The Pennsylvania State University
The Graduate School

ALGORITHMS FOR OPERATION OF POWER SYSTEMS: RISK, UNCERTAINTY, DISCRETENESS, AND NONCONVEXITY

A Dissertation in
Industrial Engineering
by
Wendian Wan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2019

The dissertation of Wendian Wan was reviewed and approved* by the following:

Uday V. Shanbhag
Gary and Sheila Bello Chair and Professor of Industrial and Manufacturing Engineering
Dissertation Advisor, Co-Chair of Committee

Mort D. Webster
Ryan Family Faculty Fellow and Professor of Energy Engineering
Co-Chair of Committee

Vittal Prabhu
Professor of Industrial and Manufacturing Engineering

Guodong Pang
Associate Professor of Industrial and Manufacturing Engineering

Robert Voigt
Professor of Industrial and Manufacturing Engineering
Graduate Program Coordinator

*Signatures are on file in the Graduate School.

Abstract

This dissertation considers the development of computational schemes for a class of operational problems in power systems, complicated by uncertainty, discreteness, and nonconvexity.

In Chapter 2, we consider a class of risk-based two-stage economic dispatch problems, a class of problems that can be captured by stochastic convex programs where the integrands are nonsmooth convex functions and each function evaluation requires solving a convex optimization problem. We proceed to show that the risk of the second-stage cost satisfies smoothability requirements under suitable assumptions. This allows for adapting a variable sample-size accelerated proximal scheme (VS-APM) for such problems. Notably, this scheme is a stochastic approximation scheme that combines smoothing, acceleration, and variance reduction. The resulting expected sub-optimality diminishes to zero at the rate of $\mathcal{O}(1/k)$. We observe that the scheme performs well in comparison with standard stochastic gradient as well as stochastic cutting-plane schemes on a range of IEEE test problem sets.

In Chapter 3, we consider a class of stochastic integer programs that arise from a two-stage stochastic unit commitment problem. We present a computational framework for addressing such a problem by combining the VS-APM scheme with a branching scheme. Such a framework is fairly adaptable and can allow for a broad range of risk-based convex models. Preliminary testing suggests that the scheme competes well with `CPLEX` when the problem has first-stage integers and the number of second-stage scenarios grows to be large. In more general problems with second-stage problems, the scheme can obtain global solutions for modest sized problems.

Finally in Chapter 4, we consider the optimal power flow problem with AC power flow constraints. The resulting problem is known to be a highly nonconvex problem and the solution of such problems is generally challenging. We consider an avenue for resolving such problems that relies on an alternating direction method of multipliers (ADMM) scheme. This scheme can be implemented in a networked setting and its performance is seen to scale with the number of scenarios when stochastic variants are considered.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express the deepest appreciation to my committee chair Professor Uday V. Shanbhag. As my advisor and mentor, he has taught me much more than I could ever imagine. His patience and kindness will always be kept in the bottom of my heart. Without his guidance and persistent help this dissertation would not be possible. I would also like to extend my deepest gratitude to my committee member Professor Mort D. Webster. His guidance and advice inspired me with more piratical and fundamental understandings of my work and career. I am also grateful to all my other committee members, Professor Guodong Pang and Professor Vittal Prabhu, for many generous and helpful suggestions and comments.

It was a great pleasure working with Afrooz Jalilzadeh, Jinlong Lei, Shisheng Cui and Yue Xie. Great advisor always brings excellent students. Working together with you will always be a wonderful experience to me. Thanks should also go to Erfan Yazdandoost, Xuan Zhang for sharing office with me. I cannot leave Penn State without mentioning Wenjing Song, Yiou Wang, Xin Li, Lan Jin, Hang Li and many others in the department for our time in Leonhard building and friends from all other departments for our time in State College.

I must also thank all my college friends, your kindness helped me settle down here far away from home, specially Xiujie Zhao who gave me home feeling from the very first day and Xue Wang who helped me go through countless ups and downs both mentally and academically throughout these years. It is never a easy task to become a phd and I shall also thank all those who have encouraged me in different ways. I am deeply indebted to my parents who supported me in any possible aspect throughout my life. Special thanks to Siyang Hao, for sharing this memorable journey together with me. It would never be such a wonderful adventure without you. It's my privilege to have your company.

I would also like to acknowledge the support of NSF Grants 1408366 and 1710974 as well as DOE ARPA-E 1952-1515.

Chapter 1
Introduction

This dissertation considers the development of computational schemes for contending with stochastic, discrete, and nonconvex optimization problems arising in the context of power systems operations. In this chapter, we provide a brief introduction to each of the subsequent chapters where the emphasis will be on providing background and motivation and a brief formulation of the problem of interest.

1.1 Two-stage risk-based economic dispatch

Chapter 2 of this dissertation focuses on developing an algorithm for two-stage economic dispatch in either risk-averse (with conditional value-at-risk measures) or risk-neutral settings.

1.1.1 Introduction to two-stage stochastic programs

We begin by considering the class of two-stage programs with following form:

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}\left[\mathcal{Q}(x,\omega)\right] \\ \text{subject to} \quad & x \in X, \end{aligned} \tag{P-S1}$$

where $X \subseteq \mathbb{R}^n, x \in \mathbb{R}^n$, $\omega \in \Omega$, and $\mathcal{Q}(x,\omega)$ denotes the optimal value of the following program:

$$\begin{aligned} \max \quad & -\tfrac{1}{2}\pi_\omega^T Q_\omega \pi_\omega + (h_\omega - T_\omega x)^T \pi_\omega \\ \text{subject to} \quad & W^T \pi_\omega \leq q_\omega, \end{aligned} \tag{LP-S2}$$

where $Q_\omega \in \mathbb{R}^{m_2 \times m_2}$, $\pi_\omega \in \mathbb{R}^{m_2}$, $T_\omega \in \mathbb{R}^{m \times n}$, $h_\omega \in \mathbb{R}^m$, $W \in \mathbb{R}^{m_2 \times n_2}$, and $q_\omega \in \mathbb{R}^{n_2}$. Before proceeding, we briefly review some techniques for resolving two-stage stochastic convex programs. We describe two broad avenues for contending with two-stage stochastic convex programs based on whether the sample-space is finite or infinite.

(i). Decomposition schemes. Decomposition techniques [1] essentially represent a class of techniques whose effort grows slowly with problem size (which is proportional to $|\Omega|$). These techniques include primal decomposition techniques as well as dual decomposition techniques, such as Benders decomposition [2] which provided the inspiration for the celebrated L-shaped method [3].
(ii). Monte-Carlo sampling schemes. Often the size of the scenario space may be large or even infinite. In such settings, Monte-Carlo sampling techniques assume relevance. Several avenues have gained in popularity over the last several decades. Amongst the earliest of these was the stochastic quasigradient methods by Ermoliev (see [4]) which was essentially a stochastic approximation scheme. An alternate avenue utilizes a sampling-based generalization of the L-shaped method [3] and is known as stochastic decomposition [5,6]. Finally, a third approach, referred to as sample average approximation, replaces the expectation by a sample-average and constructs estimators by computing a solution of this sampled problem [7].

1.1.2 Risk-based two-stage stochastic programs

The earlier model is a risk-neutral framework that employs an expectation-based framework. This can be generalized to allow for risk preferences by using the CVaR measure that captures the risk of sub-hourly dispatch cost. For a fixed level τ, the conditional value-at-risk of economic dispatch is defined as:

$$CVaR_\tau(\mathcal{Q}(x,\omega)) = \min_m \left\{ m + \frac{1}{1-\tau}\mathbb{E}\left[\mathcal{Q}(x,\omega) - m\right]^+ \right\}$$

Then the objective of risk-averse sub-hour stochastic dispatch is:

$$\begin{aligned} f_{obj} &= \min_x \left\{ f_0(x) + \mathrm{CVaR}_\tau(\mathcal{Q}(x,\omega)) \right\} \\ &= \min_{x,m} \left\{ f_0(x) + m + \frac{1}{1-\tau}\mathbb{E}\left[\mathcal{Q}(x,\omega) - m\right]^+ \right\} \end{aligned}$$

$$= \min_{x,m} \left\{ f_0(x) + \mathbb{E}\left[h(\mathcal{Q}(x,\omega), m) \right] \right\}$$

where

$$h(\mathcal{Q}(x,\omega), m) = m + \frac{1}{1-\tau}\left[\mathcal{Q}(x,\omega) - m\right]^+$$

In this dissertation, we discuss a smoothed accelerated gradient scheme for the two-stage stochastic convex problem:

$$\min_{\mathbf{z}\in Z} \left[c(\mathbf{z}) + \mathbb{E}\left[r(\mathbf{z},\omega) \right] \right], \tag{r-ED}$$

where $\mathbf{z} \triangleq (\mathbf{x}, \mathbf{m})$, $c(\mathbf{z})$ is a convex differentiable function, $r(\mathbf{z},\omega)$ is a convex and nonsmooth function, and Z is a polyhedral set.

1.1.3 Economic Dispatch and Optimal Power Flow

The optimal power flow (OPF) problem is amongst the most fundamental decision-making problems in power systems. There are many variations and generalizations of this problem including unit commitment, reserve scheduling, economic dispatch, security-constrained, DC approximations, and full AC power-flow formulations [8–12]. Our focus in this chapter lies in contending with uncertainty. One approach to address the presence of uncertainty lies in developing robust optimization models [13–17] where uncertainty sets are assumed and feasibility is ensured for every realization from such uncertainty sets. For instance, a robust formulation for AC power flow problems was provided in [18]. Uncertainty in power systems also can be dealt by adding chance constraints [12, 19] which are approximated by finite sampling of uncertain parameters from an assumed statistical model. Simulation based studies on real time dispatch are also been conducted. For instance, in [20], a simulation based framework is used in a power system with renewable resources. Our interest lies in adaptive two-stage models for such problems possibly complicated by the presence of risk measures. Such avenues have been considered in [21] and revisited in [22], where computable closed-form expressions were derived. In [23], a model for risk-limiting dispatch with generation limitation and network constraint was provided with networked variants were examined in [24].

1.1.4 Outline of Chapter 2

We consider a two-stage economic dispatch model with a conditional value-at-risk (CVaR) measure. The resulting problem is a stochastic convex program with possibly nonsmooth integrands. We make several contributions in this context.

(i) First, the conditional value-at-risk measure is a composition of two nonsmooth functions, both of which are convex. Our first set of contributions lies in showing that by smoothing each of the functions, the overall function satisfies suitable smoothness requirements.

(ii) Second, we show that a previously developed stochastic approximation (SA) scheme that combined smoothing, Nesterov acceleration, and variance reduction can be applied. As a consequence, the expected sub-optimality diminishes at a rate of $\mathcal{O}(1/k)$ while the oracle complexity to compute an ϵ-optimal solution is shown to be $\mathcal{O}(1/\epsilon^2)$.

(iii) Third, we compare this scheme with comparable cutting-plane schemes and observe that the presented scheme has significant computational benefits on a class of IEEE test systems.

1.2 Two-stage Stochastic Integer Programs

Chapter 3 of this dissertation focuses on extending VS-APM for two-stage stochastic integer programs in power system like unit-commitment problems.

1.2.1 Introduction to two-stage stochastic integer programs

Stochastic integer programming models are very important in practice with discrete decisions under uncertainty, especially in power system with unit commitment constraints and decisions. Although such models can be difficult to solve in practice, an extension of VS-APM could lead to an alternative direction of stochastic integer programs study. In this dissertation, we consider the class of two-stage programs of the following form:

$$\min \quad c^T x + \mathbb{E}[\mathcal{Q}(x,\omega)]$$

$$\text{subject to} \quad x \in \mathcal{X} \cap X$$

where $\mathcal{X}$ denotes a convex polyhedron and X denotes either the set of binary vectors $\mathbb{B}$, or integer $\mathbb{Z}$ or mixed-integer vectors $\mathbb{M} \triangleq \{x | x \geq 0, x_j \in \mathbb{Z}, j \in J\}$. $\mathcal{Q}(x, \omega)$ refers to a second stage problem that dealing with am realization of random variables ω with the following formulation:

$$\begin{aligned} \mathcal{Q}(x, \omega) = \min \quad & g_\omega^T y \\ \text{subject to} \quad & W_\omega y \geq r_\omega - T_\omega x \\ & y \geq 0, y_j \in \mathbb{Z}, j \in J_2 \end{aligned}$$

where J_2 in an index set that represents integer part of second stage variable. When J_2 is empty, then there is only integer variable in first stage, i.e prior to the realization of the random variables. In [25], a combination of sample-based function evaluations and branch and bound algorithm is proposed for such problems. In the case with non empty J_2, second stage linear programs can not be approximated via standard techniques directly. Due to its size, decomposition is a natural avenue for such large problems. One direction is also applying the idea of Benders decomposition and Cutting plane method, Caroe [26] discusses an algorithm that is motivated by deterministic cutting plane methods. Sherali and Fraticelli [27] develop cutting plane methods when all variables (first and second stage) are binary. Another direction is to use dual decomposition. In [28], an algorithm based on a dual decomposition scheme and Lagrangian relaxation was proposed to solve general linear two-stage stochastic programs with integrality requirement. By relaxing non-anticipativity constraints via Lagrangian relaxation, a separate subproblem for each scenario is formulated.

1.2.2 Outline of Chapter 3

We consider a two-stage model with integer variables. The resulting problem is a stochastic integer program. We make several contributions in this context.

(i) First, we extend our algorithm to stochastic integer programs by introducing branching techniques. We also demonstrate that this framework can be applied for more general stochastic integer programs.

(ii) Second, we enhance our framework by introducing more sophisticated integer programming techniques like cutting plane and branching rules, which lead to improvement on performance of stochastic integer framework.

(iii) Third, we compare this scheme with state-of-art integer programs solver and observe that this framework is potential to provide good quality integer solution in reasonable time.

1.3 Distributed framework for AC-OPF

Chapter 4 of this dissertation focuses on extending distributed algorithm to non-convex regime in dealing with more general power system problems.

1.3.1 Introduction to AC-OPF

With increasing penetration of distributed energy resources, centralized model and algorithm in power system can suffer from lack of information to scalability, especially for more practical Alternating current optimal flow (AC-OPF) model. An alternative approach to traditional centralized methods is to consider distributed algorithms. Rather than collecting all problem parameters and performing a central calculation, distributed algorithms are computed by many agents that obtain certain problem parameters via communication with a limited set of neighbors.

1.3.2 Distributed schemes for AC-OPF

Several distributed optimization techniques have been applied to power system optimization. Adopting from the exposition in [29], the first set of distributed optimization techniques are based on augmented Lagrangian decomposition. These include Dual Decomposition, the Alternating Direction Method of Multipliers with Proximal Message Passing, Analytical Target Cascading, and the Auxiliary Problem Principle. The second set of techniques are based on decentralized solution of the Karush-Kuhn-Tucker (KKT) necessary conditions for local optimality. These include Optimality Condition Decomposition and Consensus+Innovation. Two other approaches, Gradient Dynamics and Dynamic Programming with Message Passing. A detailed review can be found in [30].

Among those distributed methods, the Alternating Direction Method of Multipliers (ADMM) is a very commonly used distributed technique. The standard ADMM problem form is as follows:

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{subject to} \quad & Ax + By = c \end{aligned}$$

The augmented Lagrange function is as:

$$L_\rho(x, y, \lambda) = f(x) + g(y) + \lambda^T(Ax + By - c) + \frac{\rho}{2}\|Ax + By - c\|_2^2$$

The standard ADMM updating rule is as:

$$x^{k+1} \quad := \quad \arg\min_x L_\rho(x, y^k, \lambda^k) \tag{1.1}$$

$$y^{k+1} \quad := \quad \arg\min_y L_\rho(x^{k+1}, y, \lambda^k) \tag{1.2}$$

$$\lambda^{k+1} \quad := \quad \lambda^k + \rho(Ax^{k+1} + By^{k+1} - c) \tag{1.3}$$

$$\tag{1.4}$$

where x, y are two sets of variables with separable objective. By doing so, the problem is split into problem with x and problem with y.

1.3.3 Outline of Chapter 4

We consider a distributed scheme for a more general AC constrained optimal power flow problem. The resulting problem is a large-scale nonconvex problem. We make several contributions in this context.

(i) First, we reformulate the original AC-OPF model to a component-based distributable form with relatively simpler coupling constraints which leaves possibility for distributed scheme.

(ii) Second, we apply component based Alternative Direction Method of Multiplier (ADMM) framework for AC-OPF model in a distributed approach.

(iii) Third, we compare result from this scheme with centralized scheme and also observe the behavior of ADMM on different parameters.

Chapter 2 Risk-based economic dispatch

2.1 Introduction

The optimal power flow (OPF) problem is one of the most fundamental decision-making problems in power systems operations. There are a host of variants of such problems that include the modeling of reserves [10], allow for modeling security constraints [31], utilizing either DC approximations [32] or full AC formulations [33] of the power flow equations, amongst others [12]. In this chapter, we focus on the economic dispatch (ED) problem with a DC approximation of the power flow equations. With operating reserves and other regulation capacities determined in the day-ahead market [10, 11], economic dispatch decision are usually specified in a short amount of time at real time, with reserves and other regulation capacities are established. The economic dispatch of conventional generation is completed 20 minutes before the hour of delivery [9]. In power system operations, uncertainty plays a key role. Diverse formulations of stochastic optimal power flow along with different uncertainties in power system have been discussed in [8].

Most prior OPF formulations have only dealt with uncertainty in a rather rudimentary manner by choosing fixed reserve margins without using other known or estimated probabilistic information about forecast errors. Recently, one major approach to deal with such uncertainty has been through robust optimization techniques [13–17] where the uncertain parameters are assumed lie in a suitable uncertainty set and network constraints are enforced for every possible realization of uncertainty. In fact such avenues have been adopted for modeling AC power flow problems (cf. [18]).

Paper	Problem	Model	Algorithm
[34]	DC-OPF	Stochastic	`cplex`
[35]	UC	Robust	`cplex`
[36]	ED	Robust	`cplex`
[37]	ED	Robust	Alternative Direction
[38]	UC	Robust	`cplex`
[39]	UC	Robust	Cutting plane
[40]	UC	Stochastic	Review
[41]	UC	Stochastic	Benders Decomposition / Lagrangian Relaxation
[42]	UC	Stochastic	Lagrangian relaxation
[43]	UC	Stochastic	Benders Decomposition
[44]	UC	Stochastic	Dynamic Programming
[45]	UC	Stochastic	`cplex`
[46]	UC	Stochastic	`cplex`
[47]	UC	Stochastic	Importance Sampling
[48]	DC-OPF	Stochastic	Stochastic Decomposition
[33]	AC-OPF	Stochastic	Scenario reduction by clustering
[49]	DC-OPF	Stochastic	Benders Decomposition
[50]	ED	Deterministic	Gurobi
[51]	ED	Deterministic	`cplex`
[52]	ED	Deterministic	FESTIV

Table 2.1. Stochastic OPF models

Uncertainty in power systems also can be dealt with by adding chance constraints. In general, such avenues lead to possibly nonconvex problems [53] and more recently integer programming approaches have proven useful when contending with a sampled approximation [12, 19].

The authors considered power flow problems subject to chance constraints assuming that the uncertainties are Gaussian. The chance constraint can be expressed as a second-order cone constraint, which turn out to be a convex approximation. Simulation based studies on real time dispatch are also been conducted recently. In [20], a simulation based framework is used in a power system with renewable resources, with system iterates in multiple timescales.

A variety of literatures consider two-stage model in economic dispatch problem, with different sources of uncertainty. Single contingencies is taken into consideration and objective also includes cost from not meeting demand in [34], where in second stage a contingency parameter is added to transmission capacity in power flow constraint. In [54], model can be extended to muti-stage with unit commitment. A general stochastic dispatch model considering real time regulation in the presence of uncertainties in the offers was proposed in [55]. Constrain on second stage generation was in a form of a random convex set with several different forms, depending on the kind of offer involved. A two-stage economic dispatch model with stochastic producers was proposed in [56], where stochastic producer capacity constraints gets its realization in second stage. In [24], the network risk limiting dispatch problem (N-RLD) was introduced under $\epsilon = 0$, which is a two stage optimization problem under stochastic demand. On second stage, we observe the realization of random demand and balance system based on real demand. In [9], a two-stage model is built for sub-hourly dispatch decisions making. Ramping limits depending on time was also considered in [9]. In general, a two-stage stochastic model can represent the different types of uncertainty with realizations of uncertain factors reveal in second stage.

We make the following contributions in this paper. Our focus lies in solving the two-stage program in which a firm makes first-stage dispatch decisions based on forecast demand. Recourse decisions are made in the second-stage and are adapted to the realizations of uncertainty in demand and are guided by selecting first-stage decisions to minimize the sum of first-stage costs and risk-adjusted second-stage costs, where we utilize the CVaR measure to capture risk.

(I). Smoothing-based framework for CVaR-based two-stage problems. We develop a rigorous foundation for smoothing for the two-stage nonsmooth CVaR-based program that relies on providing the required regularity requirements. Notably, this allows for claiming that the CVaR-based risk-measure of second-stage satisfies (α, β) smoothability.

(II). Smoothed accelerated variance-reduced scheme for CVaR-based two-stage problems. To facilitate the application of smoothed accelerated variance reduced schemes, we analyze the bias and moment requirements of the sampled gradient of the smoothed risk-adjusted recourse function. This paves the way for formally applying the smoothed accelerated variance-reduced scheme for risk-adjusted two-stage programs. Importantly, this allows for claiming that this scheme admits a rate of convergence of $\mathcal{O}(1/k)$ while characterized by an optimal oracle complexity of $\mathcal{O}(1/\epsilon^2)$.

(III). Scenario generation and numerical studies. Finally, we apply the scheme on an IEEE 118 bus system. We utilize an autoregressive moving average model that allows for generating scenarios. Comparisons with competing schemes such as stochastic decomposition suggest the following. The proposed variance-reduced scheme provides solutions in a fraction of the time of similar accuracy.

2.2 Overview of economic dispatch problems

The basic economic dispatch problem requires satisfying load at minimal cost [57], as formulated next.

$$\begin{aligned} \min \quad & \sum_{i \in \mathcal{I}} c_i^g g_i \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}} g_i = d \\ & g_i^{\min} \leq g_i \leq g_i^{\max}, \quad \forall i \in \mathcal{I} \end{aligned} \tag{EDisp}$$

where g_i, g_i^{min}, g_i^{max} denote the generation level and the minimum, and maximum capacity level associated with generator i housed at bus i, c_i^g represents the unit cost of generation at bus i, d is the total demand, and $\mathcal{I}$ denotes the set of all buses in network. Notice that the economic dispatch problem requires specifying the minimal generation decisions while meeting demand and capacity bounds. In [57], power flow constraints are also taken into consideration together with dispatch decisions. When these power flow constraints are modeled via DC load flow approximations, the resulting bus-specific phase angles need to be considered together with transmission constraints. The resulting model is specified as follows. actual power production and transmission.

$$\begin{aligned} \min \quad & \sum_{i \in \mathcal{I}} c_i^g g_i \\ \text{subject to} \quad & g_i - d_i - \sum_{j \in \mathcal{I}} B_{i,j}(\theta_i - \theta_j) = 0, && \forall i \in \mathcal{I} \\ & B_{i,j}(\theta_i - \theta_j) \leq f_{i,j}^{max}, && \forall i, j \in \mathcal{I} \\ & g_i^{min} \leq g_i \leq g_i^{max}, && \forall i \in \mathcal{I} \\ & \theta_i^{min} \leq \theta_i \leq \theta_i^{max}, && \forall i \in \mathcal{I} \end{aligned} \tag{2.1}$$

Where θ_i denotes the phase angle at bus i, $f_{i,j}^{\max}$ represents the transmission line capacity constraint of transmission line between bus i, j, $B_{i,j}$ is the susceptance of transmission line between bus i, j, and d_i is the demand at bus i. Single contingencies may be taken into consideration the cost of unserved demand may also be modeled [34], leading to the following model.

$$
\begin{aligned}
\min \quad & \mathbb{E}\left[\sum_{i\in\mathcal{I}} \left(c_i^g g_i^\omega + c_i^r r_i^\omega - c_i^d q_i^\omega\right)\right] \\
\text{subject to} \quad & g_i^\omega - q_i^\omega - \sum B_{i,j}(\theta_i^\omega - \theta_j^\omega) = 0, && \forall i \in \mathcal{I}, \forall \omega \in \Omega \\
& B_{i,j}(\theta_i^\omega - \theta_j^\omega) \le f_{i,j}^{\max}, && \forall i \in \mathcal{I}, \forall \omega \in \Omega \\
& g_i^\omega + r_i^\omega - \chi_i^\omega g_i^{\max} = 0, && \forall i \in \mathcal{I}, \forall \omega \in \Omega \\
& \lambda_i d_i \le q_i^\omega \le d_i, && \forall i \in \mathcal{I}, \forall \omega \in \Omega
\end{aligned}
\tag{2.2}
$$

where r_i^ω and c^r denote reserve levels and the cost of reserves at bus i, q_i^ω and c_i^d represents unserved demand and the cost of unserved demand at bus i, Ω represents the scenario space, χ_i^ω denotes the proportion that generator i under outage would reduce in capacity by (i.e. the capacity of outage generator would be $\chi_i^\omega g_i^{\max}$ where $\chi_i^\omega \in [0,1]$), d_i represents the forecast demand, and $\lambda_i d_i$ denotes a critical level of demand that needs to be satisfied. In [54], this model is further extended to a multi-stage problems with unit commitment. A more general stochastic dispatch model was proposed in [55]

$$
\begin{aligned}
\min \quad & \sum_{i\in\mathcal{I}} (c_i g_i + \mathbb{E}[c_i^+ (g_i^\omega - g_i)_+ - c_i^- (g_i^\omega - g_i)_-]) \\
\text{subject to} \quad & g_i - \sum_{j\in\mathcal{I}} f_{i,j} - d_i = 0, && \forall i \in \mathcal{I} \\
& g_i^\omega - \sum_{j\in\mathcal{I}} f_{i,j}^\omega - d_i = 0, && \forall i \in \mathcal{I}, \forall \omega \in \Omega \\
& (g_i, g_i^\omega) \in C_i^\omega, && \forall \omega \in \Omega, \forall i \in \mathcal{I} \\
& f_{i,j} \le f_{i,j}^{\max}, && \forall i \in \mathcal{I}, \\
& f_{i,j}^\omega \le f_{i,j}^{\max}, && \forall i \in \mathcal{I}, \forall \omega \in \Omega.
\end{aligned}
$$

In this model, g_i^ω denotes the second-stage ("real time") decision $f_{i,j}$ and $f_{i,j}^\omega$ represents the first and scenario-specific second-stage power flow on the transmission line, and C_i^ω represents a random convex set that may take several different forms, depending on the kind of offer involved.

- *Completely inflexible generation*: First-stage dispatched quantity, denoted by

g_i cannot be varied in the second stage.

$$g_i^\omega = g_i \in [0, g_i^{\max}], \quad \forall \omega \in \Omega, \forall i \in \mathcal{I}.$$

- *Completely flexible generation*: First-stage dispatched quantity, denoted by g_i, may be varied in the second stage.

$$g_i \in [0, g_i^{\max}],\ g_i^\omega \in [0, g_i^{max}], \quad \forall \omega \in \Omega, \forall i \in \mathcal{I}.$$

- *Unpredictable or intermittent generation*: A generator with maximum capacity $g_i^{\max}$ offers a random quantity S_i.

$$g_i \in [0, g_i^{\max}],\ g_i^\omega \in [0, S_i^\omega], \quad \forall \omega \in \Omega, \forall i \in \mathcal{I}$$

- *Demand-side bid*: A quantity $-q_i \geq 0$ is bid for in the first stage while in the second-stage, this can be modified.

$$g_i \in [q_i, 0],\ g_i^\omega \in [q_i, 0], \quad \forall \omega \in \Omega, \forall i \in \mathcal{I}$$

- *Unpredictable load*: The second-stage demand-side bid g_i^ω has to be feasible with regard to a random load of size $d_i^\omega \geq 0$.

$$g_i \leq 0,\ g_i^\omega \in [-d_i^\omega, 0], \quad \forall \omega \in \Omega, \forall i \in \mathcal{I}.$$

In [21], a risk-limiting dispatch framework is introduced and this was subsequently extended in [22], where a computable closed-form formulas was derived. In [23], a related model for risk-limiting dispatch with generation limitation and network constraint was represented. In [24], a *two-stage* network risk limiting dispatch problem (N-RLD) was introduced under $\epsilon = 0$, where the first-stage represents the day-ahead scheduling while the second-stage captures real-time decisions. The overall problem requires minimizes expected cost of operation as captured by the following problem.

