Data-Driven Distributionally Robust Optimization for Power System Operations

by

Yuanyuan Guo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2019

Doctoral Committee:

Assistant Professor Ruiwei Jiang, Chair
Professor Marina A. Epelman
Assistant Professor Johanna L. Mathieu
Associate Professor Cong Shi

Yuanyuan Guo

yuanyg@umich.edu

ORCID iD:0000-0003-0942-8719

To my parents and grandparents.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor Prof. Ruiwei Jiang for his guidance and patience in the past five years. Working with him is the greatest privilege of my life. Without his support and encouragement, this dissertation would not have become possible. I cannot describe how much I have learned from him, both academically and personally. His passion, wisdom in research, kind and gracious care for students will always inspire me to work hard and become someone like him.

My appreciation also goes to my committee members. I want to thank Prof. Epelman for her remarkable advice, Prof. Mathieu for her insightful ideas and Prof. Shi for his valuable comments. Their knowledge and valuable suggestions help me a lot in improving my dissertation.

Industrial and Operations Engineering at the University of Michigan is a great home to me. I appreciate all of help and happiness I obtained here. I am grateful to have tremendous opportunities to learn from so many excellent teachers. Many thanks to my amazing graduate student colleagues. Their intelligence and support help me have a memorable Ph.D. life. Thanks also to Tina, Chris and Matt for their always kindly and patiently help to my questions and requests. Special thanks to my fiance, Xiangkun, I am so fortunate to have you in my life.

Finally, I would like to thank my parents for always believing in me and always being there. They raised me with a love of knowledge and always support and encourage me to chase my dream.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ABSTRACT

Decisions are often made in an uncertain environment. For example, in power system operations, decision makers need to schedule generators without the accurate outcome of various uncertain parameters, e.g., electricity load and renewable energy. If we have access to the (joint) probability distribution of these uncertain parameters, we can apply the stochastic programming approaches to schedule the generators. However, in practice, we can hardly have access to such (true) probability distribution. Under such circumstances, stochastic programming approaches may produce over-optimistic solutions and lead to an unreliable power system. In this dissertation, we propose data-driven optimization methods to model uncertainty directly based on the historical data. More specifically, we statistically infer key characteristics of the (ambiguous) probability distribution (e.g., support, mean, mean absolute deviation, unimodality, etc.) based on the historical data and construct an ambiguity set consisting of all probability distributions that match the inferred characteristics. Then, we make distributionally robust decisions that hedge against the worst-case distributions within the ambiguity set. We study new data-driven distributionally robust optimization models as well as their solution approaches and applications on power system operations, including optimal power flow, unit commitment, and transmission expansion planning. The specific contributions of this dissertation include (i) a distributionally robust optimization approach for unit commitment and reserve procurement and an algorithm based on generalized linear decision rule; (ii) a Wasserstein-moment ambiguity set for the distributionally robust chance-constrained optimal power flow problem and a tractable convex conservative approximation based on worst-case conditional value-at-risk; (iii) a general framework for distributionally robust optimization models to incorporate shape information of probability distributions into ambiguity sets; (iv) an adaptive two-stage robust transmission expansion planning model and an algorithm based on prioritization decision rule.

CHAPTER 1

Introduction

Decisions are often made in an uncertain environment. For example, in power system operations, decision makers need to decide generation schedule or transmission expansion plan without the accurate outcome of various uncertain parameters, e.g., electricity loads and/or renewable energy. The stochastic programming (SP) is one of the most popular approaches for making decisions under uncertainty. SP often describes uncertainty by a number of possible scenarios, each with a corresponding probability. For example, SP approaches for unit commitment problem (see, e.g., [10, 23, 135, 121, 104]) model the renewable energy uncertainty by a number of possible scenarios. Then, a two-stage stochastic program can be formulated to decide the day-ahead unit commitment and reserves in the first-stage formulation and to decide economic dispatch in the second-stage formulation, with the objective of minimizing the expected total costs. On optimal power flow problem, an SP approach, called chance-constrained programming (see, e.g., [162, 141, 113, 21]) is widely used to hedge against uncertainty. In a chance-constrained program, system operators aim to satisfy a random constraint with high probability under a prescribed distribution at the smallest cost. A basic challenge to applying SP is that, in practice, it may be hard to obtain an accurate estimation of the (true) probability distribution due to the lack of historical data and/or the fluctuating nature of the uncertainty. Under such circumstances, SP may produce over-optimistic solutions and lead to disappointing out-of-sample performance in real-world implementation. Alternatively, another widely used approach is called robust optimization (RO). Different from SP, RO only assumes an uncertainty set of the random parameters, which can be conveniently inferred by using a minimal amount of distributional information. For example, robust transmission expansion minimizes the cost of the future transmission grid with regard to the worst-case scenario within an uncertainty set of parameters (e.g., renewable energy, load, etc.) [70, 32, 101]. To compute the worst-case cost, one needs to compute the optimal power flow after the uncertainties are realized, leading to a two-stage (adaptive) RO model. Similarly, robust unit commitment approaches aim to minimize the worst-case generation and reserve costs [75, 169, 18, 167, 55] and also lead

to two-stage RO models. Various decision rules have been proposed to solve the two-stage RO models (see, e.g., continuous recourse [149, 93] and mixed-integer recourse [60, 16]). The basic idea of decision rules is to make the recourse decisions be a function of the first-stage decision and the realized uncertainties.

RO ignores commonly available distributional information, e.g., mean and variance of the uncertainty. Also, RO only focuses on the worst-case scenario while omitting all other possible (and likely) scenarios. As a result, it is often concerned that the solutions obtained by RO are over-conservative. In this dissertation, we study data-driven distributionally robust optimization (DRO) approaches for power system decision-making under uncertainty. Unlike SP, DRO does not assume perfect knowledge of the probability distribution of the uncertain parameters. Instead, it considers a family of probability distributions that share certain characteristics of the true distribution. By incorporating the distributional information, DRO makes decisions under the worst-case distribution. More specifically, we statistically infer key characteristics of the (ambiguous) probability distribution (e.g., support, mean, mean absolute deviation from the median, unimodality, etc.) based on the historical data and construct an ambiguity set consisting of all distributions that match the inferred characteristics. Then, we make decisions that hedge against the worst-case distribution within the ambiguity set. Hence, the decisions obtained from DRO approaches are less conservative than their RO counterparts. DRO methods have received growing attention in the last decade (see, e.g., [37, 52, 151, 20, 73]) and recently have also been applied in power system operations (see, e.g., [53, 155, 74, 164, 34]). For example, [155] studied a DRO model for unit commitment problem (without reserve procurement decisions) considering uncertain wind power generation. A linear decision rule was applied to approximate the recourse decisions in the second-stage problem. [166] studied contingency-constrained unit commitment and proposed an iterative algorithm based on Benders' decomposition. Distributionally robust chance-constrained optimal power flow was studied in [164]. It considered the renewable energy uncertainty as well as the load reserve uncertainty. The ambiguity sets in these works consist of statistical information such as mean and covariance. Moment-based ambiguity sets can usually lead to computationally tractable reformulations, such as semidefinite program or second-order conic program. In many practical cases, we can obtain more distributional information than moments. For example, a reference distribution can be obtained from historical data by data-fitting. Although measurement errors are inevitable, the system operators may believe that the distance between the true distribution and the reference distribution should be small. With this assumption, more recently, probability discrepancy functions have been applied to construct ambiguity set, including the Prokhorov metric [45], the ϕ-divergence [73, 67], and the Wasserstein met-

ric [105, 152, 47]. The discrepancy ambiguity set is defined as a neighborhood surrounding a reference distribution, in which all candidate distributions are close to the reference one measured by the prescribed discrepancy. The degree of conservatism of discrepancy ambiguity sets can be controlled by adjusting the discrepancy upper bound to the reference distribution. A Kullback–Leibler divergence based ambiguity set was considered for unit commitment[34]. An iterative algorithm based on a Benders' decomposition was proposed and it can guarantee global convergence within finite iterations. In [42], distributionally robust chance-constrained optimal power flow was model with Wasserstein metric. It also considered an approximated ac model instead of more tractable dc model. In this dissertation, We study new data-driven DRO approaches as well as their solution approaches and applications on power system operations, including optimal power flow, unit commitment, and transmission expansion planning. We introduce the details of our approaches and applications in each chapter as follows.

In Chapter 2, we study a data-driven approach for unit commitment and reserve procurement. The unit commitment (UC) problem [103] aims to decide an efficient schedule of generation in the power system to satisfy the forecasted loads, and required safety constraints. UC and reserve procurement decisions are to ensure sufficient generation potentials in proper locations to reliably serve the forecasted loads under transmission congestion by leveraging off-line resources if necessary. Due to the low operational costs, low green house gas emission, and government incentives (e.g., tax credits), a growing share of renewable energy is incorporated into the power systems. However, the inherent volatility of renewable energy requires the UC and reserve decisions to be made under uncertainty. Several methods have been proposed to improve the UC and reserve procurement decisions under uncertainty (see, e.g., SP [10, 23, 135, 121, 104], RO [75, 169, 18, 167, 55]). From the perspective of distributional information, SP and RO lie in opposite ends of the spectrum. In this chapter, with the objective of optimizing day-ahead UC and reserve procurement, we study a data-driven approach that relies on a set of descriptive statistics of the renewable energy that can be conveniently estimated based on historical data, e.g., mean and mean absolute deviation from the median. This approach lies in between SP and RO. It does not need the accurate information of the uncertainty and meanwhile is less conservative than RO beacuse more distributional information is considered. We introduce the generalized linear decision rules to conservatively reformulate the data-driven model as a mixed-integer linear program, which facilitates efficient commercial solvers like CPLEX. We conduct a case study based on the real-world wind data. The out-of-sample tests indicate that, with a small amount of historical data, the proposed data-driven approach can enhance the system flexibility and capability of accommodating renewable energy.

In Chapter 3, we investigate the distributionally robust chance-constrained optimal power flow problem with a Wasserstein-moment ambiguity set. Optimal power flow (OPF) is one of the most critical decision-making problems in power system operations. It aims to minimize power system costs, e.g., generation costs and power loss, while satisfying the predicted loads and safety constraints on generators and transmission lines. Integrating uncertain renewable energy creates new challenges for the power system operators to ensure safety constraints, such as the transmission capacity limits. The chance-constrained program (CCP) is an effective and convenient method to control the risk when decisions are made under uncertainty. Specifically, CCP ensures that the safety constraints are satisfied with a pre-required high probability. However, major challenges of CCP include: (i) the joint probability distribution of the uncertainties is rarely accessible in reality and (ii) CCP is in general is computationally intractable. To address these challenges, distributionally robust chance constraints (DRCC) were proposed (see, e.g., [37, 171]). In DRCC, the uncertain constraints are required to be satisfied with given probability under all distributions within an ambiguity set. The ambiguity set contains distributions characterized by certain characteristics of the (unknown) true distribution. A critical component in applying DRCC is the choice of the ambiguity set. In this chapter, we propose a Wasserstein-moment ambiguity set which takes into account both probability discrepancy and moment information of the uncertainty. DRCC model with a Wasserstein-moment ambiguity set combines predictive, prescriptive, and preventive analytics. Therefore, this approach can directly incorporate the renewable data and provide powerful out-of-sample performance guarantee even with a limited number of data samples. In addition, by using conditional value-at-risk approximation, the model can be conservatively approximated as a linear program or a second-order cone program depending on the choice of the Wasserstein metric. In the numerical experiments with real-world wind forecast data, the proposed model shows improved performance over two benchmark approaches, including the Gaussian approximation based approach and the DRCC-OPF model with a moment ambiguity set.

In Chapter 4, we investigate the distributionally robust optimization using shape information. DRO is a natural and effective way of modeling decision-making problems under ambiguous uncertainty. In reality, in addition to historical data, the decision maker often possesses certain shape information of the probability distribution (e.g., skewness, sub-Gaussian, or sub-exponential properties, etc.), which can be taken into account to further reduce the conservatism. Unfortunately, a straightforward incorporation of such shape information often results in non-convex DRO models, making them computationally prohibitive. In this chapter, we investigate a DRO framework that models shape information in a computationally tractable manner. In particular, we consider a class of concentration

inequalities, potentially leading to an infinitely constrained ambiguity set. Then, we show that the corresponding DRO model can be conservatively approximated as a stochastic program with respect to an (unambiguous) probability distribution. This facilitates efficient solution algorithms (e.g., sample average approximation) for DRO models with shape information. In addition, we show that this approximation is tight for a wide class of DRO models. Finally, we demonstrate the theoretical results via computational case studies on the appointment scheduling problem and the risk-constrained optimal power flow problem.

In Chapter 5, we investigate the adaptive robust transmission expansion planning with prioritization. The transmission expansion planning (TEP) problem aims to decide an optimal augmentation plan of an existing transmission network to serve the forthcoming electric loads while satisfying the required security constraints. Due to the long planning horizon, the TEP problem inevitably involves significant uncertainties arising from, e.g., renewable energy, loads, and installation budget. RO has been widely used for expansion decision making under uncertainty [70, 32, 101]. To compute the worst-case cost, robust TEP needs to compute the optimal power flow after the uncertainties are realized, leading to a two-stage (adaptive) RO model. Ideally, we would like to install additional lines/circuits adaptively with respect to the unveiled uncertainties (e.g., budget, load, and renewable energy). Unfortunately, the adaptive installation result in a RO model with mixed-integer recourse, which is computationally prohibitive. One way to address this challenge is by waiving the option of adaptive installation and accordingly robust TEP reduces to a two-stage RO model with continuous recourse (see, e.g., [69]). We can then apply the column-and-constraint generation algorithm [160] to efficiently solve robust TEP. The second stream of literature designs decision rules to determine the adaptive installation as a function of the first-stage installation and the realized uncertainties (see, e.g., [60, 16]). In this chapter, we adopt the decision rule of prioritization [80, 81] to the two-stage adaptive robust TEP model. In particular, we rank all possible expansions in a priority list and commit to each expansion based on its priority and the realized uncertainty (e.g., the realized budget). The prioritization decision rule is consistent with the practice in industrial and governmental decision making and provides a more interpretable expansion plan. Additionally, it enables us to efficiently compute the adaptive binary variables in the proposed robust TEP model. A tailored column-and-constraint generation algorithm is designed to solve the resulting formulation. We demonstrate the performance of the proposed approach based on the Garver 6-bus and the southern Brazilian 46-bus system with the real-world wind data.

Finally, in Chapter 6, we conclude this dissertation and discuss future research directions.

CHAPTER 2

A Data-Driven Approach for Unit Commitment and Reserve Procurement

2.1 Introduction

The last decade has witnessed a worldwide emergence of renewable energy in the energy mix [110]. For example, during years 2008-2013, the total installed wind power capacity increased from 121GW to 318GW (increasing by 162.81%), and the total grid-connected solar photovoltaic (PV) capacity increased from 16GW to 139GW (increasing by 768.75%). Up to year 2012, the renewable energy accounts for more than 10% of the world energy consumption [68]. Due to the low operational costs, low green house gas emission, and government incentives (e.g., tax credits), the share of renewable energy (e.g., wind and solar) in the U.S. energy mix is projected to continue growing at an ambitious pace [136], and the Department of Energy has also analyzed a scenario where the wind power alone provides 20% of the U.S. electricity needs by year 2030 [91]. As more wind/solar energy are integrated in the energy mix, the power systems can become more cost-effective.

However, this blueprint remains far-fetched before the power systems can effectively accommodate the fluctuating renewable energy. Due to its inherent volatility, renewable energy is often difficult to predict. Meanwhile, it can be challenging to estimate the probability distributions of renewable energy. For example, it is observed that normal distribution is not able to accurately estimate the wind forecasting error in U.S. (see, e.g., [64, 63]). Besides, the forecasted probability distributions of solar power can largely deviate from the true distributions [163]. As compared to the existing renewable resources, less amounts of historical data are available for the new ones, contributing to the distributional ambiguity of renewable energy. As a result, it can be difficult to perfectly accommodate the renewable energy when faced with transmission congestions and physical restrictions of other generation resources. In order to maintain the power systems reliability, system operators

often need to curtail a considerable portion of renewable energy [116] which equivalently increases the use of fossil resources.

One possible approach to better accommodate the volatile renewable energy is by advanced unit commitment (UC) and reserve procurement, which enhance the system flexibility to adapt to the real-time fluctuations. In power system operations, various reserves (e.g., spinning and operating) can be synchronized or put online within a short amount of time (e.g., 10 min) to provide ramping capability [125, 144, 72]. In practice, several Independent System Operators (ISOs) (e.g., ERCOT) conduct reliability unit commitment (RUC) runs after the closure of day-ahead markets [150]. RUC runs ensure sufficient generation potential in proper locations to reliably serve the forecasted loads under transmission congestion by leveraging off-line resources if necessary. Several methods have been proposed and studied in the literature to improve the RUC runs. Deterministic approaches often describe the renewable energy uncertainty by their forecasted values and impose a fixed percentage of the total online capacity (and the capacity that can be quickly online) as energy reserves [170]. Alternatively, the stochastic UC approaches (see, e.g., [10, 23, 135, 121, 104]) are popular methods that describe the renewable energy uncertainty by a number of possible scenarios, each with a corresponding probability. Then, a two-stage stochastic program is formulated to decide the day-ahead UC and reserves in the first-stage problem and to decide economic dispatch (ED) in the second-stage problem, with the objective of minimizing the *expected* total costs. More recently, robust UC approaches (see, e.g., [75, 169, 18, 167, 55]) have been proposed to ensure the system reliability. As compared to the stochastic approaches, the robust UC approaches describe the renewable energy ξ by an uncertainty set (e.g., a polytope). Meanwhile, the robust UC approaches aim to minimize the *worst-case* total costs with respect to ξ that can be realized within the uncertainty set. Linear decision rules (LDRs, also known as affine policies) have been studied in the literature of stochastic programming (see, e.g., [83]) and robust optimization (see, e.g., [13]). In power system applications, LDRs have been successfully applied in chance-constrained optimal power flow (OPF) (see, e.g., [141, 21]), robust adjustable OPF (see, e.g., [147, 148, 70]), and robust unit commitment (see, e.g., [149, 93]).

