as

$$||p||_{\bar{\Lambda}}^2 = p'\bar{\Lambda}p.$$

We define a *weighted Lagrangian function* $\tilde{L}(x, z, \mu) : \mathbb{R}^{nN} \times \mathbb{R}^W \times \mathbb{R}^W \to \mathbb{R}$ as

$$\tilde{L}(x, z, \mu) = \sum_{i=1}^{N} \frac{1}{\alpha_i} f_i(x_i) - \mu' \left(\sum_{i=1}^{N} \frac{1}{\alpha_i} D_i x + \sum_{l=1} \frac{1}{\lambda_l} H_l z \right). \tag{3.49}$$

We use the symbol $\mathcal{J}_k$ to denote the filtration up to and include iteration k, which contains information of random variables Φ^t and Ψ^t for $t \leq k$. We have $\mathcal{J}_k \subset \mathcal{J}_{k+1}$ for all $k \geq 1$.

The particular weights in $\bar{\Lambda}$-norm and the weighted Lagrangian function are chosen to relate the expectation of the norm $\frac{1}{2\beta} \left|\left|p^{k+1} - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2} \left|\left|H(z^{k+1} - v)\right|\right|_{\bar{\Lambda}}^2$ and function $\tilde{L}(x^{k+1}, z^{k+1}, \mu)$ to $\frac{1}{2\beta} \left|\left|p^k - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2} \left|\left|H(z^k - v)\right|\right|_{\bar{\Lambda}}^2$ and function $\tilde{L}(x^k, z^k, \mu)$, as we will show in the following lemma. This relation will be used in Theorem 3.3.2 to show that the scalar sequence $\left\{\frac{1}{2\beta} \left|\left|p^k - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2} \left|\left|H(z^k - v)\right|\right|_{\bar{\Lambda}}^2\right\}$ is a nonnegative supermartingale, and establish almost sure convergence of the asynchronous ADMM algorithm.

Lemma 3.3.6. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). Let $\{y^k, v^k, \mu^k\}$ be the sequence defined in Eqs. (3.16)-(3.19). Then the following hold for each iteration k:*

$$\begin{aligned}\mathbb{E}\left(\frac{1}{2\beta}\left|\left|p^{k+1}-\mu\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-v)\right|\right|_{\bar{\Lambda}}^2\middle|\mathcal{J}_k\right)&=\frac{1}{2\beta}\left|\left|\mu^{k+1}-\mu\right|\right|^2\\&+\frac{\beta}{2}\left|\left|H(v^{k+1}-v)\right|\right|^2+\frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|_{\bar{\Lambda}}^2\\&-\frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|^2-\frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|^2.\end{aligned}\tag{3.50}$$

for all μ in $\mathbb{R}^W$ and v in Z, and

$$\begin{aligned}&\mathbb{E}\left(\tilde{L}(x^{k+1},z^{k+1},\mu)\middle|\mathcal{J}_k\right)\\&=\left(F(y^{k+1})-\mu'(Dy^{k+1}+Hv^{k+1})\right)+\tilde{L}(x^k,z^k,\mu)-\left(F(x^k)-\mu'(Dx^k+Hz^k)\right),\end{aligned}\tag{3.51}$$

for all μ in $\mathbb{R}^W$.

Proof. By the definition of λ_l in Eq. (3.48, for each l, the element p_l^{k+1} can be either updated to μ_l^{k+1} with probability λ_l, or stay at previous value p_l^k with probability $1-\lambda_l$. Hence,we have the following expected value,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{2\beta}\left|\left|p^{k+1}-\mu\right|\right|_{\bar{\Lambda}}^2\middle|\mathcal{J}_k\right)=&\sum_{l=1}^{W}\frac{1}{\lambda_l}\left[\lambda_l\left(\frac{1}{2\beta}\left|\left|\mu_l^{k+1}-\mu_l\right|\right|^2\right)+(1-\lambda_l)\left(\frac{1}{2\beta}\left|\left|p_l^k-\mu_l\right|\right|^2\right)\right]\\=&\frac{1}{2\beta}\left|\left|\mu^{k+1}-\mu\right|\right|^2+\frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|_{\bar{\Lambda}}^2-\frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|^2,\end{aligned}$$

where the second equality follows from definition of $||\cdot||_{\bar{\Lambda}}$, and grouping the terms.

Similarly, z_l^{k+1} is either equal to v_l^{k+1} with probability λ_l or z_l^k with probability $1-\lambda_l$. Due to the diagonal structure of the H matrix, the vector $H_l z$ has only one non-zero element equal to $[Hz]_l$ at l^{th} position and zeros else where. Thus, we obtain

$$\mathbb{E}\left(\frac{\beta}{2}\left|\left|H(z^{k+1}-v)\right|\right|_{\bar{\Lambda}}^2 \middle| \mathcal{J}_k\right) = \sum_{l=1}^{W}\frac{1}{\lambda_l}\left[\lambda_l\left(\frac{\beta}{2}\left|\left|H_l(v^{k+1}-v)\right|\right|\right)+(1-\lambda_l)\left(\frac{\beta}{2}\left|\left|H_l(z^k-v)\right|\right|^2\right)\right]$$
$$= \frac{\beta}{2}\left|\left|H(v^{k+1}-v)\right|\right|^2 + \frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|_{\bar{\Lambda}}^2 - \frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|^2,$$

where we used the definition of $||\cdot||_{\bar{\Lambda}}$ once again. By summing the above two equations and using linearity of expectation operator, we obtain Eq. (3.50).

Using a similar line of argument, we observe that at iteration k, for each i, x_i^{k+1} has the value of y_i^{k+1} with probability α_i and its previous value x_l^k with probability $1-\alpha_i$. The expectation of function $\tilde{L}$ therefore satisfies

$$\mathbb{E}\left(\tilde{L}(x^{k+1}, z^{k+1}, \mu)\middle|\mathcal{J}_k\right) = \sum_{i=1}^{N}\frac{1}{\alpha_i}\left[\alpha_i\left(f_i(y_i^{k+1}) - \mu' D_i y^{k+1}\right)\right.$$
$$\left.+(1-\alpha_i)\left(f_i(x_i^k) - \mu' D_i x^k\right)\right] + \sum_{l=1}^{W}\frac{1}{\lambda_l}\mu'\left[\lambda_l H_l v^{k+1} + (1-\lambda_l)H_l z^k\right]$$
$$= \left(\sum_{i=1}^{N} f_i(y_i^{k+1}) - \mu'(Dy^{k+1} + Hv^{k+1})\right) + \tilde{L}(x^k, z^k, \mu)$$
$$- \left(\sum_{i=1}^{N} f_i(x_i^k) - \mu'(Dx^k + Hz^k)\right),$$

where we used the fact that $D = \sum_{i=1}^N D_i$. Using the definition $F(x) = \sum_{i=1}^N f_i(x)$ [cf. Eq. (3.3)], this shows Eq. (3.51).

□

The next lemma builds on Lemma 3.3.5 and establishes a sufficient condition for the sequence $\{x^k, z^k, p^k\}$ to converge to a saddle point of the Lagrangian. Theorem 3.3.2 will then show that this sufficient condition holds with probability 1 and thus the algorithm converges almost surely.

Lemma 3.3.7. *Let (x^*, z^*, p^*) be any saddle point of the Lagrangian function of problem (3.1) and $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM* *algorithm (3.6)-(3.8). Along any sample path of Φ^k and Ψ^k, if the scalar sequence*

$$\frac{1}{2\beta}\left|\left|p^{k+1}-p^*\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-z^*)\right|\right|_{\bar{\Lambda}}^2$$

is convergent and the scalar sequence $\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]$ converges to 0, then the sequence $\{x^k, z^k, p^k\}$ converges to a saddle point of the Lagrangian function of problem (3.1).

Proof. Since the scalar sequence $\frac{1}{2\beta}\left|\left|p^{k+1}-p^*\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-z^*)\right|\right|_{\bar{\Lambda}}^2$ converges, matrix $\bar{\Lambda}$ is positive definite, and matrix H is invertible [cf. Assumption 4], it follows that the sequences $\{p^k\}$ and $\{z^k\}$ are bounded. Lemma 3.3.5 then implies that the sequence $\{x^k, z^k, p^k\}$ has a limit point.

We next show that the sequence $\{x^k, z^k, p^k\}$ has a unique limit point. Let $(\tilde{x}, \tilde{z}, \tilde{p})$ be a limit point of the sequence $\{x^k, z^k, p^k\}$, i.e., the limit of sequence $\{x^k, z^k, p^k\}$ along a subsequence κ. We first show that the components $\tilde{z}, \tilde{p}$ are uniquely defined. By Lemma 3.3.5, the point $(\tilde{x}, \tilde{z}, \tilde{p})$ is a saddle point of the Lagrangian function. Using the assumption of the lemma for $(p^*, z^*) = (\tilde{p}, \tilde{z})$, this shows that the scalar sequence $\left\{\frac{1}{2\beta}\left|\left|p^{k+1}-\tilde{p}\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-\tilde{z})\right|\right|_{\bar{\Lambda}}^2\right\}$ is convergent. The limit of the sequence, therefore, is the same as the limit along any subsequence, implying

$$\begin{aligned}\lim_{k\to\infty}\frac{1}{2\beta}\left|\left|p^{k+1}-\tilde{p}\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-\tilde{z})\right|\right|_{\bar{\Lambda}}^2 \\ &= \lim_{k\to\infty, k\in\kappa}\frac{1}{2\beta}\left|\left|p^{k+1}-\tilde{p}\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^{k+1}-\tilde{z})\right|\right|_{\bar{\Lambda}}^2 \\ &= \frac{1}{2\beta}\left|\left|\tilde{p}-\tilde{p}\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(\tilde{z}-\tilde{z})\right|\right|_{\bar{\Lambda}}^2=0,\end{aligned}$$

Since matrix $\tilde{\Lambda}$ is positive definite and matrix H is invertible, this shows that $\lim_{k\to\infty} p^k = \tilde{p}$ and $\lim_{k\to\infty} z^k = \tilde{z}$.

Next we prove that given $(\tilde{z}, \tilde{p})$, the x component of the saddle point is uniquely determined. By Lemma 3.3.1, we have $D\tilde{x}+H\tilde{z}=0$. Since matrix D has full column rank [cf. Assumption 4], the vector $\tilde{x}$ is uniquely determined by

$$\tilde{x} = -(D'D)^{-1}D'H\tilde{z}.$$

□

The next theorem establishes almost sure convergence of the asynchronous ADMM algorithm. Our analysis uses results related to supermartingales (interested readers are referred to [79] and [80] for a comprehensive treatment of the subject).

Theorem 3.3.2. *Let $\{x^k, z^k, p^k\}$ be the sequence generated by the asynchronous ADMM algorithm (3.6)-(3.8). The sequence (x^k, z^k, p^k) converges almost surely to a saddle point of the Lagrangian function of problem (3.1).*

Proof. We will show that the conditions of Lemma 3.3.7 are satisfied almost surely. We will first focus on the scalar sequence $\frac{1}{2\beta}\left|\left|p^{k+1}-\mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1}-v)\right|\right|_{\bar{\Lambda}}^2$ and show that it is a nonnegative supermartingale. By martingale convergence theorem, this shows that it converges almost surely. We next establish that the scalar sequence $\frac{\beta}{2}\left|\left|r^{k+1}-H(v^{k+1}-z^k)\right|\right|^2$ converges to 0 almost surely by an argument similar to the one used to establish Borel-Cantelli lemma. These two results imply that the set of events where $\frac{1}{2\beta}\left|\left|p^{k+1}-p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1}-z^*)\right|\right|_{\bar{\Lambda}}^2$ is convergent and $\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]$ converges to 0 has probability 1. Hence, by Lemma 3.3.7, we have the sequence $\{x^k, z^k, p^k\}$ converges to a saddle point of the Lagrangian function almost surely.

We first show that the scalar sequence $\frac{1}{2\beta}\left|\left|p^{k+1}-\mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1}-v)\right|\right|_{\bar{\Lambda}}^2$ is a nonnegative supermartingale. Since it is a summation of two norms, it immediately follows that it is nonnegative. To see it is a supermartingale, we let vectors $y^{k+1}, v^{k+1}, \mu^{k+1}$ and r^{k+1} be those defined in Eqs. (3.16)-(3.21). Recall that the symbol $\mathcal{J}_k$ denotes the filtration up to and including iteration k. From Lemma 3.3.6, we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{2\beta}\left|\left|p^{k+1}-\mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1}-v)\right|\right|_{\bar{\Lambda}}^2 \middle| \mathcal{J}_k\right) \\
&\quad = \frac{1}{2\beta}\left|\left|\mu^{k+1}-\mu\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1}-v)\right|\right|^2 + \frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|_{\bar{\Lambda}}^2 \\
&\qquad - \frac{1}{2\beta}\left|\left|p^k-\mu\right|\right|^2 - \frac{\beta}{2}\left|\left|H(z^k-v)\right|\right|^2
\end{aligned}$$

Substituting $\mu = p^*$ and $v = z^*$ in the above expectation calculation and combining with the following inequality from Theorem 3.3.1,

$$\begin{aligned}0 \geq &\frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - p^*\right|\right|^2 - \left|\left|p^k - p^*\right|\right|^2\right) + \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) \\ &+ \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2.\end{aligned}$$

This yields

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{2\beta}\left|\left|p^{k+1} - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1} - z^*)\right|\right|_{\bar{\Lambda}}^2 \middle| \mathcal{J}_k\right) \\ &\leq \frac{1}{2\beta}\left|\left|p^k - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^k - z^*)\right|\right|_{\bar{\Lambda}}^2 - \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 - \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2.\end{aligned}$$

Hence, the sequence $\frac{1}{2\beta}\left|\left|p^{k+1} - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{k+1} - z^*)\right|\right|_{\bar{\Lambda}}^2$ is a nonnegative supermartingale in k and by martingale convergence theorem, it converges almost surely.

We next establish that the scalar sequence $\left\{\frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2\right\}$ converges to 0 almost surely. Rearranging the terms in the previous inequality and taking iterated expectation with respect to the filtration $\mathcal{J}_k$, we obtain for all T

$$\begin{aligned}&\sum_{k=1}^{T}\mathbb{E}\left(\frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 + \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2\right) \\ &\leq \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2 \\ &\quad - \mathbb{E}\left(\frac{1}{2\beta}\left|\left|p^{T+1} - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{T+1} - z^*)\right|\right|_{\bar{\Lambda}}^2\right) \\ &\leq \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2,\end{aligned} \tag{3.52}$$

where the last inequality follows from relaxing the upper bound by dropping the non-positive expected value term. Thus, the sequence

$$\left\{\mathbb{E}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2 + \left|\left|H(v^{k+1} - z^k)\right|\right|^2\right]\right)\right\}$$

is summable implying

$$\lim_{k\to\infty}\sum_{t=k}^{\infty}\mathbb{E}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\right)=0 \tag{3.53}$$

By Markov inequality, we have

$$\mathbb{P}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\geq\epsilon\right)\leq\frac{1}{\epsilon}\mathbb{E}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\right),$$

for any scalar $\epsilon > 0$ for all iterations t. Therefore, we have

$$\begin{aligned}
&\lim_{k\to\infty}\mathbb{P}\left(\sup_{t\geq k}\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\geq\epsilon\right)\\
&=\lim_{k\to\infty}\mathbb{P}\left(\bigcup_{t=k}^{\infty}\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\geq\epsilon\right)\\
&\leq\lim_{k\to\infty}\sum_{t=k}^{\infty}\mathbb{P}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\geq\epsilon\right)\\
&\leq\lim_{k\to\infty}\frac{1}{\epsilon}\sum_{t=k}^{\infty}\mathbb{E}\left(\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\right)=0,
\end{aligned}$$

where the first inequality follows from union bound on probability, the second inequality follows from the preceding relation, and the last equality follows from Eq. (3.53). This proves that the sequence $\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]$ converges to 0 almost surely.

and the scalar sequence $\{\frac{\beta}{2}\left[\left|\left|r^{k+1}\right|\right|^2+\left|\left|H(v^{k+1}-z^k)\right|\right|^2\right]\}$ converges to 0, we have the sequence $\{x^k, z^k, p^k\}$ converges to a saddle point of the Lagrangian. □

We next analyze convergence rate of the asynchronous ADMM algorithm. The rate analysis is done with respect to the time ergodic averages defined as $\bar{x}(T)$ in $\mathbb{R}^{nN}$, the time average of x^k up to and including iteration T, i.e.,

$$\bar{x}_i(T)=\frac{\sum_{1=1}^{T}x_i^k}{T}, \tag{3.54}$$

for all $i = 1, \ldots, N$,[9] and $\bar{z}(k)$ in $\mathbb{R}^W$ as

$$\bar{z}_l(T) = \frac{\sum_{k=1}^{T} z_l^k}{T}, \tag{3.55}$$

for all $l = 1, \ldots, W$.

We next introduce some scalars $Q(\mu)$, $\bar{Q}$, $\bar{\theta}$ and $\tilde{L}^0$, all of which will be used as an upper bound of the constant term which appears in the $O(1/k)$ rate analysis in the following theorem. Scalar $Q(\mu)$ is defined by

$$Q(\mu) = \max_{k, \{\Psi^k, \Phi^k\}} -\tilde{L}(x^{k+1}, z^{k+1}, \mu). \tag{3.56}$$

For the rest of the section, we adopt the following assumption, which will be used to guarantee that scalar $Q(\mu)$ is well defined and finite:

Assumption 6. *The sets X and Z are both compact.*

Since the weighted Lagrangian function $\tilde{L}$ is continuous in x and z [cf. Eq. (3.49)], and all iterates (x^k, z^k) are in the compact set $X \times Z$, by Weierstrass theorem the maximization in the preceding equality is attained and finite.

Since function $\tilde{L}$ is linear in μ, the function $Q(\mu)$ is the maximum of linear functions and is thus convex and continuous in μ. We define scalar $\bar{Q}$ as

$$\bar{Q} = \max_{\mu = p^* - \alpha, ||\alpha|| \leq 1} Q(\mu). \tag{3.57}$$

The reason that such scalar $\bar{Q} < \infty$ exists is once again by Weierstrass theorem (maximization over a compact set).

We define vector $\bar{\theta}$ in $\mathbb{R}^W$ as

$$\bar{\theta} = p^* - \operatorname*{argmax}_{||u|| \leq 1} \left|\left|p^0 - (p^* - u)\right|\right|_{\bar{\Lambda}}^2, \tag{3.58}$$

such maximizer exists due to Weierstrass theorem and the fact that the set $||u|| \leq 1$

[9]Here the notation $\bar{x}_i(T)$ denotes the vector of length n corresponding to agent i.

is compact and the function $||p^0 - (p^* - u)||_{\bar{\Lambda}}^2$ is continuous. Scalar $\tilde{L}^0$ is defined by

$$\tilde{L}^0 = \max_{\theta = p^* - \alpha, ||\alpha|| \leq 1} \tilde{L}(x^0, z^0, \theta). \tag{3.59}$$

This scalar is well defined because the constraint set is compact and the function $\tilde{L}$ is continuous in θ.