2.3 Two-stage stochastic economic dispatch

In the two-stage model for stochastic economic dispatch, first-stage decisions are given by slow-response generation decisions while second-stage decisions adapt to the realization of uncertainty and are tied to first-stage decisions. The nature of the uncertainty in the second-stage pertains to the randomness in real-time cost, randomness in real-time demand, line contingencies, and uncertainty in the availability of renewable The goal of two-stage model is to have determine a set of first decisions that minimize the sum of two costs: (i) the first-stage cost; and (ii) the risk-adjusted second-stage expected cost of contending with uncertainty.

We consider the stochastic economic dispatch problem faced over T hours while sub-hourly decisions associated with fast-response generators are made between hours (i.e. in a sub-hourly sense) to contend with uncertainty. We view the first-stage decisions as the hourly decisions from $t = 1, \ldots, T$ while the sub-hourly decisions are viewed as the recourse second-stage decisions Let g_t denote the first stage decisions at period t with a convex differentiable generation cost denoted by $f_0(g_t)$ while the cost of recourse decisions under realization ω in the sub-hourly period after t is denoted by $h(g_t^\omega, \omega)$.

2.3.1 Risk-neutral stochastic economic dispatch

In a power system network, there are several types of constraints that need consideration.

Flow balance equations

In both stages, flow balance needs to be maintained at each bus for every period. In the first stage, flow balance would be based on forecasted demand as follows:

$$g_t - B_1\theta_t - \hat{d}_t = 0,\ t = 1, \ldots, T, \tag{2.3}$$

where θ_t denotes the phase angle of each bus at period t, $B = (b_{i,j})_{n\times n}$ represents the susceptance matrix,

$$B_1 \triangleq B - \operatorname{diag}\left(\sum_j b_{i,j}\right),$$

and $\hat{d}_t$ denotes a forecast demand in period t in first stage. During the sub-hourly dispatch after period t, the recourse decisions have to satisfy the following requirement.

$$g_t + g^r_{t,\omega} - B_1\theta_{t,\omega} - d_{t,\omega} - d^u_{t,\omega} - d^w_{t,\omega} \geq 0, \tag{2.4}$$

where $g^r_{t,\omega}$ denotes the recourse generation adjustment in period t, $\theta_{t,\omega}$ represents the second-stage phase angle for period t, and $d_{t,\omega}$ denotes the observed demand in period t under realization ω. Suppose $d^u_{t,\omega}$ represents undispatchable generation while $d^w_{t,\omega}$ denotes unavailable wind power generation for period t.

Line flow constraints

In both stages, transmission capacity constraints need to be satisfied while line contingencies are considered in the second stage. The first stage line constraints can be expressed as follows:

$$-f_{\max} \leq B_2\theta_t \leq f_{\max}, t = 1, \dots, T \tag{2.5}$$

where $f_{\max}$ denotes flow bounds while B_2 is defined as follows.

$$B_2 = \begin{pmatrix} \begin{pmatrix} b_{1,\bullet} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} - \mathrm{diag}(b_{1,\bullet}) \\ \begin{pmatrix} \mathbf{0} & b_{2,\bullet} & \cdots & \mathbf{0} \end{pmatrix} - \mathrm{diag}(b_{2,\bullet}) \\ \vdots \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & b_{n,\bullet} \end{pmatrix} - \mathrm{diag}(b_{n,\bullet}) \end{pmatrix}.$$

In the second stage, line contingencies are addressed as follows:

$$-\chi_{t,\omega} f_{\max} \leq B_2\theta_{t,\omega} \leq \chi_{t,\omega} f_{\max}, \ \forall t, \tag{2.6}$$

where $\chi_{t,\omega}$ is a vector represents the realization of stochastic line contingency of all transmission lines for period t with all of its elements value in $[0, 1]$.

Generation capacity constraints

Both conventional and renewable generators we have to abide by capacity constraints.

$$g^{\min} \leq g_t \leq g^{\max} \forall t \tag{2.7}$$

$$g^{\min}_{t,\omega} \leq g_t + g^r_{t,\omega} \leq g^{\max}_{t,\omega}, \tag{2.8}$$

where $g^{\min}_{t,\omega}$ and $g^{\max}_{t,\omega}$ denote second-stage capacity bounds at period t based on renewable availability.

Ramping constraints

Both conventional and renewable generators need to satisfy ramping requirements, as specified next.

$$r^{\min}_t \leq g_t - g_{t-1} \leq r^{\max}_t, \ \forall t \tag{2.9}$$

$$r^{\min}_{t,\omega} \leq g_t + g^r_{t,\omega} - g_{t-1} - g^r_{t-1,\omega} \leq r^{\max}_{t,\omega}, \tag{2.10}$$

where $r^{\min}$ and $r^{\max}$ denote the minimum and maximum ramp limit, $g_0 = 0$, and $r^{\min}_{t,\omega}$ and $r^{\max}_{t,\omega}$ denote second-stage down and up ramping limits for period t and $g^r_{0,\omega} = 0$.

Phase angle bounds

In both stages, phase angles bound are imposed in the following fashion.

$$\theta_{\min} \leq \theta_t \leq \theta_{\max}, \ \forall t \tag{2.11}$$

$$\theta_{\min} \leq \theta_{t,\omega} \leq \theta_{\max}. \tag{2.12}$$

To summarize, the two-stage stochastic economic dispatch model is defined as follow:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{t=1}^{T} [f_0(\mathbf{x}_t) + \mathbb{E}[\mathcal{Q}(\mathbf{x}_t, \omega)]] \\ \text{subject to} \quad & (2.3), (2.5), \ldots, (2.11). \end{aligned} \tag{s-ED}$$

where $\mathbf{x} \triangleq (g, \theta)$, $\mathcal{Q}(\mathbf{x}_t, \omega)$ is defined as follows.

$$\begin{aligned} \mathcal{Q}(\mathbf{x}_t, \omega) = \min_{g^r_{t,\omega}, \theta_{t,\omega}} \quad & f(g^r_{t,\omega}, \omega) \\ \text{subject to} \quad & (2.4), (2.6), \ldots, (2.12), \end{aligned} \tag{s-ED$_2^\omega$}$$

$f(g^\omega_t, \omega)$ denotes the random second-stage generation cost and $g^r_{t,\omega}$ denotes second stage decision variable at period t under random variable realization ω.

Deterministic equivalent

Suppose the ω takes on a finite number of realizations given by $\omega_1, \ldots, \omega_K$ with probabilities $p_1, \ldots, p_K$. The resulting deterministic equivalent optimization problem is given by the following.

$$\begin{aligned} \min_{\mathbf{x}_t, \mathbf{y}_{t,\omega}} \quad & \sum_{t=1}^{T} \left[f_0(\mathbf{x}_t) + \sum_{j=1}^{K} p^{\omega_j} [f(\mathbf{y}_{t,\omega}, \omega_j)] \right] \\ \text{subject to} \quad & (2.3), (2.5), \ldots, (2.11) \\ & (2.4), (2.6), \ldots, (2.12), \qquad \forall t, \omega \in \Omega. \end{aligned} \tag{DE-ED}$$

where $\mathbf{y}_\omega \triangleq (g^r_\omega, \theta_\omega)$. Generally, (DE-ED) is not solved directly since Ω can have a large cardinality, necessitating the development of decomposition and sampling schemes.

2.3.2 Risk-averse economic Dispatch

While the prior model employs a risk-neutral framework, this can be generalized to allow for risk preferences; a commonly employed approach utilizes the conditional value-at-risk (CVaR). Recall for a fixed level τ, the conditional value-at-risk of a random loss function $Z(\omega)$ is defined as [58]:

$$\mathbf{CVaR}_\tau(Z(\omega)) \triangleq \min_m \left\{ m + \frac{1}{1-\tau} \mathbb{E}\left[Z(\omega) - m\right]^+ \right\}. \tag{2.13}$$

In this setting, we choose $Z(\omega) \triangleq \mathcal{Q}(\mathbf{x}_t, \omega)$. It may be recalled that the minimization of a CVaR-based objective can be recast as follows:

$$
\begin{aligned}
&\min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \mathbf{CVaR}_\tau(\mathcal{Q}(\mathbf{x},\omega)) \right\} \\
&= \min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \min_{\mathbf{m}} \left\{ \mathbf{m} + \frac{\mathbf{1}}{\mathbf{1}-\tau} \mathbb{E}\left[\mathcal{Q}(\mathbf{x},\omega) - \mathbf{m}\right]^+ \right\} \right\} \\
&= \min_{\mathbf{x},\mathbf{m}} \left\{ \underbrace{f_0(\mathbf{x}) + \mathbf{m}}_{\triangleq c(\mathbf{z})} + \frac{1}{1-\tau} \mathbb{E}\left[\mathcal{Q}(\mathbf{x},\omega) - \mathbf{m}\right]^+ \right\} \\
&= \min_{\mathbf{z}} \left\{ c(\mathbf{z}) + \mathbb{E}\left[r(\mathbf{z},\omega)\right] \right\}, \text{ where } r(\mathbf{z},\omega) \triangleq \frac{1}{1-\tau} \left[\mathcal{Q}(\mathbf{x},\omega) - \mathbf{m}\right]^+ . \qquad (2.14)
\end{aligned}
$$

In the next section, we discuss a smoothed accelerated gradient scheme for the two-stage stochastic convex problem:

$$
\min_{\mathbf{z} \in Z} \left[c(\mathbf{z}) + \mathbb{E}\left[r(\mathbf{z},\omega)\right] \right], \qquad \text{(r-ED)}
$$

where $\mathbf{z} \triangleq (\mathbf{x}, \mathbf{m})$, $c(\mathbf{z})$ is a convex differentiable function, $r(\mathbf{z},\omega)$ is a convex and nonsmooth function defined as (2.14), and Z is a polyhedral set defined by the constraints (2.3), (2.5), …, (2.11).

2.4 Smoothing

Consider the function $r(\mathbf{z},\omega)$, defined in (2.14). This function has two sources of nonsmoothness: (i) The function $\mathcal{Q}(\bullet,\omega)$ is a convex nonsmooth function of $(\bullet)$; and (ii) The function $[u]^+ = \max\{u, 0\}$ is a nonsmooth function of u. We intend to develop algorithms in which the gradient of a smoothed counterpart of $r(\mathbf{z},\omega)$, denoted by $r_\mu(\mathbf{z},\omega)$, is employed. Before proceeding, we recap the definition of smoothability of a convex function.

Definition 2.4.1 (Smoothable function [59]). *A convex function $d : \mathbb{R}^n \to \mathbb{R}$ is called (α,β)-smoothable if for any $\mu > 0$ there exists a convex differentiable function $d_\mu : \mathbb{R}^n \to \mathbb{R}$ such that the following holds for some $\alpha, \beta > 0$:*

(i) $d_\mu(x) \leq d(x) \leq d_\mu(x) + \beta\mu$ for all $x \in \mathbb{R}^n$.

(ii) d_μ is $\frac{\alpha}{\mu}$-smooth.

Then the function d_μ is called a $\frac{1}{\mu}$-smooth approximation of h with parameters (α, β).

2.4.1 Smoothing the recourse function $\mathcal{Q}(g, \omega)$.

Consider the recourse function $\mathcal{Q}(\mathbf{x}, \omega)$ whose evaluation requires solving the following parametrized convex problem.

$$\begin{aligned} \min_{\mathbf{y} \in Y^\omega} \quad & d(\mathbf{y}, \omega) \\ \text{subject to} \quad & W_\omega \mathbf{y} \le q_\omega - T_\omega \mathbf{x}, \qquad (\pi) \end{aligned} \tag{P-$\mathcal{Q}(\mathbf{x},\omega)$}$$

We make the following assumptions on (P-$\mathcal{Q}(\mathbf{x}, \omega)$).

Assumption 2.4.1. *Consider the recourse problem (P-$\mathcal{Q}(\mathbf{x}, \omega)$).*
(i) For every $\omega \in \Omega$, the function $d : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ is convex and Y^ω is a closed, convex, and bounded polyhedron with a nonempty interior.
(ii) For every $\omega \in \Omega$ and every $x \in X_0$, there exists a $\bar{\mathbf{y}}(\omega)$ such that $W_\omega \mathbf{y}_\omega = q_\omega - T_\omega \mathbf{x}$ and $\bar{\mathbf{y}}_\omega \in int(Y^\omega)$.

The dual of (P-$\mathcal{Q}(\mathbf{x}, \omega)$) is given by the following.

$$\max_{\pi \ge 0} \quad \underbrace{(q^\omega - T^\omega \mathbf{x})^T \pi - \bar{d}^*(W_\omega^T \pi; \omega)}_{\triangleq \varepsilon(\pi; \mathbf{x}, \omega)}, \tag{D-$\mathcal{Q}(\mathbf{x},\omega)$}$$

where $\bar{d}^*(\mathbf{y}; \omega)$ is the convex conjugate of $\bar{d}(y; \omega)$, defined as follows.

$$\bar{d}(\mathbf{y}; \omega) \triangleq \begin{cases} d(\mathbf{y}; \omega), & \mathbf{y} \in Y^\omega \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.15}$$

By convex duality, the optimal values of (P-$\mathcal{Q}(\mathbf{x}, \omega)$) and (D-$\mathcal{Q}(\mathbf{x}, \omega)$) are equal for a given $\mathbf{x}$.

Lemma 2.4.1. *Suppose $d(\mathbf{y}, \omega)$ is a convex function in $\mathbf{y}$ for every $\omega \in \Omega$. If for some $\mathbf{x}$, (P-$\mathcal{Q}(\mathbf{x}, \omega)$) has an optimal solution. Then the dual problem (D-$\mathcal{Q}(\mathbf{x}, \omega)$) has an optimal solution and the optimal values of both problems are equal.*

From the theory of stochastic programming, it is known that the recourse function $\mathcal{Q}(\mathbf{x}, \omega)$ is a convex function in $\mathbf{x}$ for every ω [60, Prop. 2.21]. Consider a modified function $\mathcal{Q}_\mu(\mathbf{x},\omega)$, defined as the optimal value of this μ-regularized dual problem (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$).

$$\max_{\pi\geq 0}\ (\varepsilon(\pi;\mathbf{x},\omega)-\mu\|\pi\|^2), \qquad \text{(D-}\mathcal{Q}_\mu(\mathbf{x},\omega))$$

Lemma 2.4.2. *Suppose v_ω and $v_{\mu,\omega}$ denote the optimal values of $(D\text{-}\mathcal{Q}(\mathbf{x};\omega))$ and $(D\text{-}\mathcal{Q}_\mu(\mathbf{x};\omega))$, respectively. In addition, suppose π_ω denotes an optimal solution of $(D\text{-}\mathcal{Q}(\mathbf{x};\omega))$. Then $v_{\mu,\omega}\geq v_\omega-\mu\|\pi_\omega\|^2$.*

Proof. The result can be concluded as follows.

$$\begin{aligned}
v_{\mu,\omega} &\triangleq \max_{\pi\geq 0}\ \{(\varepsilon(\pi;\mathbf{x},\omega)-\mu\|\pi\|^2)\}\\
&\geq\ \left\{(\varepsilon(\pi^*_\omega;\mathbf{x},\omega)-\mu\|\pi^*_\omega\|^2)\right\} \qquad (\text{where } \pi^*_\omega\in\arg\max_{\pi\geq 0}\{\varepsilon(\pi;\mathbf{x},\omega)\})\\
&=\ \left\{\max_{\pi\geq 0}\varepsilon(\pi;\mathbf{x},\omega)\right\}-\mu\|\pi^*_\omega\|^2\\
&= v_\omega-\mu\|\pi^*_\omega\|^2.
\end{aligned}$$

□

Before proceeding, we show that the solution of (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$) is bounded under the Slater regularity condition. This requires defining the Lagrangian function $\mathcal{L}(\mathbf{y},\pi,\omega)$ and the dual function $\mathcal{D}(\pi,\omega)$ associated with the primal problem, which we proceed to do next.

$$\mathcal{L}(\mathbf{y},\pi,\omega)\triangleq\left(d(\mathbf{y},\omega)+(W_\omega\mathbf{y}-q_\omega+T_\omega x)^T\pi\right) \tag{2.16}$$

$$\mathcal{D}(\pi,\omega)\triangleq\min_{\mathbf{y}}\ \mathcal{L}(\mathbf{y},\pi,\omega). \tag{2.17}$$

Lemma 2.4.3. *Consider the problem $(P\text{-}\mathcal{Q}(\mathbf{x},\omega))$ and suppose it has optimal value v_ω. Then the following hold.*

(i) Suppose for given $\mathbf{x}$ and $\omega\in\Omega$, there exists a $\bar{\mathbf{y}}\in Y^\omega$ such that $c(\bar{\mathbf{y}},\mathbf{x})<0$ where $c(\bar{\mathbf{y}},\mathbf{x})\triangleq W_\omega\bar{\mathbf{y}}-q_\omega+T_\omega\mathbf{x}$. Then the solution set of (D-$\mathcal{Q}(\mathbf{x},\omega)$) lies in a compact set.

(ii) In addition, if $\bar{\mathbf{y}}\in\cap_{\omega\in\Omega}Y^\omega$, $-c(\bar{\mathbf{y}},\omega)\geq-\bar{c}$ and $d(\bar{\mathbf{y}},\omega)-v_\omega\leq\bar{d}$ for every $\omega\in\Omega$, then the solution set of $(D\text{-}\mathcal{Q}(\mathbf{x},\omega))$ is uniformly bounded in ω, i.e. $\|\pi_\omega\|\leq\frac{\bar{d}}{\bar{c}}$ for any $\omega\in\Omega$ and for any $\pi_\omega\in SOL(D\text{-}\mathcal{Q}(\mathbf{x},\omega))$.

(iii) If $\pi_{\mu,\omega}$ denotes the optimal dual solution to (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$), then we have that

$$\|\pi_{\mu,\omega}\| \leq \frac{d(\bar{\mathbf{y}},\omega) - v_\omega}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)} + \mu\left(\frac{d(\bar{\mathbf{y}},\omega) - v_\omega}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)}\right)^2.$$

(iv) In addition, if $\bar{\mathbf{y}} \in \cap_{\omega\in\Omega} Y^\omega$, $-c_j(\bar{\mathbf{y}},\omega) \geq -\bar{c}$ for $j = 1,\ldots,m$ and $d(\bar{\mathbf{y}},\omega) - v_\omega \leq \bar{d}$ for every $\omega \in \Omega$, then the solution set of (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$ is uniformly bounded in ω and μ when $\mu \leq \bar{\mu}$, i.e. $\|\pi_{\mu,\omega}\| \leq \frac{\bar{d}}{\bar{c}} + \bar{\mu}\left(\frac{\bar{d}}{\bar{c}}\right)^2$ for any $\omega \in \Omega$, where $\pi_{\mu,\omega} = \mathrm{argmax}_{\pi\geq 0}(D\text{-}\mathcal{Q}_\mu(\mathbf{x},\omega))$.

Proof. (i) For any optimal dual solution π^*_ω, we have from strong duality,

$$\begin{aligned} v_\omega = \mathcal{D}(\pi^*_\omega,\omega) &= \inf_{\mathbf{y}\in Y^\omega} \{d(\mathbf{y},\omega) + (\pi^*_\omega)^T c_\omega(\mathbf{y};\mathbf{x})\} \\ &\leq\ d(\bar{\mathbf{y}},\omega) + (\pi^*_\omega)^T c_\omega(\bar{\mathbf{y}};\mathbf{x}) \\ &\leq d(\bar{\mathbf{y}},\omega) + \max_{1\leq j\leq m} c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x}) \sum_{j=1}^m \pi^*_{\omega,i}. \end{aligned}$$

Consequently, we have the following relationship.

$$-(\max_{1\leq j\leq m} c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x})) \sum_{j=1}^m \pi^*_{\omega,i} \leq d(\bar{\mathbf{y}},\omega) - v_\omega \tag{2.18}$$

$$\implies \|\pi^*_\omega\| \leq \sum_{j=1}^m \pi^*_{\omega,i} \leq \frac{d(\bar{\mathbf{y}},\omega) - v_\omega}{-\max_{1\leq j\leq m}(c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x}))}. \tag{2.19}$$

(ii) By hypothesis, we have that $-\max_{1\leq j\leq m}(c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x})) > \bar{c}$ and $d(\bar{\mathbf{y}},\omega) - v_\omega \leq \bar{d}$ for every $\omega \in \Omega$. Consequently, $\|\pi^*_\omega\| \leq \frac{\bar{d}}{\bar{c}}$ for all $\omega \in \Omega$.

(iii) Consider the regularized dual function $\mathcal{D}_\mu(\pi)$ and suppose its optimal value is $v_{\mu,\omega}$. Then the following sequence of inequalities hold.

$$\begin{aligned} &v_{\mu,\omega} = \mathcal{D}_\mu(\pi^*_{\omega,\mu}) \\ &= \inf_{\mathbf{y}\in Y^\omega} \{d(\mathbf{y},\omega) + (\pi^*_{\mu,\omega})^T c_\omega(\mathbf{y};\mathbf{x}) - \mu\|\pi^*_{\mu,\omega}\|^2\} \\ &\leq\ \{d(\mathbf{y},\omega) + (\pi^*_{\mu,\omega})^T c_\omega(\mathbf{y};\mathbf{x}) - \mu\|\pi^*_{\mu,\omega}\|^2\} \\ &\leq\ d(\mathbf{y},\omega) + (\pi^*_{\mu,\omega})^T c_\omega(\bar{\mathbf{y}};\mathbf{x}) \\ &\leq d(\mathbf{y},\omega) + \max_{1\leq j\leq m} c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x}) \sum_{j=1}^m \pi^*_{\mu,\omega,i}. \end{aligned}$$

It follows that

$$\begin{aligned}\|\pi^*_{\mu,\omega}\| &\leq \frac{d(\bar{\mathbf{y}},\omega) - v_{\mu,\omega}}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)} \\ &\stackrel{\text{Lemma 2.4.1}}{\leq} \frac{d(\bar{\mathbf{y}},\omega) - v_{\omega} + \mu\|\pi^*_{\omega}\|^2}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)} \\ &\stackrel{(2.19)}{\leq} \frac{d(\bar{\mathbf{y}},\omega) - v_{\omega}}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)} + \mu\left(\frac{d(\bar{\mathbf{y}},\omega) - v_{\omega}}{-\max_{1\leq j\leq m} c_j(\bar{\mathbf{y}};\omega)}\right)^2.\end{aligned}$$

(iv) By hypothesis, we have that $\min_{1\leq j\leq m}(-c_{\omega,j}(\bar{\mathbf{y}};\mathbf{x})) > \bar{c}$ and $d(\bar{\mathbf{y}},\omega) - v_\omega \leq \bar{d}$ for every $\omega \in \Omega$. Consequently, $\|\pi^*_{\mu,\omega}\| \leq \frac{\bar{d}}{\bar{c}} + \bar{\mu}\left(\frac{\bar{d}}{\bar{c}}\right)^2$ for all $\omega \in \Omega$ and for every $\mu \leq \bar{\mu}$.

□

We proceed to show that $\mathcal{Q}_\mu(\mathbf{x},\omega)$ is an (α,β)-smoothing of $\mathcal{Q}(\mathbf{x},\omega)$, where $\mathcal{Q}_\mu(\mathbf{x},\omega)$ which is defined as the optimal value of (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$).

Lemma 2.4.4. *Suppose $\mathcal{Q}(\mathbf{x},\omega)$ is defined by the optimal value of (P-$\mathcal{Q}(\mathbf{x},\omega)$). Then the following hold:*
(i) The function $\mathcal{Q}(\mathbf{x},\omega)$ is a convex function in $\mathbf{x}$ for every $\omega \in \Omega$.
(ii) The function $\mathcal{Q}_\mu(\mathbf{x},\omega)$ is a differentiable in $\mathbf{x}$ at every ω and $\nabla_{\mathbf{x}}\mathcal{Q}_\mu(\mathbf{x},\omega) = -T_\omega^T\pi^*(\mathbf{x},\omega)$, where $\pi^*(\mathbf{x},\omega)$ denotes the optimal solution of (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$).

Proof. [60, Prop. 2.22]. □

Proposition 2.4.1 ($\mathcal{Q}_\mu(\mathbf{x},\omega)$ satisfies $(\alpha(\omega),\beta(\omega))$-smoothability of $\mathcal{Q}(\mathbf{x},\omega)$)**.** *Consider the function $\mathcal{Q}_\mu(\mathbf{x},\omega)$ defined by (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$). Then this function satisfies the following:*
(i) The function $\mathcal{Q}_\mu(\mathbf{x},\omega)$ is $\frac{\|T_\omega\|^2}{\mu}$-smooth, i.e.

$$\|\nabla_{\mathbf{x}}\mathcal{Q}_\mu(\mathbf{x}_1,\omega) - \nabla_{\mathbf{x}}\mathcal{Q}_\mu(\mathbf{x}_2,\omega)\| \leq \frac{\|T_\omega\|^2}{\mu}\|\mathbf{x}_1 - \mathbf{x}_2\|, \qquad \forall \mathbf{x}_1, \mathbf{x}_2.$$

(ii) There exists a $\beta(\omega)$ such that for all $\mathbf{x}$, we have that for all $\mathbf{x}$,

$$\mathcal{Q}_\mu(\mathbf{x},\omega) \leq \mathcal{Q}(\mathbf{x},\omega) \leq \mathcal{Q}_\mu + \mu\beta(\omega).$$

Proof. (i) Consider an $\mathbf{x}_1, \mathbf{x}_2 \in X$ and let $\pi(\mathbf{x}_1)$ and $\pi(\mathbf{x}_2)$ denote the maximizers of (D-$\mathcal{Q}_\mu(\mathbf{x}_1, \omega)$) and (D-$\mathcal{Q}_\mu(\mathbf{x}_2, \omega)$), respectively. By the strong concavity of the objective, we have that

$$\begin{aligned}
&((-q_\omega + T_\omega \mathbf{x}_1 + a(\pi(\mathbf{x}_1), \omega)) - (-q_\omega + T_\omega \mathbf{x}_2 + a(\pi(\mathbf{x}_2), \omega)))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)) \\
&+ \mu(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)) \\
&\geq \mu\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\|^2, \qquad (2.20)
\end{aligned}$$

where $a(\pi(\mathbf{x}), \omega) \in \partial_\pi(\bar{d}^*_\omega(W_\omega^T \pi; \omega))$. In addition, by definition, we have that

$$(T_\omega \mathbf{x}_1 - q_\omega + a(\pi(\mathbf{x}_1), \omega) + \mu\pi(\mathbf{x}_1))^T(\pi(\mathbf{x}_2) - \pi(\mathbf{x}_1)) \geq 0. \qquad (2.21)$$

Adding (2.20) and (2.21), we obtain that

$$\begin{aligned}
&(T_\omega \mathbf{x}_2 - q_\omega + a(\pi(\mathbf{x}_2), \omega) + \mu\pi(\mathbf{x}_2))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)) \\
&\geq \mu\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\|^2. \qquad (2.22)
\end{aligned}$$

Consequently, by adding and subtracting $(T_\omega \mathbf{x}_1 - q_\omega + a(\pi(\mathbf{x}_1, \omega) + \mu\pi(\mathbf{x}_1))^T$ $(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2))$,

$$\begin{aligned}
&\underbrace{(T_\omega \mathbf{x}_1 + a(\pi(\mathbf{x}_1, \omega) + \mu\pi(\mathbf{x}_1))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2))}_{\leq 0} \\
&+ (T_\omega \mathbf{x}_2 - T_\omega \mathbf{x}_1)^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)) + \underbrace{(a(\pi(\mathbf{x}_2), \omega) - a(\pi(\mathbf{x}_1), \omega))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2))}_{\leq\ 0} \\
&+ \mu\underbrace{(\pi(\mathbf{x}_2) - \mu\pi(\mathbf{x}_1))^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2))}_{\leq\ 0} \\
&\geq \mu\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\|^2.
\end{aligned}$$

This implies that

$$\begin{aligned}
\mu\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\|^2 &\leq (T_\omega \mathbf{x}_1 - T_\omega \mathbf{x}_2)^T(\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)) \\
&\leq \|T_\omega\|\|\mathbf{x}_1 - \mathbf{x}_2\|\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\| \\
\implies \|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\| &\leq \frac{\|T_\omega\|}{\mu}\|\mathbf{x}_1 - \mathbf{x}_2\|.
\end{aligned}$$

Finally, we note that $\nabla_x \mathcal{Q}_\mu(\mathbf{x},\omega) = -T_\omega^T \pi_\omega$, where π_ω is a maximizer of (D-$\mathcal{Q}(\mathbf{x},\omega)$), implying that

$$\begin{aligned}\|\nabla_{\mathbf{x}} \mathcal{Q}_\mu(\mathbf{x}_1,\omega) - \nabla_{\mathbf{x}} \mathcal{Q}_\mu(\mathbf{x}_2,\omega)\| &\leq \|T_\omega\|\|\pi(\mathbf{x}_1) - \pi(\mathbf{x}_2)\| \\ &\leq \frac{\|T_\omega\|^2}{\mu}\|\mathbf{x}_1 - \mathbf{x}_2\|.\end{aligned}$$

(ii) We begin by noting that $\mathcal{Q}(\mathbf{x},\omega) = (q^\omega - T_\omega \mathbf{x})^T \pi - \bar{d}^*(W^T\pi,\omega)$ where π is a maximizer of (D-$\mathcal{Q}(\mathbf{x},\omega)$) while $\mathcal{Q}_\mu(\mathbf{x},\omega) = (q^\omega - T_\omega \mathbf{x})^T \pi_\mu - \bar{d}^*(W^T\pi_\mu,\omega) - \frac{1}{2}\|\pi_\mu\|^2$ where π_μ is a maximizer of (D-$\mathcal{Q}_\mu(\mathbf{x},\omega)$). Consequently, we have that

$$\begin{aligned}\mathcal{Q}(\mathbf{x},\omega) &= (q^\omega - T_\omega \mathbf{x})^T \pi - \bar{d}^*(W^T\pi,\omega) \\ &\geq (q^\omega - T_\omega \mathbf{x})^T \pi_\mu - \bar{d}^*(W^T\pi_\mu,\omega) \\ &\geq (q^\omega - T_\omega \mathbf{x})^T \pi_\mu - \bar{d}^*(W^T\pi_\mu,\omega) - \tfrac{1}{2}\mu\|\pi_\mu\|^2 \\ &= \mathcal{Q}_\mu(\mathbf{x},\omega).\end{aligned}$$

In addition, it is easily seen that

$$\begin{aligned}\mathcal{Q}_\mu(\mathbf{x},\omega) &= (q^\omega - T_\omega \mathbf{x})^T \pi_\mu - \bar{d}^*(W^T\pi_\mu,\omega) - \tfrac{1}{2}\mu\|\pi_\mu\|^2 \\ &\geq (q^\omega - T_\omega \mathbf{x})^T \pi - \bar{d}^*(W^T\pi,\omega) - \tfrac{1}{2}\mu\|\pi\|^2 = \mathcal{Q}(\mathbf{x},\omega) - \tfrac{1}{2}\mu\|\pi\|^2.\end{aligned}$$

As a result, we have that

$$\mathcal{Q}(\mathbf{x},\omega) \leq \mathcal{Q}_\mu(\mathbf{x},\omega) + \tfrac{1}{2}\mu\|\pi\|^2 \leq \mathcal{Q}_\mu(\mathbf{x},\omega) + \mu\beta(\omega),$$

where $\|\pi\|^2 \leq \beta(\omega)$ for all π, where the boundedness of π follows from the Slater regularity condition on (P-$\mathcal{Q}(\mathbf{x},\omega)$). □

2.4.2 Smoothing the max. function

From [59], recall that the smoothing of the max. function, defined as $t(u) \triangleq [u]^+$, is given by $t_\mu(x) \triangleq \mu(\log(e^{\frac{u}{\mu}} + 1) - \log(2))$ and t_μ is $(1, \log(2))$-smoothable. We prove the relatively simple result that $t'_\mu(u) \leq \bar{t}$ for all u and for any $\mu > 0$.

