Distributionally Robust Optimal Power Flow with Strengthened Ambiguity Sets

by

Bowen Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2019

Doctoral Committee:

Assistant Professor Johanna L. Mathieu, Chair
Professor Ian A. Hiskens
Assistant Professor Ruiwei Jiang
Professor Pascal R. Van Hentenryck

Bowen Li
libowen@umich.edu
ORCID id: 0000-0002-7789-6226

To my parents and friends,

for supporting and encouraging me towards

this destination

ACKNOWLEDGEMENTS

I would like to sincerely express my deepest gratitude to the many people who have supported and encouraged me to complete this journey. The last five years have been an unforgettable experience and will continue to guide me in my future career.

First, I would like to thank my advisor, Professor Johanna Mathieu. Her consistent support and patient guidance have always been the foundation of all my projects. Her outstanding leadership and thoughtfulness granted me the courage to explore new ideas and pursue higher standards. Her creativity and inspiration provided me with innovative paths to tackle challenging problems. It is my great honor and pleasure to learn power systems and conduct research with her.

Next, I would like to thank my co-advisor and collaborator Ruiwei Jiang. His vast knowledge of optimization theory benefitted me greatly by helping me to tackle the theoretical challenges of my research. Through the process, I learned the importance of critical thinking to solve challenging problems and was deeply influenced by his persistence and effort towards research.

I am also very grateful for the assistance and advice provided by the other members of my doctoral committee: Professor Ian Hiskens and Professor Pascal Van Hentenryck. Their invaluable suggestions and feedback were of great assistance in terms of improving the work. They also provided helpful career advice that enabled me to envision a future path.

In addition, I would like to thank Line Roald, Harsha Nagarajan, and Russell Bent for involving me in an exceptional summer internship at the Los Alamos National

Laboratory. This opportunity allowed me to learn important real-world problems and how to apply what I have studied to tackle challenging questions.

I would also like to thank my collaborators: Yiling Zhang, Yashen Lin, Spencer Maroukis, Jialin Liu, Maria Gabriela Martinez, Professor Siqian Shen, Professor Maria Vrakopoulou, and Professor Lindsay Anderson. Without the great insights and creativity of these people, it would be difficult to replicate the important contributions of several essential projects. I have also learned many lessons from the collaborative group discussions on how to solve different problems.

I cannot express enough gratitude for my parents' selfless sacrifices and love. Although they are living in China, I never felt alone because of their supportive texts or video messages. Whenever I feel lost or depressed, they are always there to cheer me up and guide me with their own experiences.

I would like to thank our MPEL group, which has been another big family for me in the past five years. Thank you for your helpful suggestions in the group meetings, inspirational conversations about research, and the coffee breaks and beer parties to temporarily escape from the research. I am fortunate to know you all and I sincerely wish you the best.

I would like to thank my previous advisor Juan Rivas for my first research experience, which taught me how to conduct the full research procedure independently. This experience also boosted my confidence in the following research.

Lastly, I am grateful to my advisor Johanna Mathieu, the Department of Energy, and the National Science Foundation #CCF-1442495 for financially supporting all the research.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ALGORITHMS

Algorithm

ABSTRACT

Uncertainties that result from renewable generation and load consumption can complicate the optimal power flow problem. These uncertainties normally influence the physical constraints stochastically and require special methodologies to solve. Hence, a variety of stochastic optimal power flow formulations using chance constraints have been proposed to reduce the risk of physical constraint violations and ensure a reliable dispatch solution under uncertainty. The true uncertainty distribution is required to exactly reformulate the problem, but it is generally difficult to obtain. Conventional approaches include randomized techniques (such as scenario-based methods) that provide a priori guarantees of the probability of constraint violations but generally require many scenarios and produce high-cost solutions. Another approach is to use an analytical reformulation, which assumes that the uncertainties follow specific distributions such as Gaussian distributions. However, if the actual uncertainty distributions do not follow the assumed distributions, the results often suffer from case-dependent reliability.

Recently, researchers have also explored distributionally robust optimization, which requires probabilistic constraints to be satisfied at chosen probability levels for any uncertainty distributions within a pre-defined ambiguity set. The set is constructed based on the statistical information that is extracted from historical data. Existing literature applying distributionally robust optimization to the optimal power flow problem indicates that the approach has promising performance with low objective costs as well as high reliability compared with the randomized techniques and

analytical reformulation. In this dissertation, we aim to analyze the conventional approaches and further improve the current distributionally robust methods.

In Chapter II, we derive the analytical reformulation of a multi-period optimal power flow problem with uncertain renewable generation and load-based reserve. It is assumed that the capacities of the load-based reserves are affected by outdoor temperatures through non-linear relationships. Case studies compare the analytical reformulation with the scenario-based method and demonstrate that the scenario-based method generates overly-conservative results and the analytical reformulation results in lower cost solutions but it suffers from reliability issues.

In Chapters III, IV, and V, we develop new methodologies in distributionally robust optimization by strengthening the moment-based ambiguity set by including a combination of the moment, support, and structural property information. Specifically, we consider unimodality and log-concavity as most practical uncertainties exhibit these properties. The strengthened ambiguity sets are used to develop tractable reformulations, approximations, and efficient algorithms for the optimal power flow problem. Case studies indicate that these strengthened ambiguity sets reduce the conservativeness of the solutions and result in sufficiently reliable solutions.

In Chapter VI, we compare the performance of the conventional approaches and distributionally robust approaches including moment and unimodality information on large-scale systems with high uncertainty dimensions. Through case studies, we evaluate each approach's performance by exploring its objective cost, computational scalability, and reliability. Simulation results suggest that distributionally robust optimal power flow including unimodality information produces solutions with better trade-offs between objective cost and reliability as compared to the conventional approaches or the distributionally robust approaches that do not include unimodality assumptions. However, considering unimodality also leads to longer computational times.

CHAPTER I

Introduction

1.1 Motivation and Objectives

Power systems are becoming more complicated with the integration of advanced equipment, sensing, and control techniques and the development of optimal operational approaches. Although these changes have brought significant benefits to power system operation, they have also introduced different uncertainties to the system. These uncertainties can include uncertain power injections from renewable generation, load consumption forecast errors, or random component failures. For example, more renewable resources are being incorporated into the power grid to achieve low operational costs and less environmental impact. Since the renewable generation is generally non-dispatchable and highly dependent on the ambient conditions such as the wind and solar level, the actual power output is uncertain and may deviate from the forecasts.

Similarly, in terms of load consumption, there is generally a power forecast mismatch between real-time consumption and the demand forecast. To ensure power balance between total demand and supply, these power mismatches must be compensated by backup resources or services such as spinning reserve, non-spinning reserve, and frequency regulation [57, 34]. In the following paragraphs, these services will be referred to as 'reserves' for short. If system operators fail to consider the risk

from these uncertainties, there might be disastrous outcomes such as severe power imbalance, loss of critical infrastructure, or even massive system failure. Hence, to ensure reliable operations, it is important to manage the risk of these uncertainties. In addition to scheduling more conventional backup resources, it is also necessary to harness additional types of flexible resources to provide reserves and to develop operational planning methodologies to schedule the generation and reserve capacities in a way that balances system cost and reliability.

In addition to conventional generator reserve providers such as thermal and hydropower plants, flexible loads like electric heaters, air conditioners, and electric vehicles are also capable of providing reserves [15, 16, 37, 5, 63, 105, 64, 53]. For example, recent studies have developed methods to use commercial and residential loads to provide reserves such as load following and frequency regulation [63, 64]. Compared with generator reserves, load reserves are expected to respond faster, have lower operational costs, and less environmental impact [16, 45].

Controllable loads are externally coordinated to provide reserves by changing their power consumption from normal operation. Unlike generator reserves, the flexibility of these controllable loads is uncertain and time-varying because their operations are closely related to human behaviors and ambient conditions [61, 60]. For example, in [62], an aggregated thermal energy storage model was used to estimate the power and energy flexibility of an aggregation of electric heaters as a function of outdoor temperature. Recent studies have also developed methods to dispatch controllable loads within the optimal power flow (OPF) or unit commitment problem [46, 71].

Stochastic OPF has been proposed to manage the increasing need for reserve capacity and risks from uncertainties such as renewable power production and load consumption. Stochastic OPF is an important tool for system planning because it helps to minimize system operational costs while managing the risk from uncertainties. Efficient methodologies to solve the stochastic OPF problem benefit both the system

operators and the consumers by generating solutions with low objective costs but high reliability in the out-of-sample tests. To ensure the approaches are practically applicable they must also require low computational effort. Otherwise, the approaches are difficult to apply in real time operations and the resulting solution might reduce the welfare of the consumers and pose risks to system reliability and stability.

One way to formulate the stochastic OPF problem is to use two-stage stochastic optimization [103, 76, 47]. In this formulation, the first stage contains all the decision variables and corresponding constraints that are deterministic and independent of the uncertainty. The second stage instead contains the remaining decision variables and constraints that are affected by the uncertainty realizations. The overall objective cost is a summation of the operational cost from the first stage and an expected cost from the second stage given the uncertainty. Another way to pose the stochastic OPF problem is as a chance-constrained optimization problem [94, 104, 42, 96, 79, 97]. This formulation enables physical constraints such as line limits and generation limits to be formulated as probabilistic constraints that must be satisfied for most of the realizations of system uncertainties. For example, a chance-constrained formulation to optimally determine the dispatch and reserve capacities for generators in systems with uncertain renewable production is discussed in [96]. Comparing these two stochastic OPF formulations, the two-stage formulation aims to determine the optimal operations under uncertainty while the chance-constrained formulation tries to manage the risks from constraint violation. In this dissertation, we only focus on solving different types of chance-constrained optimal power flow problems.

The true underlying distribution is required to exactly reformulate the probabilistic constraints. However, that distribution is normally unknown and hard to estimate. To solve the problem without the true distribution, researchers have proposed different solving methodologies. The first method uses randomized techniques such as the scenario approach [17] or the scenario-based method [58] that uses robust opti-

mization. Uncertainty scenarios or their probabilistically robust support are used to transform the probabilistic constraints into deterministic constraints. With enough scenarios, both methods provide a priori guarantees that the original probabilistic constraints are satisfied regardless of the true underlying distribution. However, randomized techniques can provide excessively-conservative solutions with high objective costs and reliability beyond the requirements, and there might not be enough data points to enable the reformulation. Recent work that uses the scenario-based method to solve the chance-constrained OPF problems is discussed in [95, 96].

Another method is to reformulate the constraints by assuming that the uncertainties follow specific distributions such as the Gaussian distribution [9, 79]. Since the reformulation does not change the problem dimension of the chance-constrained problem, analytical reformulation normally requires less computation effort and generates results with lower objective costs than the randomized techniques. However, analytical reformulation has no a priori guarantee of the out-of-sample reliability of the resulting solution. The empirical reliability is case-dependent and can be evaluated using the Monte Carlo analysis. Previous studies [9, 79] have used analytical reformulation by assuming that the uncertainty distributions are Gaussian and stochastic constraints are affine with respect to the random variables. It is more difficult to reformulate the problem if the constraints are nonlinear with respect to the random variables.

Researchers have also explored distributionally robust (DR) optimization to overcome the challenge of estimating the true underlying distribution [13, 22, 27, 36, 84, 101, 44, 110] and to manage the stochasticity in OPF problems [107, 80, 102, 48, 88, 33, 56]. This approach requires the probabilistic constraints to be satisfied with all the possible uncertainty distributions in a pre-defined ambiguity set. The set is normally constructed with statistical information (e.g., first and second moment) extracted from the empirical uncertainty data. Since the DR formulation evaluates the proba-

bilistic constraint using the worst-case distribution of the ambiguity set, it produces more conservative results than the approach using the true distribution. Ambiguity sets should include adequate and accurate statistical information about the true distribution. Otherwise, the solution may be conservative (if the ambiguity set is too large, i.e., more general), or the approach can become computationally intractable (if the ambiguity set is very precise). Recent studies have demonstrated that DR optimization can be used to solve chance-constrained OPF [107, 80, 102, 48, 88, 33, 56] and leads to good trade-offs between objective costs and out-of-sample reliability. However, the ambiguity sets in these works either only include moment information or consider simple univariate structural properties such as symmetry or unimodality.

A key question is how to achieve better solutions with lower costs and high reliability using DR optimization. An intuitive direction is to strengthen the ambiguity sets. The underlying reason is shown via the trade-off triangle in Fig. 1.1. In OPF problems, it is normally difficult to achieve high reliability with low costs. However, with DR optimization, both optimal cost and reliability are connected to the generality of the ambiguity set. In particular, reliability is positively correlated to the generality of the ambiguity set. Similarly, low cost and high generality must trade off because more generality means that the ambiguity set contains more distributions.

Furthermore, the optimal cost is more likely to be driven by the extreme distributions in the ambiguity set. However, the reliability is normally driven by the extreme of the typical distributions in the ambiguity set because practical uncertainties are unlikely to follow distributions with abnormal structure. Hence, if the atypical distributions could be effectively removed from the ambiguity set, solutions with low costs and high reliability could be achieved. For example, empirical wind generation forecast errors are generally unimodal and log-concave. These properties can provide more accurate structural information in terms of describing the distributions. We can strengthen the ambiguity sets by including these structural properties and develop

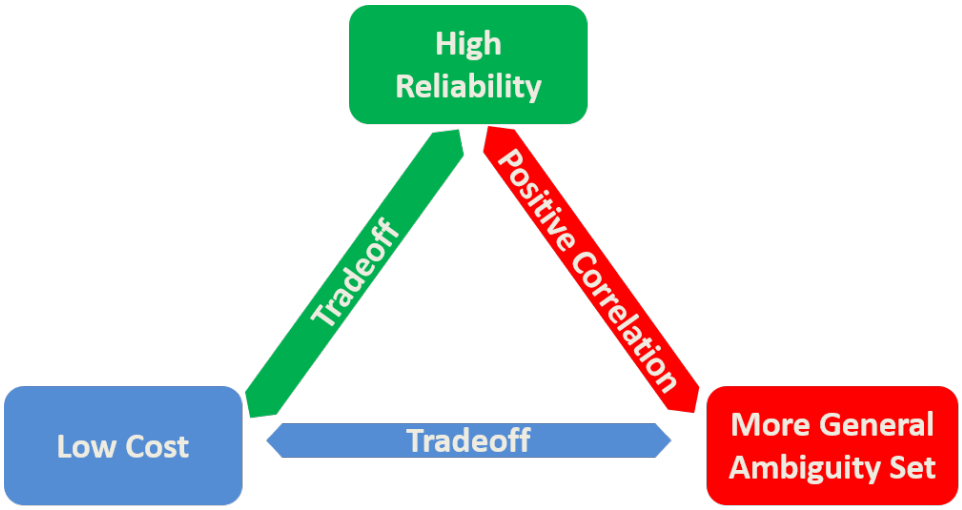


Figure 1.1: Trade-off triangle in distributionally robust optimization.

the corresponding solving methodologies.

Based on the background, the literature review, and the identified problems above, the objectives of this dissertation are as follows.

Objective 1: Incorporate uncertain load-based reserves into the stochastic OPF problem with uncertain renewable generation. Derive the analytical reformulation and compare the performance (i.e., objective cost, solution reliability, and computational tractability) with the scenario-based method.

Objective 2: Strengthen the existing moment-based ambiguity set with more structural properties such as generalized unimodality, log-concavity, and bounded support. Compare the performance with the moment-based DR approaches on a chance-constrained OPF problem.

Objective 3: Compare the performance of different conventional and newly developed approaches by performing case studies on a large-scale power network with high uncertainty dimension to gain insights on the advantages and disadvantages of each approach.

1.2 Contributions and Structure of the Dissertation

In this dissertation, we completed different tasks based on the objectives. The contributions can be summarized as follows.

For the first objective, we first derived a specific analytical reformulation for a multi-period OPF problem with uncertain wind generation and load-based reserves provided by an aggregation of electric heaters. It was assumed that the flexibility of the controllable load would be affected by the ambient temperature and that both wind and temperature uncertainties would follow multivariate Gaussian distributions.

Next, we used empirical examples and a proof sketch to demonstrate that the reformulated constraints are convex. The reformulated problem was solved using a nonlinear solver and an iterative cutting plane approach [9]. Then, we presented the scalability of the reformulation and performed case studies to compare the objective costs, computational requirements, and reliability with the scenario-based method. Analysis was also performed to determine how the different types and levels of uncertainty, cost relationship between reserve and generation, and controllable load capacity affect power system dispatch, operational costs, and CO_2 emissions.

For the second objective, the following contributions were made.

1. We first incorporated generalized multivariate unimodality assumptions into the DR chance and risk-constrained OPF problems by using moment information and then developed tractable reformulations and approximations. For risk constraints, we used conditional value-at-risk [81, 82] to measure the expectation of the constraint violation and bounded the expected risk with pre-defined tolerance. The separation approach was used to derive an efficient solving algorithm [67]. We also performed case studies on a simple OPF problem with uncertain wind generation. The simulation results were compared with existing DR approaches using only moment information.

2. We derived a tractable sandwich approximation for a DR chance-constrained OPF problem that assumes log-concave distributions. We also compared with solutions obtained from analytical reformulation under Gaussian assumptions and DR approaches with only moment information.

3. We studied an ambiguity set based on moment and unimodality information with a potentially misspecified mode location. We demonstrated that the DR chance constraints could be recast as a set of second-order conic (SOC) constraints. Furthermore, we derived an efficient iterative algorithm that guarantees global optimality. Then, we performed case studies to compare the operational cost, reliability, computational time, and optimal solutions to those obtained using ambiguity sets with the exact moment or unimodality information [49, 87, 80, 107].

4. We developed a new optimal parameter selection (OPS) technique to determine the optimal sandwich approximation of the DR chance-constrained problem with moment and unimodality information. The OPS problem is equivalent to finding the optimal piecewise linear (PWL) outer approximation of a concave function and provides us with multiple options (i.e., online or offline) to calculate the approximation. Finally, we compared the approximate solutions obtained from OPS with the solutions from the default settings [49] to demonstrate the improvements on computational time and solution quality.

For the third objective, we compared the conventional and newly developed DR approaches with unimodality and moment information on a large-dimension OPF problem with uncertain wind generation. Two different wind data sets were used to explore diverse uncertainty correlations. The performance difference and relative rankings of all the tested approaches were analyzed to demonstrate the trade-off between reliability and objective costs.

The main content of the dissertation is divided into five further chapters.

Chapter II presents the modeling technique, problem formulation, different solving methodologies, and simulation analysis for a multi-period OPF problem with uncertain wind generation and load-based reserves. The chapter also demonstrates how the types and levels of uncertainty, reserve costs, and controllable load capacity affect the dispatch solution, operational costs, and CO_2 emissions for different approaches.

Chapter III considers DR chance and risk constraints and presents the exact reformulations, asymptotic sandwich approximations, and efficient solving algorithms for a DR OPF problem with moment and unimodality information.

Chapter IV presents the sandwich approximation of a DR chance-constrained OPF formulation with the first-order moment, ellipsoidal support, and an assumption that the distribution is log-concave.

Chapter V presents a DR chance-constrained OPF problem in which the constraints are satisfied over a set of unimodal distributions with known first and second-order moments but misspecified mode values. These values are also assumed to be bounded within either a rectangular or ellipsoidal support.

Chapter VI presents the OPS technique that properly selects the parameters used in the conservative approximation that is presented in Chapter III. The chapter also presents the detailed simulation results of comparing the conventional analytical reformulation under Gaussian assumption, DR approach with only moment information and all the developed DR approaches with unimodality information.

Chapter VII summarizes the dissertation and presents potential future work.

CHAPTER II

Chance Constrained Optimal Power Flow Using Uncertain Controllable Loads

In this chapter, we solve a multi-period chance constrained optimal power flow problem to optimally schedule generation and reserves from both generators and aggregations of controllable electric loads. Unlike generator-based reserve, load-based reserve capacities are less certain because they depend on load usage patterns and ambient conditions. To handle the uncertainties from load-based reserves and renewable generation, we analytically reformulate the problem assuming the uncertainties follow multivariate normal distributions. We compare the results against the conventional scenario-based method. We find that the analytical reformulation solved using a cutting plane algorithm requires less computational time than the scenario-based method. Additionally, its solution is less costly and less conservative; however, its empirical reliability is lower, though still close to the desired reliability. With the new methodology developed, we then explore how the types and levels of uncertainty, reserve costs, and controllable load capacity affect the dispatch solution, operational costs, and CO_2 emissions. We find that different types and levels of uncertainty have significant impacts on the optimal dispatch and emissions. More controllable loads and less conservative solution methodologies lead to lower costs and emissions. The main content of this chapter is summarized in the following papers.

1. B. Li and J.L. Mathieu. Analytical reformulation of chance-constrained optimal power flow with uncertain load control. In *IEEE PowerTech*, Eindhoven, Netherlands, 2015.

2. B. Li, M. Vrakopoulou, and J.L. Mathieu. Chance constrained reserve scheduling using uncertain controllable loads part II: analytical reformulation. *IEEE Trans Smart Grid (in press)*, 2017.

3. B. Li, S.D. Maroukis, Y. Lin, and J.L. Mathieu. Impact of uncertainty from load-based reserves and renewables on dispatch costs and emissions. In *North American Power Symposium*, Denver, CO, 2016.

2.1 Introduction

As the grid penetration of fluctuating renewable energy sources increases, more reserves are needed to balance supply and demand. Thermal and hydropower plants typically provide balancing reserves to power systems, but aggregations of electric loads may be able to do so at lower cost and/or with less environmental impact [16]. Recent studies have developed methods to use commercial and residential loads to provide balancing reserves such as load following and frequency regulation, e.g., [63, 64]. Loads are coordinated to decrease and increase their consumption with respect to their baseline consumption to provide up and down balancing. Recent studies have also developed methods to dispatch controllable loads within the optimal power flow or unit commitment problem [46, 71].

To manage power system uncertainty stemming from fluctuating renewable power production and loads, we can formulate and solve stochastic optimal power flow problems. One method is to minimize the expected cost of operating the power system over a heuristically-chosen finite number of uncertainty scenarios, e.g., [72]. Another method is to formulate a chance constrained optimization problem in which constraints with random variables hold probabilistically, ensuring feasibility for a vast

majority of uncertainty scenarios, see e.g., [104, 95]. In [96] a chance constrained formulation to schedule the production levels and reserve capacities for generators in systems with uncertain renewable energy production is proposed. It is straightforward to extend this approach to additionally schedule the reserve capacities of aggregations of controllable loads if we assume the available reserve capacity is known. However, in practice, the available reserve capacity of an aggregation of loads is a function of random variables such as ambient conditions (e.g., the capacity of an aggregation of air conditioners is a function of outdoor temperature) and load usage patterns (e.g., the capacity of an aggregation of electric vehicles is a function of driving patterns) [61].

The objectives of this chapter are to i) present a multi-period chance constrained optimal power flow (CC-OPF) problem to schedule generator production and load consumption set points along with both generator and load-based reserve capacities assuming wind power and outdoor temperature forecast uncertainty, ii) solve the problem using several solution methods, and compare the results in terms of performance (objective cost, solution reliability) and computational requirements, and iii) determine how outdoor temperature forecast uncertainty, in conjunction with wind uncertainty, affects power system dispatch and system CO_2 emissions. We use the DC power flow approximation. As one solving approach, we solve the problem using the scenario-based method [58] that may require large numbers of uncertainty scenarios, but provides a priori guarantees on the probability of constraint violation, assuming no knowledge of the uncertainty distributions. As the other approach, we analytically reformulate the problem assuming that wind and outdoor temperature forecast uncertainty follow multivariate normal distributions and re-solve the problem, comparing the results against those of the scenario-based method. In reality, the errors may not follow these distributions; however, the assumption allows us to solve the problem faster and with less need for real data (i.e., uncertainty scenarios), at the cost of less

reliable solutions. This idea of analytical reformulation builds upon our preliminary work presented in [97, 51]. Other recent studies also solve CC-OPF problems using the same scenario-based method [95, 96] or via analytical reformulation [9, 79], but these studies do not model load-based reserves or their uncertainty. Ref. [59] extends the formulation in [97] to the unit commitment problem and solve it with the scenario-based and the percentile methods. Ref. [107] solves a single-period CC-OPF with uncertain renewable energy production and load-based reserve capacities using distributionally robust optimization.

The contributions of this chapter are four-fold. First, we reformulate the problem assuming the uncertainty is normally distributed, resulting in a deterministic convex nonlinear problem. Second, we solve the reformulated problem using both a nonlinear solver and an iterative cutting plane approach [9] that introduces linear approximations of the nonlinear constraints only when they are binding. Third, we compare the solutions against the scenario-based method in terms of performance and computational requirements. Fourth, we perform case studies to show how different types and levels of uncertainty, reserve costs, and controllable load capacity affect power system dispatch, operational costs, and CO_2 emissions.

The remainder of the chapter is organized as follows. Section 2.2 presents the detailed formulation of the multi-period CC-OPF problem and the conventional scenario-based method. Section 2.3 derives the analytical reformulation and presents the case study results against the scenario-based method. Section 2.4 shows the case study results on how system factors and uncertainties affect the optimal dispatch and CO_2 emissions.

2.2 Problem Formulation and Conventional Approaches

In this section, we provide the detailed formulation and introduce the scenario-based methodologies used in [93][1].

2.2.1 Modeling

In this section, we introduce the load models, reserve mechanisms and market setups we use in the problem formulation.

2.2.1.1 a. Uncertain controllable loads

In this work, we assume that aggregations of thermostatically controlled loads (TCLs) comprise a portion of the total system load. Each TCL is locally controlled to maintain its internal temperature within a narrow temperature range (e.g., 1°C) by switching on/off its power consumption. The TCLs are coordinated to shift their consumption in time (ensuring that the total amount of energy delivered over a specific longer time horizon is fixed) without violating their temperature constraints. We can model the aggregations of TCLs as thermal energy storage units [62]. We assume that the aggregator is able to broadcast control signals to all TCLs inducing on/off switching actions achieving a desired aggregate power consumption $P_{C,t}$. When the ambient conditions are constant, the energy state S_t of the aggregation evolves with time steps t of length $\Delta\tau$ as

$$S_{t+1} = S_t + (P_{C,t} - P_{B,t})\Delta\tau, \tag{2.1}$$

where $P_{B,t}$ is the baseline aggregate power consumption, i.e., the consumption that would have occurred without external scheduling. Actions which decrease (increase)

[1]These contents are from the Part I of the two-part journal [93, 52] with Maria as the primary contributor. The contents and discussions are included in this proposal so that it will help interpret the mechansim of the problem and help explain some derivations and ideas in Part II. For Part II which starts in Section 2.3, I am the primary contributor.

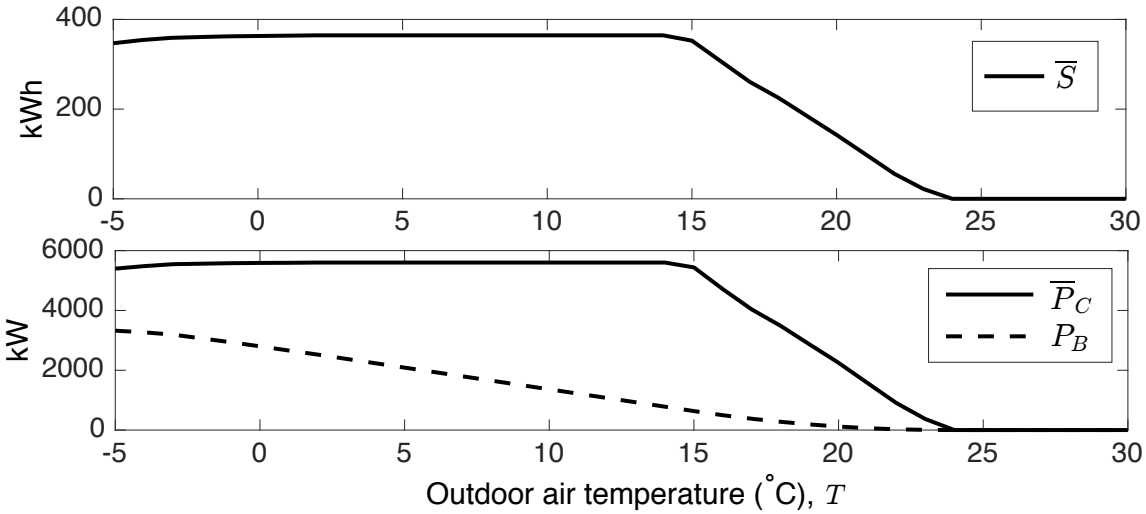


Figure 2.1: Energy capacity, power capacity, and baseline power consumption of an aggregation of electric heaters modeled as a thermal energy storage unit.

consumption relative to the baseline empty (charge) the storage unit.

A TCL aggregation's baseline consumption, power capacity (i.e., limits for $P_{C,t}$), and energy capacity (i.e., limits for S_t) are a function of a variety of time-dependent uncertain quantities such as ambient conditions and load usage patterns. Here, we assume outdoor air temperature T_t alone determines the baseline consumption $P_B(T_t)$, the power capacity $\overline{P}_C(T_t)$, and the energy capacity $\overline{S}(T_t)$, i.e.,

$$0 \leq P_{C,t} \leq \overline{P}_C(T_t),\ 0 \leq S_t \leq \overline{S}(T_t).$$

In [62], a method of computing $P_B(T_t)$, $\overline{P}_C(T_t)$, and $\overline{S}(T_t)$ for an aggregation of residential electric space heaters or air conditioners is described. Here, we use this method to compute these quantities for an aggregation of 1,000 heterogenous electric space heaters, and obtain the results shown in Fig. 2.1. We assume that the values in Fig. 2.1 are accurate for a given outdoor air temperature but that our forecasts of outdoor air temperature are uncertain. Hence Fig. 2.1 serves as a look-up table which maps an outdoor air temperature forecast to the baseline consumption, power capacity, and energy capacity of a thermal energy storage unit.

When ambient conditions are changing, the energy state equation (2.1) must be modified to take into account changes in the energy capacity. Specifically, for heating loads, when T_t increases, $\overline{S}$ may decrease (as shown in Fig. 2.1) because T_{t+1} is within or above some heaters' temperature range and those heaters are no longer available for control. We can assume that the aggregation of heaters that become unavailable in $t+1$ had approximately the same percent energy state $S_t/\overline{S}(T_t)$ as the aggregation of all heaters available in t. Therefore, when T_t increases, we assume the percent energy state at the beginning of time-step $t+1$ is the same as the percent energy state at the end of time step t. When T_t decreases, $\overline{S}(T_t)$ may increase because heaters that were previously unavailable become available. We assume that the heaters that become available have a 50% energy state. With these assumptions, the new energy state equation is

$$\begin{aligned} S_{t+1} = (S_t + (P_{C,t} - P_B(T_t))\Delta\tau) \min\left(\frac{\overline{S}(T_{t+1})}{\overline{S}(T_t)}, 1\right) \\ + 0.5 \max\left(\overline{S}(T_{t+1}) - \overline{S}(T_t), 0\right). \end{aligned} \tag{2.2}$$

2.2.1.2 b. Generation-load mismatch

We consider two types of uncertainty: uncertainty on the wind power forecast and uncertainty on the outdoor temperature forecast. If the system is operated based on the forecasts, forecast error may create generation-load mismatch, which should be compensated by reserves to maintain power balance. When the controllable loads are unscheduled, they consume the baseline power and hence the total generation-load mismatch is the sum of the total wind power forecast error and the total baseline power forecast error, i.e.,

$$P_{m,t} = \mathbf{1}^T(P_{W,t} - P_{W,t}^f) + \mathbf{1}^T(P_B(T_t) - P_B(T_t^f)), \tag{2.3}$$

where $P_{W,t}$ and $P^f_{W,t}$ are vectors including the actual and forecast wind power production of each wind power plant, respectively. Similarly, $P_B(T_t)$ and $P_B(T^f_t)$ are vectors including the actual and forecast baseline power consumption of each controllable load aggregation, respectively. The vector $\mathbf{1}$ is a unit vector with the same dimension as the vector that is multiplied with, and so $P_{m,t}$ is a scalar.

When the controllable loads are scheduled, their consumption is no longer uncertain and hence the generation-load mismatch is only the total wind power forecast error, i.e.,

$$P_{m,t} = \mathbf{1}^T (P_{W,t} - P^f_{W,t}). \tag{2.4}$$

In this case, as detailed later, both types of uncertainty will affect the energy state trajectory.

2.2.1.3 c. Reserve policies

We assume that secondary frequency control (i.e., automatic generation control) compensates power mismatches on the timescale of seconds to minutes and tertiary frequency control is applied every 5–15 minutes to redispatch the system economically. We refer to the former as "secondary reserves" and the latter as "redispatch reserves," which are comparable to real-time energy market actions in the U.S. Our CC-OPF determines the minimum-cost adequate capacities of secondary and redispatch reserves. We assume both generators and controllable loads can provide secondary reserves, but only generators provide redispatch reserves.

To determine the adequate secondary or redispatch reserve capacity, we need to model the operating point after a control action. In [95] the new generation set point P^{new}_G is modeled using an affine function of the power mismatch while [94] differentiates between positive and negative generation-load mismatch by using a

piecewise affine function. Following [97], we model P_G^{new} and the new controllable load set point P_C^{new} using piecewise affine functions, which gives us the policies

$$P_{G,t}^{new} = P_{G,t} + \overline{d}_{G,t} \max(-P_{m,t}, 0) - \underline{d}_{G,t} \max(P_{m,t}, 0), \tag{2.5}$$

$$P_{C,t}^{new} = P_{C,t} + \overline{d}_{L,t} \max(P_{m,t}, 0) - \underline{d}_{L,t} \max(-P_{m,t}, 0), \tag{2.6}$$

where $P_{G,t}$ and $P_{C,t}$ denote the generator and controllable load set points that maintain power balance for the forecasts $P_{W,t}^f$ and T_t^f. To achieve power balance under forecast error, a positive mismatch decreases the power production of generators and increases the power consumption of controllable loads as a function of the distribution vectors $d \geq 0$, where $\mathbf{1}^T \overline{d}_{G,t} + \mathbf{1}^T \underline{d}_{L,t} = 1$ and $\mathbf{1}^T \underline{d}_{G,t} + \mathbf{1}^T \overline{d}_{L,t} = 1$, and each d is treated as a decision variable. Adequate reserve capacities are determined by the amount that the generators (controllable loads) may need to increase/decrease their production (consumption) as a function of $P_{m,t}$. The exact constraints used within the optimization problem are defined in Section 2.2.2; for notational simplicity we define

$$f(d_1, d_2, x) = \; d_1 \max(-x, 0) - d_2 \max(x, 0). \tag{2.7}$$

2.2.1.4 d. Market setup

Our multi-period OPF dispatches resources hourly over a 24-hour horizon, where $\Delta\tau = 1$ hour. Generator and controllable load set points and all reserve capacities are constant within each hour. Secondary reserve may be activated at any point within the hour but redispatch is activated every $\Delta\tau/n$ minutes, where n is the number of intra-hour redispatch intervals. Redispatch compensates not only the generation-load mismatch but also unforecasted deviations in the energy state of the controllable loads, similar to the California Independent System Operator's Regulation Energy Man-

agement functionality [14]. Forecasted changes in the load energy state result from differences between $P_{C,t}$ and $P_B(T_t^f)$, as shown in (2.1). Unforecasted changes result from load-based secondary reserve actions used to compensate wind power forecast error and baseline error resulting from temperature forecast error. Hence, the actual change in load energy state in each time step will be $(P_{C,t} + R_t - P_B(T_t))\Delta\tau$, where R_t is the reserve action. When the system is redispatched, generator production and controllable load consumption are adjusted to compensate for unforecasted changes in the load energy state over the previous $\Delta\tau/n$ minutes. This ensures that the actual load energy state trajectory is closer to the forecasted one, and increases the ability of the controllable loads to continue providing reserves.

We assume temperature forecast errors, and so baseline power forecast errors, are constant within each hour whereas wind power forecast errors may appear at any time within each hour and may persist until the end of the hour. Figure 2.2 shows an example of how the forecast errors affect reserve actions. Dotted lines show the baseline power forecast error (top), the resulting load energy state deviation (bottom), the generators' redispatch action that stops the increase of this deviation (upper middle), and the loads' response (lower middle). The x-axis of the load energy state plot represents the forecasted energy trajectory. Continuous lines show the wind power forecast error (top), secondary reserve actions in the second time interval by the generators (upper middle) and the loads (lower middle) that compensate this error, the resulting load energy state deviation (bottom), and the generators' redispatch actions in third and fourth intervals that compensate the wind power forecast error and this load energy state deviation (upper middle). A third error that affects hourly decisions is any remaining load energy state deviation from the previous hour resulting from baseline power forecast error in the previous hour and/or secondary reserve actions in the last $\Delta\tau/n$-minute interval of the previous hour. The responses to this error are shown with dashed lines. We show the responses to each of the three errors

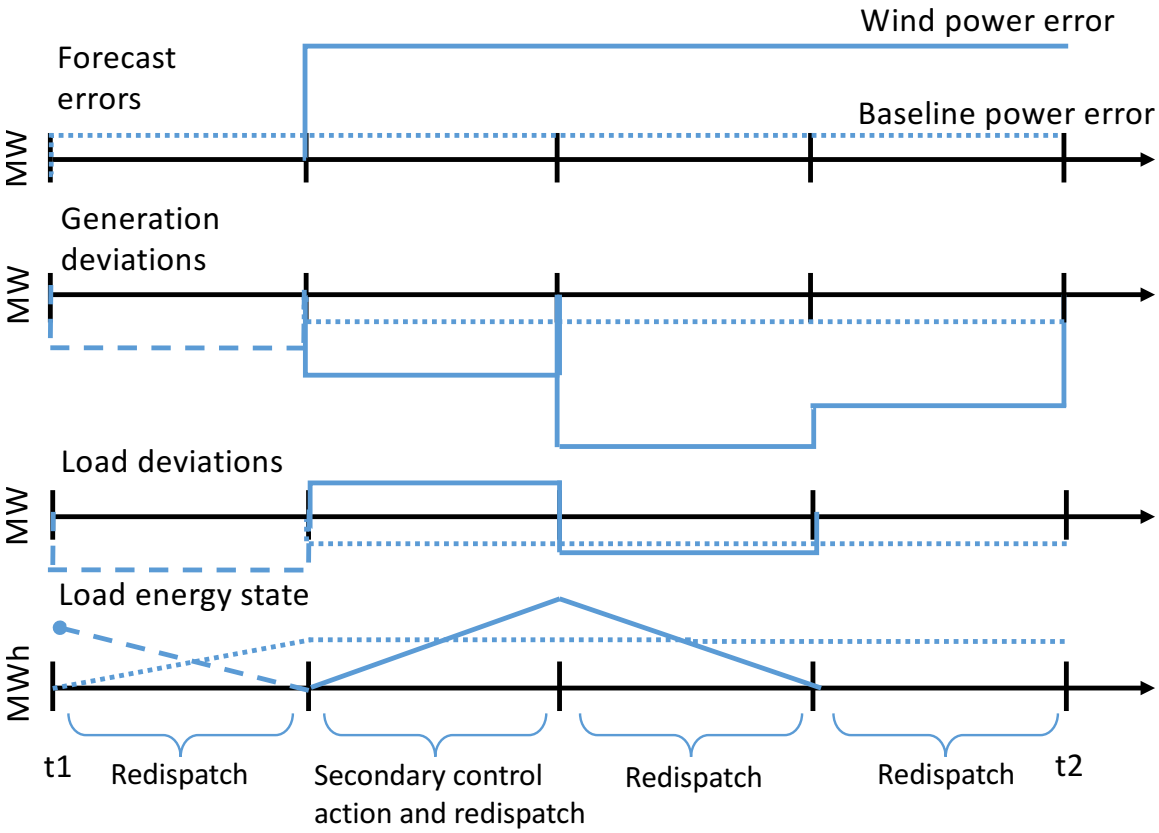


Figure 2.2: An example of how forecast errors activate secondary and redispatch reserves and influence the load energy state ($n = 4$). Dotted lines correspond to responses to baseline power forecast error, continuous lines to wind power forecast error, and dashed lines to the remaining load energy state deviation from the previous hour.

independently for clarity; in reality all three errors will occur simultaneously and the required generation/load deviations will be the sum of those shown.

Given our assumptions, we only need to check that the system constraints are satisfied at three intra-hour operating points, i.e., if the constraints are satisfied at these points they will be satisfied at all intra-hour operating points.

Operating Point 1 corresponds to the end of any $\Delta\tau/n$-minute interval, except the first, in which generators provide redispatch to compensate load energy state deviations due to the baseline power forecast error of the previous $\Delta\tau/n$-minute interval and also provide secondary reserve actions. In Fig. 2.2, this point is reached at the 30^{th} minute.

Operating Point 2 corresponds to the end of any $\Delta\tau/n$-minute interval, except the first, in which generators provide redispatch to compensate the wind power forecast error, load energy state deviations due to baseline power forecast error, and load-based secondary reserve actions of the previous $\Delta\tau/n$-minute interval, but do not provide secondary reserve actions. In Fig. 2.2, this point is reached at the 45^{th} minute.

Operating Point 3 corresponds to the end of the first $\Delta\tau/n$-minute interval, in which generators provide redispatch to compensate the remaining load energy state deviation from the previous hour and may also provide secondary reserves actions. In Fig. 2.2, this point is reached at the 15^{th} minute.

2.2.2 CC-OPF Formulation

In this section, we present a CC-OPF formulation that co-optimizes energy and reserves provided both by generators and controllable loads under the market setup described in Section 2.2.1.4 taking into account wind power and temperature forecast uncertainty. The objective is to find the generation/controllable load set points (i.e., the dispatch), reserve capacities, and distribution vectors that minimize the energy, secondary reserve, and redispatch costs such that system constraints are satisfied in a probabilistic sense.

2.2.2.1 Notation & reserve constraints

We use an optimization horizon $N_t = 24$ hours with hourly steps t. Each load is comprised of an uncontrollable portion $P_{L,t}$, which is assumed known and constant over a time step t, and a controllable portion $P_{C,t}$. For each step t, we define the vector of decision variables as $x_t = \langle P_t, d_t, \mathcal{R}_t \rangle$, where we use angle brackets to stack column vectors into a single column vector (i.e., $\langle \alpha, \beta \rangle = [\alpha^T, \beta^T]^T$) and P_t contains

the generator and load dispatch $\langle P_{G,t}, P_{C,t}\rangle$, d_t the distribution vectors

$$\langle \underline{d}_{GS,t}, \overline{d}_{GS,t}, \underline{d}_{LS,t}, \overline{d}_{LS,t}, \underline{d}_{GD,t}, \overline{d}_{GD,t}\rangle$$

and $\mathcal{R}_t$ the reserve capacities

$$\langle \underline{R}_{GS,t}, \overline{R}_{GS,t}, \underline{R}_{LS,t}, \overline{R}_{LS,t}, \underline{R}_{GD,t}, \overline{R}_{GD,t}\rangle.$$

The subscripts GS/LS correspond to generation/load secondary reserves and GD to generation redispatch reserves. Vector $\underline{d}_{GD,t} = \langle \underline{d}^{w}_{GD,t}, \underline{d}^{b}_{GD,t}, \underline{d}^{w_o}_{GD,t}, \underline{d}^{b_o}_{GD,t}\rangle$ and $\overline{d}_{GD,t} = \langle \overline{d}^{w}_{GD,t}, \overline{d}^{b}_{GD,t}, \overline{d}^{w_o}_{GD,t}, \overline{d}^{b_o}_{GD,t}\rangle$, where the component vectors correspond to redispatch actions initiated for different reasons, as described below. We use the subscript LD to denote variables related to load adjustments initiated by a redispatch. We denote the baseline power forecast error by $\Delta P_{B,t} = P_B(T_t) - P_B(T_t^f)$ and the total baseline power deviation by $P^b_{m,t} = \mathbf{1}^T \Delta P_{B,t}$. Since we choose $P_{C,t}$ (and the uncontrollable load is assumed known), the generation-load mismatch $P_{m,t}$ is given by (2.4). We next describe the four types of reserve actions.

1. Secondary reserves due to wind power forecast error Wind power forecast error activates secondary reserves and $P_{m,t}$ is distributed to generators and controllable loads by shifting their power injection by $R_{GS,t}$ and $R_{LS,t}$, respectively. Using the piecewise linear policy defined in Section 2.2.1.3, the constraints are

$$R_{GS,t} = f(\overline{d}_{GS,t}, \underline{d}_{GS,t}, P_{m,t}), \tag{2.8}$$

$$R_{LS,t} = f(\overline{d}_{LS,t}, \underline{d}_{LS,t}, -P_{m,t}), \tag{2.9}$$

$$\mathbf{1}^T \overline{d}_{GS,t} + \mathbf{1}^T \underline{d}_{LS,t} = 1, \tag{2.10}$$

$$\mathbf{1}^T \underline{d}_{GS,t} + \mathbf{1}^T \overline{d}_{LS,t} = 1. \tag{2.11}$$

2. Redispatch compensating secondary reserve actions After the secondary reserves have achieved power balance, redispatch redistributes the power mismatch $P_{m,t}$ to the generators and compensates unforecasted energy state deviations due to prior load-based secondary reserve actions. Let the power injection shift for the generators be $R^w_{GD,t}$ and for the loads be $R^w_{LD,t}$, which should be of the same magnitude and opposite sign as $R_{LS,t}$. To maintain power balance, the distribution vectors should sum to one plus the portion of the mismatch compensated by the loads. The constraints are

$$R^w_{GD,t} = f(\overline{d}^w_{GD,t}, \underline{d}^w_{GD,t}, P_{m,t}), \tag{2.12}$$

$$R^w_{LD,t} = -R_{LS,t}, \tag{2.13}$$

$$\mathbf{1}^T \overline{d}^w_{GD,t} = 1 + \mathbf{1}^T \underline{d}_{LS,t}, \tag{2.14}$$

$$\mathbf{1}^T \underline{d}^w_{GD,t} = 1 + \mathbf{1}^T \overline{d}_{LS,t}. \tag{2.15}$$

3. Redispatch due to baseline power forecast error Baseline power forecast error results in unforecasted energy state deviations, which are compensated by redispatch. Let the power injection shift for the generators be $R^b_{GD,t}$ and for the loads be $R^b_{LD,t}$. The generators should compensate $P^b_{m,t}$, while the loads should shift by $\Delta P_{B,t}$. The constraints are

$$R^b_{GD,t} = f(\overline{d}^b_{GD,t}, \underline{d}^b_{GD,t}, -P^b_{m,t}), \tag{2.16}$$

$$R^b_{LD,t} = \Delta P_{B,t}, \tag{2.17}$$

$$\mathbf{1}^T \overline{d}^b_{GD,t} = 1, \tag{2.18}$$

$$\mathbf{1}^T \underline{d}^b_{GD,t} = 1. \tag{2.19}$$

4. Redispatch at the beginning of an hour Redispatch is activated in the first $\Delta\tau/n$-minute interval to compensate energy state deviations from the previous hour

$t-1$. The constraints, similar to those above, are

$$R^{w_o}_{GD,t} = f(\overline{d}^{w_o}_{GD,t}, \underline{d}^{w_o}_{GD,t}, P_{m,t-1}), \tag{2.20}$$

$$R^{b_o}_{GD,t} = f(\overline{d}^{b_o}_{GD,t}, \underline{d}^{b_o}_{GD,t}, -P^{b}_{m,t-1}), \tag{2.21}$$

$$R^{w_o}_{LD,t} = -R_{LS,t-1} \tag{2.22}$$

$$R^{b_o}_{LD,t} = \Delta P_{B,t-1} \tag{2.23}$$

$$\mathbf{1}^T \overline{d}^{w_o}_{GD,t} = \mathbf{1}^T \underline{d}_{LS,t-1}, \tag{2.24}$$

$$\mathbf{1}^T \underline{d}^{w_o}_{GD,t} = \mathbf{1}^T \overline{d}_{LS,t-1}, \tag{2.25}$$

$$\mathbf{1}^T \overline{d}^{b_o}_{GD,t} = 1, \tag{2.26}$$

$$\mathbf{1}^T \underline{d}^{b_o}_{GD,t} = 1. \tag{2.27}$$

2.2.2.2 Optimization problem

Let c_1 and c_2 be generation cost vectors and c_R be the reserve cost vector. The optimization problem is

$$\min_{\{x_t\}_{t=1}^{N_t}} \sum_{t=1}^{N_t} \Big(P_t^T [c_2] P_t + c_1^T P_t + c_R^T R_t \Big), \tag{2.28}$$

subject to deterministic constraints and probabilistic constraints corresponding to the three points described in Section 2.2.1.4, which we describe in the following sections. All constraints should be satisfied for all $t = 1, \ldots, N_t$.

2.2.2.3 Deterministic constraints

$$\mathbf{1}^T P_{inj,t} = 0, \tag{2.29}$$

$$-P_l \leq AP_{inj,t} \leq P_l, \tag{2.30}$$

$$\underline{P}_G \leq P_{G,t} \leq \overline{P}_G, \tag{2.31}$$

$$\underline{P}_C(T_t^f) \leq P_{C,t} \leq \overline{P}_C(T_t^f), \tag{2.32}$$

$$0 \leq S_t \leq \overline{S}(T_t^f), \tag{2.33}$$

$$0 \leq S_t + (P_{C,t} - P_B(T_t^f))\Delta\tau \leq \overline{S}(T_t^f), \tag{2.34}$$

$$S_{t+1} = \left(S_t + (P_{C,t} - P_B(T_t^f))\Delta\tau\right) \min\left(\frac{\overline{S}(T_{t+1}^f)}{\overline{S}(T_t^f)}, 1\right) + 0.5 \max\left(\overline{S}(T_{t+1}^f) - \overline{S}(T_t^f), 0\right), \tag{2.35}$$

$$S_1 = 0.5\overline{S}(T_1^f),\ S_{N_t+1} = 0.5\overline{S}(T_{N_t+1}^f), \tag{2.36}$$

where $P_{inj,t} = C_G P_{G,t} + C_W P_{W,t}^f - C_L(P_{L,t} + P_{C,t})$ and the C matrices map the generators, wind power plants, and loads to the buses. Constraint (2.29) enforces power balance given the wind power forecast; (2.30), (2.31), and (2.32) encode the line, generation, and controllable load capacity limits, respectively, where A is a constant matrix that depends on the network impedances [95]; (2.33) and (2.34) ensure that the load energy state is within its limits at the beginning and end of each hour; (2.35) specifies the evolution of the load energy state; and (2.36) sets the load energy state at the beginning and end of the day to 50% of its maximum capacity (which corresponds to baseline operation). Since (2.35) is linear, (2.33) and (2.34) ensure that the load energy state is within the limits through out each hour.

2.2.2.4 Probabilistic constraints

At *Operating Point 1* generator production and controllable load consumption are

$$P_{G,t}^{new} = P_{G,t} + R_{GS,t} + R_{GD,t}^{b},$$
$$P_{C,t}^{new} = P_{C,t} + R_{LS,t} + R_{LD,t}^{b},$$

where the reserve shifts are given in (2.8), (2.9), (2.16), and (2.17). The new power injection is

$$P_{inj,t}^{new} = C_G P_{G,t}^{new} + C_W P_{W,t} - C_L(P_{L,t} + P_{C,t}^{new}) \tag{2.37}$$

and the constraints that must be enforced are

$$-P_l \leq AP_{inj,t}^{new} \leq P_l, \tag{2.38}$$
$$\underline{P}_G \leq P_{G,t}^{new} \leq \overline{P}_G, \tag{2.39}$$
$$0 \leq P_{C,t}^{new} \leq \overline{P}_C(T_t), \tag{2.40}$$
$$-\underline{R}_{GS,t} \leq R_{GS,t} \leq \overline{R}_{GS,t}, \tag{2.41}$$
$$-\underline{R}_{LS,t} \leq R_{LS,t} \leq \overline{R}_{LS,t}, \tag{2.42}$$
$$-\underline{R}_{GD,t} \leq R_{GD,t}^{b} \leq \overline{R}_{GD,t}, \tag{2.43}$$
$$\text{constraints } (2.10), (2.11), (2.18), (2.19),$$

where (2.41)–(2.43) determine the generator secondary reserve capacity, load secondary reserve capacity, and generator redispatch capacity, respectively. Note that load deviations due to redispatch actions (i.e., $R_{LD,t}^{b/w/w_o}$) are not considered reserve actions and so are not financially rewarded (i.e., $R_{LD,t}^{b/w/w_o}$ does not affect the load secondary reserve capacity).

At *Operating Point 2* generator production and controllable load consumption are

$$P_{G,t}^{new} = P_{G,t} + R_{GD,t}^{w} + R_{GD,t}^{b},$$
$$P_{C,t}^{new} = P_{C,t} + R_{LD,t}^{w} + R_{LD,t}^{b},$$

where the reserve shifts are given in (2.12), (2.13), (2.16), and (2.17). The new power injection is (2.37) and constraints that must be enforced are

$$\begin{aligned} &\text{constraints } (2.14), (2.15), (2.18), (2.19), (2.38) - (2.40), \\ &-\underline{R}_{GD,t} \leq R_{GD,t}^{w} + R_{GD,t}^{b} \leq \overline{R}_{GD,t}. \end{aligned} \tag{2.44}$$

At *Operating Point 3* generator production and controllable load consumption are

$$P_{G,t}^{new} = P_{G,t} + R_{GS,t} + R_{GD,t}^{w_o} + R_{GD,t}^{b_o},$$
$$P_{C,t}^{new} = P_{C,t} + R_{LS,t} + R_{LD,t}^{w_o} + R_{LD,t}^{b_o},$$

where the reserve shifts are given in (2.8), (2.9), (2.20) – (2.23). The new power injection is (2.37) and constraints that must be enforced are

$$\begin{aligned} &\text{constraints } (2.10), (2.11), (2.24) - (2.27), (2.38) - (2.42), \\ &-\underline{R}_{GD,t} \leq R_{GD,t}^{w_o} + R_{GD,t}^{b_o} \leq \overline{R}_{GD,t}. \end{aligned} \tag{2.45}$$

We also need to ensure that the controllable load energy capacity limits are satisfied. The following constraints are sufficient to ensure that the load energy state remains within its limits within $[t, t+1]$. Specifically, since the energy state dynamics are linear, the energy capacity limits are satisfied within each interval of the hour, if

they are satisfied at the end of the first and last interval of the hour, i.e.,

$$0 \leq S_t + (P_{C,t} + R_{LS,t} - P_B(T_t))\frac{\Delta\tau}{n} \leq \overline{S}(T_t), \tag{2.46}$$

$$0 \leq S_t + (P_{C,t} - P_B(T_t^f))\frac{(n-1)\Delta\tau}{n} + (P_{C,t} + R_{LS,t} - P_B(T_t))\frac{\Delta\tau}{n} \leq \overline{S}(T_t). \tag{2.47}$$

2.2.3 Scenario-based Solution Methodology

We solve the optimization problem using a method that does not require assumptions on the probability distributions of the uncertainty and provides guarantees on the probability of constraint satisfaction. For simplicity, define x to be a stacked version of $\{x_t\}_{t=1}^{N_t}$ and let $\delta_t \in \mathbb{R}^{N_W+N_T}$ denote the uncertainty vector in timestep t, where N_W is the number of wind power plants, i.e., $P_{W,t} \in \mathbb{R}^{N_W}$, and N_T is the number of temperature zones, i.e., $T_t \in \mathbb{R}^{N_T}$. We require constraints that are affected by δ_t to be satisfied with probability of at least $1 - \varepsilon_t$, where ε_t is the violation level. Then, the optimization problem can be formulated as a quadratic program with multiple chance constraints. For each $t = 2, \ldots, N_t$, the chance constraints can be written compactly as

$$\mathbb{P}\left(H_t(\delta_t, \delta_{t-1})x + h_t + g_t(\delta_t) \leq 0\right) \geq 1 - \varepsilon_t.$$

For $t = 1$, the chance constraint is similar except that H_1 depends only on δ_1.

To deal with the issue of multiple chance constraints we use a scenario-based optimization method [58, 106] that is a mixture of randomized and robust optimization. The method includes two steps. In the first step, for each $t = 1, \ldots, N_t$, the scenario approach [17] is used to determine, with a confidence of at least $1 - \beta_t$, the minimum volume set that contains at least $1 - \varepsilon_t$ probability mass of the uncertainty. Let this set be denoted by Δ_t. To compute this set, the number of scenarios we need to use

is given by

$$N_{s,t} \leq \frac{1}{\varepsilon_t} \frac{e}{e-1} \Big(\ln \frac{1}{\beta_t} + 4(N_W + N_T) - 1 \Big).$$

In the second step, we use Δ_t to formulate a robust problem where the uncertainty is confined in this set. For each $t = 2, \ldots, N_t$, the chance constraint is replaced by the following robust constraint

$$H_t(\delta_t, \delta_{t-1})x + h_t + g_t(\delta_t) \leq 0, \ \forall (\delta_t, \delta_{t-1}) \in \Delta_t \times \Delta_{t-1}.$$

For $t = 1$, the constraint is similar except that H_1 depends only on δ_1 and we require the constraint to be satisfied for all $\delta_1 \in \Delta_1$. Any feasible solution satisfying the robust constraints will be feasible for the chance constraints with a confidence of at least $1 - \beta_t$. To solve the resulting convex problem, standard techniques for robust optimization can be employed [8].