From the perspective of distributional information, stochastic and robust UC approaches lie in opposite ends of the spectrum. On the one hand, stochastic UC approaches need an accurate estimate of $\mathbb{P}_\xi$ so that a set of scenarios can be generated for the two-stage stochastic UC formulation. However, as more renewable energy resources are integrated in the power systems, accurately estimating $\mathbb{P}_\xi$ can be difficult due to the lack of abundant historical data. Accordingly, the UC and reserves obtained from stochastic UC approaches can be sensitive to the $\mathbb{P}_\xi$ estimates. On the other hand, robust UC approaches need a min-

imal amount of distributional information, i.e., an uncertainty set of ξ that can often be conveniently estimated. However, robust UC approaches tend to ignore other commonly available distributional information, e.g., mean of ξ. Also, robust UC approaches focus only on the worst-case scenario of ξ while omitting other possible (and likely) scenarios. As a result, it is often concerned that robust UC solutions can be over-conservative. In this work, we propose a data-driven approach that relies on a set of descriptive statistics of the renewable energy that can be conveniently estimated based on historical data, e.g., mean and mean absolute deviation from the median (MAD-median). This approach lies in between the stochastic and robust UC approaches, because it does not require an accurate estimate of $\mathbb{P}_\xi$ and meanwhile uses more distributional information than the uncertainty set. As a result, the obtained UC and reserves are immune to biased probability distribution estimates and meanwhile less conservative than the robust solutions. The proposed approach applies the methodology of distributionally robust optimization (DRO) that was first studied in [124]. DRO methods have received growing attention in the last decade (see, e.g., [37, 52, 151, 20, 73]) and recently have also been applied in power system operations (see e.g., [53, 155, 74]). As compared to the existing literature, this work investigates the reserve procurement in a DRO framework. It is demonstrated in the case study that our approach encourages to procure more reserves from the thermal units, which leads to higher utilization of renewable energy and a lower average cost than in the robust UC approach.

In the remainder of this chapter, we describe the mathematical formulations in Section 2.2. In Section 2.3, we derive a solution methodology of the data-driven approach based on generalized linear decision rules. Case studies and results are reported in Section 2.4, before we conclude this research in Section 2.5. In addition, we list the nomenclature in Section 2.6.

2.2 Mathematical Formulation

In this section, we start by reviewing the two-stage stochastic UC models [75] that can be formulated as follows:

$$\min_{y,u,v,e,r} \sum_{t=1}^{T} \sum_{i\in\mathcal{I}} \left(\mathrm{SU}_i u_{it} + \mathrm{SD}_i v_{it} + \mathrm{NL}_i y_{it} + c_i^{\mathrm{U}} r_{it}^{\mathrm{U}} + c_i^{\mathrm{D}} r_{it}^{\mathrm{D}}\right) + \mathbb{E}_{\mathbb{P}_\xi}\left[Q(e,r,\xi)\right] \tag{2.1a}$$

$$\text{s.t. } y_{i(t-1)} - y_{it} + u_{it} \geq 0,\ \forall i \in \mathcal{I}\setminus\mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.1b}$$

$$v_{it} = y_{i(t-1)} - y_{it} + u_{it}, \forall i \in \mathcal{I}\setminus\mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.1c}$$

$$y_{it} - y_{i(t-1)} \leq y_{i\tau},\ \forall i \in \mathcal{I}\setminus\mathcal{I}_R, \forall t \in \mathcal{T},$$

$$\forall \tau = t+1, \ldots, \min\{t + \mathrm{MU}_i - 1, T\}, \tag{2.1d}$$

$$y_{i(t-1)} - y_{it} \leq 1 - y_{i\tau},\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R, \forall t \in \mathcal{T},$$

$$\forall \tau = t+1, \ldots, \min\{t + \mathrm{MD}_i - 1, T\}, \tag{2.1e}$$

$$e_{it} + r^{\mathrm{U}}_{it} \leq U_i y_{it},\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.1f}$$

$$e_{it} - r^{\mathrm{D}}_{it} \geq L_i y_{it},\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.1g}$$

$$y_{it}, u_{it}, v_{it} \in \{0,1\},\ e_{it}, r^{\mathrm{U}}_{it}, r^{\mathrm{D}}_{it} \geq 0, \forall i \in \mathcal{I} \setminus \mathcal{I}_R,\ \forall t \in \mathcal{T}, \tag{2.1h}$$

where $\mathcal{I} = \{1, \ldots, I\}$ is the set of all generators, $\mathcal{I}_R \subseteq \mathcal{I}$ is the set of generators that utilize renewable energy, $\mathcal{T} = \{1, \ldots, T\}$ is the set of operating time intervals, SU_i (resp. SD_i) is the start-up (resp. shut-down) cost for generator i, NL_i is the no-load cost of thermal generator i, c^{U}_i (resp. c^{D}_i) is the unit cost of upward (resp. downward) reserve amount of generator i, MU_i (resp. MD_i) is the minimum-up (resp. minimum-down) time for thermal generator i, U_i (resp. L_i) is the maximal (resp. minimal) power output if generator i is on, binary variable u_{it} (resp. v_{it}) indicates if generator i is started up (resp. shut down) at the beginning of time period t, binary variable y_{it} indicates if generator i is on during time period t, random variable ξ_{jt} is the available renewable energy of generator j in time period t, $\mathbb{P}_\xi$ is the probability distribution of ξ, and $Q(e, r, \xi)$ represents the fuel costs with respect to given scheduled electricity generation $e = (e_{it} : i \in \mathcal{I} \setminus \mathcal{I}_R, t \in \mathcal{T})$, upward/downward reserves $r = (r^{\mathrm{U}}_{it}/r^{\mathrm{D}}_{it} : i \in \mathcal{I} \setminus \mathcal{I}_R, t \in \mathcal{T})$, and realized renewable energy $\xi = (\xi_{jt} : j \in \mathcal{I}_R, t \in \mathcal{T})$. More specifically, the value of $Q(e, r, \xi)$ equals to the optimal objective value of a second-stage ED problem that is formulated as follows:

$$\min_{g} \ \sum_{t=1}^{T} \sum_{i \in \mathcal{I} \setminus \mathcal{I}_R} f_i(g_{it}) \tag{2.2a}$$

$$\text{s.t.} \ \sum_{i \in \mathcal{I}} g_{it} = \sum_{b=1}^{B} D_{bt},\ \ \forall t \in \mathcal{T}, \tag{2.2b}$$

$$e_{it} - r^{\mathrm{D}}_{it} \leq g_{it} \leq e_{it} + r^{\mathrm{U}}_{it},\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.2c}$$

$$g_{i(t+1)} - g_{it} \leq \mathrm{RU}_i,\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R,\ \forall t = 1, \ldots, T-1, \tag{2.2d}$$

$$g_{it} - g_{i(t+1)} \leq \mathrm{RD}_i,\ \forall i \in \mathcal{I} \setminus \mathcal{I}_R,\ \forall t = 1, \ldots, T-1, \tag{2.2e}$$

$$-C_{mn} \leq \sum_{b=1}^{B} K^b_{mn} \left(\sum_{i \in \mathcal{G}_b} g_{it} - D_{bt} \right) \leq C_{mn},\ \forall (m,n) \in \mathcal{L},\ \forall t \in \mathcal{T}, \tag{2.2f}$$

$$0 \leq g_{jt} \leq \xi_{jt},\ \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.2g}$$

where $\mathcal{B} = \{1, \ldots, B\}$ is the set of buses, $\mathcal{G}_b$ is the set of generators at bus b, $\mathcal{L}$ is the set of transmission lines linking two buses, C_{mn} is the capacity of the transmission line linking bus m and bus n, D_{bt} is the load at bus b in time period t, $f_i(\cdot)$ is the fuel cost function of generator i, K^b_{mn} is the line flow distribution factor for transmission line linking bus m and bus n due to the net injection at bus b, RU_i (resp. RD_i) is the ramp-up (resp. ramp-down) limit for thermal generator i, variable g_{it} is the actual amount of electricity generated by generator i in time period t.

In formulation (2.1)–(2.2) described above, we seek to minimize the expected total costs including UC, reserve, and fuel costs. In this work, we approximate the fuel cost function $f_i(g_{it})$ by using a piecewise linear approximation. Constraints (2.1b) (respectively, (2.1c)) describe the thermal generator start-up (respectively, shut-down) operations, constraints (2.1d) (respectively, (2.1e)) describe the thermal generator minimum-up time (respectively, minimum-down time) restrictions, and constraints (2.1f) (respectively, (2.1g)) describe bounds of upward (respectively, downward) reserve amounts and scheduled generation amounts. Meanwhile, constraints (2.2b) describe the system-wide balance between generation and load amounts, constraints (2.2c) describe bounds of actual generation amounts, constraints (2.2d)–(2.2e) describe the ramp-rate limit restrictions of generation amounts, constraints (2.2f) describe transmission line capacity restrictions based on the dc approximation, and constraints (2.2g) describe the minimal and maximal amounts of renewable energy we can utilize. Based on formulation (2.2), the two-stage robust UC model can be formulated as

$$\min_{y,u,v,e,r} \sum_{t=1}^{T} \sum_{i \in \mathcal{I}} (\text{SU}_i u_{it} + \text{SD}_i v_{it} + \text{NL}_i y_{it} + c_i^{\text{U}} r_{it}^{\text{U}} + c_i^{\text{D}} r_{it}^{\text{D}}) + \max_{\xi \in \mathcal{U}} Q(e, r, \xi) \tag{2.3a}$$

$$\text{s.t. } (2.1\text{b})\text{–}(2.1\text{h}), \tag{2.3b}$$

where $\mathcal{U}$ represents the uncertainty set of ξ and the worst-case fuel costs $\max_{\xi \in \mathcal{U}} Q(e, r, \xi)$ are considered in the objective function (2.3a). In this work, we consider a polyhedral uncertainty set

$$\mathcal{U} = \Big\{ \xi : \ \xi_{jt}^{\text{L}} \le \xi_{jt} \le \xi_{jt}^{\text{U}}, \ \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.4a}$$

$$\pi_j^{\text{L}} \le \sum_{t=1}^{T} \xi_{jt} \le \pi_j^{\text{U}}, \ \forall j \in \mathcal{I}_R \Big\}, \tag{2.4b}$$

where constraints (2.4a) describe the lower and upper bounds of the renewable energy for each generator in each time unit, and constraints (2.4b) describe the lower and upper

bounds of the total renewable energy for each generator throughout the operational time intervals. We note that the polyhedral uncertainty set (2.4) is often applied in robust UC approaches (see, e.g., [75, 169]), and other cardinality-constrained (see, e.g., [19, 9]) or norm-constrained (see, e.g., [18]) polyhedral uncertainty sets can also be applied here.

In this work, we study a data-driven approach to address the challenge of ambiguous probability distribution $\mathbb{P}_\xi$. To this end, we consider descriptive statistics that can be conveniently inferred by historical data, including bounds, mean, and MAD-median of ξ. We then construct an ambiguity set containing all probability distributions that are consistent with these statistics. Mathematically, we define the ambiguity set $\mathcal{D}$ as

$$\mathcal{D} := \Big\{\mathbb{P}_\xi : \mu^{\mathrm{L}}_{jt} \le \int_{\mathcal{U}} \xi_{jt} d\mathbb{P}_\xi \le \mu^{\mathrm{U}}_{jt},\ \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.5a}$$

$$\int_{\mathcal{U}} |\xi_{jt} - \nu_{jt}| d\mathbb{P}_\xi \le \sigma_{jt},\ \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, \tag{2.5b}$$

$$\int_{\mathcal{U}} d\mathbb{P}_\xi = 1\Big\}. \tag{2.5c}$$

$\mathcal{D}$ incorporates all probability distributions $\mathbb{P}_\xi$ that have (i) mean value within interval $[\mu^{\mathrm{L}}_{jt}, \mu^{\mathrm{U}}_{jt}]$(see constraints (2.5a)), (ii) MAD-median within interval $[0, \sigma_{jt}]$, where ν_{jt} represents the median of ξ_{jt} (see constraints (2.5b)), and (iii) support on set $\mathcal{U}$ (see constraint (2.5c)). The MAD-median is a robust statistic of the variability of ξ_{jt} and less sensitive to outliers as compared to the standard deviation when the data is inadequate. The MAD-median also keeps the formulation linear. In practice, we can construct intervals $[\mu^{\mathrm{L}}_{jt}, \mu^{\mathrm{U}}_{jt}]$ and $[0, \sigma_{jt}]$ based on historical data. For example, these two intervals can be obtained based on the 95% confidence intervals (CIs) of $\mathbb{E}_{\mathbb{P}_\xi}[\xi_{jt}]$ and $\mathbb{E}_{\mathbb{P}_\xi}[|\xi_{jt}-\nu_{jt}|]$, respectively. We give a simple numerical example for obtaining 95% CIs of $\mathbb{E}_{\mathbb{P}_\xi}[\xi_{jt}]$ and $\mathbb{E}_{\mathbb{P}_\xi}[|\xi_{jt} - \nu_{jt}|]$ as follows.

Example. Suppose that $j = 1$, $t = 1$, and the samples of ξ_{11} are $\{73, 96, 83, 67, 75, 50\}$. Then, the sample mean and (corrected) sample standard deviation of ξ_{11} are 74 and 15.44, respectively. Hence, the 95% two-sided CI for $\mathbb{E}_{\mathbb{P}_\xi}[\xi_{11}]$ has a lower bound $74 - t_{0.025}(5) \times 15.44/\sqrt{6} = 57.80$ and an upper bound $74 + t_{0.025}(5) \times 15.44/\sqrt{6} = 90.20$, i.e., the two-sided 95% CI for $\mathbb{E}_{\mathbb{P}_\xi}[\xi_{11}]$ is $[57.80, 90.20]$. Note that $t_{0.025}(5) = 2.57$ represents the t-distribution critical value with 5 degrees of freedom and upper-tail probability 0.025. Likewise, the one-sided 95% CI for $\mathbb{E}_{\mathbb{P}_\xi}[|\xi_{11} - \nu_{11}|]$ can be obtained as $[0, 18.97]$.

Note that these CIs can be estimated based on marginal data, i.e., the historical data separately collected for each renewable resource. This feature is particularly desirable when the data are unsynchronous, or when the correlations among different renewable resources are unclear or not perfectly estimated. In addition, these CIs can often be obtained based on a much smaller amount of historical data than those needed to accurately estimate

$\mathbb{P}_\xi$. We propose the following data-driven UC model as an alternative of the stochastic and robust UC approaches:

$$\min_{y,u,v,e,r} \sum_{t=1}^{T} \sum_{i\in\mathcal{I}} (\mathrm{SU}_i u_{it} + \mathrm{SD}_i v_{it} + \mathrm{NL}_i y_{it} + c_i^{\mathrm{U}} r_{it}^{\mathrm{U}} + c_i^{\mathrm{D}} r_{it}^{\mathrm{D}}) + \sup_{\mathbb{P}_\xi\in\mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)] \quad (2.6a)$$

$$\text{s.t. (2.1b)–(2.1h).} \quad (2.6b)$$

In the objective function (2.6a), we consider all possible expected total costs with respect to plausible $\mathbb{P}_\xi$ in the ambiguity set, i.e., $\{\mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)] : \mathbb{P}_\xi \in \mathcal{D}\}$, and select the largest one (hence the supremum notation). It follows that formulation (2.6a)–(2.6b) offers a guarantee on the expected total costs because $\mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)] \leq \sup_{\mathbb{P}_\xi\in\mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$ if $\mathcal{D}$ contains the true probability distribution. Meanwhile, formulation (2.6a)–(2.6b) is less conservative than the robust UC approaches, which is formalized in the following observation.

Observation 2.2.1. $\sup_{\mathbb{P}_\xi\in\mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)] \leq \max_{\xi\in\mathcal{U}} Q(e,r,\xi)$. *Accordingly, the optimal objective value of the data-driven UC model* (2.6) *is less than or equal to that of the robust UC model* (2.3).

Proof. We prove that

$$\max_{\xi\in\mathcal{U}} Q(e,r,\xi) = \sup_{\mathbb{P}_\xi\in\mathcal{D}'} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)],$$

where $\mathcal{D}'$ consists of all probability distributions supported on $\mathcal{U}$, i.e.,

$$\mathcal{D}' = \left\{ \mathbb{P}_\xi : \int_{\mathcal{U}} d\mathbb{P}_\xi = 1 \right\}.$$

If this claim holds, then the conclusion follows because $\mathcal{D}'$ is a relaxation of $\mathcal{D}$ and so $\sup_{\mathbb{P}_\xi\in\mathcal{D}'} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)] \leq \sup_{\mathbb{P}_\xi\in\mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$. To see this claim, we first note that $\max_{\xi\in\mathcal{U}} Q(e,r,\xi) \geq \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$ for all distributions $\mathbb{P}_\xi$ supported on $\mathcal{U}$, and it follows that $\max_{\xi\in\mathcal{U}} Q(e,r,\xi) \geq \sup_{\mathbb{P}_\xi\in\mathcal{D}'} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$. Second, let $\xi^* \in \mathcal{U}$ represent the worst-case scenario, i.e., $\max_{\xi\in\mathcal{U}} Q(e,r,\xi) = Q(e,r,\xi^*)$. Then the Dirac measure supported on singleton $\{\xi^*\}$ belongs to $\mathcal{D}'$, i.e., $1(\{\xi^*\}) \in \mathcal{D}'$. It follows that $\max_{\xi\in\mathcal{U}} Q(e,r,\xi) = \mathbb{E}_{1(\{\xi^*\})}[Q(e,r,\xi)]$ and so $\max_{\xi\in\mathcal{U}} Q(e,r,\xi) \leq \sup_{\mathbb{P}_\xi\in\mathcal{D}'} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$. The proof is completed. □

2.3 Solution Methodology

In this section, we derive a reformulation of model (2.6a)–(2.6b) in Section 2.3.1 and a solution algorithm based on the generalized linear decision rule (GLDR) in Section 2.3.2.