Lemma 2.4.5. *Consider the function $t_\mu(x) \triangleq \mu(\log(e^{\frac{u}{\mu}} + 1) - \log(2))$. Then for any $u \in \mathbb{R}$ and any $\mu > 0$, we have that $t'_\mu(u) \leq 1$.*

Proof. It can be seen for that any u and any $\mu > 0$,

$$t'_\mu(x) = \frac{\exp\left(\frac{u}{\mu}\right)}{\exp\left(\frac{u}{\mu}\right) + 1} \leq 1.$$

□

2.4.3 Smoothing a composition of two smoothable functions

We note that $[\mathcal{Q}(\mathbf{x}, \omega) - m]^+$ denotes a composition of a nonsmooth function $h(w)$ where $h(w) = \max\{w, 0\}$ with another nonsmooth function $t(z)$ where $t(z) = z_1 - z_2$ and $z = (z_1, z_2)$. Our intent lies in showing that under if h is (α_1, β_1) smoothable and t is (α_2, β_2) smoothable, then $p = h(t)$ is (α_3, β_3) smoothable, where the smoothability of this composite function is defined as follows.

Definition 2.4.2. *Given two convex functions $h : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. Then the function $p(z) = h(t(z))$ is said to be (α, β) smoothable if the following two conditions hold and $p_\mu(z) \triangleq h_\mu(g_\mu(z))$.*
(i) There exists a constant α such that for any $\mu > 0$,

$$\|\nabla_z p_\mu(z_1) - \nabla_z p_\mu(z_2)\| \leq \tfrac{\alpha}{\mu}\|z_1 - z_2\|, \quad \forall z_1, z_2.$$

(ii) There exists a constant β such that for any $\mu > 0$,

$$p_\mu(z) \leq p(z) \leq p_\mu(z) + \beta\mu.$$

Under suitable conditions, we now prove that $p = h(t)$ is a smoothable function when h and t are smoothable.

Lemma 2.4.6. *Suppose $h : \mathbb{R} \to \mathbb{R}$ is a non-decreasing nonnegative convex function and is (α_1, β_1)-smoothable. In addition, if h_μ denotes an (α_1, β_1) smoothing of h, then h_μ is assumed to be a non-increasing and nonnegative function. Suppose $t : \mathbb{R}^n \to \mathbb{R}^m$ is (α_2, β_2)-smoothable convex function. In addition, suppose $\|\nabla_w h(w)\| \leq C_1$ for all w and $\|\nabla_x t(x)\| \leq C_2$ for all x. Then $p(z) = h(t(z))$ is (α, β)-smoothable.*

Proof. Since $\nabla_z(h_\mu(g_\mu(z)) = h'_\mu(t_\mu(z))\nabla_z t_\mu(z)$, we have the following by adding and subtracting terms and invoking the triangle inequality.

$$\begin{aligned}\|\nabla h_\mu(t_\mu(z_1)) - \nabla h_\mu(t_\mu(z_2)\| &= \|h'_\mu(t_\mu(x_1))\nabla_z t_\mu(z_1) - h'_\mu(t_\mu(z_2))\nabla_z t_\mu(z_2)\| \\ &\le \|h'_\mu(t_\mu(z_1))\nabla_z t_\mu(z_1) - h'_\mu(t_\mu(z_1))\nabla_z t_\mu(z_2)\| \\ &+ \|h'_\mu(t_\mu(z_1))\nabla_z t_\mu(z_2) - h'_\mu(t_\mu(z_2))\nabla_z t_\mu(z_2)\| \\ &\le \|h'_\mu(t_\mu(x_1))\|\|\nabla_z t_\mu(z_1) - \nabla_z t_\mu(z_2)\| \\ &+ \|\nabla_z t_\mu(x_2)\| \underbrace{\|h'_\mu(t_\mu(z_1)) - h'_\mu(t_\mu(z_2))\|}_{\textbf{Term b}}. \end{aligned} \tag{2.23}$$

Since $h : \mathbb{R} \to \mathbb{R}$ is (α_1, β_1)-smoothable and $t : \mathbb{R}^n \to \mathbb{R}$ is (α_2, β_2)-smoothable, it follows that for all z_1, z_2,

$$\begin{aligned}\|h'_\mu(t_\mu(z_1)) - h'_\mu(t_\mu(z_2))\| &\le \frac{\alpha_1}{\mu}\|t_\mu(z_1) - t_\mu(z_2)\| \\ \|\nabla t_\mu(z_1) - \nabla t_\mu(z_2)\| &\le \frac{\alpha_2}{\mu}\|z_1 - z_2\|.\end{aligned}$$

Recall that by the differentiability of $t_\mu(z)$ and the mean-value theorem, for some $\gamma \in [0, 1]$,

$$\begin{aligned} t_\mu(z_2) &= t_\mu(z_1) + \nabla_z t_\mu(z_1 + \gamma(z_2 - z_1))^T(z_2 - z_1) \\ \Rightarrow \|t_\mu(z_2) - g_\mu(z_1)\| &\le \|\nabla_z t_\mu(z_1 + \gamma(z_2 - z_1))\|\|z_1 - z_2\| \\ &\le C_2\|z_1 - z_2\|. \qquad \text{(by assumption)}. \end{aligned} \tag{2.24}$$

By (α_1, β_1)-smoothability of h and (2.24), Term b can be bounded as follows.

$$\textbf{Term b} \le \tfrac{\alpha_2}{\mu}\|t_\mu(z_1) - t_\mu(z_2)\| \le \frac{\alpha_2 C_2}{\mu}\|z_1 - z_2\|. \tag{2.25}$$

From (2.23), we have that for any z_1, z_2,

$$\begin{aligned} \|\nabla h_\mu t_\mu(z_1) - \nabla h_\mu(t_\mu(z_2)\| &\le \frac{C_1\alpha_2}{\mu}\|z_1 - z_2\| + \frac{C_2\alpha_1}{\mu}\|z_1 - z_2\| \\ \le \tfrac{\alpha}{\mu}\|z_1 - z_2\|, \text{ where } \alpha &\triangleq (C_1\alpha_2 + C_2\alpha_1). \end{aligned} \tag{2.26}$$

Since $h : \mathbb{R} \to \mathbb{R}$ is (α_1, β_1)-smoothable and $t : \mathbb{R}^n \to \mathbb{R}$ is (α_2, β_2)-smoothable, we have for any $z, \mu > 0$,

$$h_\mu(t(z)) \leq h(t(z)) \leq h_\mu(t(z)) + \beta_1\mu \tag{2.27}$$

$$t_\mu(z) \leq t(z) \leq t_\mu(z) + \beta_2\mu. \tag{2.28}$$

Since $h(\bullet)$ is a nondecreasing function,

$$\begin{aligned}
h_\mu(t_\mu(z)) &\leq h_\mu(t(z)) \leq h_\mu(t_\mu(z) + \beta_2\mu) && (h_\mu \text{ non-dec., (2.28)})\\
h(t(z)) &\leq h(t_\mu(z) + \beta_2\mu) && ((2.28),\ h \text{ nondec.})\\
h_\mu(t_\mu(z)) &\leq h_\mu(t(z))\\
&\leq h(t(z)) && (\text{From (2.27)})\\
&\leq h_\mu(t(z)) + \beta_1\mu && (\text{From (2.27)})\\
&\leq h_\mu(t_\mu(z) + \beta_2\mu) + \beta_1\mu. && (\text{From (2.28)})
\end{aligned}$$

Since h_μ is a convex and positive function, for $\kappa \in [0, 1]$, we have the following by the mean-value theorem.

$$h_\mu(t_\mu(z) + \beta_2\mu) = h_\mu(t_\mu(z)) + h'_\mu(t_\mu(z) + \kappa\beta_2\mu)\beta_2\mu.$$

Since $h'_\mu(t_\mu(z) + \kappa\beta\mu) \leq C_1$, it follows that

$$h_\mu(t_\mu(z) + \beta_2\mu) \leq h_\mu(t_\mu(z)) + C_1\beta\mu,$$

implying that

$$h_\mu(t_\mu(z)) \leq h(t(z)) \leq h_\mu(t_\mu(x)) + \underbrace{(\beta_1 + C_1\beta_2)}_{\triangleq\beta}\mu. \tag{2.29}$$

From (2.26) and (2.29), $h(t(z))$ is (α, β)-smoothable. □

2.4.4 Smoothing $r(\mathbf{z}, \omega)$

Since $r(\mathbf{z}, \omega)$ is a consequence of a composition of a nonsmooth function (specifically the max function) on an affine translation of another nonsmooth function (specifically the recourse function $\mathcal{Q}(\bullet, \omega)$), we may utilize the results from the prior subsection. Specifically, let $t(\mathbf{z},\omega) \triangleq \mathcal{Q}(\mathbf{x},\omega) - m$, where $\mathbf{z} = (g,\theta,m)$. Furthermore, suppose $h(u) = [u]^+$.

Lemma 2.4.7. *Consider the functions $t(\mathbf{z},\omega) = \mathcal{Q}(\mathbf{x},\omega)$ and $h(u) = [u]^+$. Then the following hold.*
(i) The function $h_\mu(u) = \mu \log(e^{\frac{u}{\mu}} + 1)$ represents a $(1, \log(2))$ smoothing of h and $0 \leq h'(u) \leq 1$ for all u.
(ii) The function $t(\mathbf{z},\omega) = \mathcal{Q}(\mathbf{x},\omega) - m$ is a convex (α_2,β_2)-smoothable function and $\|\nabla_{\mathbf{z}} t(\mathbf{z},\omega)\| \leq C_2$ for all ω.

Proof. (i) Follows immediately from [59] and Lemma 2.4.5.

(ii) Since $\mathcal{Q}_\mu(\mathbf{x},\omega)$ is $(\alpha_2(\omega),\beta_2(\omega))$-smoothable, we have that $\mathcal{Q}_\mu(\mathbf{x},\omega) - m$ satisfies the following for any $\mathbf{x}$.

$$\mathcal{Q}_\mu(\mathbf{x},\omega) - m \leq \mathcal{Q}(\mathbf{x},\omega) - m \leq \mathcal{Q}_\mu(\mathbf{x},\omega) - m + \beta_2(\omega)\mu. \tag{2.30}$$

In addition, we have that

$$\begin{aligned}
\|\nabla_{\mathbf{z}} r(\mathbf{z}_1,\omega) - \nabla_{\mathbf{z}} r(\mathbf{z}_2,\omega)\| &= \left\| \begin{pmatrix} \nabla_g(\mathcal{Q}_\mu(\mathbf{x}_1,\omega) - m) - \nabla_g(\mathcal{Q}_\mu(\mathbf{x}_2,\omega) - m) \\ \nabla_\theta(\mathcal{Q}_\mu(\mathbf{x}_1,\omega) - m) - \nabla_\theta(\mathcal{Q}_\mu(\mathbf{x}_2,\omega) - m) \\ \nabla_m(\mathcal{Q}_\mu(\mathbf{x}_1,\omega) - m) - \nabla_m(\mathcal{Q}_\mu(\mathbf{x}_2,\omega) - m) \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} \nabla_g \mathcal{Q}_\mu(\mathbf{x}_1,\omega) - \nabla_g \mathcal{Q}_\mu(\mathbf{x}_2,\omega) \\ 0 \\ 0 \end{pmatrix} \right\| \\
&\leq \frac{\alpha_2(\omega)}{\mu} \|g_1 - g_2\| \leq \frac{\alpha_2(\omega)}{\mu} \|\mathbf{z}_1 - \mathbf{z}_2\|.
\end{aligned}$$

Finally, it is relatively easy to see that

$$\|\nabla_{\mathbf{z}} t(\mathbf{z},\omega)\| = \left\| \begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{Q}_\mu(\mathbf{x},\omega) \\ -1 \end{pmatrix} \right\| \leq \|\nabla_{\mathbf{x}} \mathcal{Q}_\mu(\mathbf{x},\omega)\| + 1 \leq C(\omega) + 1,$$

for all $\mathbf{x}$ where the last inequality follows from observing that

$$\|\nabla_{\mathbf{x}} \mathcal{Q}_\mu(\mathbf{x},\omega)\| = \| - T_\omega^T \pi_\omega\| \leq \|T_\omega\| \left(\frac{\bar{d}}{\bar{c}} + \bar{\mu} \left(\frac{\bar{d}}{\bar{c}} \right)^2 \right).$$

$\square$

We may now claim the smoothability of $r(\mathbf{z}, \omega)$.

Proposition 2.4.2. *Suppose Assumption 2.4.1 holds. Consider function $r(\mathbf{z}, \omega)$ defined in (2.14). Then $r(\mathbf{z}, \omega)$ is a convex and (α, β)-smoothable function.*

Proof. From Lemma 2.4.7, we have that $h(u)$,$t(\mathbf{z}, \omega)$ satisfy the requirement of Lemma 2.4.6. Following Lemma 2.4.6 we have that $r(\mathbf{z}, \omega) = h(t(\mathbf{z}, \omega))$ is (α, β)-smoothable. $\square$

We may then define the smoothed approximation of $r(\mathbf{z}, \omega)$ as follows.

$$r_\mu(\mathbf{z}, \omega) \triangleq m + \frac{\mu \log\left(\exp\left(\frac{\mathcal{Q}_\mu(\mathbf{x}, \omega) - m}{\mu}\right) + 1\right) - \mu \log(2)}{1 - \tau}. \tag{2.31}$$

As a consequence, we have that

$$\nabla_{\mathbf{z}} r_\mu(\mathbf{z}, \omega) = \begin{pmatrix} \frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}}{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}+1}\right)\nabla \mathcal{Q}_\mu(\mathbf{x}, \omega) \\ 0 \\ 1 - \frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}}{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}+1}\right) \end{pmatrix}. \tag{2.32}$$

We know $\mathcal{Q}_\mu(\mathbf{x}, \omega)$ is (α, β)-smooth approximation of $\mathcal{Q}(\mathbf{x}, \omega)$. Thus, $r_\mu(\mathbf{z}, m)$ is a smooth approximation of $r(\mathbf{z}, m)$ with its gradients Lipschitz constant μ.

2.5 A variance-reduced smoothed accelerated scheme for two-stage risk-averse problems

While the prior section has analyzed the smoothing of the risk-adjusted recourse function, in this section, we utilize a variance-reduced smoothed accelerated scheme for such a class of problems. In Section 2.5.1, we provide a brief review of decomposition and Monte-Carlo sampling techniques for resolving two-stage stochastic convex problems, possibly complicated by risk-aversion. In Section 2.5.2, we introduce a recently developed variance-reduced smoothed accelerated scheme and show how it may be extended to contend with risk-averse regimes. Finally, in Section 2.5.3, we review the convergence statements inherited from this scheme.

2.5.1 A review of Monte-Carlo sampling schemes for 2-stage programs

Traditionally, schemes for resolving two-stage stochastic programs have differed based on whether the sample-space of the second-stage problem is finite or infinite. In the case of the former, decomposition techniques have proven useful in developing techniques that scale with the cardinality of Ω. Amongst the earliest of these was the L-shaped method [61] while augmented Lagrangian [62] and splitting methods [63] have also been utilized. A more comprehensive review of decomposition schemes can be found in [64]. When the sample-space is infinite, these avenues cannot be adopted and one has to resort to Monte-Carlo sampling schemes. We review three avenues for resolving such problems.

2.5.1.1 Stochastic cutting plane methods

Stochastic decomposition (SD) techniques decompose the stochastic elements of a problem from deterministic data, combining successive approximation methods from mathematical programming with sampling approaches. Unlike other sampling methods, SD leverages the special structure of linear programming problems. When the second-stage problems are linear, this implies that the second-stage recourse function is a piece-wise linear function. Cutting-plane techniques originate from the work by Kelley [65] in which the following algorithm was proposed to solve the following convex problem.

$$\min_{x \in X} \quad c^T x + \mathcal{Q}(x), \tag{2.33}$$

where $\mathcal{Q}(\cdot)$ is a convex function and X is a compact, convex, and nonempty set. The basic idea of cutting plane algorithm is as below: Such avenues have been extended to accommodate two-stage stochastic linear programs by Van Slyke and Wets [61] but only allow for finite sample-spaces . To accommodate general sample-spaces, the stochastic decomposition (SD) scheme was proposed by Higle and Sen [5] in 1991. Consider the problem (2.33) where $\mathcal{Q}(x) \triangleq \mathbb{E}[\mathcal{Q}(x,\omega)]$, where $\mathcal{Q}(x,\omega)$ is an optimal value of

$$\max_{\pi} \left\{ (h_\omega - T_\omega x)^T \pi \mid W^T \pi \le q_\omega \right\}. \tag{LP-S2D}$$

Algorithm 1 Cutting-plane scheme

initialization: $x^1, k = 0, \ell_0(x) = -\infty, u_0 = c^T x^1 + f(x^1)$ and $l_0 = -\infty$;
while $u_k - l_k > 0$ **do**
  $k = k + 1$. Find (α^k, β^k) such that

$$\mathcal{Q}(x^k) = \alpha_k + \beta_k x^k$$
$$\mathcal{Q}(x) \geq \alpha_k + \beta_k x \quad \forall x \in X$$

  Update $u_k = \min\{u_{k-1}, c^T x^k + \mathcal{Q}(x^k)\}, \ell_k(x) = \max\{\ell_{k-1}(x), \alpha_k + \beta_k x\}$.
  Update $l_k = \min_{x \in X}\{c^T x + \ell_k(x)\}$ where x_{k+1} is a solution to (2.33).
end while
$x^* = x_{k+1}$
return x^*

This scheme approximates the recourse function $\mathbb{E}\left[Q(x, \omega)\right]$ through a sequence of piecewise linear approximations. Within any given major iteration, each piece of the piecewise linear approximation is derived from a conditionally independent set of observations. As part of the scheme [5], a set V_k is constructed by solving one subproblem per iteration and dual vector obtained is added to this set. Formally, the update of V_k is defined as follows.

$$V_k := V_{k-1} \cup \pi^k_{\omega_k},$$

where $\pi^k_{\omega^k}$ is a solution to the following problem:

$$\begin{aligned} \max \quad & [h_{\omega^k} - T_{\omega^k} x^k]^T \pi \\ \text{subject to} \quad & W^T \pi \leq q_{\omega^k}. \end{aligned}$$

Akin to the cutting plane scheme, we may obtain the piecewise linear outer-approximation $\eta_k(x)$ for the recourse function with the following form:

$$\eta_k(x) := \max\{\alpha^k_t + \beta^k_t x \ \mid \ t = 1, \ldots, k\}.$$

The stochastic decomposition (SD) algorithm is formally defined in Algorithm 3 and further details can be found from [5].

Algorithm 2 Stochastic Decomposition Algorithm

initialization: $k = 0, V_0 = \emptyset, \eta_0(x) = -\infty, x^1 \in X$ L is given;
while iteration $k < K_{max}$ **do**
  $k = k+1$. Randomly generate an observation of ω, ω^k, independent of any previously generated observations;
  Solve subproblem to get solution $\pi^k_{\omega^k}$

$$\begin{aligned} \max \quad & [h_{\omega^k} - T_{\omega^k} x^k]^T \pi \\ \text{subject to} \quad & W^T \pi \leq q_{\omega^k}. \end{aligned}$$

  Update $V_k = V_{k-1} \cup \pi^k_{\omega_k}$
  Determine the coefficients of the k^{th} piecewise linear approximation of recourse function (α^k, β^k) such that

$$\alpha^k_k + \beta^k_k x = \frac{1}{k} \sum_{t=1}^{k} \pi^k_t (h_{\omega^t} - T_{\omega^t} x)$$

  where $\pi^k_t \in \arg\max\{\pi^T (h_{\omega_t} - T_{\omega_t} x^k) | \pi \in V_k\}$
  Update the coefficients of all previously generated cuts.

$$\alpha^k_t = \frac{k-1}{k} \alpha^{k-1}_t + \frac{1}{k} L \quad , \quad \beta^k_t = \frac{k-1}{k} \beta^{k-1}_t.$$

  Update $\eta_k(x) = \max\{\alpha^k_t + \beta^k_t x \ | \ t = 1, \ldots, k\}$.
  Solve $\min_{x \in X} \{c^T x + \eta_k(x)\}$ where x_{k+1} is the solution to this problem.
end while
$x^* = x_{k+1}$
return x^*

2.5.1.2 Sample-average approximation

In sample-average approximation theory (also referred to as exterior sampling), samples are generated outside of an optimization procedure. Consequently, the resulting sample average approximation (SAA) problems are solved by deterministic optimization algorithms. One of the advantages of SAA is that this method separates sampling procedures and optimization techniques. Consider the following stochastic programming problem:

$$\min_{x \in X} \quad f(x), \text{ where } f(x) \triangleq \mathbb{E}\left[F(x, \xi(\omega))\right], \tag{2.34}$$

$X \subseteq \mathbb{R}^n$ is a closed and convex set, $\xi : \Omega \to \mathbb{R}^d$ is a random vector, and the associated probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. Unless stated otherwise, the expectation is assumed to be well-defined and finite valued for all $x \in X$, which implies for every $x \in X$ the value of $F(x, \omega)$ for every $\omega \in \Omega$ is finite. Suppose we have a sample $\omega^1, \ldots, \omega^N$ of N realizations of the random vector ω. This sample is generated by Monte Carlo sampling and for any $x \in X$, we estimate the expected value $f(x)$ by the sample-average $f_N(x) \triangleq \frac{1}{N}\sum_{j=1}^N F(x, \omega^j)$ by averaging values $F(x, \omega^j), j = 1, \ldots, N$. The resulting sample average approximation (SAA) of the true problem is defined as follows.

$$\min_{x \in X} \hat{f}_N(x), \text{ where } \hat{f}_N(x) \triangleq \frac{1}{N}\sum_{j=1}^{N} F(x, \omega^j).$$

Note that $\hat{f}_N(x)$ can be viewed as the expectation taken with respect to the empirical measure associated with a probability mass function $\{\frac{1}{N}, \ldots, \frac{1}{N}\}$. By the law of large numbers (LLN), under suitable regularity conditions $\hat{f}_N(x)$ converges to $f(x)$ pointwise with probability one as $N \to \infty$. Moreover, by the classical LLN, this convergence holds if the sample is independent and identically distributed. Much of the research on SAA theory focuses on proving that the estimator for the optimal value converges to the true value as $N \to \infty$ with probability one. Related statements can be developed for the solution set. In addition, rates of convergence can also be derived for such schemes.

Consistency of SAA estimators was investigated by tools of epi-convergence analysis by King and Wets [66] and Robinson [67] while asymptotic of SAA estimators of optimal solutions of stochastic programs were discussed by King and Rockefeller [68] and Shapiro [69]. A detailed exposition of recent theoretical findings can be found in [60]. It is worth emphasizing that this avenue is not an algorithm in the conventional sense but represents an avenue for approximation.

2.5.1.3 Stochastic approximation methods

Stochastic approximation schemes originate from the seminal paper by Robbins and Monro [70] while asymptotics can be found in the research by Kushner and Clark [71] and Nevelson and Hasminskii [72]. Longer step averaging schemes was developed in Polyak [73] and these ideas were presented in a different form by Nemirovski and Yudin [74]. Stochastic quasi-gradient techniques are closely related to stochastic approximation and early work focused on the solution of two-stage stochastic linear programs [75]. This avenue saw significant subsequent study by Gaivoronski [76], Wets [77], amongst others. For a given stochastic convex optimization (2.34) where f is a differentiable function, given an $x_0 \in X$, a standard SA scheme would be based on the following update rule:

$$x_{k+1} := \Pi_X \left(x_k - \eta_k(\nabla_x f(x_k) + w_k)\right), \ k \geq 0$$

where $w_k := \nabla_x f(x_k; \omega_k) - \nabla_x f(x_k)$ and $\nabla_x f(x, \xi_\omega)$ is referred to as $\nabla_x f(x, \omega)$. An variable sample-size stochastic approximation scheme (VSSA) was proposed in [78, 79] in which the sequence $\{x_{k+1}\}$ would have the following update rule:

$$x_{k+1} := \Pi_X \left(x_k - \eta_k \frac{\sum_{j=1}^{N_k} \nabla_x f(x_k, \xi_{j,k})}{N_k}\right), \ k \geq 0.$$

In such a scheme, an increasingly unbiased estimate of the gradient is employed, leading to improved iteration complexity of the scheme. Following the idea in [78–80], we introduce a variant of this scheme for solving two-stage stochastic programs in the next section.

Specifically, in this chapter, we revisit stochastic quasigradient methods which has traditionally been plagued by the same challenges as stochastic approximation. In particular, the convergence rate was $\mathcal{O}(\frac{1}{\sqrt{k}})$ and the empirical behavior varies significantly with the choice of step length sequence. In [81], we introduce three key modifications to the standard stochastic approximation framework by (i) utilizing a two-step accelerated scheme, (ii) incorporating a smoothing of the recourse function by regularizing the second-stage dual; (iii) and leveraging variance reduction. We develop a foundation to allow for applying this framework to risk-averse two-stage problems which allows for recovering the optimal rate of $\mathcal{O}(1/k)$. Next, we describe this scheme.

2.5.2 Variance-reduced smoothed accelerated scheme

We now introduce the variable sample-size accelerated proximal method (VS-

APM) first presented in [81] and apply it to (r-ED).

$$\min_{\mathbf{z} \in Z} \mathbb{E}[h(\mathbf{z}, \omega)], \text{ where } h(\mathbf{z}, \omega) \triangleq (c(\mathbf{z}) + r(\mathbf{z}, \omega)). \tag{r-ED}$$

This framework incorporates three aspects in extending standard stochastic approximation schemes.