2.3 Analytical Reformulation Assuming Gaussian Distribution

In this section, we develop an analytical reformulation to the problem in Section 2.2 assuming that wind power forecast error and temperature forecast error follow multivariate normal distributions. In order to get tractable reformulation, we assume symmetric reserve deployment since asymmetric reserve deployment may result in a non-convex problem. We further assume that wind power forecast errors are correlated and temperature forecast errors are correlated, but wind power and temperature forecast errors are independent.

2.3.0.1 Notation

New uncertainty notation is defined in this section. The wind power forecast error vector is denoted $\Delta P_{W,t} \in \mathbb{R}^{N_W}$, with $\mu_{W,t} \in \mathbb{R}^{N_W}$ as its mean, $\Sigma_{W,t} \in \mathbb{R}^{N_W \times N_W}$ as its covariance, and $\delta_{W,t} \in \mathbb{R}^{N_W}$ as its standard deviation. The mean and standard deviation of the total wind power forecast error $P_{m,t} = \mathbf{1}^T \Delta P_{W,t} \in \mathbb{R}$ are denoted $\overline{\mu}_{W,t} \in \mathbb{R}$ and $\overline{\delta}_{W,t} \in \mathbb{R}$, respectively. The vector of temperature forecast errors corresponding to each controllable load aggregation is denoted $\Delta T_t \in \mathbb{R}^{N_C}$, where N_C is the number of controllable load aggregations. $\Delta P_{B,t} \in \mathbb{R}^{N_C}$ is the baseline power forecast error. Let $\mu_{B,t} \in \mathbb{R}^{N_C}$ be its mean, $\Sigma_{B,t} \in \mathbb{R}^{N_C \times N_C}$ be its covariance, and $\delta_{B,t} \in \mathbb{R}^{N_C}$ be its standard deviation. The mean and standard deviation of the total baseline power forecast error $P^b_{m,t} = \mathbf{1}^T \Delta P_{B,t} \in \mathbb{R}$ are denoted $\overline{\mu}_{B,t} \in \mathbb{R}$ and $\overline{\delta}_{B,t} \in \mathbb{R}$, respectively.

To simplify the equations that follow, vector operators (e.g., $\max(u)$, $\min(u)$, uv, u^2, $\sqrt{u}$, where u and v are arbitrary vectors) are applied element-wise.

2.3.0.2 Approximate controllable load capacities and baseline

To facilitate the analytical reformulation, we approximate the relationships shown in Fig. 2.1. Specifically, we assume $\overline{P}_C$ and $\overline{S}$ are piecewise linear in T_t, and $\Delta P_{B,t}$ is linear in ΔT_t:

$$\overline{P}_C(T_t) = C_1 + \min(0, k_p(T_t - T_{br})), \tag{2.48}$$

$$\overline{S}(T_t) = C_2 + \min(0, k_e(T_t - T_{br})), \tag{2.49}$$

$$\Delta P_{B,t} = a \Delta T_t, \tag{2.50}$$

where the slopes k_p, k_e, and a; intercepts C_1 and C_2; and "breaking temperature" T_{br} are shown in Fig. 2.3. The slopes k_p and k_e are computed using linear regression on the data between T_{br} and the "ending temperature" T_{end}, and a is computed

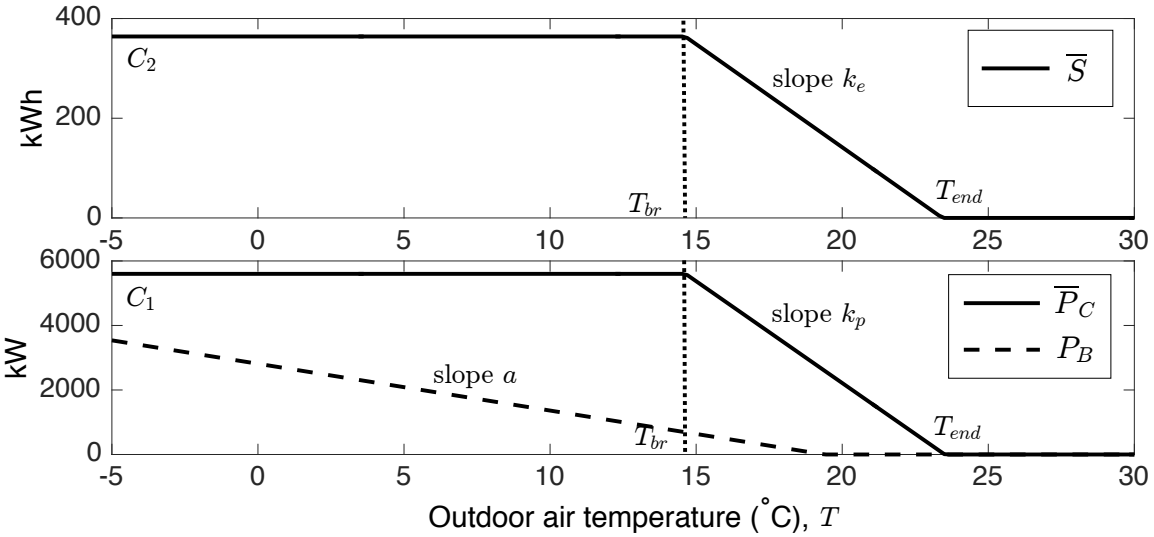


Figure 2.3: Approximate energy capacity, power capacity, and baseline power consumption of an aggregation of electric heaters.

using linear regression on the data within the temperature forecast range. We do not approximate the relationship between the baseline power and temperature forecast $P_B(T_t^f)$ but use Fig. 2.1 as a look-up table.

2.3.1 Analytical Reformulation

In this section, we analytically reformulate (2.38) – (2.47). A key difference between the formulation in Section 2.2 and the analytical reformulation is that the former satisfies chance constraints *jointly*, i.e., all constraints within an hour are satisfied with probability $1 - \epsilon$, while the latter satisfies chance constraints *individually*, i.e., each constraint is satisfied with probability $1 - \epsilon$.

2.3.1.1 General Reformulation

Assume that a random vector ξ follows a multivariate normal distribution with mean μ and covariance Σ. Then, the following constraints are equivalent [12]:

$$\mathbb{P}\left(a(x)^T\xi + b(x) \leq 0\right) \geq 1 - \epsilon, \tag{2.51}$$

$$a(x)^T\mu + b(x) + c\sqrt{a(x)^T\Sigma a(x)} \leq 0, \tag{2.52}$$

where $a(x)$ and $b(x)$ are functions of the decision variable x, and $c = \Phi_N^{-1}(1-\epsilon)$, where Φ_N denotes the cumulative distribution function (CDF) of the standard normal distribution. When $a(x)$ and $b(x)$ are affine functions of x and $\epsilon \leq 0.5$, the reformulation (2.52) is convex.

Constraint (2.52) can also be written as

$$a(x)^T\mu + b(x) + \mathcal{S} \leq 0, \tag{2.53}$$

$$\mathcal{S} \geq c\sqrt{a(x)^T\Sigma a(x)}, \tag{2.54}$$

where $\mathcal{S}$ is a slack variable and (2.54) is a second-order cone (SOC) constraint. We will reformulate the constraints into this form since it will allow us to apply a computationally-efficient cutting plane algorithm, described in Section 2.3.2.3.

2.3.1.2 Specific Reformulation

Generation constraints (2.39) for *Operating Point 1* are reformulated as

$$P_{G,t} - d_{GS,t}\overline{\mu}_{W,t} + d^b_{GD,t}\overline{\mu}_{B,t} + \mathcal{S}_{GS,t} \leq \overline{P}_G, \tag{2.55}$$

$$P_{G,t} - d_{GS,t}\overline{\mu}_{W,t} + d^b_{GD,t}\overline{\mu}_{B,t} - \mathcal{S}_{GS,t} \geq \underline{P}_G, \tag{2.56}$$

$$\mathcal{S}_{GS,t} \geq c\sqrt{(d_{GS,t}\overline{\delta}_{W,t})^2 + (d^b_{GD,t}\overline{\delta}_{B,t})^2}, \tag{2.57}$$

where $\mathcal{S}_{GS,t}$ is a slack variable. Constraints (2.39) for *Operating Points 2* and 3 can be reformulated similarly.

Next, we reformulate the controllable load power capacity constraints. The lower constraint of (2.40) for *Operating Point 1* is reformulated as

$$P_{C,t} + d_{LS,t}\overline{\mu}_{W,t} + \mu_{B,t} - \mathcal{S}_{LS,t} \geq 0, \tag{2.58}$$

$$\mathcal{S}_{LS,t} \geq c\sqrt{(d_{LS,t}\overline{\delta}_{W,t})^2 + (\delta_{B,t})^2}, \tag{2.59}$$

where $\mathcal{S}_{LS,t}$ is a slack variable. The lower constraints of (2.40) for *Operating Points 2* and 3 can be reformulated similarly.

The upper constraint of (2.40) is more difficult to reformulate. Recall that the upper constraint for (2.40) for *Operating Point 1* is

$$P_{C,t} + R_{LS,t} + R^b_{LD,t} \leq \overline{P}_C(T_t), \tag{2.60}$$

which can be rewritten as

$$P_{C,t} + Z \leq \overline{P}_C(T^f_t), \tag{2.61}$$

where Z contains the random variables, i.e., $Z = R_{LS,t} + \Delta P_{B,t} - e_{\overline{P}_C,t}$, where $e_{\overline{P}_C,t} = \overline{P}_C(T_t) - \overline{P}_C(T^f_t)$. We can rewrite Z in terms of $P_{m,t}$ and ΔT_t, by considering two cases:

Case 1, $T^f_t < T_{br}$ Referring to Fig. 2.3, denote the positive difference between the forecast and the breaking temperature as $e_{T,t} = T_{br} - T^f_t$. Then, $e_{\overline{P}_C,t} = \min(0, k_p(\Delta T_t - e_{T,t}))$ and $Z = d_{LS,t}P_{m,t} + \max(a\Delta T_t, (a-k_p)\Delta T_t + k_p e_{T,t})$.

Case 2, $T_{br} < T^f_t < T_{end}$ Referring to Fig. 2.3, $e_{\overline{P}_C,t} = \min\left(C_1 - \overline{P}_c(T^f_t), k_p\Delta T_t\right)$ and $Z = d_{LS,t}P_{m,t} + \max\left(a\Delta T_t + \overline{P}_c(T^f_t) - C_1, (a-k_p)\Delta T_t\right)$.

Both expressions for Z and the corresponding expressions for *Operating Point 2*

are of the form

$$Z = \max(k_1 X, k_2 X + h) + dY = M + dY, \tag{2.62}$$

where $X = \Delta T_t$, $Y = P_{m,t}$, $d = d_{LS,t}$, and k_1, k_2, and h are constants. In the first part of the Section 2.3.4, we show how to empirically compute a confidence bound z_1 for Z, such that $\mathbb{P}\left(Z \leq z_1(d_{LS,t})\right) \geq 1 - \epsilon$, which allows us to reformulate the upper constraint of (2.40) as

$$P_{C,t} + z_1(d_{LS,t}) \leq \overline{P}_C(T_t^f), \tag{2.63}$$

which is nonlinear but convex as shown in the second part of Section 2.3.4.

Unfortunately, the corresponding expressions for *Operating Point 3* are not of the form (2.62), specifically, $Z = R_{LS,t} + R_{LD,t}^{w_0} + R_{LD,t}^{b_0} - e_{\overline{P}_C,t} = d_{LS,t} P_{m,t} - d_{LS,t-1} P_{m,t-1} + a\Delta T_{t-1} - e_{\overline{P}_C,t}(\Delta T_t)$, which includes four random variables. While it would be possible to analytically reformulate the corresponding constraint, we instead approximate it as

$$P_{C,t} + d_{LS,t}\widetilde{P}_{m,t} + R_{LD,t}^{w_0} + R_{LD,t}^{b_0} \leq \overline{P}_C(T_t^f + \Delta\widetilde{T}_t), \tag{2.64}$$

where $\widetilde{P}_{m,t}$ uses the wind power forecast error statistics from $t-1$ rather than t and $\Delta\widetilde{T}_t = \Delta T_{t-1}$. The approximate constraint can be reformulated as (2.63), where its confidence bound can be computed using the method described in the supporting derivation. The approximation is physically justified because wind power forecast statistics and temperature forecast errors should be similar between hours. We demonstrate the impact of the approximation through case studies in Section 2.3.3.

For reserve capacity, constraints (2.41) – (2.43) are reformulated as

$$\overline{R}_{GS,t} \geq -d_{GS,t}\overline{\mu}_{W,t} + cd_{GS,t}\overline{\delta}_{W,t}, \tag{2.65}$$

$$\underline{R}_{GS,t} \geq d_{GS,t}\overline{\mu}_{W,t} + cd_{GS,t}\overline{\delta}_{W,t}, \tag{2.66}$$

$$\overline{R}_{LS,t} \geq d_{LS,t}\overline{\mu}_{W,t} + cd_{LS,t}\overline{\delta}_{W,t}, \tag{2.67}$$

$$\underline{R}_{LS,t} \geq -d_{LS,t}\overline{\mu}_{W,t} + cd_{LS,t}\overline{\delta}_{W,t}, \tag{2.68}$$

$$\overline{R}_{GD,t} \geq d^b_{GD,t}\overline{\mu}_{B,t} + cd^b_{GD,t}\overline{\delta}_{B,t}, \tag{2.69}$$

$$\underline{R}_{GD,t} \geq -d^b_{GD,t}\overline{\mu}_{B,t} + cd^b_{GS,t}\overline{\delta}_{B,t}. \tag{2.70}$$

Constraints (2.44) and (2.45) can be reformulated similarly.

For controllable load energy capacity constraints, the lower constraints of (2.46) and (2.47) are reformulated as

$$S_t + \left(P_{C,t} - P_B(T^f_t) + d_{LS,t}\overline{\mu}_{W,t} - \mu_{B,t}\right)\frac{\Delta\tau}{n} - \mathcal{S}_{EC,t} \geq 0, \tag{2.71}$$

$$S_t + \left(P_{C,t} - P_B(T^f_t)\right)\Delta\tau + \left(d_{LS,t}\overline{\mu}_{W,t} - \mu_{B,t}\right)\frac{\Delta\tau}{n} - \mathcal{S}_{EC,t} \geq 0, \tag{2.72}$$

$$\mathcal{S}_{EC,t} \geq \frac{c\Delta\tau}{n}\sqrt{(d_{LS,t}\overline{\delta}_{W,t})^2 + (\delta_{B,t})^2}, \tag{2.73}$$

where $\mathcal{S}_{EC,t}$ is a slack variable.

Like the upper power capacity constraints, the upper energy capacity constraints are more difficult to reformulate. The portion of the constraints that contains the random variables is $Z = (R_{LS,t} - \Delta P_{B,t})\Delta\tau/n - e_{\overline{S},t}$, where $e_{\overline{S},t} = \overline{S}(T_t) - \overline{S}(T^f_t)$. As before, we can rewrite Z in terms of $P_{m,t}$ and ΔT_t, by considering two cases:

Case 1, $T^f_t < T_{br}$ Referring to Fig. 2.3, $e_{\overline{S},t} = \min\left(0, k_e(\Delta T_t - e_{T,t})\right)$ and

$$Z = d_{LS,t}P_{m,t}\frac{\Delta\tau}{n} + \max\left(- a\Delta T_t\frac{\Delta\tau}{n} - k_e(\Delta T_t - e_{T,t}), -a\Delta T_t\frac{\Delta\tau}{n}\right).$$

Case 2, $T_{br} < T_t^f < T_{end}$ Referring to Fig. 2.3, $e_{\overline{S},t} = \min\Big(C_2 - \overline{S}(T_t^f), k_e \Delta T_t\Big)$ and

$$Z = d_{LS,t} P_{m,t} \frac{\Delta\tau}{n} + \max\bigg(- a\Delta T_t \frac{\Delta\tau}{n} - C_2 + \overline{S}(T_t^f), -a\Delta T_t \frac{\Delta\tau}{n} - k_e \Delta T_t \bigg).$$

Again, both expressions for Z are of the form (2.62) (though now Y is a fraction of $P_{m,t}$) and so we can reformulate the constraints as

$$S_t + \Big(P_{C,t} - P_B(T_t^f)\Big) \frac{\Delta\tau}{n} + z_2(d_{LS,t}) \leq \overline{S}(T_t^f), \tag{2.74}$$

$$S_t + \Big(P_{C,t} - P_B(T_t^f)\Big) \Delta\tau + z_2(d_{LS,t}) \leq \overline{S}(T_t^f), \tag{2.75}$$

where z_2 is the confidence bound computed using the method described in the first part of Section 2.3.4. These constraints are nonlinear but convex as shown in the second part of Section 2.3.4.

Power flow constraints (2.38) include all wind power and temperature forecast errors. First, we express (2.37) as $P_{inj,t}^{new} = P_{inj,t}^f + P_{inj,t}^w + P_{inj,t}^b$ where, for *Operating Point 1*,

$$P_{inj,t}^f = C_G P_{G,t} + C_W P_{W,t}^f - C_L(P_{L,t} + P_{C,t}), \tag{2.76}$$

$$P_{inj,t}^w = -C_G d_{GS,t} P_{m,t} + C_W \Delta P_{W,t} - C_L d_{LS,t} P_{m,t}, \tag{2.77}$$

$$P_{inj,t}^b = C_G d_{GD,t}^b P_{m,t}^b - C_L \Delta P_{B,t}. \tag{2.78}$$

Next, we define the power flows $P_t = A P_{inj,t}^{new} = P_t^f + P_t^w + P_t^b$ where, for *Operating*

Point 1,

$$P_t^f = A_G P_{G,t} + A_W P_{W,t}^f - A_L(P_{L,t} + P_{C,t}), \tag{2.79}$$

$$P_t^w = D_{W,t}(d_{GS,t}, d_{LS,t})\Delta P_{W,t}, \tag{2.80}$$

$$P_t^b = D_{B,t}(d_{GD,t}^b)\Delta P_{B,t}, \tag{2.81}$$

where $A_G = AC_G, A_W = AC_W$, and $A_L = AC_L$. The matrices $D_{W,t}$ and $D_{B,t}$ include A_G, A_W, A_L, and the distribution vectors as linear functions of the random variables.

Let subscript i refer to the i^{th} row of a vector/matrix. The power flow constraints for each line i for *Operating Point 1* are

$$P_{t,i}^f + D_{W,t,i}\mu_{W,t} + D_{B,t,i}\mu_{B,t} + \mathcal{S}_{L,t,i} \leq P_{l,i}, \tag{2.82}$$

$$P_{t,i}^f + D_{W,t,i}\mu_{W,t} + D_{B,t,i}\mu_{B,t} - \mathcal{S}_{L,t,i} \geq -P_{l,i}, \tag{2.83}$$

$$\mathcal{S}_{L,t,i} \geq c\sqrt{D_{W,t,i}^T \Sigma_{W,t} D_{W,t,i} + D_{B,t,i}^T \Sigma_{B,t} D_{B,t,i}}, \tag{2.84}$$

where $\mathcal{S}_{L,t,i}$ is a slack variable. Constraints (2.38) for *Operating Points 2* and 3 can be reformulated similarly.

2.3.2 Approximation and Solving Algorithms

The nonlinear, convex constraints (2.63), (2.74), and (2.75) include empirically-computed confidence bounds z_1 and z_2. We approximate these bounds with analytic functions and compare the performance and computational requirements of the approximations in Section 2.3.3.

2.3.2.1 Polyhedral Approximation

The polyhedral approximation introduces only linear inequalities, i.e.,

$$z(d) \geq \alpha_j d + \beta_j \, \forall \, j, \tag{2.85}$$

where α_j and β_j are parameters corresponding to linearization j.

2.3.2.2 2-Norm Approximation

If M, in (2.62), were normally distributed with mean μ_M and standard deviation δ_M, and assume Y (which is also normally distributed) has mean 0 and standard deviation δ_Y, then

$$z(d) = \mu_M + c\sqrt{(\delta_M)^2 + (d\delta_Y)^2}, \tag{2.86}$$

Since M is close to normally distributed at low violation levels, we can approximate it as such and compute μ_M and δ_M via nonlinear regression on the empirically-computed confidence bounds. Then each constraint including a confidence bound can be written as a linear constraint and a nonlinear slack variable constraint, where the latter is an SOC constraint. This approximation requires less memory than the polyhedral approximation.

2.3.2.3 Cutting Plane Algorithm

We use the cutting plane algorithm from [9] to reduce the computational effort associated with the SOC constraints, i.e., all of the slack variable constraints. In the first step of the algorithm, the CC-OPF is solved without the slack variable constraints. Then, all of the slack variable constraints are evaluated. If all are satisfied, we have obtained the solution to the full problem. Otherwise, we introduce linear constraints corresponding to first order Taylor series expansions of the unsatisfied slack variable constraints, re-solve the problem, check the slack variable constraints,

and repeat until all are satisfied.

2.3.3 Case Studies

2.3.3.1 Set-up

We use the same set-up as in [93]. All results are generated using temperature forecast profile 5 (i.e., T case #5) and four intra-hour redispatch intervals, i.e., $n = 4$. In [93], we tested the approach on the IEEE 30-bus network using cost/parameter settings from MATPOWER [108]. We modified the network to include 4 wind power plants (i.e., $N_W = 4$) connected to buses 1, 2, 22, and 27 with capacities 10, 10, 20, and 10 MW, respectively. We also increased the line capacity limits by 50%. We defined 13 temperature forecast profiles (referred to as "T cases"), where the first corresponds to Fig. 3 (a), the last to Fig. 3 (b) in [93], and intermediary cases shift all temperatures by 1° C. We used load profiles from NREL [40] and modeled all loads as partially controllable. Specifically, we assumed two-thirds of each load in the first hour of the day corresponding to T case #12 is controllable. Then, for the remainder of the day, the portion of controllable load is determined by the temperature. We assumed all loads are affected by the same temperature, i.e., $N_T = 1$, and symmetric reserve deployment. We set each generator's secondary reserve costs to be 5 times its linear energy cost and redispatch costs to be equal to its linear energy costs, and we set each load's secondary reserve costs to be 0.5 \$/MW. Wind power and temperature forecast error scenarios were generated using real data and the Markov Chain Monte Carlo mechanism described in [70]. Specifically, we used forecasted and actual hourly wind power data from Germany and forecasted and actual hourly temperature data from eleven weather stations in Switzerland to train a transition probability matrix that we then used to generate the scenarios. We set $\epsilon_t = 10\%$ and $\beta_t = 10^{-4} \quad \forall t = 1, \ldots, N_t$ and so we needed 447 uncertainty scenarios. We evaluated the empirical violation probability using $10,000$ independent scenarios.

All optimization problems were solved using the solver MOSEK via the MATLAB interface CVX.

We compute and compare solutions for four test cases that differ in formulation and solution approach.

- **Scenario 1**: the formulation/approach presented in Section 2.2, i.e., the original constraints and original load capacities/baseline (see Fig. 2.1), solved with the scenario-based method.

- **Scenario 2**: the approximate constraint (2.64) and the approximate load capacities/baseline in Fig. 2.3, solved with the scenario-based method.

- **Analytical 1**: the approximate constraint (2.64) and the approximate load capacities/baseline in Fig. 2.3, analytically reformulated using the *polyhedral approximation* of the confidence bounds and solved with i) a nonlinear solver and ii) the cutting plane algorithm.

- **Analytical 2**: the approximate constraint (2.64) and the approximate load capacities/baseline in Fig. 2.3, analytically reformulated using the *2-norm approximation* of the confidence bounds and solved with i) a nonlinear solver and ii) the cutting plane algorithm.

We use the same uncertainty scenarios used in [93]. Specifically, for each test case, we conduct five simulation runs. The first set of runs corresponding to $1-\epsilon = 90\%$ use the same 447 scenarios as used in [93] to compute the solutions for Scenario 1 and 2 and to compute the mean and covariance of the random variables for Analytical 1 and 2, and the same 10,000 scenarios as used in [93] to compute the empirical reliability, which is defined as one minus the empirical violation probability. The other four sets of runs corresponding to $1 - \epsilon = 90\%$ and five sets of runs corresponding to $1 - \epsilon = 99\%$ use different random selections of the uncertainty scenarios, where the

Table 2.1: Cost Distribution and Reserve Allocation, $1-\epsilon = 90\%$

		Scenario 1	Scenario 2	Analytical 1 & 2
Cost	Total	17320	17359	12778
	Dispatch	11941	11942	11867
	GS	904	950	0
	LS	237	232	56
	GD	4238	4235	855
Capacity (MW)	GS	181	190	0
	LS	474	464	112
	GD	2907	2905	855

Table 2.2: Cost Distribution and Reserve Allocation, $1-\epsilon = 99\%$

		Scenario 1	Scenario 2	Analytical 1 & 2
Cost	Total	19657	19721	13474
	Dispatch	11978	11978	11876
	GS	2863	2938	0
	LS	97	90	87
	GD	4719	4716	1510
Capacity (MW)	GS	572	588	0
	LS	194	179	175
	GD	3162	3158	1370

same scenarios are used to both compute the solutions for Scenario 1 and 2 and to compute the mean and covariance of the random variables for Analytical 1 and 2. Therefore, the number of scenarios used to compute the mean and covariance of the random variables for Analytical 1 and 2 is different for $1-\epsilon = 90\%$ and $1-\epsilon = 99\%$.

Table 2.3: Computational Time (s), NL = Nonlinear Solver, CPA = Cutting Plane Algorithm

$1-\epsilon$	Scenario 1	Scenario 2	Analytical 1 (NL)
90%	86.5	86.5	312.5
99%	76.1	77.7	314.4

$1-\epsilon$	Analytical 2 (NL)	Analytical 1 (CPA)	Analytical 2 (CPA)
90%	440.9	39.2	38.4
99%	433.2	38.9	38.8

2.3.3.2 Results

Tables 2.1 and 2.2 show the cost distribution and reserve allocation for the four test cases for $1-\epsilon = 90\%$ and 99%, respectively, and for the first simulation run. The solutions corresponding to Scenarios 1 and 2 are more costly than those corresponding to Analytical 1 and 2, because the scenario-based method is more conservative since it uses extreme uncertainty scenarios from a probabilistically robust set. Subsequently, the scenario-based method procures more reserves leading to higher reserve costs and achieves less load shifting leading to higher dispatch costs. Additionally, we find that the solution corresponding to Scenario 2 is more costly/conservative than that corresponding to Scenario 1 since Scenario 2 uses the approximate load capacities and baseline. The approximation error results in small suboptimal changes to $P_{C,t}$ that reduce the ability of the loads to provide secondary reserves. The different approximations and solvers used in Analytical 1 and 2 do not affect the solutions or costs. Lastly, we find that decreasing ϵ increases reserve procurement and costs for all test cases. In Scenario 1 and 2, decreasing ϵ increases the number of uncertainty scenarios needed to generate the probabilistically robust set and, in Analytical 1 and 2, it increases $c = \Phi_N^{-1}(1-\epsilon)$. Both changes tighten the feasible region.

Table 2.3 shows the computational time for the four test cases, for $1-\epsilon = 90\%$ and 99%, where Analytical 1 and 2 are solved with both the nonlinear solver (NL) and the cutting plane algorithm (CPA). Results for all test cases correspond to the average of the five simulation runs. Table 2.4 shows the breakdown of the computational time associated with solving Analytical 1 and 2 using the cutting plane algorithm. In each step, an optimization problem is solved and more constraints are added as the step increases from 1 to 4. For our problem, the algorithm converges after 4 steps. The column "Set-up" lists the the total time between all iterations, which is required to compute the cuts.

While the computational times associated with Scenario 1 and 2 are smaller than

Table 2.4: Breakdown of the Computational Time (s) Using the CPA

	$1-\epsilon$	Step 1	Step 2	Step 3	Step 4	Set-up
Analytical 1	90%	2.9	9.0	11.1	13.1	3.2
Analytical 1	99%	2.9	8.8	10.9	13.0	3.3
Analytical 2	90%	2.0	8.8	11.1	13.3	3.2
Analytical 2	99%	1.9	8.8	11.5	13.4	3.2

those associated with Analytical 1 and 2 using the nonlinear solver, the computational times associated with Analytical 1 and 2 using the cutting plane algorithm are the smallest. Comparing our results to those of [51], we find that this computational advantage increases with problem dimension. The computational time of the first step of Analytical 2 is less than that of Analytical 1, but at Step 4 the computational time of Analytical 1 is less than that of Analytical 2. This is because Analytical 2 uses the 2-norm approximation, which includes a nonlinear slack variable constraint. New linear approximations of the constraint may be introduced in each step.

Table 2.5 shows the average empirical joint and individual reliability of the solutions associated with each of the test cases, where the average is taken over the five simulation runs. Table 2.6 shows the maximum and minimum empirical joint reliability across the five simulation runs.[2] The joint reliability is the percent of hourly sets of constraints that are satisfied concurrently considering all of the scenarios, whereas the individual reliability is the percent of constraints that are satisfied individually considering all scenarios. Recall that the scenario-based method satisfies constraints jointly, while the analytical reformulation is only capable of satisfying constraints individually.

As shown in Table 2.5, Scenario 1 and 2 are conservative, achieving empirical joint reliabilities well-above the desired joint reliability. Analytical 1 and 2 achieve much lower empirical joint reliabilities, but do achieve empirical individual reliabilities well-above the desired individual reliability because, for each uncertainty scenario, few

[2]The minimum empirical joint reliability for Scenario 1, $1-\epsilon = 90\%$, i.e., 98.15%, is comparable to the worse-case hourly empirical violation probability reported in [93], i.e., 1.86%.

Table 2.5: Average Empirical Joint and Individual Reliability (%)

	$1-\epsilon$	Scenario 1	Scenario 2	Analytical 1 & 2
Joint	90%	99.66	99.67	80.42
Individual	90%	100.00	100.00	99.54
Joint	99%	99.98	99.98	95.47
Individual	99%	100.00	100.00	99.89

Table 2.6: Maximum and Minimum Empirical Joint Reliability (%)

	$1-\epsilon$	Scenario 1	Scenario 2	Analytical 1 & 2
Max	90%	100.00	100.00	88.77
Min	90%	98.15	98.49	75.51
Max	99%	100.00	100.00	97.42
Min	99%	99.87	99.87	94.38

constraints are active. Increasing $1-\epsilon$ from 90% to 99% improves the empirical joint reliability of Analytical 1 and 2 indicating that c could be tuned to achieve the desired joint reliability. As shown in Table 2.6, the empirical joint reliability of Analytical 1 and 2 varies more than that of Scenario 1 and 2, but increasing $1-\epsilon$ decreases the range.

Table 2.7 shows the empirical joint reliability by constraint type for Analytical 1 and 2. The first row (Power Flow) corresponds to all of the power flow constraints for all three operating points; the second row (Energy Capacity) corresponds to all controllable load energy capacity constraints; the third row (Operating Point 1) corresponds to all generation, controllable load power capacity, and reserve capacity constraints corresponding to *Operating Point 1*; the fourth row (Operating Point 2) corresponds to all generation, controllable load power capacity, and reserve capacity constraints corresponding to *Operating Point 2*; and the fifth row (Operating Point 3) corresponds to all generation, controllable load power capacity, and reserve capacity constraints corresponding to *Operating Point 3*, except the redundant constraints included in Operating Point 1. Power flow and energy capacity constraints have high joint reliability because they are rarely active. In contrast, reserve capacity

Table 2.7: Empirical Joint Reliability by Constraint Type for Analytical 1 & 2 (%)

	$1-\epsilon = 90\%$	$1-\epsilon = 99\%$
Power Flow	99.56	99.88
Energy Capacity	95.53	99.17
Operating Point 1	85.49	96.71
Operating Point 2	90.94	98.17
Operating Point 3	92.48	98.53

constraints, which are included within the operating point constraints, are generally active since we attempt to minimize reserve capacities while providing enough reserves to compensate for wind power and temperature forecast error. This point is also demonstrated by the large improvement in the joint reliability of Operating Point 1–3 constraints as we increase $1-\epsilon$ from 90% to 99%.

To summarize how the solution approaches scale, Tables 2.8 and 2.9 show the number of constraints per hour required by the scenario-based method and the analytical reformulations, respectively. We split the constraints by type and by form, i.e., linear v.s. SOC constraints. As before, N_W is the number of wind power plants, N_T is the number of temperature zones, N_C is the number of controllable load aggregations, and j is the number of linear inequalities introduced by the polyhedral approximation in Analytical 1. We also define N_G as the number of conventional generators and N_L as the number of transmission lines. The number of decision variables per hour is $10N_G + 4N_C$. Observe that N_W and N_T do not affect the number of constraints, but do affect the size of the matrices/vectors used to represent the constraints. These tables can be used to determine the increased computational effort required for increased system sizes and to determine the composition of constraints for Analytical 1 and 2.

Table 2.8: Number of Constraints per Hour Required for the Scenario-Based Method

	Linear	SOC
Deterministic, Equality	$6 + N_C$	0
Deterministic, Inequality	$2N_L + 11N_G + 9N_C$	0
Power Flow	$6N_L$	0
Energy Capacity	$4N_C$	0
Operating Point 1	$6N_G + 4N_C$	0
Operating Point 2	$4N_G + 2N_C$	0
Operating Point 3	$4N_G + 2N_C$	0
Total	$6 + 25N_G + 22N_C + 8N_L$	0

Table 2.9: Number and Type of Constraints per Hour Required for Analytical 1 and 2

	Analytical 1 Linear	Analytical 1 SOC
Deterministic, Equality	$6 + N_C$	0
Deterministic, Inequality	$2N_L + 11N_G + 9N_C$	0
Power Flow	0	$6N_L$
Energy Capacity	$2jN_C$	$2N_C$
Operating Point 1	$4N_G + (2 + j)N_C$	$2N_G + N_C$
Operating Point 2	jN_C	$4N_G + N_C$
Operating Point 3	0	$4N_G + (1 + j)N_C$
Total	$6 + 15N_G + (12 + 4j)N_C + 2N_L$	$10N_G + (5 + j)N_C + 6N_L$

	Analytical 2 Linear	Analytical 2 SOC
Deterministic, Equality	$6 + N_C$	0
Deterministic, Inequality	$2N_L + 11N_G + 9N_C$	0
Power Flow	0	$6N_L$
Energy Capacity	0	$4N_C$
Operating Point 1	$4N_G + 2N_C$	$2N_G + 2N_C$
Operating Point 2	0	$4N_G + 2N_C$
Operating Point 3	0	$4N_G + 2N_C$
Total	$6 + 15N_G + 12N_C + 2N_L$	$10N_G + 10N_C + 6N_L$

2.3.4 Supporting Material

2.3.4.1 Confidence Bound Derivation

In this section, we derive the confidence bound $z_1(d_{LS,t})$ corresponding to one of the N_C constraints in (2.63) for Case 1 (i.e., $T_t^f < T_{br}$) at one time interval t. Other confidence bounds can be derived similarly.

Without loss of generality, assume the mean of $X = \Delta T_t \in \mathbb{R}$ is 0 and the standard deviation is δ_X. Then, the probability density function (PDF) of X is

$$f_X(x) = \frac{1}{\delta_X\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\delta_X^2}\right), \tag{2.87}$$

and the CDF of M, defined in (2.62), is

$$\mathbb{P}(M \leq m) = F_X\left(\frac{m-h}{k_2}\right) - F_X\left(\frac{m}{k_1}\right), \tag{2.88}$$

where F_X is the CDF of X, $k_1 = a$, $k_2 = a - k_p$, and $h = k_p e_{T,t}$. The PDF of M is

$$\begin{aligned} f_M(m) = \frac{d\mathbb{P}(M \leq m)}{dm} =& \frac{f_X\left(\frac{m-h}{k_2}\right)}{k_2} - \frac{f_X\left(\frac{m}{k_1}\right)}{k_1} \\ =& \frac{f_X\left(\frac{m-h}{k_2}\right)}{k_2} + \frac{f_X\left(\frac{m}{-k_1}\right)}{-k_1} \end{aligned} \tag{2.89}$$

since f_X is symmetric with respect to zero.

Similarly, assume the mean of $dY = d_{LS,t}P_{m,t} \in \mathbb{R}$ is 0 and the standard deviation is $d\delta_Y = d_{LS,t}\overline{\delta}_{W,t}$. The PDF of dY is

$$f_{dY}(y) = \frac{1}{d\delta_Y\sqrt{2\pi}}\exp\left(-\frac{y^2}{2(d\delta_Y)^2}\right) \tag{2.90}$$

Since $h < 0$, $k_1 < 0$, and $k_2 > 0$, M is lower bounded by some value $-C_3 < 0$.

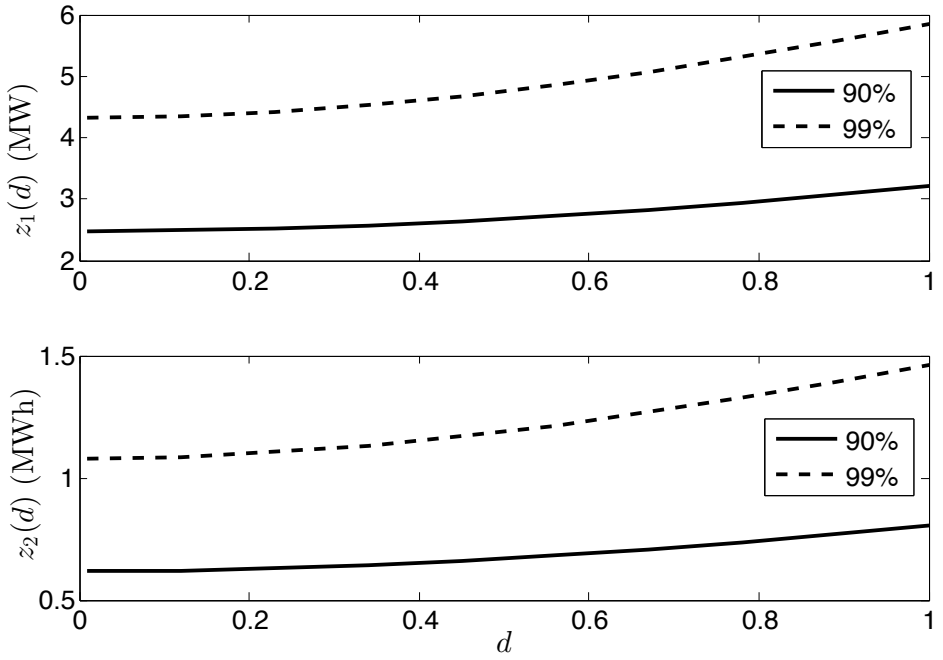


Figure 2.4: Example confidence bounds $z_1(d)$ and $z_2(d)$ for $1-\epsilon = 90\%$ and $1-\epsilon = 99\%$.

Therefore, the CDF of Z is

$$\begin{aligned}\mathbb{P}\left(Z \leq z_1(d)\right) &= \int_{-C_3}^{\infty} \int_{-\infty}^{z_1(d)-m} f_M(m) f_{dY}(y)\,\mathrm{d}y\,\mathrm{d}m \\ &= \int_{-C_3}^{\infty} f_M(m) \Phi_N\left(\frac{z_1(d)-m}{d\delta_Y}\right)\,\mathrm{d}m. \end{aligned} \tag{2.91}$$

We solve for $z_1(d)$ empirically by setting (2.91) equal to $1-\epsilon$ and sweeping discrete values of d within its domain, i.e., $[0,1]$. Examples of the confidence bounds are shown in Fig. 2.4.

2.3.4.2 Monotonicity and Convexity Proof Sketches

All constraints with the exception of those that include confidence bounds are clearly convex. Figure 2.4 shows that the empirically-computed confidence bounds appear to be convex. In this section, we sketch the proofs of monotonicity and convexity for the confidence bound $z_1(d)$ in Section 2.3.4. Similar reasoning can be

used to prove monotonicity and convexity for all other confidence bounds, and so all constraints within Section 2.3.1 that include confidence bounds are convex.

Assuming we have a solution triple (d_0, z_0, P_0) corresponding to $(d_{LS,t}, z_1(d_{LS,t}), 1-\epsilon)$ and we pick a large P_0 such that $z_0 > C_3$ and the integral in (2.91) corresponding to $m \in [2z_0 + C_3, \infty]$ is negligible (because this region corresponds to the tail of f_M), we have

$$\mathbb{P}(Z \leq z_0) = \int_{-C_3}^{2z_0+C_3} f_M(m)\Phi_N\left(\frac{z_0 - m}{d_0\delta_Y}\right) \mathrm{d}m = P_0, \tag{2.92}$$

$$\min_{m\in[-C_3,z_0]} f_M(m) = f_M(z_0), \tag{2.93}$$

$$\max_{m\in[z_0,\infty]} f_M(m) = f_M(z_0). \tag{2.94}$$

Increasing d_0 to d_1, the change in P_0 is

$$\begin{aligned}&\int_{-C_3}^{2z_0+C_3} f_M(m) \underbrace{\left[\Phi_N\left(\frac{z_0 - m}{d_1\delta_Y}\right) - \Phi_N\left(\frac{z_0 - m}{d_0\delta_Y}\right)\right]}_{\Delta\Phi} \mathrm{d}m \\ &= \int_{-C_3}^{z_0} \left(f_M(m) - f_M(2z_0 - m)\right) \Delta\Phi \, \mathrm{d}m,\end{aligned} \tag{2.95}$$

since Φ_N is symmetric with respect to z_0. From (2.93) and (2.94) we know that $f_M(m) - f_M(2z_0 - m) \geq 0 \in [-C_3, z_0]$. We also know that $\Delta\Phi < 0 \in [-C_3, z_0]$. Hence, the change in P_0 is negative and to increase P_0 we would need to increase z_0. Therefore, for a fixed P_0, which is above some threshold ensuring that the above assumptions hold, $z(d)$ will monotonically increase.

Proving convexity also requires approximations on the tails of the PDFs. Though f_M is the summation of two normal PDFs, as shown in (2.89), one tail will dominate the other for large P_0. Hence, when we increase d_0 to d_1 the change in P_0 will be approximately the same as the change that would occur if f_M corresponded to a single normal PDF. Therefore, for a fixed P_0, which is above some threshold ensuring

that the above assumptions hold, $z(d)$ will be convex and, when d is scalar, the corresponding constraint will be convex. When d is a vector, constraint convexity requires $z(d)$ to be monotonic, which was shown above. This conclusion is supported by our simulation results, which demonstrate that 2-norm approximations of the confidence bounds are accurate for large P_0.

2.4 Impact of Uncertainties on Dispatch Costs and Emissions

In this section, we perform more case studies to show how different types and levels of uncertainty, reserve costs, and controllable load (CL) capacity affect power system dispatch, operational costs, and CO_2 emissions using both the analytical reformulation and scenario-based method.

2.4.1 Simulation Setup

The testbed is a modified IEEE 9-bus system [108] with a 20 MW capacity wind power plant added to Bus 1. We assume all generators can provide secondary and re-dispatch reserves. Generator costs are taken from the Matpower case file. We choose the "base case" reserve costs as follows: the generator secondary reserve cost is five times the corresponding generator's first-order generation cost, the re-dispatch cost is equal to the generator's first-order cost, and the load reserve cost is 50% of the lowest first-order generator cost (we vary these costs in one of our case studies). To assess CO_2 emissions, we assume the generators at Buses 1 and 3 are gas turbines and at Bus 2 is a coal generator. The fuel price and emission factors can be found in [54]. We use empirical data from [40] for the load profile, which is a bimodal profile with peaks around hours at $t = 10, 20$, which represents a common winter load. We assume that 15% of the load at each bus is controllable but uncertain, and that the uncontrollable load is perfectly forecastable. We set the line flow constraint from Bus 1 to 4 at 50 MW to produce congestion in periods of high demand.

We use the same temperature and wind forecast error uncertainty set as [97], which computed the set from real data. We use two methods to solve the CC-OPF: the robust method [58] and the Gaussian method [79, 9] and use a similar formulation as in [51]. For the robust method, the number of required error samples is determined by the probability of chance constraint violation (here, 1%), a confidence parameter (here, 10^{-4}), and the uncertainty dimension (here, 2) [58]. We draw 2565 uncertainty samples from the set and use them to construct a robust set. For the Gaussian method, we compute the sample mean and covariance of the uncertainty set. Different from [51], we use a conservative approximation for confidence bound (the upper bound, which is labelled "UB" in Fig. 1 of [51]) to simplify the formulation.

To evaluate the CO_2 emissions we consider i) the emissions from fuel combustion from conventional generators and ii) the emissions from provision of reserves by generators. For i), we assume the hourly fuel consumed is $F(G,t) = f_g(P_{G,t})/k_c$, where $f_g(P_{G,t})$ is the generation cost and k_c is the fuel price. The emissions are then: $E(P_{G,t}) = k_e F(G,t)$, where k_e is the emissions factor. We assume wind generation has no emissions. When the generator is providing different upward and downward reserves, the actual average output could be different from the dispatch schedule. To evaluate the emissions, we define $P_{e,t} = P_{G,t} + (R_t^{up} - R_t^{dn})/2$, where R_t^{up} and R_t^{dn} are the upward and downward reserves provided by the generator. When the reserves are symmetric (i.e., $R_t^{up} = R_t^{dn}$), then $P_{e,t} = P_{G,t}$. Generators may produce additional emissions when they change their output quickly to track secondary reserves signals; however, the amount is not well known [54]. In this section, we let these emissions be: $E_o(R_t^{up}, R_t^{dn}) = \alpha E_m(R_t^{up} + R_t^{dn})$, where E_m is the marginal emissions from the generator at operating at $P_{G,t}$ and α is a scaling factor. In this section, we take $\alpha = 0.1$; more discussion can be found in [54].

We study the impacts of five different factors, listed below, on the system dispatch. We define both a base case and a range for each factor.

Table 2.10: Base Case Costs & Emissions Results

	Dispatch ($)	Gen. Sec. ($)	Re-dispatch ($)	Load Sec. ($)	Emissions (lbs)
Robust	43371	235.4	528.5	154.1	2.59e+06
Gaussian	43022	0	222.8	72.0	2.54e+06

Wind Forecast Error As wind forecast error increases, the system needs more reserves. We scale the wind forecast errors by values of {0.5, 0.75, 1, 1.25, 1.5}, where 1 corresponds to the base case.

Temperature Forecast Error Temperature affects the CL baseline and reserve capacity. As temperature forecast error increases, the system needs more reserves and the CL reserve capacity becomes less certain, requiring more reserves from conventional providers. We scale the temperature forecast errors by values of {0.8, 1, 1.2, 1.4, 1.6}.

Temperature Forecast The temperature forecast affects the CL baseline and energy/power capacity. We add an hourly temperature offset of {-10, -5, 0, +5, +10} to the base case forecast.

Controllable Load Energy Capacity To study the impact of the CL's energy capacity, we apply a scaling factor with values of {1, 3.25, 5.5, 7.75, 10}. Increasing the CL's energy capacity (which corresponds to increasing the temperature range within which the heat pumps are allowed to operate) increases the flexibility of the CLs.

Reserve Costs We scale the generator secondary reserve costs by applying a scaling factor with values of {0.02, 0.265, 0.51, 0.755, 1}.

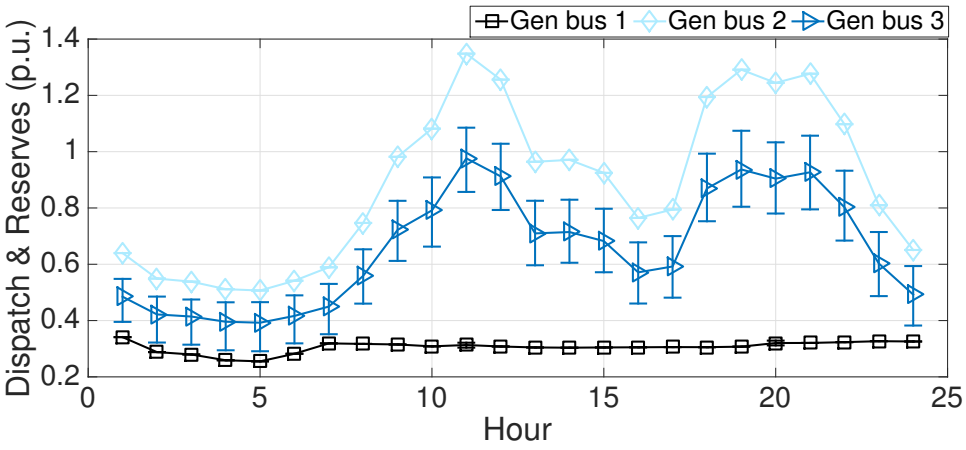


Figure 2.5: Generation schedule with re-dispatch reserves for each generator in the base case.

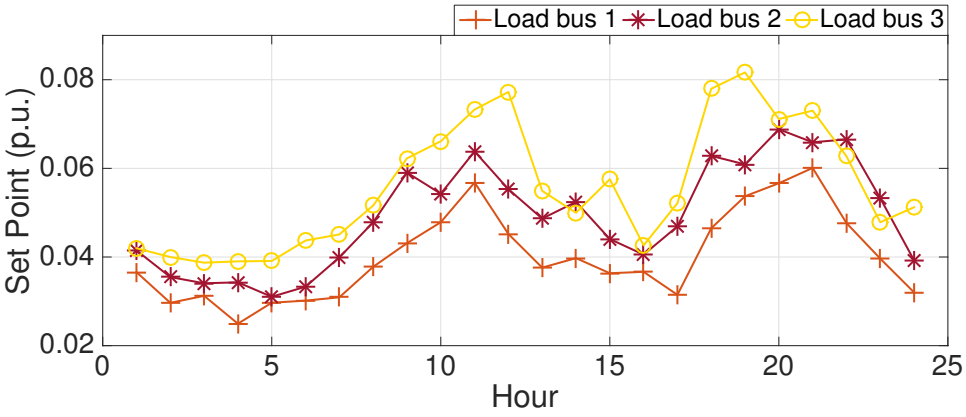


Figure 2.6: CL set point for each load in the base case.

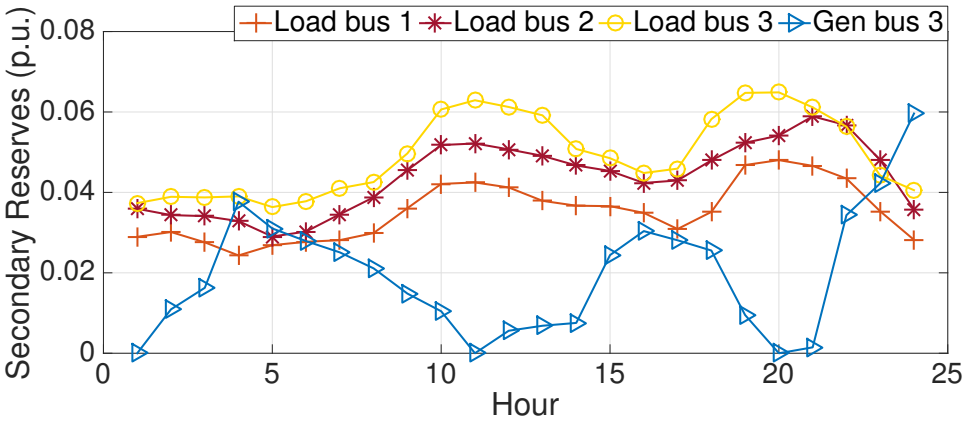


Figure 2.7: Secondary reserve capacity provided by generators and loads in the base case.

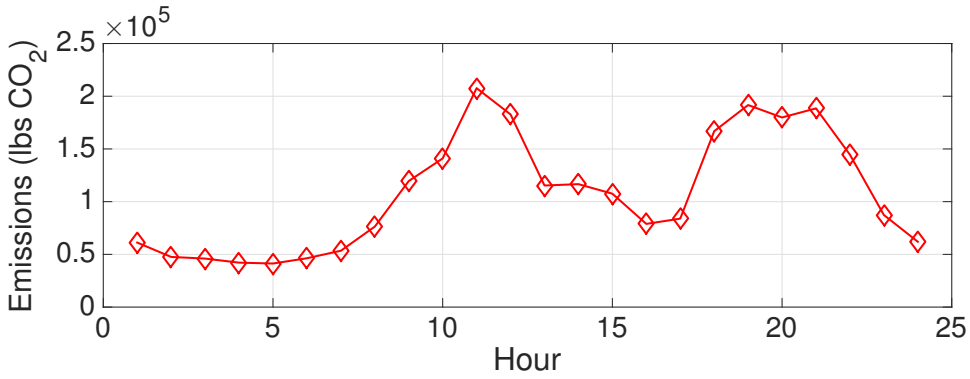


Figure 2.8: Hourly CO_2 emissions in the base case.

2.4.2 Results

Base Case The base case results using the robust method are shown in Figs. 2.5-2.8. Figure 2.5 shows the generation dispatch schedule as well as the re-dispatch reserve schedule, where "Gen bus n" refers to generator bus n.

We observe that when demand is high, line 1-4 is congested, and the generator on Gen bus 1, which is the lowest cost generator, is limited. Figure 2.6 shows CL set points for all loads. Figure 2.7 shows the secondary reserve capacity provided by each generator and load. We observe that load-based reserves are used first due their low cost. When their power or energy capacity is used up, the loads cannot provide more reserves, and so the generators are used to cover the rest of the reserve requirement (we subsequently refer to this as "saturation"). Load-based reserves cover most of the reserve requirement especially during peak hours since the CLs have larger capacities during peak hours. Figure 2.8 shows the emissions results, which follow a similar trend to the bimodal load profile since most of the emissions are from generator energy production.

Table 2.10 summarizes all costs and CO_2 emissions for both the robust method and the Gaussian method. The Gaussian method is less conservative resulting in fewer reserves, which both lowers costs and reduces emissions. The qualitative trends are similar. For example, CLs are used first for secondary reserves. Since fewer reserves

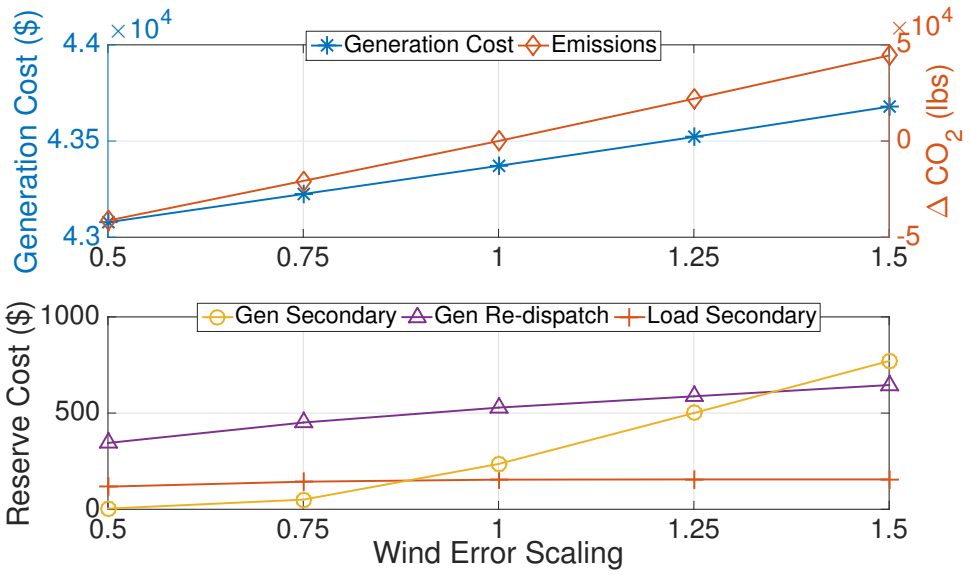


Figure 2.9: Dispatch costs and emissions (top) and reserve costs (bottom) for increasing wind forecast error. ΔCO_2 is the emissions difference from the base case.

are needed, all secondary reserves are provided by the loads.

In the following sections, we present the results associated with varying each factor listed in Section 2.4.1. In each case, the Gaussian method always produces lower costs and emissions, but similar qualitative results. Therefore, we only present the results for the robust method.

Wind Forecast Error The dispatch costs and emissions difference from the base case (ΔCO_2) and the secondary and re-dispatch reserve costs are shown in Fig. 2.9. As wind forecast error increases, the reserve requirement increases. We observe that load-based secondary reserves increase until saturation. For low uncertainty, there is no need for generator secondary reserves because the load-based reserves are able to compensate all of the error. After load-based reserves saturate, generator secondary reserves are used, and their cost increases linearly with wind error. Generator re-dispatch reserves increase with load-based reserves to recover the CLs' energy states.