Theorem 3.3.3. *Let $\{x^k, z^k, p^k\}$ be the iterates generated by the asynchronous ADMM algorithm 3.6)-(3.8) and (x^*, z^*, p^*) be a saddle point of the Lagrangian function of problem (3.1). Let the vectors $\bar{x}(T)$, $\bar{z}(T)$ be defined as in Eqs. (3.54) and (3.55), the scalars $\bar{Q}$, $\bar{\theta}$ and $\tilde{L}^0$ be defined as in Eqs. (3.57), (3.58) and (3.59) and the function $\tilde{L}$ be defined as in Eq. (3.49). Then the following relations hold:*

$$||\mathbb{E}(D\bar{x}(T) + H\bar{z}(T))|| \leq \frac{1}{T}\left[\bar{Q} + \tilde{L}^0 + \frac{1}{2\beta}\left|\left|p^0 - \bar{\theta}\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right], \tag{3.60}$$

and

$$\begin{aligned}||\mathbb{E}(F(\bar{x}(T))) - F(x^*)|| \leq \frac{||p^*||_\infty}{T}&\left[\bar{Q} + \tilde{L}^0 + \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right] \\ &+\frac{1}{T}\left[Q(p^*) + \tilde{L}(x^0, z^0, p^*) + \frac{1}{2\beta}\left|\left|p^0 - \bar{\theta}\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right].\end{aligned} \tag{3.61}$$

Proof. The proof of the theorem relies on Lemma 3.3.6 and Theorem 3.3.1. We combine these results with law of iterated expectation, telescoping cancellation and convexity of the function F to establish a bound on the value $\mathbb{E}(F(x^k) - \mu'(Dx^k + Hz^k))$ given by

$$\begin{aligned}\mathbb{E}\left[F(\bar{x}(T))\right. &\left.-\mu'(D\bar{x}(T) + H\bar{z}(T))\right] - F(x^*) \\ &\leq \frac{1}{T}\left[Q(\mu) + \tilde{L}(x^0, z^0, \mu) + \frac{1}{2\beta}\left|\left|p^0 - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right],\end{aligned} \tag{3.62}$$

for all μ in $\mathbb{R}^W$. Then by using different choices of the vector μ we can establish the desired results.

We will first prove Eq. (3.62). Recall Eq. (3.51)

$$\mathbb{E}\left(\tilde{L}(x^{k+1}, z^{k+1}, \mu)\middle|\mathcal{J}_k\right)$$
$$= \left(F(y^{k+1}) - \mu'(Dy^{k+1} + Hv^{k+1})\right) + \tilde{L}(x^k, z^k, \mu) - \left(F(x^k) - \mu'(Dx^k + Hz^k)\right),$$

We rearrange Eq. (3.36) from Theorem 3.3.1, and obtain

$$F(y^{k+1}) - \mu' r^{k+1} \leq F(x^*) - \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - \mu\right|\right|^2 - \left|\left|p^k - \mu\right|\right|^2\right)$$
$$- \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) - \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 - \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2.$$

Since $r^{k+1} = Dy^{k+1} + Hv^{k+1}$, we can apply this bound on the first term on the right-hand side of the preceding expectation calculation and obtain,

$$\mathbb{E} = \left(\tilde{L}(x^{k+1}, z^{k+1}, \mu)\middle|\mathcal{J}_k\right) \leq F(x^*) - \frac{1}{2\beta}\left(\left|\left|\mu^{k+1} - \mu\right|\right|^2 - \left|\left|p^k - \mu\right|\right|^2\right)$$
$$- \frac{\beta}{2}\left(\left|\left|H(v^{k+1} - z^*)\right|\right|^2 - \left|\left|H(z^k - z^*)\right|\right|^2\right) - \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 - \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2$$
$$+ \tilde{L}(x^k, z^k, \mu) - \left(F(x^k) - \mu'(Dx^k + Hz^k)\right),$$

We combine the above inequality with Eq. (3.50) and by linearity of expected value, have

$$\mathbb{E}\left(\tilde{L}(x^{k+1}, z^{k+1}, \mu) + \frac{1}{2\beta}\left|\left|p^{k+1} - \mu\right|\right|^2_{\bar{\Lambda}} + \frac{\beta}{2}\left|\left|H(z^{k+1} - z^*)\right|\right|^2_{\bar{\Lambda}}\middle|\mathcal{J}_k\right)$$
$$\leq F(x^*) - \frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 - \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2 - \left(F(x^k) - \mu'(Dx^k + Hz^k)\right)$$
$$+ \tilde{L}(x^k, z^k, \mu) + \frac{1}{2\beta}\left|\left|p^k - \mu\right|\right|^2_{\bar{\Lambda}} + \frac{\beta}{2}\left|\left|H(z^k - z^*)\right|\right|^2_{\bar{\Lambda}}$$
$$\leq F(x^*) - \left(F(x^k) - \mu'(Dx^k + Hz^k)\right) + \tilde{L}(x^k, z^k, \mu) + \frac{1}{2\beta}\left|\left|p^k - \mu\right|\right|^2_{\bar{\Lambda}}$$
$$+ \frac{\beta}{2}\left|\left|H(z^k - z^*)\right|\right|^2_{\bar{\Lambda}},$$

where the last inequality follows from relaxing the upper bound by dropping the non-positive term $-\frac{\beta}{2}\left|\left|r^{k+1}\right|\right|^2 - \frac{\beta}{2}\left|\left|H(v^{k+1} - z^k)\right|\right|^2$.

This relation holds for $k = 1, \dots, T$ and by the law of iterated expectation, the telescoping sum after term cancellation satisfies,

$$\begin{aligned}\mathbb{E}\left(\tilde{L}(x^{T+1}, z^{T+1}, \mu) + \frac{1}{2\beta}\left|\left|p^{T+1} - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{T+1} - z^*)\right|\right|_{\bar{\Lambda}}^2\right) \leq TF(x^*) \quad (3.63)\\ - \mathbb{E}\left[\sum_{k=1}^{T}\left(F(x^k) - \mu'(Dx^k + Hz^k)\right)\right] + \tilde{L}(x^0, z^0, \mu) + \frac{1}{2\beta}\left|\left|p^0 - \mu\right|\right|_{\bar{\Lambda}}^2\\ + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2.\end{aligned}$$

By convexity of the functions f_i, we have

$$\sum_{k=1}^{T} f_i(x_i^k) \geq T f_i(\bar{x}_i(T)) = TF(\bar{x}(T)).$$

The same results hold after taking expectation on both sides. By linearity of matrix-vector multiplication, we have

$$\sum_{k=1}^{T} Dx^k = TD\bar{x}(T), \quad \sum_{k=1}^{T} Hz^k = TH\bar{z}(T).$$

Relation (3.63) therefore implies that

$$\begin{aligned}&T\mathbb{E}\left[F(\bar{x}(T)) - \mu'(D\bar{x}(T) + H\bar{z}(T))\right] - TF(x^*)\\ &\leq \mathbb{E}\left[\sum_{k=1}^{T}\left(F(x^k) - \mu'(Dx^k + Hz^k)\right)\right] - TF(x^*)\\ &\leq - \mathbb{E}\left(\tilde{L}(x^{T+1}, z^{T+1}, \mu) + \frac{1}{2\beta}\left|\left|p^{T+1} - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^{T+1} - z^*)\right|\right|_{\bar{\Lambda}}^2\right)\\ &\quad + \tilde{L}(x^0, z^0, \mu) + \frac{1}{2\beta}\left|\left|p^0 - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2.\end{aligned}$$

Using the definition of scalar $Q(\mu)$ [cf. Eq. (3.56)] and by dropping the non-positive norm terms from the above upper bound, we obtain

$$\begin{aligned}
&T\mathbb{E}\left[F(\bar{x}(T)) - \mu'(D\bar{x}(T) + H\bar{z}(T))\right] - TF(x^*) \\
&\quad \leq Q(\mu) + \tilde{L}(x^0, z^0, \mu) + \frac{1}{2\beta}\left|\left|p^0 - \mu\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2.
\end{aligned}$$

We now divide both sides of the preceding inequality by T and obtain Eq. (3.62).

We now use Eq. (3.62) to first show that $||\mathbb{E}(D\bar{x}(T) + H\bar{z}(T))||$ converges to 0 with rate $1/k$. For each iteration T, we define a vector $\theta(T)$ as $\theta(T) = p^* - \frac{\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))}{||\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))||}$. By substituting $\mu = \theta(T)$ in Eq. (3.62), we obtain for each T,

$$\begin{aligned}
&\mathbb{E}\left[F(\bar{x}(T)) - (\theta(T))'(D\bar{x}(T) + H\bar{z}(T))\right] - F(x^*) \\
&\quad \leq \frac{1}{T}\left[Q(\theta(T)) + \tilde{L}(x^0, z^0, \theta(T)) + \frac{1}{2\beta}\left|\left|p^0 - \theta(T)\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right],
\end{aligned}$$

Since the vectors $\frac{\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))}{||\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))||}$ all have norm 1 and hence $\theta(T)$ are bounded within the unit sphere, by using the definition of $\bar{\theta}$, we have $||p^0 - \theta(T)||_{\bar{\Lambda}}^2 \leq \left|\left|p^0 - \bar{\theta}\right|\right|_{\bar{\Lambda}}^2$. Eqs. (3.57) and (3.59) implies $Q(\theta(T)) \leq \bar{Q}$ and $\tilde{L}(x^0, z^0, \theta(T)) \leq \tilde{L}^0$ for all T. Thus the above inequality suggests that the following holds true for all T,

$$\begin{aligned}
&\mathbb{E}(F(\bar{x}(T)) - (\theta(T))'\mathbb{E}(D\bar{x}(T) + H\bar{z}(T)) - F(x^*) \\
&\quad \leq \frac{1}{T}\left[\bar{Q} + \tilde{L}^0 + \frac{1}{2\beta}\left|\left|p^0 - \bar{\theta}\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right].
\end{aligned}$$

From the definition of $\theta(T)$, we have $(\theta(T))'\mathbb{E}(D\bar{x}(T)+H\bar{z}(T)) = (p^*)'\mathbb{E}(D\bar{x}(T)+H\bar{z}(T)) - ||\mathbb{E}(D\bar{x}(T) + H\bar{z}(T))||$, and thus

$$\begin{aligned}
\mathbb{E}(F(\bar{x}(T)) - &(\theta(T))'\mathbb{E}(D\bar{x}(T) + H\bar{z}(T)) - F(x^*) \\
&= \mathbb{E}(F(\bar{x}(T))) - (p^*)'\mathbb{E}\left[(D\bar{x}(T) + H\bar{z}(T))\right] \\
&\quad - F(x^*) + ||\mathbb{E}(D\bar{x}(T) + H\bar{z}(T))||.
\end{aligned}$$

Since the point (x^*, z^*, p^*) is a saddle point of the Lagrangian function and by

Lemma 3.3.1, we have,

$$\mathbb{E}L((\bar{x}(T)), \bar{z}(T), p^*) - L(x^*, z^*, p^*) \tag{3.64}$$
$$= \mathbb{E}(F(\bar{x}(T))) - F(x^*) - (p^*)'\mathbb{E}\left[(D\bar{x}(T) + H\bar{z}(T))\right] \geq 0.$$

The preceding three relations imply that

$$||\mathbb{E}(D\bar{x}(T) + H\bar{z}(T))|| \leq \frac{1}{T}\left[\bar{Q} + \tilde{L}^0 + \frac{1}{2\beta}\left|\left|p^0 - \bar{\theta}\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right],$$

which shows the first desired inequality. for all μ, we now set $\mu = p^*$, and by saddle point property, we have $p^*(Dx^* + Hz^*) = 0$ in the last equality.

To prove Eq. (3.61), we let $\mu = p^*$ in Eq. (3.62) and obtain

$$\mathbb{E}(F(\bar{x}(T)) - (p^*)'\mathbb{E}(D\bar{x}(T) + H\bar{z}(T)) - F(x^*)$$
$$\leq \frac{1}{T}\left[Q(p^*) + \tilde{L}(x^0, z^0, p^*) + \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right].$$

This inequality together with Eq. (3.64) imply

$$||\mathbb{E}(F(\bar{x}(T))) - (p^*)'\mathbb{E}(D\bar{x}(T) + H\bar{z}(T)) - F(x^*)||$$
$$\leq \frac{1}{T}\left[Q(p^*) + \tilde{L}(x^0, z^0, p^*) + \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right].$$

By triangle inequality, we obtain

$$||\mathbb{E}(F(\bar{x}(T))) - F(x^*)|| \leq \frac{1}{T}\left[Q(p^*) + \tilde{L}(x^0, z^0, p^*) + \frac{1}{2\beta}\left|\left|p^0 - p^*\right|\right|_{\bar{\Lambda}}^2 \right. \tag{3.65}$$
$$\left. + \frac{\beta}{2}\left|\left|H(z^0 - z^*)\right|\right|_{\bar{\Lambda}}^2\right] + ||\mathbb{E}((p^*)'(D\bar{x}(T) + H\bar{z}(T)))||,$$

Using definition of Euclidean and l_∞ norms,[10] the last term $||\mathbb{E}((p^*)'(D\bar{x}(T) + H\bar{z}(T)))||$

[10]We use the standard notation that $||x||_\infty = \max_i |x_i|$.

satisfies

$$||\mathbb{E}((p^*)'(D\bar{x}(T)+H\bar{z}(T)))|| = \sqrt{\sum_{l=1}^{W}(p_l^*)^2[\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))]_l^2}$$
$$\leq \sqrt{\sum_{l=1}^{W}||p^*||_\infty^2\,[\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))]_l^2} = ||p^*||_\infty\,||\mathbb{E}(D\bar{x}(T)+H\bar{z}(T))||\,.$$

The above inequality combined with Eq. (3.60) yields,

$$||\mathbb{E}((p^*)'(D\bar{x}(T)+H\bar{z}(T)))|| \leq \frac{||p^*||_\infty}{T}\left[\bar{Q}+\tilde{L}^0+\frac{1}{2\beta}\left|\left|p^0-p^*\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^0-z^*)\right|\right|_{\bar{\Lambda}}^2\right].$$

Hence, Eq. (3.65) implies

$$||\mathbb{E}(F(\bar{x}(T)))-F(x^*)|| \leq \frac{||p^*||_\infty}{T}\left[\bar{Q}+\tilde{L}^0+\frac{1}{2\beta}\left|\left|p^0-p^*\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^0-z^*)\right|\right|_{\bar{\Lambda}}^2\right]+$$
$$\frac{1}{T}\left[Q(p^*)+\tilde{L}(x^0,z^0,p^*)+\frac{1}{2\beta}\left|\left|p^0-\bar{\theta}\right|\right|_{\bar{\Lambda}}^2+\frac{\beta}{2}\left|\left|H(z^0-z^*)\right|\right|_{\bar{\Lambda}}^2\right].$$

Thus we have established the desired relation (3.61).

□

We note that by Jensen's inequality and convexity of the function F, we have

$$F(\mathbb{E}(\bar{x}(T))) \leq \mathbb{E}(F(\bar{x}(T))),$$

and the preceding results also holds true when we replace $\mathbb{E}(F(\bar{x}(T)))$ by $F(\mathbb{E}(\bar{x}(T)))$.

Note that the convergence is established with respect to the time ergodic average sequence $\{\bar{x}^k\}$. In practice, one might implement the algorithm with the ergodic average $\bar{x}_i^k$ for each agent i saved at each step to achieve the $O\left(\frac{1}{k}\right)$ rate of convergence.

3.4 Numerical Studies

In this section, we analyze the numerical performance of our proposed distributed asynchronous ADMM method. We compare this to the performance of asynchronous gossip algorithm proposed in [77]. The asynchronous gossip algorithm is based on gossip algorithm, which converges with best known rate $O(1/\sqrt{k})$. We performed the algorithms on two different 5 node networks for problem (3.2) with quadratic objective functions. The first network is the same as Figure. 3-1, the second one is a path graph, where the nodes form a line. At each iteration, an randomly picked edge becomes active for both methods. For each of these graphs, we plot the sample objective function value and the value of constraint violation Ax of both methods against iteration count, with dotted black lines denoting 10% interval of the optimal objective function value and when feasibility violation is less than 0.1. We also record the number of iterations required to reach the 10% precision and to have constraint violation less than 0.1 for both methods. Figure. 3-2 and Table. 3.1 correspond to the network as shown in Figure. 3-1. Figure. 3-3 and Table. 3.2 are for the path graph. For both of these network, the asynchronous distributed ADMM outperforms the asynchronous gossip algorithm significantly, which confirms with the intuition suggested by the faster rate of convergence associated with asynchronous ADMM method.

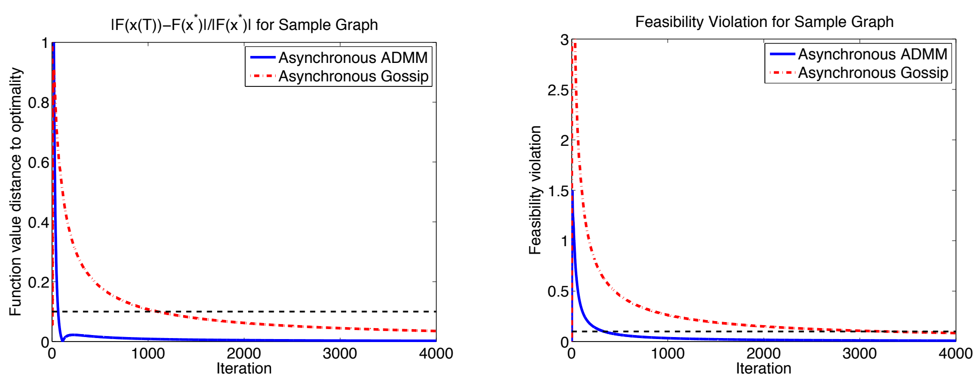

Figure 3-2: One sample objective function value and constraint violation ($Ax = 0$) of both methods against iteration count for the network as in Figure. 3-1. The dotted black lines denote 10% interval of the optimal value and feasibility violation less than 0.1.

	Number of Iterations to 10% of Optimal Function Value	Number of Iterations to Feasibility Violation < 0.1
Asynchronous ADMM	65	336
Asynchronous Gossip	1100	3252

Table 3.1: Number of iterations needed to reach 10% of optimal function value and feasibility violation less than 0.1 for the network as in Figure. 3-1.

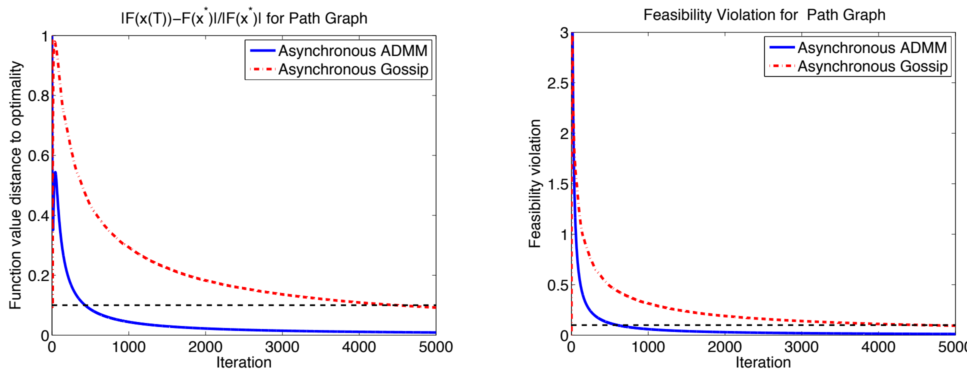

Figure 3-3: One sample objective function value and feasibility violation of both methods against iteration count for the path graph. The dotted black lines denote 10% interval of the optimal value and constraint violation ($Ax = 0$) less than 0.1.

3.5 Summaries

We developed a fully asynchronous ADMM based algorithm for a convex optimization problem with separable objective function and linear constraints. This problem is motivated by distributed multi-agent optimization problems where a (static) network of agents each with access to a privately known local objective function seek to optimize the sum of these functions using computations based on local information and communication with neighbors. We show that this algorithm converges almost surely to an optimal solution. Moreover, the rate of convergence of the objective function values and feasibility violation is given by $O(1/k)$. Future work includes investigating network effects (e.g., effects of communication noise, quantization) and time-varying network topology on the performance of the algorithm.

	Number of Iterations to 10% of Optimal Function Value	Number of Iterations to Feasibility Violation < 0.1
Asynchronous ADMM	431	596
Asynchronous Gossip	4495	4594

Table 3.2: Number of iterations needed to reach 10% of optimal function value and feasibility violation less than 0.1 for the path network.

Chapter 4

Market Analysis of Electricity Market: Competitive Equilibrium

For the next two chapters, we study the electricity market.