(i) *Smoothing.* The first change from standard stochastic approximation schemes lies in utilizing the gradient of a smoothed objective, where the smoothing parameter sequence is driven to zero. The resulting scheme can be formalized as follows, given a $z_0 \in Z$.

$$\mathbf{z}_{k+1} := \Pi_Z \left[z_k - \gamma_k (\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + w_k) \right], \qquad k \geq 0. \tag{2.35}$$

In contrast with standard stochastic approximation, we employ the sampled gradient $\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + w_k$ where $\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + w_k = \nabla_{\mathbf{z}} c(\mathbf{z}_k) + \nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k, \omega_k)$.

(ii) *Variance-reduction.* In traditional stochastic approximation schemes, a single sample $\nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k, \omega_k)$ or a fixed batch-size of samples is utilized. However, such avenues lead to biased gradients (where the conditional bias does not diminish to zero). Instead, we propose a variance-reduced scheme given by the following.

$$\mathbf{z}_{k+1} := \Pi_Z \left[z_k - \gamma_k (\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + \bar{w}_k) \right], \qquad k \geq 0 \tag{2.36}$$

where $\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + \bar{w}_k = \frac{\sum_{j=1}^{N_k} \nabla_{\mathbf{z}} c(\mathbf{z}_k) + \nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k, \omega_{j,k}):}{N_k}$. In fact, the conditional bias of the gradients diminishes to zero and this scheme starts mimicking an inexact gradient scheme.

(iii) *Acceleration.* Finally, we introduce an accelerated scheme by utilizing the following two-step rule.

$$\zeta_{k+1} := \Pi_Z \left[z_k - \gamma_k (\nabla_{\mathbf{z}} h_{\mu_k}(\mathbf{z}_k) + \bar{w}_k) \right], \qquad k \geq 0 \tag{2.37}$$

$$\mathbf{z}_{k+1} := (1 + \alpha_k) \zeta_{k+1} - \alpha_k \zeta_k, \qquad k \geq 0. \tag{2.38}$$

Note that α_k are prescribed sequences and this avenue was first suggested for solving convex programs with differentiable objectives by Nesterov [82].

The resulting accelerated scheme improved the convergence rate from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$. Similar benefits are expected to accrue here when step length sequences, smoothing sequences, and sample-size sequences are chosen appropriately. Collectively, this scheme is referred to as a variable sample-size accelerated proximal scheme (VSAPM) [81].

The scheme is formally stated in Algorithm 3. It may be recalled that

Algorithm 3 VS-APM for two-stage risk-based ED

initialization: $\lambda_1 = 1, \gamma_0, \mathbf{y}_0 = \mathbf{z}_1, M_0 = 0, N_k = 1, k = 1;$
while $k < K$ **do**
  Generate N_k samples and compute $\nabla_{\mu_k} h(\mathbf{z}_k, \omega_{1,k}), \ldots, \nabla_{\mu_k} h(\mathbf{z}_k, \omega_{N_k,k})$.
  Update

$$\mathbf{y}_{k+1} := \Pi_Z \left[\mathbf{x}_k - \eta_k \frac{\sum_{j=1}^{N_k} \nabla_{\mu_k} h(\mathbf{z}_k, \omega_{j,k})}{N_k} \right]. \tag{2.39}$$

  $\lambda_{k+1} := \frac{1+\sqrt{1+4\lambda_k^2}}{2}$; $\gamma_k := \frac{\lambda_k - 1}{\lambda_{k+1}}$,
  Update

$$\mathbf{z}_{k+1} := (1+\gamma_k)\mathbf{y}_{k+1} - \gamma_k \mathbf{y}_k). \tag{2.40}$$

  Update $k := k+1$, N_k and μ_k.
end while
return $\mathbf{z}_K$.

$$r_\mu(\mathbf{z}, \omega) = m + \frac{1}{1-\tau}\mu \left[\log \left(e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}} + 1 \right) \right], \tag{2.41}$$

where $\mathbf{z} = (\mathbf{z}, m)$. Furthermore, the gradient $\nabla_{\mathbf{z}} r_\mu(\mathbf{z}, \omega)$ is defined as follows.

$$\begin{aligned}
\nabla_{\mathbf{x}} r_\mu(\mathbf{z}, \omega) &= \left(\frac{1}{1-\tau} \right) \left(\frac{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}}{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}} + 1} \right) \nabla \mathcal{Q}_\mu(\mathbf{x}, \omega), \\
\nabla_m r_\mu(\mathbf{z}, \omega) &= 1 - \left(\frac{1}{1-\tau} \right) \left(\frac{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}}}{e^{\frac{\mathcal{Q}_\mu(\mathbf{x},\omega)-m}{\mu}} + 1} \right),
\end{aligned} \tag{2.42}$$

where $\nabla_{\mathbf{x}} \mathcal{Q}_{\mu_k}(\mathbf{x}_k, \omega_k) = -T(\omega_k)^T \pi(\mathbf{x}_k, \omega_k)$ and $\pi(\mathbf{x}_k, \omega_k)$ is a solution of the smoothed second-stage dual problem (D-$\mathcal{Q}_\mu(\mathbf{x}, \omega_{j,k})$).

2.5.3 Convergence theory

We now recall the two main assumptions for claiming convergence from (VS-APM) from [81]. Of these, the first requires that the objective of the original problem is indeed smoothable and the distance of the initial iterate to an optimal solution can be bounded.

Assumption 2.5.1. *(i) The function $\mathbb{E}[r(\mathbf{z}, \omega)]$ is (α, β) smoothable; (ii) There exists a C such that $\|\mathbf{z}_1 - \mathbf{z}^*\|^2$ where $\mathbf{z}^*$ is a solution to the original problem.*

Next, we require that the noise sequence $\bar{w}_k$, defined as follows.

$$\bar{w}_k \triangleq \nabla_{\mathbf{z}} \widehat{r}_{\mu_k}(\mathbf{z}_k) - \nabla_{\mathbf{z}} \bar{r}_{\mu_k}(\mathbf{z}_k), \text{ where } \nabla_{\mathbf{z}} \widehat{r}_{\mu_k}(\mathbf{z}_k) \triangleq \frac{\sum_{j=1}^{N_k} \nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k, \omega_{j,k})}{N_k}. \tag{2.43}$$

Assumption 2.5.2. *Consider the sequence $\{\bar{w}_k\}$ where $\bar{w}_k$ is defined as (2.43). Then there exists a scalar $\nu > 0$, such that $\mathbb{E}[\|\bar{w}_k\|^2 \mid \mathcal{F}_k] \leq \frac{\nu^2}{N_k}$ and $\mathbb{E}[\bar{w}_k \mid \mathcal{F}_k] = 0$ holds almost surely for all k, where $\mathcal{F}_k \triangleq \sigma\{\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_{k-1}\}$.*

We may now formally state the convergence statement from [81].

Proposition 2.5.1. *Consider the sequence $\{\mathbf{z}_k\}$ generated from (sVS-APM) where $\mu_k = 1/k$, $\eta_k = 1/2k$, and $N_k = \lfloor k^a \rfloor$, where $a > 1$. Suppose Assumption 2.5.1 and 2.5.2 hold. Then the following hold.*

(i) If $\bar{C} \triangleq \frac{2\nu^2 a}{a-1} + 4C^2 + B^2$, then the following holds for $K \geq 1$.

$$\mathbb{E}[h(\mathbf{z}_k, \omega_k)] - \mathbb{E}[h(\mathbf{z}^*, \omega)] \leq \frac{\bar{C}}{K}. \tag{2.44}$$

(ii) Let $\epsilon \leq \bar{C}/2$ and K is such that $\mathbb{E}[h(\mathbf{z}_k, \omega_k)] - \mathbb{E}[h(\mathbf{z}^*, \omega)] \leq \epsilon$. Then $\sum_{k=1}^{K} N_k \leq \mathcal{O}(\frac{1}{\epsilon^{1+a}})$.

We now provide some results that allows us to claim that such Prop. 2.5.1 can be invoked.

Lemma 2.5.1. *Consider the noise sequence $\bar{w}_k$ defined in (2.43). Then this sequence satisfies the following: (i) $\mathbb{E}[\bar{w}_k \mid \mathcal{F}_k] = 0$ a.s. for every $k \geq 1$; (ii) $\mathbb{E}[\bar{w}_k \mid \mathcal{F}_k] \leq \frac{\nu^2}{N_k}$ a.s. for every $k \geq 1$.*

Proof. (i) Recall that $r_\mu(\mathbf{z}) = \mathbb{E}[r_\mu(\mathbf{z},\omega)]$. Since $r_\mu(\mathbf{z},\omega)$ is a continuously differentiable convex function in $\mathbf{z}$ for every ω, it follows that we may interchange derivatives and expectations in claiming that $\nabla_{\mathbf{z}} r_\mu(\mathbf{z}) = \mathbb{E}[\nabla_{\mathbf{z}} r_\mu(\mathbf{z},\omega)]$ (cf. [60, Theorem 7.44]). Consequently, if $\mathbf{z}_k$ is adapted to $\mathcal{F}_k$ and $\bar{w}_k$ is defined as (2.43), it follows that $\mathbb{E}[\bar{w}_k \mid \mathcal{F}_k] = 0$ in an a.s. sense.

(ii) Next, we note that $\bar{w}_k$ is a sample-average of a set of i.i.d random variables with mean zero. Consequently, if $\mathbb{E}[\|w_i\|^2 \mid \mathcal{F}_k] \le \nu^2$ for $i = 1,\ldots,N_k$ in an a.s. fashion, it follows that $\mathbb{E}[\|\bar{w}_k\|^2 \mid \mathcal{F}_k] \le \frac{\nu^2}{N_k}$ in a.s. sense. It remains to show that $\mathbb{E}[\|w_i\|^2 \mid \mathcal{F}_k] \le \nu^2$ a.s. .

$$
\begin{aligned}
&\mathbb{E}\left[\|\nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k,\omega_k) - \mathbb{E}[\nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k,\omega)]\|^2\right] \le \mathbb{E}\left[\|\nabla_{\mathbf{z}} r_{\mu_k}(\mathbf{z}_k,\omega_k)\|^2\right] \\
&= \mathbb{E}\left[\|\nabla_{\mathbf{x}} r_{\mu_k}(\mathbf{z}_k,\omega_k)\|^2\right] + \mathbb{E}\left[\|\nabla_m r_{\mu_k}(\mathbf{z}_k,\omega_k)\|^2\right] \\
&= \mathbb{E}\left[\left\|\frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}\nabla\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right\|^2\right] \\
&+ \mathbb{E}\left[\left\|\left(1-\frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right)\right\|^2\right].
\end{aligned}
$$

We observe that the first term can be bounded as follows.

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}\nabla\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right\|^2\right] \\
&= \mathbb{E}\left[\left(\tfrac{1}{(1-\tau)}\right)^2\left\|\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}\nabla\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right\|^2\right] \\
&\le \mathbb{E}\left[\left(\tfrac{1}{(1-\tau)}\right)^2\left\|\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\nabla\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})\right\|^2\right] \\
&\le \mathbb{E}\left[\left(\tfrac{1}{(1-\tau)}\right)^2\|\nabla\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})\|^2\right] \\
&= \mathbb{E}\left[\left(\tfrac{1}{(1-\tau)}\right)^2\|-T^T_{\omega_{j,k}}\pi_{\omega_{j,k}}\|^2\right] \\
&\le \mathbb{E}\left[\left(\tfrac{\bar{\pi}}{(1-\tau)}\right)^2\|T_{\omega_{j,k}}\|^2\right] \le \nu_1^2.
\end{aligned}
$$

The second term can be similarly bounded.

$$\begin{aligned}
&\mathbb{E}\left[\left\|\left(1-\frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right)\right\|^2\right] \\
&\leq \mathbb{E}\left[\left\|\left(1-\frac{1}{1-\tau}\left(\frac{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}}{e^{\frac{\mathcal{Q}_{\mu_k}(\mathbf{x}_k,\omega_{j,k})-m_k}{\mu_k}}+1}\right)\right)\right\|^2\right] \\
&\leq (2+\frac{2}{(1-\tau)^2}) \triangleq \nu_2^2.
\end{aligned}$$

It follows that $\mathbb{E}[\|w_i\|^2 \mid \mathcal{F}_k] \leq \nu^2 = \nu_1^2 + \nu_2^2$. $\square$

2.6 Numerical Studies

In this section, we apply our scheme to the resolution of two-stage stochastic economic dispatch problems to two sets of problems. In Section 2.6.1, we review the model for generation of wind realizations and compare our scheme with the stochastic decomposition and stochastic quasi-gradient counterparts in Section 2.6.2 based on an IEEE 118-busy system which contains 19 generators, 35 synchronous condensers, 177 lines, 9 transformers, and 91 loads. The impact of risk is examined in Section 2.6.3. Finally, we conclude the section by examining the performance of this scheme on test problems sourced from ARPA-E's Grid Optimization competition.

2.6.1 Autoregressive Moving Average Model

In this subsection, we review the statistical model employed for developing wind forecasts and generating demand scenarios. A review of multi area wind speed and wind power scenario generation methods has been provided in [83]. It is worth noting that ARMA techniques have been used in developing stochastic optimization schemes for power system dispatch [84]. Our focus is on autoregressive moving average (ARMA) models and use data from ERCOT's hourly wind generation during 2009, 2010, and 2011 to test the resulting models. In an ARMA model [85], wind speed y_t in period t consists of the weighted sum of past observations and a

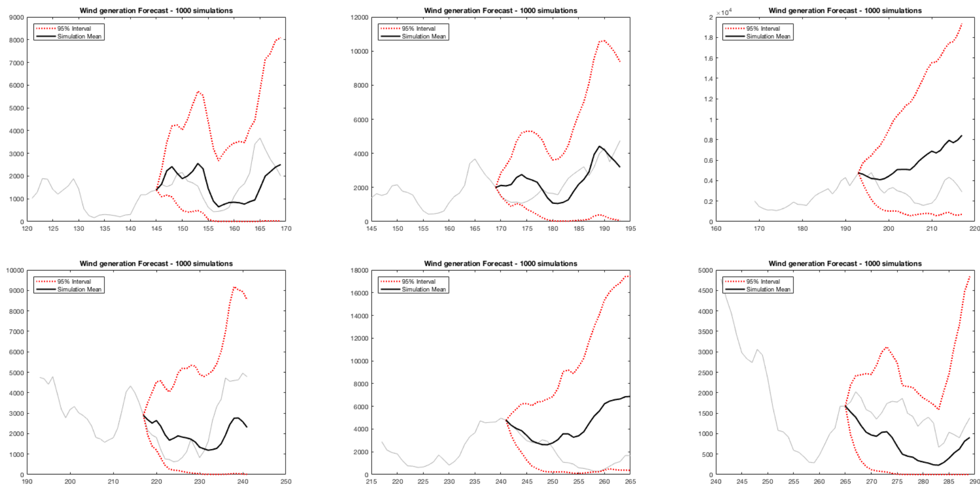

Figure 2.1. ARMA: Test data v.s. prediction

weighted sum of independent shocks defined as follows:

$$\text{ARMA}(p,q): \; y_t = \mu_0 + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} + \epsilon_t.$$

where $y_{t-1}, \ldots, y_{t-p}$ represent past observations (AR) while $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$ are past innovations (MA). All of ϵ_ts are identical and independent centered Gaussian variables (white noise processes). By solving the Yule-Walker equations, the coefficients ϕ_js and θ_ks can be estimated. As wind speed over large geographical area is generally believed to follow a Weibull distribution [86], there is a need of a normalization transformation given by $y = N^{-1}\left[F(w)\right]$, where w denotes the time series representing the wind generation data, F denotes the cumulative distribution function (CDF) of the Weibull distribution associated with the stochastic process, and N denotes the standard normal CDF. Our preliminary tests are captured in Figure 2.1 where six sets of predictions are provided with 95% confidence intervals.

2.6.2 Performance comparison for stochastic economic dispatch

We now compare the performance of the proposed VS-APM scheme with stochastic decomposition and standard stochastic quasi-gradient on an IEEE 118-bus system with 19 generators, 35 synchronous condensers, 177 lines, 9 transformers, and 91 loads. All experiments were implemented in Matlab R2017a with `cplex` employed for solving LPs and QPs.

We begin by comparing VS-APM with stochastic decomposition in a setting where the simulation budget is 1000. In Table 2.2, we compare the behavior of (VS-APM) with SD on 10 problem instances. We observe that (VS-APM) takes less than 1% of the time taken by SD while producing similar objective values. This difference is because (SD) is contending with increasingly larger first-stage problems with (VS-APM) does not have this challenge. Note that the objective value is generated by re-sampling with 20 scenarios. A comparison with standard

$\|\Omega\|$	Iter_SD	SD_mean	Time_SD	Iter_VSAPM	VSAPM_mean	ObjDiff	Time_VSAPM	TimeDiff
1000	1000	1.24E+04	4.14E+03	136	1.22E+04	-1.46%	3.06E+01	0.74%
1000	1000	3.49E+04	3.84E+03	136	3.55E+04	1.83%	2.75E+01	0.71%
1000	1000	7.12E+04	3.90E+03	136	7.16E+04	0.61%	2.89E+01	0.74%
1000	1000	1.11E+04	4.10E+03	136	1.10E+04	-0.50%	2.95E+01	0.72%
1000	1000	3.39E+04	4.35E+03	136	3.42E+04	0.88%	2.78E+01	0.64%
1000	1000	7.15E+04	4.25E+03	136	6.99E+04	-2.22%	2.79E+01	0.66%
1000	1000	1.05E+04	3.83E+03	136	1.02E+04	-1.94%	2.88E+01	0.75%
1000	1000	3.32E+04	3.82E+03	136	3.32E+04	0.07%	2.78E+01	0.73%
1000	1000	6.83E+04	3.82E+03	136	6.83E+04	-0.06%	2.86E+01	0.75%
1000	1000	1.04E+04	3.83E+03	136	1.02E+04	-2.38%	2.82E+01	0.74%

Table 2.2. VS-APM and SD

stochastic gradient provided in Table 2.3 reveals similar benefits in terms of computational time. In this instance, the key benefit lies in taking far less first-stage projection steps, a consequence of utilizing variance reduced schemes.

Iter	SA_mean	Iter	V_mean	Diff	Iter	V_mean	Diff
1000	1.24E+04	65	1.22E+04	-1.54%	136	1.22E+04	-1.58%
1000	3.51E+04	65	3.64E+04	3.67%	136	3.55E+04	1.19%
1000	7.16E+04	65	7.36E+04	2.77%	136	7.16E+04	-0.05%
1000	1.12E+04	65	1.10E+04	-1.73%	136	1.10E+04	-1.73%
1000	3.39E+04	65	3.50E+04	3.19%	136	3.42E+04	0.98%
1000	6.99E+04	65	7.16E+04	2.48%	136	6.99E+04	-0.03%
1000	1.04E+04	65	1.02E+04	-1.61%	136	1.02E+04	-1.56%
1000	3.31E+04	65	3.39E+04	2.40%	136	3.32E+04	0.44%
1000	6.83E+04	65	7.00E+04	2.48%	136	6.83E+04	-0.03%
1000	1.03E+04	65	1.02E+04	-1.52%	136	1.02E+04	-1.49%
1000	3.27E+04	65	3.33E+04	1.76%	136	3.28E+04	0.30%
1000	6.69E+04	65	6.86E+04	2.45%	136	6.70E+04	0.09%
1000	1.21E+04	65	1.21E+04	-0.09%	136	1.21E+04	-0.54%
1000	3.26E+04	65	3.33E+04	2.04%	136	3.28E+04	0.63%

Table 2.3. VS-APM and SA

2.6.3 Risk-based Economic Dispatch

We now consider the risk-based model in settings where the variance of demand is raised from 10 to 50 in steps of 10 while wind penetration is raised from 10% to 30%. We compare the risk-neutral solution with the risk-averse solution in Table 2.4 where $\tau = 0.2$. It is observed that the conditional value of risk increases as variance in demand grows. In addition, we note that the CVaR associated with the risk-neutral solution (CVaR_Mean) is significantly higher than that with the risk-averse solution (CVAR_rED). In effect, solving a risk-neutral model leads to higher risk exposure. We also observe that the value of the stochastic solution (VSS_CVaR) increases as σ and wind penetration levels grow. We conduct further tests on IEEE test networks from MATPOWER and find that risk-neutral solutions lead to higher risk exposure than risk-averse solutions (see Table 2.5). We further examine the impact of variance for the IEEE 300 bus system where demand is assumed to follow a non-normal (beta) distribution. We note that the risk exposure grows as the variance increases and risk-averse models are able to better manage this exposure.

omega_av	wind_per	sigma	fv_Mean	fv_sED	VSS	CVaR_Mean	CVaR_rED	VSS_CVaR
1000	0.2	1	5.68E+04	3.51E+04	2.18E+04	8.09E+04	5.31E+04	2.78E+04
1000	0.3	1	1.03E+05	7.08E+04	3.19E+04	1.45E+05	1.06E+05	3.86E+04
1000	0.4	1	1.58E+05	1.09E+05	4.84E+04	2.17E+05	1.63E+05	5.41E+04
1000	0.2	10	5.70E+04	3.54E+04	2.16E+04	8.45E+04	5.74E+04	2.71E+04
1000	0.3	10	1.03E+05	7.08E+04	3.20E+04	1.49E+05	1.08E+05	4.10E+04
1000	0.4	10	1.58E+05	1.09E+05	4.83E+04	2.20E+05	1.65E+05	5.46E+04
1000	0.2	20	5.77E+04	3.63E+04	2.14E+04	9.62E+04	7.05E+04	2.58E+04
1000	0.3	20	1.03E+05	7.09E+04	3.25E+04	1.61E+05	1.20E+05	4.03E+04
1000	0.4	20	1.58E+05	1.09E+05	4.84E+04	2.35E+05	1.77E+05	5.81E+04
1000	0.2	30	5.90E+04	3.75E+04	2.15E+04	1.14E+05	8.54E+04	2.86E+04
1000	0.3	30	1.04E+05	7.12E+04	3.31E+04	1.79E+05	1.35E+05	4.36E+04
1000	0.4	30	1.58E+05	1.09E+05	4.86E+04	2.52E+05	1.92E+05	6.04E+04
1000	0.2	50	6.33E+04	4.08E+04	2.24E+04	1.51E+05	1.18E+05	3.26E+04
1000	0.3	50	1.07E+05	7.28E+04	3.45E+04	2.20E+05	1.73E+05	4.68E+04
1000	0.4	50	1.60E+05	1.10E+05	4.94E+04	2.94E+05	2.28E+05	6.63E+04

Table 2.4. Value of Stochastic Solution

	Mean Cost	Two-stage Cost	Risk Cost	Mean worst 20%	Two-stage worst 20%	Risk worst 20%
IEEE118B	59533.26729	44793.66107	59530.76254	268179.7149	268179.1903	198998.0665
IEEE145	46916705.35	46851098.4	46853310.19	212284415.2	212077890.7	212070666.7
IEEE300A	47577079.24	39356294.75	39689745.57	64922699.66	56087603.98	55719991.89
IEEE300B	87018479.75	70139581.54	70877282.34	97323306.35	79622883.48	78862567.52
IEEE300C	86642730.74	69798253.99	70560254.99	89515610.86	73364330.13	72596519.46

Table 2.5. Result on different networks

Mean Cost	Two Stage Cost	Risk Cost	Mean worst 20%	Two Stage worst 20%	Risk worst 20%	Variance
86644497.45	69773646.04	70537357.47	89044111.67	72892328.35	72123658.01	0.00507185
86886563.58	70095760.05	70834083.12	92983380.85	76200148.71	75475844.68	0.01984127
87601203.2	70675269.16	71397447.66	98226025.94	80316099.73	79545661.2	0.0375
90436594.28	73754429.32	74431960.72	111269137.4	93425043.78	92685828.32	0.06857143

Table 2.6. Result on different variance for IEEE 300 system

2.6.4 Case study: ARPA-E Network

ARPA-E networks are networks that been put in use of Grid Optimization (GO) Competition. The goal is have a real-time matching of instantaneous electricity generation and demand, which requires utilities, grid operators, and other stakeholders to use a variety of sophisticated software operating across a wide range of timescales.

One test network among those was chosen and modified to be tested on. The numeric test is conducted on ARPA-E "Original Dataset Real-Time Network_01-10R". This network contains 500 buses, with 90 generators, 468 branches, 262 transformers and 371 contingencies with each contingency represents one generator failure or an branch or transformer failure. The original network is for ACOPF Some necessary modifications are made to conduct DCOPF experiment. DCOPF is an relaxation to original thus in order to tighten constraints to introduce recourse to second stage, loads are modified to double loads.

Mean Cost	Risk Cost	Two-stage Cost	Mean Risk	Risk Risk	Two-stage Risk	Samples
187878.701	187878.701	187878.701	245966.5256	245966.5256	245966.5256	0
187878.701	176647.6331	173831.485	245966.5256	200140.8446	205645.2917	7967
187878.701	176645.5224	173831.4252	245966.5256	200140.6522	205635.3819	7968
187878.701	176486.4437	173829.7477	245966.5256	200132.5389	205285.9083	15934
187878.701	176487.8997	173829.7249	245966.5256	200132.4342	205294.8453	15935
187878.701	176747.7048	173828.9865	245966.5256	200128.5912	207123.1018	23901
187878.701	176778.0333	173828.9672	245966.5256	200128.5051	207307.6371	23902
187878.701	176723.3841	173828.4069	245966.5256	200125.7028	206871.7545	31868
187878.701	176723.4163	173828.3928	245966.5256	200125.6382	206871.9366	31869
187878.701	176735.4956	173828.1031	245966.5256	200124.0864	206818.1594	39835
187878.701	176735.5698	173828.0869	245966.5256	200124.0095	206818.7758	39836

Table 2.7. Result on ARPA-E network

Comparing against mean value solution, we can find stochastic solution provides better average performance in terms of overall cost of both pre-contingency and post contingency. With mean value solution gets final average cost of 187878.70, risk-based model get 176735.57 and standard two-stage model gets 173828.09, it shows that both risk-based model and standard two-stage model can reduce expected cost but in terms of average cost standard two-stage model perform best. However, when considering worst 20% scenarios, with mean value solution gets final risk of 245966.53, risk-based model get 200124.01 and standard two-stage model gets 206818.78, it shows that both risk-based model and standard two-stage model can reduce risk but risk-based model perform best.

Chapter 3
Two-stage stochastic integer programming via Stochastic Approximation

3.1 Introduction

In the prior chapter, we focused on risk-based two-stage problems in which a CVaR-based risk-measure is employed. In this chapter, we consider the development of a framework for resolving two-stage stochastic integer programming problems defined as follows.

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}[\mathcal{Q}(x,\omega)] \\ \text{subject to} \quad & x \in X \subseteq \mathbb{R}^n, x_{\mathcal{I}_1} \in \mathbb{Z}_+, \end{aligned} \tag{SIP}$$

where X is a closed and convex set, f is a convex and continuously differentiable function, $\mathcal{I}_1 \subseteq \{1,\ldots,n\}$ denotes an index set that specifies the integer variables, while $\mathcal{Q}(x,\omega)$ is defined as follows:

$$\begin{aligned} \min_y \quad & d_\omega^T y \\ \text{subject to} \quad & W_\omega y = h_\omega - T_\omega x \\ & y \geq 0, y_{\mathcal{I}_2} \in \mathbb{Z}_+. \end{aligned} \tag{SIP-rec(ω)}$$

Finally, we assume that the sample-space is finite in that ω takes on a finite number of realizations. Consequently, the deterministic equivalent is a finite dimensional mixed-integer program.

Such problems have tremendous applicability and arise in a range of applications in transportation, supply-chain modeling, and energy systems [87–92]. Our research is motivated by the commitment of generation resources in a two-period setting, a class of problems referred to as stochastic unit commitment. These problems represent a generalization of economic dispatch problems in which the commitment decisions can be 0/1 decisions.

A key challenge that arises in two-stage stochastic integer problems of the form (SIP) is the observation that the recourse function $\mathcal{Q}(x) \triangleq \mathbb{E}[\mathcal{Q}(x, \omega)]$ is discontinuous (and nonconvex) in first-stage variables. This precludes a direct application of the schemes from the previous chapter. To address this challenge, there have been a host of schemes that have been developed to address various subclasses of two-stage stochastic integer programs including cut-based schemes [26, 93, 94], Lagrangian relaxation schemes [28, 95–97], progressive hedging techniques [98, 99], amongst others. Unfortunately such schemes are often complicated by a key shortcoming in that such techniques are often customized for mixed-integer linear programs in the second-stage. Consequently, they cannot contend with nonlinearity and risk-aversion (which leads to nonsmooth relaxations). This represents a significant gap and motivates the current research.