Increased wind error also reduces the flow scheduled on line (1-4), ensuring the line constraint will not be violated when the wind forecast uncertainty is high. This

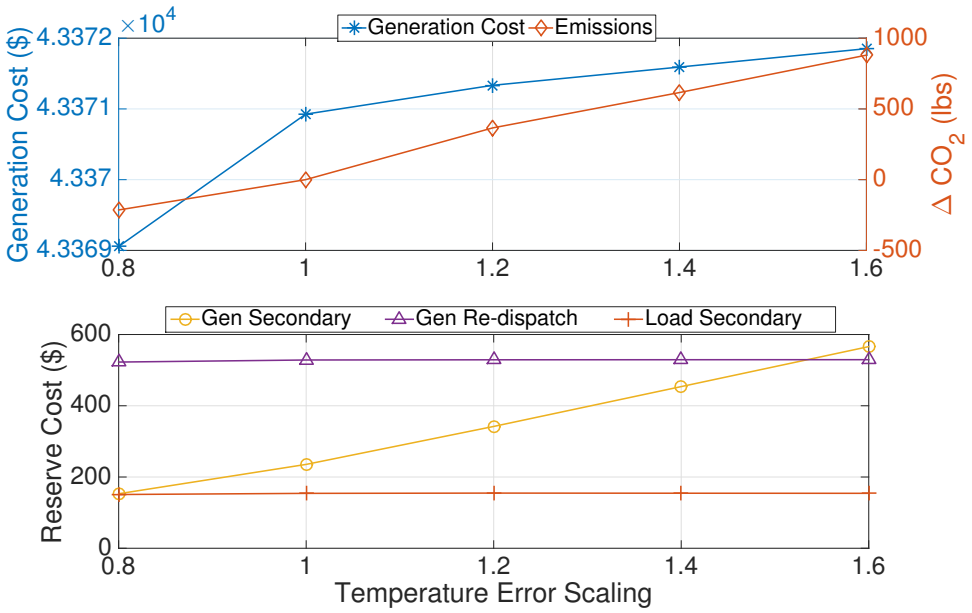


Figure 2.10: Dispatch costs and emissions (top) and reserve costs (bottom) for increasing temperature forecast error. ΔCO_2 is the emissions difference from the base case.

results in part of the generation being shifted from Gen bus 1 to the other buses. Since wind generation is also on Gen bus 1, conventional generator dispatch decreases on Gen bus 1 and increases on Gen bus 2 and 3. The CL set points do not change much as wind forecast error increases. This is because, in this case, CLs are mainly used for reserves, not load shifting.

Since generation is shifted from Gen bus 1 to Gen bus 2 and 3, emissions increase significantly as the generator on Gen bus 2 is modeled as a coal generator, which has higher emission factor. There is also an increased asymmetry to the reserves which results in a small increase in emissions.

Temperature Forecast Error The dispatch costs and emissions difference from the base case and the secondary and re-dispatch reserve costs are shown in Fig. 2.10. As temperature forecast error increases the load-based reserves quickly saturate. Also, more re-dispatch reserves are required to recover the CLs' energy states. The cost increase is smaller compared to that in the wind forecast error case, since temperature

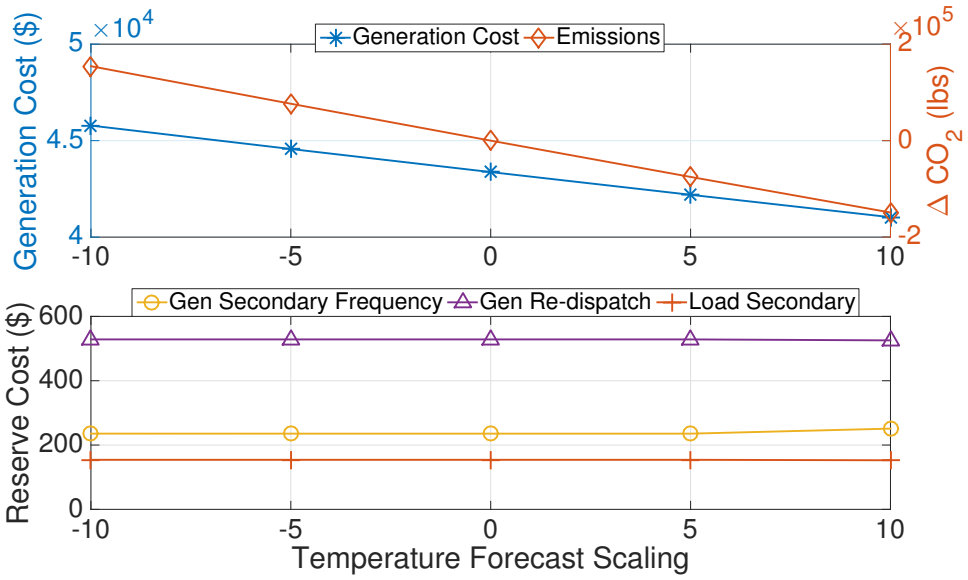


Figure 2.11: Dispatch costs and emissions (top) and reserve costs (bottom) for increasing temperature forecast. ΔCO_2 is the emissions difference from the base case.

forecast error does not induce large power mismatches. As temperature forecast error increases, a greater percentage of load-based reserve capacity is used to compensate temperature forecast error rather than wind forecast error, since the CLs are the only source for temperature foreacast error compensation. This leads to more secondary reserves required from the conventional generators to compensate the wind forecast error. Similar to the wind forecast error case, after load-based reserves saturate, secondary generation reserve costs increase linearly with temperature error.

There is a slight increase in emissions with respect to the base case (total change in emissions for the highest uncertainty case less than 900 lbs CO_2) and mostly due to more reserve usage since the generator dispatch schedule is not significantly affected by temperature error.

Temperature Forecast The dispatch costs and emissions difference from the base case and the secondary and re-dispatch reserve costs are shown in Fig. 2.11. The CL set points and load-based secondary reserve capacities for Load bus 1 are shown in Fig. 2.12. As the temperature forecast increases, the generation cost decreases since the

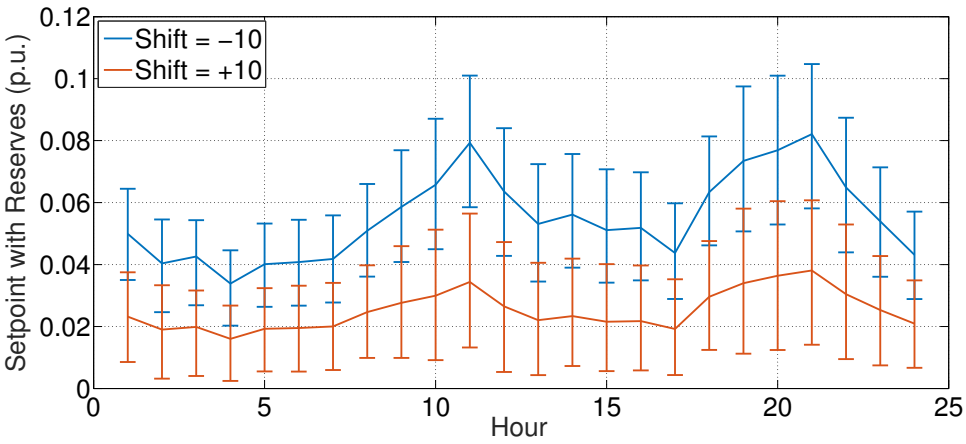


Figure 2.12: CL set point and load-based secondary reserves for lowest and highest temperature forecast on Load bus 1.

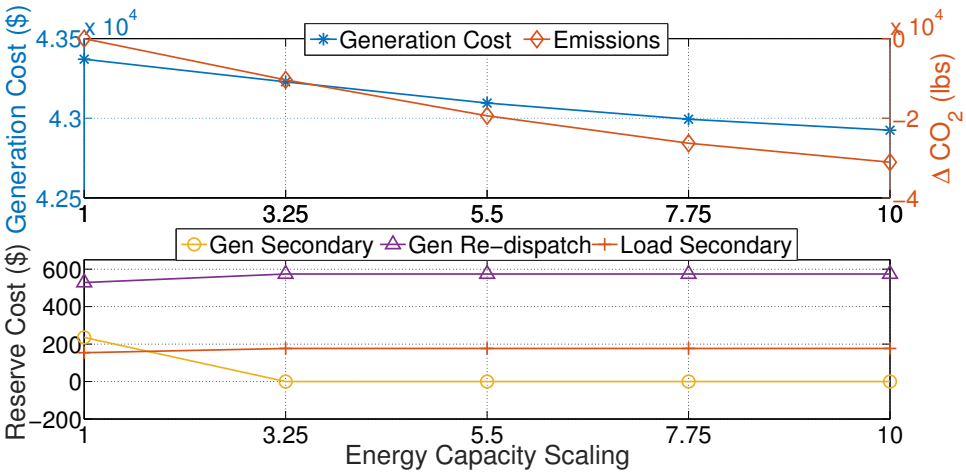


Figure 2.13: Dispatch costs and emissions (top) and reserve costs (bottom) for increasing CL energy capacity. ΔCO_2 is the emissions difference from the base case.

CL baseline decreases (i.e., heaters are used less as the temperature increases). The generator on Gen bus 1 is already limited by the line constraint, and so the generators on Gen bus 2 and 3 decrease to match the reduced demand. The load-based reserve costs decrease, the generator secondary reserve costs increase, and the re-dispatch reserves cost decrease for an offset value of 10 since the CLs' power capacities becomes very small; this can be observed in Fig. 2.12. Emissions reduce with increasing temperature since CL demand reduces, and the change in emissions is dominated by generator dispatch rather than reserves.

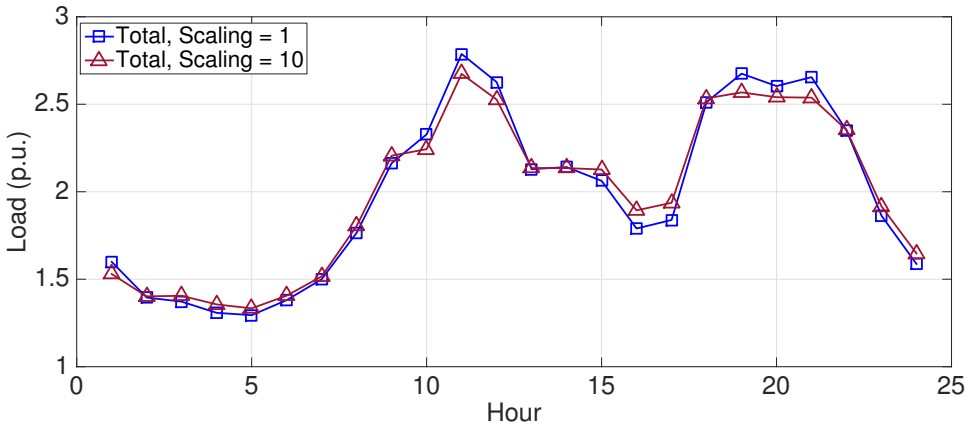


Figure 2.14: Total load schedule for lowest and highest energy capacity.

Controllable Load Energy Capacity The dispatch costs and emissions difference from the base case and the secondary and re-dispatch reserve costs are shown in Fig. 2.13. The total load (CL plus uncontrollable load) schedule is shown in Fig. 2.14. Generation dispatch costs and generation secondary reserve costs decrease as the energy capacities increase since the CLs can provide more reserves and load shifting. While the total CL over the day does not change, an increased energy capacity allows the CL to provide more load shifting, as shown in Fig. 2.14. The resulting peak shaving reduces operational costs as well as emissions since the marginal cost of generation is higher during peak hours. Load-based reserves increase with increased energy capacity until they provide all reserves. With increased energy capacities, emissions decrease for peak hours and increase for off-peak hours. The total emissions over the 24-hour period show a slight decrease as the CL energy capacity increases.

Reserve Costs The dispatch costs and emissions difference from the base case and the secondary and re-dispatch reserve costs are shown in Fig. 2.15. The generation dispatch schedule for Gen bus 1 and secondary reserves provided the generators and loads are shown in Figs. 2.16 and 2.17.

In the base case, the generator secondary reserve is more expensive than load-based secondary reserve, and at most times, it is more expensive than the marginal

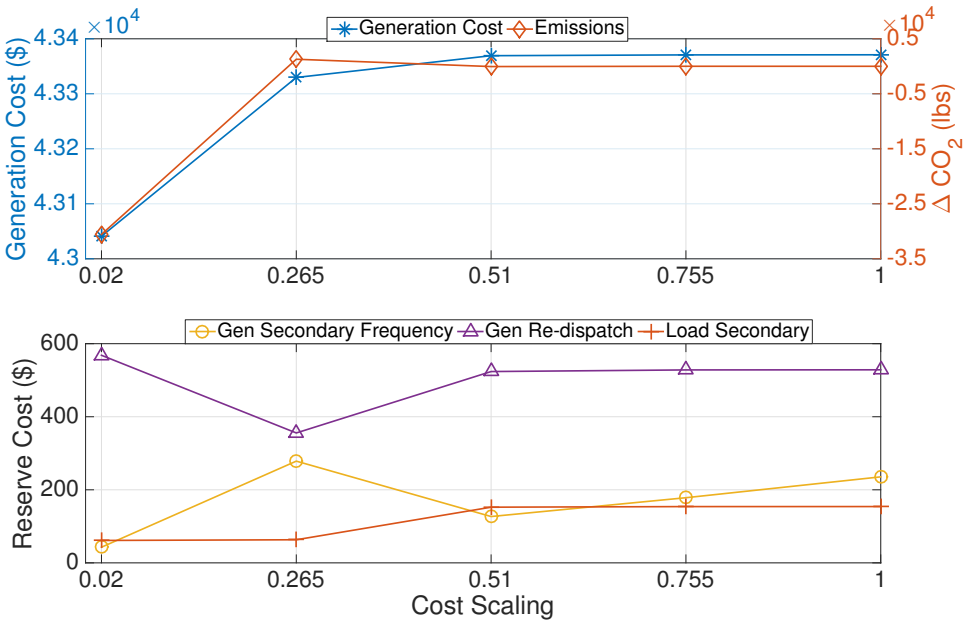


Figure 2.15: Dispatch costs and emissions (top) and reserve costs (bottom) for increasing generator secondary reserve cost. ΔCO_2 is the emissions difference from the base case.

cost of generation. When the scaling is set to 0.02, all generators have cheaper reserve costs than the CLs. Thus generators will be used first to compensate for wind forecast errors. Load-based reserves are still used as they are the only resource used to compensate temperature errors. In this case, during peak hours, the system picks a more expensive reserve provider – the generator on Gen bus 1. This is because the wind generation is also on Gen bus 1 and the line (1-4) is congested. Scheduling reserves on Gen bus 1 allows more power flow be scheduled on the congested line. This shift in generation reduces the generation cost, which is more than the increase in the reserve cost and leads to a lower total cost. CLs are also more effective at load shifting when most of the reserves are provided by generators.

As the scaling factor increases to 0.265, generator secondary reserves still compensate the majority of the wind forecast error due to its low cost. However, during peak hours, Gen bus 1 is no longer used for reserves and Gen bus 3 provides secondary and re-dispatch reserves. This is because the reserve cost increase outweighs the conges-

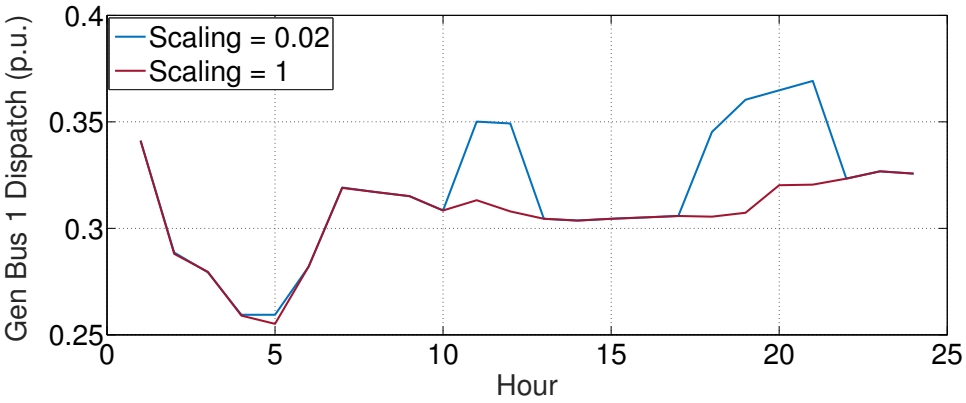


Figure 2.16: Generator dispatch for the lowest and highest generator secondary reserve costs on Gen bus 1.

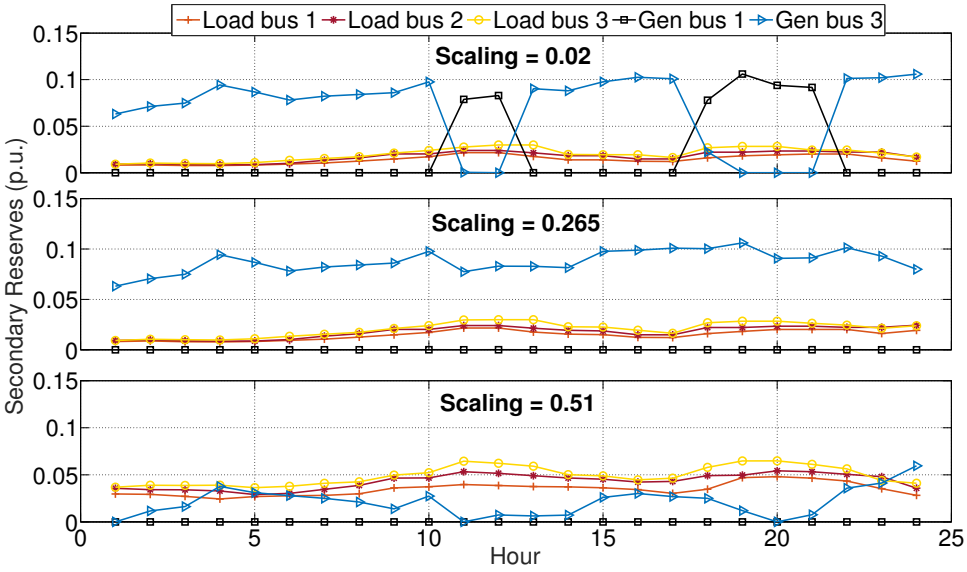


Figure 2.17: Secondary reserves for increasing generator secondary reserve costs. The cases with scaling 0.76 and 1 are the same as scaling 0.51 due to CL saturation.

tion relief. With less load-based reserve scheduled, the CLs can help more with load shifting. For scaling factors 0.51 to 1, the CLs will use their full capacity for reserves and become unable to load shift. Hence, we observe an increase in load-based reserve and dispatch costs.

For low generator secondary reserve costs, the generation dispatch is shifted from Gen bus 2 to 1 leading to a reduction in emissions. As generator secondary reserve costs increase, the dispatch is shifted back to Gen bus 2 and more load-based reserves are used. Therefore, we first observe an increase in emissions due to the dispatch shift then a reduction in emissions after load-based reserves saturate.

2.5 Conclusion

In this chapter, we first analytically reformulated the chance-constrained optimal power flow problem with uncertain controllable loads assuming temperature and wind power forecast uncertainty follow multivariate normal distributions. We proved that the nonlinear formulation is convex and we demonstrated how to compute empirical confidence bounds for the nonlinear constraints. We also showed how to approximate the confidence bounds with two different convex approximations and use an iterative cutting plane algorithm to reduce computational times. Through simulations, we compared the costs, solutions, computational times, and reliability of the approach to that of the scenario-based method. We showed that the analytical reformulation provides less conservative, lower cost solutions with empirical individual reliabilities above desired individual reliabilities. However, the approach does not guaranteed joint chance-constraint satisfaction and so empirical joint reliabilities are much lower than that of the scenario-based method. Still, joint reliabilities can be increased by heuristically tuning the desired individual reliability. Importantly, the analytical reformulation solved with the cutting plane algorithms requires less computational time than the scenario-based method for the test system used, and we provided tables

demonstrating the scalability of the approaches as a function of the problem size.

Next, we performed case studies to investigate the impacts of several factors including load-based reserve uncertainty (i.e., temperature uncertainty), wind uncertainty, reserve costs, and chance-constrain optimal power flow solution methodologies on power system dispatch, operational costs, and CO_2 emissions. We find that wind uncertainty has a larger impact on dispatch and emissions than temperature uncertainty. Changes in the temperature forecast highly affect the dispatch schedule and the emissions. Due to their low cost, controllable loads are scheduled to provide reserves before generators are, until they reach their capacity and the generators are needed. We also find that changes in generator dispatch have a more significant impact on the emissions than the changes in the reserve schedule, based on our emissions model. Higher controllable load energy capacities result in more load-based reserves, more load shifting, and reduced emissions. Complicated trade-offs were observed as the generator secondary reserve costs were varied.

CHAPTER III

Distributionally Robust Optimal Power Flow Using Moment and Unimodality Information

Optimization problems face random constraint violations when uncertainty arises in constraint parameters. Effective ways of controlling such violations include risk constraints, e.g., chance constraints and conditional Value-at-Risk (CVaR) constraints. In this chapter, we study these two types of risk constraints when the probability distribution of the uncertain parameters is ambiguous. In particular, we assume that the distributional information consists of the first two moments of the uncertainty and a generalized notion of unimodality, similar to the previous work in [48]. We find that the ambiguous risk constraints in this setting can be recast as a set of second-order cone (SOC) constraints. In order to facilitate the algorithmic implementation, we also derive efficient ways of finding violated SOC constraints. Finally, we demonstrate the theoretical results via a computational case study on power system operations. The main content of this chapter is summarized in the following papers.

1. B. Li, R. Jiang, and J. L. Mathieu. Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In *IEEE Con- ference on Decision and Control*, Las Vegas, NV, 2016

2. B. Li, R. Jiang, and J.L. Mathieu. Ambiguous risk constraints with moment and unimodality information. *Mathematical Prgramming (Accepted)*, 2017

3.1 Introduction

In an uncertain environment, optimization problems often involve making decisions before the uncertainty is realized. In this case, constraints, which may include security criteria and capacity restrictions, may face random violations. For example, we consider a constraint subject to uncertainty taking the form

$$a(x)^\top \xi \leq b(x), \tag{3.1}$$

where $x \in \{0,1\}^{n_{\mathrm{B}}} \times \mathbb{R}^{n-n_{\mathrm{B}}}$ represents an n-dimensional decision variable, $n_{\mathrm{B}} \in \{0,1,\ldots,n\}$ represents the number of binary decisions, $a(x) : \mathbb{R}^n \to \mathbb{R}^T$ and $b(x) : \mathbb{R}^n \to \mathbb{R}$ represent two affine transformations of x, and $\xi \in \mathbb{R}^T$ represents a T-dimensional random vector defined on probability space $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{P}_\xi)$ with Borel σ-algebra $\mathcal{B}^T$. An intuitive way of handling random violations of (3.1) is to employ chance constraints, which attempt to satisfy (3.1) with at least a pre-specified probability, i.e.,

$$\mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \geq 1-\epsilon, \tag{3.2}$$

where $1-\epsilon$ represents the confidence level of the chance constraint with ϵ usually taking a small value (e.g., 0.05 or 0.10) [18, 66]. Dating back to the 1950s, chance constraints have been applied in a wide range of applications including power system operations [68, 100], production planning [11, 28], and chemical processing [38, 39].

In practice, a decision maker is often interested in not only the violation probability of constraint (3.1), but also the violation magnitude if any [81, 82]. Indeed, chance constraint (3.2) offers no guarantees on the magnitude of $a(x)^\top \xi - b(x)$ when it is positive. This motivates an alternative risk measure called the conditional Value-at-Risk (CVaR) that examines the (right) tail of $a(x)^\top \xi - b(x)$. More precisely, the CVaR of a one-dimensional random variable χ with confidence level $1-\epsilon \in (0,1)$ is

defined as

$$\mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\chi}(\chi) = \inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}[\chi-\beta]_+ \right\}, \tag{3.3}$$

where $\mathbb{P}_\chi$ represents the probability distribution of χ and $[x]_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. When the infimum is attained in (3.3), β represents the Value-at-Risk of χ with confidence level $1-\epsilon$, that is, $\mathbb{P}_\chi\{\chi \leq \beta\} \geq 1-\epsilon$ [3, 82]. As a consequence, $\mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\chi}(\chi)$ measures the conditional expectation of χ on its right ϵ-tail. Hence, chance constraint (3.2) is implied by the CVaR constraint

$$\mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\xi}(a(x)^\top \xi) \leq b(x). \tag{3.4}$$

A basic challenge to using risk constraints (3.2) and (3.4) is that complete information of probability distribution $\mathbb{P}_\xi$ may not be available. Under many circumstances, we only have structural knowledge of $\mathbb{P}_\xi$ (e.g., symmetry, unimodality, etc.) and possibly a series of historical data that can be considered as samples taken from the true (while ambiguous) distribution. As a result, the solution obtained from a risk-constrained model can be biased, i.e., sensitive to the $\mathbb{P}_\xi$ we employ in constraints (3.2) and (3.4), and hence perform poorly in the out-of-sample tests. A natural way of addressing this challenge is to employ a set of plausible probability distributions, termed the ambiguity set, rather than a single estimate of $\mathbb{P}_\xi$.

3.1.1 Ambiguity Set with Unimodality Information

We consider an ambiguity set characterized by the first two moments of ξ and a structural requirement that $\mathbb{P}_\xi$ is unimodal in a generalized sense. By definition, if $T = 1$, then $\mathbb{P}_\xi$ is unimodal about 0 if function $F(z) := \mathbb{P}_\xi(\xi \leq z)$ is convex on $(-\infty, 0)$ and concave on $(0, \infty)$. If ξ admits a density function $f_\xi(z)$, then unimodality is equivalent to $f_\xi(z)$ being nondecreasing on $(-\infty, 0)$ and nonincreasing on $(0, \infty)$. In a multidimensional setting, i.e., if $T > 1$, an intuitive extension of this notion is

that $f_\xi(zd)$ is nonincreasing on $(0,\infty)$ for all $d \in \mathbb{R}^T$ and $d \neq 0$. That is, the density function of ξ is nonincreasing along any ray emanating from the mode. The following definitions extend this intuitive notion to also cover the distributions that do not admit density functions.

Definition III.1. (Star-Unimodality [23]) A set $S \subseteq \mathbb{R}^T$ is called *star-shaped* about 0 if, for all $\xi \in S$, the line segment connecting 0 and ξ is completely contained in S. A probability distribution $\mathbb{P}_\xi$ on $\mathbb{R}^T$ is called star-unimodal about 0 if it belongs to the closed convex hull of the set of all uniform distributions on sets in $\mathbb{R}^T$ which are star-shaped about 0.

In this chapter, we consider a more general notion than the star-unimodality as follows.

Definition III.2. (α-Unimodality [23]) For any given $\alpha > 0$, a probability distribution $\mathbb{P}_\xi$ is called α-unimodal about 0 if function $G(z) := z^\alpha \mathbb{P}_\xi(S/z)$ is nondecreasing on $(0,\infty)$ for all Borel set $S \in \mathcal{B}^T$.

If ξ admits a density function $f_\xi(z)$, then it can be shown that $\mathbb{P}_\xi$ is α-unimodal about 0 if and only if $z^{T-\alpha} f_\xi(zd)$ is nonincreasing on $(0,\infty)$ for all $d \in \mathbb{R}^T$ and $d \neq 0$ [23, 75]. As compared to star-unimodal distributions, the density of an α-unimodal distribution can increase along rays emanating from the mode (e.g., when $\alpha > T$), but the increasing rate is controlled by α. Indeed, along any ray d, $f_\xi(zd)$ does not increase faster than $z^{\alpha-T}$ on $(0,\infty)$. Furthermore, when $\alpha = T$, $f_\xi(zd)$ is nonincreasing on $(0,\infty)$ for all d. This implies that α-unimodality reduces to star-unimodality when $\alpha = T$.

Given the first two moments of ξ and α-unimodality, we define the following ambiguity set

$$\mathcal{D}_\xi(\mu, \Sigma, \alpha) := \left\{ \mathbb{P}_\xi \in \mathcal{M}_T : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma,\ \mathbb{P}_\xi \text{ is } \alpha\text{-unimodal about } 0 \right\}, \tag{3.5}$$

where $\mathcal{M}_T$ represents the set of all probability distributions on $(\mathbb{R}^T, \mathcal{B}^T)$, and μ and Σ represent the first and second moments of ξ, respectively. Without loss of generality, we assume that the mode of ξ is 0 in definition (3.5) and a general mode m can be modeled by shifting ξ to $\xi - m$ (see, e.g., Example 3.4.4 in [35]). For notational brevity, we often refer to ambiguity set $\mathcal{D}_\xi$ with its dependency on parameters (μ, Σ, α) omitted. Based on $\mathcal{D}_\xi$, we consider a distributionally robust chance constraint (DRCC)

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \geq 1 - \epsilon, \tag{3.6}$$

that is, we wish to satisfy chance constraint (3.2) for all probability distributions $\mathbb{P}_\xi$ in ambiguity set $\mathcal{D}_\xi$. Similarly, we define a distributionally robust risk constraint (DRRC)

$$\sup_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathrm{CVaR}^\epsilon_{\mathbb{P}_\xi}(a(x)^\top \xi) \leq b(x), \tag{3.7}$$

which requires that CVaR constraint (3.4) is satisfied for all $\mathbb{P}_\xi$ in $\mathcal{D}_\xi$.

3.1.2 Relations to the Prior Work

In recent years, distributionally robust optimization (DRO) has become an important tool to handle distributional ambiguity in stochastic programs. Using concepts similar to DRCC (3.6) and DRRC (3.7), DRO aims to optimize or protect a system from the worst-case probability distribution, which belongs to a pre-specified ambiguity set. DRO was first introduced by [83] as a minimax stochastic program for the classical newsvendor problem under an ambiguous demand with only moment information. Following this seminal work, moment information has been widely used for characterizing ambiguity sets in various DRO models [7, 22, 109]. A key merit of the DRO approach is that the model can often be recast as tractable convex programs such as semidefinite programs (SDPs) [22] or SOC programs [29]. Recently, [101] successfully identified a class of ambiguity sets that lead to tractable convex program

reformulations of general DRO models.

DRCCs with moment information (and without structural information) have been well-studied in recent years [29, 98, 13, 91, 110, 1, 19, 36]. In particular, [29], [98], and [13] showed that the DRCC can be recast as an SOC constraint if the ambiguity set is characterized by the first two moments of ξ. Later, [110] showed that DRCC and DRRC are actually equivalent if the same ambiguity set is employed. Recently, [1] and [19] extended the analysis of DRCC to the case when variable x involves binary (i.e., 0-1) decisions, and [36] made significant progress on representing the distributionally robust joint chance constraints in tractable forms. DRCCs with information on the density function have also been studied [26, 44, 27].

In contrast, DRCCs and DRRCs with both moment and structural information have received less attention. [77] considered general DRO models with ambiguity sets incorporating unimodality, symmetry, and convexity. Recently, by using the Choquet representation of α-unimodal distributions, [74] successfully derived SDPs to quantify the worst-case probability bound in DRCC. Furthermore, based on both α-unimodality and γ-monotonicity, [75] extended the analysis to quantifying the worst-case expectation in DRRC. The main focus of [74, 75] is to evaluate the worst-case expectations in DRCC or DRRC for a given decision variable x. In contrast, we adjust x to satisfy DRCC and DRRC. In our prior work [48], we derived approximations of DRRC. Here, we obtain an exact representation of DRRC and derive tighter approximations than those in [48]. To the best of our knowledge, our results on DRCC are most related to [35] (in particular, Example 3.4.4), which employs a different ambiguity set that bounds the second moment of ξ by Σ instead of matching it as in (3.5). Furthermore, [35] derived a representation of DRCC based on SDPs. In contrast, in this chapter, we employ a different approach based on projection, which allows us to represent DRCC with SOC constraints. SOC constraints are more computationally tractable than SDPs, especially when x involves binary decisions.

In addition, many off-the-shelf commercial solvers (e.g., CPLEX and GUROBI) can directly handle mixed-integer SOC programs.

We summarize our main contributions as follows.

1. We derive equivalent reformulations of DRCC (3.6) and DRRC (3.7) using both moment and unimodality information. Both reformulations are SOC constraints and so can be efficiently handled in commercial solvers. Different from previous results in [110], we find that DRCC and DRRC are not equivalent after incorporating the unimodality information.

2. Inspired by the separation approach [67], we derive efficient ways for finding violated SOC constraints in the reformulations of DRCC and DRRC. The separation procedures can be used to accelerate the algorithmic implementation of DRCC and DRRC.

3. We derive conservative and relaxed approximations of DRCC and DRRC that are asymptotically tight. As demonstrated in the computational case study, these approximations help to provide high-quality bounds for the optimal objective value of the test instances.

The remainder of this chapter is organized as follows. Section 3.2 represents DRCC (3.6) as a set of SOC constraints. Section 3.4 represents DRRC (3.7) as a set of SOC constraints. In both sections, we derive separation procedures for finding violated SOC constraints based on the golden section search. In Section 3.6, we analyze an extension of DRCC and DRRC to incorporate the linear unimodality into the ambiguity set. We present a computational case study in Section 3.7 and provide some supporting mathematical proofs in Section 3.8.

3.2 Distributionally Robust Chance Constraint

We show that DRCC (3.6) can be recast as second-order cone (SOC) constraints. To this end, we first simplify the computation of the left-hand side of (3.6), i.e., $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\}$, by projecting random vector ξ on $\mathbb{R}$ and considering a one-dimensional random variable ζ. We summarize this projection result in the following proposition, whose proof relies on the representation of α-unimodal random vectors in [23].

Proposition III.3. *Define scalars* $\mu_1 = a(x)^\top \mu$, $\Sigma_1 = a(x)^\top \Sigma a(x)$, *and ambiguity set* $\mathcal{D}_1 = \{\mathbb{P}_\zeta \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1,\ \mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1,\ \mathbb{P}_\zeta$ *is* α*-unimodal about* $0\}$. *Then*

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \ = \ \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}. \tag{3.8}$$

Proof. Theorem 3.5 in [23] states that a random vector $X \in \mathbb{R}^m$ is α-unimodal if and only if there exists a random vector $Z \in \mathbb{R}^m$ such that $X = U^{1/\alpha} Z$, where U is uniform in $(0, 1)$ and independent of Z.

First, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi$. Then, there exists Z_ξ such that $\xi = U^{1/\alpha} Z_\xi$. We define $\zeta = a(x)^\top \xi$. It follows that ζ is α-unimodal because $\zeta = a(x)^\top (U^{1/\alpha} Z_\xi) = U^{1/\alpha}(a(x)^\top Z_\xi)$. Furthermore, $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1$. Hence, $\mathbb{P}_\zeta \in \mathcal{D}_1$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \geq \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}$.

Second, pick any ζ such that $\mathbb{P}_\zeta \in \mathcal{D}_1$. Then, there exists a Z_ζ such that $\zeta = U^{1/\alpha} Z_\zeta$. It follows that $\mathbb{E}[Z_\zeta] = (\frac{\alpha+1}{\alpha})\mu_1$ and $\mathbb{E}[Z_\zeta^2] = (\frac{\alpha+2}{\alpha})\Sigma_1$. Based on Theorem 1 in [78], there exists a $Z_\xi \in \mathbb{R}^T$ such that $Z_\zeta = a(x)^\top Z_\xi$, $\mathbb{E}[Z_\xi] = (\frac{\alpha+1}{\alpha})\mu$, and $\mathbb{E}[Z_\xi Z_\xi^\top] = (\frac{\alpha+2}{\alpha})\Sigma$. We define $\xi = U^{1/\alpha} Z_\xi$. It follows that ξ is α-unimodal, and furthermore $\mathbb{E}_{\mathbb{P}_\xi}[\xi] = (\frac{\alpha}{\alpha+1})\mathbb{E}[Z_\xi] = \mu$ and $\mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = (\frac{\alpha}{\alpha+2})\mathbb{E}[Z_\xi Z_\xi^\top] = \Sigma$. Therefore, $\mathbb{P}_\xi \in \mathcal{D}_\xi$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \leq \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}$. ∎

Next, we compute the worst-case probability $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}$, for which we make the following two assumptions in the remainder of this section.

Assumption III.4. $\left(\frac{\alpha+2}{\alpha}\right)\Sigma \succ \left(\frac{\alpha+1}{\alpha}\right)^2 \mu\mu^\top$.

Assumption III.5. *Constraint $a(x)^\top\xi \leq b(x)$, and so constraint $\zeta \leq b(x)$ as well, is satisfied when ξ takes the value of its mode 0. That is, $b(x) \geq 0$.*

Assumption III.4 is standard in the literature and ensures that $\mathcal{D}_\xi \neq \emptyset$ [74]. Assumption III.5 is standard in the related literature (see, e.g., [74, 75, 35]). In fact, as DRCC (3.6) requires that $a(x)^\top\xi \leq b(x)$ holds with high probability, it is reasonable to assume that it also holds at the mode of ξ. Additionally, given DRCC (3.6) and Proposition III.3, we observe that Assumption III.5 holds if $\mathbb{P}_\zeta\{\zeta \leq 0\} < 1-\epsilon$ for each $\mathbb{P}_\zeta \in \mathcal{D}_1$, i.e., if the distributions in $\mathcal{D}_1$ are not extremely negative-skewed. To represent DRCC (3.6), we show an equivalent reformulation of $\inf_{\mathbb{P}_\zeta\in\mathcal{D}_1}\mathbb{P}_\zeta\{\zeta \leq b(x)\}$ in the following proposition that also sheds light on the worst-case probability distribution.

Proposition III.6. *Define $\mu_0 = \left(\frac{\alpha+1}{\alpha}\right)\mu_1$ and $\Sigma_0 = \left(\frac{\alpha+2}{\alpha}\right)\Sigma_1$. Then, $\inf_{\mathbb{P}_\zeta\in\mathcal{D}_1}\mathbb{P}_\zeta\{\zeta \leq b(x)\}$ is equivalent to the optimal objective value of optimization problem*

$$\min_{p_1,p_2,z_1,z_2} \quad p_1 + \left(\frac{b(x)}{z_2}\right)^\alpha p_2 \tag{3.9a}$$

$$s.t. \quad p_1 + p_2 = 1, \tag{3.9b}$$

$$p_1 z_1 + p_2 z_2 = \mu_0, \tag{3.9c}$$

$$p_1 z_1^2 + p_2 z_2^2 = \Sigma_0, \tag{3.9d}$$

$$p_1, p_2 \geq 0, \quad z_1 \in \mathbb{R}, \quad z_2 \geq b(x). \tag{3.9e}$$

Proof. First, we rewrite $\inf_{\mathbb{P}_\zeta\in\mathcal{D}_1}\mathbb{P}_\zeta\{\zeta \leq b(x)\}$ as a functional optimization problem

as follows:

$$\min_{\mathbb{P}_\zeta} \ \mathbb{P}_\zeta\{\zeta \le b(x)\} \tag{3.10a}$$

$$\text{s.t.} \ \ \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1, \tag{3.10b}$$

$$\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1, \tag{3.10c}$$

$$\mathbb{E}_{\mathbb{P}_\zeta}[1] = 1, \tag{3.10d}$$

$$\mathbb{P}_\zeta \text{ is } \alpha\text{-unimodal}, \tag{3.10e}$$

where constraints (3.10b)–(3.10c) describe the two moments of ζ, and constraint (3.10d) ensures that $\mathbb{P}_\zeta$ is a probability distribution. Using Theorem 3.5 in [23], since $\mathbb{P}_\zeta$ is α-unimodal, there exists a random variable Z such that $\zeta = U^{1/\alpha}Z$, where U is uniform in $(0,1)$ and independent of Z. It follows that $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mathbb{E}[U^{1/\alpha}]\mathbb{E}_{\mathbb{P}_Z}[Z] = (\frac{\alpha}{\alpha+1})\mathbb{E}_{\mathbb{P}_Z}[Z]$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \mathbb{E}[U^{2/\alpha}]\mathbb{E}_{\mathbb{P}_Z}[Z^2] = (\frac{\alpha}{\alpha+2})\mathbb{E}_{\mathbb{P}_Z}[Z^2]$. Furthermore,

$$\begin{aligned}
\mathbb{P}_\zeta\{\zeta \le b(x)\} &= \mathbb{P}\{U^{1/\alpha}Z \le b(x)\} \\
&= \int_{z=-\infty}^{+\infty} \mathbb{P}\{U^{1/\alpha}z \le b(x)\} d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{b(x)} 1 \ d\mathbb{P}_Z(z) + \int_{z=b(x)}^{+\infty} \mathbb{P}\left\{U^{1/\alpha} \le \frac{b(x)}{z}\right\} d\mathbb{P}_Z(z) \qquad (3.11a) \\
&= \int_{z=-\infty}^{b(x)} 1 \ d\mathbb{P}_Z(z) + \int_{z=b(x)}^{+\infty} \left(\frac{b(x)}{z}\right)^\alpha d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{+\infty} \left[\frac{b(x)}{\max\{z, b(x)\}}\right]^\alpha d\mathbb{P}_Z(z), \qquad (3.11b)
\end{aligned}$$

where equality (3.11a) is because $U^{1/\alpha}z \le b(x)$ when $z \le b(x)$ (note that $b(x) \ge 0$ due to Assumption III.5), and in (3.11b) we designate that $0/0 = 1$ in case $b(x) = 0$.

Hence, problem (3.10a)–(3.10e) can be recast as

$$\min_{\mathbb{P}_Z} \ \mathbb{E}_{\mathbb{P}_Z}\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha \tag{3.12a}$$

$$\text{s.t.} \ \ \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \tag{3.12b}$$

$$\mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \tag{3.12c}$$

$$\mathbb{E}_{\mathbb{P}_Z}[1] = 1. \tag{3.12d}$$

Second, we take the dual of problem (3.12a)–(3.12d) to obtain

$$\max_{\pi,\lambda,\gamma} \ \mu_0\pi + \Sigma_0\lambda + \gamma \tag{3.13a}$$

$$\text{s.t.} \ \ \lambda z^2 + \pi z + \gamma \leq \left[\frac{b(x)}{\max\{z, b(x)\}}\right]^\alpha, \quad \forall z \in \mathbb{R}, \tag{3.13b}$$

where dual variables π, λ, and γ are associated with primal constraints (3.12b)–(3.12d), respectively. Meanwhile, dual constraints (3.13b) are associated with primal variable $\mathbb{P}_Z$. Strong duality holds between problems (3.12a)–(3.12d) and (3.13a)–(3.13b) due to Assumption III.4 (see Proposition 3.4 in [84]). It follows that there exist an optimal solution $\mathbb{P}_Z^*$ to (3.12a)–(3.12d) that is discrete with at most 3 points of support (see Lemma 3.1 in [86]) and a finite optimal solution $(\pi^*, \lambda^*, \gamma^*)$ to (3.13a)–(3.13b) (see Proposition 3.4 in [84]).

Third, strong duality yields

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_Z^*}\left\{\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha - (\lambda^* Z^2 + \pi^* Z + \gamma^*)\right\}\\ =&\mathbb{E}_{\mathbb{P}_Z^*}\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha - (\lambda^*\Sigma_0 + \pi^*\mu_0 + \gamma^*) = 0.\end{aligned}$$

It follows that $\mathbb{P}_Z^*$ is supported at those points z such that $[b(x)/\max\{z, b(x)\}]^\alpha = \lambda^* z^2 + \pi^* z + \gamma^*$. From constraints (3.13b), we note that $\lambda \leq 0$ and so $\lambda z^2 + \pi z + \gamma$ is concave in z. Additionally, $[b(x)/\max\{z, b(x)\}]^\alpha$ is piecewise convex and consists of

two pieces, more specifically,

$$\left[\frac{b(x)}{\max\{z, b(x)\}}\right]^{\alpha} = \begin{cases} 1, & \text{if } z \leq b(x) \\ \left(\frac{b(x)}{z}\right)^{\alpha}, & \text{if } z > b(x), \end{cases}$$

and both pieces 1 and $(b(x)/z)^{\alpha}$ are convex in z. Hence, due to constraints (3.13b), $[b(x)/\max\{z, b(x)\}]^{\alpha}$ and $\lambda^* z^2 + \pi^* z + \gamma^*$ can meet at at most two points, each located at one piece of $[b(x)/\max\{z, b(x)\}]^{\alpha}$. It follows that, without loss of optimality, we can shrink the feasible region of formulation (3.12a)–(3.12d) to those discrete distributions with at most two points of support, each corresponding to one piece of $[b(x)/\max\{z, b(x)\}]^{\alpha}$. Therefore, formulations (3.12a)–(3.12d) and (3.9a)–(3.9e) have the same optimal objective value (note that we relax $z_1 \leq b(x)$ to $z_1 \in \mathbb{R}$ in (3.9a)–(3.9e) without loss of optimality, because it is suboptimal that both z_1 and z_2 correspond to the same piece of $[b(x)/\max\{z, b(x)\}]^{\alpha}$). ∎

Remark III.7. Suppose that $(p_1^*, p_2^*, z_1^*, z_2^*)$ is an optimal solution to problem (3.9a)–(3.9e). From the proof of Proposition III.6, we observe that problem (3.9a)–(3.9e) is solved for the worst-case probability distribution of a random variable Z such that $\zeta = U^{1/\alpha} Z$, where U is uniform on $(0, 1)$ and independent of Z. It follows that the worst-case distribution $\mathbb{P}_\zeta^*$ attaining $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}$ is a mixture in the form $\mathbb{P}_\zeta^* = p_1^* \mathbb{P}_\zeta^1 + p_2^* \mathbb{P}_\zeta^2$, where, for $i = 1, 2$, $\mathbb{P}_\zeta^i$ is defined on the interval connecting 0 and z_i^* (i.e., $[0, z_i^*]$ or $[z_i^*, 0]$, depending on the sign of z_i^*) and $\mathbb{P}_\zeta^i\{|\zeta| \leq t|z_i^*|\} = t^\alpha$ for all $t \in [0, 1]$.

Finally, we reformulate DRCC (3.6) by analyzing problem (3.9a)–(3.9e). We summarize the main result of this section in the following theorem.

Theorem III.8. *DRCC* (3.6) *is equivalent to a set of SOC constraints*

$$\sqrt{\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}} \left|\left|\Lambda a(x)\right|\right| \leq \tau b(x) - \left(\frac{\alpha + 1}{\alpha}\right) \mu^\top a(x), \quad \forall \tau \geq \left(\frac{1}{1 - \epsilon}\right)^{1/\alpha}, \tag{3.14}$$

where $\Lambda := [(\frac{\alpha+2}{\alpha})\Sigma - (\frac{\alpha+1}{\alpha})^2\mu\mu^\top]^{1/2}$.

Proof. We analyze the solutions to problem (3.9a)–(3.9e) and identify all possible solutions (p_1, p_2, z_1, z_2) that satisfy constraints (3.9b)–(3.9e). To this end, we analyze the following two cases.

Case 1. If $\mu_0 \leq b(x)$, then we parameterize z_2 by defining $z_2 = \tau b(x)$ for $\tau \geq 1$. Accordingly, we parameterize all solutions (p_1, p_2, z_1, z_2) that satisfy constraints (3.9b)–(3.9e) by τ as follows:

$$p_1 = \frac{(\tau b(x) - \mu_0)^2}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}, \quad p_2 = \frac{\Sigma_0 - \mu_0^2}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}, \tag{3.15a}$$

$$z_1 = \mu_0 - \frac{\Sigma_0 - \mu_0^2}{\tau b(x) - \mu_0}, \text{ and } z_2 = \tau b(x). \tag{3.15b}$$

Note that, for each $\tau \geq 1$, (p_1, p_2, z_1, z_2) satisfies constraints (3.9e) because $p_1, p_2 \geq 0$ and $z_2 = \tau b(x) \geq b(x)$. Then, problem (3.9a)–(3.9e) can be recast as

$$\min_{\tau \geq 1} \quad \frac{(\tau b(x) - \mu_0)^2 + \tau^{-\alpha}(\Sigma_0 - \mu_0^2)}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}.$$

Hence, DRCC (3.6), i.e., $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\} \geq 1 - \epsilon$, can be recast as

$$\frac{(\tau b(x) - \mu_0)^2 + \tau^{-\alpha}(\Sigma_0 - \mu_0^2)}{(\tau b(x) - \mu_0)^2 + (\Sigma_0 - \mu_0^2)} \geq 1 - \epsilon, \quad \forall \tau \geq 1.$$

After simple transformations, this is equivalent to

$$(\tau b(x) - \mu_0)^2 \geq \left(\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}\right)(\Sigma_0 - \mu_0^2), \quad \forall \tau \geq 1. \tag{3.16}$$

As $(\tau b(x) - \mu_0)^2 \geq 0$, we can assume $\tau \geq (1/(1-\epsilon))^{1/\alpha}$ without loss of generality. Furthermore, because $\tau b(x) - \mu_0 \geq 0$ for all $\tau \geq 1$, we can rewrite constraints (3.16) as (3.14), using the definitions of μ_0 and Σ_0.

Case 2. If $\mu_0 > b(x)$, then we parameterize z_2 by defining $z_2 = \tau b(x)$ for $\tau \geq 1$.

For all $\tau \geq \mu_0/b(x)$, because $z_2 \geq \mu_0$, we parameterize (p_1, p_2, z_1, z_2) by τ as in (3.15a)–(3.15b). Similar to Case 1, DRCC (3.6) can be recast as

$$\tau b(x) - \mu_0 \geq \sqrt{\left(\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}\right)_+}\sqrt{\Sigma_0 - \mu_0^2}, \quad \forall \tau \geq \frac{\mu_0}{b(x)}. \tag{3.17a}$$

For all $1 \leq \tau < \mu_0/b(x)$, because $b(x) \leq z_2 < \mu_0$, we parameterize (p_1, p_2, z_1, z_2) by τ as follows:

$$\begin{aligned} p_1 =& \frac{(\mu_0 - \tau b(x))^2}{(\mu_0 - \tau b(x))^2 + \Sigma_0 - \mu_0^2}, \quad p_2 = \frac{\Sigma_0 - \mu_0^2}{(\mu_0 - \tau b(x))^2 + \Sigma_0 - \mu_0^2}, \\ z_1 =& \mu_0 + \frac{\Sigma_0 - \mu_0^2}{\mu_0 - \tau b(x)}, \text{ and } z_2 = \tau b(x). \end{aligned}$$

Then, because $\mu_0 > \tau b(x)$, DRCC (3.6) can be recast as

$$\mu_0 - \tau b(x) \geq \sqrt{\left(\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}\right)_+}\sqrt{\Sigma_0 - \mu_0^2}, \quad \forall 1 \leq \tau < \frac{\mu_0}{b(x)}. \tag{3.17b}$$

Combining inequalities (3.17a)–(3.17b) and the fact that $(1-\epsilon-\tau^{-\alpha})/\epsilon > 0$ if and only if $\tau > [1/(1-\epsilon)]^{1/\alpha}$, we have $\mu_0/b(x) \leq [1/(1-\epsilon)]^{1/\alpha}$ because otherwise, when $\tau = \mu_0/b(x)$, the left-hand side of (3.17a) equals zero while the right-hand side is strictly positive. It follows that inequalities (3.17b) are equivalent to $\mu_0 - \tau b(x) \geq 0$ for all $1 \leq \tau < \mu_0/b(x)$ and so redundant, and inequalities (3.17a) are equivalent to (3.14), using the definitions of μ_0 and Σ_0. ∎

In computation, directly replacing DRCC with constraints (3.14) involves an infinite number of SOC constraints and so is computationally intractable. An alternative approach is by separation, i.e., (i) obtain a solution $\hat{x}$ from a relaxed formulation, (ii) find a $\hat{\tau} \geq (1/(1-\epsilon))^{1/\alpha}$ such that $\hat{x}$ violates the corresponding SOC constraint (3.14), and (iii) incorporate the violated SOC constraint to strengthen the formulation. Note

that constraints (3.14) imply that

$$\tau b(x) - \Big(\frac{\alpha+1}{\alpha}\Big)\mu^\top a(x) \geq 0, \quad \forall \tau \geq \Big(\frac{1}{1-\epsilon}\Big)^{1/\alpha}$$

because $\sqrt{(1-\epsilon-\tau^{-\alpha})/\epsilon}\ ||\Lambda a(x)|| \geq 0$. These inequalities are equivalent to a single linear constraint $(1/(1-\epsilon))^{1/\alpha} b(x) - [(\alpha+1)/\alpha]\mu^\top a(x) \geq 0$, which we assume is always incorporated in the relaxed formulation in Step (i). We discuss how to efficiently conduct Step (ii) of the separation approach, which is equivalent to solving the following problem:

Separation Problem 1: Given $\hat{x}$, does there exist a $\hat{\tau} \geq (1/(1-\epsilon))^{1/\alpha}$ such that $\hat{x}$ violates constraints (3.14)?

In the following proposition, we show that Separation Problem 1 can be solved by conducting a golden section search on the real line. This search is computationally efficient.

Proposition III.9. *Define* $\hat{\mu}_0 = (\frac{\alpha+1}{\alpha})\mu^\top a(\hat{x})$ *and* $\hat{\Sigma}_0 = (\frac{\alpha+2}{\alpha})a(\hat{x})^\top \Sigma a(\hat{x})$. *We have the following:*

1. *If* $a(\hat{x}) = 0$, *then constraints* (3.14) *are always satisfied;*

2. *If* $a(\hat{x}) \neq 0$ *and* $b(\hat{x}) = 0$, *then* $\hat{x}$ *violates constraints* (3.14) *if and only if it violates them at* $\hat{\tau} = +\infty$, *i.e.,* $\sqrt{(1-\epsilon)/\epsilon}\ ||\Lambda a(\hat{x})|| \leq -[(\alpha+1)/\alpha]\mu^\top a(\hat{x})$;

3. *If* $a(\hat{x}) \neq 0$ *and* $b(\hat{x}) > 0$, *then* $\hat{x}$ *violates constraints* (3.14) *if and only if* $\sqrt{(1-\epsilon-\hat{\tau}^{-\alpha})/\epsilon}\ ||\Lambda a(\hat{x})|| \leq \hat{\tau} b(\hat{x}) - [(\alpha+1)/\alpha]\mu^\top a(\hat{x})$, *where* $\hat{\tau}$ *represents the minimizer of the one-dimensional problem*

$$\min_{\tau \geq (1/(1-\epsilon))^{1/\alpha}} \ (b(\hat{x})\tau - \hat{\mu}_0)^2 - \left(\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}\right)(\hat{\Sigma}_0 - \hat{\mu}_0^2), \tag{3.18}$$

whose objective function is strongly convex and can be minimized via a golden section search in the interval $\big[(1/(1-\epsilon))^{1/\alpha},\ \hat{\mu}_0/b(\hat{x}) + \alpha(1-\epsilon)^{(\alpha+1)/\alpha}(\hat{\Sigma}_0 -$

$\hat{\mu}_0^2)/(2\epsilon b(\hat{x})^2)\big]$.

Proof. First, if $a(\hat{x}) = 0$, then constraints (3.14) reduce to $\tau b(x) \geq 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$, which always holds due to Assumption III.5. Second, if $a(\hat{x}) \neq 0$ and $b(\hat{x}) = 0$, then constraints (3.14) reduce to

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\big|\big|\Lambda a(\hat{x})\big|\big| \leq -\left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(\hat{x}), \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}.$$

As the left-hand side of the above inequality is increasing in τ, constraints (3.14) are violated if and only if they are violated at $\hat{\tau} = +\infty$. Third, if $a(\hat{x}) \neq 0$ and $b(\hat{x}) > 0$, then constraints (3.14) are satisfied if and only if

$$\left[\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}||\Lambda a(\hat{x})||\right]^2 \leq \left[\tau b(\hat{x}) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(\hat{x})\right]^2, \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$$

because both sides of constraints (3.14) are nonnegative. By the definitions of $\hat{\mu}_0$ and $\hat{\Sigma}_0$, this is equivalent to $(b(\hat{x})\tau - \hat{\mu}_0)^2 - [(1-\epsilon-\tau^{-\alpha})/\epsilon](\hat{\Sigma}_0 - \hat{\mu}_0^2) \geq 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$. It follows that the Separation Problem 1 can be answered by checking constraints (3.14) at the optimal solution $\hat{\tau}$ of problem (3.18).