2.3.1 Reformulation

We first provide a reformulation of the worst-case expected cost $\sup_{\mathbb{P}_\xi \in \mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e,r,\xi)]$ as follows:

$$
\begin{aligned}
& \sup_{\mathbb{P}_\xi \in \mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}\left[Q(e,r,\xi)\right] \\
= \; & \max_{\mathbb{P}_{\xi,\omega}} \int_{\mathcal{V}} Q(e,r,\xi) d\mathbb{P}_{\xi,\omega} && \text{(2.7a)} \\
& \text{s.t. } \mu^{\mathrm{L}}_{jt} \le \int_{\mathcal{V}} \xi_{jt} d\mathbb{P}_{\xi,\omega} \le \mu^{\mathrm{U}}_{jt}, \; \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, && \text{(2.7b)} \\
& \int_{\mathcal{V}} \omega_{jt} d\mathbb{P}_{\xi,\omega} \le \sigma_{jt}, \; \forall j \in \mathcal{I}_R, \forall t \in \mathcal{T}, && \text{(2.7c)} \\
& \int_{\mathcal{V}} d\mathbb{P}_{\xi,\omega} = 1, && \text{(2.7d)}
\end{aligned}
$$

where we introduce a new random variable ω_{jt} to represent the mean absolute deviation of ξ_{jt}, i.e., $|\xi_{jt} - \nu_{jt}|$ and rewrite constraints (2.5b) in the form of (2.7c). Accordingly, we lift probability distribution $\mathbb{P}_\xi$ and uncertainty set $\mathcal{U}$ to joint probability distribution $\mathbb{P}_{\xi,\omega}$ and joint uncertainty set $\mathcal{V}$ of (ξ,ω), respectively, where

$$
\mathcal{V} = \Big\{(\xi,\omega) : \xi \in \mathcal{U}, \; \omega_{jt} \ge \xi_{jt} - \nu_{jt}, \; \omega_{jt} \ge \nu_{jt} - \xi_{jt}, \forall j \in \mathcal{I}_R, \; \forall t \in \mathcal{T}\Big\}.
$$

Note that the lifting transformation (see [151, 20]) linearizes constraints (2.5b) and facilitates existing techniques on (functional) linear optimization and the GLDR. Second, we take the dual of formulation (2.7a)–(2.7d) as follows:

$$
\begin{aligned}
\min_{p,q,r} \; & \left(\mu^{\mathrm{U}}\right)^{\mathsf{T}} p^{\mathrm{U}} - \left(\mu^{\mathrm{L}}\right)^{\mathsf{T}} p^{\mathrm{L}} + \sigma^{\mathsf{T}} q + r && \text{(2.8a)} \\
\text{s.t. } & \left(p^{\mathrm{U}} - p^{\mathrm{L}}\right)^{\mathsf{T}} \xi + q^{\mathsf{T}} \omega + r \ge Q(e,r,\xi), \; \forall (\xi,\omega) \in \mathcal{V}, && \text{(2.8b)} \\
& p^{\mathrm{U}}, p^{\mathrm{L}} \ge 0, && \text{(2.8c)}
\end{aligned}
$$

where dual variables $p^{\mathrm{U/L}}$, q, and r are associated with primal constraints (2.7b)–(2.7d), respectively, and dual constraints (2.8b) are associated with primal variables $\mathbb{P}_{\xi,\omega}$. Meanwhile, we let $\mu^{\mathrm{U/L}}$, σ, $p^{\mathrm{U/L}}$, and q represent vectors $(\mu^{\mathrm{U/L}}_{jt} : j \in \mathcal{I}_R, t \in \mathcal{T})$, $(\sigma_{jt} : j \in$

$\mathcal{I}_R, t \in \mathcal{T}$), $(p_{jt}^{\mathrm{U/L}} : j \in \mathcal{I}_R, t \in \mathcal{T})$, and $(q_{jt} : j \in \mathcal{I}_R, t \in \mathcal{T})$, respectively. This dual formulation can be obtained via standard Lagrangian dual approach [126]. It can be shown that strong duality holds between the primal (2.7) and dual formulations (2.8), and so we can substitute $\sup_{\mathbb{P}_\xi \in \mathcal{D}} \mathbb{E}_{\mathbb{P}_\xi}[Q(e, r, \xi)]$ with the dual formulation in model (2.6) to obtain a monolithic formulation:

$$\begin{aligned} \min \ & \sum_{t=1}^{T} \sum_{i \in \mathcal{I}} (\mathrm{SU}_i u_{it} + \mathrm{SD}_i v_{it} + \mathrm{NL}_i y_{it} + c_i^{\mathrm{U}} r_{it}^{\mathrm{U}} + c_i^{\mathrm{D}} r_{it}^{\mathrm{D}}) \\ & + \left(\mu^{\mathrm{U}}\right)^{\mathsf{T}} p^{\mathrm{U}} - \left(\mu^{\mathrm{L}}\right)^{\mathsf{T}} p^{\mathrm{L}} + \sigma^{\mathsf{T}} q + r \\ \text{s.t.} \ & \text{(2.1b)–(2.1h), (2.8b)–(2.8c).} \end{aligned} \tag{2.9}$$

However, formulation (2.9) is computationally intractable because of the semi-infinite constraints (2.8b). In fact, it has been proved to be NP-hard (see, e.g., [17]). In this work, we apply the GLDR to simplify constraints (2.8b) and conservatively approximate formulation (2.9) by using a mixed-integer linear program (MILP).

2.3.2 Solution Algorithm based on the GLDR

To reformulate constraints (2.8b) under the GLDR, we first rewrite formulation (2.2) and uncertainty set $\mathcal{V}$ in abstract forms:

$$Q(x, \xi) \ = \ \min_{z} \ c^{\mathsf{T}} z \tag{2.10a}$$

$$\text{s.t.} \ a_{m+1}^{\mathsf{T}} z = d_{m+1}, \tag{2.10b}$$

$$a_i^{\mathsf{T}} z + e_i^{\mathsf{T}} x + h_i^{\mathsf{T}} \xi \geq d_i, \ \forall i = 1, \ldots, m, \tag{2.10c}$$

$$\mathcal{V} \ = \ \{(\xi, \omega) : \ \Gamma \xi + \Psi \omega \geq \eta\}, \tag{2.11}$$

where we let x represent the first-stage decision variables (e, r) and z represent the second-stage decision variables including g and auxiliary variables for the piecewise linear approximation of $f_i(g_{it})$. Meanwhile, constraint (2.10b) represents equality constraints (2.2b), and constraints (2.10c) subsume all inequality constraints (2.2c)–(2.2g). We also let matrices Γ and Ψ, and vectors c, d_i, a_i, e_i, h_i, and η represent given parameters. Based on formulation (2.10), the GLDR restricts the search space of variables z to affine functions of ξ and ω, i.e.,

$$z \ = \ z^0 + G_\xi \xi + G_\omega \omega, \tag{2.12}$$

where vector z^0 represents a benchmark of z, matrix G_ξ represents a linear dependence on vector ξ, and G_ω represents a linear dependence on vector ω. Note that ω_{jt} represents $|\xi_{jt} - \nu_{jt}|$, and so (2.12) designates variables z as an affine function of the (realized) renewable energy ξ and its deviation from median ν. Also, as the GLDR restricts the feasible region of the second-stage problem (2.10), it leads to a conservative approximation of the data-driven UC model. Second, we substitute (2.12) in formulation (2.10) to obtain

$$
\begin{aligned}
&Q_{\text{GLDR}}(x,\xi) \equiv c^{\mathsf{T}} z^0 + \left(G_\xi^{\mathsf{T}} c\right)^{\mathsf{T}} \xi + \left(G_\omega^{\mathsf{T}} c\right)^{\mathsf{T}} \omega, \\
&\left(G_\xi^{\mathsf{T}} a_{m+1}\right)^{\mathsf{T}} \xi + \left(G_\omega^{\mathsf{T}} a_{m+1}\right)^{\mathsf{T}} \omega = d_{m+1} - a_{m+1}^{\mathsf{T}} z^0,\ \forall (\xi,\omega) \in \mathcal{V}, && \text{(2.13a)} \\
&\left(G_\xi^{\mathsf{T}} a_i + h_i\right)^{\mathsf{T}} \xi + \left(G_\omega^{\mathsf{T}} a_i\right)^{\mathsf{T}} \omega \geq d_i - a_i^{\mathsf{T}} z^0 - e_i^{\mathsf{T}} x,\ \forall (\xi,\omega) \in \mathcal{V},\ \ \forall i = 1,\ldots,m, && \text{(2.13b)}
\end{aligned}
$$

where $Q_{\text{GLDR}}(x,\xi)$ represents the fuel costs under the GLDR, and constraints (2.13a)–(2.13b) reformulates constraints (2.10b)–(2.10c), respectively. It follows that constraints (2.8b) can be rewritten as

$$
\left(p^{\text{U}} - p^{\text{L}} - G_\xi^{\mathsf{T}} c\right)^{\mathsf{T}} \xi + \left(q - G_\omega^{\mathsf{T}} c\right)^{\mathsf{T}} \omega \ \geq\ c^{\mathsf{T}} z^0 - r,\ \forall (\xi,\omega) \in \mathcal{V}. \qquad \text{(2.13c)}
$$

(2.13a)–(2.13b).

Third, we note that constraints (2.13a)–(2.13c) are semi-infinite. To address the potential challenges in computation, we apply the techniques in robust optimization to rewrite them as finite linear constraints. For constraints (2.13a), as equality holds for arbitrary $(\xi,\omega) \in \mathcal{V}$, it is equivalent that

$$
G_\xi^{\mathsf{T}} a_{m+1} = 0,\ \ G_\omega^{\mathsf{T}} a_{m+1} = 0,\ \ a_{m+1}^{\mathsf{T}} z^0 = d_{m+1}. \qquad \text{(2.14a)}
$$

Meanwhile, constraints (2.13c) requires that a linear function of variables ξ and ω to be no less than $c^{\mathsf{T}} z^0 - r$ on polyhedron $\mathcal{V}$. It is equivalent that

$$
\min_{(\xi,\omega)\in\mathcal{V}} \left\{ \left(p^{\text{U}} - p^{\text{L}} - G_\xi^{\mathsf{T}} c\right)^{\mathsf{T}} \xi + \left(q - G_\omega^{\mathsf{T}} c\right)^{\mathsf{T}} \omega \right\} \geq\ c^{\mathsf{T}} z^0 - r. \qquad \text{(2.14b)}
$$

We then take the dual of the linear program on the left-hand side of (2.14b) to obtain

$$
\begin{aligned}
&\Psi^{\mathsf{T}} \chi_0 \leq q - G_\omega^{\mathsf{T}} c,\ \ \Gamma^{\mathsf{T}} \chi_0 \leq p^{\text{U}} - p^{\text{L}} - G_\xi^{\mathsf{T}} c, && \text{(2.14c)} \\
&\eta^{\mathsf{T}} \chi_0 \geq c^{\mathsf{T}} z^0 - r,\ \ \chi_0 \geq 0, && \text{(2.14d)}
\end{aligned}
$$

where dual variables χ_0 are associated with primal constraints (2.11) in the abstract definition of $\mathcal{V}$, and dual constraints (2.14c) are associated with primal variables ξ and ω, respectively. Similarly, constraints (2.13b) can be reformulated as

$$\Gamma^{\mathsf{T}}\chi_i \le G_\xi^{\mathsf{T}} a_i + h_i, \;\; \Psi^{\mathsf{T}}\chi_i \le G_\omega^{\mathsf{T}} a_i, \tag{2.14e}$$

$$\eta^{\mathsf{T}}\chi_i \ge d_i - a_i^{\mathsf{T}} z^0 - e_i^{\mathsf{T}} x, \;\; \chi_i \ge 0, \;\; \forall i = 1, \ldots, m. \tag{2.14f}$$

Therefore, constraints (2.8b) are represented as (finite and linear) constraints (2.14a), (2.14c)–(2.14f) and so the data-driven UC model (2.6) under the GLDR is equivalent to the following MILP:

$$\begin{aligned} \min \;\; & \sum_{t=1}^{T} \sum_{i \in \mathcal{I}} \left(\mathrm{SU}_i u_{it} + \mathrm{SD}_i v_{it} + \mathrm{NL}_i y_{it} + c_i^{\mathrm{U}} r_{it}^{\mathrm{U}} + c_i^{\mathrm{D}} r_{it}^{\mathrm{D}} \right) \\ & + \left(\mu^{\mathrm{U}}\right)^{\mathsf{T}} p^{\mathrm{U}} - \left(\mu^{\mathrm{L}}\right)^{\mathsf{T}} p^{\mathrm{L}} + \sigma^{\mathsf{T}} q + r \\ \text{s.t.} \;\; & \text{(2.1b)–(2.1h), (2.8c), (2.14a), (2.14c)–(2.14f).} \end{aligned} \tag{2.15}$$

Note that we optimize the choices of (z^0, G_ξ, G_ω) jointly with the decision variables (y, u, v, e, r). In other words, we search for an *optimal* GLDR to help make better UC and reserve procurement. Note that the GLDR (2.12) implies that, for all $i \in \mathcal{I}$ and $s \in \mathcal{T}$, generation amount g_{is} depends on vectors $(\xi_{jt} : j \in \mathcal{I}_R, t \in \mathcal{T})$ and $(\omega_{jt} : j \in \mathcal{I}_R, t \in \mathcal{T})$. To respect the nonanticipativity, i.e., making decisions in each time unit only based on the information available up to that time unit, we need to restrict the GLDR by fixing all entries of G_ξ that link g_{is} with ξ_{jt} to be zeros if $t \ge s+1$. Likewise, we fix all entries of G_ω that link g_{is} with ω_{jt} to be zeros if $t \ge s+1$. Furthermore, we can make the GLDR more flexible and sparse by designating g_{is} to depend only on the recent realizations of ξ_{jt} and ω_{jt} for $t \in \{s-L+1, \ldots, s-1, s\}$, where L represents the time length the GLDR dates back. For example, if $L = 1$, then g_{is} depends only on the most recent realizations ξ_{js} and ω_{js}. Also, if $L = s$, then g_{is} depends on all historical realizations $(\xi_{jt} : t = 1, \ldots, s)$ and $(\omega_{jt} : t = 1, \ldots, s)$. Accordingly, we fix all entries of G_ξ and G_ω that link g_{is} with ξ_{jt} and ω_{jt} to be zero if $t \le s - L$. We summarize all linear constraints that fix corresponding entries of G_ξ and G_ω to be zero in the following abstract form

$$L\left(G_\xi, G_\omega\right) = 0, \tag{2.16}$$

and we add constraints (2.16) to the MILP (2.15) for the GLDR nonanticipativity and dependency. We note that the MILP reformulation (2.15) facilitates efficient commercial

software like CPLEX. Additionally, (2.15) possesses the same number of binary variables as the deterministic UC formulation. As a consequence, the data-driven UC model (2.6) under the GLDR has a similar computational complexity as the deterministic UC approach. There are two ways of further speeding up the solution process:

1. We can solve the MILP (2.15) in a parallel computing environment [156] to accelerate the branch-and-cut algorithm.

2. We can incorporate constraints (2.14c)–(2.14f) by using a row generation algorithm. That is, we solve (2.15) with these constraints relaxed at the beginning. Then, we introduce them back to the relaxed formulation only if they are violated. This row generation algorithm can be implemented by using the LazyConstraintCallback function in CPLEX.

GLDR extends the more classical LDR

$$z \;=\; z^0 + G_\xi \xi, \tag{2.17}$$

which designates variables z as an affine function of ξ (but not ω). LDR is proposed in [13] and can be applied on robust UC model (2.3a)–(2.3b), which results in a MILP. The derivation is similar to that of the data-driven UC model with the GLDR, and so is omitted here. Finally, we note that GLDR has a close connection with the extended affine policy (EAP) proposed in [33]. Both GLDR and EAP designate that the recourse variables affinely depend on the realized random variables and some auxiliary variables. The auxiliary variables in EAP often stem from splitting the random variables as positive and negative parts along different directions, while the auxiliary variables in GLDR stem from lifting the ambiguity set (see (2.7)).

2.4 Case Study

To test the proposed data-driven UC model with GLDR and the solution algorithm, we conduct computational case studies based on a 60-bus system and an IEEE 118-bus system. We report the setup and results of the 60-bus system in Section 2.4.1, and those of the 118-bus system in Section 2.4.2. All experiments are performed on a Linux server with Intel Xeon Quad Core 2.93 GHz, 2GB memory, and CPLEX 12.3.

2.4.1 60-Bus System

Experiment Setup: The 60-bus system duplicates the IEEE 30-bus system containing 30 nodes and 41 transmission lines (see http://www.ee.washington.edu/research/pstca/ for the data set). First, we place 12 thermal generators in this system (see Tables 2.1–2.2 for the generator characteristics). In this case study, we approximate the fuel cost function of the thermal generators by linear functions and denote the slopes as c_i^{G} in Table 2.2. Meanwhile, we put 3 wind farms and 3 solar farms in the system as renewable energy resources. The 3 wind farms are based on the time series of 3 onshore sites from the National Renewable Energy Laboratory (NREL) Eastern Wind Dataset [1] and the 3 solar farms are based on the times series of 3 sites from the NREL National Solar Radiation Database [2].

Table 2.1: Generator characteristics - part I

Generator Index	U_i	L_i	RU_i	RD_i	MU_i	MD_i
1,2	60	5	30	30	1	1
3,4	100	10	50	50	1	1
5,6	120	10	60	60	1	1
7,8	350	50	150	150	6	6
9,10	300	40	120	120	7	7
11,12	400	60	180	180	7	7

Second, we characterize uncertainty set $\mathcal{U}$ in (2.4) and ambiguity set $\mathcal{D}$ in (2.5) based on the historical data from NREL. For each day we test, we collect the historical data of renewable energy in the previous week, and in the week before and after this day in the previous two years. For example, if we test Feb 1 in year 2006, then we collect data during Jan 25-31 in 2006, Jan 25-Feb 7 in 2005, and Jan 25-Feb 7 in 2004. Note that this can help reflect the seasonal patterns of renewable energy. For uncertainty set $\mathcal{U}$, for each wind/solar farm, we set ξ_{jt}^{L} and ξ_{jt}^{U} as the 0.05- and 0.95-percentile of the hourly data, respectively. Likewise, we set π_j^{L} and π_j^{U} as the 0.05- and 0.95-percentile of the sums of the hourly data within a day, respectively. For ambiguity set $\mathcal{D}$, for each wind/solar farm, we set μ_{jt}^{L} and μ_{jt}^{U} as the lower and upper limits, respectively, of the 95% CIs of $\mathbb{E}_{\mathbb{P}_\xi}[\xi_{jt}]$ based on the hourly data. Likewise, we set ν_{jt} as the median of the hourly data and σ_{jt} as the upper limit of the 95% one-sided CIs of $\mathbb{E}_{\mathbb{P}_\xi}|\xi_{jt} - \nu_{jt}|$.

Third, we test the proposed approach under four different penetration levels of renewable energy. Using the medians of renewable energy from each site as benchmarks, we scale the mean and MAD-median values of the wind and solar farms to account for 5%, 10%, 15%, and 20% of the total electricity loads, respectively.

Fourth, we simulate to schedule day-ahead UC and reserve procurement of this power system for 4 weeks, each week picked from February, May, August, and November, respectively (i.e., 28 days in total). For each day in these 4 weeks, we solve the data-driven UC model with the GLDR (2.15) with $T = 24$ hours. As a benchmark, we also test the robust UC approach with uncertainty set $\mathcal{U}$ and LDR (2.17). Depending on which approach is applied, we denote the solution obtained from the data-driven approach as DD schedule and that from the robust approach as RO schedule. For simplicity, we set $L = 1$ in both data-driven and robust UC approaches. In addition, we compare both approaches to a perfect information (PI) model, where we assume that we can perfectly predict the renewable energy before deciding the day-ahead UC and reserves. Note that the PI model is a deterministic UC model with ξ replaced by their realizations in the day we test. Hence, the PI schedule is guaranteed to perform the best.