In this chapter, we develop a novel scheme in which we employ a branching scheme which draws inspiration from standard techniques for deterministic mixed-integer programs. Specifically, branching is carried out on integer variables in both the first and second-stage at each node of this tree. The resulting problems at each node are essentially two-stage convex programs (under the caveat that a continuous relaxation of the original problem is a two-stage convex program) and these are processed by the VS-APM scheme presented in the previous chapter. The key advantage of such an avenue is that it allows for far broader models in that the first-stage and second-stage problems may be both nonsmooth and nonlinear.

The remainder of the chapter is partitioned into three sections. In Section 3.2, we survey the literature on mixed-integer programming and its stochastic variants as well as some motivating applications from power systems operations. We present the algorithm in Section 3.3 and provide a detailed set of numerics on a class of power systems operations problems in Section 3.4.

3.2 Literature review

In this section, we summarize prior work for the solution of deterministic and stochastic integer programming as well as on motivating applications from power systems operation.

3.2.1 Deterministic mixed-integer programs

One of the most successful approaches for solving mixed-integer programming problems is the branch-and-bound scheme [100] (cf. [101] for a more general survey). Applications to nonlinear integer programming with convex relaxations have also been studied [102–104]. The branch-and-cut variant [101, 105–107] significantly tighten the problem by adding polyhedral cutting planes after branching that allow for better use of fathomed nodes. We provide a brief description of branching schemes in the context of a mixed-binary quadratic program (MBQP) where Q is assumed to be positive semidefinite for the present. In order to start the branch-and-bound scheme, we first relax all integrality constraints in (MBQP) to obtain a relaxed problem (MBQP_0).

$$\begin{aligned} \min \quad & \frac{1}{2} x^T Q x + c^T x \\ \text{subject to} \quad & Ax \;=\; b, \\ & x \;\geq\; 0, \text{ for } i \;\in\; \mathcal{I}. \end{aligned} \tag{MBQP_0}$$

Since (MBQP_0) is a relaxation (with x^0 as an optimal solution), the optimal value of (MBQP_0) is a lower bound of the optimal value of (MBQP). Further, if x^0 satisfies all integrality requirements in (MBQP), it is a feasible solution to (MBQP) and also the optimal solution. Otherwise, there must be a variable $x_j \in x^0$ which is not integer. Then we may form two subproblems from (MBQP_0) by adding bound $x_j \leq [x_j]$ to one and $x_j \geq [x_j + 1]$ to the other, where $[x_j]$ represent the largest integer not greater than x_j. The subproblem (MBQP_1) is as shown below:

$$\begin{aligned} \min \quad & \frac{1}{2}x^T Q x + c^T x \\ \text{subject to} \quad & Ax \;=\; b, \\ & x_j \;\leq\; [x_j], \\ & x \;\geq\; 0. \end{aligned} \tag{MBQP_1}$$

Similarly, (MBQP_2) may be defined as follows:

$$\begin{aligned} \min \quad & \frac{1}{2}x^T Q x + c^T x \\ \text{subject to} \quad & Ax \;=\; b, \\ & x_j \;\geq\; [x_j + 1], \\ & x \;\geq\; 0. \end{aligned} \tag{MBQP_2}$$

The process of forming subproblems by utilizing constraints of the form $x_j \leq [x_j]$ and $x_j \geq [x_j] + 1$ is referred to as branching. The convex programming subproblems are continuous problems and may be solved via standard schemes and this process may be repeated for each integer variable. The entire algorithm continues as a tree search is performed with each node representing a continuous subproblem. In most cases, it is not necessary to search the entire tree. Once we obtain a feasible integer solution to one of the continuous problems, the corresponding value of the objective function represents an upper bound to the original problem. The subproblems with continuous solutions with objective value higher than the upper bound may be excluded from further consideration (a process referred to as fathoming). In addition, if it is recognized that a node leads to infeasibility, then it will also be fathomed since further branching from that node cannot lead to feasibility. The branching and fathoming continues as the upper bound keeps reducing (as one finds improved feasible integer solutions) while the lower bounds keep increasing through the addition of constraints. Thus, the gap between upper bound and lower bound keeps decreasing and scheme may be terminated when this gap is sufficiently small.

Another popular method for solving mixed-integer programming problems is

Lagrangian relaxation [108]. In 1970 [109, 110], a Lagrangian relaxation approach based on minimum spanning tree was used to devise a algorithm for the traveling salesman problem. In [111, 112], such techniques were applied to mixed-integer programming (cf. [113]). Consider the problem (MBQP) and suppose the linear constraint is relaxed using the Lagrange multiplier λ, leading to the relaxed problem LR(u).

$$\begin{aligned} L(u) = \min \quad & \frac{1}{2}x^TQx + c^Tx + u^T(Ax - b) \\ \text{subject to} \quad & x \;\geq\; 0. \end{aligned} \tag{LR(u)}$$

where u denotes the Lagrange multiplier. In many cases, the relaxation (LR$_u$) may either have a special structure (and consequently be easier to solve compared to (MBQP)). In addition, the relaxation may allow for decomposition of this problem into smaller (and possibly) structured problems that are amenable to faster solutions. Suppose we denote z as the origin optimal objective value and x^* as the optimal solution to the original problem. Then we may show that $L(u) \leq \frac{1}{2}(x^*)^TQx^* + c^Tx^* + u^T(Ax^* - b) = z$. The solution to the Lagrangian subproblem allows for updating the multiplier estimate u which in turn will provide a new primal solution. In the limit, it may be shown that the sequence of primal solutions is the solution to (P) while the sequence of dual solutions converges to the true Lagrange multiplier.

We now briefly discuss how one may contend with a mixed-binary nonlinear program (MINLP) of the following form:

$$\begin{aligned} Z \;= \min \quad & c^Ty + f(x) \\ \text{subject to} \quad & By + g(x) \;\leq\; 0, \\ & x \;\in\; \mathcal{X},\; y \;\in\; \{0, 1\}, \end{aligned} \tag{MINLP}$$

where $\mathcal{X}$ is a polyhedral set and f and g are convex functions in their arguments. The generalized Benders decomposition in [114] algorithm divides variables into sets of complicating and non complicating variables. Using a sequence of nonlinear programming (NLP) subproblems and mixed-integer linear programming (MILP) master problems to solve the original MINLP. The master problem, a MILP, can be stated as below:

$$
\begin{aligned}
z = \min \quad & \alpha \\
\text{subject to} \quad & \alpha \ \geq \ c^T y + f(x^k) + (\lambda^k)^T [By + g(x^k)], \qquad k \ = \ 1, \ldots, K_{\text{feasible}}, \\
& (\lambda^k)^T [By + g(x^k)] \ \leq \ 0, \qquad\qquad\qquad\qquad\quad k \ = \ 1, \ldots, K_{\text{infeasible}}, \\
& x \in \mathcal{X}, \ y \ \in \ \{0, 1\},
\end{aligned}
$$

where z denotes the lower bound, (x^k, λ^k) are the optimal primal and dual variables of the NLP subproblems, and K_{feasible}, $K_{\text{infeasible}}$ refer to feasible and infeasible subproblems. The solution of the master problem specifies the values of the binary variables that are then parameters in the subsequent NLP subproblem. The NLP subproblem provide a decreasing sequence of upper bounds while the master problems provides an increasing sequence of lower bounds. When these bounds lie within a suitable tolerance, the scheme terminates. In [115–117], related outer-approximation schemes for solving MINLPs are also presented.

3.2.2 Stochastic mixed-integer programming

Early efforts to resolve stochastic integer programs were restricted to first-stage integers which was easily addressed by extending cutting-plane schemes [61] to account for mixed-integer master problems (rather than purely linear programs). In this subsection, we discuss the process of adding first and second-stage cuts.

3.2.2.1 First stage cuts

For SMIP problems with binary first stage and arbitrary second stage, many algorithm are based on generating first stage cuts. The inequalities of Laporte and Louveaux [93] are used follows the idea of Benders' decomposition (or L-shaped method). That is, at each iteration k, we solve one master program, and as many subproblems as there are outcomes of the random variable. Despite the non-convexity of value functions of general optimization problems (including MIPs), the valid inequality provided by Laporte and Louveaux [93] is linear. Thus, the linearity derives from a property of the binary first-stage variables.

At iteration k, let first-stage decision x^k be given, and have the following two index sets:

$$I_k = \{i|x_i^k = 1\}, Z_k = \{1, ..., n_1\} - I_k$$

Then we can define the linear function:

$$\delta_k(x) = |I_k| - \left[\sum_{i\in I_k} x_i - \sum_{i\in Z_k} x_i\right]$$

When $x = x^k$, it is easy to verify that $\delta_k(x) = 0$, for $x \neq x^k$, there is $\delta_k(x) \geq 1$. For a well-defined recourse function with lower bound $\mathcal{Q}_l$, we have:

$$\theta \geq \mathcal{Q}(x^k) - \delta_k(x)\left[\mathcal{Q}(x^k) - \mathcal{Q}_l\right]$$

This is the optimality cut of Laporte and Louveaux [93]. Thus, the algorithm can be state as follows:

Algorithm 4 First stage cut

1: **initialization:** $k = 0, \epsilon \geq 0, x^1$ and $\mathcal{Q}_l$ be given. Define $\delta_0(x) = \mathcal{Q}_l, f_u 0\infty$;
2: **while** $f_u - f_l > \epsilon$ **do**
3: Obtain a cut: Solve a second stage problem and define a cut $\alpha + \beta x = \mathcal{Q}(x^k) - \delta_k(x)\left[\mathcal{Q}(x^k) - \mathcal{Q}_l\right]$.
4: Update approximation:

1. Define $\delta_k(x) = \max\{\delta_{k-1}(x), \alpha + \beta x\}$ and $f_k(x) = c^x + \delta_k(x)$.

2. Update the upper bound: $f_u = \min\{f_u, f_k(x^k)\}$.

5: Solve master problem: solve $\min\{f_k(x)|x \in X \cap \mathbb{B}\}$ to get x^{k+1} and $f_l = f_k(x^{k+1})$.
6: **end while**
7: Return x^k as an ϵ-optimal solution

This algorithm is very close to Kelley cutting plane method and L-shape method, with Branch and Bound involve in generating cut and solving master problem.

3.2.2.2 Cuts in both stages

For binary variables in both stage, a sequential convexification of the integer recourse problem is applied. Such sequential convexification can avoid the need to solve every subproblem from scratch in each iteration. Ntaimo and Sen [118] derived such cuts via disjunctive programming. Sherali and Fraticelli [27] also generated such cuts using Reformulation-Linearization Technique (RLT). In terms of theory of sequentially constructing approximation, one important result known as Common Cut Coefficients (C3) Theorem is introduced by Higle and Sen [119] , which allows convex approximation built recursively.

Theorem 3.2.1 (The C^3 Theorem)**.** *Consider the stochastic program with fixed recourse as stated. Given* (x, ω)*, let* $Y(x, \omega) = \{y = (u, z)|Wy \geq r_\omega - T_\omega x, u \in \mathbb{R}_+^{n_1}, z \in \mathbb{Z}_+^{n_2}\}$*, the set of mixed-integer feasible solutions for the second stage MILP. Suppose that* $\{C_h, d_h\}_{h\in H}$ *, is a finite collection of appropriately dimensioned matrices and vectors such that for all* $(x, \omega) \in X \times \Omega$*,*

$$Y(x, \omega) \subseteq \cup_{h\in H}\{y \in \mathbb{R}_+^{n_2}|C_h y \geq d_h\}$$

Let

$$S_h(x, \omega) = \{y \in \mathbb{R}_+^{n_2}|Wy \geq r_\omega - T_\omega x, C_h y \geq d_h\}$$

and let

$$S(x, \omega) = \cup_{h\in H} S_h(x, \omega)$$

Let $(\bar{x}, \bar{\omega})$ *be given, and suppose that* $S_h(\bar{x}, \bar{\omega})$ *is nonempty for all* $h \in H$ *and* $\pi^T y \geq \pi_0(\bar{x}, \bar{\omega})$ *is a valid inequality for* $S(\bar{x}, \bar{\omega})$*. Then there exists a function,* $\pi_0 : X \times \Omega \to \mathbb{R}$ *such that for all* $(x, \omega) \in X \times \Omega, \pi^T y \geq \pi_0(x, \omega)$ *is a valid inequality for* $S(x, \omega)$*.*

This theorem ensures that with a simple translation, valid inequalities derived for one pair $(\bar{x}, \bar{\omega})$ may be used to derive valid inequalities for any other pair (x, ω). Thus, a Disjunctive Decomposition, or D^2, algorithm for SMIP is introduced in [119].

Algorithm 5 D^2 Algorithm

1: **initialization:** $k = 0, \epsilon \geq 0, x^1$ and $\mathcal{Q}_l$ be given. Define $\delta_0(x) = \mathcal{Q}_l, f_u 0 \infty$;
2: **while** $f_u - f_l > \epsilon$ **do**
3: Solve one LP Subproblem for each $\omega \in \Omega$: Solve a second stage problem and define a cut $\alpha + \beta x = \mathcal{Q}(x^k) - \delta_k(x)\left[\mathcal{Q}(x^k) - \mathcal{Q}_l\right]$.
4: Solve Multiplier/Cut Generation LP:

   1. Define $\delta_k(x) = \max\{\delta_{k-1}(x), \alpha + \beta x\}$ and $f_k(x) = c^x + \delta_k(x)$.

   2. Update the upper bound: $f_u = \min\{f_u, f_k(x^k)\}$.

5: Update and Solve one LP Subproblem each $\omega \in \Omega$:
6: Update and Solve master problem: solve $\min\{f_k(x) | x \in X \cap \mathbb{B}\}$ to get x^{k+1} and $f_l = f_k(x^{k+1})$.
7: **end while**
8: Return x^k as an ϵ-optimal solution

Besides the primal methods that work with subproblems assigned to time stages, another type of decomposition methods are dual methods in which subproblems are assigned to scenarios. A dual decomposition method was proposed in [28]. For a SMIP, the deterministic equivalent can be written as:

$$\begin{aligned} \min \quad & c^T x + \sum_{j=1}^{r} p^j q^j y^j \\ \text{subject to} \quad & (x, y^j) \in S^j, \quad \forall j = 1, ..., r \end{aligned}$$

where S^j is the feasible region for scenario j. Then the idea of scenario decomposition is to introduce copies $x^1, x^2, ..., x^r$ of the first-stage variable x and the problem becomes:

$$\begin{aligned} \min \quad & \sum_{j=1}^{r} (c^T x^j + p^j q^j y^j) \\ \text{subject to} \quad & (x^j, y^j) \in S^j, \quad \forall j = 1, ..., r \\ & x^1 = ... = x^r \end{aligned}$$

where the constraint $x^1 = ... = x^r$ often referred as non-anticipativity constraint which state that the first-stage decision are independent from scenarios. Then the Lagrangian relaxation with respect to the non-anticipativity constraint is the problem of finding x^j, y^j solves:

$$\begin{aligned} D(\lambda) = \min \quad & \sum_{j=1}^{r} L_j(x^j, y^j, \lambda) \\ \text{subject to} \quad & (x^j, y^j) \in S^j, \quad \forall j = 1, ..., r \end{aligned}$$

where λ is the Lagrangian multiplier of non-anticipativity constraint and

$$L_j(x^j, y^j, \lambda) = p^j(c^T x^j + q^j y^j) + \lambda(x^r - x^1)$$

is the Lagrangian function. So the original problem becomes

$$\max_{\lambda} D(\lambda) = \sum_{j=1}^{r} D_j(\lambda)$$

where

$$\begin{aligned} D_j(\lambda) = \min \quad & L_j(x^j, y^j, \lambda) \\ \text{subject to} \quad & (x^j, y^j) \in S^j \end{aligned}$$

Each of these r subproblems is a mixed-integer problem and subgradient method was used in [28]. Under this Lagrangian relaxation, Lulli and Sen [120] propose branch-and-price (BP) algorithm using column generation procedure. A detailed comparison and computational behavior study can be found in [121].

3.2.3 Stochastic unit commitment problems

The unit commitment problem considers the determination of the optimal production schedule of power generating units, so that in a certain amount of time the operational cost may be minimized while meeting demand requirements and physical constraints. Basically, binary variables represent the status of unit. An overview of unit commitment problem in literature was provided in [87]. In [122, 123], branch and bound schemes were used to solve the unit commitment problem while the Lagrangian relaxation is also widely used in solving the unit commitment problem [88–92].

Unit commitment problems consider the scheduling of power production over a certain period of time. Define the planning horizon as T and a time step $t \in \mathcal{T} \triangleq 1, \ldots, T$ while the set of generators is denoted by $\mathcal{I} = 1, \ldots, I$. For any generator, once it is switched on, it cannot be turned off immediately. Similarly, it cannot be turn on immediately after it is switched off. Each generator must follow a minimum up and down times rule. As stated in [124,125], for any $t \in \mathcal{T}$ and generator $i \in \mathcal{I}$, we denote the generator state (on or off), start-up, and production decisions by y_{it}, z_{it}, and $g_{i,t}$ respectively. In a two-stage decision process, we note that y_{it} and z_{it} are day-ahead first-stage decisions made before real-time production while $g_{it,\omega}$ represents production decisions made during real-time during scenario ω.

Objective function The objective of unit commitment problem is to minimize the operation cost. Every time a generator is turned on, a startup cost need to be considered due to the fuel and electrical power consumed. We denote the startup cost by f^y_{it} as determined by y_{it}. During the running time of a generator, a running cost needs to be considered. We denote the running cost by f^z_{it} which is determined by z_{it}. The cost of generation may be presented as $\sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} f^g_{it}$ which is determined by g_{it}. The objective function is defined as follows:

$$f(y,z,g) = \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} f^y_{it}(y_{it}) + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} f^z_{it}(z_{it}) + \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}} \mathbb{E}[f^g_{it,\omega}(g_{it,\omega})]. \tag{3.1}$$

Startup and shutdown constraints We adopt the linear formulation of [124, 125]. We observe that the startup decision y and the operational decision z are related. The variable $y_{i,t+1}$ will be 1 only when a unit is off at time t but on at time $t+1$. Also we have $y_{it} \geq 0$, we can represent it as the following constraint

$$y_{it} \geq z_{it} - z_{i,t-1}, y_{it} \geq 0, \qquad \forall i \in \mathcal{I}, \quad \forall t \in \mathcal{T}. \tag{3.2}$$

Each unit has a minimum up and down time constraint and denote these times by L_i and l_i corresponding to generator $i \in \mathcal{I}$. Then the constraint can be written as:

$$\begin{aligned} z_{it} - z_{i,t-1} &\leq z_{i,\gamma}, 2 \leq t \leq T, \forall \gamma \in \{t+1, \ldots \min(t+L_i-1, T)\}, \forall i \in \mathcal{I}, \\ z_{i,t-1} - z_{it} &\leq 1 - z_{i,\gamma}, 2 \leq t \leq T, \forall \gamma \in \{t+1, \ldots, \min(t+l_i-1, T)\}, \forall i \in \mathcal{I}. \end{aligned} \tag{3.3}$$

Once generator i has been switched on at time t, the unit continues to be on for at least $L_i - 1$ time units. Using $z_{it} - z_{i,t-1}$ to represent y_{it} we get the above constraint. The down time constraint goes the same.

Generation constraints For each generator, there must be a minimum and a maximum generations bound. Let Q_i and q_i be the maximum and minimum generation level for generator i. We have the following constraints:

$$\underline{q}_i^{\omega} z_{it} \leq g_{it,\omega} \leq \bar{q}_i^{\omega}, \ \forall i \in \mathcal{I}, \ \forall t \in \mathcal{T}, \forall \omega. \tag{3.4}$$

Meanwhile, the predicted demand need to be satisfied.

$$\sum_{i \in \mathcal{I}} g_{it,\omega} \geq d_t^{\omega}, \ \forall t \in \mathcal{T}, \quad \forall \omega. \tag{3.5}$$

General Form By combining objective function (1.1) and constraints (1.2) - (1.5), we can get the following model of unit commitment problem:

$$\begin{aligned} \min \quad & f(y,z,g) = \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} f_{it}^{y}(y_{it}) + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} f_{it}^{z}(z_{it}) + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \mathbb{E}[\mathcal{Q}(y_{it}, z_{it}, \omega)] \\ \text{subject to} \quad & (3.1) \ldots (3.3), \end{aligned}$$

where $\mathcal{Q}(y_{it}, z_{it}, \omega)$ is defined as

$$\begin{aligned} \min \quad & \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} f_{it,\omega}^{g}(g_{it,\omega}) \\ \text{subject to} \quad & (3.4) \ldots (3.5). \end{aligned}$$

In this setting, the second-stage problem does not allow for starting up or shutting down generation assets. As a consequence, one can directly employ more standard cutting-plane schemes for resolving such problems with the caveat that the master-problem is complicated by the presence of integrality requirements. Instead, one could always extend this model to accommodate second-stage binary decisions.

3.3 VS-APM in SMIP

A natural extension on branch-and-bound is to introduce VS-APM on solving resulting continuous relaxation. Consider a standard SMIP as follows:

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}[\mathcal{Q}(x,\omega)] \\ \text{subject to} \quad & x \in \mathcal{X} \cap X \end{aligned}$$

where $\mathcal{Q}(x,\omega)$ can be represent as:

$$\begin{aligned} \mathcal{Q}(x,\omega) \quad & = \min g_\omega^T y \\ \text{subject to} \quad & W_\omega y \geq r_\omega - T_\omega x \\ & y \geq 0, y_j \in \mathbb{Z}, j \in J_2 \end{aligned}$$

Then its corresponding continuous relaxation can be represented as:

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}[\mathcal{Q}(x,\omega)] \\ \text{subject to} \quad & x \in X \end{aligned}$$

where $\mathcal{Q}(x,\omega)$ can be represent as:

$$\begin{aligned} \mathcal{Q}(x,\omega) \quad & = \min g_\omega^T y \\ \text{subject to} \quad & W_\omega y \geq r_\omega - T_\omega x \\ & y \geq 0 \end{aligned}$$

which results in a standard stochastic programs. Apply VS-APM will result in a set of solution (x, y), which serves as a foundation of integer program framework.

3.3.1 Cuts

In theory, MIP is an NP-hard problem. In order to solve it in reasonable time, special technique like adding cut can be helpful. Cutting planes were proposed by Ralph Gomory in 1958 [126] for solving integer programming and mixed-integer programming problems. Later by Padberg and Rinaldi in 1987 [105] branch-and-cut was proposed for more general use of cut in integer problems. The idea behind

Gomory's method is to initially neglect the integrality requirements and solve the corresponding linear programming problem then add corresponding cuts to reduce feasible. In [126], finite cutting plane algorithm for pure IPs was developed and it is extended to MIPs with integer objective in [127]. For a integer set as follows:

$$Q = \{x \in \mathbb{Z} : x \leq b_1, x \geq b2\}$$

Then there is a valid inequalities for such set as follow:

$$x \leq \lfloor b_1 \rfloor, x \geq \lceil b_2 \rceil$$

Thus, the following set contains all feasible integer point in Q:

$$Q_0 = \{x \in \mathbb{R} : x \leq \lfloor b_1 \rfloor, x \geq \lceil b_2 \rceil\}$$

After solving corresponding relaxation, Q_0 can be construct expression like $ax \leq b$ with all of coefficients of a are integer for integer variable x. Then the corresponding Gomory cuts are $ax \leq \lfloor b \rfloor$. Since Simplex could provide pure integer coefficient, like identity for basic variables, Gomory cut can be easily and often related to Simplex method.

For same principles as Gomory cut, Mixed integer rounding cuts (MIP) were derived and are also among the most effective cutting plane methods. The basic mixed-integer rounding idea can be found in [128]. For a mixed integer set

$$Q = \{x \in \mathbb{R}, y \in \mathbb{Z} : x + y \geq b, x \geq 0\}$$

There is a MIR inequality:

$$x \geq \hat{b}(\lceil b \rceil - y)$$

where $\hat{b} = b - \lfloor b \rfloor$, is valid for Q. In general for a multiple constraints set like:

$$P = \left\{x \in \mathbb{R}^{|C|}, y \in \mathbb{Z}^{|I|} : Ax + By \geq d, x, y \geq 0\right\}$$

where $A \in \mathbb{R}^{m \times |C|}, B \in \mathbb{R}^{m \times |I|}, d \in \mathbb{R}^m$. The MIR inequalities can be obtain by multiply original constraint by $\lambda \in \mathbb{R}^m$, then follow the idea of MIR we can get:

$$\sum_{i \in C} (\lambda A_i)^+ + \hat{b} \sum_{i \in I} \lfloor \lambda B_i \rfloor y + \sum_{i \in I} \min \left\{ \lambda A_i - \lfloor \lambda A_i \rfloor, \hat{b} \right\} y \geq \hat{b} \lceil \lambda d \rceil$$

where $\hat{b} = b - \lfloor b \rfloor$.

3.3.2 Branching

After adding cut, if the resulting solution of relaxation lies in feasible integer set of the original problem then we get the optimal solution. Otherwise, there must be an $x \notin \mathcal{X} \cap X$ or an $y_j \notin \mathbb{Z}$ for some $j \in J_2$ in some scenario. Then we can separate the feasible region X or Y into two based on infeasible integer variables. In implementation, the separation is conducted via bound on variables. The branch-and-bound tree structure is implemented through recursive functions. The algorithm can be summarized as follows:

Algorithm 6 VS-APM branch-and-bound

1: **initialization:** $k = 0, \epsilon \geq 0, f_u = \infty, f_l = -\infty$, set bound for x, y as lb_x, ub_x, lb_y, ub_y;

2: **while** $f_u - f_l > \epsilon$ **do**

3: Solve a relaxation using VS-APM:

1. If the relaxation is infeasible; Return with f_u, f_l;
2. Otherwise, get a solution x, y with relaxed optimal value f_0.

4: Examine feasibility of current solution:

1. If (x, y) is feasible; Return with $f_u = \min\{f_u, f_0\}$ and $f_l = \min\{f_l, f_0\}$;
2. Otherwise, branch on the first element of (x, y) that violates integrality and set $f_l = \min\{f_l, f_0\}$;

5: Gather information from branches:

1. If a branch is infeasible or with relaxation objective value $f_1 > f_u$, then prune this branch and return.
2. Otherwise, denote results from two branches are f_{u1}, f_{l1} and f_{u2}, f_{l2}. Then return with $f_u = \min\{f_{u1}, f_{u2}, f_u\}$ and $f_l = \min\{f_{l1}, f_{l2}, f_l\}$;

6: **end while**

7: Return x as an ϵ-optimal solution

In terms of branching, branch rules can be important to performance. In worst case, a branch process may need to travel all nodes in the tree. Several branch rules like breadth-first search (BFS), depth-first search (DFS) and best-first branch. Here we use a combination of both DFS and best-first by choosing the closest integer as branch direction and always branch to a leave to get a feasible solution as soon as possible.

3.4 Numerics

3.4.1 First stage integer

We tested on SSN problem from [129] which is a telecommunications network planning problem. The first stage decision variables are defined as capacity to be added to a certain link. The original problem has 89 first stage variables and 706 second stage variables. The scenarios are generated from a discrete distribution with total 571 values for 86 random variables. To test on integer performance, we set some of the first stage variables to be binary and compare against Cplex with it solving deterministic equivalent. The result is as follows:

Scenario#	int_size	f_op	time_cplex	bb	time_bb	diff
10	10	2.42E+02	7.03E-01	2.43E+02	1.61E+01	0.570%
20	10	2.32E+02	2.36E+00	2.32E+02	3.44E+01	0.173%
100	10	2.27E+02	6.02E+00	2.27E+02	1.75E+02	0.053%
500	10	2.40E+02	5.45E+01	2.40E+02	1.06E+03	0.100%
1000	10	2.39E+02	1.64E+02	2.39E+02	1.81E+03	0.021%
5000	10	2.36E+02	3.44E+03	2.36E+02	8.60E+03	0.004%
10000	10	2.38E+02	1.57E+04	2.38E+02	1.70E+04	0.004%
12000	10	2.36E+02	2.14E+04	2.36E+02	1.95E+04	0.004%

Table 3.1. Result on SMIP with first stage integer

From this result we can find that, as the number of scenario increases, the difference between Cplex result and VS-APM result is decreasing. Also, the computation time of Cplex is increasing near exponentially as the corresponding dimension of matrix grows, while the computation time of VS-APM increases in a steady near linear rate and out perform Cplex at 12000 scenarios.

3.4.2 Second stage integer

For integer variables on both stage, we modified the original problem to a problem with integer variables on both stage and with multiple feasible integer solutions. We tested on both pure branch-and-bound and branch-and-bound with cut. Also tested on different number of integer variables.