In this chapter, we present a competitive equilibrium framework for the electricity market, which can be shown to improve significantly over the traditional supply-follow-demand market structure. We take a systematic approach at the analysis and design of the general market dynamics in power networks, which include heterogeneous users with shiftable demand. Both the supply and demand actively participate in the market and interact through means of pricing signals. We first propose a T time period (24-hour in a day-ahead market example) model that captures these features and give a characterization of the equilibrium prices and quantities. We show that at competitive equilibrium, the market achieves the maximum social welfare. Chapter 5 builds on this framework and targets to mitigate the generation fluctuation.

This chapter is organized as follows: Section 4.1 contains our model on the electricity market. In Section 4.2, we characterize and analyze competitive equilibrium outcome and show that the competitive equilibrium improves significantly over the traditional production model. Section 4.3 summarizes and concludes this chapter.

4.1 Model

In this section, we introduce a T time horizon model describing the interaction between generators and consumers in the electricity market, which can represent the 24-period (one for each hour) day-ahead market. We assume there are many participants in the market and hence both the consumers and suppliers are price-taker.

4.1.1 Demand side

We assume there are large number of *price taking* heterogenous consumers in market, indexed by $1, \ldots, N$. Each user has a total demand of d_i with $d_i > 0$, which can be allocated over the entire time horizon $t = 1, \ldots, T$ and need not be fully satisfied. In each time period, the maximal amount of electricity user i can draw from the grid is specified by $m_i > 0$. This parameter models the maximal power consumption rate of user i, which is determined by the physical properties of the appliances and other electrical equipment. For instance, for an PEV (Plug-in Electric Vehicles), the parameter d_i is the battery size and m_i is the maximal charging rate. Each of the consumers has a private utility function $u_i^t(x_i^t)$, representing the utility the consumer derives from consumption of x_i^t units of power at time period t. If a user has preference for using power at particular time, then it is reflected in the utility function associated with that time instance. The decision each consumer makes is how to allocate the demand over the time horizon given user specific preferences and prices of electricity. [1] We assume a quasi-linear model of the consumer utility, i.e., the consumer utility is linear in the electricity price paid. Formally, the optimization problem for each

[1] This model can be applied to both elastic demand and inelastic demand depending on the utility functions.

consumer i is given as follows

$$\begin{aligned} \max_{x_i^t} \quad & \sum_{t=1}^{T} u_i^t(x_i^t) - p^t x_i^t && (4.1) \\ \text{s.t.} \quad & \sum_{t=1}^{t} x_i^t \leq d_i, \\ & 0 \leq x_i^t \leq m_i. \end{aligned}$$

where p^t denotes the market price of electricity at time t.

Assumption 7. *For all types i at all time periods t, the utility function $u_i^t(\cdot) : \mathbb{R} \to \mathbb{R}$ is concave and monotonically increasing. The scalars d_i and m_i are positive.*

4.1.2 Supply side

We model the supply side by one central benevolent supplier, which does not exercise market power. This reflects well the market interaction for a large number of suppliers. Let $\bar{x}^t$ denote the total supply at time instance t, i.e.,

$$\bar{x}^t = \sum_{i=1}^{N} x_i^t.$$

The supplier incurs a cost of production $c^t(\bar{x}^t)$ for supplying $\bar{x}^t$ unit at time t.

The supplier side optimization problem is to maximize profit, given as

$$\max_{\bar{x}^t \geq 0} \quad \sum_{t=1}^{T} p^t \bar{x}^t - c^t(\bar{x}^t), \tag{4.2}$$

where the price is the market price.

Assumption 8. *For all time periods t, the cost function $c^t(\cdot) : \mathbb{R} \to \mathbb{R}$ is convex and strictly monotonically increasing.*

4.1.3 Competitive Equilibrium

A competitive equilibrium is a sequence of prices $\{p^t\}$ and supply $\{\bar{x}^t\}$ and demand $\{x_i^t\}$ sequences such that

- x_i^t maximizes the utility of type i consumer, i.e., the vector $[x_i^t]_t$ is an optimal solution to problem (4.1).
- $\bar{x}^t$ maximizes the profit of supplier, i.e., the vector $[\bar{x}^t]_t$ is an optimal solution to problem (4.2).
- p^t is such that the market clears, i.e., $\sum_{i=1}^N x_i^t = \bar{x}^t$ for all t, i.e., $\sum_i x_i^t = \bar{x}^t$.

By the well-known first and second welfare theorems, a competitive equilibrium is Pareto optimal and every Pareto optimal allocation can be decentralized as a competitive equilibrium.

4.2 Characterizing the Competitive Equilibrium

By the second welfare theorem (and because utility is quasi-linear), a competitive equilibrium can be characterized by maximizing social welfare given as

$$\begin{aligned} \max_{x_i^t, \bar{x}^t} \quad & \sum_{t=1}^T \sum_{i=1}^N u_i^t(x_i^t) - c^t(\bar{x}^t), \\ \text{s.t.} \quad & \sum_{t=1}^T x_i^t \le d_i \text{ for } i \in \{1, \dots, N\}, \\ & \sum_{i=1}^N x_i^t = \bar{x}^t, \text{for } t \in \{1, \dots, T\}, \\ & 0 \le x_i^t \le m_i, \quad \bar{x}^t \ge 0. \end{aligned} \tag{4.3}$$

The market clearing prices emerge as the dual variables to the last constraint. The constraint forcing supply equal to demand, i.e., $\sum_{i=1}^N x_i^t = \bar{x}^t$, is referred to as the *market clearing* constraint. The constraint of $\sum_{t=1}^T x_i^t \le d_i$ is called *total demand* constraint.

4.2.1 Efficiency

Theorem 4.2.1. *Any competitive equilibrium defined in Section 4.1.3 is efficient, i.e., is an optimal solution to the social welfare maximization problem [cf. problem (4.3)]. Any optimal solution to the social welfare maximization problem [cf. problem (4.3)] can be represented by a competitive equilibrium.*

Proof. The equivalence can be established via a duality argument. Consider problem (4.3), we associate p^t with each constraint of the type $\sum_{i=1}^{N} x_i^t = \bar{x}^t$. Then the Lagrangian can be written as

$$\begin{aligned} L(x_i^t, \bar{x}^t, p^t) &= \sum_{t=1}^{T} \sum_{i=1}^{N} u_i^t(x_i^t) - c^t(\bar{x}^t) - H(\bar{x}^t, \bar{x}^{t-1}) - \sum_{t=1}^{T} p^t (\sum_{i=1}^{N} x_i^t - \bar{x}^t) \\ &= \sum_{t=1}^{T} \sum_{i=1}^{N} u_i^t(x_i^t) - \sum_{t=1}^{T} p^t \sum_{i=1}^{N} x_i^t - c^t(\bar{x}^t) - H(\bar{x}^t, \bar{x}^{t-1}) + \sum_{t=1}^{T} p^t \bar{x}^t = \end{aligned}$$

We first observe that any competitive equilibrium price-allocation pair is feasible to problem (4.3), where the first constraint is satisfied due to consumer utility maximization constraints and the market clearing constraint is satisfied due to the competitive equilibrium definition. Since problem (4.3) maximizes welfare, the social welfare can be no worse than any competitive equilibrium outcomes.

We now show that the optimal objective function value in problem (4.3) can be no better than that in a competitive equilibrium.

$$\begin{aligned}
&\max_{\{x_i^t,\bar{x}^t\}\geq 0, x_i^t\leq m_i, \sum_{t=1}^T x_i^t = d_i, \sum_{i=1}^N x_i^t = \bar{x}^t,} \sum_{t=1}^T \sum_{i=1}^N u_i^t(x_i^t) - c^t(\bar{x}^t) - H(\bar{x}^t, \bar{x}^{t-1}) \\
\leq &\max_{\{x_i^t,\bar{x}^t\}\geq 0, x_i^t\leq m_i, \sum_{t=1}^T x_i^t = d_i, \sum_{i=1}^N x_i^t = \bar{x}^t,} \sum_{t=1}^T \sum_{i=1}^N u_i^t(x_i^t) \\
+ &\max_{\{x_i^t,\bar{x}^t\}\geq 0, x_i^t\leq m_i, \sum_{t=1}^T x_i^t = d_i, \sum_{i=1}^N x_i^t = \bar{x}^t,} \sum_{t=1}^T -c^t(\bar{x}^t) - H(\bar{x}^t, \bar{x}^{t-1}) \\
\leq &\max_{\{x_i^t\}\geq 0, x_i^t\leq m_i, \sum_{t=1}^T x_i^t = d_i, \sum_{i=1}^N x_i^t = \bar{x}^t,} \sum_{t=1}^T \sum_{i=1}^N u_i^t(x_i^t) \\
+ &\max_{\{\bar{x}^t\}\geq 0, \sum_{i=1}^N x_i^t = \bar{x}^t,} \sum_{t=1}^T -c^t(\bar{x}^t) - H(\bar{x}^t, \bar{x}^{t-1}),
\end{aligned}$$

where in the first inequality we split the objective function into two problems and in the second inequality we relaxed some constraint sets. The last sum in the inequality chain is the competitive outcome, i.e., summation of problem (4.1) and (4.2) under the market clearing condition. Thus we establish the desired results. □

The above theorem also suggests that at competitive equilibrium, the market clearing price p^t is the dual variable associated with the constraint $\sum_{i=1}^N x_i^t = \bar{x}^t$.

4.2.2 Efficiency Improvement Over Traditional Grid

In this section, we show that the social welfare in the traditional grid where all demand are viewed as inelastic can be arbitrary worse than the demand response system we described above. For simplicity, we assume that the generation cost is time invariant and represented by a quadratic function $c^t(\bar{x}^t) = \frac{k}{2}(\bar{x}^t)^2$ for some positive scalar k. [2]

We consider the following two user two period simple example. Both users have linear utility, slight preference for period 1 and the utility of user 1 is slightly higher than user 2. Both consumers have the total demand equal to the maximal consumption in one period, i.e., $m_1 = m_2 = d_1 = d_2 = m$ for some m. Formally, the consumer

[2] The value of k reflects the production technology.

utilities are given by

$$u_1^1(x_1^1) = (k+\epsilon)mx_1^1, \qquad u_1^2(x_1^2) = km(1-\frac{2\epsilon}{k})x_1^2,$$
$$u_2^1(x_2^1) = kmx_2^1, \qquad u_2^2(x_2^2) = km(1-\frac{\epsilon}{2k})x_2^2,$$

for some small positive scalar ϵ.

- Traditional Grid Operation: In the traditional grid, there is no pricing signal for the consumers to respond to and therefore each consumer chooses to maximize their utility according to

$$\begin{aligned} \max_{x_i^t} \quad & \sum_{t=1}^{T} u_i^t(x_i^t) \\ \text{s.t.} \quad & \sum_{t=1}^{t} x_i^t \leq d_i, \\ & 0 \leq x_i^t \leq m_i. \end{aligned}$$

  Both consumers will prefer to allocate the entire consumption m in first time period, and obtain utilities of $(k+\epsilon)m^2$ and km^2 respectively. Generator views the demand as inelastic and has satisfy it, incurring a cost of $2km^2$. The social welfare, denoted by W^o, is given by

$$W^o = (k+\epsilon)m^2 + km^2 - 2km^2 = \epsilon m^2.$$

- In the market with demand response, at the competitive equilibrium, we have $x_1^1 = m$, $x_1^2 = 0$, $x_2^1 = 0$, $x_2^2 = m(1-\frac{\epsilon}{2k})$, $p^1 = km$, $p^2 = km(1-\frac{\epsilon}{2k})$. The utility for user 1 is $u_1^1(x_1^1) + u_1^2(x_1^2) = (k+\epsilon)m^2$ and for user 2 is $u_2^1(x_2^1) + u_x^2(x_2^2) = km^2(1-\frac{\epsilon}{2k})^2$. Cost of production is $c^1(\bar{x}^1) + c^2(\bar{x}^2) = \frac{k}{2}m^2(2-\frac{\epsilon}{2k})$. Hence the new social welfare, W^n, is given by

$$W^n = (k+\epsilon)m^2 + km^2(1-\frac{\epsilon}{2k})^2 - \frac{k}{2}m^2(2-\frac{\epsilon}{2k}) = km^2 + \frac{\epsilon}{4}m^2 + \frac{\epsilon^2 m^2}{4k}.$$

Since ϵ can be chosen to be arbitrarily small, we have that the ratio of $\frac{W^n}{W^o}$ can be arbitrarily large and therefore the improvement brought by demand response is very significant.

4.3 Summaries

In this chapter, we propose a flexible multi-period competitive equilibrium framework to analyze supply-demand interaction in the electricity market. We show that by introducing demand response, the market can improve efficiently significantly than the traditional supply-follow-demand setup, where demand is viewed as exogenous.

Chapter 5

Market Analysis of Electricity Market: Generation Fluctuation Reduction

In this chapter of the thesis, we continue the study of electricity market using the framework in Chapter 4. The focus of this chapter is to understand and control generation fluctuation, which naturally arises from competitive equilibrium. We show that the generation fluctuation and price fluctuation are proportionally correlated when the production cost structure is quadratic in Section 5.1. We then introduce an explicit penalty term on the price fluctuation, analyze properties of and develop distributed algorithm for the penalized formulation Section 5.2. In Section 5.3, we study the connection between the generation fluctuation and the storage size. Section 5.4 summarizes our development.

5.1 Price and Generation Fluctuation

We focus on the time-invariant quadratic generation cost scenario, where

$$c^t(\bar{x}) = \frac{k}{2}(\bar{x}^t)^2 \tag{5.1}$$

for some positive scalar k. We define the following two vectors both in $\mathbb{R}^{T-1}$: the price fluctuation vector $\Delta p = [p^t - p^{t+1}]_t$ and generation fluctuation vector $\Delta \bar{x} = [\bar{x}^t - \bar{x}^{t+1}]_t$. We first relate price and generation fluctuation by the following theorem.

Theorem 5.1.1. *Let $\bar{x}, p$ be the generation level and price vector be an optimal solution for the social welfare maximization problem (4.3). If the generation cost function is time-invariant and convex, we have $\Delta p^k \Delta \bar{x}^t \geq 0$, for all periods with $\bar{x}^t > 0$, $\bar{x}^{t+1} > 0$. In particular, for the quadratic generation cost in Eq. (5.1), we have we have $\Delta p^t = k \Delta \bar{x}^t$, for all periods with $\bar{x}^t > 0$, $\bar{x}^{t+1} > 0$.*

Proof. According to Theorem 4.2.1, we have the optimal solution to problem (4.3) can be represented by an optimal solution to problem (4.2) with the same price vector. Hence, for price p, $\bar{x}$ is the optimal solution for (4.2) and hence satisfies the following optimality condition:

$$p^t - \nabla c^t(\bar{x}^t) + \mu_0^t = 0, \tag{5.2}$$

where μ_0^t is the dual variable associated with constraint $\bar{x}^t \geq 0$ and satisfies complementary slackness condition, i.e., $\mu_0^t \bar{x}^t = 0$. For periods with $\bar{x}^t > 0$, we have $\mu_0^t = 0$. Hence

$$\Delta p^t = \nabla c^t(\bar{x}^t) - \nabla c^{t+1}(\bar{x}^{t+1}) = \nabla c^t(\bar{x}^t) - \nabla c^t(\bar{x}^{t+1}),$$

where the last equality follows from the time-invariant nature of generation cost functions. By convexity of generation cost function c^t, we have

$$(\bar{x}^t - \bar{x}^{t+1})(\nabla c^t(\bar{x}^t) - \nabla c^t(\bar{x}^{t+1})) \geq 0.$$

Hence, we have $\Delta p^k \Delta \bar{x}^t \geq 0$.

When the generation costs are quadratic with $\nabla c^t(\bar{x}^t) = k\bar{x}^t$, we then have for t with $\bar{x}^t > 0$,

$$p^t = k\bar{x}^t.$$

By taking difference of periods t and $t+1$ in Eq. (5.2), we establish the desired relation. □

The preceding theorem suggests that at competitive equilibrium, the price and generation fluctuations are proportionally related when the generation costs are quadratic functions.

The worst case scenario for price and generation fluctuation is when all the users have the same kind of preferences over time and therefore all rush to or avoid the same period. To analyze this scenario, we make the following assumption on the utility functions for the rest of this chapter.

Assumption 9. ***(Monotone Preference over Time)*** *For all consumers i and all periods $t < T$, the utility functions u_i^t are differentiable with $\nabla u_i^t(x) \geq \nabla u_i^{t+1}(x)$, for all x in $[0, m]$. Furthermore, for each user i either the functions u_i^t are strictly concave or $\nabla u_i^t(x) > \nabla u_i^{t+1}(x)$ for all t.*

5.2 Price Fluctuation Penalized Problem

In the previous section, we established a direct connection between generation and price fluctuation. In this section, we explore this relationship and introduce a price fluctuation penalty to the social objective function in order to control the generation level fluctuation. We will establish an alternative equivalent formulation of the price fluctuation penalized problem, which involves storage in Section 5.2.1. Section 5.2.2 derives many important properties of the optimal solution for the penalized problem and gives a simpler equivalent reformulation. Finally, we develop a distributed implementation of the price fluctuation penalized formulation in Section 5.2.3.

5.2.1 Fluctuation Penalized Problem Formulation

We recall that price is given as the dual multiplier to the market clearing constraint. To write the dual problem more compactly, we use x to denote the vector of size NT representing the consumption levels x_i^t and vector $\bar{x}$ for the generation levels $[\bar{x}^t]_t$, vectors m and d both in $\mathbb{R}^N$ with $m = [m_i]_i$ and $d = [d_i]_i$ respectively. We introduce matrix B in $\mathbb{R}^{N \times NT}$, which is element-wise either 0 or 1, and $Bx \leq d$ is equivalent to the total demand constraint. Matrix A in $\mathbb{R}^{T\times NT}$ is another element-wise 0 or 1 matrix and $Ax = \bar{x}$ is the compact representation of the market clearing condition. We let function $u : \mathbb{R}^{NT} \to \mathbb{R}$ represent the total consumer utility with

$$u(x) = \sum_{t=1}^{T}\sum_{i=1}^{N} u_i^t(x_i^t). \tag{5.3}$$

and function $c : \mathbb{R}^T \to \mathbb{R}$ is the total production cost, where

$$c(\bar{x}) = \sum_{t=1}^{T} c^t(\bar{x}^t). \tag{5.4}$$

Then the social welfare optimization can be written as

$$\begin{aligned} \max_{\{x,\bar{x}\}\geq 0, x\leq m} \quad & u(x) - c(\bar{x}), \\ \text{s.t.} \quad & Bx \leq d, \\ & Ax = \bar{x}. \end{aligned}$$

The dual problem for the above optimization problem is given as

$$\min_{\mu\geq 0, p}\ \max_{\{x,\bar{x}\}\geq 0, Bx\leq d, x\leq m} \quad u(x) - c(\bar{x}) - p'(Ax - \bar{x}).$$

We denote by e the vector of all 1 in $\mathbb{R}^T$ and by E the matrix in $\mathbb{R}^{T\times(T-1)}$ with

$$E = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ & & \vdots & 1 \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix}, \tag{5.5}$$

such that

$$p = ep_T + E\Delta p.$$

Then the price fluctuation penalty can be written as $||\Delta p||$ for some notion of norm and the price penalized problem can be written as

$$\max_{p_T,\Delta p} \min_{m\geq x\geq 0,\bar{x}\geq 0,Bx\leq d} -u(x)+c(\bar{x})+(ep_T+E\Delta p)'(Ax-\bar{x})-S\,||\Delta p||_w \,.$$

We will next derive an alternative formulation to this fluctuation penalized problem.