Table 2.2: Generator characteristics - part II

Generator Index	SU_i	SD_i	c_i^{G}	c_i^{U}	c_i^{D}
1,2	50	50	35	3	3
3,4	50	50	40	5	5
5,6	60	60	50	5	5
7,8	300	300	10	3	3
9,10	200	200	8	4	4
11,12	350	350	10	3	3

Finally, we evaluate the performance of DD schedule based on a post-optimization and out-of-sample simulation. We fix the first-stage decisions (i.e., (y, u, v, e, r)) to be the optimal solutions obtained from solving MILP (2.15). Then, we compute the actual cost of DD schedule by resolving the second-stage ED problem (2.2) based on the fixed first-stage decisions and the realized renewable energy. In other words, the GLDR helps find first-stage decisions but do not decide the actual electricity generation amounts g_{it} in this simulation. This is a more practical evaluation of DD schedule because the system operators can resolve the ED problem based on a more accurate forecast of the renewable energy shortly before dispatch. In this simulation, we evaluate the performance of DD schedule by computing the following values:

1. S_{DD}, the actual costs of DD schedule.
2. V_{PI}, the optimal objective value of the PI model.
3. $\text{Gap}_{\text{DD}} = (S_{\text{DD}} - V_{\text{PI}})/V_{\text{PI}} \times 100\%$, the relative gap between the PI and DD schedules.
4. R_{DD}, the reserve costs of DD schedule.

5. $\text{UT}_{\text{DD}} = \sum_{j\in\mathcal{I}_R} g_{jt} / \sum_{j\in\mathcal{I}_R} \xi_{jt} \times 100\%$, renewable energy utilization of DD schedule.

As a benchmark, we test RO schedule and similarly obtain S_{RO}, Gap_{RO}, R_{RO}, and UT_{RO}.

Table 2.3: Reserve cost comparison on the 60-bus system

Penetration Levels (%)	Average Reserve Cost ($)	
	R_{DD}	R_{RO}
5	4789.24	23.00
10	6987.90	45.99
15	8094.13	64.69
20	8557.62	76.79

Table 2.4: Comparisons on the out-of-sample performance on the 60-bus system

Penetration Levels (%)	Average Relative Gaps (%)		Average Renewable Utilization (%)	
	Gap_{DD}	Gap_{RO}	UT_{DD}	UT_{RO}
5	1.36	3.90	99.58	50.08
10	2.53	7.64	98.37	50.40
15	4.88	12.53	95.45	49.93
20	8.13	17.64	91.38	49.23

Experiment Results: First, we report the computational time for solving MILP (2.15). All instances can be solved within 10 mins, and on average, each instance takes less than 5 mins to achieve global optimality.

Second, we report the results of the out-of-sample simulation in Tables 2.3–2.4 and Figures 2.1–2.3. From Table 2.4 (see columns UT_{DD} and UT_{RO}) and Figure 2.1, we observe that DD schedules consistently outperform RO schedules in utilizing renewable energy. More precisely, the average UT_{DD} is consistently above 90%, while the average UT_{RO} is around 50%. Meanwhile, based on Table 2.4 (see columns Gap_{DD} and Gap_{RO}) and Figure 2.2, the average costs of DD schedules are lower than that of RO schedules. More precisely, the average Gap_{DD} is consistently less than 50% of the average Gap_{RO}. Furthermore, from Table 2.3 and Figure 2.3, we observe that DD schedules procure a larger amount of reserves as compared to RO schedules. These observations indicate that incorporating distributional information in the day-ahead scheduling can encourage procurement of system flexibility (e.g., reserves). With larger reserves, thermal units have wider ranges of adjusting their actual amounts of electricity generation according to random solar/wind power outputs. This further enhances the system capability of accommodating renewable energy and reducing average costs.

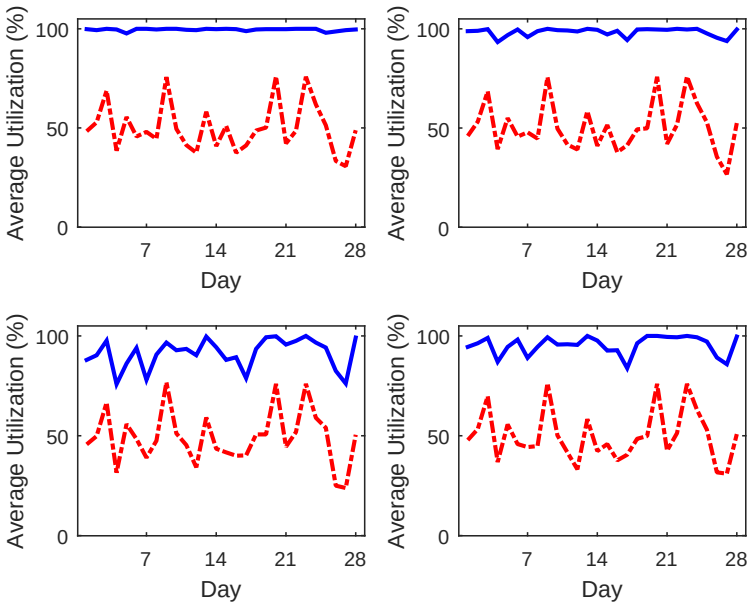


Figure 2.1: Renewable utilization comparison on the 60-bus system; Solid - DD Schedule, Dashed - RO Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

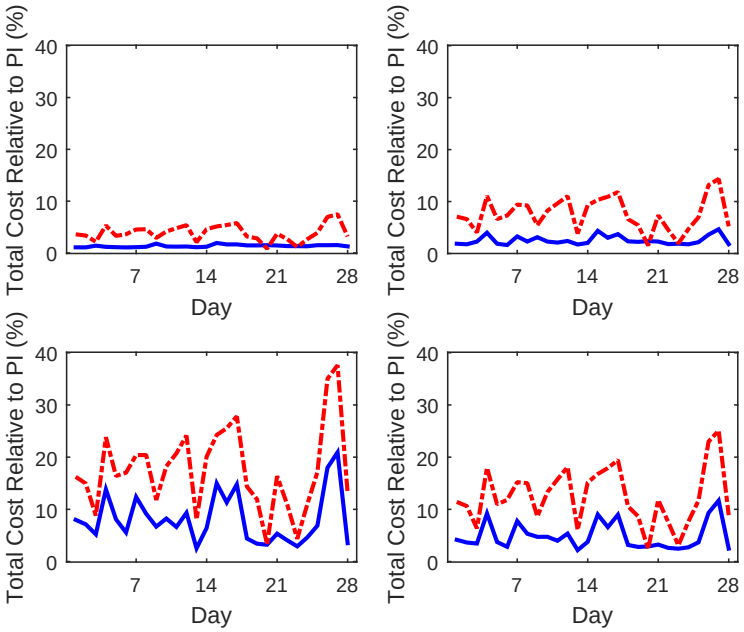


Figure 2.2: Total cost comparison on the 60-bus system; Solid - DD Schedule, Dashed - RO Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

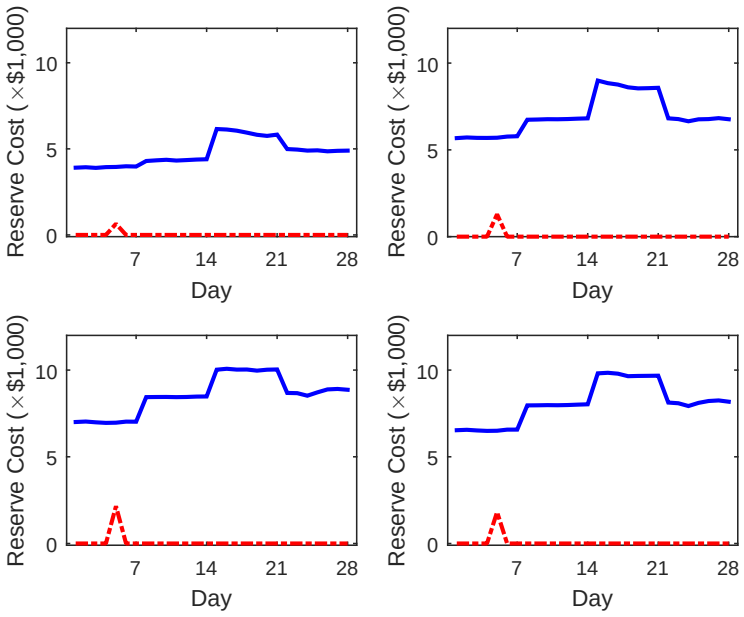


Figure 2.3: Reserve cost comparison on the 60-bus system; Solid - DD Schedule, Dashed - RO Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

2.4.2 118-Bus System

Experiment Setup: The IEEE 118-bus system contains 54 thermal units and 186 transmission lines (see http://www.ee.washington.edu/research/pstca/ for the data set). Similar to the 60-bus system, we place 3 wind farms and 3 solar farms in the system as renewable energy resources. We construct the uncertainty set $\mathcal{U}$ and ambiguity set $\mathcal{D}$ based on the NREL historical data [1, 2] as described in Section 2.4.1. Likewise, we simulate penetration levels from 5% to 20%, and schedule day-ahead UC and reserve procurement of this power system for 4 weeks, each week picked from February, May, August, and November, respectively (i.e., 28 days in total). In this case study, we compare the data-driven approach with robust and stochastic UC approaches. We denote the solution obtained from the stochastic UC approach as ST schedule.

Table 2.5: Comparisons on the reserve cost and CPU seconds on the 118-bus system

Penetration Levels (%)	Average Reserve Costs ($\times 5,000$\$)			CPU Time (s)		
	R_{DD}	R_{RO}	R_{ST}	DD	RO	ST
5	1.85	0.29	1.19	2577.43	3681.79	4586.43
10	3.60	0.44	2.31	3966.32	3966.32	6962.39
15	4.77	0.72	3.22	3958.21	7528.75	6466.14
20	5.58	0.93	4.06	3840.50	6700.86	6316.96

Table 2.6: Comparisons on the out-of-sample performance on the 118-bus system

Penetration Levels (%)	Average Relative Gaps (%)			Average Renewable Utilization (%)		
	Gap_{DD}	Gap_{RO}	Gap_{ST}	UT_{DD}	UT_{RO}	UT_{ST}
5	1.64	3.90	1.89	96.76	63.11	97.40
10	2.55	7.64	2.93	95.26	59.40	95.30
15	3.10	12.53	3.93	94.53	60.54	92.95
20	4.03	17.64	4.87	92.23	60.07	90.43

Experiment Results: First, we report the computational time for solving MILP (2.15) in Table 2.5. All instances can be solved within 3 hours, and on average, each instance takes around 1 hour to achieve global optimality (see column CPU Time–DD).

Second, we report the results of the out-of-sample simulation in Tables 2.5–2.6 and Figures 2.4–2.6. From Table 2.6 (see columns Average Renewable Utilization) and Figure 2.4, we observe that DD schedules utilize over 90% of renewable energy throughout all penetration levels. Meanwhile, the utilization of DD schedules are in general level with that of ST schedules, while both of them significantly exceed the utilization of RO schedules in most days. Furthermore, based on Table 2.6 (see columns Average Relative Gaps) and Figure 2.5, the average costs of DD schedules are lower than those of RO and ST schedules. More precisely, the average Gap_{DD} is consistently less than 50% of the average Gap_{RO} and less than 90% of the average Gap_{ST}. Together with Table 2.5 (see column Average Reserve Costs) and Figure 2.6, this confirms our observation from the 60-bus system. That is, incorporating distributional information in the day-ahead scheduling can encourage procurement of system flexibility (e.g., reserves), which reduces the average cost in out-of-sample simulations.

2.5 Conclusion

In this chapter, we proposed a data-driven approach for scheduling day-ahead UC and reserve procurement in power systems with renewable energy integration. In particular, we incorporated descriptive statistics (e.g., mean and MAD-median) of the renewable energy based on historical data, and derived a conservative approximation of the data-driven UC model by using the GLDR. We tested the proposed approach in two out-of-sample simulations based on publicly available data from the NREL. The experiment results indicated that the proposed data-driven approach enhances the system flexibility and capability of accommodating renewable energy, and a real-time economic re-dispatch can help fully utilize the system flexibility.

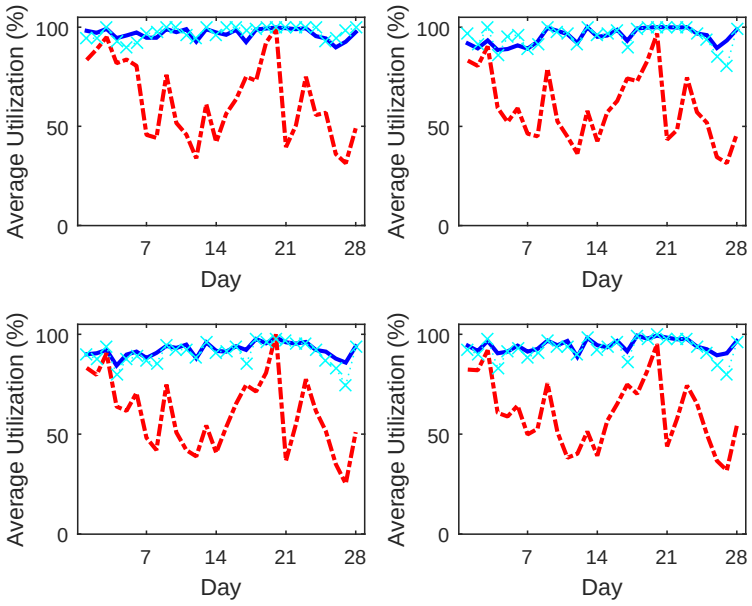


Figure 2.4: Renewable utilization comparison on the 118-bus system; Solid - DD Schedule, Dashed - RO Schedule, Cross - ST Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

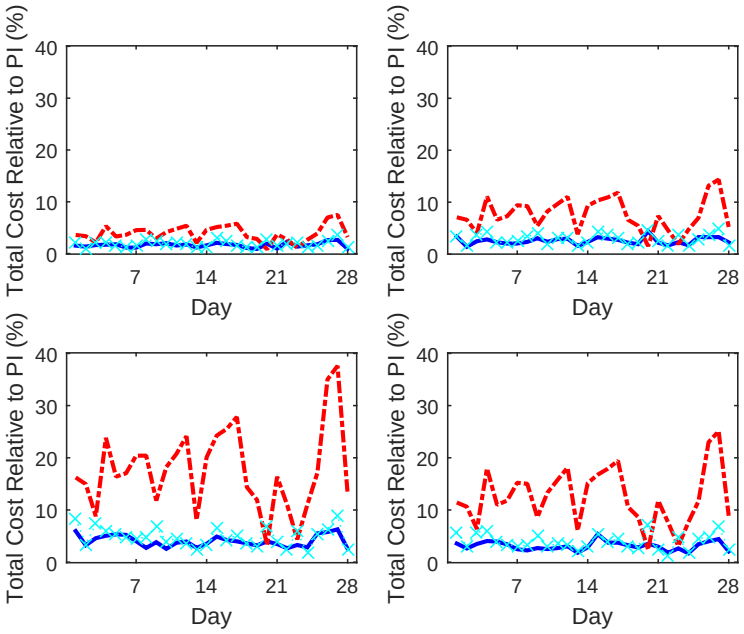


Figure 2.5: Total cost comparison on the 118-bus system; Solid - DD Schedule, Dashed - RO Schedule, Cross - ST Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

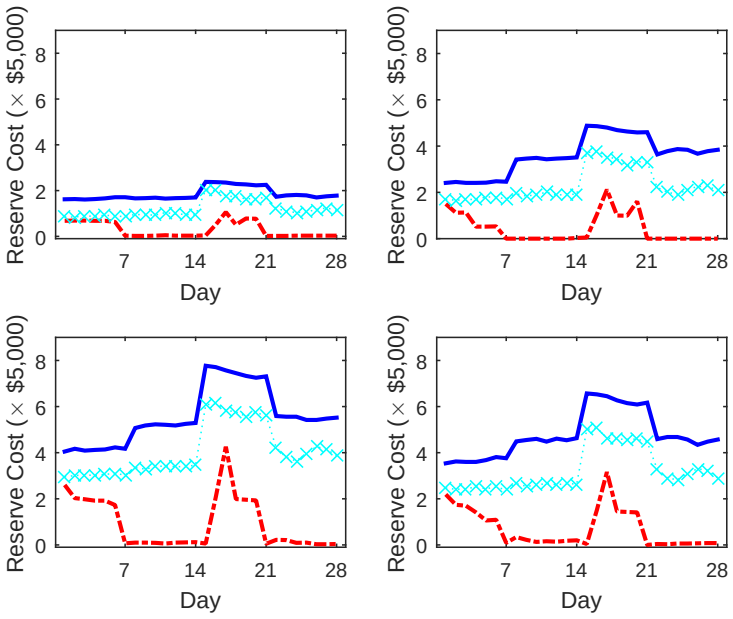


Figure 2.6: Reserve cost comparison on the 118-bus system; Solid - DD Schedule, Dashed - RO Schedule, Cross - ST Schedule; Penetration Levels - 5%, 10%, 15%, 20% Clockwise

2.6 Nomenclature

A. Sets and Indices

$\mathcal{B} = \{1, \ldots, B\}$ Set of buses.

$\mathcal{G}_b$ Set of generators at bus b.

$\mathcal{I} = \{1, \ldots, I\}$ Set of all generators.

$\mathcal{I}_R \subseteq \mathcal{I}$ Set of generators that utilize renewable energy.

$\mathcal{L}$ Set of transmission lines linking two buses.

$\mathcal{T} = \{1, \ldots, T\}$ Set of operating time intervals.

$\mathcal{U}$ Support of the renewable energy.

$\mathcal{V}$ Extended support of the renewable energy.

i Index for generators.

j Index for generators that utilize renewable energy.

B. Parameters

C_{mn} Capacity of the transmission line linking bus m and bus n.

c_i^{U}/c_i^{D} Unit cost of upward/downward reserve amount of generator i.

D_{bt} Load at bus b in time period t.

$\mathcal{D}$ Ambiguity set of $\mathbb{P}_\xi$.

$f_i(\cdot)$ Fuel cost function of generator i.

K_{mn}^b Line flow distribution factor for transmission line linking bus m and bus n due to the net injection at bus b.

L_i Minimal power output if generator i is on.