Finally, we denote the objective function of problem (3.18) as $H(\tau)$. It follows that

$$\begin{aligned} H'(\tau) &= 2b(\hat{x})(b(\hat{x})\tau - \hat{\mu}_0) - \left(\frac{\alpha}{\epsilon}\right)(\hat{\Sigma}_0 - \hat{\mu}_0^2)\tau^{-\alpha-1}, \\ H''(\tau) &= 2\big[b(\hat{x})\big]^2 + \left(\frac{\alpha^2+\alpha}{\epsilon}\right)(\hat{\Sigma}_0 - \hat{\mu}_0^2)\tau^{-\alpha-2}. \end{aligned}$$

As $H''(\tau) > 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$, $H(\tau)$ is strongly convex and can be minimized via a golden section search. More specifically, if $H'((1/(1-\epsilon))^{1/\alpha}) \geq 0$, then $(1/(1-\epsilon))^{1/\alpha}$ is optimal to problem (3.18). Otherwise, if $H'((1/(1-\epsilon))^{1/\alpha}) < 0$, then problem (3.18) is optimized at $\hat{\tau}$ such that $H'(\hat{\tau}) = 0$. It follows that $2b(\hat{x})(b(\hat{x})\hat{\tau} - \hat{\mu}_0) =$

$(\alpha/\epsilon)(\hat{\Sigma}_0 - \hat{\mu}_0^2)\hat{\tau}^{-\alpha-1}$. Since $\hat{\tau} \geq (1/(1-\epsilon))^{1/\alpha}$, we have

$$\begin{aligned} 2b(\hat{x})^2\hat{\tau} &\leq 2b(\hat{x})\hat{\mu}_0 + \left(\frac{\alpha}{\epsilon}\right)(\hat{\Sigma}_0 - \hat{\mu}_0^2)(1-\epsilon)^{(\alpha+1)/\alpha} \\ \Rightarrow \ \hat{\tau} &\leq \frac{\hat{\mu}_0}{b(\hat{x})} + \frac{\alpha(1-\epsilon)^{(\alpha+1)/\alpha}}{2\epsilon b(\hat{x})^2}(\hat{\Sigma}_0 - \hat{\mu}_0^2). \end{aligned}$$

Hence, the golden section search can be restricted to the interval $[(1/(1-\epsilon))^{1/\alpha}, \hat{\mu}_0/b(\hat{x}) + \alpha(1-\epsilon)^{(\alpha+1)/\alpha}(\hat{\Sigma}_0 - \hat{\mu}_0^2)/(2\epsilon b(\hat{x})^2)]$ without loss of optimality. ∎

3.3 Approximations of the Distributionally Robust Chance Constraint

Next, we derive relaxed and conservative approximations of DRCC (3.6) by using a finite number of SOC constraints. First, based on the exact representation (3.14) that involves all $\tau \in \left[[1/(1-\epsilon)]^{1/\alpha}, \infty\right)$, we obtain a relaxed approximation by only involving a finite number of τ. We summarize this approximation in the following proposition, whose proof is immediate and so omitted.

Proposition III.10. *For given integer $K \geq 1$ and real numbers $[1/(1-\epsilon)]^{1/\alpha} \leq n_1 < n_2 < \cdots < n_K \leq \infty$, DRCC (3.6) implies the SOC constraints*

$$\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{\epsilon}}\left|\left|\Lambda a(x)\right|\right| \leq n_k b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall k = 1, \ldots, K. \tag{3.19}$$

Second, we obtain a conservative approximation by approximating the left-hand sides of the inequalities (3.14) by using a piece-wise linear function of τ.

Proposition III.11. *Given integer $K \geq 2$ and real numbers $[1/(1-\epsilon)]^{1/\alpha} = n_1 < n_2 < \cdots < n_K = \infty$, we define a piece-wise linear function containing $(K-1)$ pieces:*

$$g(\tau) \ = \ \min_{k=2,\ldots,K}\left\{\sqrt{\frac{1}{\epsilon(1-\epsilon-n_k^{-\alpha})}}\left[\left(\frac{\alpha n_k^{-\alpha-1}}{2}\right)\tau + 1 - \epsilon - \left(1+\frac{\alpha}{2}\right)n_k^{-\alpha}\right]\right\}.$$

Then, $g(\tau) \geq \sqrt{(1-\epsilon-\tau^{-\alpha})/\epsilon}$ *for all* $\tau \geq [1/(1-\epsilon)]^{1/\alpha}$*. Furthermore, denote* $m_1 = [1/(1-\epsilon)]^{1/\alpha}$ *and let* $m_2 < \cdots < m_{K-1}$ *represent the* $(K-2)$ *breakpoints of function* $g(\tau)$*, i.e.,*

$$m_k = \frac{(1-\epsilon)\left(1-\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}}\right)+\left(1+\frac{\alpha}{2}\right)\left(n_{k+1}^{-\alpha}\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}}-n_k^{-\alpha}\right)}{\left(\frac{\alpha}{2}\right)\left(n_{k+1}^{-\alpha-1}\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}}-n_k^{-\alpha-1}\right)},$$

$\forall k = 2, \ldots, K-1.$

Then, DRCC (3.6) *is implied by the SOC constraints*

$$g(m_k)\big|\big|\Lambda a(x)\big|\big| \leq m_k b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall k = 1, \ldots, K-1. \tag{3.20}$$

Proof. Denote $h(\tau) = \sqrt{(1-\epsilon-\tau^{-\alpha})/\epsilon}$. Then, the first derivative

$$h'(\tau) = \left(\frac{\alpha\tau^{-\alpha-1}}{2}\right)\sqrt{\frac{1}{\epsilon(1-\epsilon-\tau^{-\alpha})}}$$

and the tangent of $h(\tau)$ at n_k is

$$\sqrt{\frac{1}{\epsilon(1-\epsilon-n_k^{-\alpha})}}\left[\left(\frac{\alpha n_k^{-\alpha-1}}{2}\right)\tau + 1 - \epsilon - \left(1+\frac{\alpha}{2}\right)n_k^{-\alpha}\right]$$

for all $k = 2, \ldots, K$. It follows that $g(\tau) \geq h(\tau)$ for all $\tau \geq [1/(1-\epsilon)]^{1/\alpha}$ because $h(\tau)$ is concave on the interval $\big[[1/(1-\epsilon)]^{1/\alpha}, \infty\big)$. Hence, DRCC (3.6) is implied by

$$g(\tau)\big|\big|\Lambda a(x)\big|\big| \leq \tau b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall \tau \geq [1/(1-\epsilon)]^{1/\alpha}. \tag{3.21}$$

Furthermore, given x, as the left-hand side of (3.21) is piece-wise linear in τ and the right-hand side of (3.21) is linear in τ, inequalities (3.21) hold if and only if they hold at the breakpoints of $g(\tau)$. Therefore, DRCC (3.6) is implied by constraints

(3.20). □

Remark III.12. In computation, we can use the conservative approximation (3.20) to find near-optimal solutions. More specifically, suppose that we employ the separation approach to solve problem $\min\{c(x) : x \in X,\ x \text{ satisfies (3.6)}\}$ and have finished the first K iterations. Then, from these iterations, we obtain a lower bound c_{L}^K of the optimal objective value and $\hat{\tau}_1, \ldots, \hat{\tau}_K$ by iteratively solving Separation Problem 1. By letting $n_1 = [1/(1-\epsilon)]^{1/\alpha}$, $n_{K+2} = \infty$, and $n_k = \hat{\tau}_{k-1}$ for all $k = 2, \ldots, K+1$, we obtain an upper bound c_{U}^K of the optimal objective value by solving problem $\min\{c(x) :\ x \in X,\ x \text{ satisfies (3.20) based on } n_1, \ldots, n_{K+2}\}$, whose optimal solution is denoted x_K^*. If $(c_{\mathrm{U}}^K - c_{\mathrm{L}}^K)/c_{\mathrm{L}}^K$ is small enough, then we can stop the iterations and output x_K^* as a near-optimal solution.

3.4 Distributionally Robust Risk Constraint

To recast DRRC (3.7) as SOC constraints, we adopt a similar method to that described in Section 3.2. Again, we project random vector ξ on $\mathbb{R}$ and consider a one-dimensional random variable ζ. We summarize this result in the following proposition and omit the proof due to its similarity to that of Proposition III.3.

Proposition III.13. *The following equality holds:*

$$\sup_{\mathbb{P}_\xi \in \mathcal{D}_\xi} CVaR^{\epsilon}_{\mathbb{P}_\xi}(a(x)^\top \xi) \ = \ \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} CVaR^{\epsilon}_{\mathbb{P}_\zeta}(\zeta).$$

We compute $\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta)$ by observing that $\mathbb{P}_\zeta$ is α-unimodal and so there exists a random variable Z such that $\zeta = U^{1/\alpha} Z$, where U is uniform in $(0, 1)$ and independent of Z (see Theorem 3.5 in [23]). We summarize this computation in the following proposition, and note that it can also be obtained by following Theorem 2.1 in [75].

Proposition III.14. *The following equality holds:*

$$\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} CVaR^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \;=\; \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \right\},$$

where $\mathcal{D}(\mu_0, \Sigma_0) := \{\mathbb{P}_Z \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0\}$ *and* $f(Z) = \mathbb{1}[\beta < 0] f_-(Z) + \mathbb{1}[\beta \geq 0] f_+(Z)$*, where*

$$f_+(z) \;=\; \begin{cases} 0 & \text{if } z \leq \beta \\ \left(\frac{\alpha}{\alpha+1}\right) z - \beta + \left(\frac{\beta}{\alpha+1}\right) \left(\frac{\beta}{z}\right)^{\alpha} & \text{if } z > \beta \end{cases},$$

$$f_-(z) \;=\; \begin{cases} -\left(\frac{\beta}{\alpha+1}\right) \left(\frac{\beta}{z}\right)^{\alpha} & \text{if } z < \beta \\ \left(\frac{\alpha}{\alpha+1}\right) z - \beta & \text{if } z \geq \beta \end{cases}.$$

Proof. First, based on the definition of CVaR, we have

$$\begin{aligned} \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \;&=\; \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \inf_{\beta} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\} \\ &=\; \inf_{\beta} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\}. \end{aligned} \tag{3.22}$$

To justify the switch of $\inf_\beta$ and $\sup_{\mathbb{P}_\zeta}$ in (3.22), we observe that $\beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+$ is convex in β and concave (actually affine) in $\mathbb{P}_\zeta$. Additionally, we claim that $\beta \in [\mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}, \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}]$, i.e., β belongs to a compact set, without loss of optimality. Then, the switch follows from the Sion's minimax theorem (see [85]). To prove this claim, we observe that

$$\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \;\leq\; \mathrm{argmin}_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\} \;\leq\; \mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta)$$

for all $\mathbb{P}_\zeta \in \mathcal{D}_1$, where $\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) := \inf\{\beta : \mathbb{P}_\zeta\{\zeta \leq \beta\} \geq 1 - \epsilon\}$ and $\mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta) := \inf\{\beta : \mathbb{P}_\zeta\{\zeta \leq \beta\} > 1 - \epsilon\}$ (see Theorem 10 in [82]). It follows that we can

assume $\beta \in [\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta), \mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta)]$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ without loss of optimality. But $\mathcal{D}_1 \subseteq \mathcal{D}_1^\infty := \{\mathbb{P}_\zeta \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1,\ \mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1\}$, and it follows from Cantelli's inequality that

$$\begin{aligned}
&\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\Big\{\zeta \le \mu_1 + \sqrt{\Big(\frac{2-\epsilon}{\epsilon}\Big)(\Sigma_1 - \mu_1^2)}\Big\} \\
&\qquad \ge \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1^\infty} \mathbb{P}_\zeta\Big\{\zeta \le \mu_1 + \sqrt{\Big(\frac{2-\epsilon}{\epsilon}\Big)(\Sigma_1 - \mu_1^2)}\Big\} \ \ge\ 1 - \frac{\epsilon}{2}.
\end{aligned}$$

Hence, $\mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta) \le \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ because otherwise there exists a $\mathbb{P}_\zeta \in \mathcal{D}_1$ such that $\mathbb{P}_\zeta\{\zeta \le \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}\} \le 1 - \epsilon$, which contradicts $\mathbb{P}_\zeta\{\zeta \le \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}\} \ge 1 - \epsilon/2$. Similarly, application of Cantelli's inequality gives us

$$\begin{aligned}
&\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\Big\{\zeta \le \mu_1 - \sqrt{\Big(\frac{1+\epsilon}{1-\epsilon}\Big)(\Sigma_1 - \mu_1^2)}\Big\} \\
&\qquad \le \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1^\infty} \mathbb{P}_\zeta\Big\{\zeta \le \mu_1 - \sqrt{\Big(\frac{1+\epsilon}{1-\epsilon}\Big)(\Sigma_1 - \mu_1^2)}\Big\} \ \le\ 1 - \frac{1+\epsilon}{2}.
\end{aligned}$$

Hence, $\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \ge \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ because otherwise there exists a $\mathbb{P}_\zeta \in \mathcal{D}_1$ such that $\mathbb{P}_\zeta\{\zeta \le \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}\} \ge 1 - \epsilon$, which contradicts $\mathbb{P}_\zeta\{\zeta \le \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}\} \le 1 - (1+\epsilon)/2$.

Second, based on the representation $\zeta = U^{1/\alpha} Z$ (see Theorem 3.5 in [23]), we obtain that $\mathbb{E}_{\mathbb{P}_Z}[Z] = (\frac{\alpha+1}{\alpha})\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_0$, $\mathbb{E}_{\mathbb{P}_Z}[Z^2] = (\frac{\alpha+2}{\alpha})\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_0$, and

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \ &=\ \mathbb{E}_{\mathbb{P}_Z}[U^{1/\alpha} Z - \beta]_+ \\
&=\ \int_{z=-\infty}^{+\infty} \int_{u=0}^{1} \big[u^{1/\alpha} z - \beta\big]_+ \, du \, d\mathbb{P}_Z(z).
\end{aligned}$$

It follows that, when $\beta < 0$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ &= \int_{z=-\infty}^{\beta} \int_{u=0}^{(\beta/z)^\alpha} \left(u^{1/\alpha} z - \beta\right) du \; d\mathbb{P}_Z(z) \\
&\quad + \int_{z=\beta}^{+\infty} \int_{u=0}^{1} \left(u^{1/\alpha} z - \beta\right) du \; d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{\beta} \left(-\frac{1}{\alpha+1}\right) \left(\frac{\beta^{\alpha+1}}{z^\alpha}\right) d\mathbb{P}_Z(z) \\
&\quad + \int_{z=\beta}^{+\infty} \left[\left(\frac{\alpha}{\alpha+1}\right) z - \beta\right] d\mathbb{P}_Z(z) \\
&= \mathbb{E}_{\mathbb{P}_Z}[f_-(Z)],
\end{aligned}$$

and, when $\beta \geq 0$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ &= \int_{z=\beta}^{+\infty} \int_{u=(\beta/z)^\alpha}^{1} \left(u^{1/\alpha} z - \beta\right) du \; d\mathbb{P}_Z(z) \\
&= \int_{z=\beta}^{+\infty} \left[\left(\frac{\alpha}{\alpha+1}\right) z - \beta + \left(\frac{1}{\alpha+1}\right) \left(\frac{\beta^{\alpha+1}}{z^\alpha}\right)\right] d\mathbb{P}_Z(z) \\
&= \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)]. \qquad \blacksquare
\end{aligned}$$

Proposition III.14 indicates that computing $\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathrm{CVaR}^\epsilon_{\mathbb{P}_\zeta}(\zeta)$ can be difficult because it needs to evaluate the worst-case expectation of a nonlinear function $f(z)$, i.e., $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]$. To obtain a computable form, we first present two structural properties of $f(z)$. Lemma III.15 proposes two approximations of $f(z)$ from above (termed $f_{\mathrm{U}}(z)$) and below (termed $f_{\mathrm{L}}(z)$), respectively. Both $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$ are convex and consist of two linear pieces. Furthermore, Lemma III.16 represents convex functions $f_+(z)$ and $f_-(z)$ by the supporting hyperplanes of their epigraphs.

Lemma III.15. *Define* $f_U(z) = \left(\frac{\alpha}{\alpha+1}\right)(z-\beta)_+ + \left(\frac{1}{\alpha+1}\right)(-\beta)_+$ *and* $f_L(z) = \left[\left(\frac{\alpha}{\alpha+1}\right) z - \beta\right]_+$. *Then,* $f_L(z) \leq f(z) \leq f_U(z)$ *for all* $z \in \mathbb{R}$.

Proof. First, we prove $f_{\mathrm{L}}(z) \leq f(z)$ by discussing the following four cases:

1. If $z < \beta < 0$, then $0 \leq (\beta/z) \leq 1$ and $(-\beta) \geq 0$. It follows that $f(z) =$

$-(\beta/(\alpha+1))\,(\beta/z)^{\alpha} \geq 0$. Additionally, define $H(z) := -(\beta/(\alpha+1))\,(\beta/z)^{\alpha}$ and then $H(z)$ is a convex function of z on interval $(-\infty, \beta]$. It follows that $H(z) \geq H'(\beta)(z-\beta) + H(\beta)$, i.e.,

$$-\left(\frac{\beta}{\alpha+1}\right)\left(\frac{\beta}{z}\right)^{\alpha} \;\geq\; \left(\frac{\alpha}{\alpha+1}\right)(z-\beta) + \left(-\frac{\beta}{\alpha+1}\right) \;=\; \left(\frac{\alpha}{\alpha+1}\right)z - \beta,$$

where the inequality is because $H'(z) = (\alpha/(\alpha+1))(\beta/z)^{\alpha+1}$ and $H(\beta) = (-\beta/(\alpha+1))$. Hence, $-(\beta/(\alpha+1))\,(\beta/z)^{\alpha} \geq [(\frac{\alpha}{\alpha+1})z - \beta]_+$, i.e., $f(z) \geq f_{\mathrm{L}}(z)$.

2. If $\beta < 0$ and $z \geq \beta$, then $(\frac{\alpha}{\alpha+1})z - \beta \geq 0$. It follows that $f_{\mathrm{L}}(z) = (\frac{\alpha}{\alpha+1})z - \beta = f(z)$.

3. If $\beta \geq 0$ and $z \leq \beta$, then $(\frac{\alpha}{\alpha+1})z - \beta < 0$. It follows that $f_{\mathrm{L}}(z) = 0 = f(z)$.

4. If $z > \beta \geq 0$, then $(\beta/z) \geq 0$. It follows that $f(z) = (\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq (\frac{\alpha}{\alpha+1})z - \beta$. Additionally, as $-z < -\beta \leq 0$, from Case 1 we have

$$-\left(\frac{-\beta}{\alpha+1}\right)\left(\frac{-\beta}{-z}\right)^{\alpha} \geq \left(\frac{\alpha}{\alpha+1}\right)(-z) - (-\beta).$$

In other words, $(\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq 0$. Hence, $(\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq [(\frac{\alpha}{\alpha+1})z - \beta]_+$, i.e., $f(z) \geq f_{\mathrm{L}}(z)$.

Second, we prove $f(z) \leq f_{\mathrm{U}}(z)$ by discussing the following four cases:

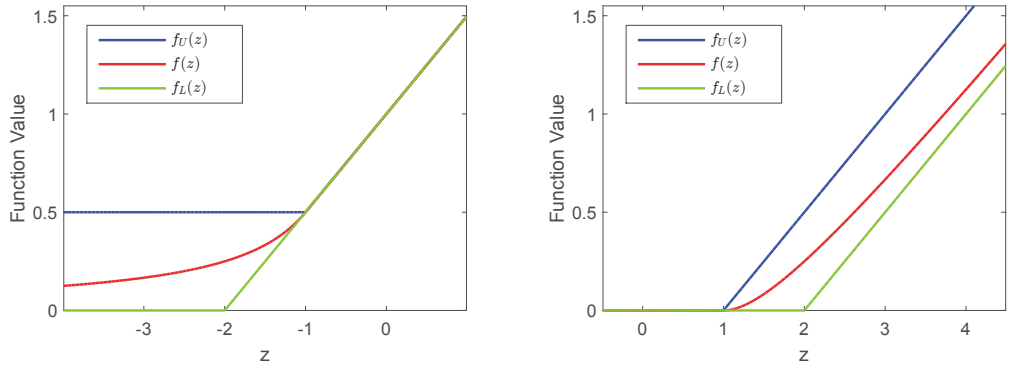
1. If $z < \beta < 0$, then $0 \leq (\beta/z) \leq 1$ and $(-\beta) \geq 0$. It follows that $f(z) = -(\beta/(\alpha+1))\,(\beta/z)^{\alpha} \leq (\frac{1}{\alpha+1})(-\beta) \leq f_{\mathrm{U}}(z)$.

2. If $\beta < 0$ and $z \geq \beta$, then $(z-\beta)_+ = z - \beta$ and $(-\beta)_+ = -\beta$. It follows that $f_{\mathrm{U}}(z) = (\frac{\alpha}{\alpha+1})(z-\beta)_+ + (\frac{1}{\alpha+1})(-\beta)_+ = (\frac{\alpha}{\alpha+1})z - \beta = f(z)$.

3. If $\beta \geq 0$ and $z \leq \beta$, then $f(z) = 0 \leq f_{\mathrm{U}}(z)$.

4. If $z > \beta \geq 0$, then $0 \leq (\beta/z) < 1$ and $(z-\beta)_+ = z - \beta$. It follows that $f(z) = (\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \leq (\alpha/(\alpha+1))z - \beta + \beta/(\alpha+1) = f_{\mathrm{U}}(z)$.

With these two discussions, the proof is complete. ∎



(a) $\beta = -1$ and $\alpha = 1$ (b) $\beta = 1$ and $\alpha = 1$

Figure 3.1: Examples of function $f(z)$ and its approximations $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$

Figs. 3.1a and 3.1b present examples of function $f(z)$ and its approximations $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$.

Lemma III.16. *The following two equalities hold:*

$$f_+(z) \;=\; \sup_{k\geq 1}\Big\{\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})z-(1-k^{-\alpha})\beta\Big\} \tag{3.23a}$$

when $\beta \geq 0$, *and*

$$f_-(z) \;=\; \sup_{k\geq 1}\Big\{\Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}z-k^{-\alpha}\beta\Big\}. \tag{3.23b}$$

when $\beta \leq 0$. *Furthermore,* $f_-(z) = f_L(z) \leq f_+(z)$ *for all* $z \in \mathbb{R}$ *when* $\beta \geq 0$ *and* $f_+(z) = f_L(z) \leq f_-(z)$ *for all* $z \in \mathbb{R}$ *when* $\beta \leq 0$.

Proof. First, we suppose that $\beta \geq 0$ and pick a $z_0 \geq \beta$. The first derivative of $f_+(z)$ at z_0 is $f'_+(z)\big|_{z=z_0} = (\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]$. It follows that the supporting hyperplane of the epigraph $\{(y,z)\in\mathbb{R}^2 : y \geq f_+(z)\}$ at z_0 is $y \geq (\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z-\big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta$. Hence, $f_+(z) = \sup_{z_0\geq\beta}\big\{(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z-\big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta\big\}$ for all $z \geq \beta$ because $f_+(z)$ is convex. Furthermore, as $f_+(z) = 0$ when $z \leq \beta$ and $(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z-\big[1-$

$(\frac{\beta}{z_0})^\alpha\big]\beta = 0$ when $z_0 = \beta$, we have $f_+(z) = \sup_{z_0 \geq \beta}\big\{(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z - \big[1-(\frac{\beta}{z_0})^\alpha\big]\beta\big\}$ for all $z \in \mathbb{R}$. Rewriting $z_0 = k\beta$ for $k \geq 1$ leads to representation (3.23a). The proof of representation (3.23b) is similar and so omitted.

Second, we suppose that $\beta \geq 0$ and define $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta$ for all $k \geq 1$. Then, $f_+(z) = \sup_{k\geq 1}\{f_+^k(z)\}$ and $f_-(z) = \sup_{k\geq 1}\{(\frac{\alpha}{\alpha+1})z - \beta - f_+^k(z)\} = (\frac{\alpha}{\alpha+1})z - \beta - \inf_{k\geq 1}\{f_+^k(z)\}$. We prove that $\inf_{k\geq 1}\{f_+^k(z)\} = -\big[(\frac{\alpha}{\alpha+1})z - \beta\big]_-$ by discussing the following two cases:

1. When $z \leq (\frac{\alpha+1}{\alpha})\beta$, we have $z \leq (\frac{\alpha+1}{\alpha})k\beta$ as $k \geq 1$ and $\beta \geq 0$. It follows that $(\frac{\alpha}{\alpha+1})(-k^{-\alpha-1})z + k^{-\alpha}\beta \geq 0$ and so $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta \geq (\frac{\alpha}{\alpha+1})z - \beta$ for all $k \geq 1$. Hence, $\inf_{k\geq 1}\{f_+^k(z)\} \geq (\frac{\alpha}{\alpha+1})z - \beta$. In addition, by letting $k \to +\infty$, we have $f_+^k(z) \to (\frac{\alpha}{\alpha+1})z - \beta$. Therefore, $\inf_{k\geq 1}\{f_+^k(z)\} = (\frac{\alpha}{\alpha+1})z - \beta$ when $z \leq (\frac{\alpha+1}{\alpha})\beta$.

2. When $z \geq (\frac{\alpha+1}{\alpha})\beta$, we have $(1-k^{-\alpha-1})z \geq (1-k^{-\alpha})(\frac{\alpha+1}{\alpha})\beta$ because $\beta \geq 0$ and $1-k^{-\alpha-1} \geq 1-k^{-\alpha} \geq 0$. It follows that $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta \geq 0$ for all $k \geq 1$. Hence, $\inf_{k\geq 1}\{f_+^k(z)\} \geq 0$. In addition, by letting $k = 1$, we have $f_+^k(z) = 0$. Therefore, $\inf_{k\geq 1}\{f_+^k(z)\} = 0$ when $z \leq (\frac{\alpha+1}{\alpha})\beta$.

It follows that $f_-(z) = (\frac{\alpha}{\alpha+1})z - \beta + \big[(\frac{\alpha}{\alpha+1})z - \beta\big]_- = \big[(\frac{\alpha}{\alpha+1})z - \beta\big]_+$. Hence, by Lemma III.15, $f_-(z) = f_{\mathrm{L}}(z) \leq f_+(z)$ for all $z \in \mathbb{R}$ when $\beta \geq 0$. The proof of $f_+(z) = f_{\mathrm{L}}(z) \leq f_-(z)$ when $\beta \leq 0$ is similar and so omitted. ∎

We are now ready to derive a reformulation of the worst-case expectation $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]$. We summarize this result in the following theorem.

Theorem III.17. *For $\beta \in \mathbb{R}$, $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] = \frac{1}{2}\max\{E_+, E_-\}$, where*

$$\begin{aligned} S_{k,\mu_0,\Sigma_0,\beta} &= \Bigg\{ \bigg[(1-k^{-\alpha})\beta - \Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\mu_0 \bigg]^2 \\ &\qquad + \Big(\frac{\alpha}{\alpha+1}\Big)^2 (1-k^{-\alpha-1})^2(\Sigma_0 - \mu_0^2) \Bigg\}^{1/2}, \\ E_+ &= \sup_{k\geq 1} \bigg\{ S_{k,\mu_0,\Sigma_0,\beta} - (1-k^{-\alpha})\beta + \Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\mu_0 \bigg\}, \quad \textit{and} \end{aligned} \tag{3.24a}$$

$$E_- = \sup_{k\geq 1} \bigg\{ S_{k,\mu_0,\Sigma_0,\beta} - (1+k^{-\alpha})\beta + \Big(\frac{\alpha}{\alpha+1}\Big)(1+k^{-\alpha-1})\mu_0 \bigg\}. \tag{3.24b}$$

Proof. To avoid clutter, throughout this proof, we assume that $\Sigma_0 > \mu_0^2$ and $\beta \neq 0$. The degenerate cases with $\Sigma_0 = \mu_0^2$ or $\beta = 0$ can be easily verified. First, we suppose that $\beta > 0$ and define $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta$ for $k \geq 1$. Then, $f(Z) = f_+(Z)$ by Proposition III.14 and $f_+(z) = \sup_{k\geq 1}\{f_+^k(z)\}$ by Lemma III.16. It follows that $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}]$. We make the following observation on switching the order of two supremum operators.

Observation 1. For $\beta \in \mathbb{R}$, we have

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\bigg[\sup_{k\geq 1}\Big\{f_+^k(Z)\Big\}\bigg] = \sup_{k\geq 1}\bigg\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\Big[f_+^k(Z)\Big]_+\bigg\}.$$

Proof of Observation 1: First, for all $k \geq 1$, it is clear that $\sup_{k\geq 1}\{f_+^k(Z)\} \geq [f_+^k(Z)]_+$ because $\sup_{k\geq 1}\{f_+^k(z)\} = f_+(z) \geq 0$ for all $z \in \mathbb{R}$. It follows that

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] \geq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$$

. We now show the opposite, i.e., $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}$

$\{f_+^k(Z)\}] \leq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$. When $\beta \leq 0$, this holds because

$$\begin{aligned}\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k\geq 1}\left\{f_+^k(Z)\right\}\right] &= \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] \\ &= \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_+^k(Z)\right]_+\right\} \\ &\leq \sup_{k\geq 1}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_+^k(Z)\right]_+\right\},\end{aligned}$$

where the first equality follows from Lemma III.16. To prove the second equality, we make the following observation on the monotonicity of function $[f_+^k(z)]_+$ in k and relegate the proof to Section 3.8.0.1.

Observation 2. $[f_+^{k+1}(z)]_+ \geq [f_+^k(z)]_+$ for all $z \in \mathbb{R}$ and $k \geq 1$.

By Observation 2, $f_{\mathrm{L}}(z) = \lim_{k\to\infty}[f_+^k(z)]_+$ for all $z \in \mathbb{R}$. It follows that, for any $\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)$,

$$\mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] \;=\; \mathbb{E}_{\mathbb{P}_Z}\left[\lim_{k\to\infty}[f_+^k(Z)]_+\right] \;=\; \lim_{k\to\infty}\mathbb{E}_{\mathbb{P}_Z}\left[[f_+^k(Z)]_+\right],$$

where the second equality follows from the monotone convergence theorem. Hence,

$$\mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] \leq \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\}.$$

As this inequality holds for all $\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)$, we have

$$\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \leq \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\}.$$

On the other hand, as $\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \geq \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+$ for all

$k \geq 1$, we have

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \geq \lim_{k \to \infty} \left\{ \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+ \right\},$$

which proves the second equality. Hence, we focus on the case when $\beta > 0$ in the remainder of this proof.

Second, we present $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k \geq 1}\{f_+^k(Z)\}]$ as the following optimization problem:

$$\begin{aligned} (\mathrm{P}): \quad v_P \;=\; \max_{\mathbb{P}_Z} \;& \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)] \\ \text{s.t.} \;& \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[1] = 1, \end{aligned}$$

whose dual is

$$\begin{aligned} (\mathrm{D}): \quad v_D \;=\; \min_{p,q,r} \;& \mu_0 p + \Sigma_0 q + r \\ \text{s.t.} \;& qz^2 + pz + r \geq f_+(z), \quad \forall z \in \mathbb{R}. \end{aligned} \tag{3.25}$$

Strong duality holds between (P) and (D) due to Assumption III.4 (see Proposition 3.4 in [84]), i.e., $v_P = v_D$. Furthermore, by Lemma 3.1 in [84], there exists a worst-case probability distribution (i.e., an optimal solution to (P)) with a finite support of at most 3 points. That is, there exists $m \in \{1, 2, 3\}$, $(z_1^*, \ldots, z_m^*) \in \mathbb{R}^m$, and $(\pi_1^*, \ldots, \pi_m^*) \in \mathbb{R}_+^m$ such that $\sum_{i=1}^m \pi_i^* z_i^* = \mu_0$, $\sum_{i=1}^m \pi_i^* (z_i^*)^2 = \Sigma_0$, and $\sum_{i=1}^m \pi_i^* = 1$. Denoting an optimal solution to (D) by (p^*, q^*, r^*), we claim that $q^*(z_i^*)^2 + p^* z_i^* + r^* = f_+(z_i^*)$ for all $i = 1, \ldots, m$, i.e., constraint (3.25) holds at equality at points $z_1^*, \ldots, z_m^*$.

Indeed, if this claim fails to hold, then we have

$$v_P \;=\; \sum_{i=1}^{m} \pi_i^* f_+(z_i^*) \;<\; \sum_{i=1}^{m} \pi_i^* [q^*(z_i^*)^2 + p^* z_i^* + r^*] \;=\; q^*\Sigma_0 + p^*\mu_0 + r^* \;=\; v_D, \quad (3.26)$$

where the inequality follows from constraint (3.25), and the second equality follows from the definitions of $(z_1^*, \ldots, z_m^*)$ and $(\pi_1^*, \ldots, \pi_m^*)$. As inequality (3.26) violates the strong duality, the claim holds. In addition, it can be shown that $f_+(z)$ and any quadratic function $qz^2 + pz + r$ satisfying constraint (3.25) intersect at most once in interval $(-\infty, \beta]$ and at most once in interval $[\beta, \infty)$. It follows that $m \leq 2$, and so $m = 2$ because $\Sigma_0 > \mu_0^2$. Without loss of generality, we assume that $z_1^* \in (-\infty, \beta]$ and $z_2^* \in [\beta, \infty)$.

Third, we define $k^* = z_2^*/\beta$ and consider function $[f_+^{k^*}(z)]_+$ that is tangent to $f_+(z)$ at z_1^* and z_2^* by Lemma III.16. Hence, $qz^2 + pz + r \geq [f_+^{k^*}(z)]_+$ for all $z \in \mathbb{R}$ with equality holding only at z_1^* and z_2^*. Consider the primal and dual formulations of $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+$ as follows:

$$\begin{aligned} (\mathrm{P}_{k^*}): \quad v_P^{k^*} \;=\; \max_{\mathbb{P}_Z} \;& \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+ \\ \text{s.t.} \;& \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[1] = 1, \end{aligned}$$

$$\begin{aligned} (\mathrm{D}_{k^*}): \quad v_D^{k^*} \;=\; \min_{p,q,r} \;& \mu_0 p + \Sigma_0 q + r \\ \text{s.t.} \;& qz^2 + pz + r \geq [f_+^{k^*}(z)]_+, \quad \forall z \in \mathbb{R}. \end{aligned}$$

It is clear that the pair (z_1^*, z_2^*) and (π_1^*, π_2^*) provide a primal feasible solution to (P_{k^*}), and (p^*, q^*, r^*) is a dual feasible solution to (D_{k^*}) because $f_+(z) \geq [f_+^{k^*}(z)]_+$

for all $z \in \mathbb{R}$. Meanwhile, these two solutions share the same objective function value because $\sum_{i=1}^{2} \pi_i^* [f_+^{k^*}(z_i^*)]_+ = \sum_{i=1}^{2} \pi_i^* f_+(z_i^*) = \mu_0 p^* + \Sigma_0 q^* + r^*$, where the first equality follows from the definition of $[f_+^{k^*}(z)]_+$ and the second equality is due to $v_P = v_D$. It follows that strong duality holds between (P_{k^*}) and (D_{k^*}) and $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+$. Therefore, $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] \leq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$ and so the proof is completed. ∎

(Proof of Theorem III.17 continued) By Observation 1, we have

$$
\begin{aligned}
&\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \\
= \;& \sup_{k\geq 1}\left\{ \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\Big[\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})Z - (1-k^{-\alpha})\beta\Big]_+ \right\} \\
= \;& \sup_{k\geq 1}\left\{ \Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1}) \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[Z - \Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big)\beta \right]_+ \right\} \\
= \;& \sup_{k\geq 1}\left\{ \Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\Big(\frac{1}{2}\Big)\left[\sqrt{\Big[\Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big)\beta - \mu_0\Big]^2 + (\Sigma_0 - \mu_0^2)} \right.\right. \\
& \left.\left. - \Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big)\beta + \mu_0 \right] \right\} \qquad (3.27) \\
= \;& \frac{1}{2} E_+,
\end{aligned}
$$

where equality (3.27) follows from Observation 3 presented in Section 3.8.0.2.

Second, we suppose that $\beta < 0$. Then, $f(Z) = f_-(Z)$ by Proposition III.14. It

follows that

$$
\begin{aligned}
& \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \\
= & \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k \geq 1}\left\{\Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}Z - k^{-\alpha}\beta\right\}\right] && \text{(3.28a)} \\
= & \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k \geq 1}\left\{\max\left\{\Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}Z - k^{-\alpha}\beta,\ \Big(\frac{\alpha}{\alpha+1}\Big)Z - \beta\right\}\right\}\right] && \text{(3.28b)} \\
= & \sup_{k \geq 1}\left\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\max\left\{\Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}Z - k^{-\alpha}\beta,\ \Big(\frac{\alpha}{\alpha+1}\Big)Z - \beta\right\}\right]\right\} && \text{(3.28c)} \\
= & \sup_{k \geq 1}\left\{\Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}\mu_0 - k^{-\alpha}\beta + \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\} && \text{(3.28d)} \\
= & \frac{1}{2}E_-, && \text{(3.28e)}
\end{aligned}
$$

where equality (3.28a) follows from Lemma III.16, equality (3.28b) is because $(\frac{\alpha}{\alpha+1})k^{-\alpha-1}z - k^{-\alpha}\beta = (\frac{\alpha}{\alpha+1})z - \beta$ when $k = 1$, equality (3.28c) is parallel to Observation 1 and can be similarly proved, and equality (3.28d) follows from the definition of $f_+^k(z)$.

Finally, it remains to prove that $E_+ \geq E_-$ when $\beta > 0$ and $E_+ \leq E_-$ when $\beta < 0$. Due to the similarity of proof, we only show the former case, i.e., when $\beta > 0$. To that end, we note that the equalities (3.28b)–(3.28e) are independent of the sign of β and so still hold when $\beta > 0$. It follows that $\frac{1}{2}E_- = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_-(Z)]$ when $\beta > 0$. Similarly, we have $\frac{1}{2}E_+ = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)]$. But $f_-(z) \leq f_+(z)$ for all $z \in \mathbb{R}$ by Lemma III.16, and so $\frac{1}{2}E_- \leq \frac{1}{2}E_+$ when $\beta > 0$. ∎

Theorem III.17 leads to an equivalent reformulation of DRRC (3.7). We summarize the main result of this section in the following theorem.

Theorem III.18. *DRRC* (3.7) *is equivalent to a set of SOC constraints*

$$\left\| \begin{bmatrix} (1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x) \end{bmatrix} \right\| \leq 2\epsilon b(x) - (1-k^{-\alpha-1})\mu^\top a(x) + (1-k^{-\alpha}-2\epsilon)\beta, \tag{3.29a}$$

$$\left\| \begin{bmatrix} (1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x) \end{bmatrix} \right\| \leq 2\epsilon b(x) - (1+k^{-\alpha-1})\mu^\top a(x) + (1+k^{-\alpha}-2\epsilon)\beta, \tag{3.29b}$$

for all $k \geq 1$.

Proof. By Propositions III.13–III.14, DRRC (3.7) is equivalent to

$$\inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \right\} \leq b(x).$$

Meanwhile, the proof of Proposition III.14 shows that there exists a finite β that attains the above infimum. It follows that DRRC (3.7) is satisfied if and only if there exists a $\beta \in \mathbb{R}$ such that $\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \leq b(x)$. Then, the conclusion follows from Theorem III.17 by the definition of μ_0, Λ, and that

$$S_{k,\mu_0,\Sigma_0,\beta} = \left\| \begin{bmatrix} (1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x) \end{bmatrix} \right\|, \quad \forall k \geq 1. \quad \blacksquare$$

In computation, directly replacing DRRC with constraints (3.29a)–(3.29b) requires an infinite number of SOC constraints and is so computationally intractable. Like what we described for DRCC in Section 3.2, we adopt the separation approach and solve the following problem:

Separation Problem 2: Given $\hat{\beta}$ and $\hat{x}$, does there exist a $\hat{k}$ such that $(\hat{\beta}, \hat{x})$ violate

constraints (3.29a)–(3.29b)?

In the following proposition, we show that Separation Problem 2 can be solved by conducting a golden section search on the real line. This search is computationally efficient.

Proposition III.19. *Define* $\hat{\mu}_0 = (\frac{\alpha+1}{\alpha})\mu^\top a(\hat{x})$, $\hat{\Sigma}_0 = (\frac{\alpha+2}{\alpha})a(\hat{x})^\top \Sigma a(\hat{x})$. *We have the following:*

1. *If* $\hat{\beta} = 0$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (3.29a)–(3.29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at* $\hat{k} = \infty$;

2. *If* $\hat{\beta} \neq 0$ *and* $\hat{\Sigma}_0 = \hat{\mu}_0^2$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (3.29a)–(3.29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at* $\hat{k} = \max\{\hat{\mu}_0/\hat{\beta}, 1\}$;

3. *If* $\hat{\beta} \neq 0$ *and* $\hat{\Sigma}_0 > \hat{\mu}_0^2$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (3.29a)–(3.29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at the unique root of equation*

$$2\left[\Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big) - \mu_\beta\right] \;=\; (k-\mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)} \tag{3.30}$$

lying within the interval $\Big[1+\sqrt{(1-\mu_\beta)^2+\Gamma_\beta},\ 1+1/\alpha+\sqrt{(1-\mu_\beta+1/\alpha)^2+\Gamma_\beta}\Big]$, *where* $\mu_\beta = \hat{\mu}_0/\hat{\beta}$ *and* $\Gamma_\beta = (\hat{\Sigma}_0 - \hat{\mu}_0^2)/\hat{\beta}^2$.

Proof. For a given $(\hat{\beta}, \hat{x})$, solving Separation Problem 2 is equivalent to finding $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]$, i.e., $1/2 \max\{E_+, E_-\}$ defined in Theorem III.17. First, if $\hat{\beta} = 0$, then

$$\begin{aligned} S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}} &= \sqrt{\Big[\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0\Big]^2 + \Big(\frac{\alpha}{\alpha+1}\Big)^2(1-k^{-\alpha-1})^2(\hat{\Sigma}_0 - \hat{\mu}_0^2)} \\ &= \Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}. \end{aligned}$$

It follows that

$$\begin{aligned}\frac{1}{2}E_+ &= \frac{1}{2}\sup_{k\geq 1}\left\{\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}+\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \frac{1}{2}\sup_{k\geq 1}\left\{\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\Big(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\Big)\right\} \qquad (3.31a)\\ &= \frac{1}{2}\Big(\frac{\alpha}{\alpha+1}\Big)\Big(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\Big), \qquad (3.31b)\end{aligned}$$

where equality (3.31b) is because $\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0 \geq 0$ and so $k=\infty$ maximizes (3.31a). Additionally,

$$\begin{aligned}\frac{1}{2}E_- &= \frac{1}{2}\sup_{k\geq 1}\left\{\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}+\Big(\frac{\alpha}{\alpha+1}\Big)(1+k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \frac{1}{2}\Big(\frac{\alpha}{\alpha+1}\Big)\sup_{k\geq 1}\left\{\Big(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\Big)+k^{-\alpha-1}\Big(\hat{\mu}_0-\sqrt{\hat{\Sigma}_0}\Big)\right\} \qquad (3.31c)\\ &= \frac{1}{2}\Big(\frac{\alpha}{\alpha+1}\Big)\Big(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\Big), \qquad (3.31d)\end{aligned}$$

where equality (3.31d) is because $\hat{\mu}_0-\sqrt{\hat{\Sigma}_0}\leq 0$ and so $k=\infty$ maximizes (3.31c). Summing up the above two cases, we have $\hat{k}=\infty$ if $\hat{\beta}=0$.

Second, if $\hat{\beta}\neq 0$ and $\hat{\Sigma}_0=\hat{\mu}_0^2$, then $S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}}=|(1-k^{-\alpha})\hat{\beta}-(\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})\hat{\mu}_0|$. It follows that

$$\begin{aligned}\frac{1}{2}E_+ &= \frac{1}{2}\sup_{k\geq 1}\left\{\Big|(1-k^{-\alpha})\hat{\beta}-\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0\Big|-(1-k^{-\alpha})\hat{\beta}\right.\\ &\qquad\qquad \left.+\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \sup_{k\geq 1}\left\{\Big[\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0-(1-k^{-\alpha})\hat{\beta}\Big]_+\right\} \qquad (3.32a)\\ &= f_+(\hat{\mu}_0), \qquad (3.32b)\end{aligned}$$

where equality (3.32b) results from Lemma III.16 and so $k=\max\{\hat{\mu}_0/\hat{\beta},1\}$ maximizes

(3.32a). Meanwhile,

$$\begin{aligned}\frac{1}{2}E_- &= \frac{1}{2}\sup_{k\geq 1}\Bigg\{\Big|(1-k^{-\alpha})\hat{\beta}-\Big(\frac{\alpha}{\alpha+1}\Big)(1-k^{-\alpha-1})\hat{\mu}_0\Big|-(1+k^{-\alpha})\hat{\beta} \\ &\qquad\qquad + \Big(\frac{\alpha}{\alpha+1}\Big)(1+k^{-\alpha-1})\hat{\mu}_0\Bigg\} \\ &= \sup_{k\geq 1}\Bigg\{\max\Big\{\Big(\frac{\alpha}{\alpha+1}\Big)\hat{\mu}_0-\hat{\beta},\ \Big(\frac{\alpha}{\alpha+1}\Big)k^{-\alpha-1}\hat{\mu}_0-k^{-\alpha}\hat{\beta}\Big\}\Bigg\} \qquad (3.32c) \\ &= f_-(\hat{\mu}_0), \qquad (3.32d)\end{aligned}$$

where equality (3.32d) results from Lemma III.16 and so $k=\max\{\hat{\mu}_0/\hat{\beta},1\}$ maximizes (3.32c). Summing up the above two cases, we have $\hat{k}=\max\{\hat{\mu}_0/\hat{\beta},1\}$ if $\hat{\beta}\neq 0$ and $\hat{\Sigma}_0=\hat{\mu}_0^2$.

Third, suppose that $\hat{\beta}\neq 0$ and $\hat{\Sigma}_0>\hat{\mu}_0^2$. As the case when $\hat{\beta}<0$ can be similarly derived, we focus on the case when $\hat{\beta}>0$. In this case, solving Separation Problem 2 is equivalent to finding the maximizer of optimization problem (3.24a) that defines E_+. To this end, we let $F(k)$ represent the objective function of (3.24a), i.e., $F(k):=S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}}-(1-k^{-\alpha})\hat{\beta}+(\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})\hat{\mu}_0$. It follows that

$$F'(k) \;=\; \alpha\hat{\beta}k^{-\alpha-2}\left\{\frac{\Big[\Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big)-\mu_\beta\Big](k-\mu_\beta)+\Gamma_\beta}{\sqrt{\Big[\Big(\frac{\alpha+1}{\alpha}\Big)\Big(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\Big)-\mu_\beta\Big]^2+\Gamma_\beta}}-(k-\mu_\beta)\right\}.$$

We prove that $F(k)$ is unimodal, and in particular, $F(k)$ is nondecreasing on $[1,\hat{k}]$ and nonincreasing on $[\hat{k},\infty)$, where $\hat{k}$ represents the root of equation (3.30). The conclusion of this proposition then follows because $\hat{k}$ is the maximizer of $F(k)$ on $[1,\infty)$. To that end, it suffices to show that (i) $\lim_{k\to 1^+}F'(k)>0$, (ii) there exists a $k\in[1,\infty)$ such that $F'(k)<0$, and (iii) $\hat{k}$ is the unique root of equation $F'(k)=0$. We show (i)–(iii) as follows.

(i) As $\lim_{k\to 1^+}\{\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\} = \frac{\alpha}{\alpha+1}$ and $\Gamma_\beta > 0$, we have

$$\lim_{k\to 1^+} F'(k) = \alpha\hat{\beta}\big[\sqrt{(1-\mu_\beta)^2+\Gamma_\beta} - (1-\mu_\beta)\big] > 0$$

(ii) We have

$$\begin{aligned}\frac{F'(k)}{\alpha\hat{\beta}k^{-\alpha-2}} &= \left(\frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta\right]^2+\Gamma_\beta}}-1\right)(k-\mu_\beta)\\ &\quad+\frac{\Gamma_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta\right]^2+\Gamma_\beta}}.\end{aligned}$$

As $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}} \in \big[\frac{\alpha}{\alpha+1}, 1\big]$ and

$$\frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta\right]^2+\Gamma_\beta}}-1 \;<\; 0,$$

there exists a sufficiently large k such that $F'(k) < 0$.

(iii) We consider the roots of equation $F'(k) = 0$. As $F'(k) = 0$ is equivalent to

$$\begin{aligned}&\left(1-\frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta\right]^2+\Gamma_\beta}}\right)(k-\mu_\beta)\\ =\;&\frac{\Gamma_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)-\mu_\beta\right]^2+\Gamma_\beta}},\end{aligned}$$

any root k satisfies $k - \mu_\beta > 0$ because $\Gamma_\beta > 0$. The above equation can be further simplified to equation (3.30) and so any roots k of equation (3.30) also satisfy $F'(k) = 0$. We now prove the uniqueness of the root. We note that the first derivative of $2\big[(\frac{\alpha+1}{\alpha})(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}) - \mu_\beta\big]$, i.e., the left-hand side of equation (3.30), is always less than 1. To see this, we take the first derivative and denote

it

$$Q(k) := 2\Big(\frac{\alpha+1}{\alpha}\Big)\frac{k^{-2\alpha-2}+\alpha k^{-\alpha-1}-(\alpha+1)k^{-\alpha-2}}{(1-k^{-\alpha-1})^2}.$$

Through basic algebraic manipulations, it follows that $Q(k) \leq 1$ if and only if $\overline{Q}(k) := (\alpha+2)k^{-2\alpha-2} + (2\alpha^2+4\alpha)k^{-\alpha-1} - (2\alpha^2+4\alpha+2)k^{-\alpha-2} - \alpha \leq 0$. As $\overline{Q}(1) = 0$, it suffices to show that $\overline{Q}'(k) \leq 0$ for all $k \geq 1$. Noting that $\overline{Q}'(k) = 2(\alpha+1)(\alpha+2)k^{-2\alpha-3}[(\alpha+1)k^{\alpha} - \alpha k^{\alpha+1} - 1]$, we need to show that $\hat{Q}(k) := (\alpha+1)k^{\alpha} - \alpha k^{\alpha+1} - 1 \leq 0$ for all $k \geq 1$, which holds because $\hat{Q}(1) = 0$ and $\hat{Q}'(k) = \alpha(\alpha+1)k^{\alpha-1}(1-k) \leq 0$. Meanwhile, the first derivative of $(k-\mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)}$, i.e., the right-hand side of equation (3.30), is always greater than 1. Furthermore, $2\big[(\frac{\alpha+1}{\alpha})(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}) - \mu_\beta\big] \in [2(1-\mu_\beta), 2(\frac{\alpha+1}{\alpha} - \mu_\beta)]$, while the range of function $(k-\mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)}$ is $(-\infty, \infty)$ for $k \in (\mu_\beta, \infty)$. It follows that the two sides of equation (3.30) can meet only once, i.e., this equation has a unique root.

Finally, we provide lower and upper bounds of root $\hat{k}$. As $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}} \in \big[\frac{\alpha}{\alpha+1}, 1\big]$, we have $2(1-\mu_\beta) \leq (\hat{k}-\mu_\beta) - \frac{\Gamma_\beta}{(\hat{k}-\mu_\beta)} \leq 2(\frac{\alpha+1}{\alpha} - \mu_\beta)$. It follows that $\hat{k} \in \Big[1+\sqrt{(1-\mu_\beta)^2+\Gamma_\beta},\ 1+1/\alpha+\sqrt{(1-\mu_\beta+1/\alpha)^2+\Gamma_\beta}\Big]$. ∎

3.5 Approximations of the Distributionally Robust Risk Constraint

Next, we derive approximations of DRRC (3.7). First, in the following proposition, we present a conservative approximation based on $f_{\mathrm{U}}(z)$ and a relaxed one based on $f_{\mathrm{L}}(z)$, both of which are in the form of SOC constraints.

Proposition III.20. *DRRC* (3.7) *is implied by SOC constraints*

$$\left\| \begin{bmatrix} \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x) \\ \Lambda a(x) \end{bmatrix} \right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1\right] \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x), \tag{3.33a}$$

$$\begin{aligned} &\left\| \begin{bmatrix} \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x) \\ \Lambda a(x) \end{bmatrix} \right\| \\ &\quad \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha}\right] \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x). \end{aligned} \tag{3.33b}$$

Furthermore, DRRC (3.7) *implies SOC constraint*

$$\begin{aligned} &\left\| \begin{bmatrix} \left(\frac{\alpha+1}{\alpha}\right) \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x) \\ \Lambda a(x) \end{bmatrix} \right\| \\ &\quad \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha}\right] \beta - \left(\frac{\alpha+1}{\alpha}\right) \mu^\top a(x). \end{aligned} \tag{3.33c}$$

Proof. First, based on Propositions III.13–III.14 and Lemma III.15, DRRC (3.7) is implied by constraint $\beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{U}}(Z)] \leq b(x)$. Furthermore, we have

$$\begin{aligned} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{U}}(Z)] &= \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z} \left[\left(\frac{\alpha}{\alpha+1}\right) [Z-\beta]_+ + \left(-\frac{\beta}{\alpha+1}\right)_+ \right] \\ &= \left(-\frac{\beta}{\alpha+1}\right)_+ + \left(\frac{\alpha}{\alpha+1}\right) \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ \\ &= \left(-\frac{\beta}{\alpha+1}\right)_+ + \\ &\qquad \left(\frac{\alpha}{\alpha+1}\right) \left(\frac{1}{2}\right) \left[\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} - \beta + \mu_0 \right], \end{aligned}$$

where the last equality is due to Observation 3 presented in Section 3.8.0.2. It follows

that DRRC (3.7) is implied by

$$\begin{aligned}
&\beta + \left(\frac{1}{\epsilon}\right)\left\{\left(-\frac{\beta}{\alpha+1}\right)_+ + \left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right)\left[\sqrt{(\beta-\mu_0)^2+(\Sigma_0-\mu_0^2)} - \beta + \mu_0\right]\right\} \\
\leq & b(x) \Leftrightarrow \sqrt{(\beta-\mu_0)^2+(\Sigma_0-\mu_0^2)} \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1\right]\beta \\
& \quad - \left(\frac{2}{\alpha}\right)(-\beta)_+ - \mu_0.
\end{aligned}$$

This is equivalent to constraints (3.33a)–(3.33b) by the definition of μ_0 and observing that

$$\sqrt{(\beta-\mu_0)^2+(\Sigma_0-\mu_0^2)} \;=\; \left\| \begin{bmatrix} \beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x) \\ \Lambda a(x) \end{bmatrix} \right\|.$$

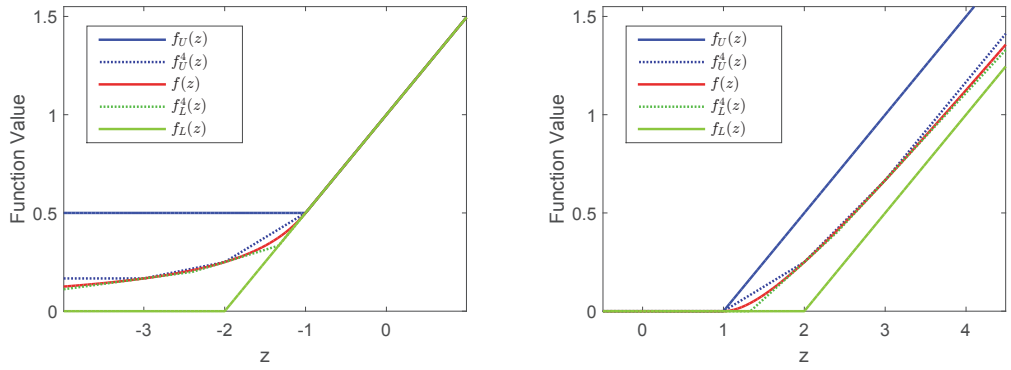
Second, based on Propositions III.13–III.14 and Lemma III.15, DRRC (3.7) implies constraint

$\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \leq b(x)$. Furthermore, we have

$$\begin{aligned}
&\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \;=\; \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\left(\frac{\alpha}{\alpha+1}\right)Z - \beta\right]_+ \\
&= \left(\frac{\alpha}{\alpha+1}\right) \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[Z - \left(\frac{\alpha+1}{\alpha}\right)\beta\right]_+ \\
&= \left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right)\left[\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta - \mu_0\right)^2 + (\Sigma_0-\mu_0^2)} - \left(\frac{\alpha+1}{\alpha}\right)\beta + \mu_0\right],
\end{aligned}$$

where the last equality is due to Observation 3. It follows that DRRC (3.7) implies

$$\begin{aligned}
&\beta + \left(\frac{1}{\epsilon}\right)\left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right)\left[\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta - \mu_0\right)^2 + (\Sigma_0-\mu_0^2)} - \left(\frac{\alpha+1}{\alpha}\right)\beta + \mu_0\right] \\
&\leq b(x) \Leftrightarrow \sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta - \mu_0\right)^2 + (\Sigma_0-\mu_0^2)} \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) \\
&\qquad - \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha}\right]\beta - \mu_0.
\end{aligned}$$



(a) $\beta = -1$ and $\alpha = 1$ (b) $\beta = 1$ and $\alpha = 1$

Figure 3.2: K-piece approximations of $f(z)$ with $K = 4$, $n_1 = 1$, $n_2 = 2$, $n_3 = 3$, and $n_4 = \infty$

This is equivalent to constraints (3.33c) by the definition of μ_0 and observing that

$$\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta - \mu_0\right)^2 + (\Sigma_0 - \mu_0^2)} = \left\| \begin{bmatrix} \left(\frac{\alpha+1}{\alpha}\right)\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x) \\ \Lambda a(x) \end{bmatrix} \right\|. \blacksquare$$

Second, we derive tighter approximations of DRRC (3.7) based on tighter approximations of function $f(z)$. Note that both $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$ approximate $f(z)$ based on two linear pieces (see Figs. 3.2a–3.2b). We generalize $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$ by defining K-piece approximations as follows.