For pure branch and bound, we set the time limit to be 50 mins for each

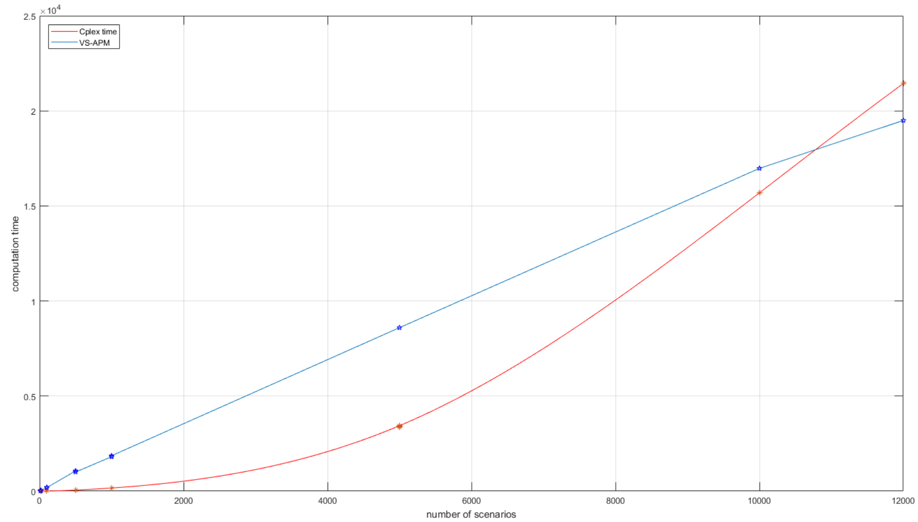

Figure 3.1. Computation time comparison

Scenario#	Nodes	LB	UB	Gap	Cplex Solution	difference
100	7581	-180.2646	-172.558	4.47%	-179.7069	3.98%
200	4185	-179.4568	-170.933	4.99%	-178.9130	4.46%
300	2818	-179.1501	-171.417	4.51%	-178.4725	3.95%
400	1975	-178.9119	-171.554	4.29%	-178.2710	3.77%
500	1726	-178.7943	-170.976	4.57%	-178.0647	3.98%

Table 3.2. Pure BnB for 2 stage integer

problem, and each problem has 5 first stage integer (binary) variables and 1 second stage integer (binary) variables for each scenario. From results in 3.2 we can find pure branch and bound can give a feasible solution but the optimality gap is not small enough to stop. Thus, it still needs to travel through many nodes which can be extremely time consuming as number of scenarios and number of integer variables grows.

Hence, we introduced MIP cuts and best first and DFS branching rule into branch and bound. For the same 5 problems, we have the result as 3.3. From this table we can find with cut and branching rule, branch and bound with VS-APM could provide good enough solution as cplex within few nodes. With the introduction of integer programs techniques, VS-APM can be extended to much complicated stochastic mixed integer programs.

Scenario#	Nodes	LB	UB	Gap	Cplex Solution	difference
100	39	-180.2646	-179.7069	0.31%	-179.7069	0.00%
200	84	-179.4568	-178.9130	0.30%	-178.9130	0.00%
300	124	-179.1501	-178.4729	0.38%	-178.4725	0.00%
400	167	-178.9119	-178.2711	0.36%	-178.2710	0.00%
500	166	-178.7943	-178.0649	0.41%	-178.0647	0.00%

Table 3.3. Bnb with cut and branching rule for 2 stage integer

Integer#	Nodes	LB	UB	Gap	Cplex Solution	difference
105	39	-180.265	-179.7069	0.31%	-179.7069213	0.00%
205	41	-178.882	-178.166	0.40%	-178.1649934	0.00%
305	45	-178.055	-177.424	0.36%	-177.4231244	0.00%
405	1904	-176.717	-164.16	7.65%	-170.8388657	3.91%
505	1929	-177.754	-162.103	9.65%	-172.1984897	5.86%

Table 3.4. Result on different number of integer variables

For same 100 scenarios problem, we also tested its sensitivity on different level of integrality, with result shown in 3.4. From this table we can find as number of integer grows, the complexity of the problem grows and simple cut and branch rule may not always be as impactive as it is for smaller problem which is a common complexity for all integer related problems. But within certain range, current algorithm can be very efficient.

For two-stage model, these results show VS-APM with integer programing techniques like cutting plane and branch and bound can also provide good enough solution for problems with integer variables in both stages and have the potential to be utilized in more general problems.

Chapter 4
A distributed framework for economic dispatch problems with AC power flow constraints

4.1 Introduction

In the prior chapters, we focused primarily on problems motivated by optimal flow under DC power flow approximation, which leads to convex programming problems. Such an approximation is widely used in power systems modeling but reliant on the following four approximations:

- All voltage magnitudes are close to one per unit (p.u.).
- Conductances are negligible relative to susceptances.
- Voltage angle difference are small enough in magnitude that they occupy the nearly linear region of the sine function (i.e. $\sin(x) \approx x$).
- Reactive power flows are negligible relative to real power flows.

However, these assumptions may not always hold in practice and this necessitates extending our DC power flow model to the case of AC power flow constraints. However, the resulting ACOPF problem can be highly nonconvex. We consider the development of distributed schemes in this chapter with a view towards addressing large-scale instances in a scalable fashion as well as from the standpoint of developing networked schemes. In fact, networked schemes respect privacy requirements.

The remainder of this chapter is organized as follows. In Section 4.2, we present a literature review on optimal power flow problems with AC power flow constraints. A model of optimal power flow problems with AC power flow constraints is introduced in Section 4.3. An alternating direction method of multipliers (ADMM) scheme is presented in Section 4.4 as well as an accelerated variant. We conclude the chapter with a set of numerics in Section 4.5.

4.2 Literature review

4.2.1 AC power flow problems

The optimal power flow problem for managing real and reactive power dispatch (referred to as ACOPF) to promote reliable operation was first introduced in [130]. The classical ACOPF formulation used in early research may be cast as a nonlinear program [131]. Such models capture power system behavior to a reasonable level of accuracy. However, experimentation conducted with standard nonlinear programming solvers suggest that there may be convergence issues in certain settings [132]. Recent related work on diverse formulations and schemes can be found in [133–136]. These challenges have motivated the development of convexification techniques where it was shown that a suitable convex relaxation is known to be tight [137].Recently, convexification techniques have been applied to ACOPF [138–143]. For example, in [140, 142, 143], the authors propose a semidefinite programming relaxation approach for which it has been shown by Lavaei and Low [137] that a globally optimal solution for the ACOPF problem (under certain conditions) can be obtained by a suitable semidefinite programs. A shortcoming of these convexifications is that there is no mechanism to recover an ACOPF feasible solution when the sufficient condition is not satisfied. As a generalization of semidefinite relaxation, moment relaxation methods [144, 145] were proposed to obtain tighter lower bounds which however become intractable to compute for large scale networks. There has been some effort to develop second-order conic relaxation (SOCR) formulations for the ACOPF with radial (tree) networks (see e.g. [141, 146, 147]). The SOCR solutions are often inexact but with a finite optimality gap; closing the gap may require stronger bounds (which could guarantee a globally optimal outcome when exact) or a local solution method in order to achieve ACOPF feasibility.

4.2.2 Distributed schemes

When faced with structured optimization problems, there has been a tremendous amount of effort in developing decomposition schemes that have allowed for decomposing the problem into smaller tractable problems. For instance, in [148], a distributed approach for solving a DCOPF problem employed a dual decomposition technique reliant on adding fictitious buses at the interconnections between independently coordinated areas. In [149]. two decompositions for the SDP relaxation were developed from the primal and dual formulations was proposed. For the more general ACOPF model, distributed approaches have been to be capable of solving a range of nonconvex problems. Amongst these, the Alternating Direction Method of Multipliers (ADMM) is a commonly used distributed technique.

ADMM schemes, originating from the seminal work by Eckstein [63, 150], have been recently employed for addressing problems complicated by uncertainty [151–153] and nonconvexity [154, 155]. Past work has considered accelerated stochastic ADMM schemes for stochastic convex optimization [156] as well as a tractable ADMM framework for ℓ_0-regularized problems [157] (i.e. nonconvex). In [158], coupling constraints on the rectangular voltage components are relaxed while dual variable updates are computed locally by each agent. In [159], auxiliary variables that represent the sums and differences of voltage phasors between the terminals of lines that are shared by multiple regions are used for decomposition. An extensive numerical study to a variety of test cases on ADMM can be found in [160].

4.3 Model

In this section, we present both a simple AC power flow model and then extend it to a more general model with transformers.

4.3.1 A simple AC power flow model

In a general AC optimal power flow model, the voltage is specified as complex. Consequently, if v_i denotes voltage at node (bus) i and $\mathcal{I}$ denotes set of buses as well, then the magnitude of the voltage must satisfy the following limits.

$$v_i^{min} \leq |v_i| \leq v_i^{max}, \quad \forall i \in \mathcal{I}.$$

If there is a transmission line between bus i and j, the real and reactive power flows across this line are denoted by $p_{i,j}$ and $q_{i,j}$, respectively. As a result, the magnitude of the power on a transmission lines must satisfy the capacity limits:

$$p_{i,j}^2 + q_{i,j}^2 \leq s_{i,j}^2, \quad \forall i, j \in \mathcal{I},$$

which leads to a convex quadratic constraint for each line. The real and reactive power balance needs to be maintained at each bus, as specified by the next requirement.

$$\sum_j p_{i,j} = g_i - d_i, \qquad \text{and} \qquad \sum_j q_{i,j} = g_i^q - d_i^q, \quad \forall i \in \mathcal{I}.$$

In addition, bound constraints are imposed at each bus as well.

$$g_i^{\mathrm{min}} \leq g_i \leq g_i^{\mathrm{max}}, \qquad \text{and} \qquad g_i^{\mathrm{qmin}} \leq g_i^q \leq g_i^{\mathrm{qmax}}, \quad \forall i \in \mathcal{I}.$$

For each transmission line, the impedance $z_{i,j}$ is complex-valued and is the reciprocal of the admittance $y_{i,j}$. Furthermore, these parameters determine the relationship between power flow and voltage.

$$p_{i,j} + iq_{i,j} = v_i(v_i^* - v_j^*)y_{i,j}^*, \quad \forall i, j \in \mathcal{I},$$

where * denotes the complex conjugate. In fact, these equality constraints represent one of the sources of nonconvexity in the AC optimal power flow problem. When employing a linear cost of real-power generation, the resulting AC power flow problem can be cast as the following nonconvex quadratically constrained optimization problem.

$$
\begin{aligned}
\min_{g,p,q,v} \quad & c_g^T g \\
\text{subject to} \quad & p_{i,j} + iq_{i,j} = v_i(v_i^* - v_j^*)y_{i,j}^*, && \forall i,j \in \mathcal{I} \\
& \sum_j p_{i,j} = g_i - d_i, && \forall i \in \mathcal{I} \\
& \sum_j q_{i,j} = g_i^q - d_i^q, && \forall i \in \mathcal{I} \\
& g_i^{\min} \le g_i \le g_i^{\max}, && \forall i \in \mathcal{I} \\
& g_i^{\text{qmin}} \le g_i^q \le g_i^{\text{qmax}}, && \forall i \in \mathcal{I} \\
& p_{i,j}^2 + q_{i,j}^2 \le s_{i,j}^2, && \forall i,j \in \mathcal{I} \\
& v_i^{\min} \le |v_i| \le v_i^{\max}, && \forall i \in \mathcal{I}.
\end{aligned}
$$

4.3.2 Security-constrained OPF models

We now consider a more general security-constrained ACOPF problem articulated as a two-stage stochastic optimization problem. In the first-stage, a pre-contingency decision is made while second stage represents a re-balancing process under a contingency (post-contingency). The goal is to determining the optimal dispatch and control settings for power generation and grid control equipment in order to minimize the cost of operation, subject to pre- and post-contingency constraints. The underline subproblem of this model is a nonconvex AC-OPF problem, for which we use distributed scheme to dealt with.

Objective The objective c is a sum of generator real power output costs in the base case, and a weighted sum of soft constraint violation penalties in the base case and contingencies and is defined as

$$c = c_0 + \delta c^\sigma + (1-\delta)/|K| \sum_{k \in K} c_k^\delta$$

with c_0 represent base case and c_k denote for contingency k and for either case the cost c_i can be written as:

$$c_i = \sum_{g \in G} c_g + c^\sigma, \text{ where } c_g = \sum_{h \in H_g} c_{gh} t_{gh},$$

$$
\begin{aligned}
p_g &= \sum_{h \in H_g} p_{gh} t_{gh}, \\
t_{gh} &\geq 0, \quad \forall g \in G, \forall h \in H_g, \\
\sum_{h \in H_g} t_{gh} &= 1, \quad \forall g \in G.
\end{aligned}
$$

Note that c_g represents a piecewise linear generation cost and c^σ denotes a violation cost. The piecewise linear generation cost is modeled as a pure linear objective by adding interpolation coefficient variables t_{gh} with $h \in H_g$. Similarly, the violation penalty is also given by a piecewise linear cost function for the violation cost terms where a small penalty is applied to minor violations followed by a more stringent penalty for moderate violations and a severe penalty for remaining violations. Similarly, similarly λ_n represents different levels of penalty cost.

$$
c^\sigma = \sum_{n \in N} \left[\lambda_n^P \sum_{i \in I} \left(\sigma_{in}^{P+} + \sigma_{in}^{P-} \right) + \lambda_n^Q \sum_{i \in I} \left(\sigma_{in}^{Q+} + \sigma_{in}^{Q-} \right) + \lambda_n^S \sum_{e \in E} \sigma_{en}^S + \lambda_n^S \sum_{f \in F} \sigma_{fn}^S \right].
$$

$$
\begin{aligned}
c_k^\sigma = &\sum_{n \in N} \left[\lambda_n^P \sum_{i \in I, k \in K} \left(\sigma_{ikn}^{P+} + \sigma_{ikn}^{P-} \right) \right] \\
&+ \sum_{n \in N} \left[\lambda_n^Q \sum_{i \in I, k \in K} \left(\sigma_{ikn}^{Q+} + \sigma_{ikn}^{Q-} \right) + \lambda_n^S \sum_{e \in E, k \in K} \sigma_{ekn}^S + \lambda_n^S \sum_{f \in F, k \in K} \sigma_{fkn}^S \right], \quad \forall k \in K
\end{aligned}
$$

Note that I, E, and F denote the set of buses, branches and transformers while λ^P, λ^Q are penalties for real and reactive power, λ^S represent penalties for overloads on branches and transformers, while n represents the index for the set of kinks in the piecewise linear formulation. For both pre-contingency state and every contingency k, similar operation constraints as follows apply.

Line flow constraints There are two type of transmission lines in this model: branches and transformers. The flow equations are modeled in polar form with a few differences. Real and reactive power flows into line e at the origin buses are defined by the following equations.

$$
\begin{aligned}
p_e^o &= g_e v_{i_e^o}^2 + \left(-g_e \cos(\theta_{i_e^o} - \theta_{i_e^d}) - b_e \sin(\theta_{i_e^o} - \theta_{i_e^d}) \right) v_{i_e^o} v_{i_e^d}, && \forall e \in E \\
q_e^o &= -\left(b_e + b_e^{CH}/2 \right) v_{i_e^o}^2 + \left(b_e \cos(\theta_{i_e^o} - \theta_{i_e^d}) - g_e \sin(\theta_{i_e^o} - \theta_{i_e^d}) \right) v_{i_e^o} v_{i_e^d}, && \forall e \in E
\end{aligned}
$$

Similarly real and reactive power flows into line e at the destination buses are defined as follows.

$$p_e^d = g_e v_{i_e^d}^2 + \left(-g_e \cos(\theta_{i_e^d} - \theta_{i_e^o}) - b_e \sin(\theta_{i_e^d} - \theta_{i_e^o}))\right) v_{i_e^o} v_{i_e^d}, \qquad \forall e \in E$$
$$q_e^d = -\left(b_e + b_e^{CH}/2\right) v_{i_e^d}^2 + \left(b_e \cos(\theta_{i_e^d} - \theta_{i_e^o}) - g_e \sin(\theta_{i_e^d} - \theta_{i_e^o}))\right) v_{i_e^o} v_{i_e^d}, \quad \forall e \in E,$$

where b_e^{CH} denotes the total charging susceptance for line e. In the case of transformers, the phase angle deviation is different in the flow equations. Specifically, the real and reactive power flows into transformer f at the origin buses are defined by the following.

$$\begin{aligned} p_f^o = {} & \left(g_f/\tau_f^2 + g_f^M\right) v_{i_f^o}^2 \\ & + \left(-g_f/\tau_f \cos(\theta_{i_f^o} - \theta_{i_f^d} - \theta_f) - b_f/\tau_f \sin(\theta_{i_f^o} - \theta_{i_f^d} - \theta_f))\right) v_{i_f^o} v_{i_f^d}, \quad && \forall e \in E \\ q_f^o = {} & -\left(b_f/\tau_f^2 + b_f^M\right) v_{i_f^o}^2 \\ & + \left(b_f/\tau_f \cos(\theta_{i_f^o} - \theta_{i_f^d} - \theta_f) - g_f/\tau_f \sin(\theta_{i_f^o} - \theta_{i_f^d}) - \theta_f)\right) v_{i_f^o} v_{i_f^d}, \quad && \forall e \in E. \end{aligned}$$

Similarly, real and reactive power flows into transformer f line at the destination buses are defined as follows.

$$\begin{aligned} p_f^d = {} & g_f v_{i_f^d}^2 \\ & + \left(-g_f/\tau_f \cos(\theta_{i_f^d} - \theta_{i_f^o} + \theta_f) - b_f/\tau_f \sin(\theta_{i_f^d} - \theta_{i_f^o} + \theta_f))\right) v_{i_f^o} v_{i_f^d}, \quad && \forall f \in F \\ q_f^d = {} & -b_f v_{i_f^d}^2 \\ & + \left(b_f/\tau_f \cos(\theta_{i_f^d} - \theta_{i_f^o} + \theta_f) - g_f/\tau_f \sin(\theta_{i_f^d} - \theta_{i_f^o} + \theta_f))\right) v_{i_f^o} v_{i_f^d}, \quad && \forall f \in F, \end{aligned}$$

where τ_f is tap ratio of transformer f, g_f^M represents the magnetizing conductance of transformer f, and θ_f denotes the phase angle of transformer f.

Bus balance constraints The bus balance constraints differ because of the introduction of shunt susceptance. Consequently, the bus real power balance equation are defined as follows.

$$\sum_{g \in G_i} p_g - p_i^L - g_i^{FS} v_i^2$$

$$- \sum_{e \in E_i^o} p_e^o - \sum_{e \in E_i^d} p_e^d - \sum_{f \in F_i^o} p_f^o - \sum_{f \in F_i^o} p_f^o = \sigma^{P+} - \sigma^{P-}, \forall i \in I$$

Similarly, the reactive power balance at bus i is defined as follows.

$$\begin{aligned} &\sum_{g \in G_i} q_g - q_i^L - (-b_i^{FS} - b_i^{CS}) v_i^2 \\ &\qquad - \sum_{e \in E_i^o} q_e^o - \sum_{e \in E_i^d} q_e^d - \sum_{f \in F_i^o} q_f^o - \sum_{f \in F_i^o} q_f^o = \sigma^{Q+} - \sigma^{Q-}, \forall i \in I \end{aligned}$$

where g_i^{FS} is fixed shunt conductance of bus i, b_i^{FS} is fixed shunt susceptance of bus i and b_i^{CS} is the controllable shunt susceptance of bus i.

4.4 A Distributed ADMM framework

In this section, we provide a brief introduction to ADMM schemes and then describe how one may apply such a scheme to the security-constrained OPF problem.

4.4.1 An introduction to ADMM schemes

Consider a structured optimization problem of the form:

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y) \\ \text{subject to} \quad & Ax + By = c \\ & x \in \mathcal{X}, \quad y \in \mathcal{Y}. \end{aligned}$$

One may then define the augmented Lagrangian function as follows.

$$\mathcal{L}_\rho(x, y, \lambda) = f(x) + g(y) + \lambda^T (Ax + By - c) + \frac{\rho}{2} \|Ax + By - c\|_2^2,$$

where λ denotes the Lagrange multiplier associated with $Ax + By = c$ and ρ represents a fixed penalty parameter. The ADMM update rule requires minimizing the augmented Lagrangian function in a step-wise fashion, given an initial iterate (x_0, y_0, λ_0).

$$x^{k+1} \quad := \quad \arg\min_{x \in \mathcal{X}} \mathcal{L}_\rho(x, y^k, \lambda^k) \tag{4.1}$$

$$
\begin{aligned}
y^{k+1} &:= \arg\min_{y\in\mathcal{Y}} \ \mathcal{L}_\rho(x^{k+1}, y, \lambda^k) && (4.2)\\
\lambda^{k+1} &:= \lambda^k + \rho(Ax^{k+1} + By^{k+1} - c), && (4.3)
\end{aligned}
$$

where x and y denote two sets of variables with separable objective. By doing so, the problem is split into problem with x and problem with y. In a power network, there are multiple choices of which constraints may be relaxed. We relax the flow balance constraints from which we may easily decompose the entire problem network into components such as the generators, branches, and transformers. The detailed framework is introduced in the next section.

4.4.2 A distributed ADMM framework for SC-OPF

The decision variables fall into 3 category: generator related, branch related and transformer related. Thus, we have original decision variables x represented as:

$$x \triangleq [x_G, x_E, x_F],$$

where

$$
\begin{aligned}
x_G &\triangleq (t_{gh}, p_g, q_g, c_g)_G,\\
x_E &\triangleq (p_{eo}, p_{ed}, q_{eo}, q_{ed}, \sigma_{en}, v_{eo}, v_{ed}, \theta_{eo}, \theta_{ed})_E,\\
x_F &\triangleq (p_{fo}, p_{fd}, q_{fo}, q_{fd}, \sigma_{fn}, v_{fo}, v_{fd}, \theta_{fo}, \theta_{fd})_F.
\end{aligned}
$$

Similarly, we may also generate partition the added variable z into three similar classes as well as an additional class pertaining to buses.

$$z \triangleq [z_G, z_E, z_F, z_I],$$

where

$$
\begin{aligned}
z_G &\triangleq (p_{g(i)}, q_{g(i)})_G,\\
z_E &\triangleq (p_{eo(i)}, p_{ed(i)}, q_{eo(i)}, q_{ed(i)})_E,\\
z_F &\triangleq (p_{fo(i)}, p_{fd(i)}, q_{fo(i)}, q_{fd(i)})_F,\\
z_I &\triangleq (v, \theta, b_{cs}, \sigma_{p+}, \sigma_{p-}, \sigma_{q+}, \sigma_{q-})_I,
\end{aligned}
$$

where subscript i represents a copy of the original, i.e. $p_g(i)$ is a copy of p_g. Thus, the dimension of variables are as follow:

$$
\begin{aligned}
N_x &\triangleq (3+|H|)|G| + (8+|N|)|E| + (8+|N|)|F| \\
N_z &\triangleq 2|G| + 4|E| + 4|F| + 5|I|,
\end{aligned}
$$

where $|H|$ denotes number of different levels of piecewise generation cost, $|N|$ denotes number of different levels of over load penalty on branches and transformers. $|G|, |E|, |F|, |I|$ represents number of generators, branches, transformers and buses. Finally the coupling constraints are given by

$$Ax - Bz = 0.$$

Consequently, the augmented Lagrange function is

$$\mathcal{L}_\rho(x, z, \lambda) = f(x) + g(z) + \lambda^T(Ax - Bz) + \frac{\rho}{2}\|Ax - Bz\|^2,$$

where $\rho > 0$ and

$$
\begin{aligned}
\lambda \triangleq [&(\lambda_{pg}, \lambda_{qg})_G, (\lambda_{peo}, \lambda_{ped}, \lambda_{qeo}, \lambda_{qed}, \lambda_{veo}, \lambda_{ved}, \lambda_{\theta eo}, \lambda_{\theta ed})_E, \\
&\qquad (\lambda_{pfo}, \lambda_{pfd}, \lambda_{qfo}, \lambda_{qfd}, \lambda_{vfo}, \lambda_{vfd}, \lambda_{\theta fo}, \lambda_{\theta fd})_F]
\end{aligned}
$$

denotes the vector of dual variable associated with coupling constraints with dimension $N_\lambda = 2|G| + 8|E| + 8|F|$.

4.4.3 Algorithm

Recall ADMM generates new iterates (x^k, z^k, λ^k) as follows:

$$x^{k+1} := \arg\min_x L_\rho(x, z^k, \lambda^k) \tag{4.4}$$

$$z^{k+1} := \arg\min_z L_\rho(x^{k+1}, z, \lambda^k) \tag{4.5}$$

$$\lambda^{k+1} := \lambda^k + \rho(Ax^{k+1} - Bz^{k+1}) \tag{4.6}$$

Specifically, for generator g, we have the objective for generator subproblem as follows:

$$f_g = \sum_{g \in G} \left\{ f_g(x_g) + \langle \lambda_g^k, x'_g \rangle + \frac{\rho}{2} \left[\left(p_g - p_{g(i)}^k \right)^2 + \left(q_g - q_{g(i)}^k \right)^2 \right] \right\}$$

With generators subproblem as follow:

$$\begin{aligned}
\min_{x_g} \quad & f_g \\
\text{subject to} \quad & \underline{p}_g \leq p_g \leq \bar{p}_g \quad \forall g \in \mathcal{G}, \\
& \underline{q}_g \leq q_g \leq \bar{q}_g \quad \forall g \in \mathcal{G}, \\
& c_g = \sum_{\forall h \in \mathcal{H}} c_{gh} t_{gh} \quad \forall g \in \mathcal{G}, \\
& \sum_{\forall h \in \mathcal{H}} p_{gh} t_{gh} = p_g \quad \forall g \in \mathcal{G}, \\
& 0 \leq t_{gh} \quad \forall g \in \mathcal{G}, \forall h \in \mathcal{H} \\
& \sum_{\forall h \in \mathcal{H}} t_{gh} = 1 \quad \forall g \in \mathcal{G}.
\end{aligned} \tag{Generator subproblem}$$

where

$$x_g := [t_{gh}, p_g, q_g, c_g]$$

are variables for generator g,

$$x'_g := [p_g, q_g]$$

are parts of x_g in coupling constraints and

$$\lambda_g := [\lambda_{pg}, \lambda_{qg}]$$

are parts of λ corresponds to coupling constraints with generators.

For branch line e from bus i to j, we have branch subproblem as follows:

$$\begin{aligned}
f_e := & \sum_{e \in E} f_e(\sigma_{en}) + \langle \lambda_e^k, x'_e \rangle \\
& + \frac{\rho}{2} \left(p_{eo} - p_{eo(i)}^k \right)^2 + \frac{\rho}{2} \left(p_{ed} - p_{ed(i)}^k \right)^2 + \frac{\rho}{2} \left(q_{eo} - q_{eo(i)}^k \right)^2 + \frac{\rho}{2} \left(q_{ed} - q_{ed(i)}^k \right)^2 \\
& + \frac{\rho}{2} \left(v_{eo} - v_i^k \right)^2 + \frac{\rho}{2} \left(\theta_{eo} - \theta_i^k \right)^2 + \frac{\rho}{2} \left(v_{ed} - v_j^k \right)^2 + \frac{\rho}{2} \left(\theta_{ed} - \theta_j^k \right)^2
\end{aligned}$$

The branch subproblem is as follows:

$$
\begin{aligned}
\min_{x_e} \quad & f_e, \\
\text{subject to} \quad & \underline{v}_{ed} \le v_{ed} \le \bar{v}_{ed} \quad \forall e \in \mathcal{E}, \\
& p_{eo} = g_e v_{eo}^2 + (-g_e \cos(\theta_{eo} - \theta_{ed}) - b_e \sin(\theta_{eo} - \theta_{ed})) v_{eo} v_{ed} \\
& q_{eo} = -(b_e + b_e^{CH}/2) v_{eo}^2 + (b_e \cos(\theta_{eo} - \theta_{ed}) - g_e \sin(\theta_{eo} - \theta_{ed})) v_{eo} v_{ed} \\
& p_{ed} = g_e v_{ed}^2 + (-g_e \cos(\theta_{ed} - \theta_{eo}) - b_e \sin(\theta_{ed} - \theta_{eo})) v_{eo} v_{ed} \\
& q_{ed} = -(b_e + b_e^{CH}/2) v_{ed}^2 + (b_e \cos(\theta_{ed} - \theta_{eo}) - g_e \sin(\theta_{ed} - \theta_{eo})) v_{eo} v_{ed} \\
& \sqrt{p_{eo}^2 + q_{eo}^2} \le \overline{R}_e v_{eo} + \sigma_e \quad \forall e \in \mathcal{E} \\
& \sqrt{p_{ed}^2 + q_{ed}^2} \le \overline{R}_e v_{ed} + \sigma_e \quad \forall e \in \mathcal{E} \\
& \sigma_e = \sum_{\forall n \in \mathcal{N}} \sigma_{en} \quad \forall e \in \mathcal{E}, \\
& 0 \le \sigma_{en} \le \overline{\sigma}_{en} \quad \forall e \in \mathcal{E}, \forall n \in \mathcal{N}. \qquad \text{(Branch subproblem)}
\end{aligned}
$$

where

$$x_e := [p_{eo}, p_{ed}, q_{eo}, q_{ed}, \sigma_{en}, v_{eo}, v_{ed}, \theta_{eo}, \theta_{ed}]$$

are variables for branch e,

$$x'_e := [p_{eo}, p_{ed}, q_{eo}, q_{ed}, v_{eo}, v_{ed}, \theta_{eo}, \theta_{ed}]$$

are parts of x_e in coupling constraints and

$$\lambda_e := [\lambda_{peo}, \lambda_{ped}, \lambda_{qeo}, \lambda_{qed}, \lambda_{veo}, \lambda_{ved}, \lambda_{\theta eo}, \lambda_{\theta ed}]$$

are parts of λ corresponds to coupling constraints with branch.