MU_i Minimum-up time for thermal generator i.

MD_i Minimum-down time for thermal generator i.

NL_i No-load cost of thermal generator i.

$\mathbb{P}_\xi$ Probability distribution of ξ.

$\mathbb{P}_{\xi,\omega}$ Joint probability distribution of ξ and ω.

RU_i Ramp-up limit for thermal generator i.

RD_i Ramp-down limit for thermal generator i.

SU_i Start-up cost for generator i.

SD_i Shut-down cost for generator i.

U_i Maximal power output if generator i is on.

ξ_{jt} Available renewable energy of generator j in time period t.

μ_{jt}^{L} Lower bound of the mean value of ξ_{jt}.

μ_{jt}^{U} Upper bound of the mean value of ξ_{jt}.

ν_{jt} Median of ξ_{jt}.

σ_{jt} Upper bound of the mean absolute deviation of ξ_{jt}.

ω_{jt} Auxiliary random variables for the mean absolute deviation from median of ξ_{jt}.

C. Decision Variables

e_{it} Scheduled amount of electricity generated by generator i in time period t.

g_{it} Actual amount of electricity generated by generator i in time period t.

r_{it}^{U}/r_{it}^{D} Upward/downward reserve amount of generator i in time period t.

u_{it} Binary decision variable to indicate if generator i is started up at the beginning of time period t.

v_{it} Binary decision variable to indicate if generator i is shut down at the beginning of time period t.

y_{it} Binary decision variable to indicate if generator i is on during time period t.

CHAPTER 3

Distributionally Robust Chance-Constrained Optimal Power Flow with a Wasserstein-Moment Ambiguity Set

3.1 Introduction

Optimal power flow (OPF) problem is one of the most important optimization problems in power system operations. It aims to minimize power system costs including, e.g., generation cost or power loss, while satisfying network and physical constraints on generators and transmission lines. During the last decade, renewable energy (e.g., wind and solar) has been steadily penetrating into power systems all over the world [110]. For example, the U.S. Department of Energy has analyzed a scenario, in which the wind power contributes 20% of the total electricity utility by 2030 [91]. However, the inherent intermittency and volatility of renewable energy introduce significant uncertainty into the OPF problem, rendering its safety constraints (e.g., transmission line capacity limits and reserve capacity limits) vulnerable.

Chance-constrained programming has been widely studied to hedge against uncertainties in the OPF problem (see e.g., [161, 162, 70, 141, 113, 21, 142, 88]). For a constraint subject to uncertainty, its corresponding chance constraint ensures that the constraint is satisfied with certain probability under a prescribed distribution of the uncertain parameters. The solution approaches of the chance-constrained program include scenario approximation [26], probabilistically robust methods [97], and analytical reformulations relying on specific forms of the probability distributions (see e.g., [113, 21]). However, scenario approximation is often computationally heavy and conservative in general, (classical) robust optimization is even more conservative than scenario approximation, and the analytical reformulations may result in unreliable solutions if the uncertain parameters do not follow the assumed distributions (see, e.g., [86]). Another basic challenge of chance constraints lies

in estimating the probability distribution of the renewable energy due to its non-stationary nature and/or a lack of historical data (see, e.g., [64]). To address these drawbacks, distributionally robust chance constraints (DRCC) were proposed (see, e.g., [37, 171]). In DRCC, the uncertain constraints are required to be satisfied with given probability under all distributions within an ambiguity set. The ambiguity set contains distributions characterized by certain characteristics of the (unknown) true distribution. The ambiguity set plays a crucial role in the DRCC model. Most existing literature focuses on the moment ambiguity set, i.e., a set of probability distributions that share the same moments of certain functions (see, e.g., [37, 52, 151, 171]). For example, [171] studied DRCC under a moment ambiguity set based on the mean and covariance matrix of the uncertain parameters. In addition, [86, 164] demonstrated that DRCC under moment ambiguity sets lead to improved OPF solutions.

More recently, probability discrepancy functions have been applied to construct ambiguity sets, including the Prokhorov metric [45], the ϕ-divergence [73, 67], and the Wasserstein metric [105, 152, 47]. The discrepancy ambiguity set is defined as a neighborhood surrounding a reference distribution, in which all candidate distributions are close to the reference one measured by the prescribed discrepancy. The degree of conservatism of discrepancy ambiguity sets can be controlled by adjusting the discrepancy upper bound to the reference distribution. If we set the bound to be zero, then the discrepancy ambiguity set reduces to a singleton (i.e., the reference distribution) and the DRCC reduces to the corresponding (non-robust) chance constraint. Wasserstein metric has received growing attention in recent literature (see, e.g., [165, 59, 42, 47, 153]). In particular, [165] showed that the Wasserstein ambiguity set enjoys asymptotic guarantee as the data size increases to infinity, [47] derived a tractable reformation of the distributionally robust optimization problem with the Wasserstein ambiguity set, and [153] recast Wasserstein DRCCs as a deterministic mixed-integer linear program with the help of the big-M coefficients. The contributions of this study include:

1. We propose a Wasserstein-moment (W-M) ambiguity set that contains both moment and discrepancy information. We apply this strengthened ambiguity set on DRCC-OPF to reduce its conservatism, especially under a limited amount of data.

2. We derive a conservative convex approximation of the proposed DRCC-OPF model using conditional Value-at-Risk (CVaR). We show that this approximation can be recast as linear constraints if the ℓ_1- or the ℓ_∞-norm is applied to characterize the (W-M) ambiguity set, and as second order conic constraints if the ℓ_2-norm is applied. This improves the scalability of DRCC.

3. We demonstrate the proposed approach in a case study on the IEEE 30-bus system. The

results show that the DRCC-OPF model with the W-M ambiguity set provides better balance between cost-effectiveness and reliability, as compared to two state-of-the-art benchmarks.

The remainder of this chpater is organized as follows. We present the DRCC-OPF model under Wasserstein and W-M ambiguity sets in Section 3.2. In Section 3.3, we derive CVaR-based approximations of the proposed models. Section 3.4 reports the case study and the comparisons with the two benchmark approaches. Section 3.5 summarizes this chapter. In addition, Section 3.6 lists the nomenclature of this chapter.

3.2 Mathematical Formulation

3.2.1 Multi-Period Optimal Power Flow

We start by describing the deterministic multi-period OPF problem [86] in the following formulation (3.1).

$$\min_{P_{G,t},\overline{R}_{G,t},\underline{R}_{G,t},d_t} \quad \sum_{t=1}^{N_t} \left(c_G^\top P_{G,t} + c_R^\top(\overline{R}_{G,t} + \underline{R}_{G,t})\right) \tag{3.1a}$$

$$\begin{aligned}
\text{s.t.}\ \ & P_{I,t} = B_G(P_{G,t} + R_t) + B_w P_{w,t} - B_L P_{L,t}, && t \in [N_t], && \text{(3.1b)}\\
& P_{w,t} = P_{w,t}^f + w_t, && t \in [N_t], && \text{(3.1c)}\\
& R_t = -(\mathbf{1}^\top w_t) d_t, && t \in [N_t], && \text{(3.1d)}\\
& \mathbf{1}^\top d_t = 1, && t \in [N_t], && \text{(3.1e)}\\
& \mathbf{1}^\top (B_G P_{G,t} + B_w P_{w,t}^f - B_L P_{L,t}) = 0, && t \in [N_t], && \text{(3.1f)}\\
& P_{G,t+1} - P_{G,t} \le \mathbf{RU}, && t \in [N_t - 1], && \text{(3.1g)}\\
& P_{G,t} - P_{G,t+1} \le \mathbf{RD}, && t \in [N_t - 1], && \text{(3.1h)}\\
& -P_\ell \le A_\ell P_{I,t} \le P_\ell, && t \in [N_t], && \text{(3.1i)}\\
& \mathbf{0} \le P_{G,t} + R_t \le \overline{P}_G, && t \in [N_t], && \text{(3.1j)}\\
& \underline{R}_{G,t} \le R_t \le \overline{R}_{G,t}, && t \in [N_t], && \text{(3.1k)}\\
& P_{G,t}, \overline{R}_{G,t}, \underline{R}_{G,t}, d_t \ge \mathbf{0}, && t \in [N_t]. && \text{(3.1l)}
\end{aligned}$$

where $[N] := \{1, \ldots, N\}$ for integer N. $\mathbf{1}$ (resp. $\mathbf{0}$) denotes an all one (resp. zero) vector of proper dimension. Let N_B, N_G, N_L, N_ℓ, N_w, and N_t be the number of buses, thermal generators, loads, transmission lines, wind farms, and operating time intervals, respectively. $c_G, c_R \in \mathbb{R}^{N_G}$ are the unit costs for energy generation and providing reserve, respectively.

$B_G \in \mathbb{R}^{N_B \times N_G}$, $B_w \in \mathbb{R}^{N_B \times N_w}$, $B_L \in \mathbb{R}^{N_B \times N_L}$ are the bus-generator incidence, bus-wind incidence matrix, and bus-load incidence matrix, respectively. $P^f_{w,t} \in \mathbb{R}^{N_w}$ is the forecasted wind power during time interval t. RU, RD $\in \mathbb{R}^{N_G}$ are the ramp-up and -down capacities of the thermal generators, respectively. $A_\ell \in \mathbb{R}^{N_\ell \times N_B}$ is the distribution factor matrix. $P_\ell \in \mathbb{R}^{N_\ell}$ and $\overline{P}_G \in \mathbb{R}^{N_G}$ are the transmission line and thermal generator production capacities, respectively. $P_{L,t} \in \mathbb{R}^{N_L}$ is the loads during time interval t. Random variables $w_t \in \mathbb{R}^{N_w}$ represent the forecast errors of wind energy during time interval t. Variables $d_t \in \mathbb{R}^{N_G}$ represent the distribution vector during time interval t, which parameterizes an affine response from the generator-providing reserves to the real-time supply/demand mismatch. Variables $P_{G,t} \in \mathbb{R}^{N_G}$ represent the planned thermal generation amounts during time interval t. Variables $\overline{R}_{G,t}$, $\underline{R}_{G,t} \in \mathbb{R}^{N_G}$ represent the generator upward and downward reserve capacities during time interval t, respectively. Auxiliary variables $P_{I,t} \in \mathbb{R}^{N_B}$ represent the net power injection at buses during time interval t. Auxiliary variables $P_{w,t} \in \mathbb{R}^{N_w}$ represent the actual wind energy during time interval t. Auxiliary variables $R_t \in \mathbb{R}^{N_G}$ represent the generator reserve dispatch during time interval t.

The objective function (3.1a) is to minimize the power production and the reserve costs while satisfying the network and physical constraints. As in [141, 113, 21], power flows are modeled as a system of linear equations, i.e., based on the dc power flow approximation. In constraints (3.1b), for each time interval t, matrices B_G, B_w, and B_L map real generators production, wind energy, and loads to $P_{I,t}$, the net power injection at buses. Constraints (3.1c) describe the actual wind energy during each time interval t. Constrains (3.1d) compute the actual power reserve amount provided by all thermal generators during each time interval t. Here distribution vector d_t specifies the portion of the real-time supply/demand mismatch to be provided by each thermal generator. Constraints (3.1e) ensure that distribution vectors d_t, for all $t \in [N_t]$, are valid. Constraints (3.1f) enforce the system balance between the total generation and total load. Constraints (3.1g)–(3.1h) describe the ramp-rate limit restrictions on generation amounts. Constraints (3.1i) describe transmission line capacity limits. Finally, constraints (3.1j) and (3.1k) enforce bounds on the actual generation amounts and reserve amounts of the thermal generators, respectively.

3.2.2 Distributionally Robust Chance-Constrained Optimal Power Flow

Chance-constrained program has been considered in the OPF problem in face of uncertain wind power (see e.g., [162, 113, 21]). It ensures that each constraint subject to uncertainty is satisfied with at least a pre-specified probability with regard to a prescribed distribution. When wind power is incorporated, constraints (3.1i), (3.1j), and (3.1k) in the OPF formu-

lation are subject to uncertainty. For presentation brevity, we present these constraints in the following abstract form:

$$c_i(w)^\top x \leq d_i(w), \qquad \forall i \in [I], \tag{3.2}$$

where $x \in \mathbb{R}^{N_d}$ represents decision variables $(P_G, \overline{R}_G, \underline{R}_G, d)$ in formulation (3.1) and N_d denotes dimension of the decision variables. To handle the random constraint violations, chance constraints attempt to satisfy (3.2) with at least pre-specified probabilities $1-\epsilon_i,\ \forall i \in [I]$, i.e.,

$$\mathbb{Q}\left(c_i(w)^\top x \leq d_i(w)\right) \geq 1-\epsilon_i, \qquad \forall i \in [I], \tag{3.3}$$

where uncertain coefficients $c_i(w)$, $d_i(w)$ depend on the random variables w, $\mathbb{Q}$ denotes the true joint probability distribution of random variables w, and $1-\epsilon_i$ represents the risk threshold of the ith chance constraint, with $\epsilon_i \in [0,1]$ usually taking a small value (e.g., 0.05 or 0.10). In a chance-constrained OPF (CC-OPF) model, constraints (3.1i), (3.1j), and (3.1k) are replaced with the chance constraint counterpart (3.3). In this work, we assume that $c_i(w)$ and $d_i(w)$ depend affinely on w (see [127]):

$$c_i(w) = c_i^0 + \sum_{k=1}^{N_w} c_i^k w_k, \qquad d_i(w) = d_i^0 + \sum_{k=1}^{N_w} d_i^k w_k, \qquad i \in [I]. \tag{3.4}$$

By introducing auxiliary functions $a_i^k(x) : \mathbb{R}^{N_d} \to \mathbb{R}$ and $b_i(x) : \mathbb{R}^{N_d} \to \mathbb{R}, k \in [N_w], i \in [I]$ defined through

$$a_i^k(x) = (c_i^k)^\top x - d_i^k, \qquad b_i(x) = (c_i^0)^\top x - d_i^0, \qquad k \in [N_w], \quad i \in [I] \tag{3.5}$$

and letting $a_i(x) = \left[a_i^1(x), \ldots, a_i^{N_w}(x)\right]^\top$ for all $i \in [I]$, we rewrite constraints (3.3) as

$$\mathbb{Q}\left(a_i(x)^\top w + b_i(x) \leq 0\right) \geq 1-\epsilon_i, \qquad \forall i \in [I]. \tag{3.6}$$

Chance-constrained program was first proposed by Charnes et al. [30], Miller and Wagner [100] and Prékopa [109]. Although having been studied for a long time, chance-constrained program still has the following limitations [171]:

1. In general, the chance-constrained program is computationally intractable. In fact, even to check if a given solution x is feasible requires computing a multi-dimensional integral, which is notoriously difficult when the dimension of the random variables is large. In addition, the feasible region of chance constraint is typically nonconvex

and even disconnected in some cases.

2. Chance-constrained program assumes that full and accurate information of the true distribution $\mathbb{Q}$ is known. However, in practice, this assumption rarely holds. In most situations, $\mathbb{Q}$ has to be estimated from historical data. Similar to the overfitting effects in statistics [129], replacing $\mathbb{Q}$ with a (possibly biased) estimate may lead to over-optimistic solutions.

To address these limitations, DRCC is proposed (see, e.g., [3, 25, 50, 45, 102, 139, 171, 172, 73]). It forces the constraint to be satisfied with a prescribed probability under the worst-case distribution. Here, the worst case is with respect to an ambiguity set $\mathcal{D}$, which contains all probability distributions that are consistent with certain known characteristics of $\mathbb{Q}$ (e.g., moment information). Then, the distributionally robust version of constraints (3.3) is formulated as

$$\inf_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left(a_i(x)^\top w + b_i(x) \leq 0\right) \geq 1-\epsilon_i, \qquad \forall i \in [I]. \tag{3.7}$$

Formulation (3.1) becomes a DRCC-OPF model if constraints (3.1i), (3.1j), and (3.1k) are replaced with (3.7).

3.2.3 Wasserstein-Moment Ambiguity Set

To guarantee the performance of the model as well as to prevent over-conservative decisions, the ambiguity set should contain the true distribution with high confidence while avoiding pathological distributions. In this work, we propose a Wasserstein-moment ambiguity set which incorporates moment information into a Wasserstein ball. The underlying support set $\mathcal{U}$ of w is assumed to be a polyhedron in the form

$$\mathcal{U} := \left\{w \in \mathbb{R}^{N_w} : Hw \leq h\right\}. \tag{3.8}$$

In addition, we use $\mathcal{M}(\mathcal{U})$ to represent all probability distributions supported on $\mathcal{U}$.

3.2.3.1 Wasserstein ambiguity set

The Wasserstein metric is defined on the probability space $\mathcal{M}(\mathcal{U})$ of all probability distributions $\mathbb{P}$ supported on $\mathcal{U}$ with $\mathbb{E}_{\mathbb{P}}[\|w\|] = \int_{\mathcal{U}} \|w\| \mathbb{P}(dw) < \infty$.

Definition 3.2.1 (Wasserstein metric [78])**.** *The Wasserstein metric* $d_W : \mathcal{M}(\mathcal{U}) \times \mathcal{M}(\mathcal{U}) \to$

$\mathbb{R}$ *is defined as*

$$d_W(\mathbb{P}_1, \mathbb{P}_2) := \inf \left\{ \int_{\mathcal{U}^2} \|w^{(1)} - w^{(2)}\| \Pi(dw^{(1)}, dw^{(2)}) : \begin{array}{r} \Pi \textit{ is a joint distribution of } w^{(1)} \textit{ and } w^{(2)} \\ \textit{with marginals } \mathbb{P}_1 \textit{ and } \mathbb{P}_2, \textit{ respectively} \end{array} \right\}$$

for all distributions $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{M}(\mathcal{U})$*, and* $\|\cdot\|$ *represents an arbitrary norm on* $\mathbb{R}^{N_w}$.

Accordingly, the Wasserstein ambiguity set is defined as

$$\mathcal{D}_W := \{\mathbb{P} \in \mathcal{M}(\mathcal{U}) : d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta\}, \tag{3.9}$$

that is, $\mathcal{D}_W$ is a Wasserstein ball of radius θ centered at a reference distribution $\mathbb{P}_0$. A natural approach to estimating the reference distribution $\mathbb{P}_0$ is by using the discrete empirical probability distribution based on N historical data samples $w_0^{(1)}, w_0^{(2)}, \ldots, w_0^{(N)}$ [47], i.e.,

$$\mathbb{P}_0(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{w_0^{(i)}}(x), \tag{3.10}$$

where $\delta_{w_0^{(i)}}(x)$ denotes the indicator variable that equals 1 if $w_0^{(i)} \leq x$ and 0 otherwise. The following theorem provides a probabilistic guarantee of the true probability distribution $\mathbb{Q}$ lying within the Wasserstein ambiguity set $\mathcal{D}_W$.