Definition III.21. Given integer $K \geq 3$ and real numbers $1 = n_1 < n_2 < \cdots < n_K = \infty$, we define $f_{\mathrm{U}}^K(z) = \mathbb{1}[\beta < 0] f_{\mathrm{U}-}^K(z) + \mathbb{1}[\beta \geq 0] f_{\mathrm{U}+}^K(z)$ and $f_{\mathrm{L}}^K(z) = \mathbb{1}[\beta <$

$0]f^K_{\mathrm{L}-}(z) + \mathbb{1}[\beta \geq 0]f^K_{\mathrm{L}+}(z)$, where

$$
\begin{aligned}
f^K_{\mathrm{U}+}(z) &= \max\{0, \overline{f^K_{\mathrm{U}+}}(z)\}, \\
\overline{f^K_{\mathrm{U}+}}(z) &= \\
&\max_{k=1,\ldots,K-1} \left\{ \Big[\Big(\frac{\alpha}{\alpha+1}\Big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big] z + \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} - 1\Big]\beta \right\}, \\
f^K_{\mathrm{U}-}(z) &= \max\left\{ \Big(\frac{\alpha}{\alpha+1}\Big) z - \beta, \overline{f^K_{\mathrm{U}-}}(z) \right\}, \\
\overline{f^K_{\mathrm{U}-}}(z) &= \max_{k=1,\ldots,K-1} \left\{ \Big[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big] z - \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\beta \right\}, \\
f^K_{\mathrm{L}+}(z) &= \max_{k=1,\ldots,K} \left\{ \Big(\frac{\alpha}{\alpha+1}\Big)(1 - n_k^{-\alpha-1}) z - (1 - n_k^{-\alpha})\beta \right\}, \quad \text{and} \\
f^K_{\mathrm{L}-}(z) &= \max_{k=1,\ldots,K} \left\{ \Big(\frac{\alpha}{\alpha+1}\Big) n_k^{-\alpha-1} z - n_k^{-\alpha}\beta \right\}.
\end{aligned}
$$

We note that $f^K_{\mathrm{U}}(z)$ is the linear interpolation of points $\{(n_k, f(n_k\beta))\}_{k=1,\ldots,K}$ and $f^K_{\mathrm{L}}(z)$ is the pointwise maximum of the tangents of $f(z)$ at these points (see Figs. 3.2a–3.2b). Due to the convexity of $f(z)$, it follows that $f^K_{\mathrm{U}}(z)$ and $f^K_{\mathrm{L}}(z)$ are convex, and $f^K_{\mathrm{L}}(z) \leq f(z) \leq f^K_{\mathrm{U}}(z)$. Furthermore, we observe that $f^K_{\mathrm{L}+}(z) \leq f_+(z)$ by definition. Based on Lemma III.16, $f^K_{\mathrm{L}+}(z) \leq f_{\mathrm{L}}(z) \leq f^K_{\mathrm{L}-}(z)$ when $\beta < 0$. Similarly, we have $f^K_{\mathrm{L}-}(z) \leq f^K_{\mathrm{L}+}(z)$ when $\beta \geq 0$. It follows that $f^K_{\mathrm{L}}(z) = \max\{f^K_{\mathrm{L}+}(z), f^K_{\mathrm{L}-}(z)\}$. We formalize and extend this observation to $f^K_{\mathrm{U}}(z)$ in the following lemma.

Lemma III.22. *We have $f^K_L(z) = \max\{f^K_{L+}(z), f^K_{L-}(z)\}$ for all $z \in \mathbb{R}$. Furthermore, $f^K_{U+}(z) \leq f(z)$ when $\beta < 0$ and $f^K_{U-}(z) \leq f(z)$ when $\beta \geq 0$. It follows that $f^K_U(z) = \max\{f^K_{U+}(z), f^K_{U-}(z)\}$.*

Proof. We first show that $f^K_{\mathrm{U}+}(z) \leq f_{\mathrm{L}}(z) = \big[\big(\frac{\alpha}{\alpha+1}\big)z - \beta\big]_+$ when $\beta < 0$. Assuming this is true, we have $f^K_{\mathrm{U}+}(z) \leq f(z)$ based on Lemma III.16. To this end, we define $f^K_{\mathrm{U}+}(z) = \max\big\{0, \max_{k=1,\ldots,K-1}\ g^k_{\mathrm{U}+}(z)\big\}$, where

$$
g^k_{\mathrm{U}+}(z) := \Big[\Big(\frac{\alpha}{\alpha+1}\Big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big] z + \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} - 1\Big]\beta.
$$

For each $k = 1, \ldots, K-1$, we prove $g^k_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) \le 0$ by the following chain of equivalences:

$$\begin{aligned}
&\Big[\Big(\frac{\alpha}{\alpha+1}\Big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\Big(\frac{\alpha+1}{\alpha}\Big)\beta + \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} - 1\Big]\beta \;\le\; 0 \\
\Leftrightarrow\;& \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{\alpha(n_{k+1}-n_k)} + \frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \;\ge\; 0 \\
\Leftrightarrow\;& n_{k+1}^{\alpha}(\alpha n_{k+1} - \alpha - 1) \;\ge\; n_k^{\alpha}(\alpha n_k - \alpha - 1),
\end{aligned}$$

where the last line holds because function $g(y) := y^\alpha(\alpha y - \alpha - 1)$ is nondecreasing when $y \ge 1$. Indeed, $g'(y) = (\alpha^2+\alpha)y^{\alpha-1}(y-1) \ge 0$ when $y \ge 1$. It follows that $f^K_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) = 0$. In addition, we note that $0 \le \big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \le \frac{\alpha}{\alpha+1}$. On the one hand, $\big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \le \frac{\alpha}{\alpha+1}$ because $n_{k+1} > n_k$ and $n_{k+1}^{-\alpha} - n_k^{-\alpha} < 0$. On the other hand, $\big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \ge 0$ follows from the following equivalence:

$$\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \le \frac{\alpha}{\alpha+1} \quad\Leftrightarrow\quad n_{k+1}^{-\alpha} + \alpha n_{k+1} \;\ge\; n_k^{-\alpha} + \alpha n_k,$$

where the right-hand side holds because function $h(y) := y^{-\alpha} + \alpha y$ is nondecreasing when $y \ge 1$. Indeed, $h'(y) = \alpha(1 - y^{-\alpha-1}) \ge 0$ when $y \ge 1$. Hence, the slope of $g^k_{\mathrm{U}+}(z)$ is within interval $[0, \frac{\alpha+1}{\alpha}]$ for all k. It follows that $f^K_{\mathrm{U}+}(z) \le f_{\mathrm{L}}(z)$ for all $z \in \mathbb{R}$ because (i) $f^K_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) = f_{\mathrm{L}}(\frac{\alpha+1}{\alpha}\beta) = 0$, (ii) $f^K_{\mathrm{U}+}(z) \le 0 = f_{\mathrm{L}}(z)$ when $z < \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}+}(z)$ are nonnegative, and (iii) $f^K_{\mathrm{U}+}(z) \le f_{\mathrm{L}}(z)$ when $z > \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}+}(z)$ are smaller than or equal to that of $f_{\mathrm{L}}(z)$, i.e., $\frac{\alpha}{\alpha+1}$.

Second, we show that $f^K_{\mathrm{U}-}(z) \le f_{\mathrm{L}}(z) = \big[\big(\frac{\alpha}{\alpha+1}\big)z - \beta\big]_+$ when $\beta \ge 0$. Assuming this is true, we have $f^K_{\mathrm{U}-}(z) \le f(z)$ based on Lemma III.16. To this end, we define

$f^K_{\mathrm{U}-}(z) = \max\{\frac{\alpha}{\alpha+1}z - \beta, \max_{k=1,\ldots,K-1}\ g^k_{\mathrm{U}-}(z)\}$, where

$$g^k_{\mathrm{U}-}(z) := \Big[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]z - \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\beta.$$

For each $k = 1, \ldots, K-1$, we prove $g^k_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) \leq 0$ by the following chain of equivalences:

$$\begin{aligned}
&\Big[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\Big(\frac{\alpha+1}{\alpha}\Big)\beta - \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\beta \ \leq\ 0 \\
\Leftrightarrow\ & (\alpha+1)(n_k^{-\alpha} - n_{k+1}^{-\alpha}) \ \leq\ \alpha(n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}) \\
\Leftrightarrow\ & n_{k+1}^{\alpha}(\alpha n_{k+1} - \alpha - 1) \ \geq\ n_k^{\alpha}(\alpha n_k - \alpha - 1),
\end{aligned}$$

where the last line has been shown above. It follows that $f^K_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) = 0$. In addition, we note that $0 \leq \frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \leq \frac{\alpha}{\alpha+1}$ since $0 \leq \big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \leq \frac{\alpha}{\alpha+1}$. It follows that $f^K_{\mathrm{U}-}(z) \leq f_{\mathrm{L}}(z)$ for all $z \in \mathbb{R}$ because (i) $f^K_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) = f_{\mathrm{L}}(\frac{\alpha+1}{\alpha}\beta) = 0$, (ii) $f^K_{\mathrm{U}-}(z) \leq 0 = f_{\mathrm{L}}(z)$ when $z < \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}-}(z)$ are nonnegative, and (iii) $f^K_{\mathrm{U}-}(z) \leq f_{\mathrm{L}}(z)$ when $z > \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}-}(z)$ are smaller than or equal to that of $f_{\mathrm{L}}(z)$, i.e., $\frac{\alpha}{\alpha+1}$. □

In the following proposition, we present conservative approximations based on $f^K_{\mathrm{U}}(z)$ and relaxed ones based on $f^K_{\mathrm{L}}(z)$, both of which are in the form of linear matrix inequalities. We note that these approximations are asymptotically tight as K grows to infinity. We omit the proof here because it follows from the standard duality approach. Interested readers are referred to [29] and [110].

Proposition III.23. *Define* $(T+1)\times(T+1)$ *matrix* $\Omega := \begin{bmatrix} \big(\frac{\alpha+2}{\alpha}\big)\Sigma & \big(\frac{\alpha+1}{\alpha}\big)\mu \\ \big(\frac{\alpha+1}{\alpha}\big)\mu^\top & 1 \end{bmatrix}$. *Then, for given integer* $K \geq 3$ *and real numbers* $1 = n_1 < n_2 < \cdots < n_K = \infty$, *DRRC* (3.7) *is satisfied if there exists a symmetric matrix* $M_U \in \mathbb{R}^{(T+1)\times(T+1)}$ *such*

that

$$\beta + \frac{1}{\epsilon} M_U \cdot \Omega \;\leq\; b(x), \quad M_U \;\succeq\; 0, \quad M_U \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)a(x) \\ \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)a(x)^\top & -\beta \end{bmatrix}, \tag{3.34a}$$

$$M_U \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\Big[\big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]a(x) \\ \frac{1}{2}\Big[\big(\frac{\alpha}{\alpha+1}\big) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]a(x)^\top & \Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} - 1\Big]\beta \end{bmatrix},$$

$$\forall k = 1, \ldots, K-1, \tag{3.34b}$$

$$M_U \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\Big[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]a(x) \\ \frac{1}{2}\Big[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]a(x)^\top & -\Big[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Big]\beta \end{bmatrix}, \quad \forall k = 1, \ldots, K-1, \tag{3.34c}$$

where $\cdot$ *represents the Frobenius product of matrices. Furthermore, DRRC* (3.7) *implies that there exists a symmetric matrix* $M_L \in \mathbb{R}^{(T+1)\times(T+1)}$ *such that*

$$\beta + \frac{1}{\epsilon} M_L \cdot \Omega \;\leq\; b(x), \quad M_L \;\succeq\; 0, \quad M_L \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)a(x) \\ \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)a(x)^\top & -\beta \end{bmatrix},$$

$$M_L \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)(1 - n_k^{-\alpha-1})a(x) \\ \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)(1 - n_k^{-\alpha-1})a(x)^\top & -(1 - n_k^{-\alpha})\beta \end{bmatrix}, \; \forall k = 1, \ldots, K-1,$$

$$M_L \;\succeq\; \begin{bmatrix} 0 & \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)n_k^{-\alpha-1}a(x) \\ \frac{1}{2}\big(\frac{\alpha}{\alpha+1}\big)n_k^{-\alpha-1}a(x)^\top & -n_k^{-\alpha}\beta \end{bmatrix}, \; \forall k = 1, \ldots, K-1.$$

Note that, as $n_K = \infty$, constraints (3.34b)–(3.34c) reduce to

$$M_{\mathrm{U}} \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x)^\top & \left(\frac{n_{K-1}^{-\alpha}}{\alpha+1} - 1\right)\beta \end{bmatrix},$$

$$M_{\mathrm{U}} \succeq \begin{bmatrix} 0 & 0 \\ 0 & -\left(\frac{n_{K-1}^{-\alpha}}{\alpha+1}\right)\beta \end{bmatrix}$$

when $k = K - 1$.

Remark III.24. In computation, we can use the conservative approximation (3.34a)–(3.34c) to find near-optimal solutions. More specifically, suppose that we employ the separation approach to solve problem $\min\{c(x) : x \in X,\ x \text{ satisfies (3.7)}\}$ and have finished the first K iterations. Then, from these iterations, we obtain a lower bound c_{L}^K of the optimal objective value and K outputs, denoted $\varphi_1, \ldots, \varphi_K$, by iteratively solving Separation Problem 2. By letting $n_1 = 1$, $n_{K+2} = \infty$, and $n_k = \varphi_{k-1}$ for all $k = 2, \ldots, K+1$, we obtain an upper bound c_{U}^K of the optimal objective value by solving problem $\min\{c(x) :\ x \in X,\ x \text{ satisfies (3.34a)–(3.34c) based on } n_1, \ldots, n_{K+2}\}$, whose optimal solution is denoted x_K^*. If $(c_{\mathrm{U}}^K - c_{\mathrm{L}}^K)/c_{\mathrm{L}}^K$ is small enough, then we can stop the iterations and output x_K^* as a near-optimal solution.

3.6 Extension to Linear Unimodality

In this section, we consider an extension of DRCC (3.6) and DRRC (3.7) based on a related structural property called *linear unimodality*.

Definition III.25. (Linear Unimodality; see [23]) A probability distribution $\mathbb{P}_\xi$ is called linear unimodal about 0 if for all $a \in \mathbb{R}^T$, the linear combination $a^\top \xi$ is univariate unimodal about 0.

Analogous to (3.5), we define the alternative ambiguity set based on linear uni-

modality as

$$\mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma) := \left\{\mathbb{P}_\xi \in \mathcal{M}_T : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma,\ \mathbb{P}_\xi \text{ is linear unimodal about } 0\right\}. \tag{3.35}$$

We now show an equivalence between ambiguity sets $\mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)$ and $\mathcal{D}_\xi(\mu,\Sigma,\alpha)$ with $\alpha = 1$. It follows that all results derived in Sections 3.2–3.4, with α set to be 1, remain valid under $\mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)$.

Proposition III.26. *For any Borel measurable function $h : \mathbb{R} \to \mathbb{R}$, we have*

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}\left[h(a(x)^\top \xi)\right] \ = \ \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{LU}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}\left[h(a(x)^\top \xi)\right].$$

Proof. By Theorem 3.5 in [23], a random variable X is 1-unimodal if and only if there exists a random variable Z such that $X = UZ$, where U is uniform in $(0,1)$ and independent of Z.

First, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)$. As $a^\top\xi$ is univariate 1-unimodal for all $a \in \mathbb{R}^T$ because $\mathbb{P}_\xi$ is 1-unimodal, $\mathbb{P}_\xi \in \mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)$. It follows that $\mathcal{D}_\xi(\mu,\Sigma,1) \subseteq \mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)$ and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top\xi)] \geq \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top\xi)]$.

Second, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)$. Then, $\zeta := a(x)^\top\xi$ is 1-unimodal because $\mathbb{P}_\xi$ is linear unimodal. Hence, there exists a Z_ζ such that $\zeta = UZ_\zeta$. It follows that $\mathbb{E}[Z_\zeta] = 2\mu_1$ and $\mathbb{E}[Z_\zeta^2] = 3\Sigma_1$. Based on Theorem 1 in [78], there exists a $Z_\xi \in \mathbb{R}^T$ such that $Z_\zeta = a(x)^\top Z_\xi$, $\mathbb{E}[Z_\xi] = 2\mu$, and $\mathbb{E}[Z_\xi Z_\xi^\top] = 3\Sigma$. It follows that UZ_ξ is 1-unimodal, and meanwhile $\mathbb{E}_{\mathbb{P}_\xi}[UZ_\xi] = \frac{1}{2}\mathbb{E}[Z_\xi] = \mu$ and $\mathbb{E}_{\mathbb{P}_\xi}[(UZ_\xi)(UZ_\xi)^\top] = \frac{1}{3}\mathbb{E}[Z_\xi Z_\xi^\top] = \Sigma$. Furthermore, $a(x)^\top\xi = a(x)^\top(UZ_\xi)$. Therefore, the probability distribution of UZ_ξ belongs to $\mathcal{D}_\xi(\mu,\Sigma,1)$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top\xi)] \leq \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{\text{LU}}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top\xi)]$. ∎

3.7 Case Studies

In this section, we evaluate the theoretical results derived in Sections 3.2–3.4 based on a risk-constrained optimal power flow (RCED) problem in power system operation. We present a nominal RCED model as follows:

$$\min_{g,d,r^{\mathrm{U}},r^{\mathrm{D}}} \quad \sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}} \left[c_{i2}g_i^2 + c_{i1}g_i + c_i^{\mathrm{R}}(r_i^{\mathrm{U}} + r_i^{\mathrm{D}})\right] \tag{3.36a}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}} g_i + \sum_{i\in\mathcal{I}_{\mathrm{R}}} f_i = \sum_{b=1}^{B} L_b, \tag{3.36b}$$

$$r_i = -\Big(\sum_{i\in\mathcal{I}_{\mathrm{R}}} w_i\Big) d_i, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}, \tag{3.36c}$$

$$\sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}} d_i \;=\; 1, \tag{3.36d}$$

$$-r_i^{\mathrm{D}} \le r_i \le r_i^{\mathrm{U}}, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}, \tag{3.36e}$$

$$g_i^{\mathrm{MIN}} \le g_i + r_i \le g_i^{\mathrm{MAX}}, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}, \tag{3.36f}$$

$$-C_\ell \le \sum_{b=1}^{B} D_\ell^b \Big[\sum_{i\in\mathcal{G}_b}(g_i + r_i) + \sum_{i\in\mathcal{H}_b}(f_i + w_i) - L_b\Big] \le C_\ell, \quad \forall \ell \in \mathcal{L}, \tag{3.36g}$$

where B represents the number of buses in the power system, $\mathcal{I}$ represents the set of generating units (conventional and renewable), $\mathcal{I}_{\mathrm{R}}$ represents the set of renewable units, $\mathcal{L}$ represents the set of transmission lines, $\mathcal{G}_b$ represents the set of conventional units at bus b, $\mathcal{H}_b$ represents the set of renewable units at bus b, c_{i2} and c_{i1} represent cost parameters of conventional unit i, c_i^{R} represents the unit cost for up/down reserve capacity of conventional unit i, L_b represents the load at bus b, and C_ℓ represents the capacity of transmission line ℓ. For each renewable unit $i \in \mathcal{I}_{\mathrm{R}}$, f_i and w_i represent the forecasted power output and the forecast error, respectively. For each conventional unit $i \in \mathcal{I}\setminus\mathcal{I}_{\mathrm{R}}$, g_i and r_i represent the planned generation amount and the adjustment amount, respectively, and d_i represents the portion of total generation-load mismatch to be offset by this unit [96, 9]. Constraint (3.36b) describes the power balance

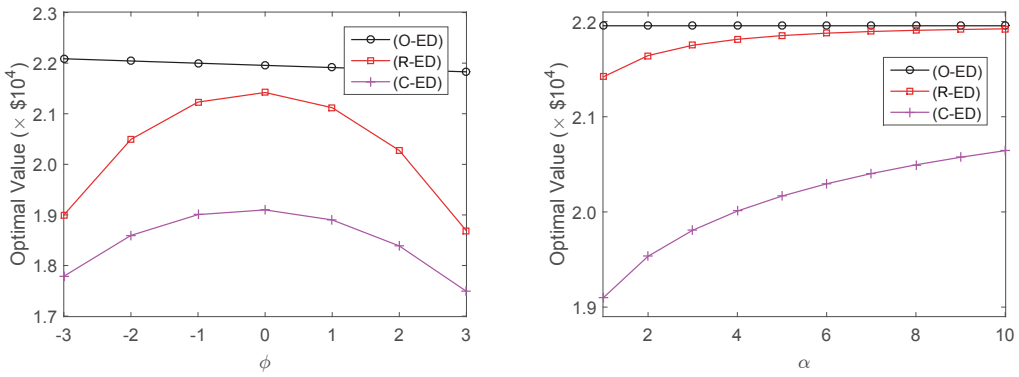
requirement for generation and loads (we assume that the loads are deterministic), constraints (3.36c) describe the proportional distribution of mismatches, constraint (3.36d) requires that all proportions sum up to be 1, constraints (3.36e) limit the adjustment amount by the reserve capacities r^{U} and r^{D}, constraints (3.36f) bound the generation amount by the generation capacity, and constraints (3.36g) describe the transmission capacity limits based on the dc approximation where D_ℓ^b maps power injections to power flows (see, e.g., [6] and [30]).

Our case study uses the IEEE 30-bus system [108]. We increase all electricity loads by 50% and add two wind farms at buses 5 and 22. The forecasted power output from each wind farm is 30 MW. The transmission line between buses 1 and 2 has a capacity of 30 MW, while all other line flows are unconstrained. Other cost and capacity coefficients are reported in Table 3.1. We assume random forecast errors and describe the uncertainty by an uncorrelated random vector $w := [w_1, w_2]^\top$ with mean μ_{w} and covariance matrix $\Gamma_{\mathrm{w}} = \mathrm{diag}(9, 9)$. Additionally, we assume that w is α-unimodal about $[0, 0]^\top$. To handle random violations of constraints (3.36e)–(3.36g), we replace them by DRCC (3.6) and DRRC (3.7), and term the resultant RCED model (C-ED) and (R-ED), respectively. For example, in (C-ED), we replace constraints (3.36e) by $\inf_{\mathbb{P}_{\mathrm{w}}\in\mathcal{D}_{\mathrm{w}}}\{d_i \sum_{i\in\mathcal{I}_{\mathrm{R}}} w_i \le r_i^{\mathrm{D}}\} \ge 1-\epsilon$ and $\inf_{\mathbb{P}_{\mathrm{w}}\in\mathcal{D}_{\mathrm{w}}}\{-d_i \sum_{i\in\mathcal{I}_{\mathrm{R}}} w_i \le r_i^{\mathrm{U}}\} \ge 1-\epsilon$, where $\mathcal{D}_{\mathrm{w}} = \{\mathbb{P}_{\mathrm{w}} \in \mathcal{M}_2 : \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[w] = \mu_{\mathrm{w}}, \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[ww^\top] = \mu_{\mathrm{w}}\mu_{\mathrm{w}}^\top + \mathrm{diag}(9,9), \mathbb{P}_{\mathrm{w}} \text{ is } \alpha\text{-unimodal about } 0\}$. In contrast, in (R-ED), we replace constraints (3.36e) by $\sup_{\mathbb{P}_{\mathrm{w}}\in\mathcal{D}_{\mathrm{w}}} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_{\mathrm{w}}}\Big(d_i \sum_{i\in\mathcal{I}_{\mathrm{R}}} w_i\Big) \le r_i^{\mathrm{D}}$ and $\sup_{\mathbb{P}_{\mathrm{w}}\in\mathcal{D}_{\mathrm{w}}} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_{\mathrm{w}}}\Big(-d_i \sum_{i\in\mathcal{I}_{\mathrm{R}}} w_i\Big) \le r_i^{\mathrm{U}}$. Throughout this case study, we set $1-\epsilon = 95\%$. Lastly, when the requirement of α-unimodality is relaxed from $\mathcal{D}_{\mathrm{w}}$, (C-ED) and (R-ED) become equivalent and we term this model (O-ED).

By using (O-ED) as a benchmark, we test (C-ED) and (R-ED) under various selections of μ_{w} and α values. First, we fix $\alpha = 1$ and let $\mu_{\mathrm{w}} = \phi[1, 1]^\top$ with $\phi \in \{-3, -2, \ldots, 3\}$. We report the optimal objective values of the three models

Conventional Unit	Bus Index	c_{i1} (\$/MW)	c_{i2} (\$/MW^2)	c_i^{R} (\$/MW)	g_i^{MIN} (MW)	g_i^{MAX} (MW)
1	1	20	0.04	200	0	360
2	2	40	0.25	400	0	140
3	5	40	0.01	400	0	100
4	8	40	0.01	400	0	100
5	11	40	0.01	400	0	100
6	13	40	0.01	400	0	100

Table 3.1: Coefficients of the Case Study



(a) Optimal Value vs. ϕ (b) Optimal Value vs. α

Figure 3.3: Optimal values of (O-ED), (C-ED), and (R-ED) with various ϕ and α

in Fig. 3.3a. From this figure, we observe that the optimal value of (O-ED) is consistently larger than that of (R-ED), which is consistently larger than that of (C-ED). This demonstrates that incorporating α-unimodality makes the RCED model less conservative and hence decreases the cost of economic dispatch. Meanwhile, unlike in (O-ED), DRCC (3.6) and DRRC (3.7) are not equivalent when α-unimodality is incorporated in the ambiguity set. Furthermore, we observe that the discrepancy between (O-ED) and (C-ED)/(R-ED) amplifies as ϕ deviates from 0. This indicates that α-unimodality plays a more important role in $\mathcal{D}_{\mathrm{w}}$ as the difference between μ_{w} and the mode increases.

Second, we fix $\mu_{\mathrm{w}} = [0, 0]^\top$ and let α increase from 1 to 10. We report the optimal objective values of the three models in Fig. 3.3b. From this figure, we observe that the discrepancy between (O-ED) and (C-ED)/(R-ED) shrinks as α grows. This is as

expected because the requirement of α-unimodality weakens as α grows. Although not shown in this figure, the convergence of (R-ED) to (O-ED) takes place when $\alpha \geq 40$, while the convergence of (C-ED) takes place when $\alpha \geq 10^4$. The slow convergence indicates that unimodality information can significantly influence the structure of $\mathcal{D}_{\mathrm{w}}$ and the worst-case probability distribution.

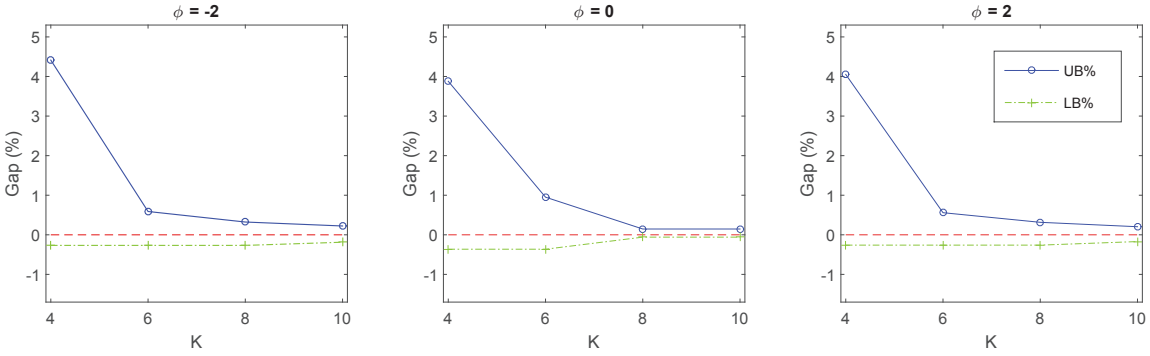


Figure 3.4: Gaps between the Optimal Objective Value and the Relaxed and Conservative Approximations of (C-ED)

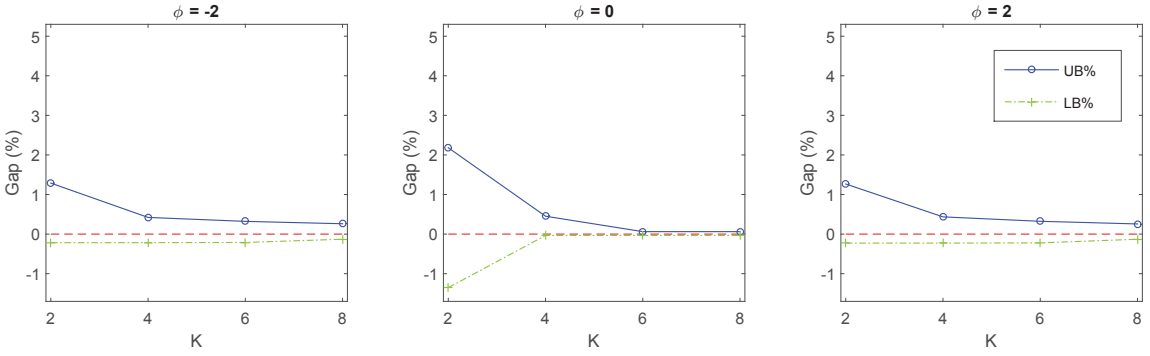


Figure 3.5: Gaps between the Optimal Objective Value and the Relaxed and Conservative Approximations of (R-ED)

Third, we let $\alpha = 1$, $\mu_{\mathrm{w}} = \phi[1,1]^\top$ with $\phi \in \{-2, 0, 2\}$, and evaluate the tightness of the approximations of DRCC and DRRC derived in Propositions III.10–III.11 and Proposition III.23, respectively. In this test, we follow Remarks III.12–III.24 to choose the interpolation points $n_1, \ldots, n_K$ in these approximations. In Fig. 3.4, we report the gap between the optimal objective value $v^*_{(\text{C-ED})}$ of (C-ED) and the upper bound v_{UB} obtained from the conservative approximation, and the gap between $v^*_{(\text{C-ED})}$ and the

lower bound v_{LB} obtained from the relaxed approximation, for $K \in \{4, 6, 8, 10\}$. The gaps are obtained by computing $\text{UB}\% = (v_{\text{UB}} - v^*_{\text{(C-ED)}})/v^*_{\text{(C-ED)}} \times 100\%$ and $\text{LB}\% = (v^*_{\text{(C-ED)}} - v_{\text{LB}})/v^*_{\text{(C-ED)}} \times 100\%$. Similarly, in Fig. 3.5, we report the gap between the optimal objective value $v^*_{\text{(R-ED)}}$ of (R-ED) and those of its K-piece approximations with $K \in \{2, 4, 6, 8\}$. From Figs. 3.4–3.5, we observe that the gaps quickly shrink as K increases and the approximations become near-optimal (e.g., $\text{UB}\% + \text{LB}\% < 1\%$) when $K \geq 8$.

3.8 Supporting Material

In this section, we prove the proof of the Observations 2 and another observation on the worst case expectation.

3.8.0.1 Proof of Observation 2

Proof. As $f^k_+(z) = \left(\frac{\alpha}{\alpha+1}\right)(1 - k^{-\alpha-1})z - (1 - k^{-\alpha})\beta$, we have

$$\left[f^k_+(z)\right]_+ = \begin{cases} 0, & \text{if } z < z_0(k) := \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta \\ \left(\frac{\alpha}{\alpha+1}\right)(1 - k^{-\alpha-1})z - (1 - k^{-\alpha})\beta, & \text{if } z \geq z_0(k) \end{cases}$$

for all $k \geq 1$. As $[f^{k+1}_+(z)]_+ \geq 0$ for all $z \in \mathbb{R}$, to show that $[f^{k+1}_+(z)]_+ \geq [f^k_+(z)]_+$, it suffices to prove that $[f^{k+1}_+(z)]_+ \geq f^k_+(z)$ for all $z \in \mathbb{R}$. First, as $\beta \leq 0$ and $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}$ increases in k, we have $\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta \geq \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-(k+1)^{-\alpha}}{1-(k+1)^{-\alpha-1}}\right)\beta$, i.e., $z_0(k) \geq z_0(k+1)$. It follows that, when $z < z_0(k)$, $f^k_+(z) \leq 0$ and hence $[f^{k+1}_+(z)]_+ \geq f^k_+(z)$. Second, when $z \geq z_0(k)$, $f^{k+1}_+(z) \geq 0$ because $z \geq z_0(k) \geq z_0(k+1)$ and $f^{k+1}_+(z)$ increases in z. As both $f^k_+(z)$ and $f^{k+1}_+(z)$ are affine functions of z, we have $f^{k+1}_+(z) = f^{k+1}_+(z_0(k)) + (\frac{\alpha}{\alpha+1})(1-(k+1)^{-\alpha-1})(z-z_0(k))$ and $f^k_+(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})(z-z_0(k))$ for $z \geq z_0(k)$. It follows that $f^{k+1}_+(z) - f^k_+(z) = f^{k+1}_+(z_0(k)) + (\frac{\alpha}{\alpha+1})[k^{-\alpha-1} - (k+1)^{-\alpha-1}](z - z_0(k)) \geq 0$. Hence, $[f^{k+1}_+(z)]_+ \geq f^k_+(z)$ when $z \geq z_0(k)$ and the proof is

complete. ∎

3.8.0.2 Observation on Worst Case Expectation

For random variable Z and constant $\beta \in \mathbb{R}$, we make the following observation on the worst-case expectation $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+$. Note that this observation can be made following the derivations in [83], and we present a proof below for completeness.

Observation 3. Given $\beta \in \mathbb{R}$, we have

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ \;=\; \frac{1}{2}\left[\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} - \beta + \mu_0\right].$$

Proof. We represent $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+$ as the following optimization problem

$$\begin{array}{rrll} v_P \;=\; \max\limits_{\mathbb{P}_Z} & \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ \\ \text{(P)} \qquad \text{s.t.} & \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\ & \mathbb{E}_{\mathbb{P}_Z}[1] = 1, \end{array}$$

whose dual is

$$\begin{array}{rrll} v_D \;=\; \min\limits_{q,p,r} & \mu_0 p + \Sigma_0 q + r \\ \text{(D)} \qquad \text{s.t.} & qz^2 + pz + r \geq [z-\beta]_+, & \forall z \in \mathbb{R}. \end{array}$$

The weak duality between (P) and (D), i.e., $v_D \leq v_P$, holds because $\mu_0 p + \Sigma_0 q + r = \mathbb{E}_{\mathbb{P}_Z}[qZ^2 + pZ + r] \leq \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+$ for any feasible solution (q,p,r) to (D) and feasible solution $\mathbb{P}_Z$ to (P). Now we prove the strong duality by constructing two feasible solutions to (P) and (D), respectively, that have the same objective value. On the one hand, the primal solution $\hat{\mathbb{P}}_Z$ is supported on two points z_1 and z_2 with probability

masses p_1 and p_2, respectively, where $\Delta = \sqrt{(\beta - \mu_0)^2 + (\Sigma_0 - \mu_0^2)}$ and

$$p_1 = \frac{\beta - \mu_0 + \Delta}{2\Delta},\ p_2 = \frac{\mu_0 - \beta + \Delta}{2\Delta},\ z_1 = \beta - \Delta,\ \text{and}\ z_2 = \beta + \Delta.$$

We have $p_1, p_2 \geq 0$ because $\Delta \geq |\beta - \mu_0|$. Meanwhile, we have

$$p_1 z_1 + p_2 z_2 \ = \ \frac{(\beta - \mu_0 + \Delta)(\beta - \Delta)}{2\Delta} + \frac{(\mu_0 - \beta + \Delta)(\beta + \Delta)}{2\Delta} \ = \ \mu_0,$$

and

$$\begin{aligned} p_1 z_1^2 + p_2 z_2^2 \ &= \ \frac{(\beta - \mu_0 + \Delta)(\beta - \Delta)^2}{2\Delta} + \frac{(\mu_0 - \beta + \Delta)(\beta + \Delta)^2}{2\Delta} \\ &= \frac{(\beta - \mu_0)\left[(\beta - \Delta)^2 - (\beta + \Delta)^2\right] + \Delta\left[(\beta - \Delta)^2 + (\beta + \Delta)^2\right]}{2\Delta} \\ &= \ -\beta^2 + 2\mu_0\beta + \Delta^2 \ = \ -\beta^2 + 2\mu_0\beta + (\beta - \mu_0)^2 + (\Sigma_0 - \mu_0^2) \ = \ \Sigma_0. \end{aligned}$$

Hence, $\hat{\mathbb{P}}_Z$ is feasible to (P). On the other hand, the dual solution $(\hat{q}, \hat{p}, \hat{r})$ is such that

$$\hat{q} = \frac{1}{4\Delta},\ \hat{p} = \frac{\Delta - \beta}{2\Delta},\ \text{and}\ \hat{r} = \frac{(\Delta - \beta)^2}{4\Delta}.$$

Hence, $\hat{q}z^2 + \hat{p}z + \hat{r} = \frac{1}{4\Delta}(z + \Delta - \beta)^2$. It follows that $\hat{q}z^2 + \hat{p}z + \hat{r} \geq 0$ for all $z \in \mathbb{R}$. Meanwhile, $(\hat{q}z^2 + \hat{p}z + \hat{r}) - (z - \beta) = \frac{1}{4\Delta}(z - \beta - \Delta)^2 \geq 0$, i.e., $\hat{q}z^2 + \hat{p}z + \hat{r} \geq z - \beta$. Thus, $\hat{q}z^2 + \hat{p}z + \hat{r} \geq [z - \beta]_+$ and so $(\hat{q}, \hat{p}, \hat{r})$ is feasible to (D).

Finally, the primal objective value associated with $\hat{\mathbb{P}}_Z$ is $p_2(z_2 - \beta) \ = \ \frac{(\mu_0 - \beta + \Delta)\Delta}{2\Delta} \ = \ \frac{1}{2}(\Delta - \beta + \mu_0)$. Meanwhile, the dual objective value

associated with $(\hat{q}, \hat{p}, \hat{r})$ is

$$\begin{aligned}
&\mu_0 \left(\frac{\Delta - \beta}{2\Delta} \right) + \Sigma_0 \left(\frac{1}{4\Delta} \right) + \frac{(\Delta - \beta)^2}{4\Delta} \\
&\qquad = \frac{\Delta^2 + (\beta^2 - 2\mu_0\beta + \mu_0^2) + (\Sigma_0 - \mu_0^2) + 2\mu_0\Delta - 2\Delta\beta}{4\Delta} \\
&\qquad = \frac{2\Delta^2 + 2\mu_0\Delta - 2\Delta\beta}{4\Delta} \;=\; \frac{1}{2}(\Delta - \beta + \mu_0),
\end{aligned}$$

which coincides with the primal objective value associated with $\hat{\mathbb{P}}_Z$. ∎

CHAPTER IV

Distributionally Robust Optimal Power Flow Assuming Log-Concave Distributions

Optimization formulations with chance constraints have been widely proposed to operate the power system under various uncertainties, such as renewable production and load consumption. Constraints like the system's physical limits are required to be satisfied at high confidence levels. Conventional solving methodologies either make assumptions on the underlying uncertainty distributions or give overly-conservative results. In this chapter, we develop a new distributionally robust (DR) chance constrained optimal power flow formulation in which the chance constraints are satisfied over a family of distributions with known first-order moments, ellipsoidal support, and an assumption that the probability distributions are log-concave. Since most practical uncertainties have log-concave probability distributions, including this assumption in the formulation reduces the objective costs as compared to traditional DR approaches without sacrificing reliability. We derive second-order cone approximations of the DR chance constraints, resulting in a tractable formulation that can be solved with commercial solvers. We perform case studies to compare our approach to standard approaches and find that our approach produces solutions that are sufficiently reliable and less costly than traditional DR approaches. The main content of this chapter is summarized in the following paper.

1. B. Li, R. Jiang, and J.L. Mathieu. Distributionally robust chance constrained optimal power flow assuming log-concave distributions. In *Power Systems Computation Conference*, Dublin, Ireland, 2018.

4.1 Introduction

Uncertainties resulting from, e.g., renewable generation or load consumption, complicate the optimal operation of power systems. Optimal power flow (OPF) formulations using chance constraints [104, 42, 96, 79, 97] have been proposed to limit the chance of violating physical constraints, such as generation and line limits. The key difficulty in solving chance constrained problems is that we usually do not know the true underlying probability distributions of the uncertain variables. Most existing work either assumes that the uncertainties follow known, empirical distributions [9, 51] or use the randomized techniques that require constraint satisfaction for a large number of scenarios [17]. The former often fail to guarantee reliability (unless the distributions are modeled perfectly) while the latter often give overly-conservative results. Recently, distributionally robust chance constrained (DRCC) OPF formulations have been developed, e.g., [107, 80, 102, 48, 88, 33, 56]. This new approach requires the chance constraints to be satisfied for all possible distributions with known statistical parameters (e.g., first and second-order moments [107, 102, 56] or likelihood to a data-based distribution [33]) producing highly reliable solutions generally at a lower cost than those obtained with the randomized techniques. In addition to moment information, common structural properties of practical uncertainties, like unimodality [48, 88] and symmetry [80], can be enforced to further lower the costs. To solve DRCC OPF problems, the constraints are either exactly reformulated or approximated resulting in second-order cone programs (SOCPs) [49] or semidefinite programs (SDPs) [44, 110], which can both be directly solved using commercial solvers.

In this chapter, we propose a new DRCC OPF formulation in which chance constraints are satisfied over a family of distributions (i.e., an ambiguity set) with known first-order moments, an ellipsoidal support, and an assumption that the probability distributions are log-concave [23, 4]. Most practical uncertainties follow log-concave distributions; Gaussian, Beta and Weibull distributions are all log-concave. For example, wind forecast errors are generally modeled as specific distributions in the log-concave family [73, 24, 10]. Including this assumption limits the distributions over which the chance constraints should be satisfied, reducing the conservatism and the objective cost of the solution. Meanwhile, assuming the real uncertainty distributions are log-concave, the solutions will be sufficiently reliable. We benchmark our approach against a DRCC approach that uses an ambiguity set with only moment and support requirements and a chance constrained approach that assumes all uncertainty follows multivariate normal distributions, which is a special type of log-concave distribution.

The contributions of the chapter are as follows. 1) We derive a projection property to simplify the multi-dimensional ambiguity set into an equivalent single dimensional one. 2) We derive second-order cone (SOC) relaxing and conservative approximations (i.e., a sandwich approximation) of the distributionally robust chance constraints under our ambiguity set. Using the DC power flow equations, the resulting DRCC OPF is an SOCP. We also derive exact SOC constraints for the simpler ambiguity set. 3) We apply the theoretical results to solve a DC OPF problem on a modified IEEE 9-bus system with wind uncertainty. 4) We compare our results to those produced by the two benchmark approaches described above and report the objective costs and reliability of all approaches. To the best of our knowledge, this is the first work to include the log-concavity structure in distributionally robust optimization.

In the remainder of the chapter, the DR chance constraints and ambiguity set are introduced in Section 4.2. In Section 4.3, we derive the projection property. In

Section 4.4, we give our main theoretical results. Finally, in Section 4.5, we present the case studies.

4.2 Ambiguity Sets with Log-Concave Information

Here, we follow the same assumption as in Chapter IV. Assume that ξ represents an n-dimensional random vector defined on a probability space $(\mathbb{R}^n, \mathcal{B}, \mathbb{P}_\xi)$ with Borel σ-algebra $\mathcal{B}$ and probability distribution $\mathbb{P}_\xi$. We define $x \in \mathbb{R}^m$ as the design variable vector and $1-\epsilon$ as the confidence level, and obtain the following chance constraint

$$\mathbb{P}_\xi \left(f(x,\xi) \leq 0 \right) \geq 1 - \epsilon. \tag{4.1}$$

Furthermore, by considering a distributional ambiguity set $\mathcal{D}_\xi$ that incorporates plausible candidates of the true distribution $\mathbb{P}_\xi$, we obtain the following DR chance constraint:

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi \left(f(x,\xi) \leq 0 \right) \geq 1 - \epsilon. \tag{4.2}$$

Next, we further specify that the constraint function $f(x,\xi)$ is a bilinear function, i.e., $f(x,\xi) = a(x)^T\xi - b(x)$, where both $a(x) : \mathbb{R}^m \to \mathbb{R}^n$ and $b(x) : \mathbb{R}^m \to \mathbb{R}$ are affine functions of x. The DR chance constraint is then in the form:

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi \left(a(x)^T\xi \leq b(x) \right) \geq 1 - \epsilon. \tag{4.3}$$

For ambiguity sets, we assume $\mathcal{D}_\xi$ consists of all probability distributions $\mathbb{P}_\xi$ that live in an ellipsoid around its mean value μ and satisfies structural properties specified in set $\mathcal{P}^{\text{S}}$, i.e.,

$$\begin{aligned} \mathcal{D}_\xi := \{ \mathbb{P}_\xi \in \mathcal{P}^{\text{S}} : \quad & \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu, \\ & \|\Sigma^{-\frac{1}{2}}(\xi - \mu)\|_2 \leq r \text{ almost surely} \}, \end{aligned} \tag{4.4}$$

where matrix $\Sigma \succ 0$ defines the ellipsoid and r represents its radius. For example, we could choose Σ to be the empirical covariance matrix obtained from the data corresponding to ξ. In this chapter, we consider the following two options for $\mathcal{D}_\xi$:

- Case 1: $\mathcal{P}_1^{\mathrm{S}} = \{\mathbb{P}_\xi \text{ is log-concave}\}$,
- Case 2: $\mathcal{P}_2^{\mathrm{S}} = \{\mathbb{P}_\xi \text{ is any probability distribution}\}$,

where Case 2 is our DRCC benchmark. We denote the ambiguity set corresponding to Case 1 as $\mathcal{D}_\xi^1$ and corresponding to Case 2 as $\mathcal{D}_\xi^2$, where $\mathcal{D}_\xi^1 \subseteq \mathcal{D}_\xi^2$. We formally define the log-concavity [23] as follows.

Definition IV.1. A probability distribution $\mathbb{P}$ is log-concave if and only if for all non-empty sets $A, B \in \mathcal{B}$ and for all $\theta \in (0,1)$, we have

$$\mathbb{P}\left(\theta A + (1-\theta)B\right) \geq [\mathbb{P}(A)]^\theta [\mathbb{P}(B)]^{1-\theta}. \tag{4.5}$$

A large family of probability distributions are log-concave, including Gaussian, Beta, and Weibull distributions and log-concavity is commonly assumed for many practical uncertainty distributions.

4.3 Projection Property

In this section, we derive a projection property that transforms the ambiguity set $\mathcal{D}_\xi^i$ of a random vector $\xi \in \mathbb{R}^n$ into an equivalent ambiguity set $\mathcal{D}_\zeta^i$ of a random variable $\zeta \in \mathbb{R}$, for $i = 1, 2$.

Lemma IV.2. *For $i = 1, 2$, the following equality holds:*

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^i} \mathbb{P}_\xi\left(a(x)^T \xi \leq b(x)\right) = \inf_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta^i} \mathbb{P}_\zeta\left(\zeta \leq g(x)\right), \tag{4.6}$$

where

$$g(x) = b(x) - a(x)^T\mu + r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \tag{4.7}$$

and

$$\begin{aligned}\mathcal{D}_\zeta^i := \{\mathbb{P}_\zeta \in \mathcal{P}_i^s :\ & \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = r\|\Sigma^{\frac{1}{2}}a(x)\|_2, \\ & 0 \le \zeta \le 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \ \textit{almost surely}\}.\end{aligned} \tag{4.8}$$

Proof: We provide the proof for $i = 1$; the proof for $i = 2$ is similar and so omitted.

On the one hand, we can pick any ξ with $\mathbb{P}_\xi \in \mathcal{D}_\xi^1$. Define $\zeta = a(x)^T(\xi - \mu) + r\|\Sigma^{\frac{1}{2}}a(x)\|_2$, we have $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = r\|\Sigma^{\frac{1}{2}}a(x)\|_2$ and $0 \le \zeta \le 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2$ almost surely, where the bounds of ζ are valid because

$$\begin{aligned}\zeta &\le r\|\Sigma^{\frac{1}{2}}a(x)\|_2 + \max_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le r} a(x)^T(\xi - \mu) \\ &= r\|\Sigma^{\frac{1}{2}}a(x)\|_2 + \max_{y:\|y\|_2\le r} (\Sigma^{\frac{1}{2}}a(x))^T y \\ &= r\|\Sigma^{\frac{1}{2}}a(x)\|_2 + r\|\Sigma^{\frac{1}{2}}a(x)\|_2 = 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2\end{aligned}$$

and

$$\begin{aligned}\zeta &\ge r\|\Sigma^{\frac{1}{2}}a(x)\|_2 + \min_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le r} a(x)^T(\xi - \mu) \\ &= r\|\Sigma^{\frac{1}{2}}a(x)\|_2 - r\|\Sigma^{\frac{1}{2}}a(x)\|_2 = 0.\end{aligned}$$

Furthermore, from Lemma 2.1 of [23], we know that $\mathbb{P}_\zeta$ is log-concave. Hence, $\mathbb{P}_\zeta \in \mathcal{D}_\zeta^1$ and

$$\inf_{\mathbb{P}_\xi\in\mathcal{D}_\xi^1} \mathbb{P}_\xi\left(a(x)^T\xi \le b(x)\right) \ge \inf_{\mathbb{P}_\zeta\in\mathcal{D}_\zeta^1} \mathbb{P}_\zeta\left(\zeta \le g(x)\right).$$

On the other hand, we can pick any ζ with $\mathbb{P}_\zeta \in \mathcal{D}_\zeta^1$ and define $\xi = \mu + (\zeta - r\|\Sigma^{\frac{1}{2}}a(x)\|_2)\frac{\Sigma a(x)}{a(x)^T\Sigma a(x)}$. We have $\mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu$ and $\|\Sigma^{-\frac{1}{2}}(\xi - \mu)\|_2 \le r$ almost surely,

which follows from

$$\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2 = \frac{\|\Sigma^{\frac{1}{2}}a(x)\|_2}{a(x)^T\Sigma a(x)}\left|\zeta - r\|\Sigma^{\frac{1}{2}}a(x)\|_2\right| \le r.$$

Furthermore, from Lemma 2.1 of [23], we know that $\mathbb{P}_\xi$ is log-concave. Hence, $\mathbb{P}_\xi \in \mathcal{D}_\xi^1$ and

$$\inf_{\mathbb{P}_\xi\in\mathcal{D}_\xi^1} \mathbb{P}_\xi\left(a(x)^T\xi \le b(x)\right) \le \inf_{\mathbb{P}_\zeta\in\mathcal{D}_\zeta^1} \mathbb{P}_\zeta\left(\zeta \le g(x)\right).$$

■

4.4 Reformulation and Approximation

We next present an SOC sandwich approximation of (4.3) under ambiguity set $\mathcal{D}_\xi^1$ and an SOC exact reformulation under ambiguity set $\mathcal{D}_\xi^2$.

4.4.1 Sandwich Approximation

First, we derive a conservative approximation for (4.3) under $\mathcal{D}_\xi^1$ by relaxing $\mathcal{P}_1^{\text{S}}$ to a set consisting of all $\mathbb{P}_\xi$ with a log-concave cumulative distribution functions (CDFs). Letting $\mathcal{P}^{\text{L}}$ represent the set of all $\mathbb{P}_\xi$ with a log-concave CDFs, we define

$$\begin{aligned}\mathcal{D}_\zeta^{\text{L}} := \{\mathbb{P}_\zeta \in \mathcal{P}^{\text{L}} :\ & \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = r\|\Sigma^{\frac{1}{2}}a(x)\|_2, \\ & 0 \le \zeta \le 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \text{ almost surely}\}. \end{aligned} \tag{4.9}$$

From Theorem 1 of [4], we have $\mathcal{D}_\zeta^1 \subseteq \mathcal{D}_\zeta^{\text{L}}$ and so

$$\inf_{\mathbb{P}_\zeta\in\mathcal{D}_\zeta^{\text{L}}} \mathbb{P}_\zeta\left(\zeta \le g(x)\right) \le \inf_{\mathbb{P}_\zeta\in\mathcal{D}_\zeta^1} \mathbb{P}_\zeta\left(\zeta \le g(x)\right). \tag{4.10}$$

Theorem IV.3. *If $\epsilon \le \frac{1}{4}$, then (4.3) under $\mathcal{D}_\xi^1$ is implied by the following SOC*

constraint:

$$a(x)^T\mu + \left[1 - \frac{2\log(1-\epsilon)}{d^*}\right] r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \leq b(x), \tag{4.11}$$

where d^* *is the unique root of function* $\exp\{d\} - d/2 = 1$ *on the interval* $(-\infty, 0)$ *and* log *represents the natural logarithm.*

Proof: From the above discussion, it is clear that (4.3) under $\mathcal{D}_\xi^1$ is implied by

$$\inf_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta^{\text{L}}} \mathbb{P}_\zeta\left(\zeta \leq g(x)\right) \geq 1 - \epsilon. \tag{4.12}$$

Define $\pi = r\|\Sigma^{\frac{1}{2}}a(x)\|_2$ and the CDF of ζ as $F_\zeta(z) = \mathbb{P}_\zeta(\zeta \leq z)$. Then, $F_\zeta(z)$ is log-concave in z for any $\mathbb{P}_\zeta \in \mathcal{D}_\zeta^{\text{L}}$.

We claim that, without loss of optimality, we can focus on those $\mathbb{P}_\zeta$ with $\log(F_\zeta(z))$ being the minimum of an affine function of z and the constant-zero function when computing $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta^{\text{L}}} \mathbb{P}_\zeta\left(\zeta \leq g(x)\right)$. To see this, we pick any ζ with $\mathbb{P}_\zeta \in \mathcal{D}_\zeta^{\text{L}}$. Then $\log(F_\zeta(z))$ is concave and non-decreasing. Consider a tangent of $\log(F_\zeta(z))$ at $z^* = g(x)$, i.e., $G(z) = k(z - z^*) + \log(F_\zeta(z^*))$ with $k \in \partial \log(F_\zeta(z^*))$. As $\log(F_\zeta(z))$ is nondecreasing and concave, we have $k \geq 0$ and $G(z) \geq \log(F_\zeta(z))$ for all $z \in [0, 2\pi]$. Define $\hat{F}_\zeta(z) = \min\{1, \exp(G(z))\}$, which satisfies the property of a log-concave CDF and yields a probability measure $\hat{\mathbb{P}}_\zeta$. In addition, we have $\hat{\mathbb{P}}_\zeta\left(\zeta \leq z^*\right) = \mathbb{P}_\zeta\left(\zeta \leq z^*\right)$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] \geq \mathbb{E}_{\hat{\mathbb{P}}_\zeta}[\zeta]$. Then, we manipulate $G(z)$ via the following two steps.

1. Increase the horizontal intercept to 2π with vertical intercept fixed, and

2. decrease the vertical intercept towards $-\infty$ with the horizontal intercept fixed.

Both actions will decrease $\hat{\mathbb{P}}_\zeta\left(\zeta \leq z^*\right)$ and increase $\mathbb{E}_{\hat{\mathbb{P}}_\zeta}[\zeta]$. So we could stop as soon as $\hat{\mathbb{P}}_\zeta\left(\zeta \leq z^*\right) \leq \mathbb{P}_\zeta\left(\zeta \leq z^*\right)$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mathbb{E}_{\hat{\mathbb{P}}_\zeta}[\zeta]$. This proves the claim.