Similarly, for transformer f from bus i to j, we have the subproblem objective:

$$
\begin{aligned}
f_f := & \sum_{f \in E} f_f(\sigma_{fn}) + \langle \lambda_f^k, x'_f \rangle \\
& + \frac{\rho}{2} \left(p_{fo} - p_{fo(i)}^k \right)^2 + \frac{\rho}{2} \left(p_{fd} - p_{fd(i)}^k \right)^2 + \frac{\rho}{2} \left(q_{fo} - q_{fo(i)}^k \right)^2 + \frac{\rho}{2} \left(q_{fd} - q_{fd(i)}^k \right)^2 \\
& + \frac{\rho}{2} \left(v_{fo} - v_i^k \right)^2 + \frac{\rho}{2} \left(\theta_{fo} - \theta_i^k \right)^2 + \frac{\rho}{2} \left(v_{fd} - v_j^k \right)^2 + \frac{\rho}{2} \left(\theta_{fd} - \theta_j^k \right)^2
\end{aligned}
$$

The corresponding subproblem

$$\begin{aligned}
\min_{x_f} \quad & f_f, \\
\text{subject to} \quad & \underline{v}_{fd} \leq v_{fd} \leq \bar{v}_{fd} \quad \forall f \in \mathcal{F}, \\
& p_{fo} = (g_f/\tau_f^2 + g_f^M) v_{fo}^2 \\
& \qquad + (-g_f/\tau_f \cos(\theta_{fo} - \theta_{fd} - \theta_f) - b_f/\tau_f \sin(\theta_{fo} - \theta_{fd} - \theta_f)) v_{fo} v_{fd} \\
& q_{fo} = -(b_f/\tau_f^2 + b_f^M) v_{fo}^2 \\
& \qquad + (b_f/\tau_f \cos(\theta_{fo} - \theta_{fd} - \theta_f) - g_f/\tau_f \sin(\theta_{fo} - \theta_{fd} - \theta_f)) v_{fo} v_{fd} \\
& p_{fd} = g_f v_{fd}^2 \\
& \qquad + (-g_f/\tau_f \cos(\theta_{fd} - \theta_{fo} + \theta_f) - b_f \sin(\theta_{fd} - \theta_{fo} + \theta_f)) v_{fo} v_{fd} \\
& q_{fd} = -b_f v_{fd}^2 \\
& \qquad + (b_f/\tau_f \cos(\theta_{fd} - \theta_{fo} + \theta_f) - g_f \sin(\theta_{fd} - \theta_{fo} + \theta_f)) v_{fo} v_{fd} \\
& \sqrt{p_{fo}^2 + q_{fo}^2} \leq \overline{s}_f + \sigma_f \quad \forall f \in \mathcal{F} \\
& \sqrt{p_{fd}^2 + q_{fd}^2} \leq \overline{s}_f + \sigma_f \quad \forall f \in \mathcal{F} \\
& \sigma_f = \sum_{\forall n \in \mathcal{N}} \sigma_{fn} \quad \forall f \in \mathcal{F}, \\
& 0 \leq \sigma_{fn} \leq \overline{\sigma}_{fn} \quad \forall f \in \mathcal{F}, \forall n \in \mathcal{N}. \qquad \text{(Transformer subproblem)}
\end{aligned}$$

where

$$x_f := [p_{fo}, p_{fd}, q_{fo}, q_{fd}, \sigma_{fn}, v_{fo}, v_{fd}, \theta_{fo}, \theta_{fd}]$$

are variables for transformer f,

$$x'_f := [p_{fo}, p_{fd}, q_{fo}, q_{fd}, v_{fo}, v_{fd}, \theta_{fo}, \theta_{fd}]$$

are parts of x_f in coupling constraints and

$$\lambda_f := [\lambda_{pfo}, \lambda_{pfd}, \lambda_{qfo}, \lambda_{qfd}, \lambda_{vfo}, \lambda_{vfd}, \lambda_{\theta fo}, \lambda_{\theta fd}]$$

are parts of λ corresponds to coupling constraints with transformer.

Thu bus subproblem is about bus flow balance with the following objective:

$$f_b := \sum_{g \in G} f'_g + \sum_{e \in E} f'_e + \sum_{f \in F} f'_f + \sum_{i \in I} f_i(\sigma_{p+}, \sigma_{p-}, \sigma_{q+}, \sigma_{q-})$$

where

$$
\begin{aligned}
f'_g &= -\langle \lambda_g^k, z_g \rangle + \frac{\rho}{2}\left(\left(p_g^{k+1} - p_{g(i)}\right)^2 + \left(q_g^{k+1} - q_{g(i)}\right)^2\right) \\
f'_e &= -\langle \lambda_e^k, z_e \rangle + \frac{\rho}{2}\left(p_{eo}^{k+1} - p_{eo(i)}\right)^2 + \frac{\rho}{2}\left(p_{ed}^{k+1} - p_{ed(i)}\right)^2 \\
&\quad + \frac{\rho}{2}\left(q_{eo}^{k+1} - q_{eo(i)}\right)^2 + \frac{\rho}{2}\left(q_{ed}^{k+1} - q_{ed(i)}\right)^2 + \frac{\rho}{2}\left(v_{eo}^{k+1} - v_i\right)^2 + \frac{\rho}{2}\left(\theta_{eo}^{k+1} - \theta_i\right)^2 \\
&\quad + \frac{\rho}{2}\left(v_{ed}^{k+1} - v_j\right)^2 + \frac{\rho}{2}\left(\theta_{ed}^{k+1} - \theta_j\right)^2 \\
f'_f &= -\langle \lambda_f^k, z_f \rangle + \frac{\rho}{2}\left(p_{fo}^{k+1} - p_{fo(i)}\right)^2 + \frac{\rho}{2}\left(p_{fd}^{k+1} - p_{fd(i)}\right)^2 \\
&\quad + \frac{\rho}{2}\left(q_{fo}^{k+1} - q_{fo(i)}\right)^2 + \frac{\rho}{2}\left(q_{fd}^{k+1} - q_{fd(i)}\right)^2 + \frac{\rho}{2}\left(v_{fo}^{k+1} - v_i\right)^2 + \frac{\rho}{2}\left(\theta_{fo}^{k+1} - \theta_i\right)^2 \\
&\quad + \frac{\rho}{2}\left(v_{fd}^{k+1} - v_j\right)^2 + \frac{\rho}{2}\left(\theta_{fd}^{k+1} - \theta_j\right)^2
\end{aligned}
$$

The subproblem is as follows:

$$
\begin{aligned}
& z^{k+1} := \arg\min_z f_b \\
\text{s.t.} \quad & \sum_{g \in G} p_{g(i)} - p_i^L - g_i^{FS} v_i^2 \\
& - \sum_{e \in E_{io}} p_{eo(i)} - \sum_{e \in E_{id}} p_{ed(i)} - \sum_{f \in F_{io}} p_{fo(i)} - \sum_{f \in F_{id}} p_{fd(i)} = \sigma_{p+} - \sigma_{p-} \quad \forall i \in \mathcal{I} \\
& \sum_{g \in G} q_{g(i)} - q_i^L - (-b_i^{FS} - b_i^{CS}) v_i^2 \\
& - \sum_{e \in E_{io}} q_{eo(i)} - \sum_{e \in E_{id}} q_{ed(i)} - \sum_{f \in F_{io}} q_{fo(i)} - \sum_{f \in F_{id}} q_{fd(i)} = \sigma_{q+} - \sigma_{q-} \quad \forall i \in \mathcal{I} \\
& \sigma_{p+} = \sum_{\forall n \in \mathcal{N}} \sigma_{pn+} \quad \forall i \in \mathcal{I}, \\
& 0 \le \sigma_{pn+} \le \overline{\sigma}_{pn+} \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N}, \\
& \sigma_{p-} = \sum_{\forall n \in \mathcal{N}} \sigma_{pn-} \quad \forall i \in \mathcal{I}, \\
& 0 \le \sigma_{pn-} \le \overline{\sigma}_{pn-} \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N}, \\
& \sigma_{q+} = \sum_{\forall n \in \mathcal{N}} \sigma_{qn+} \quad \forall i \in \mathcal{I}, \\
& 0 \le \sigma_{qn+} \le \overline{\sigma}_{qn+} \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N},
\end{aligned}
$$

$$\sigma_{q-} = \sum_{\forall n \in \mathcal{N}} \sigma_{qn-} \quad \forall i \in \mathcal{I},$$

$$0 \leq \sigma_{qn-} \leq \overline{\sigma}_{qn-} \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N}. \tag{Bus subproblem}$$

where

$$z_g := [p_{g(i)}, q_{g(i)}]$$

$$z_e := [p_{eo(i)}, p_{ed(i)}, q_{eo(i)}, q_{ed(i)}, v_{eo}, v_{ed}, \theta_{eo}, \theta_{ed}]$$

$$z_f := [p_{fo(i)}, p_{fd(i)}, q_{fo(i)}, q_{fd(i)}, v_{fo}, v_{fd}, \theta_{fo}, \theta_{fd}]$$

Finally, after solving all subproblems we can update Lagrangian multiplier λ following (4.6). A common stopping criterion is the feasibility of coupling constraints which is also the updating step of λ. As updating process of λ stops, we get a feasible solution in terms of coupling condition. Since all variables are solved in subproblem solver, the resulting solution is a KKT conditions satisfied solution.

4.4.4 Stochastic ADMM

The overall security constraints model can be represented in a general form as follows:

$$\begin{aligned}
\min \quad & f(x) + \sum_{i \in K} f_i(y_i) \\
\text{subject to} \quad & x \in X, \\
& y_i \in Y_i, && \forall i \in K \\
& Ax + B_i y_i = b_i, && \forall i \in K
\end{aligned}$$

with X and Y_i represents pre- and post- contingency feasible region as a set of nonconvex constraints as Section 4.3. The resulting problem is a natural form for a stochastic ADMM framework. As the augmented Lagrangian function associated is defined as

$$\mathcal{L}(x_k, y_k, \lambda_k, \rho_k, N_k) \triangleq f(x_k) + \frac{1}{N_k} \sum_{j=1}^{N_k} \mathcal{L}_j(x_k, y_{j,k}, \lambda_{j,k}, \rho_k)$$

where

$$\mathcal{L}_j(x_k, y_{j,k}, \lambda_{j,k}, \rho_k) \triangleq$$
$$\Big(g_j(y_{j,k}) - \lambda_j^T(A_j x_k + B_j y_{j,k} - b_j) + \rho_k^2 \|A_j x_k + B_j y_{j,k} - b_j\|^2\Big)$$

The stochastic ADMM framework can be stated as follows for sequences $\{\rho_k, N_k\}$:

$$x_{k+1} \in \arg\min_{x \in X} \mathcal{L}(x, y_k, \lambda_k, \rho_k, N_k) \tag{x-update}$$

$$y_{j,k+1} \in \begin{cases} \arg\min_{y_j \in Y_j} \mathcal{L}_j(x_k, y_j, \lambda_{j,k}, \rho_k), & j \in \mathcal{N}_k \subseteq \mathcal{N} \\ \{y_{j,k}\}, & j \in \mathcal{N} \backslash \mathcal{N}_k \end{cases} \tag{y-update}$$

$$\lambda_{j,k+1} := \lambda_{j,k} + \rho_k(A_j x_k + B_j y_{j,k} - b_j), \qquad j \in \mathcal{N}. \tag{λ-update}$$

We develop a stochastic ADMM scheme in which the possibility of large N is addressed by utilizing an increasing sequence $\{N_k\}$ of contingency scenarios (sampled without replacement). Consequently, the x decision relies on an average over N_k terms while N_k of the N scenario-based y problems are solved while recourse decisions y_j corresponding to unsampled scenarios are kept invariant.

4.5 Numerics

4.5.1 Distributed ADMM

We reformulated ARPA-E model into distributed ADMM (DADMM) computable model. Numerical test was conducted on ARPA-E "Original Dataset Real-Time Network_01-10R". This network contains 500 buses, with 90 generators, 468 branches, 262 transformers. We applied DADMM algorithm with component subproblem solved in Ipopt while a centralized problem also solved in Ipopt for comparison As we can find from this result, since DADMM relaxed coupling constraints, it could yield better solution in terms of original objective. The infeasiblity represents the violation of coupling constraints. DADMM is steadily reducing infeasiblity along iterations and stopped at a level of 2e-2.

We also compared behaviors of different penalty parameters. As we can observe from this plot, the behavior of ADMM is very sensitive to the selection of penalty parameters.

Case	Ipopt cost	DADMM_cost	DADMM_inf	DADMM_rho	time
1	27525.81236	15573.86761	0.021402498	2.273736754	4030.185
2	39355.71734	23139.50358	0.018151319	2.273736754	3927.135
3	28303.43332	15606.93508	0.021434079	2.273736754	4426.232
4	32027.2072	16742.80309	0.021661662	1.818989404	4237.188
5	33982.43563	23231.74579	0.026799315	2.273736754	4417.986
6	34190.90127	22033.42446	0.019681965	2.273736754	2249.481
7	40963.65905	22656.06014	0.020325284	2.273736754	3700.074
8	37080.89701	21875.25697	0.031438247	2.273736754	3855.665
9	51854.69583	28675.80271	0.024224493	2.273736754	3796.785
10	50987.64473	25070.04499	0.024336965	2.273736754	3862.696

Table 4.1. Result of DADMM on ARPA-E network

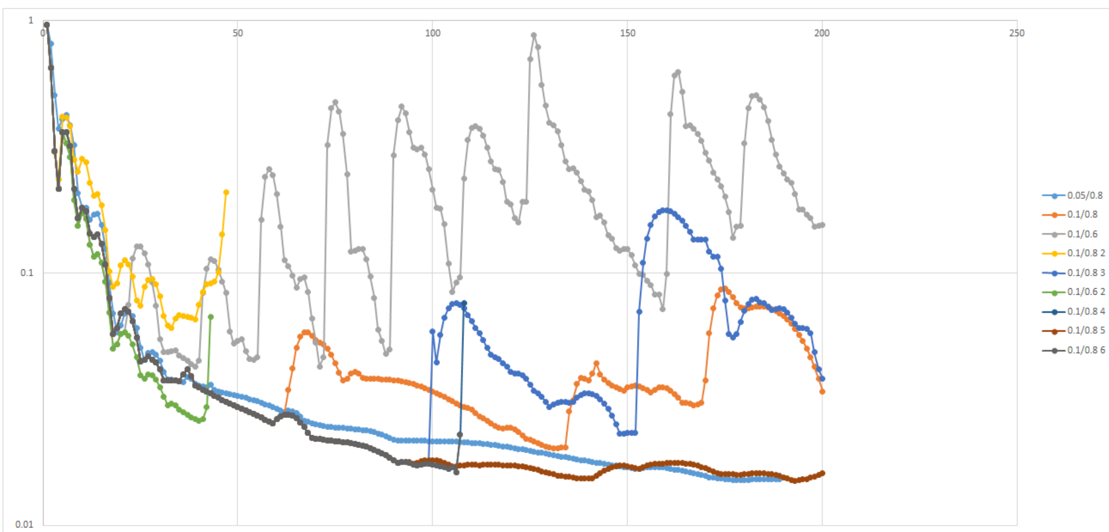

Figure 4.1. Different penalty value comparison

4.5.2 Stochastic ADMM

To examine the compatibility of ADMM to two stage stochastic programs, we also conducted numerical test of stochastic ADMM on two-stage stochastic AC-OPF model with randomly generated real power demand. Test network is IEEE 9 bus system with two-stage AC-OPF problem. The test is conducted under MATLAB R2016a with MATLAB Optimization Toolbox as nonlinear subproblem solver and BARON as deterministic equivalent solver.

Feasibility residual represents the feasibility of coupling constraint between stages, while KKT residual represents the optimality of subproblems. As we can

K	iter	time_admm	Feasibility	KKT_residual	obj_admm	obj_det	diff
50	34	1.11E+03	9.49E-05	8.53E-04	2.44E+04	2.44E+04	0.12%
70	35	1.62E+03	7.85E-05	8.39E-04	2.45E+04	2.45E+04	0.21%
90	34	2.04E+03	9.30E-05	7.04E-04	2.46E+04	2.45E+04	0.28%
100	34	1.45E+03	9.09E-05	2.93E-03	2.47E+04	2.47E+04	0.04%
110	35	2.55E+03	8.67E-05	6.30E-04	2.47E+04	2.47E+04	0.11%
130	37	3.17E+03	8.73E-05	3.10E-04	2.49E+04	2.48E+04	0.41%
150	36	3.62E+03	6.44E-05	6.28E-04	2.49E+04	2.48E+04	0.22%
170	42	4.63E+03	9.85E-05	1.08E-03	2.50E+04	2.49E+04	0.02%
190	38	4.71E+03	9.76E-05	6.63E-04	2.48E+04	2.48E+04	0.18%
200	41	3.88E+03	6.85E-05	5.11E-04	2.47E+04	2.47E+04	0.00%
210	39	5.21E+03	9.46E-05	7.34E-04	2.47E+04	2.46E+04	0.13%
230	55	7.79E+03	9.78E-05	1.36E-02	2.47E+04	2.47E+04	0.00%
500	500	9.98E+04	1.80E-04	1.31E-03	2.46E+04	2.46E+04	0.07%

Table 4.2. Result of ADMM on stochastic ACOPF

observe from Table 4.2, with relatively low feasibility and KKT residual, ADMM could yield reasonable result comparing with centralized result. And from Figure 4.2 we can find ADMM computational time grows near linearly as number of scenarios grows, which also presents its potential to be applied to large scale problems.

When taking updating step, similar acceleration technique as chapter 2 can be applied to updating process. Instead of using the endpoint, we can take a extrapolation of the sequence to accelerate the convergence process. The acceleration of ADMM can be found in [161] in which Nesterov's smoothing technique [82] is applied to ADMM with rate of convergence $\mathcal{O}(1/N)$. The comparison of accelerated ADMM and standard ADMM is shown in Figure 4.3.

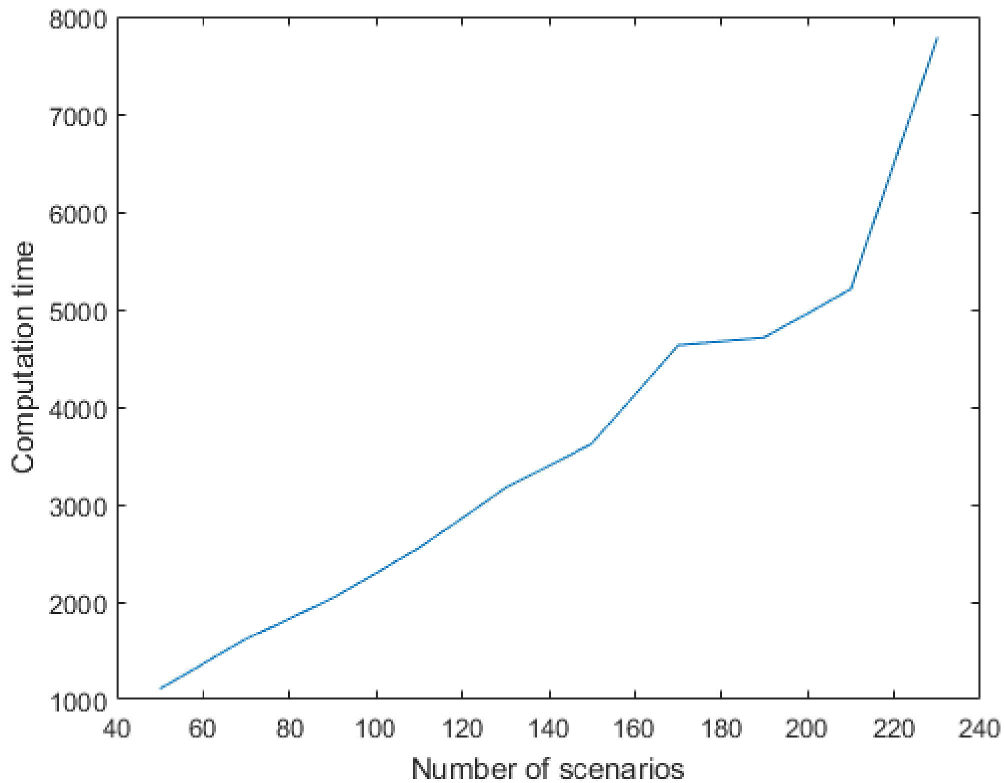

Figure 4.2. ADMM computation time

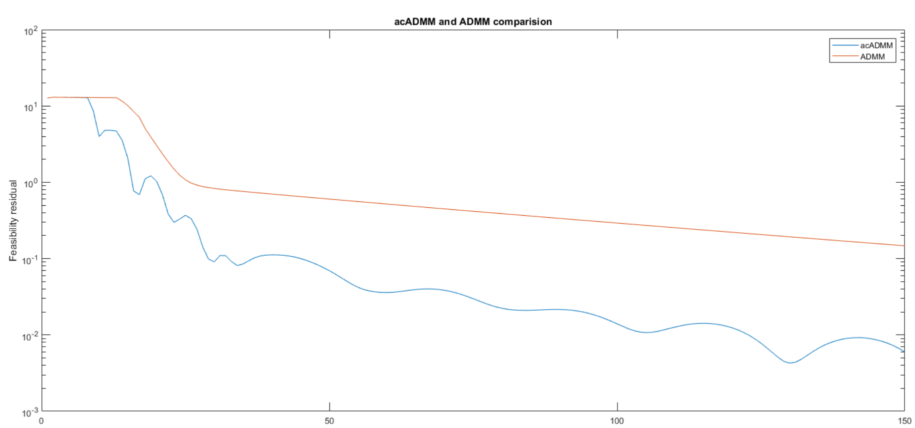

Figure 4.3. Accelerate ADMM

Chapter 5 Concluding remarks and future work

In this section, we provide some concluding remarks regarding each chapter and discuss some furture work in this regard.

5.1 Risk-based economic dispatch

In Chapter 2, we consider the problem of two-stage risk-based economic dispatch when employing a conditional value-at-risk (CVaR) measure. In this setting, we develop a stochastic approximation scheme reliant on smoothing, acceleration, and variance reduction (by utilizing an increasing batch-size of gradients). The resulting sequence of iterates. As a consequence, the expected sub-optimality diminishes at a rate of $\mathcal{O}(1/k)$ while the oracle complexity to compute an ϵ-optimal solution is shown to be $\mathcal{O}(1/\epsilon^2)$. We compare this scheme with comparable cutting-plane schemes and observe that the presented scheme has significant computational benefits on a class of IEEE test systems.

Current risk-based economic dispatch is modeled under DC-OPF. Future works consider more general optimal flow models. Among a variety of convex relaxation of OPF model, Semidefinite programming (SDP) [162] or a secondorder cone programming (SOCP) [163] relaxation are two major convex approximations that are promising in solving large scale AC-OPF models. The resulting two-stage convex programs are nature extension of VS-APM. These relaxation often provide significantly more accurate solutions than previous linear approximations, which could lead to more general practical use.

5.2 Two-stage stochastic integer programming via Stochastic Approximation

In Chapter 3, we discuss an extension of VS-APM to integer programs. With introduction of integer programming techniques like branch and bound, VS-APM can be applied to more general stochastic programs which lead to broader area of piratical problems such as unit commitment problems in power system operation. With more cut and branching rules, this framework is shown to provide comparable solution with state of art integer solver with potential to be applied to large scale problems.

Current two-stage stochastic integer framework utilizes limited integer programs techniques. Future works consider introducing more sophisticated cutting plane procedure like introducing Gomory cut [126, 127] associated with Simplex basis, and branch and cut process [105] with more suitable cuts for each node to reduces the time spend on searching. Numerical study for more general practical problems like unit commitment problem can be conducted along with other uncertainties in power system operation introduced in previous chapter.

5.3 A distributed framework for economic dispatch problems with AC power flow constraints

In Chapter 4, we consider dealing with large scale nonconvex power flow model in a distributed fashion. By reformulating AC-OPF model into component based decomposable model, a component based distributed ADMM scheme is applied to ARPA-E power system model. For security constrained optimal power flow problem, a stochastic ADMM is developed for this two-stage nonconvex program. Numerical results show that ADMM provide possibility for solving large-scale nonconvex stochastic AC-OPF.

Present study demonstrate the relation between performance of ADMM algorithm with corresponding penalty parameters. Future study consider further investigation of penalty parameter ρ and its updating rule. The acceleration of ADMM from [161] in which Nesterov's smoothing technique [82] can be also applied to distributed ADMM. For higher precision required problems, further study on coordinating subproblem agents could lead to a larger variety of practical application of this framework.

Bibliography

[1] Ruszczynski, A. (2003) "Decomposition Methods," in **Handbook in Operations Research and Management Science**, vol. 10, Elsevier Science, Amsterdam, pp. 141–212.

[2] Benders, J. F. (1962) "Partitioning procedures for solving mixed-variables programming problems," **Numer. Math.**, **4**, pp. 238–252.

[3] Van Slyke, R. M. and R. Wets (1969) "L-shaped linear programs with applications to optimal control and stochastic programming," **SIAM J. Appl. Math.**, **17**, pp. 638–663.

[4] Birge, J. R. and F. Louveaux (1997) **Introduction to Stochastic Programming: Springer Series in Operations Research**, Springer.

[5] Higle, J. L. and S. Sen (1991) "Stochastic decomposition: An algorithm for two-stage linear programs with recourse," **Mathematics of operations research**, **16**(3), pp. 650–669.

[6] Higle, J. and S. Sen (1996) **Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming**, Kluwer Academic Publishers, Boston, MA.

[7] Shapiro, A. (2003) "Monte Carlo Sampling Methods," in **Handbook in Operations Research and Management Science**, vol. 10, Elsevier Science, Amsterdam, pp. 353–426.

[8] Yong, T. and R. Lasseter (2000) "Stochastic optimal power flow: formulation and solution," in **Power Engineering Society Summer Meeting, 2000. IEEE**, vol. 1, IEEE, pp. 237–242.

[9] Gangammanavar, H., S. Sen, and V. M. Zavala (2016) "Stochastic optimization of sub-hourly economic dispatch with wind energy," **IEEE Transactions on Power Systems**, **31**(2), pp. 949–959.

[10] Doherty, R. and M. O'malley (2005) "A new approach to quantify reserve demand in systems with significant installed wind capacity," **IEEE Transactions on Power Systems**, **20**(2), pp. 587–595.

[11] Morales, J. M., A. J. Conejo, and J. Pérez-Ruiz (2009) "Economic valuation of reserves in power systems with high penetration of wind power," **IEEE Transactions on Power Systems**, **24**(2), pp. 900–910.

[12] Summers, T., J. Warrington, M. Morari, and J. Lygeros (2014) "Stochastic optimal power flow based on convex approximations of chance constraints," in **Power Systems Computation Conference (PSCC), 2014**, IEEE, pp. 1–7.

[13] Warrington, J., P. Goulart, S. Mariéthoz, and M. Morari (2013) "Policy-based reserves for power systems," **IEEE Transactions on Power Systems**, **28**(4), pp. 4427–4437.

[14] Jabr, R. A., S. Karaki, and J. A. Korbane (2015) "Robust multi-period OPF with storage and renewables," **IEEE Transactions on Power Systems**, **30**(5), pp. 2790–2799.

[15] Jabr, R. A. (2013) "Adjustable robust OPF with renewable energy sources," **IEEE Transactions on Power Systems**, **28**(4), pp. 4742–4751.