Theorem 3.2.2 ([165])**.** *Let* N *be the number of data points and* $\varnothing$ *be the diameter of* N_w*-dimensional support space* $\mathcal{U}$*, we have*

$$P(d_W(\mathbb{P}_0, \mathbb{Q}) \leq \theta) \geq 1 - \exp\left(-\frac{\theta^2}{2\varnothing^2} N\right).$$

Setting the confidence level to be $\eta = 1 - \exp(-\frac{\theta^2}{2\varnothing^2} N)$*, the radius*

$$\theta = \varnothing \sqrt{\frac{2}{N} \log\left(\frac{1}{1-\eta}\right)} \tag{3.11}$$

of the Wasserstein ambiguity set $\mathcal{D}_W$ *ensures that* $\mathbb{Q} \in \mathcal{D}_W$ *with probability at least* η *when the diameter* $\varnothing$ *of the support space* $\mathcal{U}$ *is defined by [51]:*

$$\varnothing = \sup\left\{d_W(w^{(i)}, w^{(j)}) : w^{(i)}, w^{(j)} \in \mathcal{U}\right\}.$$

3.2.3.2 Incorporating additional moment information

A Wasserstein ball contains all distributions within a certain distance from a reference distribution $\mathbb{P}_0$ measured by the Wasserstein metric. Theorem 3.2.2 proves a strong probabilistic guarantee. However, the Wasserstein ambiguity set ignores conveniently available distributional information, e.g., mean and mean absolution deviation of random variables. In this section, we incorporate the moment information into the Wasserstein ball to further reduce the conservativeness of this ambiguity set. Given the first moment information and the Wasserstein ball of random variables w , we define the Wasserstein-moment ambiguity set as following:

$$\mathcal{D}_{WM} := \left\{ \mathbb{P} \in \mathcal{M}(\mathcal{U}) : \begin{array}{c} d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta, \\ \mathbb{E}_{\mathbb{P}}[w] = \mu, \\ \mathbb{E}_{\mathbb{P}}[V^\top (w-\mu)]^+ \leq \delta^+, \\ \mathbb{E}_{\mathbb{P}}[V^\top (w-\mu)]^- \leq \delta^-, \end{array} \right\} \tag{3.12}$$

where $(x)^+ := \max\{x, 0\}$, and $(x)^- := \max\{-x, 0\}$. In (3.12), we consider all probability distributions of w that (i) are supported on the set $\mathcal{U}$; (ii) are in the Wasserstein ball of radius θ centered at the reference distribution $\mathbb{P}_0$; (iii) have mean value μ; and (iv) have at most δ^+ (resp. δ^-) mean positive (resp. negative) deviation along perturbation directions $V \in \mathbb{R}^{N_w \times N_v}$.

3.3 Solution Methodology

3.3.1 CVaR Approximations

In this section, we conservatively approximate DRCC under Wasserstein and W-M ambiguity sets as convex programs based on CVaR.

Definition 3.3.1 (CVaR [114]). *For a given measurable loss function $L(x, w) : \mathbb{R}^{N_d} \times \mathbb{R}^{N_w} \to \mathbb{R}$, probability distribution $\mathbb{P}$, and the tolerance $\epsilon \in (0, 1)$, the CVaR at level ϵ with respect to $\mathbb{P}$ is defined as*

$$\mathbb{P}\text{-}\boldsymbol{CVaR}_\epsilon \left(L(x, w) \right) := \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[(L(x, w) - \beta)^+ \right] \right\},$$

where $\mathbb{E}_{\mathbb{P}}[\cdot]$ denotes the expectation with respect to distribution $\mathbb{P}$ and $(x)^+$ represents $\max\{x, 0\}$.

Essentially, $\mathbb{P}$-CVaR$_\epsilon(L(x,w))$ evaluates the conditional expectation of loss function $L(x,w)$ on the upper ϵ-tail part of its distribution $\mathbb{P}$. Therefore, the following inequality holds valid:

$$\mathbb{P}\big(L(x,w) \leq \mathbb{P}\text{-CVaR}_\epsilon\left(L(x,w)\right)\big) \geq 1-\epsilon. \tag{3.13}$$

Since CVaR is a convex function with respect to its random input $L(x,w)$, we can use it to obtain a convex conservation approximation of DRCC. Let $L_i(x,w) = a_i(x)^\top w + b_i(x)$ for all $i \in [I]$ be the loss functions. If $\mathbb{P}$-CVaR$_\epsilon\left(L_i(x,w)\right) \leq 0, \forall i \in [I]$ hold for every probability $\mathbb{P} \in \mathcal{D}$, then DRCC (3.7) will be satisfied. Formally,

$$\begin{aligned} &\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\text{-CVaR}_{\epsilon_i}\left(a_i(x)^\top w + b_i(x)\right) \leq 0,\ \forall i \in [I], \\ \Rightarrow\ &\inf_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left(a_i(x)^\top w + b_i(x) \leq 0\right) \geq 1-\epsilon_i,\ \forall i \in [I]. \end{aligned} \tag{3.14}$$

Replacing DRCC (3.7) by constraints (3.14) yields a conservative convex approximation for the DRCC-OPF problem. Based on Definition 3.3.1, we recast the CVaR approximation (3.14) as follows:

$$\begin{aligned} &\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\text{-CVaR}_{\epsilon_i}\left(a_i(x)^\top w + b_i(x)\right) \leq 0,\ \forall i \in [I], \\ \Leftrightarrow\ &\sup_{\mathbb{P}\in\mathcal{D}} \inf_{\beta\in\mathbb{R}} \left\{\beta + \frac{1}{\epsilon_i}\mathbb{E}_\mathbb{P}\left[(a_i(x)^\top w + b_i(x) - \beta)^+\right]\right\} \leq 0,\ \forall i \in [I], \\ \Leftrightarrow\ &\inf_{\beta\in\mathbb{R}} \left\{\beta + \frac{1}{\epsilon_i}\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_\mathbb{P}\left[(a_i(x)^\top w + b_i(x) - \beta)^+\right]\right\} \leq 0,\ \forall i \in [I], \\ \Leftrightarrow\ &\exists \beta_i \in \mathbb{R}:\ \beta_i + \frac{1}{\epsilon_i}\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_\mathbb{P}\left[(a_i(x)^\top w + b_i(x) - \beta_i)^+\right] \leq 0,\ \forall i \in [I]. \end{aligned} \tag{3.15}$$

In the following sections, we present tractable reformulations for the CVaR approximation under $\mathcal{D}_W$ and $\mathcal{D}_{WM}$.

3.3.1.1 CVaR approximation with the Wasserstein ambiguity set

Theorem 3.3.2 gives an exact reformulation of the worst-case expectation in (3.15) over $\mathcal{D}_W$.

Theorem 3.3.2 ([47])**.** *Suppose that the loss function is in the form of maximal of several affine functions, that is, $L(x,w) := \max_{k\in[K]} l_k(x,w)$, where $l_k(x,w) = g_k(x)^\top w + e_k(x), \forall k \in [K]$. And suppose that the uncertainty set is a polyhedron of form* (3.8)*. Let N be the number of data samples and $w_0^{(n)}, n \in [N]$ be the data samples. Then the worst-case*

expectation

$$\sup_{\mathbb{P}\in\mathcal{D}_W} \mathbb{E}_{\mathbb{P}}[L(x,w)] \tag{3.16}$$

with the Wasserstein ambiguity set (3.9) *equals the optimal value of the following convex program*

$$\inf_{\lambda,s_n,\gamma_{nk}} \quad \lambda\theta + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{3.17a}$$

$$\text{s.t.} \quad e_k(x) + g_k(x)^\top w_0^{(n)} + \gamma_{nk}^\top \left(h - H w_0^{(n)}\right) \le s_n, \quad \forall n \in [N], \quad \forall k \in [K], \tag{3.17b}$$

$$\|H^\top \gamma_{nk} - g_k(x)\|_* \le \lambda, \qquad \forall n \in [N], \quad \forall k \in [K], \tag{3.17c}$$

$$\gamma_{nk} \ge 0, \qquad \forall n \in [N], \quad \forall k \in [K]. \tag{3.17d}$$

where $\|\cdot\|_*$ *is the dual norm of the norm used in the Wasserstein mertic.*

By Theorem 3.3.2, for each $i \in [I]$, the CVaR constraint (3.15) can be reformulated as:

$$\beta_i + \frac{1}{\epsilon_i}\left(\lambda_i\theta + \frac{1}{N}\sum_{n=1}^{N} s_{in}\right) \le 0, \qquad \forall n \in [N], \quad \forall k \in [K], \tag{3.18a}$$

$$e_k(x) + g_k(x)^\top w_0^{(n)} + \gamma_{ink}^\top \left(h - H w_0^{(n)}\right) \le s_{in}, \qquad \forall n \in [N], \quad \forall k \in [K], \tag{3.18b}$$

$$\|H^\top \gamma_{ink} - g_k(x)\|_* \le \lambda_i, \qquad \forall n \in [N], \quad \forall k \in [K], \tag{3.18c}$$

$$\gamma_{ink} \ge 0, \qquad \forall n \in [N], \quad \forall k \in [K], \tag{3.18d}$$

where x are decision variables and $\lambda_i, s_{in}, \gamma_{ink}$ are auxiliary variables.

3.3.1.2 CVaR with the W-M ambiguity set

Theorem 3.3.3 gives an exact reformulation of the worst-case expectation in (3.15) over $\mathcal{D}_{WM}$.

Theorem 3.3.3. *Suppose that the loss function is in the form of maximal of several affine functions, that is,* $L(x,w) := \max_{k\in[K]} l_k(x,w)$*, where* $l_k(x,w) = g_k(x)^\top w + e_k(x), \forall k \in [K]$*. And suppose that the uncertainty set is a polyhedron* (3.8)*. Let* N *be the number of data samples and* $w_0^{(n)}, n \in [N]$ *be the data samples. Then the worst-case expectation*

$$\sup_{\mathbb{P}\in\mathcal{D}_{WM}} \mathbb{E}_{\mathbb{P}}[L(x,w)] \tag{3.19}$$

with the W-M ambiguity set (3.12) *equals the optimal value of following convex program*

$$\inf_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N,\, z_{nk}\in\mathbb{R}^{N_w} \\ d_{nk}^{\pm}\in\mathbb{R}_+^{N_v}, d_{nk}^{h}\in\mathbb{R}_+^{N_u}}} \quad \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{3.20a}$$

$$\text{s.t.} \;\; \mu^\top V d_{nk}^+ - \mu^\top V d_{nk}^- + h^\top d_{nk}^h + e_k(x) + z_{nk}^\top w_0^{(n)} \le s_n, \quad \forall n\in[N], \forall k\in[K], \tag{3.20b}$$

$$V d_{nk}^+ - V d_{nk}^- + H^\top d_{nk}^h = g_k(x) - z_{nk} - p, \quad \forall n\in[N], \forall k\in[K], \tag{3.20c}$$

$$d_{nk}^+ \le q^+, \quad \forall n\in[N], \forall k\in[K], \tag{3.20d}$$

$$d_{nk}^- \le q^-, \quad \forall n\in[N], \forall k\in[K], \tag{3.20e}$$

$$\|z_{nk}\|_* \le \lambda, \quad \forall n\in[N], \forall k\in[K]. \tag{3.20f}$$

where $\|\cdot\|_*$ *is the dual norm of the norm used in the Wasserstein mertic.*

Proof. First, the worst-case expectation (3.19) can be written as a functional optimization problem as follows:

$$\sup_{\substack{\mathbb{P}\in\mathcal{M}(\mathcal{U}) \\ \Pi\in\mathcal{M}(\mathcal{U})\times\mathcal{M}(\mathcal{U})}} \quad \int_{\mathcal{U}} L(x,w)\mathbb{P}(dw) \tag{3.21a}$$

$$\text{s.t.} \quad \int_{\mathcal{U}^2} \|w-w'\|\Pi(dw,dw') \le \theta, \tag{3.21b}$$

$$\Pi \text{ is a joint distribution of } w \text{ and } w' \text{with marginals } \mathbb{P} \text{ and } \mathbb{P}_0, \text{ respectively,} \tag{3.21c}$$

$$\int_{\mathcal{U}} w\mathbb{P}(dw) = \mu, \tag{3.21d}$$

$$\int_{\mathcal{U}} [V^\top(w-\mu)]^+\mathbb{P}(dw) \le \delta^+, \tag{3.21e}$$

$$\int_{\mathcal{U}} [V^\top(w-\mu)]^-\mathbb{P}(dw) \le \delta^-. \tag{3.21f}$$

By the law of total probability, the joint probability distribution Π of w and w' can be constructed from the marginal distribution $\mathbb{P}_0$ of w' and the conditional distribution $\mathbb{P}^{(n)}$ of w given $w' = w_0^{(n)},\ n\in[N]$. Then problem (3.21) can be reformulated as

$$\sup_{\mathbb{P}^{(n)}\in\mathcal{M}(\mathcal{U})} \quad \frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}} L(x,w)\mathbb{P}^{(n)}(dw) \tag{3.22a}$$

$$\text{s.t.}\ \ \frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}\|w-w_0^{(n)}\|\mathbb{P}^{(n)}(dw)\le\theta, \tag{3.22b}$$

$$\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}w\mathbb{P}^{(n)}(dw)=\mu, \tag{3.22c}$$

$$\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}[V^\top(w-\mu)]^+\mathbb{P}^{(n)}(dw)\le\delta^+, \tag{3.22d}$$

$$\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}[V^\top(w-\mu)]^-\mathbb{P}^{(n)}(dw)\le\delta^-. \tag{3.22e}$$

Take the Lagrangian dual of (3.22), and strong duality holds due to an extended version of Proposition 3.4 in [126].

$$\begin{aligned}
=\Bigg\{&\inf_{\substack{\lambda\in\mathbb{R}_+,\ q^\pm\in\mathbb{R}_+^{N_v}\\ p\in\mathbb{R}^{N_w}}}\ \sup_{\mathbb{P}^{(n)}\in\mathcal{M}(\mathcal{U})}\ \ \frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}L(x,w)\mathbb{P}^{(n)}(dw)\\
&\qquad+\lambda\left[\theta-\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}\|w-w_0^{(n)}\|\mathbb{P}^{(n)}(dw)\right]\\
&\qquad+p^\top\left[\mu-\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}w\mathbb{P}^{(n)}(dw)\right]\\
&\qquad+q^{+\top}\left[\delta^+-\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}[V^\top(w-\mu)]^+\mathbb{P}^{(n)}(dw)\right]\\
&\qquad+q^{-\top}\left[\delta^--\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}[V^\top(w-\mu)]^-\mathbb{P}^{(n)}(dw)\right]
\end{aligned}\Bigg\} \tag{3.23a}$$

$$\begin{aligned}
=\Bigg\{&\inf_{\substack{\lambda\in\mathbb{R}_+,\ q^\pm\in\mathbb{R}_+^{N_v}\\ p\in\mathbb{R}^{N_w}}}\ \theta\lambda+p^\top\mu+q^{+\top}\delta^++q^{-\top}\delta^-\\
&\qquad+\sup_{\mathbb{P}^{(n)}\in\mathcal{M}(\mathcal{U})}\frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{U}}\Big[L(x,w)-\lambda\|w-w_0^{(n)}\|-p^\top w\\
&\qquad-q^{+\top}[V^\top(w-\mu)]^+-q^{-\top}[V^\top(w-\mu)]^-\Big]\mathbb{P}^{(n)}(dw)
\end{aligned}\Bigg\} \tag{3.23b}$$

$$
= \begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w}}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- \\ & + \dfrac{1}{N}\sum\limits_{n=1}^{N} \sup\limits_{w\in\mathcal{U}} \Big[L(x,w) - \lambda\|w - w_0^{(n)}\| - p^\top w \\ & - q^{+\top}[V^\top(w-\mu)]^+ - q^{-\top}[V^\top(w-\mu)]^- \Big] \end{cases} \tag{3.23c}
$$

$$
= \begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \dfrac{1}{N}\sum\limits_{n=1}^{N} s_n \\ \text{s.t.} & \sup\limits_{w\in\mathcal{U}} \Big[L(x,w) - \lambda\|w - w_0^{(n)}\| - p^\top w - q^{+\top}[V^\top(w-\mu)]^+ \\ & - q^{-\top}[V^\top(w-\mu)]^- \Big] \le s_n, \forall n \in [N], \end{cases} \tag{3.23d}
$$

$$
= \begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \dfrac{1}{N}\sum\limits_{n=1}^{N} s_n \\ \text{s.t.} & \sup\limits_{\substack{w\in\mathcal{U} \\ t_{nk}^{\pm}\in\mathbb{R}_+^{N_v}}} \Big[l_k(x,w) - \max\limits_{\|z_{nk}\|_*\le\lambda} z_{nk}^\top(w - w_0^{(n)}) - p^\top w \\ & - q^{+\top}t_{nk}^+ - q^{-\top}t_{nk}^- \Big] \le s_n, \forall n \in [N],\ \forall k \in [K], \\ & \quad \text{s.t. } t_{nk}^+ \ge V^\top(w-\mu), \\ & \quad\quad t_{nk}^- \ge -V^\top(w-\mu), \end{cases} \tag{3.23e}
$$

$$
= \begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \dfrac{1}{N}\sum\limits_{n=1}^{N} s_n \\ \text{s.t.} & \min\limits_{\|z_{nk}\|_*\le\lambda} \sup\limits_{\substack{w\in\mathcal{U} \\ t_{nk}^{\pm}\in\mathbb{R}_+^{N_v}}} \Big[l_k(x,w) - z_{nk}^\top(w - w_0^{(n)}) - p^\top w \\ & - q^{+\top}t_{nk}^+ - q^{-\top}t_{nk}^- \Big] \le s_n, \forall n \in [N],\ \forall k \in [K], \\ & \quad \text{s.t. } t_{nk}^+ \ge V^\top(w-\mu), \\ & \quad\quad t_{nk}^- \ge -V^\top(w-\mu), \end{cases} \tag{3.23f}
$$

$$
= \begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N,\, z_{nk}\in\mathbb{R}^{N_w}}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} s_n \\ \text{s.t.} & \|z_{nk}\|_* \le \lambda,\ \forall n \in [N],\ \forall k \in [K], \\ & \sup\limits_{\substack{w\in\mathcal{U}, \\ t_{nk}^{\pm}\in\mathbb{R}_+^{N_v}}} \Big[l_k(x,w) - z_{nk}^\top(w - w_0^{(n)}) - p^\top w \\ & -q^{+\top}t_{nk}^+ - q^{-\top}t_{nk}^- \Big] \le s_n, \forall n \in [N],\ \forall k \in [K], \\ & \quad \text{s.t. } t_{nk}^+ \ge V^\top(w-\mu), \\ & \quad\quad t_{nk}^- \ge -V^\top(w-\mu). \end{cases} \tag{3.23g}
$$

Equality (3.23c) holds because the worst-case distribution will occur on a singleton distribution, i.e., $\mathbb{P}^{(n)}(w) = 1$ for some $w \in \mathcal{U}$ [47]. In (3.23d), we introduce auxiliary variables s_n. Equality (3.23e) holds because: (i) $L(x,w) := \max_{k\in[K]} l_k(x,w)$; (ii) the definition of dual norm is used and (iii) auxiliary variables t^+ and t^- are introduced to linearize $(\cdot)^+$ and $(\cdot)^-$, respectively. Equality (3.23f) follows from the classical minimax theorem [15] and the fact that (i) the set $\{z_{nk} : \|z_{nk}\|_* \le \lambda\}$ is compact for any finite $\lambda \ge 0$; (ii) $\mathcal{U}$ is a polyhedral uncertainty set hence $\mathcal{U}$ is convex and closed and (iii) the function is linear in z_{nk}, w and $t_{nk}^{\pm}$.