Theorem 5.2.1. *The following two problems are equivalent*

$$\max_{p_T,\Delta p} \min_{m\geq x\geq 0,\bar{x}\geq 0,Bx\leq d} -u(x)+c(\bar{x})+(ep_T+E\Delta p)'(Ax-\bar{x})-S\,||\Delta p||_w \,. \tag{5.6}$$

$$\begin{aligned} \min_{x,\bar{x}} \quad & -u(x)+c(\bar{x}), \\ s.t. \quad & ||E'(Ax-\bar{x})||_q \leq S, \\ & e'(Ax-\bar{x})=0, \\ & Bx\leq d, \\ & m\geq x\geq 0, \quad \bar{x}\geq 0. \end{aligned}$$

where the two norms $||\cdot||_w$ *and* $||\cdot||_q$ *are dual operators and* S *is a nonnegative scalar.*

Proof. We first rearrange problem (5.6) as

$$\max_{p_T,\Delta p} \min_{m\geq x\geq 0,\bar{x}\geq 0,Bx\leq d} -u(x)+c(\bar{x})+p_Te'(Ax-\bar{x})+\Delta p'E'(Ax-\bar{x})-S\,||\Delta p||_w \,.$$

By Saddle Point Theorem (Proposition 2.6.4 in [81]), we can write the problem equivalently as

$$\min_{m\geq x\geq 0,\bar{x}\geq 0,Bx\leq d} \max_{p_T,\Delta p} -u(x)+c(\bar{x})+p_Te'(Ax-\bar{x})+\Delta p'E'(Ax-\bar{x})-S\,||\Delta p||_w \,.$$

For any solution pair with $e'(Ax-\bar{x})\neq 0$, there exists p_T such that the value of the previous problem attains infinity and thus cannot be the optimal solution. Therefore we conclude $e'(Ax - \bar{x}) = 0$.

We next analyze the terms involving Δp: $\Delta p' E'(Ax - \bar{x}) - S\,||\Delta p||_w$. From Höler's inequality, we have

$$\Delta p' E'(Ax - \bar{x}) - S\,||\Delta p||_w \leq ||\Delta p||_w\,(||E'(Ax - \bar{x})||_q - S),$$

i.e.,

$$\max_{\Delta p} \Delta p' E'(Ax - \bar{x}) - S\,||\Delta p||_w = ||\Delta p^*||_w\,(||E'(Ax - \bar{x})||_q - S),$$

where Δp^* is the maximizer. Therefore problem (5.6) is equivalent to

$$\max_{\Delta p} \min_{m \geq x \geq 0, \bar{x} \geq 0, Bx \leq d, e'(Ax - \bar{x}) = 0} -u(x) + c(\bar{x}) + ||\Delta p||_w\,(||E'(Ax - \bar{x})||_q - S).$$

Since Δp is unconstrained, the scalar $||\Delta p||$ can take any nonnegative values, which yields another equivalent formulation of

$$\max_{\alpha \geq 0} \min_{m \geq x \geq 0, \bar{x} \geq 0, Bx \leq d, e'(Ax - \bar{x}) = 0} -u(x) + c(\bar{x}) + \alpha(||E'(Ax - \bar{x})||_q - S).$$

If $||E'(Ax - \bar{x})||_q - S \geq 0$, then the problem attains value of infinity and thus cannot be optimal, and therefore the preceding formulation is equivalent to (PW2). □

For the rest of the chapter, we will focus on the L_1 norm of the fluctuation vectors as a metric of total fluctuation, i.e., the following problem

$$\begin{aligned} \min_{x, \bar{x}} \quad & -u(x) + c(\bar{x}), \\ \text{s.t.} \quad & -S \leq \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau \leq S, \quad \text{for } t = 1, \ldots, T-1, \\ & e'(Ax - \bar{x}) = 0, \\ & Bx \leq d, \\ & m \geq x \geq 0, \quad \bar{x} \geq 0, \end{aligned} \tag{PW}$$

where we used definition of matrix E [c.f. (5.5)].

We note that quantity $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau$ represents the difference between total supply and total supply (accumulated up to time t). We refer to the difference $[Ax-\bar{x}]_t$ as *the supply-demand mismatch* for period t. The constraint $-S \leq \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \leq S, \quad$ for $t = 1, \ldots, T-1$, ensures that the total supply-demand mismatch is upper bounded by S. This can be guaranteed by providing a storage of size $2S$ with initial charge S, which may be charged and discharged during the T periods. Constraint $||E'(Ax - \bar{x})||_q \leq S$ guarantees the storage will stay with nonnegative charge and within its limit, while constraint $e'(Ax - \bar{x}) = 0$ ensures at the end of the T cycle, the storage returns to be charged with S and is ready for the next cycle. Thus the price penalty parameter S is precisely the storage size necessary to implement the penalized problem. We will refer to S as the storage size, the constraint involving S as the *storage size constraint.* We call the constraint $e'(Ax - \bar{x}) = 0$ the *total market balancing constraint.*

5.2.2 Properties of the Fluctuation Penalized Problem

In this section, we derive some important properties of the fluctuation penalized problem and its optimal solution.

We associate dual variable Δq with constraints $-S \leq \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau$, Δp with $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \leq S$, p_T with $e'(Ax - \bar{x}) = 0$, μ_d with $Bx \leq d$, μ_0 with $x \geq 0$, μ_m with $x \geq m$ and λ_0 with $\bar{x} \geq 0$. Lemma 5.2.1 characterizes the optimal solution of problem (PW).

Lemma 5.2.1. *The following conditions are necessary and sufficient for a primal-dual solution*

$(x, \bar{x}, \Delta p, \Delta q, p_T, \mu_d, \mu_0, \mu_m, \lambda_0)$ *to be optimal for problem (PW).*

$$
\begin{aligned}
&-\nabla u(x) - A'E\Delta q + A'E\Delta p + A'ep_T + B\mu_d + \mu_m - \mu_0 = 0, \\
&k\bar{x} + E\Delta q - E\Delta p - ep_T - \lambda_0 = 0, && \text{(FOC)} \\
&-S \leq \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau \leq S, \quad \textit{for } t = 1, \ldots, T-1, \\
&e'(Ax - \bar{x}) = 0, \quad Bx \leq d, m \geq x \geq 0, \quad \bar{x} \geq 0, \\
&\Delta q, \Delta p, \mu_m, \mu_0, \mu_d, \lambda_0 \geq 0 \\
&\Delta q' E'(Ax - \bar{x} + S) = 0, \quad \Delta p' E'(Ax - \bar{x} - S) = 0, && \text{(CS)} \\
&\mu_m'(x - m) = 0, \quad \mu_0' x = 0, \quad \mu_d'(Bx - d) = 0, \quad \lambda_0' \bar{x} = 0.
\end{aligned}
$$

Proof. Since the utility functions are concave and the cost functions are convex, problem (PW) is convex with differentiable objective function. Therefore the KKT conditions are the necessary and sufficient conditions to characterize an optimal solution, which includes first order stationary condition (FOC), primal feasibility, dual feasibility and complementary slackness conditions (CS) as above. □

The next two lemmas establish conditions when the optimal solution will be interior to constraint $\bar{x} \geq 0$ and $Bx \leq d$ respectively.

Lemma 5.2.2. *For problem (PW) with* $S > 0$*, if there exists one consumer* i *with* $\nabla u_i^\tau(0) > 0$*, then* $\bar{x}^t > 0$ *for all* t *at any optimal solution.*

Proof. We first argue that at optimal solution there exists at least one t with $\bar{x}^t > 0$. We assume the contrary, $\bar{x}^t = 0$ and by total market clearing constraints, we have $\sum_{i,t} x_i^t = 0$, hence $x_i^t = 0$ for all i and t. From condition (FOC) in Lemma 5.2.1, we have

$$\nabla u_i^t(x_i^t) - [\mu_d]_i - [\mu_m]_i^t + [\mu_0]_i^t = k\bar{x}^t - \lambda_0^t,$$

for all τ and i. For this all 0 solution, the constraints $x_i^t \leq m$ and $\sum_{t=1}^T x_i^t \leq d_i$ are not tight, and hence $mu_d = \mu_m = 0$. We have for user j at time τ

$$\nabla u_i^\tau(x_i^\tau) = -[\mu_0]_i^\tau - \lambda_0^\tau \leq 0,$$

which is a contradiction. Hence we have shown the existence of some t with $\bar{x}^t > 0$.

We next argue by contradiction that $\bar{x}^t > 0$ for all t. If $T = 1$ the statement holds trivially. For $T \geq 2$, assume there exists a period θ with $\bar{x}^\theta = 0$ in an optimal solution $(x, \bar{x})$, whose (at least one) neighboring period β satisfies $\bar{x}^\beta > 0$. We construct a new solution $(y, \bar{y})$, which is identical to $(x, \bar{x})$, except $\bar{y}^\beta = \bar{x}^\beta - \epsilon$ and $\bar{y}^\theta = \bar{y}^\theta + \epsilon$ for some positive scalar ϵ with $\epsilon \leq \frac{\bar{x}^\beta - \bar{x}^\theta}{2}$. From strict convexity of generation cost function and Lemma 5.2.6, we obtain

$$\frac{k(\bar{x}^\theta)^2}{2} + \frac{k(\bar{x}^\beta)^2}{2} > \frac{k(\bar{x}^\theta + \epsilon)^2}{2} + \frac{k(\bar{x}^\beta - \epsilon)^2}{2}.$$

If solution $(y, \bar{y})$ is feasible, then the proof is completed. Otherwise, we have either

$$\sum_{\tau=1}^{\theta} [Ax - \bar{x}]_\tau = -S \quad \text{and } \theta < \beta, \quad \text{or} \quad \sum_{\tau=1}^{\beta} [Ax - \bar{x}]_\tau = S \quad \text{and } \theta > \beta.$$

In the first situation, we have $\sum_{\tau=1}^{\theta} [Ax]_\tau + S = \sum_{\tau=1}^{\theta} \bar{x}^\tau$, therefore, there exists a $t < \theta$ with $\bar{x}^t > 0$, we set β equal to the largest such t and repeat the previous ϵ-shifting procedure to obtain a feasible solution with better objective function value.

In the second situation, by the constraint $e'(Ax - \bar{x}) = 0$, we have $\beta < T$ and there exists a time $t > \beta$ with $0 \leq [Ax]_t < \bar{x}^t$, we set β to be the smallest such t with $t > \theta$ and repeat the previous ϵ-shifting procedure to obtain a feasible solution with better objective function value.

Hence we conclude that there is another feasible solution with better objective function value and therefore the original solution with $\bar{x}^\theta = 0$ cannot be optimal.

□

Lemma 5.2.3. *Let $(x, \bar{x})$ be any optimal solution for problem (PW). Then for consumer i with $d_i > (T - 1)m$ and*

$$\sum_{t=1}^{T} \nabla u_i^t(0) < kd_i, \tag{5.7}$$

the constraint $\sum_{t=1}^{T} x_i^t \leq d_i$ is not tight, i.e., $\sum_{t=1}^{T} x_i^t < d_i$ for any $S \geq 0$.

Proof. We assume the contrary, i.e., $\sum_{t=1}^{T} x_i^t = d_i$. From condition (FOC) in Lemma 5.2.1, we have

$$\nabla u_i^t(x_i^t) - [\mu_d]_i - [\mu_m]_i^t + [\mu_0]_i^t = k\bar{x}^t - \lambda_0^t,$$

for all t. Since $d_i > (T-1)m$, we have $x_i^t > 0$ for all periods t. Assumption 10 and Lemma 5.2.2 guarantee that $\bar{x}^t > 0$. Hence by complementary slackness condition $[\mu_0]_i^t = 0$ and $\lambda_0^t = 0$ for all t. Therefore the previous relation can be simplified to

$$\nabla u_i^t(x_i^t) - [\mu_d]_i - [\mu_m]_i^t = k\bar{x}^t,$$

We sum the above equation over t and have

$$\sum_{t=1}^{T} \nabla u_i^t(x_i^t) = \sum_{t=1}^{T} k\bar{x}^t + [\mu_d]_i + [\mu_m]_i^t.$$

From the constraint $e'(Ax - \bar{x}) = 0$, we have $\sum_{t=1}^{T} k\bar{x}^t \geq \sum_{t=1}^{T} kx_i^t = d$. Therefore, by nonnegativity of the dual variables μ_d and μ_m, we have

$$\sum_{t=1}^{T} k\bar{x}^t + [\mu_d]_i + [\mu_m]_i^t \geq kd.$$

Concavity of utility functions u_i^t implies that $\nabla u_i^t(x_i^t) \leq \nabla u_i^t(0)$, and thus we can sum over t and have

$$\sum_{t=1}^{T} \nabla u_i^t(x_i^t) \leq \sum_{t=1}^{T} \nabla u_i^t(0) < kd,$$

The previous three relations combined gives a contradiction and therefore we conclude $\sum_{t=1}^{T} x_i^t < d_i$. □

To avoid trivial solution of producing 0 electricity for all periods, we assume the following holds true for the rest of the chapter.

Assumption 10. *There exists one consumer i with $\nabla u_i^t(0) > 0$ for some time period t.*

By Lemma 5.2.2, this condition guarantees that the production is positive for all periods when $S > 0$.

We can now derive some monotonicity results of problem (PW). The next lemma states that price fluctuation is monotonically decreasing with parameter S, which will be used later to show that the generation fluctuation decreases with S in Corollary 5.2.1.

Lemma 5.2.4. *The price fluctuation* $||\Delta p||_w$ *at the optimal solution of problem (5.6) is a nonincreasing function in* S.

Proof. For notational clarity, we let

$$g(p_T, \Delta p) = \min_{m \geq x \geq 0, \bar{x} \geq 0, Bx \leq d} -u(x) + c(\bar{x}) + (ep_T + E\Delta p)'(Ax - \bar{x}).$$

Hence problem (5.6) can be written as

$$\max_{p_T, \Delta p} g(p_T, \Delta p) - S\,||\Delta p||_w\,.$$

Consider two nonnegative scalars $S_1 < S_2$. We denote by $p_T, \Delta p$ the optimal solution to the previous problem with $S = S_1$ and $q_T, \Delta q$ the optimal solution associated with $S = S_2$. Using optimality conditions of two problem instances, we have

$$g(p_T, \Delta p) - S_1\,||\Delta p||_w \geq g(q_T, \Delta q) - S_1\,||\Delta q||_w\,,$$

and

$$g(q_T, \Delta q) - S_2\,||\Delta q||_w \geq g(p_T, \Delta p) - S_2\,||\Delta p||_w\,.$$

By summing the previous two inequalities and canceling the terms involving function g, we obtain

$$(S_2 - S_1)\,||\Delta p||_w \geq (S_2 - S_1)\,||\Delta q||_w\,.$$

By assumption, we have $S_1 < S_2$ and we can divide both side by positive scalar $S_2 - S_1$ and conclude

$$||\Delta p||_w \geq ||\Delta q||_w\,,$$

i.e., the price fluctuation associated with the smaller S is at least as large as the fluctuation with bigger S, which establishes the claim. □

The above lemma holds for any $w \geq 0$ and therefor applies to (PW) where we used L_1 norm.

Lemma 5.2.5. *At the optimal solution of problem (PW), we have* $\Delta p = k\Delta\bar{x}$ *for* $S > 0$.

Proof. Assumption 10 guarantees that $\bar{x}^t > 0$ and thus by complementary slackness condition, we have $\mu_0^t = 0$ for all t. We can therefore drop μ_0^t from Eq. (FOC) in Lemma 5.2.1 and obtain

$$k\bar{x}^t = p^t,$$

which leads to the above result. □

Corollary 5.2.1. *The total generation fluctuation* $||\Delta\bar{x}||_1$ *is nonincreasing function in storage size* S.

Proof. We consider the L_1 norm of generation fluctuation and have

$$||\Delta\bar{x}||_1 = \sum_{t=1}^{T-1} |\bar{x}^{t+1} - \bar{x}^t| = x^0 - x^T,$$

where the last equality follows from the monotonicity shown in the preceding theorem.

The result then follows from Lemmas 5.1.1, 5.2.5 and 5.2.4. □

Remark: An alternative way to reduce generation fluctuation is to penalize directly $||\Delta\bar{x}||$. We chose instead to focus on the penalty involving $||\Delta p||$, because this formulation preserves two important properties from the competitive equilibrium formulation, i.e., monotonicity and the direct connection between price and generation fluctuation as shown in the previous two theorems. The monotonicity property is very intuitive and easy to explain to all stakeholders. Hence, it is a desirable feature in designing future electricity market, which will need support from various consumers, generators and government organizations. Even though we have been focusing on generation fluctuation, heavy price fluctuation also imposes difficulties in budgeting process for both large industrial consumers, such as aluminum factories, and generators who are operating on very thin margins and therefore is also an undesirable characteristic. The connection between generation and price fluctuations from Lemmas 5.2.5 and 5.2.4 can address this problem, since it implies a reduction in both generation and price fluctuation as a result of introducing storage.

The following simple lemma states a powerful property of convex functions and will play a key role in the rest of the analysis.

Lemma 5.2.6. *For a convex function $f : \mathbb{R} \to \mathbb{R}$ and $a < b$, $\epsilon > 0$ we have*

$$f(a) + f(b) \geq f(a + \epsilon) + f(b - \epsilon).$$

If function f is strictly convex, then

$$f(a) + f(b) > f(a + \epsilon) + f(b - \epsilon).$$

Proof. We first assume that function f is convex. We write scalar α to denote $\beta = \frac{\epsilon}{b-a}$, and therefore by simple calculation, we have

$$\beta a + (1 - \beta)b = \frac{\epsilon a}{b - a} + \frac{(b - a - \epsilon)b}{b - a} = \frac{(b - a)(b - \epsilon)}{b - a} = b - \epsilon, \tag{5.8}$$

and

$$(1 - \beta)a + \beta b = \frac{(b - a - \epsilon)a}{b - a} + \frac{\epsilon b}{b - a} = \frac{(b - a)(a + \epsilon)}{b - a} = a + \epsilon. \tag{5.9}$$

From convexity of function f, we have

$$\beta f(a) + (1 - \beta)f(b) \geq f(\beta a + (1 - \beta)b),$$

and

$$(1 - \beta)f(a) + \beta f(b) \geq f((1 - \beta)a + \beta b).$$

We can add the preceding two inequalities and have

$$f(a) + f(b) \geq f(\beta a + (1-\beta)b) + f((1-\beta)a + \beta b).$$

By substituting Eqs. (5.8) and (5.9) into the above relation, we can establish the desired result.

If function f is strictly convex, we have that all the inequalities above become strict, which shows the claim. □

We can now use the existing set of results to establish monotonicity of generation, consumption, supply-demand mismatch at optimal solution, which will then be used in Theorem 5.2.3 to show that the constraint $-S \leq \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau$ is redundant, so that we can simplify the problem and derive a distributed implementation in the next section.

Theorem 5.2.2. ***(Monotonicity of Total Consumption and Production)*** *Consider problem (PW) with a strict convex time-invariant generation cost function c^t. Then for any optimal solution $(x, \bar{x})$, both the total consumption level $[Ax]_t$ and generation level $[\bar{x}^t]^t$ are monotone nonincreasing in t, i.e., $[Ax]_t \geq [Ax]_{t+1}$, $\bar{x}^t \geq \bar{x}^{t+1}$ for all $1 \leq t < T$.*

Proof. We prove the claims by contradiction.