[16] Munoz-Alvarez, D., E. Bitar, L. Tong, and J. Wang (2014) "Piecewise affine dispatch policies for economic dispatch under uncertainty," in **PES General Meeting| Conference & Exposition, 2014 IEEE**, IEEE, pp. 1–5.

[17] Bai, X., L. Qu, and W. Qiao (2016) "Robust AC Optimal Power Flow for Power Networks With Wind Power Generation," **IEEE Transactions on Power Systems**, **31**(5), pp. 4163–4164.

[18] Louca, R. and E. Bitar (2016) "Stochastic AC optimal power flow with affine recourse," in **Decision and Control (CDC), 2016 IEEE 55th Conference on**, IEEE, pp. 2431–2436.

[19] Bienstock, D., M. Chertkov, and S. Harnett (2014) "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," **SIAM Review**, **56**(3), pp. 461–495.

[20] Ela, E. and M. O'Malley (2012) "Studying the variability and uncertainty impacts of variable generation at multiple timescales," **IEEE Transactions on Power Systems**, **27**(3), pp. 1324–1333.

[21] VARAIYA, P. P., F. F. WU, and J. W. BIALEK (2011) "Smart operation of smart grid: Risk-limiting dispatch," **Proceedings of the IEEE**, **99**(1), pp. 40–57.

[22] RAJAGOPAL, R., E. BITAR, P. VARAIYA, and F. WU (2013) "Risk-limiting dispatch for integrating renewable power," **International Journal of Electrical Power & Energy Systems**, **44**(1), pp. 615–628.

[23] PENG, C. and Y. HOU (2014) "Risk-limiting dispatch with operation constraints," in **PES General Meeting| Conference & Exposition, 2014 IEEE**, IEEE, pp. 1–5.

[24] ZHANG, B., R. RAJAGOPAL, and D. TSE (2014) "Network risk limiting dispatch: Optimal control and price of uncertainty," **IEEE Transactions on Automatic Control**, **59**(9), pp. 2442–2456.

[25] NORKIN, V. I., Y. M. ERMOLIEV, and A. RUSZCZYŃSKI (1998) "On optimal allocation of indivisibles under uncertainty," **Operations Research**, **46**(3), pp. 381–395.

[26] CARØE, C. C. and J. TIND (1998) "L-shaped decomposition of two-stage stochastic programs with integer recourse," **Mathematical Programming**, **83**(1-3), pp. 451–464.

[27] SHERALI, H. D. and B. M. FRATICELLI (2002) "A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse," **Journal of Global Optimization**, **22**(1-4), pp. 319–342.

[28] CARØE, C. C. and R. SCHULTZ (1999) "Dual decomposition in stochastic integer programming," **Operations Research Letters**, **24**(1-2), pp. 37–45.

[29] KARGARIAN, A., J. MOHAMMADI, J. GUO, S. CHAKRABARTI, M. BARATI, G. HUG, S. KAR, and R. BALDICK (2016) "Toward distributed/decentralized DC optimal power flow implementation in future electric power systems," **IEEE Transactions on Smart Grid**, **9**(4), pp. 2574–2594.

[30] MOLZAHN, D. K., F. DÖRFLER, H. SANDBERG, S. H. LOW, S. CHAKRABARTI, R. BALDICK, and J. LAVAEI (2017) "A survey of distributed optimization and control algorithms for electric power systems," **IEEE Transactions on Smart Grid**, **8**(6), pp. 2941–2962.

[31] CAPITANESCU, F., J. M. RAMOS, P. PANCIATICI, D. KIRSCHEN, A. M. MARCOLINI, L. PLATBROOD, and L. WEHENKEL (2011) "State-of-the-art,

challenges, and future trends in security constrained optimal power flow," **Electric Power Systems Research**, **81**(8), pp. 1731–1741.

[32] Stott, B., J. Jardim, and O. Alsaç (2009) "DC power flow revisited," **IEEE Transactions on Power Systems**, **24**(3), pp. 1290–1300.

[33] Bie, P., H.-D. Chiang, B. Zhang, and N. Zhou (2018) "Online Multiperiod Power Dispatch With Renewable Uncertainty and Storage: A Two-Parameter Homotopy-Enhanced Methodology," **IEEE Transactions on Power Systems**, **33**(6), pp. 6321–6331.

[34] Wong, S. and J. D. Fuller (2007) "Pricing energy and reserves using stochastic optimization in an alternative electricity market," **IEEE Transactions on Power Systems**, **22**(2), pp. 631–638.

[35] Jabr, R. A., S. Karaki, and J. A. Korbane (2014) "Robust multiperiod OPF with storage and renewables," **IEEE Transactions on Power Systems**, **30**(5), pp. 2790–2799.

[36] Thatte, A. A. and L. Xie (2015) "A metric and market construct of inter-temporal flexibility in time-coupled economic dispatch," **IEEE Transactions on Power Systems**, **31**(5), pp. 3437–3446.

[37] Lorca, A. and X. A. Sun (2014) "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," **IEEE Transactions on Power Systems**, **30**(4), pp. 1702–1713.

[38] Pandžić, H., Y. Dvorkin, T. Qiu, Y. Wang, and D. S. Kirschen (2015) "Toward cost-efficient and reliable unit commitment under uncertainty," **IEEE Transactions on Power Systems**, **31**(2), pp. 970–982.

[39] Cho, Y., T. Ishizaki, N. Ramdani, and J.-i. Imura (2019) "Box-based Temporal Decomposition of Multi-period Economic Dispatch for Two-stage Robust Unit Commitment," **IEEE Transactions on Power Systems**.

[40] Zheng, Q. P., J. Wang, and A. L. Liu (2014) "Stochastic optimization for unit commitmentâĂŤA review," **IEEE Transactions on Power Systems**, **30**(4), pp. 1913–1924.

[41] Cerisola, S., Á. Baíllo, J. M. Fernández-López, A. Ramos, and R. Gollmer (2009) "Stochastic power generation unit commitment in electricity markets: A novel formulation and a comparison of solution methods," **Operations Research**, **57**(1), pp. 32–46.

[42] Wu, L., M. Shahidehpour, and T. Li (2007) "Stochastic security-constrained unit commitment," **IEEE Transactions on Power Systems**, **22**(2), pp. 800–811.

[43] Wang, J., M. Shahidehpour, and Z. Li (2008) "Security-constrained unit commitment with volatile wind power generation," **IEEE Transactions on Power Systems**, **23**(3), pp. 1319–1327.

[44] Zou, J., S. Ahmed, and X. A. Sun (2018) "Multistage stochastic unit commitment using stochastic dual dynamic integer programming," **IEEE Transactions on Power Systems**, **34**(3), pp. 1814–1823.

[45] Kalantari, A., J. F. Restrepo, and F. D. Galiana (2012) "Security-constrained unit commitment with uncertain wind generation: The loadability set approach," **IEEE Transactions on Power Systems**, **28**(2), pp. 1787–1796.

[46] Papavasiliou, A., S. S. Oren, and R. P. O'Neill (2011) "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," **IEEE Transactions on Power Systems**, **26**(4), pp. 2197–2206.

[47] Papavasiliou, A. and S. S. Oren (2013) "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," **Operations Research**, **61**(3), pp. 578–592.

[48] Gangammanavar, H., S. Sen, and V. M. Zavala (2015) "Stochastic optimization of sub-hourly economic dispatch with wind energy," **IEEE Transactions on Power Systems**, **31**(2), pp. 949–959.

[49] Mégel, O., G. Andersson, and J. L. Mathieu (2016) "Reducing the computational effort of stochastic multi-period dc optimal power flow with storage," in **2016 Power Systems Computation Conference (PSCC)**, IEEE, pp. 1–7.

[50] Hua, B., D. Schiro, T. Zheng, R. Baldick, and E. Litvinov (2019) "Pricing in Multi-Interval Real-Time Markets," **IEEE Transactions on Power Systems**.

[51] Zhao, J., T. Zheng, and E. Litvinov (2018) "A Multi-Period Market Design for Markets with Intertemporal Constraints," **arXiv preprint arXiv:1812.07034**.

[52] Ela, E. and M. O'Malley (2015) "Scheduling and pricing for expected ramp capability in real-time power markets," **IEEE Transactions on Power Systems**, **31**(3), pp. 1681–1691.

[53] Zhang, H. and P. Li (2011) "Chance constrained programming for optimal power flow under uncertainty," **IEEE Transactions on Power Systems**, **26**(4), pp. 2417–2424.

[54] Bouffard, F., F. D. Galiana, and A. J. Conejo (2005) "Market-clearing with stochastic security-part I: formulation," **IEEE Transactions on Power Systems**, **20**(4), pp. 1818–1826.

[55] Pritchard, G., G. Zakeri, and A. Philpott (2010) "A single-settlement, energy-only electric power market for unpredictable and intermittent participants," **Operations research**, **58**(4-part-2), pp. 1210–1219.

[56] Morales, J. M., M. Zugno, S. Pineda, and P. Pinson (2014) "Electricity market clearing with improved scheduling of stochastic production," **European Journal of Operational Research**, **235**(3), pp. 765–774.

[57] Kirschen, D. S. and G. Strbac (2004) **Fundamentals of power system economics**, John Wiley & Sons.

[58] Pflug, G. C. (2000) "Some remarks on the value-at-risk and the conditional value-at-risk," in **Probabilistic constrained optimization**, Springer, pp. 272–281.

[59] Beck, A. (2017) **First-Order Methods in Optimization**, vol. 25, SIAM.

[60] Shapiro, A., D. Dentcheva, and A. Ruszczyński (2009) **Lectures on stochastic programming: modeling and theory**, SIAM.

[61] Van Slyke, R. M. and R. Wets (1969) "L-shaped linear programs with applications to optimal control and stochastic programming," **SIAM Journal on Applied Mathematics**, **17**(4), pp. 638–663.

[62] Mulvey, J. M. and A. Ruszczyński (1995) "A new scenario decomposition method for large-scale stochastic optimization," **Operations research**, **43**(3), pp. 477–490.

[63] Eckstein, J. and D. P. Bertsekas (1992) "On the DouglasâĂŤRachford splitting method and the proximal point algorithm for maximal monotone operators," **Mathematical Programming**, **55**(1), pp. 293–318.

[64] Ruszczyński, A. (1997) "Decomposition methods in stochastic programming," **Mathematical programming**, **79**(1), pp. 333–353.

[65] Kelley, J. E., Jr (1960) "The cutting-plane method for solving convex programs," **Journal of the society for Industrial and Applied Mathematics**, **8**(4), pp. 703–712.

[66] King, A. J. and R. J. Wets* (1991) "Epi-consistency of convex stochastic programs," **Stochastics and Stochastic Reports**, **34**(1-2), pp. 83–92.

[67] Robinson, S. M. (1996) "Analysis of sample-path optimization," **Mathematics of Operations Research**, **21**(3), pp. 513–528.

[68] King, A. J. and R. T. Rockafellar (1993) "Asymptotic theory for solutions in statistical estimation and stochastic programming," **Mathematics of Operations Research**, **18**(1), pp. 148–162.

[69] Shapiro, A. (1989) "Asymptotic properties of statistical estimators in stochastic programming," **The Annals of Statistics**, pp. 841–858.

[70] Robbins, H. and S. Monro (1951) "A stochastic approximation method," **The annals of mathematical statistics**, pp. 400–407.

[71] Kushner, H. J. and D. S. Clark (1978) "Stochastic approximation methods for constrained and unconstrained systems," .

[72] Nevel'son, M. B. and R. Z. Khas' minskii (1976) **Stochastic approximation and recursive estimation**, vol. 47, American Mathematical Society Providence.

[73] Polyak, B. T. and A. B. Juditsky (1992) "Acceleration of stochastic approximation by averaging," **SIAM Journal on Control and Optimization**, **30**(4), pp. 838–855.

[74] Nemirovski, A. and D. Yudin (1978) "On CezariâĂŹs convergence of the steepest descent method for approximating saddle point of convex-concave functions," in **Soviet Math. Dokl**, vol. 19.

[75] Ermoliev, Y. M. (1969) "On the stochastic quasi-gradient method and stochastic quasi-Feyer sequences," **Kibernetika**, **2**, pp. 72–83.

[76] Gavironski, A. (1988) "Implementation of stochastic quasigradient methods," **Numerical Techniques for Stochastic Optimization. Springer-Verlag**.

[77] Wets, R. (1983) "Stochastic programming: Solution techniques and approximation schemes," in **Mathematical programming The state of the art**, Springer, pp. 566–603.

[78] Shanbhag, U. V. and J. H. Blanchet (2015) "Budget-constrained stochastic approximation," in **Winter Simulation Conference (WSC), 2015**, IEEE, pp. 368–379.

[79] Jalilzadeh, A. and U. V. Shanbhag (2016) "eg-VSSA: an extragradient variable sample-size stochastic approximation scheme: error analysis and complexity trade-offs," in **Proceedings of the 2016 Winter Simulation Conference**, IEEE Press, pp. 690–701.

[80] Jalilzadeh, A., U. V. Shanbhag, J. H. Blanchet, and P. W. Glynn (2017) "On the analysis of unaccelerated and accelerated variable sample-size stochastic approximation (VSSA) schemes," Manuscript submitted for publication.

[81] ——— (2018) "Optimal smoothed variable sample-size accelerated proximal methods for structured nonsmooth stochastic convex programs," **arXiv preprint arXiv:1803.00718**.

[82] Nesterov, Y. (1983) "A method for unconstrained convex minimization problem with the rate of convergence O (1/kˆ 2)," in **Doklady AN USSR**, vol. 269, pp. 543–547.

[83] Zhang, Y., J. Wang, and X. Wang (2014) "Review on probabilistic forecasting of wind power generation," **Renewable and Sustainable Energy Reviews**, **32**, pp. 255–270.

[84] Gangammanavar, H. (2013) **Multiple timescale stochastic optimization with application to integrating renewable resources in power systems**, Ph.D. thesis, The Ohio State University.

[85] Brown, B. G., R. W. Katz, and A. H. Murphy (1984) "Time series models to simulate and forecast wind speed and wind power," **Journal of climate and applied meteorology**, **23**(8), pp. 1184–1195.

[86] Garcia, A., J. Torres, E. Prieto, and A. De Francisco (1998) "Fitting wind speed distributions: a case study," **Solar energy**, **62**(2), pp. 139–144.

[87] Sheble, G. B. and G. N. Fahd (1994) "Unit commitment literature synopsis," **Power Systems, IEEE Transactions on**, **9**(1), pp. 128–135.

[88] Merlin, A. and P. Sandrin (1983) "A new method for unit commitment at Electricité de France," **IEEE Trans. Power Appar. Syst.;(United States)**, **102**(5).

[89] Tong, S. and S. Shahidehpour (1989) "Combination of Lagrangian-relaxation and linear-programming approaches for fuel-constrained unit-commitment problems," in **IEE Proceedings C (Generation, Transmission and Distribution)**, vol. 136, IET, pp. 162–174.

[90] Aoki, K., M. Itoh, T. Satoh, K. Nara, and M. Kanezashi (1989) "Optimal long-term unit commitment in large scale systems including fuel constrained thermal and pumped-storage hydro," **Power Systems, IEEE Transactions on**, **4**(3), pp. 1065–1073.

[91] Feltenmark, S., K. C. Kiwiel, and P.-O. Lindberg (1997) "Solving unit commitment problems in power production planning," in **Operations Research Proceedings 1996**, Springer, pp. 236–241.

[92] Zheng, Q. P., J. Wang, P. M. Pardalos, and Y. Guan (2013) "A decomposition approach to the two-stage stochastic unit commitment problem," **Annals of Operations Research**, **210**(1), pp. 387–410.

[93] Laporte, G. and F. V. Louveaux (1993) "The integer L-shaped method for stochastic integer programs with complete recourse," **Operations research letters**, **13**(3), pp. 133–142.

[94] Ahmed, S., M. Tawarmalani, and N. V. Sahinidis (2004) "A finite branch-and-bound algorithm for two-stage stochastic integer programs," **Mathematical Programming**, **100**(2), pp. 355–377.

[95] Gröwe-Kuska, N., K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner (2002) "Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation," in **Decision making under uncertainty**, Springer, pp. 39–70.

[96] Nowak, M. P. and W. Römisch (2000) "Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty," **Annals of Operations Research**, **100**(1-4), pp. 251–272.

[97] Takriti, S. and J. R. Birge (2000) "Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs," **Operations Research**, **48**(1), pp. 91–98.

[98] Watson, J.-P. and D. L. Woodruff (2011) "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," **Computational Management Science**, **8**(4), pp. 355–370.

[99] Gade, D., G. Hackebeil, S. M. Ryan, J.-P. Watson, R. J.-B. Wets, and D. L. Woodruff (2016) "Obtaining lower bounds from the progressive hedging algorithm for stochastic mixed-integer programs," **Mathematical Programming**, **157**(1), pp. 47–67.

[100] Lawler, E. L. and D. E. Wood (1966) "Branch-and-bound methods: A survey," **Operations research**, **14**(4), pp. 699–719.

[101] Wolsey, L. A. and G. L. Nemhauser (2014) **Integer and combinatorial optimization**, John Wiley & Sons.

[102] Gupta, O. K. and A. Ravindran (1985) "Branch and bound experiments in convex nonlinear integer programming," **Management science**, **31**(12), pp. 1533–1546.

[103] Quesada, I. and I. E. Grossmann (1992) "An LP/NLP based branch and bound algorithm for convex MINLP optimization problems," **Computers & chemical engineering**, **16**(10), pp. 937–947.

[104] Fletcher, R. and S. Leyffer (1998) "Numerical experience with lower bounds for MIQP branch-and-bound," **SIAM Journal on Optimization**, **8**(2), pp. 604–616.

[105] Padberg, M. and G. Rinaldi (1987) "Optimization of a 532-city symmetric traveling salesman problem by branch and cut," **Operations Research Letters**, **6**(1), pp. 1–7.

[106] ——— (1991) "A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems," **SIAM review**, **33**(1), pp. 60–100.

[107] Stubbs, R. A. and S. Mehrotra (1999) "A branch-and-cut method for 0-1 mixed convex programming," **Mathematical programming**, **86**(3), pp. 515–532.

[108] Geoffrion, A. M. (1974) **Lagrangean relaxation for integer programming**, Springer.

[109] Held, M. and R. M. Karp (1970) "The traveling-salesman problem and minimum spanning trees," **Operations Research**, **18**(6), pp. 1138–1162.

[110] ——— (1971) "The traveling-salesman problem and minimum spanning trees: Part II," **Mathematical programming**, **1**(1), pp. 6–25.

[111] Shapiro, J. F. (1971) "Generalized Lagrange multipliers in integer programming," **Operations Research**, **19**(1), pp. 68–76.

[112] Fisher, M. L. and J. F. Shapiro (1974) "Constructive duality in integer programming," **SIAM Journal on Applied Mathematics**, **27**(1), pp. 31–52.

[113] Fisher, M. L. (2004) "The Lagrangian relaxation method for solving integer programming problems," **Management science**, **50**(12_supplement), pp. 1861–1871.

[114] Geoffrion, A. M. (1972) "Generalized benders decomposition," **Journal of optimization theory and applications**, **10**(4), pp. 237–260.

[115] Duran, M. A. and I. E. Grossmann (1986) "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," **Mathematical programming**, **36**(3), pp. 307–339.

[116] Viswanathan, J. and I. E. Grossmann (1990) "A combined penalty function and outer-approximation method for MINLP optimization," **Computers & Chemical Engineering**, **14**(7), pp. 769–782.

[117] Fletcher, R. and S. Leyffer (1994) "Solving mixed integer nonlinear programs by outer approximation," **Mathematical programming**, **66**(1-3), pp. 327–349.

[118] Ntaimo, L. and S. Sen (2005) "The million-variable âĂIJmarchâĂİ for stochastic combinatorial optimization," **Journal of Global Optimization**, **32**(3), pp. 385–400.

[119] Sen, S. and J. L. Higle (2005) "The C 3 theorem and a D 2 algorithm for large scale stochastic mixed-integer programming: set convexification," **Mathematical Programming**, **104**(1), pp. 1–20.

[120] Lulli, G. and S. Sen (2004) "A branch-and-price algorithm for multistage stochastic integer programming with application to stochastic batch-sizing problems," **Management Science**, **50**(6), pp. 786–796.

[121] Lubin, M., K. Martin, C. G. Petra, and B. Sandikçi (2013) "On parallelizing dual decomposition in stochastic integer programming," **Operations Research Letters**, **41**(3), pp. 252–258.

[122] Lauer, G., N. Sandell Jr, D. Bertsekas, and T. Posbergh (1982) "Solution of large-scale optimal unit commitment problems," **Power Apparatus and Systems, IEEE Transactions on**, (1), pp. 79–86.

[123] Cohen, A. I. and M. Yoshimura (1983) "A branch-and-bound algorithm for unit commitment," **IEEE Trans. Power Appar. Syst.;(United States)**, **102**(2).

[124] Goez, J., J. Luedtke, D. Rajan, and J. Kalagnanam (2008) **Stochasic Unit Committment Problem**, **Tech. rep.**, IBM.

[125] Abhyankar, K., T. Fogarty, J. L. Kimber, A. Lin, S. Seo, and S. Takriti (1998) "Deterministic and stochastic models for the unit commitment problem," .

[126] Gomory, R. E. et al. (1958) "Outline of an algorithm for integer solutions to linear programs," **Bulletin of the American Mathematical society**, **64**(5), pp. 275–278.

[127] Gomory, R. (1960) **An algorithm for the mixed integer problem**, **Tech. rep.**, RAND CORP SANTA MONICA CA.

[128] Wolsey, L. A. (1998) **Integer programming**, Wiley.

[129] Sen, S., R. D. Doverspike, and S. Cosares (1994) "Network planning with random demand," **Telecommunication systems**, **3**(1), pp. 11–30.

[130] Carpentier, J. (1962) "Contribution to the economic dispatch problem," **Bulletin de la Societe Francoise des Electriciens**, **3**(8), pp. 431–447.

[131] Dommel, H. W. and W. F. Tinney (1968) "Optimal power flow solutions," **IEEE Transactions on power apparatus and systems**, (10), pp. 1866–1876.

[132] Capitanescu, F. and L. Wehenkel (2013) "Experiments with the interior-point method for solving large scale optimal power flow problems," **Electric Power Systems Research**, **95**, pp. 276–283.

[133] Cain, M. B., R. P. OâĂŹNeill, and A. Castillo (2012) "History of optimal power flow and formulations," **Federal Energy Regulatory Commission**, pp. 1–36.

[134] Frank, S., I. Steponavice, and S. Rebennack (2012) "Optimal power flow: a bibliographic survey I," **Energy Systems**, **3**(3), pp. 221–258.

[135] Castillo, A., P. Lipka, J.-P. Watson, S. S. Oren, and R. P. OâĂŹNeill (2016) "A successive linear programming approach to solving the IV-ACOPF," **IEEE Trans. Power Syst**, **31**(4), pp. 2752–2763.

[136] Frank, S. and S. Rebennack (2016) "An introduction to optimal power flow: Theory, formulation, and examples," **IIE Transactions**, **48**(12), pp. 1172–1197.

[137] Lavaei, J. and S. H. Low (2012) "Zero duality gap in optimal power flow problem," **IEEE Transactions on Power Systems**, **27**(1), p. 92.

[138] Low, S. H. (2014) "Convex relaxation of optimal power flow–Part I: Formulations and equivalence," **IEEE Transactions on Control of Network Systems**, **1**(1), pp. 15–27.

[139] ——— (2014) "Convex relaxation of optimal power flow–Part II: Exactness," **IEEE Transactions on Control of Network Systems**, **1**(2), pp. 177–189.

[140] Gan, L., N. Li, U. Topcu, and S. H. Low (2015) "Exact convex relaxation of optimal power flow in radial networks," **IEEE Transactions on Automatic Control**, **60**(1), pp. 72–87.

[141] Madani, R., S. Sojoudi, and J. Lavaei (2015) "Convex relaxation for optimal power flow problem: Mesh networks," **IEEE Transactions on Power Systems**, **30**(1), pp. 199–211.

[142] Molzahn, D. K., J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco (2013) "Implementation of a large-scale optimal power flow solver based on semidefinite programming," **IEEE Transactions on Power Systems**, **28**(4), pp. 3987–3998.

[143] Coffrin, C., H. L. Hijazi, and P. Van Hentenryck (2017) "Strengthening the SDP relaxation of AC power flows with convex envelopes, bound tightening, and valid inequalities," **IEEE Transactions on Power Systems**, **32**(5), pp. 3549–3558.

[144] Molzahn, D. K. and I. A. Hiskens (2014) "Moment-based relaxation of the optimal power flow problem," in **Power Systems Computation Conference (PSCC), 2014**, IEEE, pp. 1–7.

[145] Molzahn, D. K., C. Josz, I. A. Hiskens, and P. Panciatici (2017) "A Laplacian-based approach for finding near globally optimal solutions to OPF problems," **IEEE Transactions on Power Systems**, **32**(1), pp. 305–315.

[146] Jabr, R. A. (2006) "Radial distribution load flow using conic programming," **IEEE transactions on power systems**, **21**(3), pp. 1458–1459.

[147] Kocuk, B., S. S. Dey, and X. A. Sun (2016) "Strong SOCP relaxations for the optimal power flow problem," **Operations Research**, **64**(6), pp. 1177–1196.

[148] Conejo, A. J. and J. A. Aguado (1998) "Multi-area coordinated decentralized DC optimal power flow," **IEEE Transactions on Power Systems**, **13**(4), pp. 1272–1278.

[149] Lam, A. Y., B. Zhang, and N. T. David (2012) "Distributed algorithms for optimal power flow problem," in **2012 IEEE 51st IEEE Conference on Decision and Control (CDC)**, IEEE, pp. 430–437.

[150] Eckstein, J. (1989) **Splitting methods for monotone operators with applications to parallel optimization**, Ph.D. thesis, Massachusetts Institute of Technology.

[151] Ouyang, H., N. He, L. Tran, and A. Gray (2013) "Stochastic alternating direction method of multipliers," in **International Conference on Machine Learning**, pp. 80–88.

[152] Liu, Y., F. Shang, and J. Cheng (2017) "Accelerated variance reduced stochastic ADMM," in **Thirty-First AAAI Conference on Artificial Intelligence**.

[153] Zheng, S. and J. T. Kwok (2016) "Fast-and-Light Stochastic ADMM." in **IJCAI**, pp. 2407–2613.

[154] Hong, M., Z.-Q. Luo, and M. Razaviyayn (2016) "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," **SIAM Journal on Optimization**, **26**(1), pp. 337–364.

[155] Wang, Y., W. Yin, and J. Zeng (2019) "Global convergence of ADMM in nonconvex nonsmooth optimization," **Journal of Scientific Computing**, **78**(1), pp. 29–63.

[156] Xie, Y. and U. V. Shanbhag (2016) "SI-ADMM: A stochastic inexact ADMM framework for resolving structured stochastic convex programs," in **Proceedings of the 2016 Winter Simulation Conference**, IEEE Press, pp. 714–725.

[157] ——— (2017) "Tractable ADMM Schemes for Computing KKT Points and Local Minimizers for ℓ_0 Minimization Problems," **arXiv preprint arXiv:1710.04613**.

[158] Sun, A. X., D. T. Phan, and S. Ghosh (2013) "Fully decentralized AC optimal power flow algorithms," in **2013 IEEE Power & Energy Society General Meeting**, IEEE, pp. 1–5.

[159] Erseghe, T. (2015) "A distributed approach to the OPF problem," **EURASIP Journal on Advances in Signal Processing**, **2015**(1), p. 45.

[160] Mhanna, S., A. C. Chapman, and G. Verbič (2018) "Component-based dual decomposition methods for the OPF problem," **Sustainable Energy, Grids and Networks**, **16**, pp. 91–110.

[161] Ouyang, Y., Y. Chen, G. Lan, and E. Pasiliao Jr (2015) "An accelerated linearized alternating direction method of multipliers," **SIAM Journal on Imaging Sciences**, **8**(1), pp. 644–681.

[162] Bai, X., H. Wei, K. Fujisawa, and Y. Wang (2008) "Semidefinite programming for optimal power flow problems," **International Journal of Electrical Power & Energy Systems**, **30**(6), pp. 383–392.

[163] Baradar, M. and M. R. Hesamzadeh (2014) "AC power flow representation in conic format," **IEEE Transactions on Power Systems**, **30**(1), pp. 546–547.

Vita

Wendian Wan

Wendian Wan was born and raised in Wuhan, China. He received his B.S. in Industrial Engineering in 2013 from Tsinghua University in Beijing, China and M.S. in Industrial Engineering and Operations Research from Penn State University in 2015. He is currently a Ph.D. candidate in Industrial Engineering and Operations Research at The Pennsylvania State University. Wendian's research interests includes stochastic programming and its application in power system optimization.