Take the dual of each supreme problem in (3.23g), we have

$$
\begin{cases} \sup\limits_{\substack{w\in\mathbb{R}^{N_w} \\ t_{nk}^{\pm}\in\mathbb{R}_+^{N_v}}} & \Big[g_k(x)^\top w + e_k(x) - z_{nk}^\top(w - w_0^{(n)}) - p^\top w - q^{+\top}t_{nk}^+ - q^{-\top}t_{nk}^- \Big] \\ \text{s.t.} & t_{nk}^+ \ge V^\top(w-\mu), \\ & t_{nk}^- \ge -V^\top(w-\mu), \\ & Hw \le h, \end{cases} \tag{3.24a}
$$

$$
= \begin{cases} \inf\limits_{\substack{d_{nk}^{\pm}\in\mathbb{R}_+^{N_v} \\ d_{nk}^{h}\in\mathbb{R}_+^{N_u}}} & \Big[\mu^\top V d_{nk}^+ - \mu^\top V d_{nk}^- + h^T d_{nk}^h \Big] + e_k(x) + z_{nk}^\top w_0^{(n)} \\ \text{s.t.} & V d_{nk}^+ - V d_{nk}^- + H^\top d_{nk}^h = g_k(x) - z_{nk} - p, \\ & d_{nk}^+ \le q^+, \\ & d_{nk}^- \le q^-, \end{cases} \tag{3.24b}
$$

Strong duality holds since this is a linear formulation. Plug formulation (3.24b) into (3.23g),

we have

$$\begin{cases} \inf\limits_{\substack{\lambda\in\mathbb{R}_+,\, q^{\pm}\in\mathbb{R}_+^{N_v} \\ p\in\mathbb{R}^{N_w},\, s\in\mathbb{R}^N,\, z_{nk}\in\mathbb{R}^{N_w} \\ d_{nk}^{\pm}\in\mathbb{R}_+^{N_v}, d_{nk}^{h}\in\mathbb{R}_+^{N_u}}} & \theta\lambda + p^\top\mu + q^{+\top}\delta^+ + q^{-\top}\delta^- + \dfrac{1}{N}\sum\limits_{n=1}^{N} s_n \\ \text{s.t.} & \mu^\top V d_{nk}^+ - \mu^\top V d_{nk}^- + h^T d_{nk}^h + e_k(x) + z_{nk}^\top w_0^{(n)} \le s_n, \\ & \forall n\in[N],\ \forall k\in[K], \\ & V d_{nk}^+ - V d_{nk}^- + H^\top d_{nk}^h = g_k(x) - z_{nk} - p, \\ & \forall n\in[N],\ \forall k\in[K], \\ & d_{nk}^+ \le q^+,\ \forall n\in[N],\ \forall k\in[K], \\ & d_{nk}^- \le q^-,\ \forall n\in[N],\ \forall k\in[K], \\ & \|z_{nk}\|_* \le \lambda,\ \forall n\in[N],\ \forall k\in[K]. \end{cases} \tag{3.25a}$$

which completes the proof. □

By Theorem 3.3.3, for each $i\in[I]$, the CVaR constraint (3.15) can be recast as:

$$\beta_i + \frac{1}{\epsilon_i}\left(\lambda_i\theta + p_i^\top\mu + q_i^{+\top}\delta^+ + q_i^{-\top}\delta^- + \frac{1}{N}\sum_{n=1}^{N} s_{in}\right) \le 0, \forall n\in[N], \forall k\in[K], \tag{3.26a}$$

$$\mu^\top V d_{ink}^+ - \mu^\top V d_{ink}^- + h^\top d_{ink}^h + e_k(x) + z_{ink}^\top w_0^{(n)} \le s_{in}, \forall n\in[N], \forall k\in[K], \tag{3.26b}$$

$$V d_{ink}^+ - V d_{ink}^- + H^\top d_{ink}^h = g_k(x) - z_{ink} - p_i, \quad \forall n\in[N], \forall k\in[K], \tag{3.26c}$$

$$d_{ink}^+ \le q_i^+, \quad \forall n\in[N], \forall k\in[K], \tag{3.26d}$$

$$d_{ink}^- \le q_i^-, \quad \forall n\in[N], \forall k\in[K], \tag{3.26e}$$

$$\|z_{ink}\|_* \le \lambda_i, \quad \forall n\in[N], \forall k\in[K], \tag{3.26f}$$

where x are decision variables and $\lambda_i, s_{in}, p_i, q_i^{\pm}, d_{ink}^{\pm}, d_{ink}^h, z_{ink}$ are auxiliary variables.

When the ℓ_1- or the ℓ_∞-norm is applied in the Wasserstein metric, constraints (3.18c) and (3.26f) can be reformulated as linear constraints; and when ℓ_2-norm is applied, (3.18c) and (3.26f) are second-order conic constraints. Thus, the CVaR approximation under both Wasserstein and W-M ambiguity sets lead to tractable convex program of the DRCC-OPF problem. These approximations can then be efficiently solved by the off-the-shelf commercial softwares (e.g., CPLEX).

3.4 Case Study

To test the DRCC-OPF models under $\mathcal{D}_W$ and $\mathcal{D}_{WM}$, we conduct computational case studies on the IEEE 30-bus system. All instances are solved by CPLEX 12.7.1 on a 64-bit Windows Server 2012 R2 Standard machine with two Intel(R) Xeon(R) E5-2630 v4 processors, running at 2.20 GHz with 128 GB memory.

3.4.1 Experiment Setting

The IEEE 30-bus system contains 30 nodes and 41 transmission lines (see http://www.ee.washington.edu/research/pstca/ for the data set). We place 9 thermal generators (see Tables 3.1 for the generator characteristics) and 3 wind farms in this system. The 3 wind farms are based on 3 onshore sites from the National Renewable Energy Laboratory (NREL).

Table 3.1: Generator characteristics

Generator Index	Bus Index	U_i	RU_i	RD_i	c_i^{G}	c_i^{R}
1	4	20	10	10	2.5	25
2	6	10	5	5	2.4	24
3	10	30	15	15	1.4	14
4	12	20	10	10	1.3	13
5	15	30	15	15	2.8	28
6	18	30	15	15	1.8	18
7	24	20	10	10	2.2	22
8	26	10	5	5	1.1	11
9	27	20	10	10	2.6	26

We characterize the uncertainty set $\mathcal{U}$ defined in (3.8), the Wasserstein ambiguity set $\mathcal{D}_W$ defined in (3.9), and the W-M ambiguity set $\mathcal{D}_{WM}$ defined in (3.12) based on the historical data from NREL Wind Integration National Dataset Toolkit [41, 40, 90, 79]. For each day we test, we collect the historical data of the forecast errors of wind energy in the previous 20 days, and 20 days before and after this day in the previous six years. For example, if we test Feb. 1 in the year 2013, then we collect data during Jan. 12–31 in 2013, Jan. 12–Feb. 20 in 2007–2012. For the uncertainty set $\mathcal{U}$, we consider upper and lower bounds of the hourly forecast errors of each wind farm as well as the upper and lower bounds of the hourly total forecast errors of all three wind farms. The ℓ_1-norm is applied to characterize both $\mathcal{D}_W$ and $\mathcal{D}_{WM}$. Two types of Wasserstein ball radii θ are used to test our approaches:

1. The theoretical radius calculated based on Theorem 3.2.2 in [165], with regard to the confidence level $\eta = 0.95$.

2. A statistical radius inspired by the concept of SAA. The theoretical radius guarantees that the true probability distribution $\mathbb{Q} \in \mathcal{D}_W$ with confidence level at least η, but this radius is often too conservative and can lead to unnecessarily robust solutions. The statistical radius is estimated by calculating the Wasserstein distance between the reference distribution $\mathbb{P}_0$ and another reference distribution $\mathbb{P}_e$ constructed from much more data samples than $\mathbb{P}_0$. That is, the radius is set to be $d_W(\mathbb{P}_0, \mathbb{P}_e)$, where $\mathbb{P}_e$ is a uniform distribution on $N_c \gg N$ data samples.

The proposed approaches are tested under a 20% penetration level of renewable energy. Using the median of renewable energy from each site during each hour as a benchmark, we scale the total capacity of the wind farms to take up 20% of the total electricity load. We schedule the power flow of this power system for 30 days. In each day, we solve the DRCC-OPF model with $N_t = 6$ hours. We compare the DRCC-OPF under $\mathcal{D}_W$ and $\mathcal{D}_{WM}$ with the moment-based DRCC-OPF model in [164] and a benchmark approach based on Gaussian approximation [113, 21]. Both approaches lead to second-order conic programming reformulations. In particular, the moment ambiguity set used in [164] is

$$\mathcal{D}_M := \left\{ \mathbb{P} : \begin{array}{l} \int_{w \in \mathbb{R}^{N_w}} \mathbb{P}(dw) = 1 \\ \mathbb{E}[w] = \mu \\ \mathbb{E}[(w-\mu)(w-\mu)^\top] = \Sigma \end{array} \right\}, \tag{3.27}$$

where μ and Σ are the empirical mean and covariance matrix of the random variables w, respectively.

3.4.2 Out-of-Sample Performance

We first compare solutions of the existing approaches with those of our approaches. We refer to them as

- Gaussian (A1): The Gaussian approximation based approach [113, 21].
- DR-moment (A2): The DR approach with $\mathcal{D}_M$ [164].
- DR-Wasserstein-T (A3): The DR approach with $\mathcal{D}_W$ using the theoretical radius.
- DR-Wasserstein-S (A4): The DR approach with $\mathcal{D}_W$ using the statistical radius.

- DR-Wasserstein-Moment-S (A5): The DR approach with $\mathcal{D}_{WM}$ using the statistical radius.

In the experiments, we evaluate the performance of the approaches by the following three criteria:

1. Cost: The average objective cost (3.1a) out of the 30 days.
2. Reliability: The fraction of days out of the 30 days in which all the constraints are satisfied in the out-of-sample simulation.
3. Time: The average CPU seconds for solving each instance.

We report the comparison of the performance of A1–A5 in Table 3.2 and Figure 3.1. In this simulation, all approaches (A1–A5) are solved based on 260 data samples of $\hat{w}$. Note that A1 and A2 only use data to get the empirical mean and covariance matrix while A3–A5 directly incorporate the data samples. First, the average computational time (see the column Time) is around 5 minutes for solving A3 and A4 to global optimality and 90 minutes for solving A5. A1 and A2 can be solved within 1 second. The main difficulty of solving A3, A4, and A5 is that the number of constraints in these models is related to the number of data samples. As compared with A3 and A4, A5 is computationally heavier due to the incorporation of moment information in $\mathcal{U}_{WM}$. Although A5 takes longer time, it still can be solved within a reasonable amount of time for practical applications. Second, we report the results of the out-of-sample performance of each approach (see the columns Cost and Reliability). Overall, A5 balances the robustness and cost-efficiency the best among all the five approaches. A1 is the most optimistic, leading to the lowest cost and reliability. This is mainly because the wind forecast errors typically do not follow a Gaussian distribution [64]. Compared with A3–A5, A2 is much more sensitive to the risk threshold ($1 - \epsilon_i$). When the threshold is high, A2 provides over-conservative solutions. On the other hand, when the threshold is low, A2 becomes less conservative than A3 and A4, but still more conservative than A5. Note that A3 provides the same solution in all the experiments under all the risk thresholds. This is because the theoretical radius is so large that the ambiguity set even includes some singletons, i.e., probability distributions that put all the weight on certain worst-case scenarios. In this case, A3 performs similarly as a classical robust OPF model. In contrast, the statistical radius produces a less conservative ambiguity set so that the ambiguity set can exclude such extremal distribution. A4 provides relatively less conservative solutions which have 100% reliability under all risk thresholds. By incorporating the moment information into the Wasserstein ambiguity set, for all risk

thresholds, A5 shows a delicate balance between reliability and cost-effectiveness. It provides a less conservative solution while always satisfying the uncertain constraints above the required risk thresholds.

Table 3.2: Comparisons on the performance of A1–A5

	A1: Gaussion			A2: DR-moment					
$1-\epsilon_i$ (%)	Cost	Reliability (%)	Time (s)	Cost	Reliability (%)	Time (s)			
99	1401.20	60%	0.29	5769.14	100%	0.43			
95	1299.53	43%	0.31	2786.56	100%	0.36			
90	1236.65	30%	0.34	2115.92	100%	0.32			
85	1188.81	17%	0.28	1811.64	93%	0.32			
	A3: DR-Wasserstein-T			A4: DR-Wasserstein-S			A5: DR-Wasserstein-Moment-S		
$1-\epsilon_i$ (%)	Cost	Reliability (%)	Time (s)	Cost	Reliability (%)	Time (s)	Cost	Reliability (%)	Time (s)
99	2135.52	100%	478.08	2135.52	100%	352.78	2135.52	100%	2619.64
95	2135.52	100%	406.94	2134.39	100%	248.60	2130.30	100%	6185.00
90	2135.52	100%	392.15	2116.88	100%	243.36	1915.35	100%	6450.25
85	2135.52	100%	375.43	1945.44	100%	284.18	1680.60	87%	6320.47

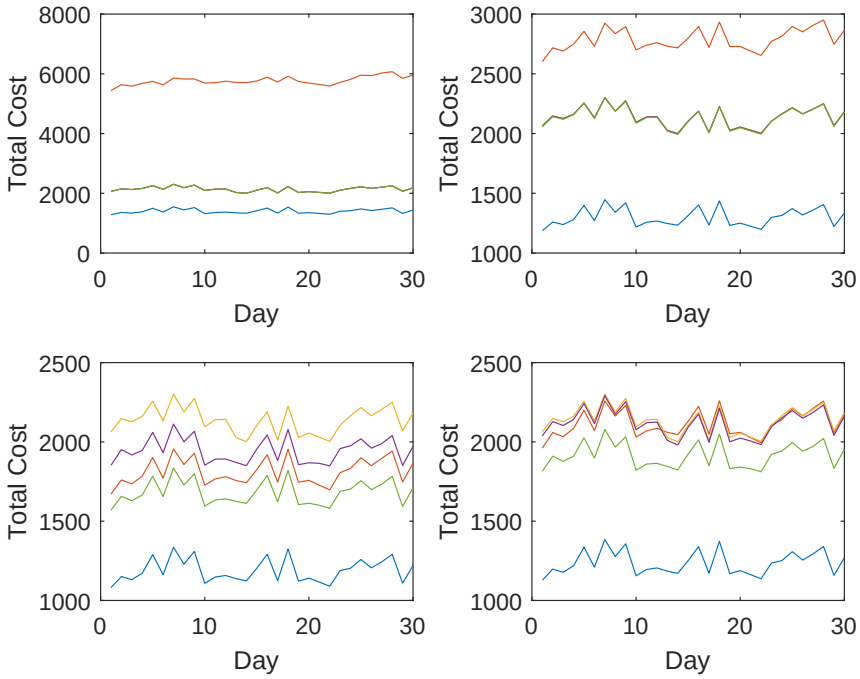


Figure 3.1: Average total cost comparison on the 30-bus System; Blue - A1, Red - A2, Yellow - A3, Purple - A4, Green - A5; Risk threshold - 99%, 95%, 90%, 85% Clockwise[1]

The Table 3.3 reports the influence of the data sizes on the performance of A3–A5. First, the computation time decreases nearly linearly with the data size decreases (see the column Time). The column Radius reports the average radius of the 30 days for each

[1]When $1-\epsilon_i = 99\%$ and 95%, the average total costs of A3, A4, and A5 are too close to distinguish.

approach and each data size. Note that the radii of all approaches increase as the data size decreases and the theoretical radii are much larger than the corresponding statistical radii. Second, since the theoretical radii are over-conservative, A3 yields the same solution for all the data sizes. A4 and A5 provide less conservative solutions than A3 because the corresponding ambiguity sets defined by the statistical radii are much smaller. The costs in A4 and A5 increase as the data size decreases because less data leads to a larger radius. By incorporating moment information, the Wasserstein ambiguity set is further strengthened and A5 shows the best balance between cost-effectiveness and reliability.

Table 3.3: Comparisons on the performance of A3, A4, and A5 with different data sizes

$1-\epsilon_i = 85\%$	A3: DR-Wasserstein-T				A4: DR-Wasserstein-S				A5: DR-Wasserstein-Moment-S			
Number of samples (n)	Cost	Reliability (%)	Time (s)	Radius	Cost	Reliability (%)	Time (s)	Radius	Cost	Reliability (%)	Time (s)	Radius
260	2135.52	100	375.43	3.24	1945.44	100	284.18	0.52	1680.60	87	6320.47	0.52
195	2135.52	100	244.47	3.61	2077.44	100	197.57	0.73	1788.43	97	3846.04	0.73
130	2135.52	100	112.24	4.18	2128.50	100	106.06	1.02	1924.70	100	1927.72	1.02
65	2135.52	100	47.39	5.34	2132.53	100	35.17	1.48	2061.73	100	581.22	1.48

3.5 Conclusion

In this chapter, we proposed a Wasserstein-moment ambiguity set and formulated a multi-period DRCC-OPF model with both the Wasserstein and the Wasserstein-moment ambiguity sets. These two models can be conservatively recast as an linear program if the ℓ_1- or the ℓ_∞-norm is applied in the Wasserstein metric, or as a second-order conic program if the ℓ_2-norm is applied. We also proposed a statistical radius to overcome the over-conservativeness of the theoretical radius. We compared the out-of-sample performance of the proposed approaches with an existing moment-based DR approach and a benchmark approach based on Gaussian approximation on the IEEE 30-bus system. We found that the DRCC-OPF model with the Wasserstein-moment ambiguity set using statistical radius performs the best in term of balancing cost-effectiveness and reliability.