Monotonicity of $\bar{x}^t$:

We first establish the monotonicity in $\bar{x}^t$. Assume the contrary, i.e., there exists a period t with $\bar{x}^t < \bar{x}^{t+1}$. There are two possibilities: $[Ax]_t < [Ax]_{t+1}$ and $[Ax]_t \geq [Ax]_{t+1}$

(a) $[Ax]_t < [Ax]_{t+1}$

By using the definition of

$$[Ax]_t = \sum_{i=1}^{N} x_i^t,$$

and the fact that $x_i^t \geq 0$, we can infer that there exists a user i, with $x_i^t < x_i^{t+1}$ in this case. Let $\epsilon = \min\{\frac{x_i^{t+1} - x_i^t}{2}, \frac{\bar{x}^{t+1} - \bar{x}^t}{2}\}$, we will derive a contradiction by constructing another feasible solution with better objective function value than $(x, \bar{x})$. This new solution $(y, \bar{y})$ is constructed by shifting ϵ consumption from time $t+1$ to t for user i and adjust production by same amount, i.e., $y_i^t = x_i^t + \epsilon$, $y_i^{t+1} = x_i^{t+1} - \epsilon$, $\bar{y}^t = \bar{x}^t + \epsilon$, $\bar{y}^{t+1} = \bar{x}^{t+1} - \epsilon$, all other y_j^τ and $\bar{x}^\tau$ same as $(x, \bar{x})$. From definition of ϵ, we have

$$x_i^t < x_i^t + \epsilon \leq x_i^{t+1} - \epsilon < x_i^{t+1}. \tag{5.10}$$

Since the utility functions u_i^t are concave, the functions $-u_i^t$ are convex, by Lemma 5.2.6 and we have

$$u_i^t(x_i^t + \epsilon) - u_i^t(x_i^t) \geq u_i^t(x_i^{t+1}) - u_i^t(x_i^{t+1} - \epsilon).$$

The functions u_i^t are differentiable, therefore by fundamental theorem of calculus, the right hand side can be written as

$$u_i^t(x_i^{t+1}) - u_i^t(x_i^{t+1} - \epsilon) = \int_{x_i^{t+1}-\epsilon}^{x_i^{t+1}} \nabla u_i^t(\tau) d\tau \geq \int_{x_i^{t+1}-\epsilon}^{x_i^{t+1}} \nabla u_i^{t+1}(\tau) d\tau = u_i^{t+1}(x_i^{t+1}) - u_i^{t+1}(x_i^{t+1} - \epsilon),$$

where the inequality follows from Assumption 9. Assumption 9 guarantees at least one of the above two inequalities will be strict. We can therefore add these two inequalities together and deduce

$$u_i^{t+1}(x_i^{t+1} - \epsilon) + u_i^t(x_i^t + \epsilon) > u_i^{t+1}(x_i^{t+1}) + u_i^t(x_i^t). \tag{5.11}$$

We use the definition of ϵ one more time and have

$$\bar{x}^t < \bar{x}^t + \epsilon \leq \bar{x}^{t+1} - \epsilon < x_i^{t+1}. \tag{5.12}$$

The functions c^t are time invariant and strictly convex. Hence, we can apply

Lemma 5.2.6 to generation cost, and obtain

$$c^t(\bar{x}^t) + c^{t+1}(\bar{x}^{t+1}) > c^t(\bar{x}^t + \epsilon) + c^{t+1}(\bar{x}^{t+1} - \epsilon).$$

By adding the preceding inequality and Eq. (5.11), we get

$$-u_i^{t+1}(x_i^{t+1}-\epsilon)-u_i^t(x_i^t+\epsilon)+c^t(\bar{x}^t+\epsilon)+c^{t+1}(\bar{x}^{t+1}-\epsilon) < -u_i^{t+1}(x_i^{t+1})-u_i^t(x_i^t)+c^t(\bar{x}^t)+c^{t+1}(\bar{x}^{t+1}),$$

and thus

$$-u(y) + c(\bar{y}) < -u(x) + c(\bar{x}).$$

Based on the construction of $(y, \bar{y})$ and feasibility of $(x, \bar{x})$, we have $Ay - \bar{y} = Ax - \bar{x}$, $By = Bx \leq d$. From Eqs. (5.10) and (5.12), we have $\bar{y} \geq 0$, $0 \geq y \geq 0$. Thus the solution $(y, \bar{y})$ is feasible for problem (PW) and achieves a better objective function value than the optimal solution, which is a contradiction.

(b) $[Ax]_t \geq [Ax]_{t+1}$

We will use a similar argument as before, this time keeping the consumption the same, while shifting some generation from period $t+1$ to t. Formally, for some positive scalar ϵ, we construct a new solution $(y, \bar{y})$ with $\bar{y}^t = \bar{x}^t + \epsilon$ and $\bar{y}^{t+1} = \bar{x}^{t+1} - \epsilon$, and all other y_j^τ and $\bar{y}^\tau$ same as $(x, \bar{x})$.

We will use the choice of ϵ given by $\epsilon = \min\{\frac{\bar{x}^{t+1}-\bar{x}^t}{2}, \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau + S\}$. We first show that $\epsilon > 0$ and the solution constructed above is feasible. To prove $\epsilon > 0$, we assume the contrary, i.e., $\epsilon \leq 0$, which could only happen when

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \leq -S,$$

since $\bar{x}^{t+1} > \bar{x}^t$ by assumption. From feasibility of $(x, \bar{x})$, we have

$$\sum_{\tau=1}^{\beta}[Ax - \bar{x}]_\tau \geq -S, \tag{5.13}$$

for all $\beta < T$, including $\beta = t$ and therefore

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau = -S. \tag{5.14}$$

We use Eq. (5.13) with $\beta = t - 1$ and have

$$\sum_{\tau=1}^{t-1}[Ax - \bar{x}]_\tau = \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau - [Ax - \bar{x}]_t \geq -S.$$

Summing the previous two relations gives

$$[Ax - \bar{x}]_t \leq 0.$$

From the assumptions that $\bar{x}^t < \bar{x}^{t+1}$ and $[Ax]_t \geq [Ax]_{t+1}$, we have $[Ax - \bar{x}]_t > [Ax - \bar{x}]_{t+1}$, and thus

$$[Ax - \bar{x}]_{t+1} < 0.$$

This inequality and Eq. (5.14) imply that

$$\sum_{\tau=1}^{t+1}[Ax - \bar{x}]_\tau = \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau + [Ax - \bar{x}]_{t+1} < -S,$$

which violates Eq. (5.13) for $\beta = t + 1$ and is a contradiction. Hence we have $\epsilon > 0$.

We next verify that solution $(y, \bar{y})$ is feasible for problem (PW). The constraints $m \geq y \geq 0$, $\bar{y} \geq 0$ and $e'(Ay - \bar{y}) = 0$ can all be easily verified based on definition of $(y, \bar{y})$. We also have

$$-S \leq \sum_{\tau=1}^{\pi}[Ay - \bar{y}]_\tau = \sum_{\tau=1}^{\pi}[Ax - \bar{x}]_\tau \leq S,$$

for $\pi \neq t$ by definition of $(y, \bar{y})$, and hence we only need to show

$$-S \leq \sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau \leq S. \tag{5.15}$$

Since we are increasing production at period t, we have $\bar{y}^t > \bar{x}^t$ and

$$\sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau = \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau - \epsilon \leq S,$$

where we used Eq. (5.13) for $\beta = t$. To show $-S \leq \sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau$, based on definition of ϵ, we have

$$\sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau = \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau - \epsilon \geq \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau - (\sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau + S) = -S.$$

Thus Eq. (5.15) is satisfied and the solution $(y, \bar{y})$ is feasible.

We now show that $(y, \bar{y})$ achieves better objective function value than $(x, \bar{x})$. By using strict convexity of functions c^t and Lemma 5.2.6, we have

$$\bar{x}^t < \bar{x}^t + \epsilon \leq \bar{x}^{t+1} - \epsilon < \bar{x}^{t+1}, \quad c^t(\bar{x}^t) + c^{t+1}(\bar{x}^{t+1}) > c^t(\bar{x}^t + \epsilon) + c^{t+1}(\bar{x}^{t+1} - \epsilon).$$

Since all the other generation and consumption are the same as before, we have

$$-u(y) + c(\bar{y}) < -u(x) + c(\bar{x}).$$

Thus the solution $(y, \bar{y})$ achieves a strictly better objective function value.

To conclude this part of the proof, we have shown that in both of the possible scenarios, there exists a feasible solution $(y, \bar{y})$ with a better objective function value, which is a contradiction to the optimality of solution $(x, \bar{x})$ and thus the generations $\bar{x}^t$ is monotonically nonincreasing.

Monotonicity of $[Ax]_t$:

We now show monotonicity in $[Ax]_t$, also by contradiction. Assume there exists a period t with $[Ax]_t < [Ax]_{t+1}$. By using the definition of

$$[Ax]_t = \sum_{i=1}^{N} x_i^t,$$

and the fact that $x_i^t \geq 0$, we can infer that there exists a user i, with $x_i^t < x_i^{t+1}$. We use a similar argument of creating another feasible solution with better objective function value as before. This time we keep the generation schedule the same, while shift some consumption for user i from period $t+1$ to t. Formally, for some scalar ϵ with $\epsilon = \min\{\frac{x_i^t - x_i^{t+1}}{2}, S - \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau\}$, we construct a new solution $(y, \bar{y})$ with $y_i^t = x_i^t + \epsilon$ and $y_i^{t+1} = x_i^{t+1} - \epsilon$, and all other y_j^τ and $\bar{y}^\tau$ same as $(x, \bar{x})$. We first show that $\epsilon > 0$ and the solution $(y, \bar{y})$ is feasible.

To prove $\epsilon > 0$, we assume the contrary, i.e., $\epsilon \leq 0$, which could only happen when

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \geq S,$$

since $\bar{x}^{t+1} > \bar{x}^t$ by assumption. From feasibility of $(x, \bar{x})$, we have

$$\sum_{\tau=1}^{\beta}[Ax - \bar{x}]_\tau \leq S, \tag{5.16}$$

for all $\beta < T$, including $\beta = t$ and therefore

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau = S. \tag{5.17}$$

We use Eq. (5.16) with $\beta = t - 1$ and have

$$\sum_{\tau=1}^{t-1}[Ax - \bar{x}]_\tau = \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau - [Ax - \bar{x}]_t \leq S.$$

Summing the previous two relations gives

$$[Ax - \bar{x}]_t \geq 0.$$

From the assumptions that $\bar{x}^t > \bar{x}^{t+1}$ and $[Ax]_t < [Ax]_{t+1}$, we have $[Ax - \bar{x}]_t < [Ax - \bar{x}]_{t+1}$, and thus

$$[Ax - \bar{x}]_{t+1} > 0.$$

This inequality and Eq. (5.17) imply that

$$\sum_{\tau=1}^{t+1} [Ax - \bar{x}]_\tau = \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau + [Ax - \bar{x}]_{t+1} > S,$$

which violates Eq. (5.16) for $\beta = t + 1$ and is a contradiction. Hence we have $\epsilon > 0$.

We next verify that solution $(y, \bar{y})$ is feasible for problem (PW). The constraints $m \geq y \geq 0$, $\bar{y} \geq 0$ and $e'(Ay - \bar{y}) = 0$ can all be easily verified based on definition of $(y, \bar{y})$. We also have

$$-S \leq \sum_{\tau=1}^{\pi} [Ay - \bar{y}]_\tau = \sum_{\tau=1}^{\pi} [Ax - \bar{x}]_\tau \leq S,$$

for $\pi \neq t$ by definition of $(y, \bar{y})$, and hence we only need to show

$$-S \leq \sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau \leq S. \tag{5.18}$$

Since we are decreasing production at period t, we have $\bar{y}^t > \bar{x}^t$ and

$$\sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau = \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau + \epsilon \geq -S,$$

where we used Eq. (5.16) for $\beta = t$. To show $\sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau \leq S$, we use definition of ϵ and have

$$\sum_{\tau=1}^{t} [Ay - \bar{y}]_\tau = \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau + \epsilon \leq \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau + (S - \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau) = S.$$

Thus Eq. (5.18) is satisfied and the solution $(y, \bar{y})$ is feasible.

We now show that $(y, \bar{y})$ achieves better objective function value than $(x, \bar{x})$, for $x_i^t < x_i^t + \epsilon \leq x_i^{t+1} - \epsilon < x_i^{t+1}$. Then by concavity of function u and Lemma 5.2.6, we have

$$u_i^t(x_i^t + \epsilon) - u_i^t(x_i^t) \geq u_i^t(x_i^{t+1}) - u_i^t(x_i^{t+1} - \epsilon).$$

Assumption 9 guarantees

$$u_i^t(x_i^{t+1}) - u_i^t(x_i^{t+1} - \epsilon) = \int_{x_i^{t+1}-\epsilon}^{x_i^{t+1}} \nabla u_i^t(\tau) d\tau \geq \int_{x_i^{t+1}-\epsilon}^{x_i^{t+1}} \nabla u_i^{t+1}(\tau) d\tau = u_i^{t+1}(x_i^{t+1}) - u_i^{t+1}(x_i^{t+1} - \epsilon),$$

and at least one of these inequalities is strict.

Thus, we can combine these two inequalities and have

$$u_i^{t+1}(x_i^{t+1} - \epsilon) + u_i^t(x_i^t + \epsilon) > u_i^{t+1}(x_i^{t+1}) + u_i^t(x_i^t).$$

Since all the other generation and consumption are the same as before, we have

$$-u(y) + c(\bar{y}) < -u(x) + c(\bar{x}).$$

Thus the feasible solution $(y, \bar{y})$ achieves a strictly better objective function value than the optimal solution $(x, \bar{x})$, which is a contradiction. Therefore, we have shown that $[Ax]_t$ is monotone.

□

Lemma 5.2.7. ***(Monotonicity of Supply-Demand Mismatch)*** *Let $(x, \bar{x})$ be an optimal solution to problem (PW) with $S > 0$, then the supply-demand mismatch $[Ax - \bar{x}]_t$ is monotone nonincreasing in t, i.e.,*

$$[Ax - \bar{x}]_t \geq [Ax - \bar{x}]_{t+1}$$

for all $1 \leq t < T$.

Proof. We show this by contraction. Let t be an time with $[Ax - \bar{x}]_t < [Ax - \bar{x}]_{t+1}$.

By monotonicity of $[Ax]_t$ from Theorem 5.2.2, we have $[Ax]_t \geq [Ax]_{t+1}$, therefore $\bar{x}^t > \bar{x}^{t+1}$. We will construct another solution $(y, \bar{y})$ and show that it is feasible and achieves a better objective function value than $(x, \bar{x})$. This new solution is created by keeping the consumptions the same and shifting some generation from period $t+1$ to t. Formally, for some scalar ϵ with $\epsilon = \min\{\frac{\bar{x}^t - \bar{x}^{t+1}}{2}, S - \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau\}$, we construct a new solution $(y, \bar{y})$ with $\bar{y}^t = \bar{x}^t - \epsilon$ and $\bar{y}^{t+1} = \bar{x}^{t+1} + \epsilon$, and all other y_j^τ and $\bar{y}^\tau$ same as $(x, \bar{x})$. We first show that $\epsilon > 0$ and the solution $(y, \bar{y})$ is feasible.

To prove $\epsilon > 0$, we assume the contrary, i.e., $\epsilon \leq 0$, which could only happen when

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \geq S,$$

since $\bar{x}^t > \bar{x}^{t+1}$. From feasibility of $(x, \bar{x})$, we have

$$\sum_{\tau=1}^{\beta}[Ax - \bar{x}]_\tau \leq S, \tag{5.19}$$

for all $\beta < T$, including $\beta = t$ and therefore

$$\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau = S. \tag{5.20}$$

We use Eq. (5.19) with $\beta = t - 1$ and have

$$\sum_{\tau=1}^{t-1}[Ax - \bar{x}]_\tau = \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau - [Ax - \bar{x}]_t \leq S.$$

The previous two relations imply that

$$[Ax - \bar{x}]_t \geq 0.$$

Recall that $[Ax - \bar{x}]_t < [Ax - \bar{x}]_{t+1}$, and thus

$$[Ax - \bar{x}]_{t+1} > 0.$$

The above inequality and Eq. (5.20) imply that

$$\sum_{\tau=1}^{t+1}[Ax-\bar{x}]_\tau = \sum_{\tau=1}^{t}[Ax-\bar{x}]_\tau + [Ax-\bar{x}]_{t+1} > 0,$$

which violates Eq. (5.13) for $\beta = t+1$ and is a contradiction. Hence we have $\epsilon > 0$.

We next verify that solution $(y, \bar{y})$ is feasible for problem (PW). The constraints $m \geq y \geq 0$, $\bar{y} \geq 0$ and $e'(Ay - \bar{y}) = 0$ can all be easily verified based on definition of $(y, \bar{y})$. We also have

$$\sum_{\tau=1}^{\pi}[Ay-\bar{y}]_\tau = \sum_{\tau=1}^{\pi}[Ax-\bar{x}]_\tau,$$

for $\pi \neq t$ by definition of $(y, \bar{y})$, and hence we only need to show

$$-S \leq \sum_{\tau=1}^{t}[Ay-\bar{y}]_\tau \leq S. \tag{5.21}$$

Since we are decreasing production at period t, we have $\bar{y}^t > \bar{x}^t$ and

$$\sum_{\tau=1}^{t}[Ay-\bar{y}]_\tau = \sum_{\tau=1}^{t}[Ax-\bar{x}]_\tau + \epsilon \geq -S,$$

where we used Eq. (5.19) for $\beta = t$. To show $-S \leq \sum_{\tau=1}^{t}[Ay-\bar{y}]_\tau$, based on definition of ϵ, we have

$$\sum_{\tau=1}^{t}[Ay-\bar{y}]_\tau = \sum_{\tau=1}^{t}[Ax-\bar{x}]_\tau + \epsilon \leq \sum_{\tau=1}^{t}[Ax-\bar{x}]_\tau + (S - \sum_{\tau=1}^{t}[Ax-\bar{x}]_\tau) = S.$$

Thus Eq. (5.21) is satisfied and the solution $(y, \bar{y})$ is feasible.

We now show that $(y, \bar{y})$ achieves better objective function value than $(x, \bar{x})$. By using strict convexity of functions c^t and Lemma 5.2.6, we have

$$\bar{x}^t < \bar{x}^t + \epsilon \leq \bar{x}^{t+1} - \epsilon < \bar{x}^{t+1}, \quad c^t(\bar{x}^t) + c^{t+1}(\bar{x}^{t+1}) > c^t(\bar{x}^t + \epsilon) + c^{t+1}(\bar{x}^{t+1} - \epsilon).$$

Due to the fact that all the other generation and consumption are the same as before, we have

$$-u(y) + c(\bar{y}) < -u(x) + c(\bar{x}).$$

Thus the solution $(y, \bar{y})$ achieves a strictly better objective function value. We now arrive at a contradiction and conclude that $[Ax - \bar{x}]_t \geq [Ax - \bar{x}]_{t+1}$ for all $1 \leq t < T$. □

Theorem 5.2.3. *Let $(x, \bar{x})$ be an optimal solution of (PW), then we have*

$$\sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau \geq 0$$

for all $t \leq T - 1$.

Proof. Assume there exists a time $t \geq T-1$, with $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau < 0$. From Lemma 5.2.7, we have at optimal solution $(x, \bar{x})$ of problem (PW), $[Ax - \bar{x}]_\tau$ is monotonically nonincreasing in τ. Therefore, we have $[Ax - \bar{x}]_t < 0$ and $[Ax - \bar{x}]_\tau < 0$ for all $\tau \geq t$. This implies that

$$\sum_{\tau=1}^{T} [Ax - \bar{x}]_\tau = e'(Ax - \bar{x}) < 0,$$

which contradicts with the constraint $e'(Ax - \bar{x}) = 0$. Thus we arrive at a contradiction and conclude that $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \geq 0$ for all $t \leq T - 1$. □

The next lemma states that any constraints what are not binding can be dropped from the optimization problem.

Lemma 5.2.8. *Consider the following two optimization problems,*

$$\begin{aligned} \min_x \quad & f(x), \\ s.t. \quad & Hx \leq h, Gx \leq g. \end{aligned} \qquad\qquad \begin{aligned} \min_y \quad & f(y), \\ s.t. \quad & Hy \leq h, \end{aligned}$$

where f is strictly convex. Suppose the optimal solutions for these problems exists, denoted by x^*, y^* respectively. If $Gx^* < g$, then the two problems have the same optimal solution.