Based on the claim, we recast the left side of (4.12) as follows.

$$\min_{c,d} \quad \min\{1, \exp\{cz^* + d\}\} \tag{4.13}$$

$$\text{s.t.} \quad \int_0^{2\pi} [1 - \min\{1, \exp\{cz + d\}\}]\, dz = \pi \tag{4.14}$$

$$\exp\{c(2\pi) + d\} \geq 1 \tag{4.15}$$

$$c > 0, \quad d < 0, \tag{4.16}$$

where (4.14) enforces $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \pi$, (4.15) enforces $0 \leq \zeta \leq 2\pi$ almost surely, and (4.16) makes sure that CDF $\min\{1, \exp\{cz+d\}\}$ is nondecreasing and nontrivial. Note that (4.15) is equivalent to $-d/c \leq 2\pi$. Then, it follows that (4.14) is equivalent to

$$\begin{aligned}
\pi &= \int_0^{2\pi} [1 - \min\{1, \exp\{cz + d\}\}]\, dz \\
&= \int_0^{-d/c} [1 - \exp\{cz + d\}]\, dz \\
&= \frac{1}{c} [\exp\{d\} - d - 1],
\end{aligned}$$

and so $c = (1/\pi)[\exp\{d\} - d - 1]$. Thus, (4.15) is equivalent to $2\pi \geq -d/c = -d\pi/[\exp\{d\} - d - 1]$, or $\exp\{d\} \geq 1 + d/2$ because $\exp\{d\} - d - 1 > 0$ when $d < 0$. It follows that the optimal objective value of problem (4.13)–(4.16) equals that of the following problem:

$$\min_{d<0} \quad \min\left\{1, \exp\left\{\frac{\exp\{d\} - d - 1}{\pi} z^* + d\right\}\right\} \tag{4.17}$$

$$\text{s.t.} \quad \exp\{d\} \geq \frac{d}{2} + 1. \tag{4.18}$$

We analyze the objective function in (4.17) by considering the following two scenarios.

- Scenario 1. If $z^* \geq 2\pi$, then

$$\begin{aligned}
&(\exp\{d\} - d - 1)\left(\frac{z^*}{\pi}\right) + d \\
\geq &(\exp\{d\} - d - 1)\left(\frac{2\pi}{\pi}\right) + d \\
= &2\left(\exp\{d\} - \frac{d}{2} - 1\right) \geq 0
\end{aligned}$$

  for all $d < 0$ and $\exp\{d\} \geq d/2 + 1$. Then, the optimal objective value of problem in (4.13)–(4.16) equals 1. Note that this makes sense because when $z^* \geq 2\pi$, we always have $\zeta \leq z^*$ for any $\zeta \in [0, 2\pi]$.

- Scenario 2. If $z^* < 2\pi$, then by the definition of d^*

$$\begin{aligned}
&(\exp\{d^*\} - d^* - 1)\left(\frac{z^*}{\pi}\right) + d^* \\
< &(\exp\{d^*\} - d^* - 1)\left(\frac{2\pi}{\pi}\right) + d^* \\
= &2\left(\exp\{d^*\} - \frac{d^*}{2} - 1\right) = 0
\end{aligned}$$

  Hence, there exists a $d < 0$ with $\exp\{d\} \geq d/2 + 1$ such that $(\exp\{d\} - d - 1)(z^*/\pi) + d < 0$. It follows that the objective function in (4.17) is equivalent to

$$\min_{d<0} \ \exp\left\{(\exp\{d\} - d - 1)\left(\frac{z^*}{\pi}\right) + d\right\}.$$

Finally, we recast (4.12) by discussing the following two scenarios.

- Scenario 1. If $z^* \geq 2\pi$, then (4.12) always holds.

- Scenario 2. If $z^* < 2\pi$, then (4.12) holds if and only if the optimal objective value of problem (4.17)–(4.18) is greater than or equal to $1 - \epsilon$, or equivalently,

$$z^* \geq \max_{d<0;\ \exp\{d\} \geq d/2+1} \left\{\frac{\log(1-\epsilon) - d}{\exp\{d\} - d - 1}\right\} \pi.$$

Note that $d < 0$ and $\exp\{d\} \geq d/2 + 1$ is equivalent to $d \leq d^*$ because $h(d) = \exp\{d\} - d/2 - 1$ is convex in d with roots $d = 0$ and $d = d^* \approx -1.59$. In addition, $H(d) = \frac{\log(1-\epsilon)-d}{\exp\{d\}-d-1}$ is nondecreasing in d because $\epsilon \leq 1/4$. It follows that (4.12) is equivalent to

$$z^* \geq \left\{ \frac{\log(1-\epsilon) - d^*}{\exp\{d^*\} - d^* - 1} \right\} \pi = \left\{ 2 - \frac{2\log(1-\epsilon)}{d^*} \right\} \pi.$$

Note that $2 - \frac{2\log(1-\epsilon)}{d^*} < 2$ as $d^* < 0$ and $1 - \epsilon \in (0, 1)$.

Summarizing the above two scenarios, (4.12) is equivalent to

$$z^* \geq \left\{ 2 - \frac{2\log(1-\epsilon)}{d^*} \right\} \pi.$$

The proof is complete given the definition of z^*. ∎

Second, we derive a relaxing approximation for (4.3) by focusing on a particular distribution in $\mathcal{D}_\zeta^1$. More specifically, let $\mathbb{P}_\zeta^{\mathrm{U}}$ represent the uniform distribution on the interval $[0, 2r\|\Sigma^{1/2}a(x)\|_2]$, then (4.3) implies

$$\mathbb{P}_\zeta^{\mathrm{U}}\{\zeta \leq g(x)\} \geq 1 - \epsilon. \tag{4.19}$$

which can be recast as

$$a(x)^T\mu + r(1-2\epsilon)\|\Sigma^{\frac{1}{2}}a(x)\|_2 \leq b(x). \tag{4.20}$$

The conservative approximation and this relaxing approximation are used as the sandwich approximation.

Third, we consider another relaxing approximation of (4.3) by restricting $\mathbb{P}_\xi$ to be normally-distributed with mean μ and covariance matrix Σ. Then, based on existing

results (e.g., [6]), (4.3) implies the following SOC constraint:

$$a(x)^T\mu + \Phi_N^{-1}(1-\epsilon)\|\Sigma^{\frac{1}{2}}a(x)\|_2 \leq b(x). \tag{4.21}$$

where Φ_N is the inverse CDF of the standard normal distribution. This relaxing approximation is used as a benchmark.

4.4.2 Exact Reformulation

We next derive an exact reformulation for (4.3) under $\mathcal{D}_\xi^2$, which we use as the DRCC benchmark.

Theorem IV.4. *If $\epsilon < \frac{1}{2}$, then* (4.3) *under $\mathcal{D}_\xi^2$ is equivalent to the following SOC constraint:*

$$a(x)^T\mu + r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \leq b(x). \tag{4.22}$$

Proof: We prove by contradiction.

First, if $g(x) < r\|\Sigma^{\frac{1}{2}}a(x)\|_2$, then we consider $\hat{\mathcal{P}}_\zeta \in \mathcal{D}_\zeta^2$ such that $\zeta = r\|\Sigma^{1/2}a(x)\|_2$ almost surely. It follows that $\hat{\mathcal{P}}_\zeta(\zeta \leq g(x)) = 0$ and so $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta^1} \mathbb{P}_\zeta\left(\zeta \leq g(x)\right) = 0 < 1-\epsilon$.

Second, if $r\|\Sigma^{\frac{1}{2}}a(x)\|_2 \leq g(x) < 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2$, then there exists a point $\tilde{\zeta}$ such that $g(x) < \tilde{\zeta} < 2r\|\Sigma^{\frac{1}{2}}a(x)\|_2$. Consider a specific distribution $\hat{\mathbb{P}}_\zeta \in \mathcal{D}_\zeta^2$ that puts equal weight at $\tilde{\zeta}$ and $2r\|\Sigma^{\frac{1}{2}}a(x)\|_2 - \tilde{\zeta}$. We have

$$\inf_{\mathbb{P}_\zeta \in \mathcal{D}_\zeta^2} \mathbb{P}_\zeta\left(\zeta \leq g(x)\right) \leq \hat{\mathbb{P}}_\zeta\left(\zeta \leq g(x)\right) = \frac{1}{2} < 1-\epsilon.$$

Hence, (4.3) is valid if and only if $g(x) \geq 2r\|\Sigma^{1/2}a(x)\|_2$. ∎

Note that (4.3) under $\mathcal{D}_\xi^2$ is conservative because it is invariant for any $\epsilon \in (0, 1/2)$.

4.5 Case Studies

4.5.1 Simulation Setup

We consider the DCOPF problem from [48] and Chapter III, which is similar to that of [9, 96]. Assuming two wind power plants, the random variables are the wind forecast error $\widetilde{W} = [W_1, W_2]^T$. Assuming the system has N_G generators and N_B buses, the design variables are generation $P_G \in \mathbb{R}^{N_G}$, up and down reserve capacities $R^{up} \in \mathbb{R}^{N_G}, R^{dn} \in \mathbb{R}^{N_G}$, and a "distribution vector" $d \in \mathbb{R}^{N_G}$, which allocates real-time supply/demand mismatch to generators providing reserves. The problem is

$$\min\ P_G^T[C_1]P_G + C_2^T P_G + C_R^T(R^{up} + R^{dn}) \tag{4.23}$$

$$\text{s.t.} - P_l \leq A_s P_{\text{inj}} \leq P_l \tag{4.24}$$

$$R = -d(W_1 + W_2) \tag{4.25}$$

$$P_{\text{inj}} = C_G(P_G + R) + C_W(P_W^f + \widetilde{W}) - C_L P_L \tag{4.26}$$

$$\underline{P}_G \leq P_G + R \leq \overline{P}_G \tag{4.27}$$

$$-R^{dn} \leq R \leq R^{up} \tag{4.28}$$

$$\mathbf{1}_{1\times N_G} d = 1 \tag{4.29}$$

$$\mathbf{1}_{1\times N_B}(C_G P_G + C_W P_W^f - C_L P_L) = 0 \tag{4.30}$$

$$P_G \geq \mathbf{0}_{N_G\times 1},\ d \geq \mathbf{0}_{N_G\times 1}, \tag{4.31}$$

$$R^{up} \geq \mathbf{0}_{N_G\times 1},\ R^{dn} \geq \mathbf{0}_{N_G\times 1} \tag{4.32}$$

where $[C_1] \in \mathbb{R}^{N_G\times N_G}$, $C_2 \in \mathbb{R}^{N_G}$, and $C_R \in \mathbb{R}^{N_G}$ are cost parameters. Constraint (6.16a) uses the DC power flow approximation to relate the power injections P_{inj} to the line flows using parameter matrix A_s, which includes the network impedances, and constrains the flows to below to the line limits P_l. Constraint (6.16b) computes the real-time reserve actions R; (6.16c) defines the real-time power injections, where P_W^f is the wind forecast and C_G, C_W, and C_L are matrices that map generators, wind

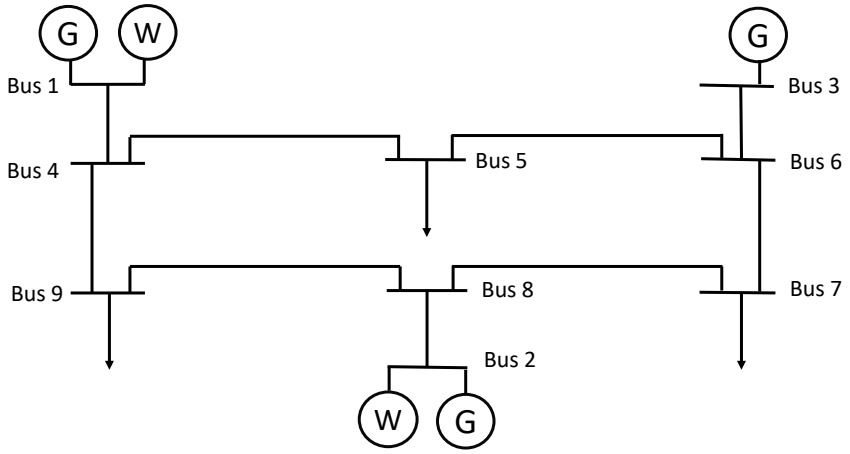


Figure 4.1: Modified IEEE 9-bus system.

power plants, and loads to their buses; (6.16d) constrains generator production to within its limits $[\underline{P}_G, \overline{P}_G]$; (6.16e) requires the real-time reserve actions be within the reserve capacity; (6.16f) normalizes the distribution vector; (6.16g) enforces power balance for the wind forecast; and (6.16h), (6.16i) ensures all decision variables are non-negative. Real-time power balance is enforced by (6.16b) and (6.16f) which require reserve actions to exactly compensate for the wind forecast error. Constraints (6.16a), (6.16d), and (6.16e) are reformulated as chance constraints. Each constraint is enforced individually, not jointly.

We test our approach on the modified IEEE 9-bus system shown in Fig. 4.1. We use the network parameters and generation costs from [108], and set the reserve cost $C_R = 10C_2$. We set the wind power forecasts to $P_W^f = [66.8, 68.1]$ MW. We use the same wind power forecast uncertainty data as in [93], which was generated using a Markov Chain Monte Carlo mechanism [70] applied to real data. We congest the system by increasing each load by 50% and reducing the line limit of the line connecting buses 2 and 8 from 250 MW to 200 MW. We pick this transmission line because it connects a wind power plant to the system and its forecasted power flow is close to its limit. All optimization problems are solved using CVX with the Mosek

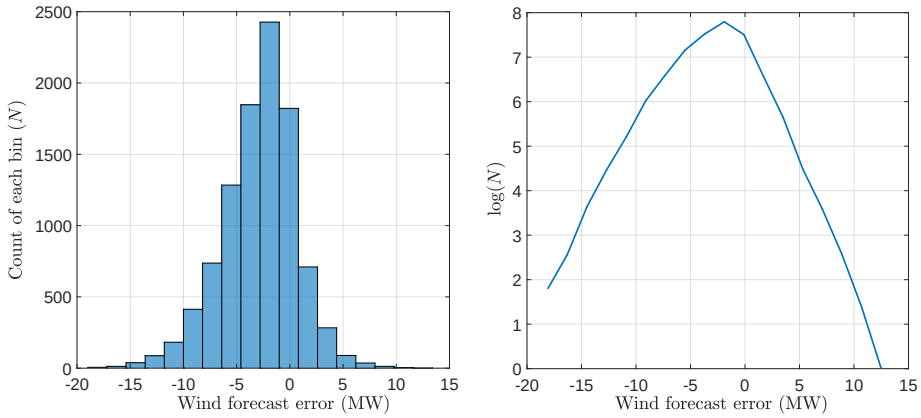


Figure 4.2: Histogram of univariate wind forecast errors and its logarithmic profile.

solver [32, 31].

4.5.1.1 Empirical Wind Forecast Error Distributions

We verify the log-concavity of the wind forecast error distributions using the full data set (10,000 scenarios) with the exception of statistical outliers (total probability$<$ 0.1%). Note that the outliers are also not used when empirically estimating the first-order moment μ, covariance Σ, and radius r, but they are used when evaluating the reliability of the solution.

In Fig. 4.2, we show the histogram of univariate wind forecast errors (i.e., all errors analyzed together) and its logarithmic profile, which appears to be concave. Figure 4.3 shows the histogram of the bivariate wind forecast error (i.e., errors were used to generate paired forecasts for the two wind power plants) and its logarithmic profile, which also appears to be concave. Figures 4.2 and 4.3 empirically justify our assumption that $\mathbb{P}_\xi$ in the ambiguity set $\mathcal{D}_\xi^1$ is log-concave.

4.5.2 Results

1. Objective Cost We first compare the optimal objective cost across the following four DRCC OPF formulations:

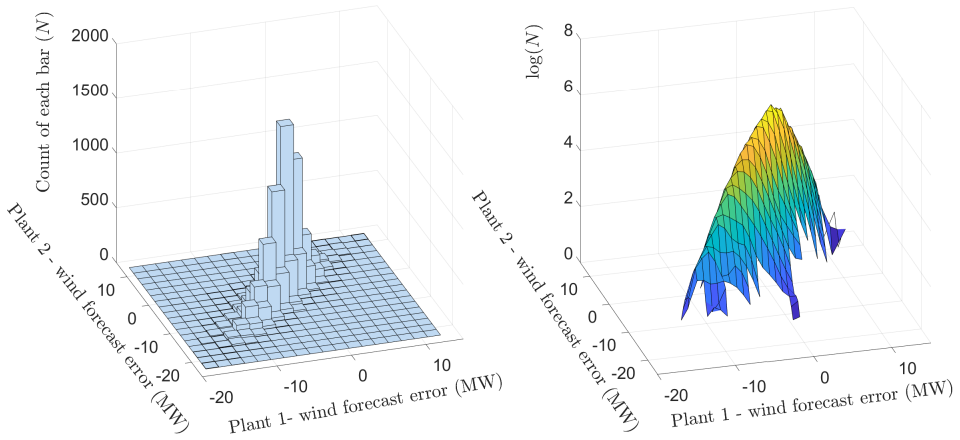


Figure 4.3: Histogram of bivariate wind forecast errors and its logarithmic profile.

- Gaussian: relaxing approximation in which $\mathbb{P}_\xi$ is normally-distributed, given in (4.21).
- RA: relaxing approximation of $\mathcal{D}_\xi^1$, given in (4.20).
- CA: conservative approximation of $\mathcal{D}_\xi^1$, given in (4.11).
- BM: benchmark based on $\mathcal{D}_\xi^2$, given in (4.22).

For each formulation, we test two confidence levels $\epsilon = 5\%, 10\%$ and four data sizes $N = 500$, 1000, 2000, and 5000. We replicate the test associated with each (ϵ, N) combination 20 times by re-drawing N error scenarios. Table 4.1 displays the minimum, average, and maximum objective costs for each formulation over all replicates. BM yields the same optimal objective costs at different confidence levels as it is invariant under any $\epsilon \in (0, 1/2)$ (see Theorem IV.4). We also observe that, for all formulations, the range of the optimal objective cost becomes narrower as the data size grows because of better estimation of μ, Σ, and r. In addition, we observe that the optimal objective cost of a purely moment-based formulation (i.e., Gaussian) depends less on the data size as compared with the formulations using both moment and support information. The reason is that the support is more sensitive to outliers than the moments are. We observe that the optimal objective cost increases in the

Table 4.1: Optimal Objective Cost of DRCC OPF under Various Formulations, Data Sizes, and Confidence Levels

	Data size	500		1000		2000		5000	
	1-ϵ	95%	90%	95%	90%	95%	90%	95%	90%
Gaussian	min	4891	4849	4894	4851	4895	4852	4903	4858
	avg	4903	4858	4907	4862	4907	4861	4906	4861
	max	4927	4877	4925	4876	4915	4868	4913	4866
RA	min	5078	5035	5192	5134	5748	5623	5985	5830
	avg	5468	5377	5918	5772	6001	5845	6046	5885
	max	6069	5904	6119	5948	6110	5940	6121	5950
CA	min	5094	5064	5212	5173	5793	5708	6040	5934
	avg	5501	5439	5970	5870	6057	5950	6104	5994
	max	6127	6015	6180	6063	6171	6054	6183	6065
BM	min	5122	5122	5250	5250	5875	5875	6141	6141
	avg	5561	5561	6066	6066	6159	6159	6211	6211
	max	6235	6235	6292	6292	6282	6282	6295	6295

same order as the list above due to the increasing size of the corresponding ambiguity set.

We next employ the full data set to construct the ambiguity sets. Then, we solve all formulations with varying confidence levels from 75% to 99% (i.e., $\epsilon \in [1, 25]\%$). Figure 4.4 displays the optimal objective costs and the generation cost as a percent of the total cost for each formulation. We observe that the optimal objective costs of RA and CA increase at a faster rate than that of Gaussian. Consistent with Table 4.1, BM yields a constant optimal objective cost independent of ϵ. At high confidence levels ($\epsilon \leq 5\%$), the objective costs of RA and CA become closer and they both converge to that of BM as ϵ decreases. This indicates that the sandwich approximations become tighter as ϵ decreases. Meanwhile, the impact of the log-concavity assumption on the ambiguity sets becomes weaker. The optimal objective cost increases in the same order as in Table 4.1 (i.e., Gaussian, RA, CA, and BM). We also observe that the generation cost percentage of all formulations except BM decreases as the confidence level increases because we need to procure more reserve capacity to balance wind forecast error at a higher confidence levels. More reserve

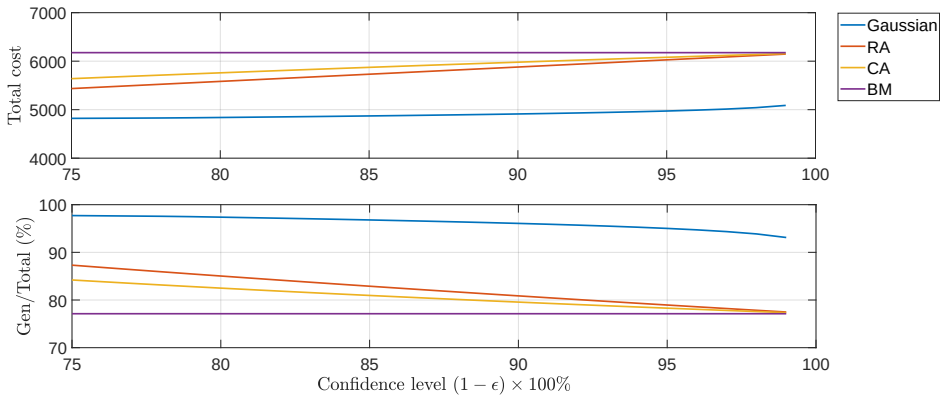


Figure 4.4: Optimal objective costs and generation cost percentage at various confidence levels $(1-\epsilon)$.

capacity is required by more conservative approaches (i.e., reserve capacity increases and generation cost percentage decreases in the following order: Gaussian, RA, CA, and BM). We also observe a similar convergence on the generation cost percentages associated with RA, CA, and BM as the confidence level increases.

2. Reliability We evaluate the empirical reliability of all formulations via an out-of-sample Monte Carlo analysis for data sizes $N = 500,\ 5000$ (used to generate the statistical information needed in the formulation) and confidence levels $1-\epsilon = 95\%, 75\%$. For each formulation, we select the solution associated with the minimum optimal objective value reported in Table 4.1 and test it on 20 randomly-generated groups of 1000 out-of-sample wind forecast errors. We define the empirical reliability as the percentage of errors for which all chance constraints are satisfied by the selected optimal solution. Tables 4.2 and 4.3 show the results. With a data size of 5000, we get an accurate estimate of the statistical information and hence RA, CA, and BM all achieve high overall reliability, while Gaussian does not meet the reliability requirement. With a data size of 500, the reliability drops since the statistical information estimated is less accurate, but RA and CA still give sufficiently high reliability and are less conservative than BM. This demonstrates that the sandwich approximation

Table 4.2: Overall Reliability (%) with Data Size 5000

$1-\epsilon$		Gaussian	RA	CA	BM
95%	min	87.5	99.7	99.8	99.8
	avg	88.7	99.9	99.9	100
	max	90.2	100	100	100
75%	min	53.2	99.5	99.5	99.8
	avg	55.2	99.8	99.8	100
	max	57.5	100	100	100

Table 4.3: Overall Reliability (%) with Data Size 500

$1-\epsilon$		Gaussian	RA	CA	BM
95%	min	85.4	98.3	98.5	98.9
	avg	86.9	99.1	99.3	99.5
	max	89.2	99.6	99.8	99.8
75%	min	51.6	88.7	94.7	98.9
	avg	53.4	90	95.9	99.5
	max	55.4	92.1	96.7	99.8

(RA and CA) provides a good trade-off between cost and reliability.

4.6 Conclusion

In this chapter, we derived distributionally robust chance constraints corresponding to log-concave uncertainty distributions with known first-order moment and ellipsoidal support. We derived a projection property and a tractable sandwich approximation (i.e., relaxing and conservative approximations) of the distributionally robust chance constraints as second-order cone constraints. We compared the approximations to a benchmark using only moment and support information and another that assumed normally-distributed uncertainty. The optimal objective costs of the sandwich approximation depend on the accuracy of the moment and support information. As the confidence level increases, the gap between the conservative and relaxing approximation reduce and the effect of the log-concavity assumption becomes weaker. The sandwich approximation provides a better trade-off between optimal objective cost and reliability than either benchmark, where the benchmark assuming normal

distributions is cheap but not sufficiently reliable and the benchmark assuming known first-order moment and support (without log-concavity) is expensive. Our approach works well even if only a small number of data points are available to estimate the statistical parameters.

CHAPTER V

Distributionally Robust Chance Constrained Optimal Power Flow Assuming Unimodal Distributions with Misspecified Modes

Chance constrained optimal power flow (CC-OPF) formulations have been proposed to minimize operational costs while controlling the risk arising from uncertainties like renewable generation and load consumption. To solve CC-OPF, we often need access to the (true) joint probability distribution of all uncertainties, which is rarely known in practice. A solution based on a biased estimate of the distribution can result in poor reliability. To overcome this challenge, recent work has explored distributionally robust chance constraints, in which the chance constraints are satisfied over a family of distributions called the ambiguity set. Commonly, ambiguity sets are only based on moment information (e.g., mean and covariance) of the random variables; however, specifying additional characteristics of the random variables reduces conservatism and cost. In this chapter, we consider ambiguity sets that additionally incorporate unimodality information. In practice, it is difficult to estimate the mode location from the data and so we allow it to be potentially misspecified. We formulate the problem and derive a separation-based algorithm to efficiently solve it. Finally, we evaluate the performance of the proposed approach on a modified IEEE-30

bus network with wind uncertainty and compare with other distributionally robust approaches. We find that a misspecified mode significantly affects the reliability of the solution and the proposed model demonstrates a good trade-off between cost and reliability. The main content of this chapter is summarized in the following paper.

1. B. Li, R. Jiang, and J.L. Mathieu. Distributionally robust chance constrained optimal power flow assuming unimodal distributions with misspecified modes. *IEEE Trans Control of Network Systems (Submitted)*, 2018

5.1 Introduction

With higher penetrations of renewable generation, uncertainties have increasing influence on power system operation and hence need to be carefully considered in scheduling problems, such as optimal power flow (OPF). To manage the risk arising from uncertainties, different stochastic OPF approaches have been studied. Among these formulations, CC-OPF has been proposed to directly control the constraint violation probability below a pre-defined threshold [104, 42, 96, 9, 79, 93, 52]. Traditional methods to solve chance constrained programs require knowledge of the joint probability distribution of all uncertainties, which may be unavailable or inaccurate. However, biased estimate may yield poor out-of-sample performance. Randomized techniques such as scenario approximation [17, 58], which provides a priori guarantees on reliability, require the constraints to be satisfied over a large number of uncertainty samples. The solutions from these approaches are usually overly conservative with high costs [107, 52]. Another popular approach is to assume that the uncertainties follow a parametric distribution such as Gaussian [9, 79, 52]. The resulting CC-OPF is often easier to solve but the solution may have low reliability unless the assumed probability distribution happens to be close to the true one.

As an alternative, distributionally robust chance constrained (DRCC) OPF models do not depend on a single estimate of the probability distribution [102, 33, 88,

56, 50, 49, 80, 107]. More specifically, DRCC models consider a family of distributions, called the ambiguity set, that share certain statistical characteristics and requires that the chance constraint holds with respect to all distributions within the ambiguity set [29, 22, 87, 44]. Most existing work characterizes the ambiguity set based on moment information obtained from historical data of the uncertainty (see, e.g., [107, 102, 56, 88]). For example, a commonly adopted ambiguity set consists of all distributions whose mean and covariance agree with their corresponding sample estimates [107, 102, 88]. Many uncertainty distributions (e.g., those associated with wind forecast error) are unimodal and so, recently, unimodality has been incorporated to strengthen the ambiguity set and reduce the conservatism of DRCC models [88, 49, 80]. However, as compared to the moments, the mode location is more likely to be misspecified in sample-based estimation.

In this chapter, we study a DRCC model with an ambiguity set based on moment and unimodality information with a potentially misspecified mode location. To the best of our knowledge, this is the first work discussing misspecification of a value related to a structural property, though others have considered misspecification of moments [22, 56, 29, 44, 107] and misspecification of distributions [33, 90]. Our main theoretical result shows that the distributionally robust chance constraints can be recast as a set of second-order conic (SOC) constraints. Furthermore, we derive an iterative algorithm to accelerate solving the reformulation. In this algorithm, we begin with a relaxed formulation, and in each iteration, we efficiently find the most violated SOC constraint, if any, or terminate with a globally optimal solution. We apply the theoretical results to a direct current (DC) OPF problem and conduct a case study using a modified IEEE 30-bus system with wind power. We compare our results (operational cost, reliability, computational time, and optimal solutions) to those obtained using four alternative ambiguity sets [49, 87, 80, 107].

The remainder of this chapter is organized as follows. Section 5.2 empirically

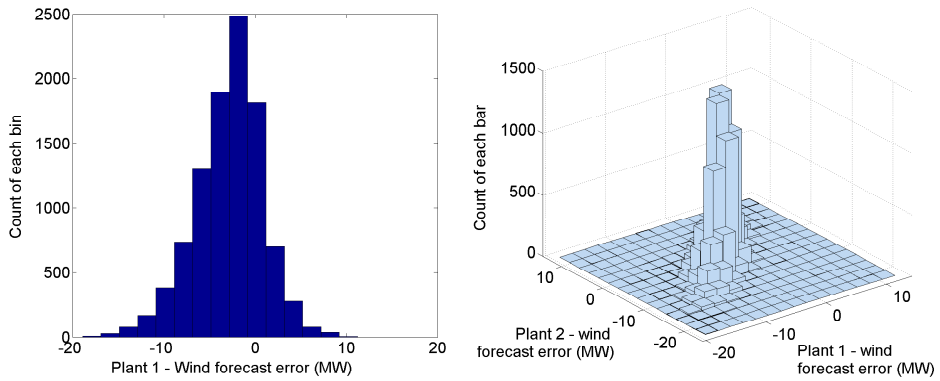


Figure 5.1: Histograms of univariate and bivariate wind forecast errors (15 bins).

verifies the (multivariate) unimodality of wind forecast errors and explores misspecification of the mode location. The proposed DRCC model and ambiguity set are introduced in Section 5.3 and the main theoretical results are presented in Section 5.4. Section 5.5 includes the case studies and Section 5.6 provides the supporting mathematical proofs.

5.2 Validation and Misspecification Error on Unimodality

In this section, we first empirically verify the unimodality of wind forecast error distributions using 10,000 data samples from [93, 52] with statistical outliers omitted (total probability $< 0.1\%$). The samples were generated using a Markov Chain Monte Carlo mechanism [70] based on real data that includes both hourly forecast and actual wind generation in Germany. In Fig. 5.1, we depict the histograms of univariate and bivariate wind forecast errors with 15 bins. Both histograms empirically justify our assumption that the probability distribution of wind forecast errors is unimodal.

Next, we empirically evaluate the errors of mean and mode estimates (i.e., the peak location in the histogram). We randomly extract 100 groups of samples, each group containing 500 data points, from the wind forecast error data pool. For each group of samples, we estimate the mean by taking sample averages and estimate the

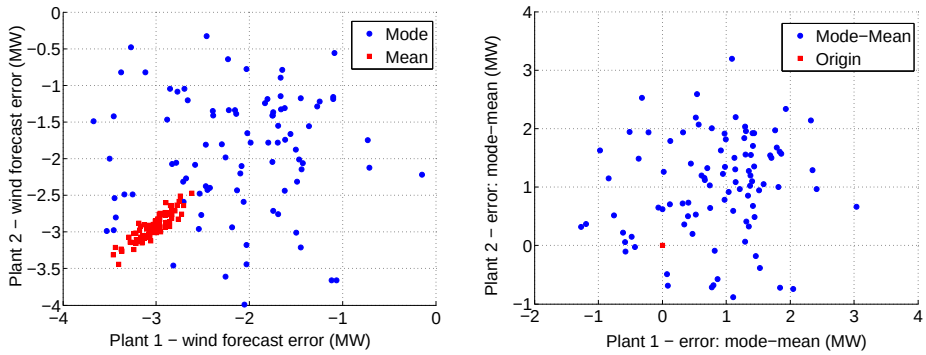


Figure 5.2: Scatter plots of mode and mean estimates from data samples (left) and mode vs. mean differences (right).

mode by identifying the center of the highest bin in the 15-bin histogram. In Fig. 5.2, we plot all the mean and mode estimates and the differences between them. From the left subfigure, we observe that sampling errors have larger impacts on mode estimates than on mean estimates. From the right subfigure, we observe that the mode estimate can deviate from the corresponding mean estimate in all directions. This indicates the importance of considering the misspecification of mode location in DRCC models, because the mode-mean deviation shows the skewness of the uncertainty. As a result, if we misspecify the mode location (e.g., by modeling a right-skewed distribution as a left-skewed one, see Section 5.3.0.2 for an example), then we may mistakenly relax the chance constraint and get poor out-of-sample performance.

5.3 Ambiguity Sets with Misspecified Mode

Similar to other chapters, we consider the following physical constraint under uncertainty:

$$a(x)^\top \xi \leq b(x), \tag{5.1}$$

where $x \in \mathbb{R}^l$ represents an l-dimensional decision variable, and $a(x) : \mathbb{R}^l \to \mathbb{R}^n$ and $b(x) : \mathbb{R}^l \to \mathbb{R}$ represent two affine functions of x. Uncertainty $\xi \in \mathbb{R}^n$ represents an n-dimensional random vector defined on probability space $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\xi)$ with Borel

σ-algebra $\mathcal{B}^n$ and probability distribution $\mathbb{P}_\xi$.

To manage constraint violations due to uncertainty, we define a probability threshold $1-\epsilon$ and obtain the following chance constraint [18, 66]:

$$\mathbb{P}_\xi\left(a(x)^\top \xi \le b(x)\right) \ge 1-\epsilon, \tag{5.2}$$

where $1-\epsilon$ normally takes a large value (e.g., 0.99). With an ambiguity set $\mathcal{D}_\xi$ consisting of plausible candidates of $\mathbb{P}_\xi$, we further obtain the following distributionally robust chance constraint:

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\left(a(x)^\top \xi \le b(x)\right) \ge 1-\epsilon. \tag{5.3}$$

5.3.0.1 Ambiguity Sets

In this chapter, we consider three ambiguity sets, denoted as $\mathcal{D}_\xi^i$ for $i = 1, 2, 3$, that are defined by a combination of moment and unimodality information. Precisely, we consider α-Unimodality as in Chapter III.

From the definition, we notice that α parameterizes the "degree of unimodality." When $\alpha = n = 1$, the definition coincides with the classical univariate unimodality with mode 0. When $\alpha = n > 1$, the density function of ξ (if exists) peaks at the mode and is non-increasing in any directions moving away from the mode. As $\alpha \to \infty$, the requirement of unimodality gradually relaxes and eventually vanishes. Next, we define the following three ambiguity sets:

Ambiguity set 1: (moment information only)

$$\mathcal{D}_\xi^1 := \left\{\mathbb{P}_\xi \in \mathcal{P}^n : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma\right\}, \tag{5.4}$$

Ambiguity set 2: (moment and α-unimodality, fixed mode)

$$\mathcal{D}_\xi^2 := \left\{\mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 :\ \mathcal{M}(\xi) = m_t\right\}, \tag{5.5}$$

Ambiguity set 3: (moment and α-unimodality, misspecified mode)

$$\mathcal{D}_\xi^3 := \left\{\mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 :\ \mathcal{M}(\xi) \in \Xi\right\}, \tag{5.6}$$

where $\mathcal{P}_\alpha^n$ and $\mathcal{P}^n$ denote all probability distributions on $\mathbb{R}^n$ with and without the requirement of α-unimodality respectively; μ and Σ denote the first and second moments of ξ; and $\mathcal{M}(\xi)$ denotes a function returning the true mode location of ξ with m_t and Ξ representing a single mode value and a connected and compact set. The compact set can be constructed using possible mode estimates calculated from samples of historical data.

Among these three ambiguity sets, we use $\mathcal{D}_\xi^1$ as a benchmark. Set $\mathcal{D}_\xi^2$ is a special case of $\mathcal{D}_\xi^3$, i.e., Ξ only contains a single value m_t. In practice, since the mode estimate is influenced by sampling errors, the mode estimates from data samples are not the same single values but distribute around a certain area. The shape of this area decides the underlying structural skewness in the uncertainty distribution. Hence, we compare $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^3$ to see how misspecified mode estimates affect the DRCC problem. In this work, we do not additionally consider misspecified moments since this topic has been well-studied [22, 56, 29, 44] and our main results can be easily extended based on these existing works.

5.3.0.2 Impact of Inaccurate Mode Estimate

We use a simple example to illustrate the impact of an inaccurate mode estimate. We assume random variable ζ follows distribution $\mathbb{P}_{\zeta_1}$. $\mathbb{P}_{\zeta_2}$ is a biased estimate of $\mathbb{P}_{\zeta_1}$ due to sampling errors. Both distributions are illustrated in Fig. 5.3, where each has

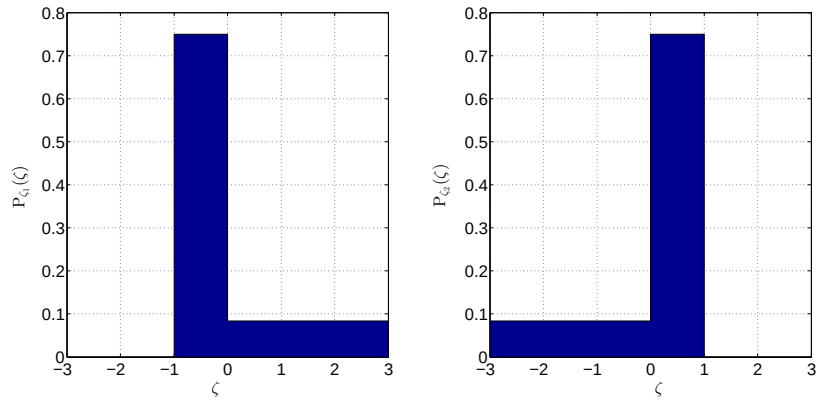


Figure 5.3: True estimate $\mathbb{P}_{\zeta_1}$ and biased estimate $\mathbb{P}_{\zeta_2}$

zero mean and unit variance. However, $\mathbb{P}_{\zeta_1}$ is right-skewed with mode at -1 and $\mathbb{P}_{\zeta_2}$ is left-skewed with mode at 1. Suppose that we try to reformulate $\mathbb{P}_{\zeta}(\zeta \leq z) \geq 90\%$. Based on the given distributions, we find $z \geq 1.8$ from the correct distribution $\mathbb{P}_{\zeta_1}$ and $z \geq 0.925$ from the biased distribution $\mathbb{P}_{\zeta_2}$. In this example, we observe that a misspecified mode estimate could shrink the 90% confidence bound by almost a half and significantly decrease the reliability of the solution to the chance constraint.

5.4 Iterative Solving Algorithm with Separation

5.4.0.3 Assumptions and previous results

To compute the exact reformulation of distributionally robust chance constraints with various ambiguity sets, we make the following assumptions same to Chapter IV.

Assumption V.1. *For $\mathcal{D}_{\xi}^{2}$, we assume that*

$$\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^{\top}) \succ \frac{1}{\alpha^{2}}(\mu - m_t)(\mu - m_t)^{\top}.$$

Similarly, for $\mathcal{D}_\xi^3$, we assume that, $\forall m \in \Xi$,

$$\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) \succ \frac{1}{\alpha^2}(\mu - m)(\mu - m)^\top.$$

Assumption V.2. *For $\mathcal{D}_\xi^2$, we assume that $a(x)^\top m_t \leq b(x)$. Similarly, for $\mathcal{D}_\xi^3$, we assume that $a(x)^\top m \leq b(x)$, $\forall m \in \Xi$.*

Both assumptions are standard in the related literature [35, 74, 75, 49]. Assumption V.1 ensures that the corresponding $\mathcal{D}_\xi^i \neq \emptyset$. Assumption V.2 ensures that the constraint is satisfied at the mode. Furthermore, we assume $\epsilon < 0.5$ and $\alpha \geq 1$, since in practice the uncertainties will at least be univariate-unimodal.

Reformulations of (6.7) under $\mathcal{D}_\xi^1$ and $\mathcal{D}_\xi^2$ are derived in previous work.

Theorem V.3. *(Theorem 2.2 in [98]) With $\mathcal{D}_\xi^1$, (6.7) can be exactly reformulated as*

$$\sqrt{\left(\frac{1-\epsilon}{\epsilon}\right)a(x)^\top(\Sigma - \mu\mu^\top)a(x)} \leq b(x) - a(x)^\top\mu. \tag{5.7}$$

Theorem V.4. *With $\mathcal{D}_\xi^2$, (6.7) can be exactly reformulated as*

$$\begin{aligned}&\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda_t a(x)\| \leq \tau\left(b(x) - \mu^\top a(x)\right)\\ &+\left(\tau - \frac{\alpha+1}{\alpha}\right)(\mu - m_t)^\top a(x), \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha},\end{aligned} \tag{5.8}$$

where $\Lambda_t := \left(\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) - \frac{1}{\alpha^2}(\mu - m_t)(\mu - m_t)^\top\right)^{1/2}$. This result is adapted from Theorem 1 in [49] and Theorem III.8 in Chapter III by assuming $m_t \neq 0$.

Since parameter τ has an infinite number of choices, the reformulation in Theorem V.4 also involves an infinite number of SOC constraints. Here we obtain a similar result for the generalized ambiguity set $\mathcal{D}_\xi^3$.

5.4.0.4 Reformulation for $\mathcal{D}_\xi^3$

We now present the reformation with $\mathcal{D}_\xi^3$, which is based on Theorem V.4:

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda a(x)\| \leq \left(\tau - \frac{\alpha+1}{\alpha}\right)(\mu - m)^\top a(x)$$
$$+\tau\left(b(x) - \mu^\top a(x)\right),\ \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha},\ \forall m \in \Xi, \tag{5.9}$$

$$a(x)^\top m \leq b(x),\ \forall m \in \Xi, \tag{5.10}$$

where $\Lambda := \left(\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) - \frac{1}{\alpha^2}(\mu - m)(\mu - m)^\top\right)^{1/2}$ and (5.10) comes from Assumption V.2.

Compared to (6.8), (5.9) is more complicated with two parameters m and τ each with an infinite number of choices. To solve an optimization problem with (5.9), we propose an iterative solving algorithm given in Algorithm 1.

Algorithm 1: Iterative solving algorithm

Initialization: $i = 1$, $\tau_0 = \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$, and $m_0 = \{\text{any singular point in } \Xi\}$;
Iteration i:
Step 1: Solve the reformulated optimization problem with (5.9) using τ_j and m_j for all $j = 0, \ldots, i-1$ and obtain optimal solution x_i^*. All τ_j and m_j values are collected from previous iterations;
Step 2: Find worst case τ^* and m^* that result in the largest violation of (5.9) under x_i^*: **IF** m^* and τ^* does not exist, **STOP** and **RETURN** x_i^* as optimal solution; **ELSE GOTO** Step 3;
Step 3: Set $\tau_i = \tau^*$, $m_i = m^*$, and $i = i + 1$;

Note that the reformulated optimization problem in Step 1 contains only SOC constraints.

5.4.0.5 Step 2 of Algorithm 1

The challenge is how to efficiently perform Step 2 of Algorithm 1. In the following, we assume $a(x_i^*) \neq 0$, otherwise (5.9) is satisfied with x_i^* regardless of the values of τ

and m. Next, we define the following terms

$$h = a(x_i^*)^\top(\mu - m)/\alpha,\ \tilde{c} = b(x_i^*) - \mu^\top a(x_i^*),$$
$$\tilde{R} = \sqrt{a(x_i^*)^\top \left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top)a(x_i^*)},$$
$$g(\tau) = \sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}},\ f(\tau) = -(\alpha\tau - \alpha - 1).$$

Since $m \in \Xi$, we have $h \in [\underline{h}, \overline{h}]$ where

$$\overline{h} = \max_{m\in\Xi} a(x_i^*)^\top(\mu - m)/\alpha,\ \underline{h} = \min_{m\in\Xi} a(x_i^*)^\top(\mu - m)/\alpha. \tag{5.11}$$

From Assumption V.1, we have $[\underline{h}, \overline{h}] \in (-\tilde{R}, \tilde{R})$ and transform (5.9) into

$$\left[g(\tau)\sqrt{\tilde{R}^2 - h^2} + f(\tau)h\right] - \tilde{c}\tau \leq 0,\ \forall h \in [\underline{h}, \overline{h}],\ \forall \tau \geq \tau_0. \tag{5.12}$$

Since the left side of (5.12) is not jointly convex or concave in h and τ (see a proof in Section 5.6.0.5), we can not find the global maximum value for the left side by simply checking the boundary values or stationary points. Therefore, we propose the following algorithm to efficiently find the global maximum.

We notice that for given a τ and if $h \in [-\tilde{R}, \tilde{R}]$, the maximum value of $g(\tau)\sqrt{\tilde{R}^2 - h^2} + f(\tau)h$ equals $\tilde{R}\sqrt{g(\tau)^2 + f(\tau)^2}$ with maximizer $\hat{h}(\tau) = \frac{f(\tau)}{\sqrt{g(\tau)^2+f(\tau)^2}}\tilde{R}$. Next, by taking the derivative of $\hat{h}$, we observe that $\hat{h}$ is a strictly decreasing function of τ. Hence, we can compute $\underline{\tau}$ and $\overline{\tau}$ that cause h to reach its boundary values by solving $\hat{h}(\underline{\tau}) = \overline{h}$ and $\hat{h}(\overline{\tau}) = \underline{h}$. Since $\epsilon < 1/2$ and $\alpha \geq 1$, we have $\tau_0 < (\alpha+1)/\alpha$ and hence $\hat{h}(\tau_0) = \tilde{R}$. Then we know $[\underline{\tau}, \overline{\tau}] > \tau_0$ as $[\underline{h}, \overline{h}] < \tilde{R}$.

To efficiently solve these two equalities, we will use a golden section search by first solving for $\overline{\tau}$ on $[\tau_0, \infty]$ and then for $\underline{\tau}$ on $[\tau_0, \overline{\tau}]$. To efficiently apply a golden section

search on $\overline{\tau}$, we need to find a finite upper bound instead of ∞. The following lemma describes the selection of the finite upper bound τ_1 and the best region to conduct the golden section search.

Lemma V.5. *If $\underline{h} \geq 0$, $\tau_1 = \frac{\alpha+1}{\alpha}$. The golden section search of $\overline{\tau}$ can be conducted on $[\tau_0, \frac{\alpha+1}{\alpha}]$. If $\underline{h} < 0$, $\tau_1 = -\left(\underline{h}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2-\underline{h}^2)}} - (\alpha+1)\right)/\alpha$. The search can be conducted on $[\frac{\alpha+1}{\alpha}, \tau_1]$. The proof is given in Section 5.6.0.6.*

Furthermore, from Assumption V.2, we have $\tilde{c} \geq -\alpha\underline{h} \geq -\alpha\overline{h}$.

Based on the threshold values $\overline{\tau}$ and $\underline{\tau}$, we divide our discussion into three cases.

Case 1: If $\tau \in [\tau_0, \underline{\tau}]$, $h^* = \overline{h}$. Then from (5.12), we find

$$g(\tau)\sqrt{\tilde{R}^2 - \overline{h}^2} + f(\tau)\overline{h} - \tilde{c}\tau \leq 0.$$

Then, we transform the above constraint into the following equivalent form:

$$F_1(\tau) = C_1 g(\tau) - (\tilde{c} + \alpha\overline{h})\tau + (\alpha+1)\overline{h} \leq 0, \tag{5.13}$$

where $C_1 = \sqrt{\tilde{R}^2 - \overline{h}^2} \geq 0$. The left side of (5.13) is concave on τ. Define the derivative of the left side as $F_1'(\tau) = C_1 g'(\tau) - (\tilde{c} + \alpha\overline{h})$. We observe that $F_1'(\tau_0) > 0$ as $g'(\tau_0) \to \infty$ and

1. if $F_1'(\underline{\tau}) \leq 0$, τ^* is the unique solution of $F_1'(\tau) = 0$ within the domain $[\tau_0, \underline{\tau}]$;
2. else if $F_1'(\underline{\tau}) > 0$, $\tau^* = \underline{\tau}$.

Case 2: If $\tau \in [\underline{\tau}, \overline{\tau}]$, $h^* = \hat{h}(\tau)$. Then from (5.12), we find

$$\tilde{R}\sqrt{g(\tau)^2 + f(\tau)^2} - \tilde{c}\tau \leq 0.$$

The above problem is a one-dimensional problem on τ. We transform it into the

following form:

$$F_2(\tau) = \tilde{R}^2 \left(g(\tau)^2 + f(\tau)^2 \right) - \tilde{c}^2 \tau^2 \leq 0. \tag{5.14}$$

We observe that $F_2(\tau)$ is differentiable on $[\underline{\tau}, \overline{\tau}]$. Then, we know that the extreme value of $F_2(\tau)$ happens at the critical points (boundary points $\underline{\tau}$, $\overline{\tau}$ or τ_i such that that $F_2'(\tau_i) = 0$). In the following numerical analysis, we present efficient ways to find τ^* which maximize the left side of (5.14).

The first and second derivative of the left side of (5.14) are

$$\begin{aligned} F_2'(\tau) &= \tilde{R}^2 \left(\frac{\alpha}{\epsilon} \tau^{-\alpha-1} + 2\alpha(\alpha\tau - \alpha - 1) \right) - 2\tilde{c}^2 \tau \\ &= \frac{\alpha \tilde{R}^2}{\epsilon} \tau^{-\alpha-1} + (2\alpha^2 \tilde{R}^2 - 2\tilde{c^2})\tau - 2\tilde{R}^2 \alpha(\alpha+1), \\ F_2''(\tau) &= -\frac{\alpha \tilde{R}^2(\alpha+1)}{\epsilon} \tau^{-\alpha-2} + (2\alpha^2 \tilde{R}^2 - 2\tilde{c^2}). \end{aligned}$$

Given this, there are two conditions.

Condition 1: If $2\alpha^2 \tilde{R}^2 - 2\tilde{c^2} \leq 0$, $F_2'(\tau)$ is monotonically decreasing on τ and $F_2(\tau)$ is concave on τ. Then,

1. if $F_2'(\underline{\tau}) \leq 0$, $\tau^* = \underline{\tau}$;

2. else if $F_2'(\underline{\tau}) > 0$ and $F_2'(\overline{\tau}) \leq 0$, τ^* is the unique solution of $F_2'(\tau) = 0$ within the domain $[\underline{\tau}, \overline{\tau}]$.

3. else if $F_2'(\overline{\tau}) > 0$, $\tau^* = \overline{\tau}$.

Condition 2: If $2\alpha^2 \tilde{R}^2 - 2\tilde{c^2} > 0$, $F_2''(\tau)$ is monotonically increasing on τ and $F_2'(\tau)$ is convex on τ. Then,

1. if $F_2''(\overline{\tau}) \leq 0$, $F_2'(\tau)$ is decreasing within the domain. To find τ^*, we follow the same discussions as in Condition 1;

2. else if $F_2''(\overline{\tau}) > 0$ and $F_2''(\underline{\tau}) \leq 0$, $F_2'(\tau)$ is first decreasing and then increasing.

Define $F_s = F_2'(\tau_s)$ where $F_2''(\tau_s) = 0$ within the domain $[\underline{\tau}, \overline{\tau}]$, $F_l = F_2'(\underline{\tau})$, and $F_u = F_2'(\overline{\tau})$. Then,

(a) If $0 \leq F_s$, $\tau^* = \overline{\tau}$.

(b) If $F_s \leq 0 \leq F_l \leq F_u$ or $F_s \leq 0 \leq F_u \leq F_l$, $\tau* = \overline{\tau}$ or the unique solution of $F_2'(\tau) = 0$ within the domain $[\underline{\tau}, \tau_s]$ that maximizes $F_2(\tau)$.

(c) If $F_s \leq F_l \leq 0 \leq F_u$, $\tau^* = \underline{\tau}$ or $\overline{\tau}$ that maximizes $F_2(\tau)$.

(d) If $F_s \leq F_u \leq 0 \leq F_l$, τ^* equals the unique solution of $F'(\tau) = 0$ within the domain $[\underline{\tau}, \tau_s]$.

(e) If $F_s \leq F_l \leq F_u \leq 0$ or $F_s \leq F_u \leq F_l \leq 0$, $\tau^* = \underline{\tau}$.

3. else if $F_2''(\underline{\tau}) > 0$, $F_2(\tau)$ is convex on τ. $\tau^* = \underline{\tau}$ or $\overline{\tau}$ that maximizes $F_2(\tau)$.

Case 3: If $\tau \in [\overline{\tau}, \infty]$, $h^* = \underline{h}$. Then from (5.12), we find

$$g(\tau)\sqrt{\tilde{R}^2 - \underline{h}^2} + f(\tau)\underline{h} - \tilde{c}\tau \leq 0,$$

which we transform into the following equivalent form

$$F_3(\tau) = C_3 g(\tau) - (\tilde{c} + \alpha\underline{h})\tau + (\alpha + 1)\underline{h} \leq 0, \tag{5.15}$$

where $C_3 = \sqrt{\tilde{R}^2 - \underline{h}^2}$. Define the derivative of the left hand side of (5.15) as $F_3'(\tau) = C_3 g'(\tau) - (\tilde{c} + \alpha\underline{h})$. Then $F_3(\tau)$ is concave on τ and as $\tau \to \infty$, $F_3'(\tau) \leq 0$. Then,

1. if $\tilde{c} + \alpha\underline{h} = 0$, $F_3(\tau)$ is an increasing function and $\tau^* = \infty$;

2. else if $\tilde{c} + \alpha\underline{h} > 0$, as $\tau \to \infty$, $F_3'(\tau) < 0$. Based on the concavity of $F_3(\tau)$, we find

(a) if $F_3'(\overline{\tau}) \leq 0$, $\tau^* = \overline{\tau}$;

(b) else if $F_3'(\overline{\tau}) > 0$, τ^* equals the unique solution of $F_3'(\tau) = 0$ within the domain $[\overline{\tau}, \infty]$.

To efficiently apply the golden section search, we determine an effective finite upper bound instead of ∞. Let the effective upper bound be τ_2, we have

$$F_3'(\tau_2) = C_3 g'(\tau_2) - (\tilde{c} + \alpha \underline{h}) \leq 0.$$

Lemma V.6. *A feasible selection of τ_2 is*

$$\tau_2 = \left[\frac{-1 + \sqrt{1 + 4(1-\epsilon)C_2}}{2C_2} \right]^{-\frac{1}{\alpha}},$$

where $C_2 = \frac{\alpha^2 C_3^2}{4\epsilon(\tilde{c}+\alpha\underline{h})^2}$. The proof is given in Section 5.6.0.7.

Then, instead of a search on $[\overline{\tau}, \infty]$, we only need to search on $[\overline{\tau}, \tau_2]$.

Combining all three cases, we can find the overall worst case τ^* and h^* given x_i^*. If (5.12) is satisfied with these parameters, then there is no violated constraint in Step 2 of Algorithm 1. If (5.12) is not satisfied, we need to use the worst case τ^* and m^* in Step 3 and the iteration continues. Depending on how we define Ξ, m^* are different functions of h^*.

5.4.0.6 Candidates of Ξ

In this section, we demonstrate how the selection of Ξ affects the determination of $\underline{h}$, $\overline{h}$, and m^*. Specifically, we give two examples of Ξ and show how to exactly reformulate (5.10) (i.e., Assumption V.2) and how to calculate $\underline{h}$ and $\overline{h}$, given x_i^*. Furthermore, we show how to find the worst case m^* from h^*.

Rectangular Support: We assume that $\Xi = [\underline{k}, \overline{k}]$ and hence we can reformulate

(5.10) as

$$a(x)^\top \left(\frac{\underline{k}+\overline{k}}{2}\right) + |a(x)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right) \leq b(x). \tag{5.16}$$

Furthermore, given x_i^*, we have the following relationships due to (5.11).

$$\underline{h} = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) - |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha, \tag{5.17}$$

$$\overline{h} = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) + |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha. \tag{5.18}$$

Based on (5.17) and (5.18), if we have the worst case h^*, we find the worst case m^* by solving (5.19) for λ_r and substituting in (5.20):

$$h^* = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) + \lambda_r |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha, \tag{5.19}$$

$$m^* = \left(\frac{\underline{k}+\overline{k}}{2}\right) - \lambda_r \mathbf{sign}\left(a(x_i^*)\right) \left(\frac{\overline{k}-\underline{k}}{2}\right), \tag{5.20}$$

where $\mathbf{sign}\left(a(x)\right)$ returns a diagonal matrix whose diagonal elements equal the sign of each elements in $a(x)$.