Proof. We first show that the two problems have the same optimal objective function value. Since the problem on the right is more constrained, we have

$$f(x^*) \geq f(y^*).$$

By optimality condition of x^*, we have

$$x^* \in \underset{Hx \leq h}{\operatorname{argmin}}\, f(x) + (z^*)'(Gx - g),$$

where z^* is the optimal dual variable associated with $Gx \leq g$. In view of $Gx^* < g$, by complementary slackness condition, we have $z^* = 0$, and thus

$$x^* \in \underset{Hx \leq h}{\operatorname{argmin}}\, f(x),$$

i.e.,

$$f(x^*) \leq f(y^*).$$

Hence we have

$$f(x^*) = f(y^*).$$

We can now use strict convexity of function f and conclude that the two optimal solutions are the same.

□

Theorem 5.2.3 implies that constraint $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \geq -S$ is not active at optimal solution of (PW) and therefore by the previous lemma and strict convexity of generation cost function [c.f. Assumption 8], this constraint can be dropped from the problem formulation. Similarly, Lemma 5.2.2 and Assumption 10 implies that the constraint $\bar{x} \geq 0$ can be ignored from the problem formulation. For the rest of the chapter, we focus on the following problem

$$\begin{aligned}
\min_{x,\bar{x}} \quad & -u(x) + c(\bar{x}), && \text{(PW2)} \\
\text{s.t.} \quad & \sum_{\tau=1}^{t} [Ax - \bar{x}]_\tau \leq S, \quad \text{for } t = 1, \ldots, T-1, \\
& e'(Ax - \bar{x}) = 0, \\
& Bx \leq d, \\
& m \geq x \geq 0.
\end{aligned}$$

We note that the properties we derived thus far for the optimal solution of (PW) carry over to (PW2) by Lemma 5.2.8. We can now reinterpret the physical meaning of S. For a storage of size S starting fully charged, the level of energy remaining in storage at time t is given by $S - \sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau$. The constraint $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \leq S$ guarantees that the energy level will stay nonnegative. Theorem 5.2.3 implies that the storage level will not exceed S at the optimal solution. Hence the optimal solution can be implemented with a storage of size S starting fully charged. Constraint $e'(Ax - \bar{x}) = 0$ also guarantees that at the end of T periods, the storage level will come back to S and be ready for the next T period cycle (assuming the storage is friction free).

5.2.3 Distributed Implementation

As the competitive equilibrium framework before, this problem (PW) can also be decomposed into the following two problems, one for the generator (PG), one for the demand/consumers (PD), where the total market balancing constraint and the storage size constraint are implemented via price.

$$\min_{\bar{x} \geq 0} \quad c(\bar{x}) - p'\bar{x}, \qquad \text{(PG)}$$

$$\begin{aligned} \min_{x} \quad & p'x - u(x), \\ \text{s.t.} \quad & Bx \le d, \\ & m \ge x \ge 0. \end{aligned} \tag{PD}$$

The following theorem states formally the equivalence between problems (PG), (PD) and (PW). This is the counterpart of Theorem 4.2.1, when storage is introduced.

Theorem 5.2.4. *Any optimal solution of problem (PW2) can be represented by problems (PD) and (PG) with $p = E\Delta p + p_T$, where Δp and p_T are the optimal dual variables for problem (PW2). Conversely, any solution pair $(x, \bar{x})$ for problem (PD) and (PW) where $\sum_{\tau=1}^{t}[Ax - \bar{x}]_\tau \le S$ for $t = 1, \ldots, T-1$, $e'(Ax - \bar{x}) = 0$ and the complementary slackness conditions for $\Delta p' E'(Ax - \bar{x} - S) = 0$, is satisfied where $[\Delta p]_t = p_t - p_{t+1}$, is an optimal solution for problem (PW2).*

Proof. This can be shown based on the necessary and sufficient optimality conditions. Using an argument similar to Theorem 4.2.1. □

The preceding theorem gives us an equivalent form for problem (PW2), which leads to the following distributed implementation of problem (PW2).

Lemma 5.2.9. *Let function*

$$q(p_T, \Delta p) = \max_{\{x,\bar{x}\}\ge 0, m\ge x, Bx\le d} \quad u(x) - c(\bar{x}) - p_T e'(Ax - \bar{x}) - \Delta p'[E'(Ax - \bar{x}) - S].$$

Function $q(p_T, \Delta p)$ is convex in p_T, Δp and if the variable pair $(x^*, \bar{x}^*)$ attains the maximum, then the vector $(e'(\bar{x}^* - Ax^*), E'(Ax^* - \bar{x}^*) - S)$ is a subgradient for function $q(p_T, \Delta p)$.

Proof. The function $q(p_T, \Delta p)$ is pointwise maximum of a convex function in p_T and Δp, therefore is convex in both of the variables.

We use the definition of subgradient to show the second part of the lemma, for some $j_T, \Delta j$ in $\mathbb{R}$, $\mathbb{R}^{T-1}$ respectively.

$$\begin{aligned}
q(j_T, \Delta j) = & \max_{\{x,\bar{x}\}\geq 0, m\geq x, Bx\leq d} \quad u(x) - c(\bar{x}) - j_T e'(Ax - \bar{x}) - \Delta j'[E'(Ax - \bar{x}) - S] \\
& \geq u(x^*) - c(\bar{x}^*) - j_T e'(Ax^* - \bar{x}^*) - \Delta j'[E'(Ax^* - \bar{x}^*) - S] \\
& = u(x^*) - c(\bar{x}^*) - p_T e'(Ax^* - \bar{x}^*) - \Delta p'[E'(Ax^* - \bar{x}^*) - S] \\
& - (j_T - p_T)e'(Ax^* - \bar{x}^*) - (\Delta j - \Delta p)'[E'(Ax^* - \bar{x}^*) - S], \\
& = q(p_t, \Delta p) - (j_T - p_T)e'(Ax^* - \bar{x}^*) - (\Delta j - \Delta p)'[E'(Ax^* - \bar{x}^*) - S],
\end{aligned}$$

where the inequality follows from the maximization operation. Thus the preceding relation implies that the vector $(e'(\bar{x}^* - Ax^*), S - E'(Ax^* - \bar{x}^*))$ is a subgradient for function $q(p_T, \Delta p)$. □

By using subgradient descent algorithm on problem (PW2), we have the following iteration, where we use notation $x(\tau)$ to denote the value of variable x at iteration τ.

Distributed Implementaion for Price Fluctuation Penalized Social Welfare:

A Initialization: the central planner chooses some arbitrary price vector $p_T(0)$ in $\mathbb{R}$ and $\Delta p(0)$ in $\mathbb{R}^{T-1}$.

B At iteration k:

  a Consumers solve problem (PD) and suppliers solve problem (PG) independently using price vector $p_T(\tau), \Delta p(\tau)$ to obtain quantity demanded vector $x(\tau)$ and quantity supplied $\bar{x}(\tau)$.

  b The central planner updates price as

$$p_T(\tau + 1) = p_T(\tau) - \theta(\tau)e'(\bar{x}^* - Ax^*),$$

$$\Delta p(\tau + 1) = \Delta p(\tau) - \theta(\tau)[S - E'(Ax^* - \bar{x}^*)],$$

  where $\theta(\tau)$ is some positive stepsize.

The above algorithm is subgradient descent of a convex problem, therefore converges for $s(k)$ sufficiently small.

Remarks: The variable $p_T e + E\Delta p$ is the market price vector (note in case of excess supply, the supplier receives same compensation for supplying to consumers and the storage and in case of excess demand, the consumer pays the same price for using power from storage or the generator). When this algorithm converges, the solution satisfies total market clearing constraint and storage constraint. Problem (PD) and (PG) are the same as their counterparts in previous chapter without fluctuation penalty. The only difference is that price adjustment in the algorithm now involves storage size S.

We also note that when implementing step B.b, only aggregate quantity demanded for each time period is necessary, instead of individual consumer choices of x_i^t, which can be used to protect consumer privacy if a third party who can aggregate the demand is present, such as distribution center.

5.3 Social Welfare and Generation Fluctuation Analysis

We next analyze the effect of storage on social welfare and price, generation fluctuations, i.e., $u(x) - c(\bar{x})$ and $\Delta\bar{x}$, Δp as a function of the nonnegative scalar S in problem (PW2). Section 5.3.1, we show that social welfare is concavely nondecreasing in storage size. In Section 5.3.2, we focus on the case where demand is inelastic and we give an explicit characterization of the optimal storage access policy and production level. In Section 5.3.3, a two period case is analyzed, where we relate the production variation with concavity of the utility functions and show that a more concave utility function results in less production fluctuation.

5.3.1 Social Welfare Analysis

In this section, we study the properties of the social welfare function $u(x) - c(\bar{x})$ as a function of the parameter S for the general problem formulation with L_q norm, i.e.,

$$\begin{aligned} \max_{x,\bar{x}} \quad & u(x) - c(\bar{x}), \\ \text{s.t.} \quad & \left|\left|E'(Ax - \bar{x})\right|\right|_q \leq S, \\ & Bx \leq d, \\ & x \geq 0, \quad \bar{x} \geq 0, \\ & e'(Ax - \bar{x}) = 0, \end{aligned} \tag{PWq}$$

and show that the social welfare is a concave nondecreasing function of S. This is a more general problem than (PW2) and we have shown in the previous section, when L_∞ norm is used, the two are equivalent.

We will need to use the following lemma to show the concavity.

Lemma 5.3.1. *Consider the following problem*

$$\begin{aligned} \min_{y} \quad & f(y), \\ s.t. \quad & g_i(y) \leq b_i, \quad i = 1, \ldots, m \\ & Ly = h, \end{aligned}$$

where each of the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i is convex. Let $f^*(\bar{b})$ denote the optimal function value where vector $b = [b_i]_i = \bar{b}$, then we have for α in $[0, 1]$

$$f^*\left(\alpha b_1 + (1 - \alpha) b_2\right) \leq \alpha f^*(b_1) + (1 - \alpha) f^*(b_2),$$

i.e., function f^* is convex.

Proof. We let y_1 denote an optimal solution associated with $b = b_1$, i.e., $f^*(b_1) = f(y_1)$. Similarly y_2 denote an optimal solution associated with $b = b_2$. We proceed by constructing a feasible solution $\alpha y_1 + (1 - \alpha) y_2$ for $b = \alpha b_1 + (1 - \alpha) b_2$ and then use optimality of $f^*(\alpha b_1 + (1-\alpha)b_2)$.

We first show that $\alpha y_1 + (1-\alpha)y_2$ is a feasible solution. Based on feasibility of solutions y_1 and y_2, we have

$$g_i(y_1) \leq b_1, \quad g_i(y_2) \leq b_2, \quad Ly_1 = h, \quad Ly_2 = h$$

Hence, by using convexity of the functions g_i, we have

$$g_i(\alpha y_1 + (1-\alpha)y_2) \leq \alpha g_i(y_1) + (1-\alpha)g_i(y_2) \leq \alpha b_1 + (1-\alpha)b_2.$$

By linearity of the matrix vector product, we have

$$L(\alpha y_1 + (1-\alpha)y_2) = \alpha Ly_1 + (1-\alpha)Ly_2 = \alpha h + (1-\alpha)h = h.$$

Therefore $\alpha y_1 + (1-\alpha)y_2$ is a feasible solution for $b = \alpha b_1 + (1-\alpha)b_2$.

From optimality of the value $f^*(\alpha b_1 + (1-\alpha)b_2)$, we have

$$f^*(\alpha b_1+(1-\alpha)b_2) \leq f(\alpha y_1+(1-\alpha)y_2) \leq \alpha f(y_1)+(1-\alpha)f(y_2) = \alpha f^*(b_1)+(1-\alpha)f^*(b_2),$$

where the second inequality follows from convexity of function f and the equality is based on definitions of y_1 and y_2. □

The following lemma will be necessary to show that our constraint involving L_q norm in (PWq) has a convex function on the left hand side.

Lemma 5.3.2. *The L_q norm i.e., $||y||_q$ for y in $\mathbb{R}^n$, is convex in y for $q \geq 1$.*

Proof. From definition of L_q norm and algebraic manipulation, we have

$$||\alpha y||_q = \left(\sum_{i=1}^{n} \alpha^q y_i^q\right)^{\frac{1}{q}} = \alpha \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}} = \alpha \, ||y||_q \, .$$

Thus, by using Minkowski inequality, we have

$$||\alpha y_1 + (1-\alpha)y_2||_q \leq ||\alpha y_1||_q + ||(1-\alpha)y_2||_q = \alpha \, ||y_1||_q + (1-\alpha) \, ||y_2||_q \, ,$$

where the equality follows from the preceding relation. □

Theorem 5.3.1. *The optimal social welfare $u(x) - c(\bar{x})$ in problem (PW2) as a function of S, denoted by $W(S)$, is concave and monotonically nondecreasing.*

Proof. We first establish the concave part using results from Lemma 5.3.1. We let $y = [x', \bar{x}']'$, $f(y) = -u(x) + c(\bar{x})$, $Ly = h$ represent the linear equality constraint and define g_i for each of the inequality constraint. We first show that all the inequality constraints are convex. This claim is clearly true for the linear inequality constraints. We then focus on the left hand side of the constraint

$$||E'(Ax - \bar{x})||_q \leq S.$$

The function $||E'(Ax - \bar{x})||_q$ is a composition of a linear function and the norm operator, which is convex by the previous lemma, and therefore the composition function is also convex. We can now apply Lemma 5.3.1 by letting the right hand side of the inequality take values $b_1 = [S_1, d', 0', 0']'$ and $b_2 = [S_2, d', 0', 0']'$ with $\alpha b_1 + (1-\alpha)b_2 = [\alpha S_1 + (1-\alpha)S_2, d', 0', 0']$ for α in $[0, 1]$ and Lemma 5.3.1 shows that the optimal social welfare satisfies

$$-W(\alpha S_1 + (1-\alpha)S_2) \leq -\alpha W(S_1) - (1-\alpha)W(S_2).$$

We can multiply both sides by -1 and establish concavity of function $W(S)$.

We next show that function $W(S)$ is an increasing function in S. We note that the feasible set is expanding as S increases and therefore the maximal objective function value is nondecreasing in S. □

The previous theorem suggests that social welfare is non-decreasing in S and the increasing in social welfare slows down, as S grows bigger. The result holds for any general L_q norm. In case of L_∞ norm, the scalar S can be viewed as the storage size as shown in the previous section. Thus far we have assumed the storage is freely available, however the cost of storage in reality is a convex increasing function in size. When we combine both the social welfare and the cost of storage together, the previous theorem implies the existence of an optimal storage size (under the assumption of a perfectly frictionless storage). The value of such storage size depends on the actual functional form of the cost of storage, consumer utility and generator cost functions.

5.3.2 Inelastic Demand Analysis

For this part of analysis, we assume the individual demands are inelastic (nonnegative and satisfying Assumption 9), denoted by x and aggregate demand $y = Ax$, [1] and analyze the effect of storage on generation and price fluctuation. [2] Since the demand is inelastic the consumer utility $u(x)$ and consumption vector x remain constant as we vary storage size S. We can then simply problem (PW2) to

$$\begin{aligned} \min_{\bar{x}} \quad & \sum_{t=1}^{T} c^t(\bar{x}^t), && \text{(PC)} \\ \text{s.t.} \quad & \sum_{q=1}^{t} [y - \bar{x}]_q \leq S, \quad \text{for } t = 1, \dots, T-1, \\ & e'(y - \bar{x}) = 0. \end{aligned}$$

To derive the optimal generation schedule, we will rely on the following lemma, which states an important property of convex functions.

Lemma 5.3.3. *For any convex function $f : \mathbb{R}^n \to \mathbb{R}$, the following relation holds for integer any $M \geq 2$,*

$$Mf\left(\frac{\sum_{j=1}^{M} x_j}{M}\right) \leq \sum_{j=1}^{M} f(x_j).$$

Proof. We prove by induction. For $M = 2$ the desired result is equivalent to

$$f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) \leq \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2),$$

[1] Since the individual demands satisfy the monotonicity assumption, Theorem 5.2.2 states that the aggregate demand is monotone, i.e. $y^t \geq y^{t+1}$ for all t.

[2] From previous section, we have that as the size of storage increases, both the price and generation fluctuation decrease and the total demand would fluctuate more if the demand is elastic. Therefore, the inelastic case considered in this section provides a best case bound on how much improvement in price and generation fluctuation the introduction of storage can bring.

which holds by convexity of function f.

Now we assume the desired relation holds for M and prove it for $M+1$. We observe that

$$\frac{\sum_{j=1}^{M+1} x_j}{M+1} = \frac{\sum_{j=1}^{M} x_j}{M+1} + \frac{x_{M+1}}{M+1} = \frac{M}{M+1}\frac{\sum_{j=1}^{M} x_j}{M} + \frac{x_{M+1}}{M+1},$$

where the second equality follows by multiplying the first term by $1 = \frac{M}{M}$. Therefore, by convexity of function f, we have

$$f\left(\frac{\sum_{j=1}^{M+1} x_j}{M+1}\right) \leq \frac{M}{M+1} f\left(\frac{\sum_{j=1}^{M} x_j}{M}\right) + \frac{1}{M+1} f\left(x_{M+1}\right).$$

By the induction hypothesis, we have

$$f\left(\frac{\sum_{j=1}^{M} x_j}{M}\right) \leq \frac{1}{M}\sum_{j=1}^{M} f(x_j).$$

Therefore, we can combine the preceding two relations and obtain

$$f\left(\frac{\sum_{j=1}^{M+1} x_j}{M+1}\right) \leq \frac{1}{M+1}\sum_{j=1}^{M} f(x_j) + \frac{1}{M+1} f\left(x_{M+1}\right),$$

which by multiplying both sides with factor $M+1$ completes the induction step. □

The intuition from the preceding lemma suggests the minimal cost generation is when all periods are close to the mean. We next formally derive the optimal production schedule based on this intuition. For notational convenience, we denote by

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y^t \tag{5.22}$$

the average aggregate demand over time. We let τ to denote the largest time index with $y^t \geq \bar{y}$ and scalar

$$\theta = \sum_{t \leq \tau} [y^t - \bar{y}], \tag{5.23}$$

represents the one sided total mean deviation. We note that by definition of $\bar{y}$, we have

$\sum_{t \notin \Phi}[y^t - \bar{y}] = -\theta$. We can now establish the following theorem, which characterizes the optimal generation level.

Theorem 5.3.2. *For problem (PW2) with any time invariant strictly convex differentiable cost function c^t, storage size $S \geq 0$, with scalars $\bar{y}$, θ, τ defined as above. Then when $S \geq \theta$, the unique optimal solution $\bar{x}$ is given as*

$$\bar{x}^t = \bar{y}.$$

Otherwise, we let ξ denote the smallest index with

$$\sum_{t=1}^{\xi} y^t - S > \xi y^{\xi+1}, \tag{5.24}$$

and β denote the largest index with

$$\sum_{t=\beta+1}^{T} y^t + S < (T - \beta) y^{\beta}. \tag{5.25}$$

The optimal solution is given by

$$\bar{x}^t = \begin{cases} \frac{\sum_{t=1}^{\xi} y^t - S}{\xi} & \text{for} \quad t \leq \xi, \\ y^t & \text{for} \quad \xi < t \leq \beta, \\ \frac{\sum_{t=\beta+1}^{T} y^t + S}{T - \beta} & \text{for} \quad t > \beta. \end{cases} \tag{5.26}$$

Proof. Due to strict convexity of c^t, the optimal solution is unique. Since the cost functions are convex, the sufficient condition for $\bar{x}$ to be an optimal solution of problem (PC) is

$$\nabla c^t(\bar{x}^t) - \sum_{q=t}^{T} \Delta p - p_T = 0, \tag{5.27}$$

$$\Delta p \geq 0, \quad \sum_{q=1}^{t} [y - \bar{x}]_q \leq S, \quad e'(\bar{x} - y) = 0, \tag{5.28}$$

$$\Delta p' \left(\sum_{q=1}^{t} [y - \bar{x}]_q - S \right) = 0, \tag{5.29}$$

for some dual variables $[\Delta p^t]_t = \Delta p$ and p_T. The dual variable Δp^t is associated with constraints $\sum_{q=1}^t [y - \bar{x}]_q \leq S$ and p_T with the constraint $e'(y - \bar{x})$. The three relations above represent stationary condition, primal, dual feasibility and complementary slackness condition respectively. We next verify solution $\bar{x}^t$, for both $S \geq \theta$ and $S < \theta$, defined in the theorem satisfies all these conditions.