Ellipsoidal Support: We assume that $\Xi = \{m : m = m_c + P^{1/2}u, ||u||_2 \leq 1\}$, where $P \succ 0$. Then we can reformulate (5.10) as

$$a(x)^\top m_c + \left\|P^{1/2}a(x)\right\|_2 \leq b(x). \tag{5.21}$$

Furthermore, due to (5.11), we have the following relationships:

$$\underline{h} = \left[a(x_i^*)^\top(\mu - m_c) - \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha, \tag{5.22}$$

$$\overline{h} = \left[a(x_i^*)^\top(\mu - m_c) + \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha. \tag{5.23}$$

Next, if we have the worst case h^*, we find the worst case m^* directly by solving

(5.24) for λ_e and substituting in (5.25):

$$h^* = \left[a(x_i^*)^\top(\mu - m_c) - \lambda_e \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha, \tag{5.24}$$

$$m^* = m_c + \lambda_e \frac{Pa(x_i^*)}{\|P^{1/2}a(x_i^*)\|_2}. \tag{5.25}$$

5.5 Case Studies

5.5.1 Simulation Setup

For problem formulation, we consider the same DC OPF problem as in Chapter IV and test our approach on a modified IEEE 30-bus system with network and cost parameters from [108]. We set $C_R = 10C_2$. We add the wind power plants to buses 22 and 5 and set $P_W^f = [66.8, 68.1]$ MW. We use the same wind power forecast uncertainty data (10000 scenarios) as in Section 5.2. We congest the system by increasing each load by 50% and reducing the limit of the line connecting buses 1 and 2 to 30 MW. All optimization problems are solved using CVX with the Mosek solver [32, 31].

To construct the ambiguity sets, unlike in Section 5.2, the outliers are used when estimating the statistical parameters (first moment μ, second moment Σ, and the set of the mode Ξ) and evaluating the reliability of the solution. We set $\epsilon = 5\%$, $\alpha = 1$, and assume Ξ is a rectangular set.

5.5.1.1 Additional Ambiguity Sets

We benchmark against two additional ambiguity sets from related work.

Ambiguity set 4: (moment and unimodality with fixed mode at the mean [87])

$$\mathcal{D}_\xi^4 := \left\{\mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 : \ \mathcal{M}(\xi) = \mu\right\}. \tag{5.26}$$

Ambiguity set 5: (moment and unimodality with $\alpha = 1$ and arbitrary mode [80])

$$\mathcal{D}_\xi^5 := \left\{ \mathbb{P}_\xi \in \mathcal{P}_1^n \cap \mathcal{D}_\xi^1 \right\}. \tag{5.27}$$

Set $\mathcal{D}_\xi^4$ is a special case of $\mathcal{D}_\xi^2$ with the mode at the mean, while $\mathcal{D}_\xi^5$ is a special case of $\mathcal{D}_\xi^3$ with $\alpha = 1$ and $\Xi := \{\text{all possible values of } \mathcal{M}(\xi)\}$ that is an ellipsoidal set based on μ and Σ as shown in Assumption V.1. In other words, our $\mathcal{D}_\xi^3$ is more general than $\mathcal{D}_\xi^2$, $\mathcal{D}_\xi^4$, and $\mathcal{D}_\xi^5$. The reformulations of $\mathcal{D}_\xi^4$ and $\mathcal{D}_\xi^5$ are simpler than $\mathcal{D}_\xi^3$ with a single SOC constraint

$$\mathcal{K}\sqrt{a(x)^\top(\Sigma - \mu\mu^\top)a(x)} \leq b(x) - a(x)^\top\mu, \tag{5.28}$$

where $\mathcal{K}$ can be found in [87] for $\mathcal{D}_\xi^4$ and in [80] for $\mathcal{D}_\xi^5$.

5.5.2 Results

5.5.2.1 Estimation of Ξ

We next analyze how the data size of each sample N_{data} and the number of bins within the histogram N_{bin} affect the estimate of the mode support. Figure 5.4 shows that if we change N_{bin} from 15 to 30 the histograms no longer show a unimodal distribution, as compared to Fig. 5.1. The problem is exacerbated as N_{bin} grows.

We next explore the impact of the size of the data pool. We first use the entire data pool to select 100 samples with different data sizes N_{data} (100 and 1000) and number of bins N_{bin} (15 and 30) and show scatter plots of the mode values in Fig. 5.5. As N_{data} gets larger, the mode values are more condensed and hence more accurate. When $N_{bin} = 30$ and $N_{data} = 100$ mode values appear in several disjoint regions, but this disjointness is mitigated as N_{data} increases to 1000. Based on the scatter plots, we determined the parameters $\underline{k}, \overline{k}$ of the four rectangular sets Ξ used in $\mathcal{D}_\xi^3$. The

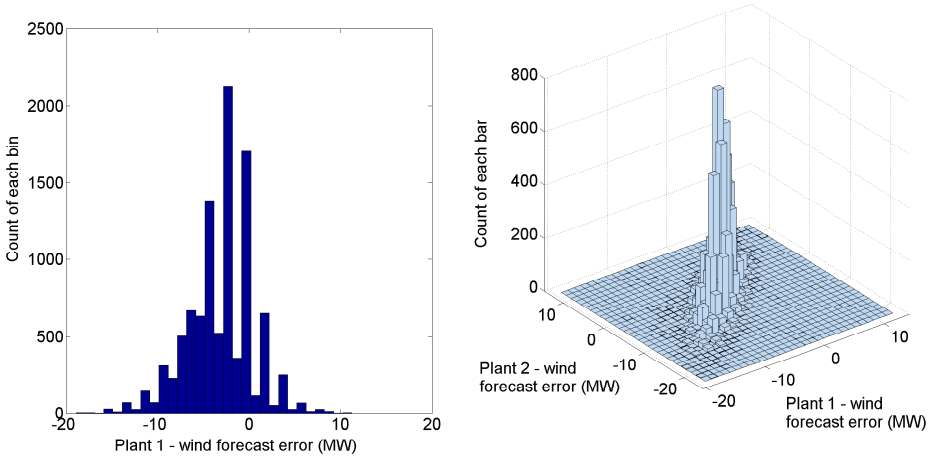


Figure 5.4: Histogram of univariate and bivariate wind forecast errors (30 bins).

Table 5.1: Full pool: $\underline{k}$ and $\overline{k}$ (MW) of Four Rectangular Sets Ξ.

N_{bin}		$N_{data} = 100$ $\underline{k}$	$\overline{k}$	$N_{data} = 1000$ $\underline{k}$	$\overline{k}$
15	Plant 1	-4.44	0.10	-3.45	0.17
	Plant 2	-4.45	0.24	-3.69	-0.11
30	Plant 1	-4.36	0.19	-3.02	-0.93
	Plant 2	-4.22	0.22	-3.06	-0.39

results are given in Table 5.1.

We repeated the analysis using only a partial data pool, specifically, we randomly selected 1000 data from the full pool to comprise the partial pool. We also use different choices of N_{data} and N_{bin}. The scatter plots are shown in Fig. 5.6 and parameter values for Ξ are given in Table 5.2.

Table 5.2: Partial pool: $\underline{k}$ and $\overline{k}$ (MW) of Four Rectangular Sets Ξ

N_{bin}		$N_{data} = 50$ $\underline{k}$	$\overline{k}$	$N_{data} = 200$ $\underline{k}$	$\overline{k}$
10	Plant 1	-4.77	0.58	-3.52	0.09
	Plant 2	-5.05	0.44	-4.43	0.06
20	Plant 1	-5.82	0.06	-4.36	-0.09
	Plant 2	-5.76	0.04	-3.86	0.19

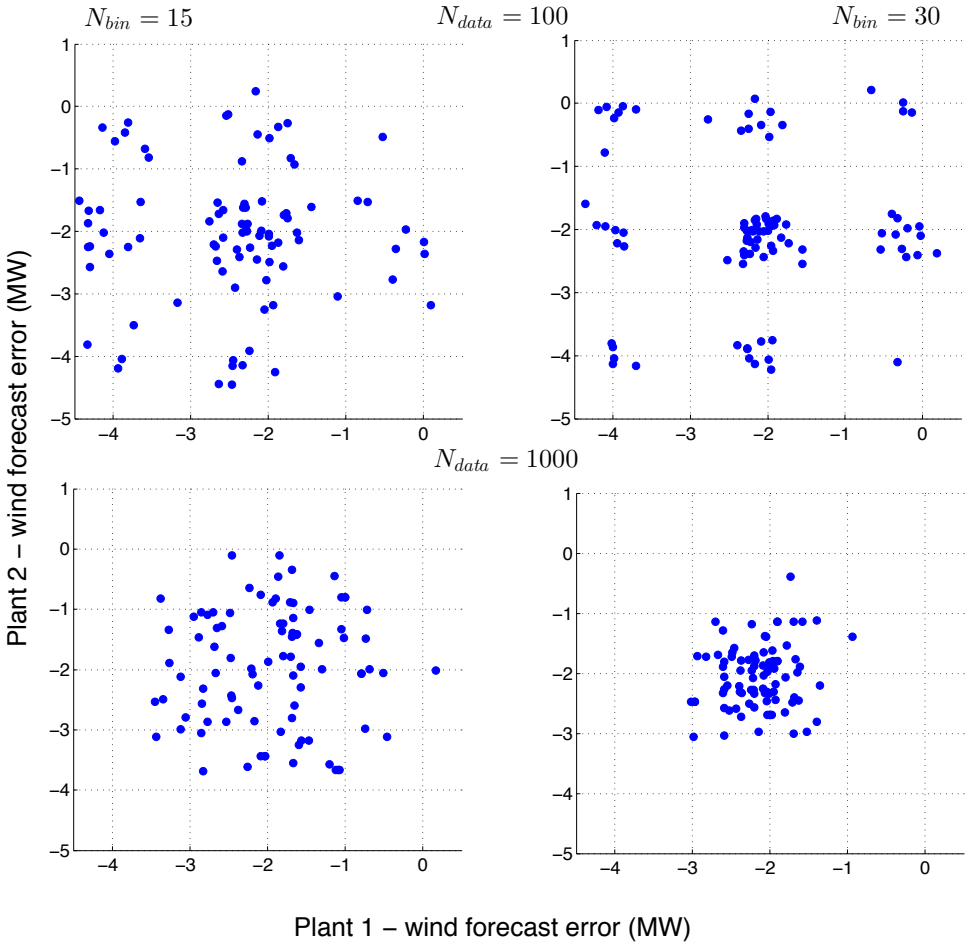


Figure 5.5: Mode values from samples with different N_{data} and N_{bin}. Data is sampled from the full data pool.

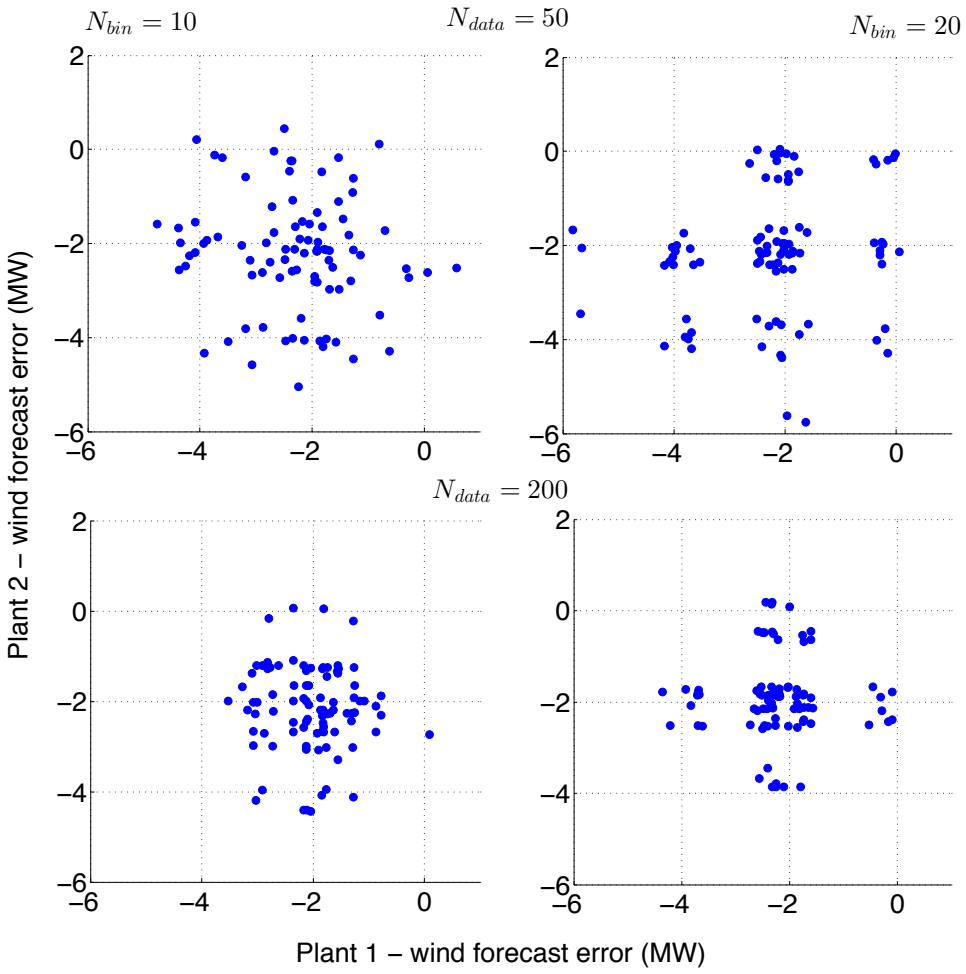


Figure 5.6: Mode values from samples with different N_{data} and N_{bin}. Data is sampled from the partial data pool.

5.5.2.2 Objective Costs

We next analyze the objective costs and the optimal reserve capacities using different ambiguity sets. The results are summarized in Table 5.3. In all case studies, since we focus on mode misspecification not moment misspecification, moments are calculated using the full or partial data pool and all ambiguity sets use the same moments.

For ambiguity set $\mathcal{D}_\xi^2$, we perform tests with the following six fixed mode estimates.

- M1: mode determined using the full (partial) data pool with histogram of 15 (10) bins. This case demonstrates the performance of $\mathcal{D}_\xi^2$ with an accurate mode estimate.
- M2: mode determined using the full (partial) data pool with histogram of 30 (20) bins. This case shows how N_{bin} affects the result.
- M3-6: combinations of the largest $\overline{k}$ and the smallest $\underline{k}$ of both plants from Table 5.1 (full pool) and Table 5.2 (partial pool). These cases demonstrate the affect of outlying data samples.

For ambiguity set $\mathcal{D}_\xi^3$, we perform tests with different Ξ, specifically, $\Xi_1 : 100 \times 15, \Xi_2 : 1000 \times 15, \Xi_3 : 100 \times 30, \Xi_4 : 1000 \times 30, \Xi_5 : 50 \times 10, \Xi_6 : 200 \times 10, \Xi_7 : 50 \times 20$, and $\Xi_8 : 200 \times 20$, where the first number refers to N_{data} and the second number refers to N_{bin}. In each case we use the parameters $\underline{k}, \overline{k}$ from Tables 5.1 and 5.2.

As shown in Table 5.3, $\mathcal{D}_\xi^1$ has the highest objective cost since it does not include the assumption of unimodality. The cost of $\mathcal{D}_\xi^2$ varies with the mode estimate. We observe opposite variations on the total up and down reserve capacities since different mode estimates lead to different estimates of the skewness of the uncertainty distribution. Comparing M1 and M2 to M3-6 we see that inaccurate estimation of the mode could lead to either higher or lower costs. Furthermore, results for M1 and M2 are significantly different demonstrating the effect of different choices of N_{bin}.

Table 5.3: Objective Costs and Reserve Capacities

Full pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4		
Total Cost	26160	19440	19546	18993	19547	19526	19542	19949	19818	19982	19896	19818	19982
Generation Cost	13032	11515	11504	11506	11481	11491	11506	11522	11522	11522	11522	11514	11522
Reserve Cost	13129	7925	8042	7488	8065	8035	8036	8427	8296	8460	8373	8304	8460
Up Reserve (MW)	38.8	26.8	26.2	26.3	25.1	26.1	26.1	27.1	26.8	27.1	27.0	26.7	27.1
Down Reserve (MW)	26.9	12.9	14.0	11.1	15.2	14.1	14.1	15.1	14.7	15.2	14.9	14.8	15.2

Partial pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8		
Total Cost	21845	16759	16740	15250	15915	16735	16718	17883	17909	17876	17909	16791	17949
Generation Cost	11713	11376	11376	11324	11295	11336	11367	11420	11420	11420	11420	11371	11420
Reserve Cost	10132	5383	5364	3926	4620	5399	5351	6463	6489	6456	6489	5420	6529
Up Reserve (MW)	31.1	20.1	20.1	17.5	16.1	19.4	19.1	22.0	22.0	21.9	22.0	19.2	22.1
Down Reserve (MW)	19.6	6.8	6.7	2.1	7.0	7.6	7.7	10.3	10.4	10.4	10.4	7.9	10.6

The costs of $\mathcal{D}_\xi^3$ are higher than those of $\mathcal{D}_\xi^2$ since the solution is designed to cope with mode misspecification. The costs do not vary significantly as a function of N_{bin} and N_{data}. For a given N_{bin}, as N_{data} increases, the costs decrease since the mode estimates are more closely clustered.

The cost of $\mathcal{D}_\xi^4$ is higher than the costs of $\mathcal{D}_\xi^2$ with M1, demonstrating the benefit in allowing the mode to be different than the mean. The cost of $\mathcal{D}_\xi^5$ is close to that of $\mathcal{D}_\xi^3$ with Ξ_3 since the mode estimates are widely distributed in this case; however, the cost of all other $\mathcal{D}_\xi^3$ is below that of $\mathcal{D}_\xi^5$. As expected, $\mathcal{D}_\xi^3$ is lower bounded by the fixed mode ambiguity sets $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$, and upper bounded by $\mathcal{D}_\xi^5$.

5.5.2.3 Reliability

Using the solutions we generated, we run out-of-sample test with 20 samples of 5000 wind forecast errors to evaluate the joint reliability of each optimal solution. We define the joint reliability as the percentage of wind forecast errors for which all chance constraints are satisfied. Then, we compare the reliability results with our pre-defined probability level ($1 - \epsilon = 95\%$). The results are summarized in Table 5.4.

We observe that reliability ranking almost always matches the cost ranking. Ambiguity sets $\mathcal{D}_\xi^1$ and $\mathcal{D}_\xi^5$ have the most conservative solutions and hence higher reliability and costs. The reliability of $\mathcal{D}_\xi^3$ is lower bounded by the reliability of $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$, and upper bounded by the reliability of $\mathcal{D}_\xi^5$. It also shows robustness against the selection of N_{data} and N_{bin}. For the full pool, all ambiguity sets achieve constraint satisfaction above 95%. For the partial pool, $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$ fail to meet the threshold, while ambiguity sets with misspecified modes $\mathcal{D}_\xi^3$, arbitrary modes $\mathcal{D}_\xi^5$, or no unimodality assumptions $\mathcal{D}_\xi^1$ achieve constraint satisfaction above 95%.

In this example, $\mathcal{D}_\xi^5$ can be use to approximate $\mathcal{D}_\xi^3$ since they have similar reliability. However, $\mathcal{D}_\xi^3$ is less conservative than $\mathcal{D}_\xi^5$ if Ξ does not include the global worst

Table 5.4: Joint Reliability (%) for $1 - \epsilon = 95\%$

Full pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4		
min	99.78	98.22	98.08	97.72	98.14	98.02	98.06	98.54	98.48	98.58	98.52	98.44	98.58
avg	99.87	98.61	98.53	98.21	98.48	98.47	98.54	98.94	98.86	98.96	98.91	98.81	98.96
max	99.94	98.84	98.84	98.42	98.74	98.84	98.86	99.14	99.10	99.14	99.12	99.04	99.14

Partial pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8		
min	99.46	92.32	92.16	82.60	88.20	91.58	92.42	95.56	95.70	95.64	95.70	92.72	95.78
avg	99.64	93.13	93.01	83.42	88.92	92.20	93.16	96.24	96.29	96.24	96.29	93.48	96.40
max	99.78	93.68	93.62	84.20	89.54	92.80	93.58	96.64	96.64	96.62	96.64	93.80	96.78

Table 5.5: Iteration Count and Computational Time for $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^3$

Full pool	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$			
	M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4
Iterations	4	4	8	8	4	4	9	8	9	6
Time (s)	16.73	16.65	40.25	39.34	17.39	16.93	33.53	31.64	33.58	19.53

Partial pool	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$			
	M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8
Iterations	4	4	6	7	4	4	9	9	9	9
Time (s)	16.78	17.05	27.30	33.35	16.79	16.91	34.08	36.72	36.21	36.08

case mode. Set $\mathcal{D}_\xi^3$ is also more applicable to multivariate unimodality as $\mathcal{D}_\xi^5$ is only defined for $\alpha = 1$.

5.5.2.4 Computational Effort

Table 5.5 shows the iteration count and computational time for $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^3$. The problems can be solved within 10 iterations and the computational time grows linearly with the number of iterations. Set $\mathcal{D}_\xi^3$ requires more iterations than $\mathcal{D}_\xi^2$. Problems using ambiguity sets $\mathcal{D}_\xi^1$, $\mathcal{D}_\xi^4$, and $\mathcal{D}_\xi^5$ can each be solved in a single run, and each takes less than one second.

5.6 Supporting Material

5.6.0.5 Convexity and Concavity of (5.12)

Here we prove the left side of (5.12) is neither jointly convex nor concave in h and τ through counter examples. We first pick $\alpha = \tilde{R} = 1$, $\epsilon = 0.05$, and $\tilde{c} = 0$ without loss of generality. Then we select two groups of points and calculate the left-side values v. Group 1: $[h, \tau, v] = (0.1, 2, 2.985)$ and $(0.3, 3, 3.05)$, then the midpoint $(0.2, 2.5, 3.15)$ has a value higher than line segment value 3.0175 (concave). Group 2: $[h, \tau, v] = (0.4, 11, 0.1990)$ and $(0.6, 10, -1.5015)$, then the midpoint $(0.5, 10.5, -0.6693)$ has a value lower than line segment value -0.65125 (convex).

5.6.0.6 Proof of Lemma V.5

We first check if $\underline{h} \geq 0$. If so, we know $\tau_1 \in \left[\tau_0, \frac{\alpha+1}{\alpha}\right]$ as $\hat{h}\left(\frac{\alpha+1}{\alpha}\right) = 0$ and $\hat{h}$ is decreasing. Hence, we can conduct the golden section search on $[\tau_0, \frac{\alpha+1}{\alpha}]$.

Next, if $\underline{h} < 0$, we know $\tau_1 > \frac{\alpha+1}{\alpha}$ and we have

$$\hat{h}(\tau_1) < h_2(\tau_1) = \frac{f(\tau_1)}{\sqrt{\frac{1-\epsilon}{\epsilon} + f(\tau_1)^2}}\tilde{R}.$$

If we further force $h_2(\tau_1) = \underline{h}$, we have $\hat{h}(\tau_1) < h_2(\tau_1) = \underline{h}$ and $\overline{\tau} \in [\frac{\alpha+1}{\alpha}, \tau_1]$. The equality $h_2(\tau_1) = \underline{h}$ will always have a solution on $\left[\frac{\alpha+1}{\alpha}, \infty\right]$ as $h_2\left(\frac{\alpha+1}{\alpha}\right) = 0$ and as $\tau \to \infty$, $h_2(\tau) = -\tilde{R}$.

Next we solve the equality and find

$$f(\tau_1) = \underline{h}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2 - \underline{h}^2)}} \Rightarrow$$
$$\tau_1 = -\left(\underline{h}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2 - \underline{h}^2)}} - (\alpha+1)\right)/\alpha.$$

5.6.0.7 Proof of Lemma V.6

We have the following relationship because $\tau \geq \tau_0 > 1$.

$$g'(\tau) = \frac{\frac{\alpha}{\epsilon}\tau^{-\alpha-1}}{2g(\tau)} \leq g_2(\tau) = \frac{\frac{\alpha}{\epsilon}\tau^{-\alpha}}{2g(\tau)}.$$

Then, we have the following relationship where τ_2 is the effective upper bound.

$$F_3'(\tau_2) \leq F_4(\tau_2) = C_3 g_2(\tau_2) - (\tilde{c} + \alpha\underline{h}) = 0.$$

The last equality will always have solution on $[\overline{\tau}, \infty]$ since $F_4(\overline{\tau}) \geq F_3'(\overline{\tau}) > 0$ and as $\tau \to \infty$, $F_4(\tau) < 0$. By solving the equality, we obtain

$$C_2(\tau_2^{-\alpha})^2 + \tau_2^{-\alpha} - (1-\epsilon) = 0,$$

where $C_2 = \frac{\alpha^2 C_3^2}{4\epsilon(\bar{c}+\alpha\underline{h})^2}$. This is a quadratic equation of $\tau_2^{-\alpha}$ and we find

$$\tau_2 = \left[\frac{-1+\sqrt{1+4(1-\epsilon)C_2}}{2C_2}\right]^{-\frac{1}{\alpha}}.$$

5.7 Conclusion

In this chapter, we proposed a distributionally robust chance constrained optimal power flow formulation considering uncertainty distributions with known moments and generalized unimodality with misspecified modes. We derived an efficient solving algorithm using the separation approach. In each iteration of the algorithm, the problem contains only SOC constraints and hence can be solved with commercial solvers. Using wind forecast errors, we found that the distribution of mode estimates are highly dependent on the data pool size, the data size of each sample, and the number of bins used in the histogram. We tested our approach on a modified IEEE 30-bus system and compared our results to those generated with other ambiguity sets. Without the assumption of unimodality, we obtain overly conservative results as unrealistic distributions are included in the ambiguity set. Considering unimodality, but with fixed mode, the results are highly dependent on the quality of the mode estimate. Considering unimodality with misspecified mode, the results are relatively consistent across different mode supports and the performance is bounded by that of the fixed-mode model and that of the arbitrary-mode model. With univariate

unimodality and large mode deviations, the misspecified-mode model can be well approximated by the arbitrary-mode model.

CHAPTER VI

Assessing the Value of Including Unimodality Information in Distributionally Robust Optimization Applied to Optimal Power Flow

To manage uncertainty, chance-constrained optimal power flow formulations and various solution methodologies have been proposed. However, conventional approaches either provide overly-conservative results or rely on accurate estimates of uncertainty distributions, which may not exist. Chapter IV considered a distributionally robust optimal power flow problem with both moment and unimodality information because most practical uncertainties follow unimodal distributions. The problem is formulated using chance constraints. Additionally, reformulations, approximations, and efficient solving techniques were provided. This chapter improves on the previous work with a new optimal parameter algorithm that searches for an optimal approximation. The algorithm improves the computational time and quality of the approximate solution. Additionally, we evaluate the performance of the aforementioned approaches in solving the chance constrained OPF problem using modified IEEE 118-bus and 300-bus systems with high penetrations of renewable generation. Results show that including unimodality information improves the solution quality but increases the computational requirements. The main content of this chapter is

summarized in the following paper.

1. B. Li, R. Jiang, and J.L. Mathieu. Assessing the value of including unimodality information in distributionally robust optimization applied to optimal power flow. *Working Paper*, 2018.

6.1 Introduction

Previous literature has sought to ensure the reliability of power system operation under uncertainties such as renewable generation forecast error by solving the chance constrained optimal power flow (CC-OPF) problem in which physical constraints are required to be satisfied at high probability levels [104, 42, 96, 9, 79, 93, 52]. Conventional approaches to solving the CC-OPF problem include randomized techniques with empirical uncertainty scenarios [17, 58], various analytical reformulations that assume known distributions [9, 79, 52, 80], and sample average approximation (SAA) [69, 2]. Randomized techniques such as the scenario-based method [58] require a large number of scenarios and often provide overly-conservative results.

Analytical reformulations incur less computational efforts because the reformulations often involve estimates of statistical information like moments. However, the solutions can be unreliable because estimates might not be accurate. SAA performs better as the number of samples increases, but it also introduces heavier computational burden as more binary variables and constraints are needed when recasting the SAA formulation as a mixed-integer program. The objective of this work is to identify new approaches that are efficient to solve and generate high-quality solutions with low objective costs and high reliability.

A distributionally robust (DR) optimization formulation has recently been proposed to mitigate the ambiguity from estimating the uncertainty distribution. This formulation incorporates chance constraints with regard to the worst-case probability distributions within a data-driven ambiguity set [29, 22, 87, 44]. This new formulation

is closely related to robust and stochastic optimization because (1) it reduces to a robust optimization model if the ambiguity set includes only the support information and (2) it reduces to a stochastic optimization model if the ambiguity set includes only a single distribution. By incorporating statistical information (e.g., mean, covariance, etc.) into the ambiguity set, DR formulations can achieve a better trade-off between objective costs and reliability than the existing approaches.

In previous studies on DR-OPF problems, tractable reformulations have been derived if the ambiguity set is moment-based [56, 88, 65, 107, 102, 55, 89], statistical density-based [33, 25, 99], or if the ambiguity set additionally includes structural information such as symmetry [80], unimodality [80, 48, 49, 88], and log-concavity [50]. Reference [107] used an ambiguity set with the first and second moments to solve a single-period OPF problem and obtained a good trade-off between objective performance and computational tractability. Similarly, [102] considered two-sided joint chance constraints for generator and transmission line limit constraints. References [88, 80] derive analytical reformulations when the ambiguity set incorporates structural properties such as symmetry and unimodality. References [33, 99, 25] construct the ambiguity set based on the discrepancy between the real distribution and the empirical distribution. In this chapter, we consider an ambiguity set that incorporates the first two moments and a generalized unimodality property ([49, 48]). In practice, most uncertainties such as wind power forecast error follow a "bell-shaped" unimodal distribution, whose density function has a single peak and decaying tails. Specifically, our ambiguity set employs the concept of α-unimodality, where parameter α adjusts the degree of unimodality and the shape of the distribution.

This chapter seeks to solve a distributionally robust chance-constrained (DRCC) OPF problem by extending the DRCC optimization methodologies with both moment and unimodality information from [49] and Chapter III. In particular, we propose an optimal parameter selection (OPS) approach to find a compact and conservative

approximation of the DRCC, in order to find high-quality solutions efficiently. The OPS problem is equivalent to finding the closest piecewise linear (PWL) outer approximation of a concave function that is irrelevant to the design variables. We further propose multiple options, both online and offline, to exploit the optimal parameters in the approximation and compare with the default parameter selection from [49] and Chapter III. Additionally, we provide the mathematical proofs for the optimality condition, the existence guarantee of these parameters, and a heuristic solving algorithm. Finally, we demonstrate the proposed DRCC-OPF model on modified 118-bus and 300-bus systems with high renewable penetration, and benchmark against the Gaussian approximation, scenario approximation, and a DR model that does not consider unimodality. We compare these approaches in objective costs, reliability, and computational tractability. Specifically, we assess the value of including unimodality information in the DR approaches. In addition, we demonstrate the performance and computational tractability of the proposed OPS approach.

The remainder of the chapter is organized as follows. In Section 6.2, we introduce some fundamental concepts and generalize the results regarding DRCC formulations in [49] and Chapter III. In Section 6.3, we introduce the OPS approach for the sandwich approximation of the DR chance constraints. In Section 6.4, we test all the proposed approaches, algorithms, and newly developed techniques and compare them with conventional methodologies. We specifically discuss the trade-offs and value of adding unimodality into a moment-based DR optimization. Section 6.6 provides the supporting mathematical proofs and statistical analysis.

6.2 Fundamentals and Ambiguity Sets

In this section, we first introduce the key fundamentals and the ambiguity sets considered in the DR reformulation. Note that the theoretical results in this section are adapted from [49] and Chapter III with minor generalizations.

We assume the constraints under the uncertainty can be transformed into

$$a(x)^\top \xi \leq b(x), \tag{6.1}$$

where $x \in \mathbb{R}^n$ represents the vector of decision variables and $a(x) : \mathbb{R}^n \to \mathbb{R}^l$ and $b(x) : \mathbb{R}^n \to \mathbb{R}$ represent two affine functions of x. Uncertainty $\xi \in \mathbb{R}^l$ represents a random vector defined on probability space $(\mathbb{R}^l, \mathcal{B}^l, \mathbb{P}_\xi)$ with Borel σ-algebra $\mathcal{B}^l$ and probability distribution $\mathbb{P}_\xi$. To manage uncertain violations in (6.1), we use the following chance constraint with probability threshold $1 - \epsilon$:

$$\mathbb{P}_\xi \left(a(x)^\top \xi \leq b(x) \right) \geq 1 - \epsilon. \tag{6.2}$$

6.2.0.8 Ambiguity Sets

Here, we introduce the two types of ambiguity sets considered in the DR reformulation. These sets are defined as combinations of moment and unimodality information (i.e., α unimodality introduced in Chapter III with parameter α).

Next, we define our ambiguity sets:

Moment information only:

$$\mathcal{D}_\xi := \left\{ \mathbb{P}_\xi \in \mathcal{P}^l : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma \right\}. \tag{6.3}$$

Moment and unimodality information:

$$\mathcal{U}_\xi := \left\{ \mathbb{P}_\xi \in \mathcal{P}_\alpha^l \cap \mathcal{D}_\xi :\ \mathcal{M}(\xi) = m \right\}, \tag{6.4}$$

where $\mathcal{P}_\alpha^l$ and $\mathcal{P}^l$ denote all probability distributions on $\mathbb{R}^l$ with and without the requirement of α-unimodality respectively; μ and Σ denote the first and second moments of ξ; and $\mathcal{M}(\xi) = m$ means that the true mode value of ξ is m.

Next, before we present the DR reformulations, we discuss the minor generalizations. In [49] and Chapter III, the results are derived assuming the mode is at the origin. Without loss of generality, we can rewrite (6.1) as $a(x)^\top(\xi - m) \leq b(x) - a(x)^\top m$ with $\xi - m$ as our new random vector whose mode is at the origin and apply the results accordingly. Meanwhile, we also require Assumptions V.1 and V.2 as in Chapter V with $\mathcal{D}_\xi^2$. Furthermore, we assume $\epsilon < 0.5$ and $\alpha \geq 1$, since in practice the uncertainties will at least be univariate-unimodal.

6.2.0.9 DR reformulations with $\mathcal{D}_\xi$

In this section, we consider the DR chance constraint with $\mathcal{D}_\xi$:

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi \left(a(x)^\top \xi \leq b(x) \right) \geq 1 - \epsilon, \tag{6.5}$$

and have the exact reformulation in the following SOC constraint base on Theorem V.3.

$$\sqrt{\left(\frac{1-\epsilon}{\epsilon} \right) a(x)^\top (\Sigma - \mu\mu^\top) a(x)} \leq b(x) - a(x)^\top \mu. \tag{6.6}$$

6.2.0.10 DR Chance Constraint with $\mathcal{U}_\xi$

In this section, we consider the DR chance constraint with $\mathcal{U}_\xi$:

$$\inf_{\mathbb{P}_\xi \in \mathcal{U}_\xi} \mathbb{P}_\xi \left(a(x)^\top \xi \leq b(x) \right) \geq 1 - \epsilon, \tag{6.7}$$

6.2.0.11 Exact Reformulations

From Theorem V.4, DR chance constraint (6.7) can be exactly reformulated as

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda a(x)\| \leq \tau\left(b(x)-a(x)^\top m\right) - \left(\frac{\alpha+1}{\alpha}\right)(\mu-m)^\top a(x), \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}, \tag{6.8}$$

where $\Lambda := \left(\left(\frac{\alpha+2}{\alpha}\right)(\Sigma-\mu\mu^\top) - \frac{1}{\alpha^2}(\mu-m)(\mu-m)^\top\right)^{1/2}$.

Since parameter τ has an infinite number of choices, the reformulation in Theorem V.4 also involves an infinite number of SOC constraints. To solve an optimization problem with (6.8), we give the following iterative algorithm, i.e., Algorithm 2 below, based on the separation approach in [49] and Chapter III. Note that the reformulated optimization problem in Step 1 can be solved directly.

Algorithm 2: Iterative solving algorithm

Initialization: $i=1$, $\tau_0 = \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$;

Iteration i:

Step 1: Solve the reformulated optimization problem with (6.8) using τ_j for all $j=0,\ldots,i-1$ and obtain optimal solution x_i^*. All τ_j values are collected from previous iterations;

Step 2 (**Separation [49]**): Find worst case τ^* that result in the largest violation of (6.8) under x_i^*: **IF** τ^* does not exist, **STOP** and **RETURN** x_i^* as optimal solution; **ELSE GOTO** Step 3;

Step 3: Set $\tau_i = \tau^*$ and $i = i+1$, **GOTO** Step 1;

To efficiently perform Step 2 in Algorithm 2, we follow the proposition below.

Proposition VI.1. *Define* $\mu_0^* = \left(\frac{\alpha+1}{\alpha}\right)(\mu-m)^\top a(x_i^*)$ *and*

$$\Sigma_0^* = \left(\frac{\alpha+2}{\alpha}\right) a(x_i^*)^\top(\Sigma + mm^\top - m\mu^\top - \mu m^\top)a(x_i^*).$$

Then we have the following:

1. *If* $a(x_i^*)=0$, *then constraints* (6.8) *are always satisfied;*

2. *If* $a(x_i^*) \neq 0$ *and* $b(x_i^*) - a(x_i^*)^\top m = 0$*, then* x_i^* *violates constraints* (6.8) *if and only if it violates them at* $\tau^* = \infty$*.*

3. *If* $a(x_i^*) \neq 0$ *and* $b(x_i^*) - a(x_i^*)^\top m > 0$*, then* x_i^* *violates constraints* (6.8) *if and only if it violates them at* $\tau^* = \hat{\tau}$*, where* $\hat{\tau}$ *represents the minimizer of the one-dimensional problem*

$$\min_{\tau \geq \tau_0} \left((b(x_i^*) - a(x_i^*)^\top m)\tau - \mu_0^*\right)^2 - \left(\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}\right)(\Sigma_0^* - \mu_0^{*2}), \tag{6.9}$$

whose objective function is strongly convex. The minimizer can be efficiently found through golden section search in the interval $[\tau_0, \tau_u]$ *where*

$$\tau_u = \frac{\mu^*}{b(x_i^*) - a(x_i^*)^\top m} + \frac{\alpha(1-\epsilon)^{\frac{\alpha+1}{\alpha}}(\Sigma_0^* - \mu_0^{*2})}{2\epsilon(b(x_i^*) - a(x_i^*)^\top m)^2}. \tag{6.10}$$

The result is adapted from Proposition 3 in [49] and Proposition III.9 in Chapter III by assuming $m \neq 0$*.*

In other words, if x_i^* can violate (6.8), the largest violation happens at τ^* given in Proposition VI.1.

6.2.0.12 Asymptotic Sandwich Approximations

We notice that solving the exact reformulation can be cumbersome since we might have to deal with many separation problems and iterations. Hence, it is reasonable to have the sandwich approximations to bound the true objective value from both below and above. The approximations are asymptotic since they will converge to the true objective costs with more parameters considered. The results are as follows.

Proposition VI.2. *Relaxed Approximation For given integer $K \geq 1$, and real number $\tau_0 \leq n_1 < n_2 \ldots < n_K \leq \infty$, (6.7) implies the SOC constraints*

$$\sqrt{\frac{1-\epsilon-{n_k}^{-\alpha}}{\epsilon}}\|\Lambda a(x)\| \leq n_k \left(b(x) - a(x)^\top m\right) - \left(\frac{\alpha+1}{\alpha}\right)(\mu - m)^\top a(x), \quad \forall k = 1, \ldots, K. \tag{6.11}$$

The result is adapted from Proposition 4 in [49] and Proposition III.10 in Chapter III by assuming $m \neq 0$.

Proposition VI.3. *Conservative Approximation For given integer $K \geq 2$, and real number $\tau_0 = n_1 < n_2 \ldots < n_K = \infty$, we define a piece-wise linear function containing $(K-1)$ pieces:*

$$g(\tau) = \min_{k=2,\ldots,K} \left\{ \sqrt{\frac{1}{\epsilon(1-\epsilon-n_k^{-\alpha})}} \left[\left(\frac{\alpha n_k^{-\alpha-1}}{2}\right)\tau + 1 - \epsilon - \left(1+\frac{\alpha}{2}\right) n_k^{-\alpha} \right] \right\}. \tag{6.12}$$

Denote $q_1 = \tau_0$ and $q_2 < \ldots < q_{K-1}$ represent the $(K-2)$ breakpoints of function $g(\tau)$. Then, (6.7) is implied by the SOC constraints

$$g(q_k)\|\Lambda a(x)\| \leq q_k \left(b(x) - a(x)^\top m\right) - \left(\frac{\alpha+1}{\alpha}\right)(\mu - m)^\top a(x), \quad \forall k = 1, \ldots, K-1. \tag{6.13}$$

The result is adapted from Proposition 4 in [49] and Proposition III.11 in Chapter III by assuming $m \neq 0$.

The convergence of the sandwich approximation is directly affected by the selection of n_k for all $k = 1, \ldots, K$. In [49], we proposed to use a *default* online parameter selection scheme by using the worst case τ^* values on the affected constraints deter-

mined by the separation problem in Algorithm 2. However, these values are only critical to the relaxed approximation (see Theorem V.4) but they do not have any direct connections to the conservative approximation. In Section 6.3, we will propose a new offline OPS approach to help improve the conservative approximation.

Remark: If the decision makers are more interested in the violation magnitude rather than violation probability, we can use other risk measures such as conditional value at risk (CVaR) which evaluates the conditional expectation of $a(x)^\top \xi - b(x)$ on the right tail of its distribution. The details of solving DR CVaR constrained problem under $\mathcal{D}_\xi$ or $\mathcal{U}_\xi$ are discussed in [110, 49] and Chapter III respectively.

6.3 Optimal Parameter Selection

In this section, we propose an OPS approach to develop the conservative approximations of the DR chance constraint. Based on [49] and Chapter III, q_k for all k in Proposition VI.3 define the break points of a concave PWL function $g(\tau)$ that outer approximates the nonlinear function

$$v(\tau) = \sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}} \text{ where } \tau \in [\tau_0, \infty). \tag{6.14}$$

Hence, equivalently we can define the OPS problem as to find the optimal PWL outer approximation of $v(\tau)$. Conventional approaches on finding optimal PWL approximation have been thoroughly discussed in [21, 41, 92]. However, they are not applicable to our problem as they do not consider outer approximations and assume the function has a bounded domain. Based on their work, we make the following specific extensions for our OPS problem:

1. We prove the optimality condition and existence guarantee for the optimal PWL outer approximation of $v(\tau)$.

2. We develop a heuristic searching algorithm to find the optimal PWL outer approximation.

6.3.1 Optimality and Existence

First we define what is an optimal PWL approximation in our problem. Denote an $|\mathcal{S}|$-piece PWL outer approximation of $v(\tau)$ as $h(\tau) = \min_{s\in\mathcal{S}}\{d_s\tau + f_s\}$ where $\mathcal{S}$ represents the set of indices representing the pieces and d_s is non-increasing with increasing s. We further define $h_s(\tau)$ representing the s-th piece in $h(\tau)$. Suppose the domain of each piece in $h(\tau)$ as $\mathcal{H}_s$, then the error $e^{\max}$ of the PWL outer approximation can be defined as the largest distance between these two functions

$$e^{\max} = \max_{s\in\mathcal{S}} \max_{\tau\in\mathcal{H}_s}\{d_s\tau + f_s - v(\tau)\}. \tag{6.15}$$

Hence, the optimal PWL outer approximation describes the $|\mathcal{S}|$-piece PWL approximation that minimizes $e^{\max}$.

Next, we discuss the optimality conditions such that: if there exists a PWL solution satisfying these conditions then this solution is optimal. When $|\mathcal{S}| \geq 1$, we have the following theorem.

Theorem VI.4. *(Optimality) An $|\mathcal{S}|$-piece PWL function $h(\tau) = \min_{s\in\mathcal{S}}\{d_s\tau + f_s\}$ is an optimal PWL outer approximation of $v(\tau)$ defined in* (6.14) *if the following three conditions hold.*

1. $h_{|S|}(\tau) = \sqrt{\frac{1-\epsilon}{\epsilon}}$.

2. $h_s(\tau)$ *is tangent to* $v(\tau)$ *for all* $s \in \mathcal{S}$.

3. *Denote all* $|\mathcal{S}|$ *break points of* $h(\tau)$ *(including* τ_0*) as* $\{B_s,\ s \in \mathcal{S}\}$. *Then*

$$h(B_{s_1}) - v(B_{s_1}) = h(B_{s_2}) - v(B_{s_2}),\ \forall s_1, s_2 \in \mathcal{S}.$$

The proof is given in Appendix 6.6.0.5. Next, for given $v(\tau)$, we show there always exists an $h(\tau)$ that satisfies all three conditions in Theorem VI.4.

Theorem VI.5. *(Existence) There always exists an $|\mathcal{S}|$-piece PWL function $h(\tau) = \min_{s\in\mathcal{S}}\{d_s\tau + f_s\}$ that satisfies all three conditions in Theorem VI.4.*

The proof is given in Appendix 6.6.0.6.

6.3.2 Searching Algorithm

Here we provide a heuristic searching algorithm Algorithm 3 to solve for the optimal $|\mathcal{S}|$-piece PWL approximation of $v(\tau)$ on $[\tau_0, \infty)$ in (6.14). The algorithm is adapted from the recursive descent algorithm in [41]. Before we give the algorithm, we define the following notation:

- $\mathcal{I}$: max number of iterations;
- δ: percentage tolerance as termination criteria;
- $h^i(\tau)$: For iteration i, the current first $|\mathcal{S}|-1$ pieces of the $|\mathcal{S}|$-piece PWL outer approximation of $v(\tau)$. We exclude the last zero-slope piece since it is trivial. $h^i_s(\tau)$ represents the s-th linear function in the PWL approximation.
- $B^i \in \mathbb{R}^{|\mathcal{S}|}$: For iteration i, all the break points and end points in $h^i(\tau)$. B^i_s represents the s-th entry and we have $B^i_1 = \tau_0$;
- $T^i \in \mathbb{R}^{|\mathcal{S}|-1}$: For iteration i, all the tangent points between $h(\tau)$ and $v(\tau)$. T^i_s represents the s-th entry;
- $E^i \in \mathbb{R}^{|\mathcal{S}|}$: For iteration i, distance between $h(\tau)$ and $v(\tau)$ at all B^i's. Define the error between last zero-slope piece and $v(\tau)$ at $B^i_{|\mathcal{S}|}$ as $e^i_T = \sqrt{(1-\epsilon)/\epsilon} - v(B^i_{|\mathcal{S}|})$.
- $\gamma^i \in \mathbb{R}^{|\mathcal{S}|-1}$: For iteration i, the modification on T^i. γ^i_s represents the s-th entry;

- $\Delta \in \mathbb{R}^{\mathcal{I}}$: the step size for all iterations. Δ^i represents the step size for i-th iteration.

Algorithm 3: Heuristic searching algorithm

Initialization: $i = 1$, $\Delta = 1$, $\delta = 0.01$, $\mathcal{I} = 50$, $B^1_{|\mathcal{S}|} = 10$, calculate T^1 by evenly divide $[\tau_0, B^1_{|\mathcal{S}|}]$ into $|\mathcal{S}|$ pieces;

Iteration i:

Step 1: **IF** $i \leq \mathcal{I}$, calculate the internal B^i for $i = 2, ..., |\mathcal{S}| - 1$ based on end points τ_0 and $B^i_{|\mathcal{S}|}$ as well as all T^i;

ELSE STOP and **RETURN** no convergence under current initialization.

Step 2: Calculate E^i and e^i_T:

- **IF** $(1+\delta)e^i_T < E^i_{|\mathcal{S}|}$ or $(1+\delta)E^i_{|\mathcal{S}|} < e^i_T$, set $B^i_{|\mathcal{S}|} = 0.5(\tau^* + B^i_{|\mathcal{S}|})$ where $h_{|\mathcal{S}|-1}(\tau^*) = \sqrt{(1-\epsilon)/\epsilon}$. **GOTO** Step 1;
- **ELSE GOTO** Step 3;

Step 3: **IF** $\max(E^i) \leq (1+\delta)\min(E^i)$, **STOP** and **RETURN** $h^i(\tau)$ combined with the last zero-slope piece as the optimal solution; **ELSE GOTO** Step 4;

Step 4: **IF** $i = 1$, **GOTO** Step 5; **ELSEIF** $\max(E^i) > \max(E^{i-1})$, $i = i - 1$, $\Delta^i = \Delta^i/2$, and **GOTO** Step 5; **ELSE GOTO** Step 5;

Step 5: For all $s = 1, ..., |\mathcal{S}| - 1$,

$$\gamma^i_s = \frac{\Delta^i(E^i_{s+1} - E^i_s)}{\frac{E^i_{s+1}}{B^i_{s+1} - T^i_s} + \frac{E^i_s}{T^i_s - B^i_s}}.$$

Then, set $T^{i+1} = T^{i+1} + \gamma^i$ and $i = i + 1$. **GOTO** Step 1;

In the initialization step, δ, Δ, $B^1_{|\mathcal{S}|}$, and $\mathcal{I}$ can use other reasonable values. When $|\mathcal{S}| = 2$, an illustrative example is shown in Fig. 6.1. In each iteration i, Step 1 first estimate T^i with in $[\tau_0, B^i_{|\mathcal{S}|}]$. In Step 2, we first coarsely adjust the current last break point $B^i_{|\mathcal{S}|}$ by comparing $E^i_{|\mathcal{S}|}$ and e^i_T. Generally, if $E^i_{|\mathcal{S}|}$ is larger, we should reduce $B^i_{|\mathcal{S}|}$ and vice versa. In Step 4, we further adjust $T^i i$ with γ^i, that is based on the same derivations as in [41]. If the algorithm successfully terminates from Step 3 at iteration i^*, based on Theorem VI.4, we have $E^{i^*}_s = e^{i^*}_T$ for all $s \in \mathcal{S}$. The optimal break points B^{i^*} can be used in the Proposition VI.3 to establish the corresponding conservative approximation.

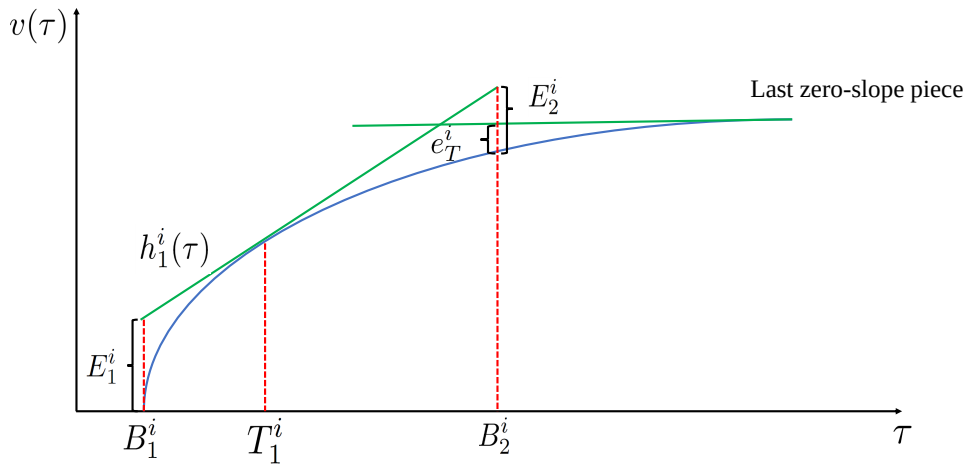


Figure 6.1: Illustrative Example of Iteration i in Algorithm 3 when $|\mathcal{S}| = 2$.

Remark: We observe that the resulting optimal parameter from Algorithm 3 is uniquely determined once the required dimension $|\mathcal{S}|$ is determined. Specifically, it is independent of the decision variable values but only dependent on the system parameter. Hence, these optimal parameters can be efficiently determined offline.

6.4 Case Studies

6.4.1 Simulation Setup

We consider a similar DC OPF problem as in [49] and Chapter III with more uncertain wind power plants. We assume that the system has N_W wind power plants with wind power forecast error $\tilde{w} \in \mathbb{R}^{N_W}$ (each element is represented as $\tilde{w}_i$), N_G generators, and N_B buses. The wind power forecast errors are calculated as the difference between actual realizations and their corresponding forecasts and will be compensated by the generator reserves. Design variables include generation output $P_G \in \mathbb{R}^{N_G}$, up and down reserve capacities $R_G^{up} \in \mathbb{R}^{N_G}$, $R_G^{dn} \in \mathbb{R}^{N_G}$, and a distribution vector $d_G \in \mathbb{R}^{N_G}$, which determines how much reserve does each generator provide to balance the overall wind forecast forecast error. The full problem formulation is

given below.

$$\min\ P_G^T[C_1]P_G + C_2^T P_G + C_R^T(R_G^{up} + R_G^{dn})$$

$$\text{s.t.} - P_l \le AP_{\text{inj}} \le P_l, \tag{6.16a}$$

$$R_G = -d_G(\Sigma_{i=1}^{N_W}\tilde{w}_i), \tag{6.16b}$$

$$P_{\text{inj}} = C_G(P_G + R_G) + C_W(P_W^f + \tilde{w}) - C_L P_L, \tag{6.16c}$$

$$\underline{P}_G \le P_G + R_G \le \overline{P}_G, \tag{6.16d}$$

$$-R_G^{dn} \le R_G \le R_G^{up}, \tag{6.16e}$$

$$\mathbf{1}_{1\times N_G} d_G = 1, \tag{6.16f}$$

$$\mathbf{1}_{1\times N_B}(C_G P_G + C_W P_W^f - C_L P_L) = 0, \tag{6.16g}$$

$$P_G \ge \mathbf{0}_{N_G\times 1},\ d_G \ge \mathbf{0}_{N_G\times 1}, \tag{6.16h}$$

$$R_G^{up} \ge \mathbf{0}_{N_G\times 1},\ R_G^{dn} \ge \mathbf{0}_{N_G\times 1}, \tag{6.16i}$$

where $[C_1] \in \mathbb{R}^{N_G\times N_G}$, $C_2 \in \mathbb{R}^{N_G}$, and $C_R \in \mathbb{R}^{N_G}$ are cost parameters. Constraint (6.16a) bounds the power flow by the line limits P_l. The power flow is calculated from the power injections P_{inj} in (6.16c) and the parameter matrix A based on admittance and network connection. Constraint (6.16b) computes the real-time reserve value R_G that is bounded by the reserve capacities R_G^{dn} and R_G^{up} in (6.16e). In (6.16c), P_W^f is the wind power forecast, P_L is the load, and C_G, C_W, and C_L are matrices that map generators, wind power plants, and loads to buses; (6.16d) bounds generation within its limits $[\underline{P}_G, \overline{P}_G]$; (6.16f), (6.16g) enforce power balance with and without wind power forecast error; and (6.16h), (6.16i) ensure all decision variables are non-negative.

We test our approaches on the modified IEEE 118-bus and 300-bus systems with network and cost parameters from [20]. We set $C_R = 10C_2$ and add wind power plants to all generation buses to achieve high uncertainty dimension. Meanwhile, to realize

high wind penetration, we add in total 400 and 2000 MW forecast generation for the 118-bus and 300-bus systems respectively. Specifically, we scale the wind power forecast so that the forecast generation on each generation bus is proportional to its generation limit.

For uncertainty data, we define the forest error ratio that is calculated as the ratio between the wind power forecast error and the corresponding forecast. Then, we consider the following two data sets with distinct properties:

Data Set 1 (DS1) We use the same wind power scenarios as in [93]. The data set is generated using the Markov-Chain Monte Carlo mechanism [70] on real wind power forecasts and realizations from Germany. The wind power is well-forecasted with small forecast error ratios ($-30\% \sim 60\%$). For each wind bus, we randomly select the scenarios from the same wind data pool without considering spatial correlation.

Data Set 2 (DS2) The RE-Europe data set [43] contains hourly wind power forecasts and realizations based on the European energy system. The data set includes strong spatio-temporal correlation. However, the data set also contains poor forecasts with extreme forecast error ratios, up to 5300%. Hence, we apply an additional filter to exclude the outliers and scale down the forecast error while maintaining its distribution and correlation. The details are given in Section 6.6.0.7.

We solve all the optimization problems using CVX with the Mosek solver [32, 31]. We set $\epsilon = 5\%$ and $\alpha = 1$. The latter is valid because, in general, wind power forecast error is marginally unimodal (details in Section 6.6.0.8). We use 5000 and 8000 randomly generated data points for the 118-bus and 300-bus systems to construct $\mathcal{D}_\xi$ and $\mathcal{U}_\xi$. More data is needed for the 300-bus system since the uncertainty dimension is larger. In addition, we use histograms with 15 and 20 bins to determine the locations of mode m for DS1 and DS2 by identifying the bin with the most points. Further, we randomly select 5000 and 8000 data points to conduct out-of-sample tests to evaluate

reliability for the 118-bus and 300-bus systems respectively. We define the reliability as the percentage of wind power forecast errors for which all chance constraints are satisfied. To guarantee the credibility of the result, we perform 3 parallel tests by randomly reselecting the data used to construct the ambiguity sets.

As benchmarks, we consider the following conventional approaches:

- Scenario-based method [58]: The method enforces the constraints affected by uncertainties to be robust against a probabilistically robust set. This set is constructed using a sufficient number of randomly selected uncertainty realizations.
- Analytical reformulation with Gaussian assumption [9, 52, 79]: Here, we assume the uncertainty follows a multivariate normal distribution with moments determined by the data. Then all the chance constraints can be exactly reformulated as SOC constraints.