A. When $S \geq \theta$:

We first verify that when $S \geq \theta$, the solution of $\bar{x}^t = \bar{y}$ satisfies the above condition. We will use the dual variables $p_T = \nabla c^t(\bar{y})$, $\Delta p = 0$. All the conditions other than $\sum_{q=1}^t [y - \bar{x}]_q \leq S$ are clearly satisfied. We now verify that $\sum_{q=1}^t [y - \bar{x}]_q \leq S$ is also satisfied. In view of definition of θ, we have for all $t \leq \tau$,

$$0 \leq \sum_{q=1}^{t} y^q - \bar{x}^q \leq \theta \leq S,$$

and $t > \tau$,

$$0 \geq \sum_{q=\tau+1}^{t} y^q - \bar{x}^q \geq -\theta.$$

Also by definition of τ, we have and

$$\sum_{q=1}^{\tau} y^q - \bar{x}^q = \theta.$$

The preceding two relations combined yield for $t > \tau$,

$$S \geq \theta \geq \sum_{q=1}^{t} y^q - \bar{x}^q \geq 0.$$

Thus the constraint $\sum_{q=1}^t y^q - \bar{x}^q \leq S$ is satisfied by $\bar{x}^t = \bar{y}$ for all t.

B. When $S < \theta$:

We now analyze the case when $S < \theta$. We let the dual variables be defined as

$$p_T = \nabla c^t(x_T), \quad \Delta p^t = \nabla c^t(\bar{x}^t) - \nabla c^t(\bar{x}^{t+1}). \tag{5.30}$$

(a) Scalars ξ and β are well defined and $\xi \leq \beta$:

We start by showing that τ satisfies Eqs. (5.24) and (5.25), and therefore scalars ξ and β are well defined. We will then establish that $\xi \leq \beta$. Since $S < \theta$, we have $\theta > 0$ and $\tau < T$, therefore $y^\tau \geq \bar{y} > y^{\tau+1}$. By definition of τ the following relation holds,

$$\sum_{t=1}^{\tau} y^t - \tau y^{q+1} = \sum_{t=1}^{\tau}(y^t - y^{q+1}) \geq \sum_{t=1}^{\tau}(y^t - \bar{y}) = \theta > S,$$

hence τ satisfies Eq. (5.24). Similarly, τ satisfies

$$(T-\tau)y^\tau - \sum_{t=\tau+1}^{T} y^t = \sum_{t=\tau+1}^{T}(y^\tau - y^t) \geq \sum_{t=\tau+1}^{T}(\bar{y} - y^t) = \theta > S,$$

where the last equality follows from the fact that $\theta = \sum_{t\leq\tau}[y^t - \bar{y}]$ and $\sum_{t\leq T}[y^t - \bar{y}] = 0$ [c.f. Eqs. (5.22), (5.23)]. Thus, the indices ξ and β are well defined and satisfies $\xi \leq \tau \leq \beta$.

We have thus far proven that the solution proposed in the theorem is well defined.

(b) The solution $\bar{x}$ and dual variables defined in Eq. (5.30) satisfy stationary condition in Eq. (5.27).

Based on Eq. (5.30), Eq. (5.27) is clearly satisfied.

(c) The solution $\bar{x}$ and dual variables defined in Eq. (5.30) is feasible, i.e., satisfies Eq. (5.28).

We first verify $\Delta p \geq 0$. Based on definition of $\bar{x}$, we have $\bar{x}^t = \bar{x}^{t+1}$ for $t < \xi$ and $t > \beta$. Therefore based on definition of Δp^t, we have

$$\Delta p^t = 0, \quad \text{for } t < \xi \text{ or } t > \beta. \tag{5.31}$$

By monotonicity of y^t and Eq. (5.26), for $\xi < t \leq \beta - 1$, we have $\bar{x}^t < \bar{x}^{t+1}$

and hence $\Delta p^t \geq 0$. For $t = \xi$, from definition of Δ^t, we have

$$\Delta p^\xi = \nabla c^t \left(\frac{\sum_{t=1}^{\xi} y^t - S}{\xi} \right) - \nabla c^t(\bar{x}^{t+1}).$$

By definition of ξ, we have $\frac{\sum_{t=1}^{\xi} y^t - S}{\xi} > y^{\xi+1}$ and thus by strict convexity of function c^t, we have $\Delta p^\xi > 0$. Similarly, using definition of β, we have $\Delta p^\beta > 0$. Hence, we have $\Delta p = 0$.

We next verify that the constraint $\sum_{q=1}^{t} [y - \bar{x}]_q \leq S$ is satisfied. From definition of ξ, we have

$$\sum_{q=1}^{\xi} [y^q - \bar{x}^q] = \sum_{t=1}^{\xi} y^t - \left(\sum_{t=1}^{\xi} y^t - S \right) = S. \tag{5.32}$$

We next show that $y^t - \bar{x}^t \geq 0$ and thus $\sum_{q=1}^{t} [y - \bar{x}]_q$ is monotonically increasing in t and does not exceed S.

From definition of ξ, we have

$$\sum_{t=1}^{\xi-1} y^t - S \leq (\xi - 1) y^\xi.$$

We can add the scalar y^ξ to both side and have

$$\sum_{t=1}^{\xi} y^t - S \leq \xi y^\xi.$$

Therefore, by Eq. (5.26), we have

$$y^\xi - \bar{x}^\xi = \frac{1}{\xi} \left(\xi y^\xi - \sum_{t=1}^{\xi} y^t + S \right) \geq 0.$$

Since $\bar{x}^t = \bar{x}^\xi$ for all $t \leq \xi$, by monotonicity of y^t in t, we have

$$y^t - \bar{x}^t = y^t - \bar{x}^\xi \geq y^\xi - \bar{x}^\xi \geq 0.$$

Thus, we have shown that $\sum_{q=1}^{t}[y-\bar{x}]_q \le S$ for $t \le \xi$. For $\xi < t \le \beta$, we have $\bar{x}^t = y^t$ and therefore

$$\sum_{q=1}^{t}[y-\bar{x}]_q = \sum_{q=1}^{\xi}[y-\bar{x}]_q = S, \quad \text{for } \xi < t \le \beta, \tag{5.33}$$

where the last equality follows from Eq. (5.32). We next show that for $t > \beta$ the constraint is also satisfied, by showing that $y^t - \bar{x}^t \le 0$.

From the way β is defined, we have

$$\sum_{t=\beta+2}^{T} y^t + S \ge (T-\beta-1)y^{\beta+1},$$

by adding $y^{\beta+1}$ to both sides, we have

$$\sum_{t=\beta+1}^{T} y^t + S \ge (T-\beta)y^{\beta+1}.$$

By Eq. (5.26), we have

$$y^{\beta+1} - \bar{x}^{\beta+1} = \frac{1}{T-\beta}\left[(T-\beta)y^{\beta+1} - \sum_{t=\beta+1}^{T} y^t - S\right] \le 0,$$

where we used the inequality above to derive the non-positivity.

Hence, for all $t > \beta$, by monotonicity of y^t and the fact that $\bar{x}^t = \bar{x}^{\beta+1}$, we have

$$y^t - \bar{x}^t \ge y^{\beta+1} - \bar{x}^t = y^{\beta+1} - \bar{x}^{\beta+1} \le 0.$$

We combine this relation with Eq. (5.33) and have

$$\sum_{q=1}^{t}[y-\bar{x}]_q \le \sum_{q=1}^{\xi}[y-\bar{x}]_q = S.$$

Hence, we have shown that for all $1 \le t < T$, $\sum_{q=1}^{t}[y-\bar{x}]_q \le S$.

We now move on to verify $e'(\bar{x} - y)$. We definition of $\bar{x}$ and have

$$\sum_{t=1}^{T} \bar{x}^t = \xi \frac{\sum_{t=1}^{\xi} y^t - S}{\xi} + \sum_{t=\xi+1}^{\beta} y^t + (T - \beta) \frac{\sum_{t=\beta+1}^{T} y^t + S}{T - \beta} = \sum_{t=1}^{T} y^t.$$

Thus, all conditions in Eq. (5.28) are satisfied.

(d) The solution $\bar{x}$ and dual variables defined in Eq. (5.30) satisfies complementary slackness condition in Eq. (5.29).
From Eqs. (5.31) and (5.33), we have either $\Delta p^t = 0$ or $\sum_{q=1}^{t} [y - \bar{x}]_q = 0$. Therefore complementary slackness condition is satisfied.

Therefore all the optimality conditions are satisfied, $\bar{x}$ is an optimal solution for $S < \theta$.

We can now combine the two cases and conclude that $\bar{x}$ given by Eq. (5.26) is optimal.

□

The above theorem characterizes an interesting phase transition in storage size. Once storage size S reaches the level of one sided total mean deviation θ, there are no additional benefits. On the other hand, when $S < \theta$, meaning the quantity demanded changes a lot over time, then the storage is used to smooth out fluctuation by targeting very large and very small [cf. Eqs. (5.24) and (5.25)]. The effect on each production period is linear. When the demand across time does not change significantly, we have that $\xi = \beta$ and there will be only one drop in price as well as production level, whereas when the demand changes drastically over time, then the for each period between ξ and β there may be production level changes.

Remark: The inelastic demand can be realized when the user utility functions have very large $\nabla^2 u(x)$.

5.3.3 Elastic Demand Two Period Analysis

In this section, we consider the scenario where demand is elastic and analyze the effect of elasticity, measured by the second derivative of the consumer utility function, on generation fluctuation for a given storage size. We assume the cost function is quadratic given by $c(\bar{x}) = \frac{k}{2}\bar{x}^2$. We will focus on a two period example with one representative consumption agent, i.e., $T = 2, N = 1$ to illustrate that more concave utility functions induce less fluctuations. Thus the inelastic demand (can be viewed as utility function with infinity concavity) gives the least fluctuation in production level, i.e., the results in the previous section serves as a best case bound. For notational simplicity, we will drop the index i. The social welfare maximizing problem can be written as following,

$$\begin{aligned}
\min_{x^t,\bar{x}^t} \quad & -u^1(x^1) - u^2(x^2) + \frac{k}{2}(\bar{x}^1)^2 + \frac{k}{2}(\bar{x}^2)^2 \\
\text{s.t.} \quad & x^1 - \bar{x}^1 \leq S, \quad x^1 + x^2 = \bar{x}^1 + \bar{x}^2, \\
& x^1 + x^2 \leq d, \\
& 0 \leq x^t \leq m, \quad 0 \leq \bar{x}^t, \qquad t = 1, 2.
\end{aligned} \tag{5.34}$$

The first period is assumed to be the preferred time slot. To model the fact that utility derived from a later period is discounted, we assume the utility functions in two periods u^1, u^2 satisfy

$$u^2(x) = \alpha u^1(x),$$

where $\alpha \leq 1$ is the discounting factor. We consider two set of twice differentiable concave utility functions $u, \alpha u$ and $g, \alpha g$ with

$$0 \geq \nabla^2 u(x) > \nabla^2 g(y), \tag{5.35}$$

for any x, y in $[0, m]$. We denote the optimal solution associated with functions $u, \alpha u$ and $g, \alpha g$ by $(x^t(S), \bar{x}^t(S))$ and $(y^t(S), \bar{y}^t(S))$ respectively. To normalize the utility function choices, we assume when $S = 0$, the optimal solution satisfies

$$x^1(0) + x^2(0) = y^1(0) + y^2(0). \tag{5.36}$$

To isolate the effect of concavity, we will focus on the interior point case where

$$kd > (1+\alpha)\nabla u(0), \quad kd > (1+\alpha)\nabla g(0) \tag{5.37}$$

and $0 < x^t, y^t < m$.

Theorem 5.3.3. *Consider the two twice differentiable concave utility functions u and g defined as above, the optimal solutions $(x^t(0), \bar{x}^t(0))$ and $(y^t(0), \bar{y}^t(0))$ associated with $S = 0$ satisfy*

$$\bar{x}^1(0) - \bar{x}^2(0) \geq \bar{y}^1(0) - \bar{y}^2(0).$$

Proof. Theorem 5.2.2 states that $\bar{x}^1(0) \geq \bar{x}^2(0)$ and $\bar{y}^1(0) \geq \bar{y}^2(0)$. By using the fact that $x^1(0) + x^2 = y^1 + y^2(0)$, the desired result can be written equivalently as

$$\bar{x}^1(0) \geq \bar{y}^1(0).$$

In the following proof, we establish this relation by contradiction. Assume $\bar{x}^1(0) < \bar{y}^1(0)$. This assumption and monotonicity of $\bar{x}^t$ implies that

$$\bar{y}^2(0) < \bar{x}^2(0) \leq \bar{x}^1(0) < \bar{y}^1(0).$$

With $S = 0$, we also have $x^t = \bar{x}^t$. The previous inequality also implies that

$$y^2(0) < x^2(0) \leq x^1(0) < y^1(0) \tag{5.38}$$

Eq. (5.37) and Lemma 5.2.3 implies that the constraint $x^1(0) + x^2(0) \leq d$ is not tight at either optimal solutions. By using complementary slackness condition, we can simplify (FOC) condition from Lemma 5.2.1 to

$$-\nabla u(x^1(0)) + \Delta p + p_2 = 0, \qquad k\bar{x}^1(0) - \Delta p - p_2 = 0, \tag{5.39}$$
$$-\alpha \nabla u(x^2(0)) + p_2 = 0, \qquad k\bar{x}^2(0) - p_2 = 0.$$

We use the condition $\bar{x}^t = x^t$ one more time to rewrite the previous equations as

$$\nabla u(x^1(0)) = kx^1(0), \quad \alpha \nabla u(x^2(0)) = kx^2(0).$$

Similarly, the optimality conditions for $(y^t, \bar{y}^t)$ imply that

$$\nabla g(y^1(0)) = ky^1(0), \quad \alpha \nabla g(y^2(0)) = ky^2(0).$$

These two relations imply that

$$\nabla u(x^1(0)) - \nabla u(x^2(0)) = kx^1(0) - \frac{k}{\alpha}x^2(0), \quad \nabla g(y^1(0)) - \nabla g(y^2(0)) = ky^1(0) - \frac{k}{\alpha}y^2(0).$$

By Eq. (5.38), we have

$$kx^1(0) - \frac{k}{\alpha}x^2(0) < ky^1(0) - \frac{k}{\alpha}y^2(0).$$

By Fundamental Theorem of Calculus, we have

$$\begin{aligned}\nabla u(x^1(0)) - \nabla u(x^2(0)) = \int_{x^2(0)}^{x^1(0)} \nabla^2 u(z)dz > \int_{x^2(0)}^{x^1(0)} \nabla^2 g(z)dz \geq \int_{y^2(0)}^{y^1(0)} \nabla^2 g(z)dz \\ = \nabla g(y^1(0)) - \nabla g(y^2(0)).\end{aligned}$$

where the first inequality follows from Eq. (5.35) and the second one is by the fact that $\nabla^2 g(z) \leq 0$ and Eq. (5.38). The previous three relations lead to a contradiction and therefore Eq. (5.38) cannot hold. This proves the desired result. □

The above theorem implies that utility function g, which is more concave, is associated with less fluctuation for $S = 0$. We next build on this result and study the case where $S > 0$.

Theorem 5.3.4. *Consider the two twice differentiable concave utility functions u and g defined as above. Then for the optimal solutions $(x(S), \bar{x}(S)$, $(y(S), \bar{y}(S))$ of problem (5.34) for all $S > 0$ where the constraints $x^1 - \bar{x}^1 \leq S$, $y^1 - \bar{y}^1 \leq S$ are tight, the following condition holds*

$$\bar{x}^1(S) - \bar{x}^2(S) \geq \bar{y}^1(S) - \bar{y}^2(S).$$

Proof. Since the storage constraint is tight, we have $\bar{x}^1(S) = x^1(S) - S$ and $\bar{x}^2(S) = x^2(S) + S$, which gives

$$\bar{x}^1(S) - \bar{x}^2(S) = x^1(S) - x^2(S) - 2S.$$

Similarly for utility function g, we have that the optimal solution satisfies

$$\bar{y}^1(S) - \bar{y}^2(S) = y^1(S) - y^2(S) - 2S.$$

We will establish the desired relation by showing

$$x^1(S) - x^2(S) \geq y^1(S) - y^2(S).$$

These terms can be equivalently represented as

$$x^1(S) - x^2(S) = x^1(0) - x^2(0) + (x^1(S) - x^1(0)) + (x^2(0) - x^2(S)),$$

and

$$y^1(S) - y^2(S) = y^1(0) - y^2(0) + (y^1(S) - y^1(0)) + (y^2(0) - y^2(S))$$

respectively. Theorem 5.3.3 and the fact that market clears at each period for $S = 0$, i.e., $x^t = \bar{x}^t$ and $y^t = \bar{y}^t$, guarantee

$$x^1(0) - x^2(0) \geq y^1(0) - y^2(0).$$

Next we will show that

$$x^1(S) - x^1(0) > y^1(S) - y^1(0), \quad x^2(0) - x^2(S) > y^2(0) - y^2(S). \tag{5.40}$$

We first show that both sides of the preceding inequalities are nonnegative.

Eq. (5.37) and Lemma 5.2.3 implies that the constraint $x^1(S) + x^2(S) \leq d$ is not tight at the optimal solutions for all $S \geq 0$ with either utility functions. By using complementary slackness condition, we can simplify (FOC) condition from Lemma 5.2.1 to

$$\begin{aligned} -\nabla u(x^1(0)) + \Delta p + p_2 &= 0, & k\bar{x}^1(0) - \Delta p - p_2 &= 0, \\ -\alpha \nabla u(x^2(0)) + p_2 &= 0, & k\bar{x}^2(0) - p_2 &= 0. \end{aligned}$$

By using the market clearing constraint, i.e., $x^t = \bar{x}^t$, this implies that

$$\nabla u(x^1(0)) = kx^1(0), \quad \alpha \nabla u(x^2(0)) = kx^2(0). \tag{5.41}$$

Similarly, by using the fact that the storage constraint is tight, we have for $S > 0$,

$$\begin{aligned} -\nabla u(x^1(S)) + \Delta p + p_2 &= 0, & k(x^1(S) - S) - \Delta p - p_2 &= 0, \\ -\alpha \nabla u(x^2(S)) + p_2 &= 0, & k(x^2(S) + S) - p_2 &= 0. \end{aligned}$$

This yields that

$$\nabla u(x^1(S)) = k(x^1(S) - S), \quad \alpha \nabla u(x^2(S)) = k(x^2(S) + S). \tag{5.42}$$

If $x^1(S) < x^1(0)$, then we have $k(x^1(S) - S) < kx^1(0)$. Since function u is concave, its gradient is monotonically nonincreasing and we have that $\nabla u(x^1(S)) \geq \nabla u(x^1(0))$, which leads to a contradiction with Eqs. (5.41) and (5.42). Hence we conclude that

$$x^1(S) \geq x^1(0).$$

The same result holds for utility function g and thus we also have

$$y^1(S) \geq y^1(0).$$

We now consider the second period. If $x^2(S) > x^2(0)$, then we have $k(x^2(S) + S) > kx^2(0)$. Once again, by the concavity of function u, we have $\alpha \nabla u(x^2(S)) \leq \alpha \nabla u(x^2(0))$, which is another contradiction with the preceding two equality systems. Hence we conclude

$$x^2(S) \leq x^2(0)$$

and similarly

$$y^2(S) \leq y^2(0).$$

For notational simplicity, we define nonnegative scalars δ^1, δ^2, ϵ^1, ϵ^2 as

$$x^1(0) + \delta^1 = x^1(S), \quad y^1(0) + \epsilon^1 = y^1(S),$$
$$x^2(S) + \delta^2 = x^2(0), \quad y^2(S) + \epsilon^2 = y^2(0).$$

The desired condition in Eq. (5.40) can be written compactly as

$$\delta^1 > \epsilon^1, \quad \delta^2 > \epsilon^2.$$

We first show that $\delta^1 > \epsilon^1$. Eqs. (5.41) and (5.42) imply that

$$\nabla u(x^1(0)) = kx^1(0), \quad \nabla u(x^1(0) + \delta^1) = k(x^1(0) + \delta^1 - S).$$

We can take the difference of these two equalities and obtain that

$$\nabla u(x^1(0)) - \nabla u(x^1(0) + \delta^1) = -k\delta^1 + kS.$$

Since function u is twice differentiable, the preceding equality can be written equivalently as

$$\int_{x^1(0)+\delta^1}^{x^1(0)} \nabla^2 u(z)dz = -k\delta^1 + kS.$$

Similarly, for utility function g, we have

$$\int_{y^1(0)+\epsilon^1}^{y^1(0)} \nabla^2 g(z)dz = -k\epsilon^1 + kS.$$

We will prove $\delta^1 > \epsilon^1$ by contradiction, assume $\delta^1 \leq \epsilon^1$, and thus

$$-k\delta^1 + kS \geq -k\epsilon^1 + kS.$$

Eq. (5.35) suggests that

$$\int_{x^1(0)+\delta^1}^{x^1(0)} \nabla^2 u(z)dz < \int_{y^1(0)+\delta^1}^{y^1(0)} \nabla^2 g(z)dz \leq \int_{y^1(0)+\epsilon^1}^{y^1(0)} \nabla^2 g(z)dz,$$

where the second relation follows from $\delta^1 \leq \epsilon^1$ and $\nabla^2 g(z) \leq 0$. The above four relations lead to a contradiction and hence we conclude

$$\delta^1 > \epsilon^1.$$

We next show that $\delta^2 > \epsilon^2$. Eqs. (5.41) and (5.42) imply that

$$\alpha \nabla u(x^2(0)) = kx^2(0), \quad \alpha \nabla u(x^2(0) - \delta^2) = k(x^2(0) - \delta^1 + S).$$

We can take the difference of these two equalities and obtain that

$$\nabla u(x^2(0)) - \nabla u(x^2(0) - \delta^2) = k\delta^1 - kS.$$

Since function u is twice differentiable, the preceding equality can be written equivalently as

$$\alpha \int_{x^2(0)-\delta^2}^{x^2(0)} \nabla^2 u(z)dz = k\delta^2 - kS.$$

Similarly, for utility function g, we have

$$\alpha \int_{y^2(0)-\epsilon^2}^{y^2(0)} \nabla^2 g(z)dz = k\epsilon^2 - kS.$$

If $\delta^2 \leq \epsilon^2$, we have

$$k\delta^2 - kS \leq k\epsilon^2 - kS.$$

By using Eq. (5.35), we have

$$\alpha \int_{y^2(0)-\epsilon^2}^{y^2(0)} \nabla^2 g(z)dz < \alpha \int_{x^2(0)-\epsilon^2}^{x^2(0)} \nabla^2 u(z)dz \leq \alpha \int_{x^2(0)-\delta^2}^{x^2(0)} \nabla^2 u(z)dz,$$

where the last inequality follows from the assumption that $\delta^2 \leq \epsilon^2$ and $\nabla^2 u(z) \leq 0$. The preceding four relations combined gives a contradiction and therefore we have

$$\delta^2 > \epsilon^2.$$

This completes the proof of Eq. (5.40), which in turn shows the desired result. □

The preceding theorem states that for positive storage, the optimal solution associated with more concave utility functions have less generation fluctuation. The less concave utility functions are more responsive to price changes. When we introduce storage into the system, the peak time production goes down and as a result, the peak time price goes down too. With demand responding to the price change, at the optimal solution, demand for peak time goes up (compared to $S = 0$). Intuitively, the more responsive demand (with more concave utility function) will raise peak time demand by more than a less responsive demand. The opposite effects happen in the off-peak period, with generation going up and price going down. The more responsive demand will reduce total demand for the off-peak period. Thus the overall fluctuation associated with a more responsive demand is higher than the less responsive one.

5.4 Summaries

In this chapter, we also address the issue of undesirable price and generation fluctuations, which imposes significant challenges in maintaining reliability of the electricity grid. We first establish that the two fluctuations are correlated. Then in order to reduce both fluctuations, we introduce an explicit penalty on the price fluctuation, where the penalized problem is equivalent to the existing system with storage. We give a distributed implementation of the new system, where each agent locally responds to price signals. Lastly, we analyze the connection between the size of storage, demand properties and generation fluctuation in two scenarios: when demand is inelastic, we can explicitly characterize the optimal storage access policy and the generation fluctuation; when demand is elastic, the relationship between concavity of generation fluctuation is studied and we show that concavity of consumer demand may not always reduce or increase generation fluctuation even in a two period example. The multi-period problem is left as a future research direction.

Chapter 6

Conclusions

In this chapter, we conclude the thesis by summarizing its main contributions and provide some interesting future directions.

6.1 Summary

In this thesis, we study two classes of multi-agent systems with local information. In the first class of systems, the agents are cooperative and aiming to optimize a sum of convex local functions. We developed both synchronous and asynchronous ADMM (Alternating Direction Method of Multipliers) based distributed methods to solve this problem and have shown that both method obtain a $O(1/k)$ rate of convergence, where k is the number of iterations. This rate is the best known rate for this class of problems and is the first rate of convergence guarantee for asynchronous methods. For the synchronous methods, we also relate the algorithm performance to the underlying graph topology.

For the second class of networks, where the agents are only interested in local objectives, we study the market interaction in the electricity market. We propose a systematic framework to capture demand response and show that the new market interaction at competitive equilibrium is efficient and the improvement in social welfare over the traditional market can be arbitrarily large. The resulting system, however, may feature undesirable price and generation fluctuations, which imposes significant challenges in maintaining reliability of the electricity grid. We first establish that the two fluctuations are correlated. Then in order to reduce both fluctuations, we introduce an explicit penalty on the price fluctuation, where the penalized problem is equivalent to the existing system with storage. We give a distributed implementation of the new system, where each agent locally responds to price signals. Lastly, we analyze the connection between the size of storage, demand properties and generation fluctuation in two scenarios: when demand is inelastic, we can explicitly characterize the optimal storage access policy and the generation fluctuation; when demand is elastic, the relationship between concavity of generation fluctuation is studied and we show that concavity of consumer demand may not always reduce or increase generation fluctuation.

6.2 Future Directions

On the optimization fronts, some interesting future works include to generalize the effect of network topology to general asynchronous methods and to analyze the algorithm on a time-varying network topology. For the electricity market study, the multi-period, multi-user storage access characterization is left open. We have also assumed all the agents are price-taking, the case where agents are strategic will be very interesting to study.

Bibliography

[1] C. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer, 2006.

[2] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation.* PhD thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Belmont, MA: Athena Scientific, 1997.

[4] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," *Proceedings of IEEE Conference on Decision and Control (CDC)*, 2005.

[5] A. Jadbabaie, J. Lin, and S. Morse, "Coordination of Groups of Mobile Autonomous Agents using Nearest Neighbor Rules,," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[6] R. Olfati-Saber and R. M. Murray, "Consensus Problems in Networks of Agents with Switching Topology and Time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[7] A. Olshevsky and J. N. Tsitsiklis, "Convergence Speed in Distributed Consensus and Averaging," *SIAM Journal on Control and Optimization*, vol. 48(1), pp. 33–35, 2009.

[8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2508–2530, 2006.

[9] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic Gossip: Efficient Aggregation for Sensor Networks," *Proceedings of International Conference on Information Processing in Sensor Networks*, pp. 69–76, 2006.

[10] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Transactions on Automatic Control*, vol. 31(9), pp. 803–812, 1986.

[11] A. Nedić and A. Ozdaglar, "Distributed Subgradient Methods for Multi-agent Optimization," *IEEE Transactions on Automatic Control*, vol. 54(1), pp. 48–61, 2009.

[12] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient Methods and Consensus Algorithms for Solving Convex Optimization Problems," *Proceedings of IEEE Conference on Decision and Control(CDC)*, pp. 4185–4190, 2008.

[13] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained Consensus and Optimization in Multi-agent Networks," *IEEE Transactions on Automatic Control*, vol. 55(4), pp. 922–938, 2010.

[14] I. Lobel and A. Ozdaglar, "Convergence Analysis of Distributed Subgradient Methods over Random Networks," *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[15] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed Multi-agent Optimization with State-Dependent Communication," *Mathematical Programming, special issue in honor of Paul Tseng*, vol. 129, no. 2, pp. 255–284, 2011.

[16] I. Matei and J. S. Baras, "Performance Evaluation of the Consensus-Based Distributed Subgradient Method Under Random Communication Topologies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 754–771, 2011.

[17] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[18] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "Distributed Subgradient Algorithms and Quantization Effects," *Proceedings of IEEE Conference on Decision and Control (CDC)*, 2008.

[19] Y. Nesterov, "Primal-dual subgradient methods for convex problems," *Mathematical Programming*, vol. 120, no. 1, pp. 221–259, 2009.

[20] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual Averaging for Distributed Optimization : Convergence Analysis and Network Scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[21] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Fast Distributed Gradient Methods," *Arxiv preprint*, 2011.

[22] R. Glowinski and A. Marrocco, "Sur l'Approximation, par Elements Finis d'Ordre un, et la Resolution, par Penalisation-dualité, d'une Classe de Problems de Dirichlet non Lineares," *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, vol. 9, pp. 41âĂŞ–76, 1975.

[23] D. Gabay and B. Mercier, "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation ," *Computers and Mathematics with Applications*, vol. 2, no. 1, pp. 17– 40, 1976.

[24] M. Fortin and R. Glowinski, "On Decomposition-Coordination Methods Using an Augmented Lagrangian," *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems, (M. Fortin and R. Glowinski, eds.)*, 1983.

[25] J. Douglas and H. H. Rachford, "On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables," *Transactions of the American Mathematical Mathematical Society*, vol. 82, pp. 421–439, 1956.

[26] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators," *LIDS Report 1919*, 1989.

[27] P. L. Lions and B. Mercier, " Splitting Algorithms for the Sum of Two Nonlinear Operators," *SIAM Journal on Numerical Analysis*, vol. 16, pp. 964–979, 1979.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[29] J. Mota, J. Xavier, P. Aguiar, and M. Püschel, "ADMM For Consensus On Colored Networks," *Proceedings of IEEE Conference on Decision and Control (CDC)*, 2012.

[30] J. Mota, J. Xavier, P. Aguiar, and M. Püschel, "D-ADMM : A Communication-Efficient Distributed Algorithm For Separable Optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2718–2723, 2013.

[31] I. D. Schizas, R. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs with Noisy Links - Part I: Distributed Estimation of Deterministic Signals," *IEEE Transactions on Singal Processing*, vol. 56, pp. 350–364, 2008.

[32] H. Zhu, A. Cano, and G. B. Giannakis, "In-Network Channel Decoding Using Consensus on Log-Likelihood Ratio Averages," *Proceedings of Conference on Information Sciences and Systems (CISS)*, 2008.

[33] J. Eckstein, "Augmented Lagrangian and Alternating Direction Methods for Convex Optimization: A Tutorial and Some Illustrative Computational Results," *Rutcor Research Report*, 2012.

[34] B. He and X. Yuan, "On the O(1/n) Convergence Rate of the Douglas-Rachford Alternating Direction Method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.

[35] W. Deng and W. Yin, "On the Global and Linear Convergence of the Generalized Alternating Direction Method of Multipliers," *Rice University CAAM, Technical Report TR12-14*, 2012.

[36] D. Goldfarb, S. Ma, and K. Scheinberg, "Fast Alternating Linearization Methods for Minimizing the Sum of Two Convex Functions," *Mathematical Programming*, pp. 1–34, 2012.

[37] D. Goldfarb and S. Ma, "Fast Multiple Splitting Algorithms for Convex Optimization," *Under revision in SIAM Journal on Optimization*, 2010.

[38] D. Han and X. Yuan, "A Note on the Alternating Direction Method of Multipliers," *Journal of Optimization Theory and Applications*, vol. 155, no. 1, pp. 227–238, 2012.

[39] M. Hong and Z. Luo, "On the Linear Convergence of the Alternating Direction Method of Multipliers," *Arxiv preprint*, 2012.

[40] C. I. S. Operator, "Integration of Renewable Resources: Operational Requirements and Generation Fleet Capability at 20% RPS," *California Independent System Operator Report*, 2010.

[41] M. Ortega-Vazquez and D. Kirschen, "Assessing the Impact of Wind Power Generation on Operating Costs," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 295–301, 2010.

[42] S. Meyn, M. Negrete-Pincetic, G. Wang, A. Kowli, and E. Shafieepoorfard, "The Value of Volatile Resources in Electricity Markets," in *Proceedings of IEEE Conference on Decision and Control (CDC), 2010*, pp. 1029–1036, 2010.

[43] M. R. Milligan and B. Kirby, *Calculating Wind Integration Costs: Separating Wind Energy Value from Integration Cost Impacts*. National Renewable Energy Laboratory Golden, 2009.

[44] T. Lee, M. Cho, Y. Hsiao, P. Chao, and F. Fang, "Optimization and implementation of a load control scheduler using relaxed dynamic programming for large air conditioner loads," *Power Systems, IEEE Transactions on*, vol. 23, no. 2, pp. 691–702, 2008.

[45] C. Tan and P. Varaiya, "Interruptible Electric Power Service Contracts," *Journal of Economic Dynamics and Control*, vol. 17, no. 3, pp. 495–517, 1993.

[46] M. Alizadeh, Y. Xiao, A. Scaglione, and M. van der Schaar, "Incentive Design for Direct Load Control Programs," *arXiv preprint arXiv:1310.0402*, 2013.

[47] P. Ahlstrand, "Demand-side Real-time Pricing," *1986 annual report (Pacific Gas & Electric Co., San Francisco, CA)*, 1987.

[48] T. Chang, M. Alizadeh, and A. Scaglione, "Real-Time Power Balancing Via Decentralized Coordinated Home Energy Scheduling," *Smart Grid, IEEE Transactions on*, vol. 4, pp. 1490–1504, Sept 2013.

[49] L. Gan, U. Topcu, and S. Low, "Stochastic Distributed Protocol for Electric Vehicle Charging with Discrete Charging Rate," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–8, IEEE, 2012.

[50] W. Tushar, W. Saad, H. V. Poor, and D. B. Smith, "Economics of Electric Vehicle Charging: A Game Theoretic Approach," *Smart Grid, IEEE Transactions on*, vol. 3, no. 4, pp. 1767–1778, 2012.

[51] N. Rotering and M. Ilic, "Optimal Charge Control of Plug-in Hybrid Electric Vehicles in Deregulated Electricity Markets," *Power Systems, IEEE Transactions on*, vol. 26, no. 3, pp. 1021–1029, 2011.

[52] Z. Ma, D. Callaway, and I. Hiskens, "Optimal charging control for plug-in electric vehicles," in *Control and Optimization Methods for Electric Smart Grids*, Springer, 2012.

[53] G. Schwartz, T. Hamidou, S. Amin, and S. Sastry, "Electricity demand shaping via randomized rewards: a mean field game approach," in *Allerton Conference on Communication, Control, and Computing*, IEEE, 2012.

[54] P. Loiseau, G. A. Schwartz, J. Musacchio, S. Amin, and S. S. Sastry, "Congestion pricing using a raffle-based scheme," in *NetGCoop*, pp. 1–8, 2011.

[55] D. Merugu, B. Prabhakar, and N. Rama, "An incentive mechanism for decongesting the roads:a pilot program in bangalore," in *NetEcon Workshop on the Economics of Networked Systems*, ACM, 2009.

[56] M. P. Nowak and W. Römisch, "Stochastic Lagrangian Relaxation Applied to Power Scheduling in a Hydro-thermal System under Uncertainty," *Annals of Operations Research*, vol. 100, no. 1-4, pp. 251–272, 2000.

[57] P. Carpentier, G. Gohen, J.-C. Culioli, and A. Renaud, "Stochastic optimization of unit commitment: a new decomposition framework," *Power Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 1067–1073, 1996.

[58] I. Cho and S. Meyn, "A dynamic newsboy model for optimal reserve management in electricity markets," 2010.

[59] D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng, "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem," *Power Systems, IEEE Transactions on*, vol. 28, no. 1, pp. 52–63, 2013.

[60] E. T. Mansur, "Measuring Welfare in Restructured Electricity Markets," *The Review of Economics and Statistics*, vol. 90, no. 2, pp. 369–386, 2008.

[61] I. Cho and S. P. Meyn, "Efficiency and marginal cost pricing in dynamic competitive markets with friction," *Theoretical Economics*, vol. 5, no. 2, 2010.

[62] A. Kizilkale and S. Mannor, "Regulation and efficiency in markets with friction," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, pp. 4137–4144, Dec 2010.

[63] M. Roozbehani, M. A. Dahleh, and S. K. Mitter, "Volatility of power grids under real-time pricing," *CoRR*, 2011.

[64] Q. Huang, M. Roozbehani, and M. Dahleh, "Efficiency-risk tradeoffs in dynamic oligopoly markets - with application to electricity markets," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 2388–2394, 2012.

[65] E. Bitar and S. Low, "Deadline differentiated pricing of deferrable electric power service," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 4991–4997, IEEE, 2012.

[66] J. N. Tsitsiklis and Y. Xu, "Pricing of Fluctuations in Electricity Markets.," in *CDC*, pp. 457–464, IEEE, 2012.

[67] A. Nayyar, M. Negrete-Pincetic, K. Poolla, and P. Varaiya, "Duration-differentiated services in electricity," *arXiv preprint arXiv:1404.1112*, 2014.

[68] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, 2008.

[69] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *Automatic Control, IEEE Transactions on*, vol. 54, pp. 48 –61, Jan 2009.

[70] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *Journal of Machine Learning Research*, vol. 11, pp. 1663–1707, 2010.

[71] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A Distributed Newton Method for Network Utility Maximization, I: Algorithm," pp. 2162 – 2175, 2013.

[72] E. Wei and A. Ozdaglar, "Distributed Alternating Direction Method of Multipliers," *Proceedings of IEEE Conference on Decision and Control (CDC)*, 2012.

[73] E. Wei and A. Ozdaglar, "On the $O(1/k)$ Convergence of Asynchronous Distributed Alternating Direction Method of Multipliers," *LIDS report 2906, submitted to Mathematical Programming*, 2013.

[74] S. Shtern, E. Wei, and A. Ozdaglar, "Distributed Alternating Direction Method of Multipliers (ADMM): Performance and Network Effects," *LIDS report 2917*, 2014.

[75] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization.* Cambridge, MA: Athena Scientific, 2003.

[76] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics).* No. 92, American Mathematical Society, 1997.

[77] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous Gossip Algorithms for Stochastic Optimization," *Proceedings of IEEE Conference on Decision and Control (CDC)*, 2009.

[78] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Asynchronous Distributed Optimization using a Randomized Alternating Direction Method of Multipliers," *Submitted to IEEE Conference on Decision and Control (CDC)*, 2013.

[79] G. Grimmett and D. Stirzaker, *Probability and Random Processes.* Oxford: Oxford Univeristy Press, 2001.

[80] D. Williams, *Probability with Martingales.* Cambridge: Cambridge University Press, 1991.

[81] D. P. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization.* Cambridge, MA: Athena Scientific, 2003.