6.4.2 Results

6.4.2.1 Convergence of Algorithm 3

Here, we demonstrate the convergence performance of Algorithm 3 under different $|\mathcal{S}|$ values. In Fig. 6.2, we observe that when $|\mathcal{S}|$ increases, the optimal approximation error $e^{\max}$ will decrease and the total number of iterations grows almost linearly. In addition, the algorithm also demonstrates a fast convergence rate at different $|\mathcal{S}|$ values.

6.4.2.2 Costs, Computational time, and Reliability

Here, we compare the scenario-based method (SC), analytical reformulation under Gaussian assumption (GA), DR approach with only moment information (DR-M), and DR approach with both moment and unimodality information (DR-U) in terms of objective costs, reliability, and the overall computational requirement. The results

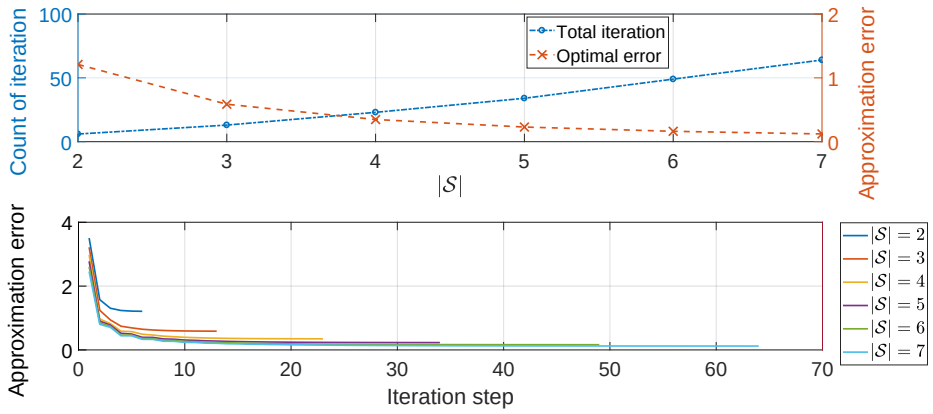


Figure 6.2: Total iteration and optimal approximation error at different $|\mathcal{S}|$ (top); convergence of the approximation error through iterations at different $|\mathcal{S}|$ (bottom).

are summarized in Tables 6.1 and 6.2. To quantify the comparisons, we define a percentage difference (Diff) on objective cost and reliability against our benchmarks SC and GA as they normally bound the other approaches. Specifically, we define Diff of GA to be 0 and Diff of SC to be 100%. Diffs of other approaches are calculated as (X-GA)/(SC-GA) where X can refer to DR-M or DR-U. Next, we define the improvement (Improv) of a certain approach to be the ratio of its Diff on reliability to its Diff on cost. Both SC and GA have Improvs of 1 (i.e., we assume $0/0 = 1$ for GA). If the Improv of a certain approach is large, we conclude that this approach has better trade-off between cost and reliability than the SC and GA benchmarks. Specifically, it means the approach achieves high reliability with low relative cost.

From Tables 6.1 and 6.2, we see that, at any system dimensions or data sets, SC provides overly conservative results with the highest objective costs and 100% reliability far from our pre-defined probability level of 95%. GA provides the least conservative results with the lowest objective costs and the lowest reliability that always below the 95% requirement. For DR-U and DR-M, their costs and reliability are between SC and GA and satisfy the 95% requirement. Specifically, DR-U provides higher objective costs and higher reliability since it only considers moment information in the ambiguity set. If we compare the Diffs and Improvs of DR-U

Table 6.1: Objective Costs, Reliability (%), and Computational Time (Second)

Bus/Data Set		GA			DR-M					
		Cost	Reliability	Time	Cost	Cost Diff	Reliability	Reliability Diff	Improv	Time
118/DS1	min	3309	81.7	11.0	3466	9.6	99.7	98.3	10.1	11.1
	avg	3310	81.8	11.4	3467	9.6	99.7	98.3	10.2	11.3
	max	3310	81.9	11.8	3468	9.7	99.7	98.4	10.2	11.4
118/DS2	min	3491	79.5	11.0	4064	23.5	98.9	94.3	3.3	11.1
	avg	3520	81.8	11.7	4141	25.9	99.2	95.8	3.7	11.6
	max	3564	85.4	12.4	4261	29.3	99.7	97.9	4.1	12.3
300/DS1	min	14408	72.9	125.4	14579	4.7	99.6	98.5	20.9	124.5
	avg	14409	73.6	127.0	14580	4.7	99.6	98.6	21.0	125.3
	max	14410	74.1	130.2	14581	4.7	99.7	98.9	21.0	125.9
300/DS2	min	15125	85.7	234.8	16956	29.0	99.7	97.7	3.1	236.5
	avg	15161	86.2	236.1	17052	30.5	99.7	97.8	3.2	237.7
	max	15191	87.1	237.1	17128	32.0	99.7	97.9	3.4	239.4

Table 6.2: Objective Costs, Reliability (%), and Computational Time (Second)

Bus/Data Set		SC			DR-U					
		Cost	Reliability	Time	Cost	Cost Diff	Reliability	Reliability Diff	Improv	Time
118/DS1	min	4935	100.0	11.2	3340	1.9	97.0	83.6	40.4	475.1
	avg	4937	100.0	14.4	3343	2.0	97.1	84.2	41.6	478.4
	max	4942	100.0	18.7	3344	2.1	97.2	84.5	44.0	483.4
118/DS2	min	5902	100.0	10.5	3703	8.7	95.0	74.7	8.1	2490.2
	avg	5926	100.0	11.6	3736	9.0	95.6	75.7	8.4	3160.7
	max	5942	100.0	13.3	3780	9.2	96.6	76.7	8.7	3815.9
300/DS1	min	18032	100.0	134.0	14479	1.9	96.4	86.7	44.4	1948.5
	avg	18038	100.0	136.9	14479	1.9	96.6	87.0	44.9	2158.5
	max	18046	100.0	140.2	14480	2.0	96.7	87.4	45.4	2263.6
300/DS2	min	21249	100.0	141.7	15724	8.9	96.7	76.9	6.8	6547.7
	avg	21373	100.0	143.1	15777	9.9	97.1	78.8	8.0	10880.4
	max	21436	100.0	144.9	15880	11.4	97.4	81.8	8.7	15720.6

Table 6.3: Percentage Time of Step 2 (%) and Iteration Number

Bus/Data Set	118/DS1			118/DS2			300/DS1			300/DS2		
	min	avg	max	min	avg	max	min	avg	max	min	avg	max
Percentage	86.2	86.8	87.4	85.4	85.6	85.9	59.1	59.1	59.2	43.1	43.6	43.9
Iteration	5	5	5	26	32	36	6	7	7	14	23	33

and DR-M, we can conclude that DR-U provides the best trade-off in terms of cost and reliability and hence has the best solution quality. In terms of the computation time, GA, SC, and DR-M finish in a single run. However, DR-U requires an iterative solving algorithm and hence takes the longest computational time. As the system dimension increases, we observe that the resulting computational burdens become severe but similar relationship between approaches still hold. Comparing DS1 and DS2, solutions from DS1 are more stable with less variability across the parallel tests. Meanwhile, cost Diffs and solving time from DS1 are much smaller than the ones from DS2, but reliability Diffs and Improvs of DS1 are much larger.

6.4.2.3 Convergence of Algorithm 2

Here, we analyze the computational performance of Algorithm 2 as well as the approximate solutions of its intermediate iterations. Table 6.3 summarizes the percent time it takes to complete Step 2 in Algorithm 2 (i.e., the separation problem) and the total number of iterations. The rest of the percent time is used to complete Step 1, i.e., to solve for optimal solutions. We observe that generally DS2 requires a larger number of iterations than DS1 and the 118-bus system requires more computational effort to complete Step 2 than the 300-bus system.

We further use the 4 test cases (i.e., different combinations of systems and data sets) to analyze the time decomposition in each iteration as shown in Fig. 6.3. We see that, in all the cases, the time used for solving Step 2 almost remains constant. With larger number of iterations, the overall solving time slightly increases as more constraints are added in Step 1 of Algorithm 2. Additionally, cases with the 300-bus

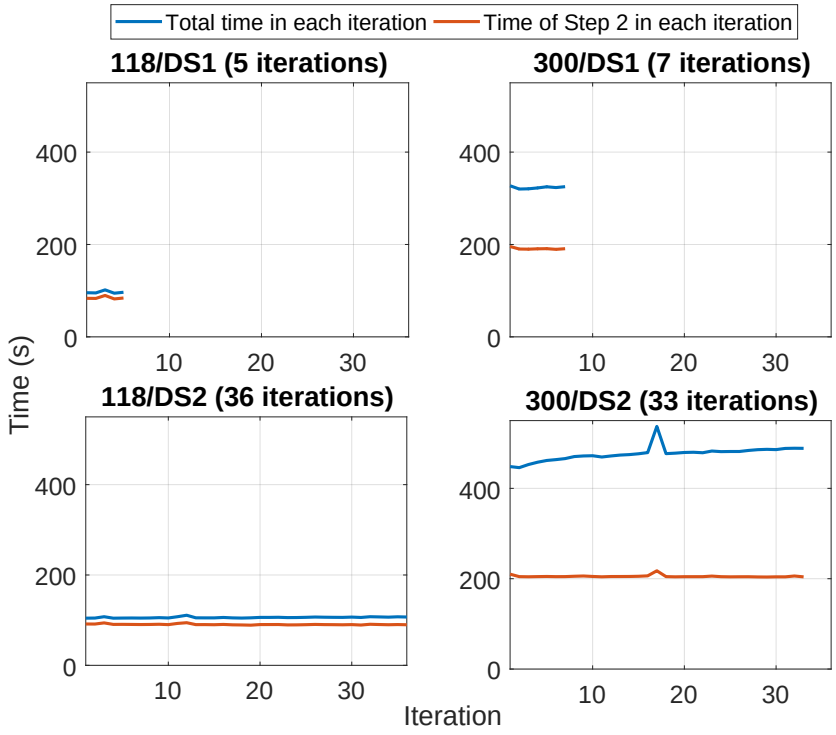


Figure 6.3: Total Solving Time and Time of Step 2 in Each Iteration.

system or DS2 require longer solving time in each iteration.

We have shown that the exact solution of DR-M normally requires long solving time. Next, we check if the solutions from the intermediate iterations can be good approximates of the final optimal solution. As an example, we use 36-iteration test case in 118/DS2 and calculate its reliability and optimality gap as shown in Fig. 6.4. We observe that solutions from the intermediate iterations (i.e., lower bound approximations) are not suitable to approximate the optimal solution. The reason is that even at low optimality gap ($< 1\%$), it may have disastrous reliability ($< 70\%$). In addition, we also observe that higher objective costs do not always guarantee higher reliability in the out-of-sample tests.

6.4.2.4 Conservative Approximations and OPS

In this section, we provide several other options to approximate the optimal solutions by using the conservative approximation together with the OPS solutions. We

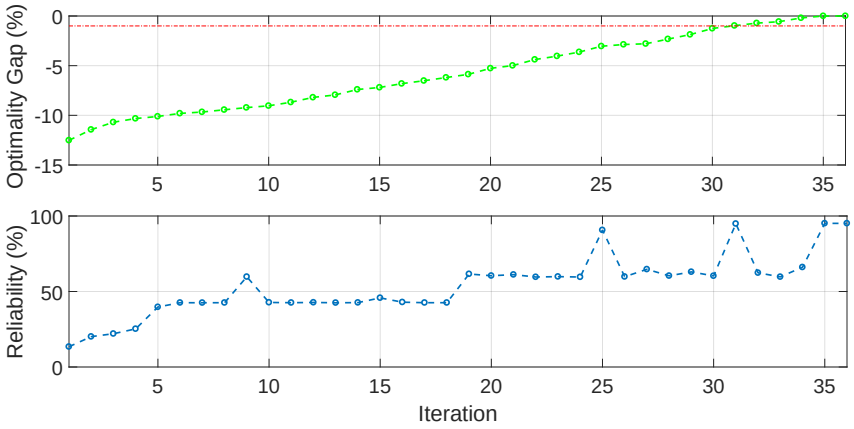


Figure 6.4: Optimality Gap and reliability of intermediate-iteration solutions of Algorithm 2 (red dash marks 1% threshold optimality gap).

propose 5 different options to carry out the conservative approximation: Given K in Proposition VI.3:

- UB ([49] and Chapter III): an online approach that uses the worst case τ^* on the violated constraints from the separation problem in Algorithm 2 for all iterations before $K-1$.
- OPS0: an online approach that uses the OPS solutions $|\mathcal{S}| = K-1$ on the violated constraints from the separation problem in Algorithm 2.
- OPS1: an offline approach that uses the OPS solutions $|\mathcal{S}| = K-1$ on all the chance constraints.
- OPS2: an aggregated version of OPS0 that use all the OPS solutions with $1 \leq |\mathcal{S}| \leq K-1$.
- OPS3: an aggregated version of OPS1 that use all the OPS solutions with $1 \leq |\mathcal{S}| \leq K-1$.

Note that if $|\mathcal{S}| = 1$, the solution of Algorithm 2 trivially coincides with Proposition VI.3 with $K = 2$. Comparing the online (OPS0 and OPS2) and offline (OPS1 and OPS3) options, online options require the information about which DR chance

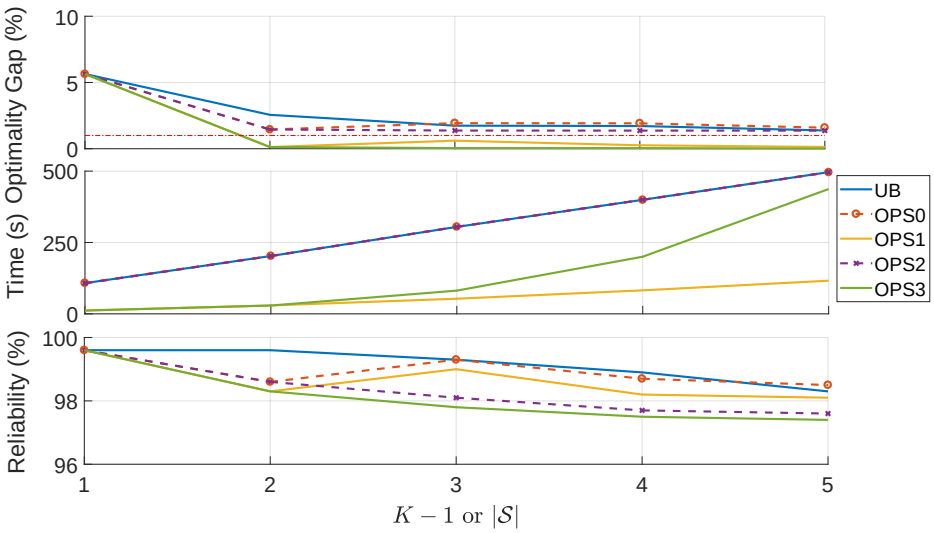


Figure 6.5: Overall performance of the conservative approximations on the 118-bus system with DS1 (red dash marks 1% threshold optimality gap).

constraints are violated in the separation problem in Algorithm 2, but offline options apply the conservative approximations with OPS on all of the DR chance constraints. Further, if we compare normal (OPS0 and OPS1) and aggregated versions (OPS2 and OPS3), the aggregated versions take advantage of the OPS solutions from smaller parameter dimension and hence the OPS solutions from smaller $|\mathcal{S}|$ are subsets of the ones from larger $|\mathcal{S}|$. However, the normal versions do not include the OPS solutions from smaller $|\mathcal{S}|$.

Next, we use the same 4 test cases and analyze their overall performance. For UB, we check the resulting solutions from all iterations in the separation problem. For OPS0/1/2/3 options, we limit $|\mathcal{S}| \leq 5$ for the 118-bus system and $|\mathcal{S}| \leq 4$ for the 300-bus system.

Figure 6.5 shows the comparison of the conservative approximations on the 118-bus system with DS1. UB fails to converge into the 1% optimality gap as K increases to the maximum number of iterations. OPS0 and OPS1 demonstrate better convergence rates and optimality gaps but the cost does not continue to decrease as $|\mathcal{S}|$ increases. The reason is that the OPS solutions for smaller $|\mathcal{S}|$ values are not

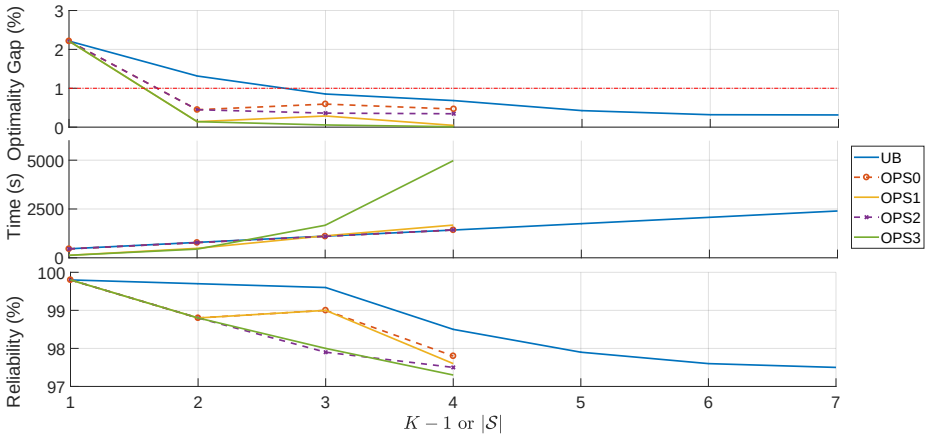


Figure 6.6: Overall performance of the conservative approximations on the 300-bus system with DS1 (red dash marks 1% threshold optimality gap).

a subset of the OPS solutions for larger $|\mathcal{S}|$. On the other hand, we also see that OPS2 and OPS3 show similar convergence rates as well as non-increasing costs as $|\mathcal{S}|$ increases since they aggregate the OPS solutions. In addition, OPS1 and OPS3 successfully converge to the 1% optimality gap when $|\mathcal{S}| \geq 2$. In terms of the solving time, we observe that all the approximation solutions except the last solutions of UB/OPS0/OPS1 take less time than exactly solving the DR-U (483.4s). Specifically, the offline OPS options: OPS1 and OPS3 show overall faster convergence than the other online options. Actually, the first few solutions in OPS1 and OPS3 even take similar amount of time as the single-run approaches (i.e., GA, SC, and DR-M). Online options UB/OPS0/OPS2 show linear relationships between solving time and $K-1$ or $|\mathcal{S}|$. However, the solving time in OPS1 and OPS3 grow faster as $|\mathcal{S}|$ increases. In terms of reliability, all the approximation solutions satisfy the 95% level and follow the tendencies of the optimality gaps since the objective costs from the conservative approximations are higher than the exact optimal cost. Hence, solutions with larger optimality gaps are more conservative with higher reliability than our requirement. Next, if we shift to the 300-bus system as shown in Fig. 6.6, similar properties can be observed except that the convergence time of OPS1 and OPS3 grow beyond that of the offline options when $|\mathcal{S}| \geq 3$. Hence, to approximate the optimal solution of

DR-U, we can either use offline options with small $|\mathcal{S}|$ or online options. Both of the options have small optimality gaps, high reliability and much less convergence time than solving the DR-U exactly (2263.6s).

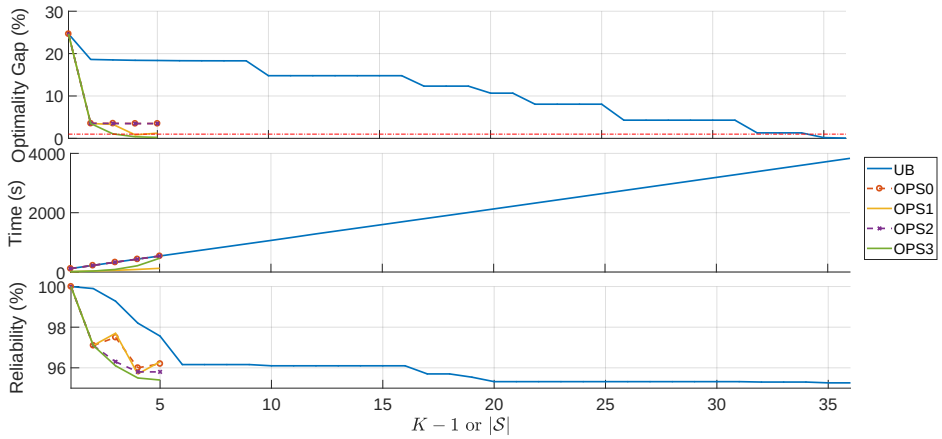


Figure 6.7: Overall performance of the conservative approximations on the 118-bus system with DS2 (red dash marks 1% threshold optimality gap).

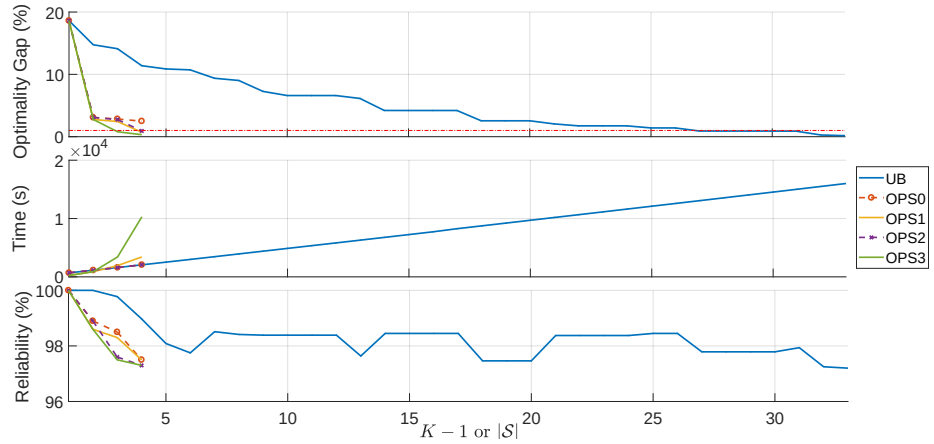


Figure 6.8: Overall performance of the conservative approximations on the 300-bus system with DS2 (red dash marks 1% threshold optimality gap).

Figure 6.7 shows similar test results on the 118-bus system with DS2. Since more iterations are required to solve DR-U, UB also shows much slower convergence rate and larger initial optimality gap 25% that finally converges to 1% optimality gap when $K-1=35$. Meanwhile, all OPS options show excellent convergence rates even if we limit $|\mathcal{S}| \leq 5$. Online options OPS0 and OPS2 converge when $|\mathcal{S}| = 2$

with optimality gaps $< 5\%$. Offline options OPS1 and OPS3 achieve optimality gaps around the 1% threshold as $|\mathcal{S}|$ increases to the limits. For convergence rate, all OPS options are much faster than the UB or exactly solving the DR-U (3815.9s). Among the OPS options, similar linear and nonlinear properties in terms of the solving times are observed as in Fig. 6.5. For reliability, we see all options still satisfy the 95% probability level and the UB reliability converges faster than its optimality gap. The reliability of all the OPS options follow the same trend as the optimality gap results. However, we also observe two interesting results that: OPS0 and OPS2 have the same costs when $|\mathcal{S}| \geq 2$ but OPS0 has higher reliability. OPS1 has lower cost than OPS0 and OPS2 but has higher reliability. Next, we check the 300-bus system with DS2 as shown in Fig. 6.8. We observe similar properties as in Fig. 6.7 and all online and offline options give solutions with better performance much faster than exactly solving the DR-U (15720.6s). We also see changes on the solving time difference between the online and offline options similar to Fig. 6.6 due to the increase on the problem dimension. Further, we also see an oscillation on the reliability of UB which is similar to Fig. 6.4 which shows that reliability is not always driven by the cost.

In addition to the results above, we also observe that the conservative approximations from DS1 generally have smaller optimality gaps than DS2. The solving time advantages of the offline options start to decrease when system dimension increases. In general, we can see that, conservative approximations are suitable to approximate the optimal solution of DR-U since they provide solutions with smaller optimality gaps and higher reliability. Especially with OPS, we are able to use the offline approximations or improve the online approximations to achieve better convergence rate and high-quality approximate solutions with less solving time.

6.5 Conclusions

In this chapter, we developed an optimal parameter selection approach to solve a DRCC OPF problem. We further derived optimality and existence guarantee as well as an efficient solving algorithm. The approach extended the existing DRCC methodology that considers the moment and unimodality information by effectively seek high-quality approximate solutions.

To evaluate the approaches and new ideas, we compared the scenario-based method, analytical reformulation under Gaussian assumption, and DRCC approaches with moment or unimodality information. We performed case studies on modified IEEE 118-bus and 300-bus systems using two wind uncertainty data sets, and observed that including unimodality information in a moment-based DRCC approach greatly improves the trade-off between objective costs and reliability compared with other approaches. However, this approach also requires larger computational efforts because the separation algorithm takes more iterations than the single-run approaches.

Detailed analysis indicates that the separation algorithm's computation effort is approximately the same in each iteration. Additionally, conservative approximations proved to be more reliable than relaxed approximation for efficient determination of optimal approximates. Furthermore, optimal parameter selection enabled the development of multiple online and offline options that outperformed the existing conservative approximation approach by providing solutions with low optimality gaps that satisfied reliability level and that required much less solving time.

Potential areas for future research include comparing DR approaches using different ambiguity sets, solving optimization problems other than OPF, and improving the performance of current DR approaches.

6.6 Supporting Material

6.6.0.5 Proof of Theorem VI.4

Condition 1: The last piece of $h(\tau)$, i.e., $h_{|S|}(\tau)$ must have zero slope because otherwise the error is infinite (if the slope is strictly positive) or $h_{|S|}(\tau) < v(\tau)$ for a sufficiently large τ (if the slope is strictly negative). It follows that $h_{|S|}(\tau) = \lim_{\tau\to\infty} v(\tau) = \sqrt{(1-\epsilon)/\epsilon}$ because this is the constant function that dominates $v(\tau)$ with the smallest error.

Condition 2: Since $v(\tau)$ is non-decreasing and concave, we have $d_s \geq 0$ and d_s is non-increasing in s. If a piece of $h(\tau)$ is not tangent to $v(\tau)$, then we decrease the intercept of this piece until it meets $v(\tau)$. Note that this does not increase $e^{\max}$. Then, all the pieces of $h(\tau)$ are tangent to $v(\tau)$, with the only exception that $h_1(\tau_0) = v(\tau_0)$ and $h_1'(\tau_0) > v'(\tau_0)$. In the case of this exception, we rotate $h_1(\tau)$ clockwise around the point $(\tau_0, v(\tau_0))$ until $h_1(\tau)$ becomes tangent to $v(\tau)$ at τ_0. Note that the rotation does not increase the error $e^{\max}$.

Condition 3: We prove by contradiction. Assume that $h^t(\tau)$ satisfies all three conditions and has an error $e^{t,\max}$, and there exists an $h^c(\tau)$ that satisfies Conditions 1 and 2 and has an error $e^{c,\max} < e^{t,\max}$.

If $|\mathcal{S}| = 1$, then $h^t(\tau) = h^c(\tau)$ due to Condition 1. This contradicts the assumption.

If $|\mathcal{S}| > 1$, since $e^{c,\max} < e^{t,\max} = E_2^t = E_{|\mathcal{S}|}^t$, we have $B_2^t > B_2^c$ (note that B_2^t and B_2^c represent the first break points of h^t and h^c other than τ_0, respectively) and $B_{|\mathcal{S}|}^t < B_{|\mathcal{S}|}^c$ (last break points of h^t and h^c, respectively). If $|\mathcal{S}| = 2$, this is a clear contradiction. If $|\mathcal{S}| > 2$, then there exists an $s \in [2, |S|-1]$ such that $[B_s^t, B_{s+1}^t] \subsetneqq [B_s^c, B_{s+1}^c]$, i.e., there exists a pair of pieces $h_s^t(\tau)$ and $h_s^c(\tau)$ with the same index s such that the domain of $h_s^t(\tau)$ is a strict subset of that of $h_s^c(\tau)$, because $h^t(\tau)$ and $h^c(\tau)$ have the same domain $[\tau_0, \infty)$ and the same number of pieces. According to Condition 2, both $h_s^t(\tau)$ and $h_s^c(\tau)$ are tangent to $v(\tau)$ and hence $e^{t,\max} = E_s^t = E_{s+1}^t \leq \max\{E_s^c, E_{s+1}^c\} \leq e^{c,\max}$,

contradicting the assumption.

6.6.0.6 Proof of Theorem VI.5

First, with a similar proof to that of Theorem VI.4, we can show that an $|S|$-piece PWL function $h(\tau)$ is an optimal outer approximation of $v(\tau)$ on a bounded interval $[\tau_0, \overline{\tau}]$ if it satisfies Conditions 2 and 3 in Theorem VI.4. We term this result Theorem VI.4-finite.

Second, we use mathematical induction to prove that there exists an $|\mathcal{S}|$-piece PWL approximation that satisfies the conditions of Theorem VI.4-finite, if $v(\tau)$ is defined on a bounded interval $[\tau_0, \overline{\tau}]$. When $|\mathcal{S}| = 1$, Condition 3 becomes $h(\tau_0) - v(\tau_0) = h(\overline{\tau}) - v(\overline{\tau})$. The single-piece optimal PWL approximation exists by simply searching for a point in $[\tau_0, \overline{\tau}]$ where $h(\tau)$ and $v(\tau)$ are tangent at.

Next, we show that if a C-piece optimal PWL approximation exists and satisfies the conditions of Theorem VI.4-finite, then so does a $(C+1)$-piece optimal PWL approximation. In the C-piece approximation, denote the second largest break point as B_F. According to the induction assumption, there exists a C-piece approximation on $[\tau_0, B_F]$ and a single-piece approximation on $[B_F, \overline{\tau}]$. As we move B_F from τ_0 to $\overline{\tau}$, the error of the C-piece approximation continuously increases from zero to a finite positive number (i.e., the optimal error for a C-piece approximation on $[\tau_0, \overline{\tau}]$). Accordingly, the error of the single-piece approximation continuously decreases from a finite positive value (i.e., the optimal error for a single-piece approximation on $[\tau_0, \overline{\tau}]$) to zero. It follows that there exists a $B_F^* \in [\tau_0, \overline{\tau}]$ such that the corresponding error of the C-piece approximation (on $[\tau_0, B_F^*]$) equals that of the single-piece approximation (on $[B_F^*, \overline{\tau}]$). The resultant $(C+1)$-piece PWL approximation satisfies the conditions of Theorem VI.4-finite. The proof of the case with $\overline{\tau} = +\infty$ is similar and so omitted here.

6.6.0.7 Statistical analysis and filter on DS2

First, we demonstrate the poor forecast quality in DS2 by showing the distribution of the forecast error ratio for an instance node. Since the wind power realization has to be non-negative, this ratio is practically lower bounded by -100%. In the boxplot of Fig. 6.9, we see that the original data contains extreme outliers. Further, from the histogram of Fig. 6.9, we observe that the forecast error ratio is distributed with large magnitudes ($> 200\%$). As one of data points in DS2, a wind plant with a forecast of 2 MW actually produces an output of 20 MW with a forecast error ratio of 900%. Hence, if we directly use the data from DS2, we might get extreme wind power forecast errors after we scale the forecast up based on our testing systems.

To obtain appropriate data for the simulations while maintaining the distribution of the wind power forecast error, we scale all the forecast errors down by 60% and then filter out the extreme points with forecast error ratios larger than 100%. The resulting histogram and the boxplot are shown in Fig. 6.10 and we see that the new distribution is much more reasonable than Fig. 6.9.

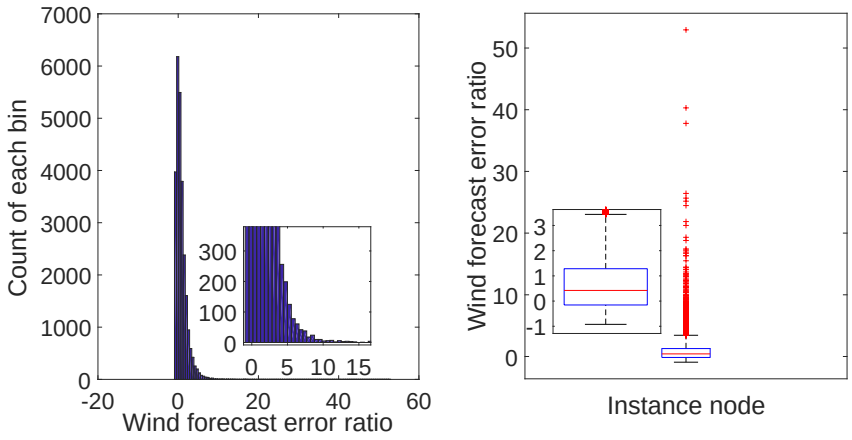


Figure 6.9: Histogram (100 bins) and the boxplot of the original wind power forecast error ratio of an instance node.

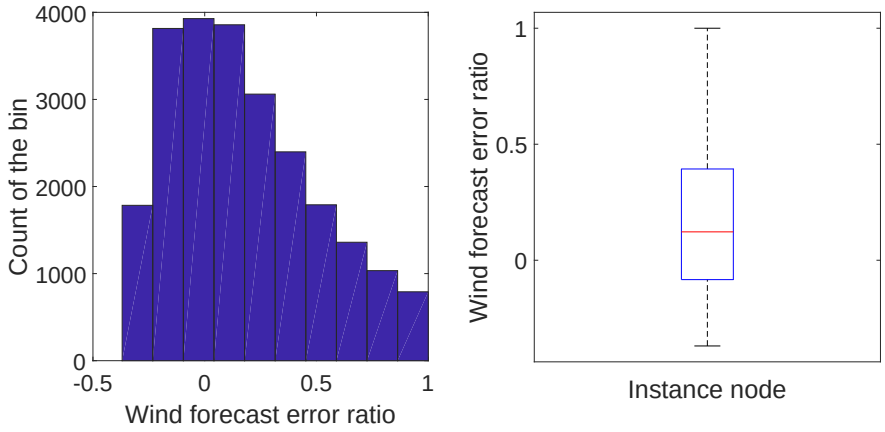


Figure 6.10: Histogram (10 bins) and the boxplot of the filtered wind power forecast error ratio of an instance node.

6.6.0.8 Validation of Unimodality

Here, we validate the unimodality of the wind power forecast error from DS1 and DS2 using 5000 data samples. In Figs. 6.11 and 6.12, we present the histograms of univariate and bivariate forecast errors with 15 and 20 bins for DS1 and DS2 respectively.

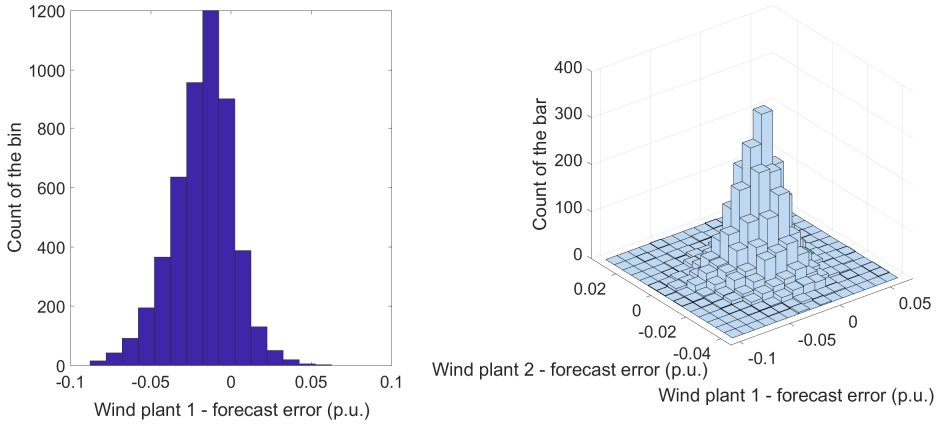


Figure 6.11: Histograms of univariate and bivariate wind power forecast errors of DS1 (15 bins).

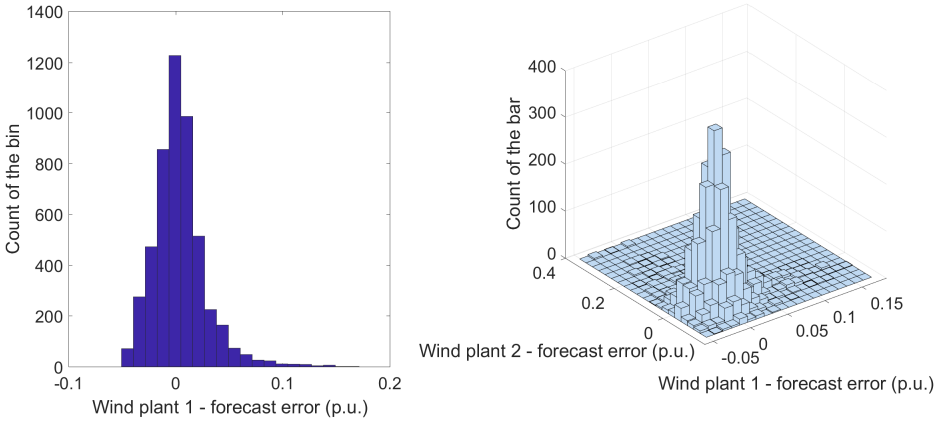


Figure 6.12: Histograms of univariate and bivariate wind power forecast errors of DS2 (20 bins).

In general, both figures empirically justify our assumption that the probability distribution of wind power forecast errors is unimodal and hence satisfies Assumption V.1. Furthermore, we also observe that the distribution of DS2 is more skewed than DS1.

CHAPTER VII

Conclusion and Future Work

With an increasing amount of uncertainties involved with power system operation, special techniques have to be developed to alleviate the influence from the uncertainties to the system feasibility. For example, uncertain power output from increasing penetration of renewable generation might lead to large variability to the network power flow and severe supply and demand mismatch across the system. Stochastic optimal power flow problems with chance constraints are formulated to not only help reduce the risk of physical constraint violations but also minimize the operational cost of generations and reserves. To solve the problem, most existing methodologies either produce conservative results with high objective costs or suffer from solutions with low reliability against uncertainty. Based on these facts, new methodologies are needed to achieve better performance with low cost and high reliability for realistic optimal power flow problems under different uncertainties. Recent research demonstrates that distributionally robust optimization can outperform the conventional approaches by using properly defined ambiguity set that summarizes the potential probability distributions of the uncertainty.

This dissertation analyzed the performance of the conventional approaches, developed new distributionally robust approaches with strengthened ambiguity sets and performed related case studies to illustrate the performance of those approaches.

The first part of the dissertation dealt with the difficulties of applying analytical reformulations assuming Gaussian distributions to an optimal power flow problem with uncertain load-based reserves. Chapter II discussed the development of new techniques to reformulate the problem and proved that the reformulation is convex. These techniques were compared with the scenario-based method, and the results indicated that analytical reformulation provides less conservative results but the optimal solutions suffer from low reliability in the out-of-sample tests. Case studies were also performed to investigate the impacts of several factors including uncertainty levels, reserve costs, and solution methodologies on power system dispatch, operational costs, and CO_2 emissions and illustrated complicated trade-offs from different factors.

The second part of the dissertation dealt with developing new distributionally robust formulations with strengthened ambiguity sets. Chapter III explained how to strengthened a moment-based ambiguity set with generalized unimodality information and derived tractable reformulations and approximations as well as efficient solving algorithms. Case studies compared with the original moment-based ambiguity set demonstrated that the new ambiguity set including the unimodality information results in less conservative solutions but had similar computational requirements.

Similarly, Chapter IV demonstrated how to strengthened a moment and support-based ambiguity set with log-concavity information and derived tractable sandwich approximations. Case studies illustrated a better trade-off between objective cost and reliability compared with the conventional ambiguity set excluding log-concavity information. Chapter V extended the work of Chapter III by assuming that the mode value from the unimodality information can be misspecified but still be bounded by either an ellipsoidal or rectangular support. Case studies demonstrated that the solution of the distributionally robust approaches including unimodality information depend significantly on the quality of mode estimate. Considering misspecified modes in the ambiguity sets effectively reduces the risk of generating solutions with low

reliability against uncertainty.

Chapter VI proposed an optimal parameter selection scheme to improve the conservative approximation of the distributionally robust approach with unimodality by efficiently finding high-quality approximates of the global optimal solution. In addition, simulations were conducted to compare the conventional approaches with the newly developed distributionally robust approach with unimodality information. Case studies demonstrated that strengthening ambiguity sets with unimodality information helps to significantly improve the trade-off between objective cost and reliability. However, this approach also requires larger computational efforts. In addition, by using optimal parameter selection, high-quality approximates of the global optimal solution can be efficiently found with a low optimality gap and high reliability.

7.0.0.9 Future Work

This section will introduce two potential directions for future work after the final defense.

Improvements on the distributionally robust approaches Based on the dissertation results, we propose the following potential areas to improve the developed distributionally robust approaches:

- To improve the solution quality, the exact reformulation of the distributionally robust approaches using log-concavity information will be derived. This result will further improve the trade-off between objective cost and reliability compared with Chapter IV.

- To reduce the computational time, parallel or distributed algorithms will be developed to tackle the computational burden of implementing the current iterative algorithm on large-scale systems. For example, in the distributionally robust approach with unimodality information, a high percentage of simulation

time is used to solve the separation problem; however, it can be parallelized to reduce computation effort.

- Distributionally robust optimization with joint chance constraints (e.g., two-sided constraint with lower and upper bounds) or nonlinear constraints will be derived.

Power System Adaptation under Uncertain Natural Disasters This project will be in collaboration with the Los Alamos National Laboratory. In this project, I propose to develop efficient algorithms to solve a multi-period power system adaptation problem (i.e., bus hardening and expansion) that is scalable in the problem dimension and scenario size. The objective of this project is to improve the resiliency of coastal power networks that face potential threats to sea level changes and storm surge events. In terms of the optimization model, the original problem is formulated as a two-stage joint chance-constrained problem with finite support and feasible mixed binary recourse. The problem is difficult to solve because it is a combination of stochastic programming and mixed integer programming. The only algorithm available is a scenario-based approach with finite convergence and a global optimality guarantee. However, the algorithm suffers from computational intractability when there are many scenarios. To improve computational performance, I first reformulate the problem to avoid faulty decisions under extreme scenarios while providing scenario dominance relationships. Next, to speed up the algorithm, I will develop techniques that are based on warm starts, bound tightening, smart constraints based on network topology, and a scalable algorithm with improved computational tractability.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Ahmed and D. Papageorgiou. Probabilistic set covering with correlations. *Operations Research*, 61(2):438–452, 2013.

[2] S. Ahmed and A. Shapiro. Solving chance-constrained stochastic programs via sampling and integer programming. *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, pages 261–269, 2008.

[3] P. Artznera, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[4] M. Bagnoli and T. Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26(2):445–469, 2005.

[5] S. Bashash and H. K. Fathy. Modeling and control of aggregate air conditioning loads for robust renewable power management. *IEEE Trans Control Systems Technology*, 21(4):1318–1327, 2013.

[6] A.R. Bergen and V. Vittal. *Power Systems Analysis*. Prentice Hall, 1999.

[7] D. Bertsimas, X.V. Doan, K. Natarajan, and C.P. Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[8] D. Bertsimas and M. Sim. Tractable Approximations to Robust Conic Optimization Problems. *Mathematical Programming, Series B*, 107:5–36, 2006.

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

[10] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart. Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems*, 23(3):983–991, 2008.

[11] J.H. Bookbinder and J.Y. Tan. Strategies for the probabilistic lot-sizing problem with service-level constraints. *Management Science*, 34(9):1096–1108, 1988.

[12] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[13] G. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[14] California ISO. Regulation energy management draft final proposal, January 2011.

[15] D.S. Callaway. Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Conversion and Management*, 50:1389–1400, 2009.

[16] D.S. Callaway and I.A. Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184–199, 2011.

[17] M. Campi, G. Calafiore, and Maria Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149–157, 2009.

[18] A. Charnes, W. Cooper, and G. Symonds. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.

[19] J. Cheng, E. Delage, and A. Lisser. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506, 2014.

[20] C. Coffrin, D. Gordon, and P. Scott. Nesta, the NICTA energy system test case archive. *arXiv preprint arXiv:1411.0359v5*, 2016.

[21] M. G. Cox. An algorithm for approximating convex functions by means by first degree splines. *The Computer Journal*, 14(3):272–275, 1971.

[22] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[23] S. W. Dharmadhikari and K. Joag-Dev. *Unimodality, convexity, and applications*. Academic Press, 1988.

[24] R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Trans Power Systems*, 20(2):587–595, 2005.

[25] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu. Distributionally robust chance-constrained approximate ac-opf with wasserstein metric. *IEEE Trans Power Systems*, 33(5):4924–4936, 2018.

[26] E. Erdogan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.

[27] P.M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable re-

formulations. Available at Optimization Online: http://www.optimization-online.org/DB_FILE/2015/05/4899.pdf.

[28] D. Gade and S. Küçükyavuz. Formulations for dynamic lot sizing with service levels. *Naval Research Logistics*, 60(2):87–101, 2013.

[29] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[30] A. Gómez-Expósito, A.J. Conejo, and C. Cañizares. *Electric Energy Systems: Analysis and Operation*. CRC Press, 2008.

[31] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences. 2008.

[32] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[33] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. Summers. Stochastic optimal power flow based on data-driven distributionally robust optimization. In *IEEE Annual American Control Conference*, Milwaukee, WI, 2018.

[34] D.A. Halamay, T.K.A. Brekken, A. Simmons, and S. McArthur. Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation. *IEEE Trans Sustainable Energy*, 2(3):321–328, Jul. 2011.

[35] G.A. Hanasusanto. *Decision Making under Uncertainty: Robust and Data-Driven Approaches*. PhD thesis, Imperial College London, 2015.

[36] G.A. Hanasusanto, V. Roitch, and D. Kuhn W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion informa-

tion. Available at Optimization Online: http://www.optimization-online.org/DB_FILE/2015/11/5199.pdf.

[37] H. Hao, T. Middelkoop, P. Barooah, and S. Meyn. How demand response from commercial buildings will provide the regulation needs of the grid. In *Allerton Conference on Communication, Control, and Computing*, Monticello, IL, 2012.

[38] R. Henrion, P. Li, A. Möller, M.C. Steinbach, M. Wendt, and G. Wozny. Stochastic optimization for operating chemical processes under uncertainty. *Online Optimization of Large Scale Systems*, pages 457–478, 2001.

[39] R. Henrion and A. Möller. Optimization of a continuous distillation process under random inflow rate. *Computers & Mathematics with Applications*, 45(1):247–262, 2003.

[40] A. Hoke, R. Butler, J. Hambrick, and B. Kroposki. Steady-state analysis of maximum photovoltaic penetration levels on typical distribution feeders. *IEEE Trans Sustainable Energy*, 4(2):350–357, 2013.

[41] A. Imamoto and B. Tang. A recursive descent algorithm for finding the optimal minimax piecewise linear approximation of convex functions. In *World Congress on Engineering and Computer Science*, San Francisco, CA, 2008.

[42] R. A. Jabr. Adjustable robust OPF with renewable energy sources. *IEEE Trans Power Systems*, 28(4):4742–4751, 2013.

[43] T. V. Jensen and P. Pinson. RE-Europe, a large-scale dataset for modeling a highly renewable european electricity system. *Scientific data*, 4:170–175, 2017.

[44] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1):291–327, 2016.

[45] W. Katzenstein and J. Apt. Air emissions due to wind and solar power. *Environmental Science and Technology*, 43:253–258, 2009.

[46] A. Khodaei, M. Shahidehpour, and S. Bahramirad. SCUC with hourly demand response considering intertemporal load characteristics. *IEEE Trans Smart Grid*, 2(3):564–571, Sept 2011.

[47] L. M. Kimball, K. A. Clements, and P. W. Davis. Stochastic opf via bender's method. In *IEEE PowerTech Conference*, Porto, Portugal, 2001.

[48] B. Li, R. Jiang, and J. L. Mathieu. Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In *IEEE Conference on Decision and Control*, Las Vegas, NV, 2016.

[49] B. Li, R. Jiang, and J.L. Mathieu. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming (Accepted)*, 2017.

[50] B. Li, J. L. Mathieu, and R. Jiang. Distributionally robust chance constrained optimal power flow assuming log-concave distributions. In *Power Systems Computation Conference*, Dublin, Ireland, 2018.

[51] B. Li and J.L. Mathieu. Analytical reformulation of chance-constrained optimal power flow with uncertain load control. In *IEEE PowerTech*, Eindhoven, Netherlands, 2015.

[52] B. Li, M. Vrakopoulou, and J.L. Mathieu. Chance constrained reserve scheduling using uncertain controllable loads Part II: Analytical reformulation. *IEEE Trans Smart Grid (in press)*, 2017.

[53] Chiao-Ting Li, Changsun Ahn, Huei Peng, and Jing Sun. Synergistic control of plug-in vehicle charging and wind power scheduling. *IEEE Trans Power Systems*, 28(2):1113–1121, May 2013.

[54] Y. Lin, J.X. Johnson, and J.L. Mathieu. Emissions impacts of using energy storage for power system reserves. *Applied Energy*, 168:444–456, 2016.

[55] X. Lu, K. W. Chan, S. Xia, B. Zhou, and X. Luo. Security-constrained multi-period economic dispatch with renewable energy utilizing distributionally robust optimization. *IEEE Trans Sustainable Energy*, 2018.

[56] M. Lubin, Y. Dvorkin, and S. Backhaus. A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Trans Power Systems*, 31(5):3840–3849, 2016.

[57] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello. Operational impacts of wind generation on california power systems. *IEEE Trans Power Systems*, 24(2):1039–1050, 2009.

[58] K. Margellos, P. Goulart, and J. Lygeros. On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Trans Automatic Control*, 59(8):2258–2263, 2014.

[59] G. Martinez, J. Liu, B. Li, J.L Mathieu, and C.L. Anderson. Enabling renewable resource integration: The balance between robustness and flexibility. In *Allerton Conference on Communication, Control, and Computing*, Monticello, IL, 2015.

[60] J.L. Mathieu, D.S. Callaway, and S. Kiliccote. Variability in automated responses of commercial buildings and industrial facilities to dynamic electricity prices. *Energy and Buildings*, 43:3322–3330, 2011.

[61] J.L. Mathieu, M. González Vayá, and G. Andersson. Uncertainty in the flexibility of aggregations of demand response resources. In *IEEE Industrial Electronics Society Annual Conference*, Vienna, Austria, 2013.

[62] J.L. Mathieu, M. Kamgarpour, J. Lygeros, G. Andersson, and D.S. Callaway. Arbitraging intraday wholesale energy market prices with aggregations of thermostatic loads. *IEEE Trans Power Systems*, 30(2):763–772, 2015.

[63] J.L. Mathieu, S. Koch, and D.S. Callaway. State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Trans Power Systems*, 28(1):430–440, 2013.

[64] S.P. Meyn, P. Barooah, A. Busic, Yue Chen, and J. Ehren. Ancillary service to the grid using intelligent deferrable loads. *IEEE Trans Automatic Control*, 60(11):2847–2862, Nov 2015.

[65] R. Mieth and Y. Dvorkin. Data-driven distributionally robust optimal power flow for distribution systems. *IEEE Control Systems Letters*, 2(3):363–368, 2018.

[66] B. Miller and H. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[67] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley & Sons, 1999.

[68] U. Ozturk, M. Mazumdar, and B. Norman. A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Trans on Power Systems*, 19(3):1589–1598, 2004.

[69] B. K. Pagnoncelli, , S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.

[70] G. Papaefthymiou and B. Klockl. MCMC for wind power simulation. *IEEE Trans Energy Conversion*, 23(1):234–240, 2008.

[71] A. Papavasiliou and S. Oren. A stochastic unit commitment model for integrating renewable supply and demand response. In *IEEE PES General Meeting*, San Diego, CA, 2012.

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

[73] V. S. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski. A stochastic model for the optimal operation of a wind-thermal power system. *IEEE Trans Power Systems*, 24(2):940–950, 2009.

[74] B.P.G. Van Parys, P.J. Goulart, and D. Kuhn. Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming*, 156(1):271–302, 2016.

[75] B.P.G. Van Parys, P.J. Goulart, and M. Morari. Distributionally robust expectation inequalities for structured distributions. *Mathematical Programming (in press)*, 2017.

[76] D. Phan and S. Ghosh. Two-stage stochastic optimization for optimal power flow under renewable generation uncertainty. *ACM Trans Modeling and Computer Simulation*, 24(1):2:1–2:22, 2014.

[77] I. Popescu. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.

[78] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[79] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson. Analytical reformulation of security constrained optimal power flow with probabilistic constraints. In *IEEE PowerTech Conference*, Grenoble, France, 2013.

[80] L. Roald, F. Oldewurtel, B. Van Parys, and G. Andersson. Security constrained optimal power flow with distributionally robust chance constraints. *arXiv preprint arXiv:1508.06061*, 2015.

[81] R.T. Rockafellar and S. Uryasev. Optimization of conditional Value-at-Risk. *Journal of Risk*, 2:21–42, 2000.

[82] R.T. Rockafellar and S. Uryasev. Conditional Value-at-Risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

[83] H. Scarf. A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209.

[84] A. Shapiro. On duality theory of conic linear problems. *Semi-Infinite Programming*, pages 135–165.

[85] A. Shapiro and A. Kleywegt. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[86] A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

[87] B. Stellato. *Data-driven chance constrained optimization*. Master thesis, ETH Zurich, 2014.

[88] T. Summers, J. Warrington, M. Morari, and J. Lygeros. Stochastic optimal power flow based on conditional value at risk and distributional robustness. *International Journal of Electrical Power & Energy Systems*, 72:116 – 125, 2015.

[89] X. Tong, H. Sun, X. Luo, and Q. Zheng. Distributionally robust chance constrained optimization for economic dispatch in renewable energy integrated systems. *Journal of Global Optimization*, 70(1):131–158, 2018.

[90] S. H. Tseng, E. Bitar, and A. Tang. Random convex approximations of ambiguous chance constrained programs. In *IEEE Conference on Decision and Control*, Las Vegas, NV, 2016.

[91] L. Vandenberghe, S. Boyd, and K. K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.

[92] J. Vandewalle. On the calculation of the piecewise linear approximation to a discrete function. *IEEE Trans on Computers*, 24:843–846, 1975.

[93] M. Vrakopoulou, B. Li, and J.L. Mathieu. Chance constrained reserve scheduling using uncertain controllable loads Part I: Formulation and scenario-based analysis. *IEEE Trans Smart Grid (in press)*, 2017.

[94] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. A probabilistic framework for security constrained reserve scheduling of networks with wind power generation. In *IEEE EnergyCon*, Florence, Italy, 2012.

[95] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Probabilistic guarantees for the N-1 security of systems with wind power generation. In *International Conference on Probabilistic Methods Applied to Power Systems*, Istanbul, Turkey, 2012.

[96] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration. *IEEE Trans Power Systems*, 28(4), 2013.

[97] M. Vrakopoulou, J.L. Mathieu, and G. Andersson. Stochastic optimal power flow with uncertain reserves from demand response. In *Hawaii International Conference on System Science*, Waikoloa, HI, 2014.

[98] M. Wagner. Stochastic 0–1 linear programming under limited distributional information. *Operations Research Letters*, 36(2):150–156, 2008.

[99] C. Wang, R. Gao, F. Qiu, J. Wang, and L. Xin. Risk-based distributionally robust optimal power flow with dynamic line rating. *IEEE Trans Power Systems*, pages 1–1, 2018.

[100] Q. Wang, Y. Guan, and J. Wang. A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. *IEEE Trans on Power Systems*, 27(1):206–215, 2012.

[101] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[102] W. Xie and S. Ahmed. Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Trans Power Systems*, 33(2):1860–1867, 2018.

[103] T. Yong and R. H. Lasseter. Stochastic optimal power flow: formulation and solution. In *Power Engineering Society Summer Meeting*, Seattle, WA, 2000.

[104] H. Zhang and P. Li. Chance constrained programming for optimal power flow under uncertainty. *IEEE Trans Power Systems*, 26(4):2417–2424, 2011.

[105] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi. Aggregated modeling and control of air conditioning loads for demand response. *IEEE Trans Power Systems*, 28(4):4655–4664, 2013.

[106] X. Zhang, K. Margellos, P. Goulart, and J. Lygeros. Stochastic model predictive control using a combination of randomized and robust optimization. In *IEEE Conference on Decision and Control*, Florence, Italy, 2013.

[107] Y. Zhang, S. Shen, and J. L. Mathieu. Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. *IEEE Trans Power Systems*, 32(2):1378–1388, 2017.

[108] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans Power Systems*, 26(1):12–19, 2011.

[109] S. Zymler, D. Kuhn, and B. Rustem. Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188, 2013.

[110] S. Zymler, D.D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1):167–198, 2013.