3.6 Nomenclature

A. Sets and Indices

N_B Number of buses.

N_G Number of thermal generators.

N_L Number of loads.

N_ℓ Number of transmission lines.

N_w Number of wind farms.

N_t Number of operating time intervals.

$\left[N\right]$ Set $\{1, \ldots, N\}$ for integer N.

B. Parameters

$A_\ell \in \mathbb{R}^{N_\ell \times N_B}$ The distribution factor matrix, in which each component $[A_\ell]_{ij}$ represents the shift distribution factor of bus j to line i in the dc power flow model.

$B_G \in \mathbb{R}^{N_B \times N_G}$ The bus-generator incidence matrix, in which component $[B_G]_{ij} = 1$ if generator j locates at bus i and $[B_G]_{ij} = 0$ otherwise.

$B_L \in \mathbb{R}^{N_B \times N_L}$ The bus-load incidence matrix, in which component $[B_L]_{ij} = 1$ if load j locates at bus i and $[B_L]_{ij} = 0$ otherwise.

$B_w \in \mathbb{R}^{N_B \times N_w}$ The bus-wind incidence matrix, in which component $[B_w]_{ij} = 1$ if wind farm j locates at bus i and $[B_w]_{ij} = 0$ otherwise.

$c_G \in \mathbb{R}^{N_G}$ The unit costs for energy generation.

$c_R \in \mathbb{R}^{N_G}$ The unit costs for providing reserve.

$P^f_{w,t} \in \mathbb{R}^{N_w}$ Forecasted wind power during time interval t.

$\overline{P}_G \in \mathbb{R}^{N_G}$ Thermal generator production capacities.

$P_{L,t} \in \mathbb{R}^{N_L}$ Loads during time interval t.

$P_\ell \in \mathbb{R}^{N_\ell}$ Transmission line capacities.

RU, **RD** $\in \mathbb{R}^{N_G}$ Ramp-up/-down capacities of the thermal generators.

C. Random Variables

$w_t \in \mathbb{R}^{N_w}$ Forecast errors of wind energy during time interval t.

D. Decision Variables

$d_t \in \mathbb{R}^{N_G}$ Distribution vector during time interval t, which parameterizes an affine response from the generator-providing reserves to the real-time supply/demand mismatch.

$P_{G,t} \in \mathbb{R}^{N_G}$ Planed thermal generation amounts during time interval t.

$\overline{R}_{G,t}$, $\underline{R}_{G,t} \in \mathbb{R}^{N_G}$ Generator upward/downward reserve capacities during time interval t.

E. Auxiliary Variables

$P_{I,t} \in \mathbb{R}^{N_B}$ Net power injection at buses during time interval t.

$P_{w,t} \in \mathbb{R}^{N_w}$ Actual wind energy during time interval t.

$R_t \in \mathbb{R}^{N_G}$ Generator reserve dispatch amounts during time interval t.

CHAPTER 4

Distributionally Robust Optimization Using Shape Information

4.1 Introduction

In an uncertain environment, crucial decisions often need to be made before the uncertain parameters are fully realized. If we have access to the probability distribution of these parameters, then we can adopt a classical stochastic program (SP) that optimizes the expected value of the cost:

$$z_{\text{SP}} := \min_{x} \ \mathbb{E}_{\mathbb{P}_\xi}\left[f(x,\xi)\right], \tag{SP}$$

where $x \in \mathbb{R}^n$ represents an n-dimensional decision vector, $\xi \in \mathbb{R}^T$ represents a T-dimensional random vector, $f(x,\xi)$ represents a cost function involving x and ξ, and $\mathbb{P}_\xi$ represents the probability distribution of ξ. (SP) is an effective approach and has been successfully implemented in a wide range of applications (e.g., appointment scheduling [46] and power system operations [134]). Indeed, suppose that a decision maker needs to repeatedly implement a decision in face of uncertainty that is stationary over time. Then, the objective function of (SP) becomes an unbiased representative of the long-term average cost, and as a result, an optimal solution x_{SP} to (SP) leads to a minimal cost in the long run.

A basic challenge to the (SP) model is that the decision maker may not have perfect knowledge of $\mathbb{P}_\xi$. For example, in appointment scheduling, the service durations may be non-stationary due to patient heterogeneity or seasonality; and in power system operations, the data on the new renewable energy resources may be limited. As a consequence, x_{SP} can lead to unsatisfactory performance in reality. As an alternative to SP, distributionally robust optimization (DRO) proposes to waive the perfect-knowledge assumption on $\mathbb{P}_\xi$ by considering an ambiguity set, consisting of a family of plausible distributions that satisfy certain characteristics of the true (yet ambiguous) $\mathbb{P}_\xi$. A convenient ambiguity set is constructed

based on moments, e.g.,

$$\mathcal{D} := \Big\{\mathbb{P} \in \mathcal{P}(\Xi) : \mathbb{E}_{\mathbb{P}}[A\xi] = m\Big\},$$

where Ξ represents the support of ξ, $\mathcal{P}(\Xi)$ represents the set of distributions supported on Ξ, $A \in \mathbb{R}^{L\times T}$ and $m \in \mathbb{R}^L$ respectively represent a matrix of coefficients and a vector of moments that are matched by $\mathbb{E}_{\mathbb{P}}[\xi]$ under the linear transformation defined by A. For example, if $L = T$ and $A = I_{T\times T}$ then $\mathcal{D}$ incorporates all distributions under which the mean of ξ equals m. DRO seeks to minimize the expected cost with regard to the worst-case distribution in $\mathcal{D}$, i.e.,

$$z_{\text{DRO}} := \min_{x} \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]. \tag{DRO}$$

In a data-driven context, the matched moments m can be inferred from the historical data of ξ. In that case, if ambiguity set $\mathcal{D}$ includes $\mathbb{P}_\xi$ as a member with high confidence, then z_{DRO} provides a conservative, but high-confidence, upper bound on z_{SP}, which is unknown when $\mathbb{P}_\xi$ is ambiguous.

In reality, in addition to the moments, a decision maker often possesses certain shape information of $\mathbb{P}_\xi$. For example, multiple classes of distributions have been proposed to model wind prediction errors, including normal [39], Cauchy [65], and hyperbolic [64], all of which are unimodal. Likewise, multiple classes of distributions have been proposed to model appointment durations, including exponential [145], Gamma [130], and normal [31], all of which are log-concave. Incorporation of the shape information into the definition of $\mathcal{D}$ can make (DRO) less conservative (see, e.g., [89, 87]). Unfortunately, such incorporation often makes DRO models computationally intractable. For example, when unimodality is incorporated into $\mathcal{D}$ and $f(x,\xi)$ is convex and piecewise affine in ξ, [138] indicates that the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]$ may be computationally prohibitive even when x is fixed. For another example, incorporating log-concavity may deprive the convexity of $\mathcal{D}$, making (DRO) non-convex. Indeed, the mixture of two log-concave distributions is in general not log-concave anymore.

In this chapter, we investigate a general framework to indirectly take the shape information into account. Specifically, we consider concentration inequalities in the form

$$\mathbb{P}\{\xi \in \Xi_r\} \geq 1 - \epsilon(r) \ \ \forall r \in \mathcal{R}, \tag{4.1}$$

where $\mathcal{R}$ represents a set of real numbers, $\{\Xi_r\}_{r\in\mathcal{R}}$ represent a series of nondecreasing confidence sets parameterized by r, i.e., $\Xi_{r_1} \subseteq \Xi_{r_2}$ whenever $r_1 \leq r_2$, and $1-\epsilon(r)$ represents a nondecreasing function in r on $\mathcal{R}$. In Section 4.2, we make it concrete how to infer

concentration inequalities in the above form from higher moments and shape information, including unimodality, log-concavity, and sub-Gaussian property. Using (4.1), we re-define the ambiguity set as follows and stick with this definition in the remainder of this chapter:

$$\mathcal{D} \;=\; \Big\{\mathbb{P} \in \mathcal{P}(\Xi) :\; \mathbb{E}_{\mathbb{P}}[A\xi] = m,\;\; \mathbb{P}\big\{\xi \in \Xi_r\big\} \geq 1 - \epsilon(r),\; \forall r \in \mathcal{R}\Big\}. \tag{4.2}$$

When $\mathcal{R}$ is chosen to be a continuous interval of $\mathbb{R}$, $\mathcal{D}$ becomes an *infinitely constrained* ambiguity set, potentially making (DRO) intractable. Perhaps surprisingly, we shall show that, even under an infinitely constrained $\mathcal{D}$ as in (4.2), (DRO) with shape information admits a SP reformulation with respect to an *unambiguous* probability distribution. This facilitates efficient solution algorithms, e.g., sample average approximation.

4.1.1 Literature Review

DRO was first introduced by [124] as a minimax stochastic program for the newsvendor problem under an ambiguous demand with the first two moments. Following this seminal work, moment information has been widely used for characterizing ambiguity sets in various DRO models (see, e.g., [44, 107, 25, 108, 139, 143, 17, 37, 171, 3, 172, 35, 61]). A key merit of the moment-based DRO is that the model can often be recast as tractable convex programs such as semidefinite programs (see, e.g., [107, 139, 37, 171, 35]) or conic programs (see, e.g., [44, 25, 143, 61]) that can be efficiently solved by commercial solvers. Recently, [151] successfully identified a class of moment-based ambiguity sets that lead to tractable convex program reformulations of general DRO models.

An alternative type of ambiguity sets consist of probability distributions that are within a neighborhood of a reference distribution. To define the neighborhood, various divergences including the Prohorov metric [45], the ϕ-divergence (see, e.g., [14, 11, 73]), the ζ-structure probability metrics [165], and the Wasserstein distance (see, e.g., [47]) have been employed in DRO models. Based on a series of data of ξ, the divergence-based DRO can asymptotically recover the (SP) model if the ambiguity set shrinks to the true distribution $\mathbb{P}_\xi$, e.g., as the data size increases to infinity. Nevertheless, the divergence-based ambiguity sets usually rely on *joint* data of ξ, i.e., all components of ξ need to be realized in all data samples. In many applications, only *marginal* data is available due to asynchronous data collection and/or data privacy consideration. In that case, it is unclear how divergence-based ambiguity sets can be constructed, while the (marginal) moments of ξ can still be inferred from the data.

As compared to the aforementioned ambiguity sets, shape information has received much less attention in the existing literature. For example, [137] generalized the Gauss in-

equality when $\mathbb{P}_\xi$ is α-unimodal, [58] generalized the newsvendor model in [124] to a multi-item setting under a multi-modal ambiguity set in the sense that $\mathbb{P}_\xi$ is a mixture of multiple distributions with known moments, [131, 86] studied distributionally robust chance constraints when $\mathbb{P}_\xi$ is α-unimodal, [138] considered ambiguity sets with α-unimodality or γ-monotonicity in distributionally robust expectation inequalities, and [89] incorporated log-concavity into distributionally robust chance constraints. To the best of our knowledge, this study takes the first step to study DRO models with shape information in a general and computationally tractable manner. We highlight our main contributions as follows:

1. We propose a new class of ambiguity sets using shape information. We derive concentration inequalities from various shape information and incorporate these inequalities to strengthen the ambiguity set. This provides a general framework for DRO models with shape information.

2. We derive an upper bound of the worst-case expectation under the new ambiguity set. Under mild conditions, we show that this upper bound is tight for a wide class of cost functions $f(x, \xi)$, both when $\mathcal{D}$ is finitely constrained and when it is infinitely constrained. In addition, we show that the (DRO) model is computationally tractable when the worst-case expectation is replaced by the tight upper bound.

3. We further extend the framework in multiple directions, including adaptive DRO models with fixed recourse and cost functions that violate the identified conditions. In particular, we show that these models admit tractable reformulations when shape information is taken into account.

Notation. Throughout this chapter, we adopt the convention that $\infty \cdot 0 = 0 \cdot \infty = 0/0 = 0$. For integers $M, N \in \mathbb{N}$, we denote $[N] := \{1, \ldots, N\}$ and $[M, N]_{\mathbb{Z}} := \{n \in \mathbb{N} : M \leq n \leq N\}$. Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, its dual norm is defined as $\|x\|_* := \sup_{\xi:\|\xi\|\leq 1} x^\top \xi$. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, its conjugate $f^* : \mathbb{R}^n \to \mathbb{R}$ is defined as $f^*(x) := \sup_{\xi\in\mathbb{R}^n} \left\{x^\top \xi - f(\xi)\right\}$. For a set $\Xi \subseteq \mathbb{R}^n$, its indicator function $\mathbb{1}_\Xi : \mathbb{R}^n \to \{0, 1\}$ is defined as $\mathbb{1}_\Xi(\xi) := \begin{cases} 1, & \text{if } \xi \in \Xi, \\ 0, & \text{otherwise} \end{cases}$, the characteristic function $\chi_\Xi : \mathbb{R}^n \to \{0, \infty\}$ is defined as $\chi_\Xi(\xi) := \begin{cases} 0, & \text{if } \xi \in \Xi \\ \infty, & \text{otherwise} \end{cases}$, and the support function $\sigma_\Xi : \mathbb{R}^n \to \mathbb{R}$ is defined as $\sigma_\Xi(x) := \sup_{\xi\in\Xi} x^\top \xi$.

Organization. The remainder of this chapter is organized as follows. In Section 4.2, we discuss how the concentration inequalities in $\mathcal{D}$ can be inferred from higher moments and

shape information. In Section 4.3, we derive an upper bound of the worst-case expectation in (DRO) for two cases: (i) $\mathcal{R}$ is discrete and (ii) $\mathcal{R}$ is a continuous interval. In Section 4.4, we show that this upper bound is sharp for a wide class of cost functions. Then, we study extensions of the proposed framework in Section 4.5. Finally, we conduct computational case studies on appointment scheduling and optimal power flow in Section 4.6.

4.2 Inferring Concentration Inequalities

In this section, we discuss how the concentration inequalities (4.1) can be inferred from higher moments and shape information.

Higher moments: Suppose that we know the k^{th} marginal moments $m_{kt} := \mathbb{E}_{\mathbb{P}_\xi}[|\xi_t|^k]$ for some $k \geq 2$ and all $t \in [T]$. Then, the Markov's inequality implies that $\mathbb{P}_\xi\{|\xi_t| > r\} \leq m_{kt}/r^k$ for all $r > 0$. It follows that

$$\mathbb{P}_\xi\{\|\xi\|_\infty \leq r\} \geq 1 - \sum_{t=1}^{T} \frac{m_{kt}}{r^k}, \ \ \forall r > 0.$$

In other words, these k^{th} moments infer concentration inequalities (4.1) with $\Xi_r = \{\xi : \|\xi\|_\infty \leq r\}$, $\epsilon(r) = \sum_{t=1}^{T}(m_{kt}/r^k)$, and $\mathcal{R} = (0, \infty)$.

Unimodality: Suppose that, for all $t \in [T]$, the t^{th} component of ξ is unimodal about a mode o_t. Then, the Gauss inequality implies that $\mathbb{P}_\xi\{|\xi_t - o_t| > r\} \leq (2s_t/3r)^2$ for all $r \geq 2s_t/\sqrt{3}$, where $s_t^2 := \mathbb{E}_{\mathbb{P}_\xi}[(\xi_t - o_t)^2]$. It follows that

$$\mathbb{P}_\xi\{\|\xi - o\|_\infty \leq r\} \geq 1 - \sum_{t=1}^{T} \left(\frac{2s_t}{3r}\right)^2, \ \ \forall r \geq \max_{t\in[T]} \left\{\frac{2s_t}{\sqrt{3}}\right\}.$$

In other words, the unimodality infers concentration inequalities (4.1) with $\Xi_r = \{\xi : \|\xi - o\|_\infty \leq r\}$, $\epsilon(r) = \sum_{t=1}^{T}(2s_t/(3r))^2$, and $\mathcal{R} = \big[\max_{t\in[T]}\{2s_t/\sqrt{3}\}, \ \infty\big)$.

Star-unimodality: Suppose that ξ is star-unimodal[1] about its mean μ. Then, a generalized Gauss inequality by Lemma 4 in [137] implies that

$$\mathbb{P}_\xi\left\{\|\Sigma^{-1/2}(\xi - \mu)\|_\infty \leq \sqrt{T}r\right\} \geq 1 - \left(\frac{2}{T+2}\right)^{\frac{2}{T}} \frac{1}{r^2}, \ \ \forall r > \left(\frac{2}{T+2}\right)^{\frac{1}{T}} \left(\frac{T+2}{T}\right)^{\frac{1}{2}},$$

[1] Star-unimodality is a generalized concept of unimodality for random vectors; see [38].

where Σ represents the covariance matrix of ξ. In other words, the star-unimodality infers concentration inequalities (4.1) with $\Xi_r = \{\xi : \|\Sigma^{-1/2}(\xi - \mu)\|_\infty \leq \sqrt{T}r\}$, $\epsilon(r) = (2/(T+2))^{2/T}(1/r^2)$, and $\mathcal{R} = \big((2/(T+2))^{1/T}(1+2/T)^{1/2},\ \infty\big)$.

Log-concavity: Suppose that ξ is log-concave and $\|\cdot\|$ represent an arbitrary norm on $\mathbb{R}^T$. Then, Proposition 5.15 in [56] implies that

$$\mathbb{P}_\xi\{\|\xi\| \geq 4\Gamma r\} \leq e^{-\frac{r}{2}},\ \ \forall r \geq 1,$$

where $\Gamma := \mathbb{E}_{\mathbb{P}_\xi}[\|\xi\|]$. This is equivalent to concentration inequalities (4.1) with $\Xi_r = \{\xi : \|\xi\| \leq 4\Gamma r\}$, $\epsilon(r) = e^{-r/2}$, and $\mathcal{R} = [1, \infty)$.

Sub-Gaussian: Suppose that each component of ξ is sub-Gaussian, i.e., there exists constants $c_t \geq 1$ and $\tau_t \in \mathbb{R}$ such that $\mathbb{P}_\xi\{|\xi_t - \mu_t| \geq r\} \leq c_t\mathbb{P}\{|\mathcal{N}(\mu_t, \tau_t^2) - \mu_t| \geq r\}$ for all $t \in [T]$ and $r \geq 0$, where $\mu := \mathbb{E}_{\mathbb{P}_\xi}[\xi]$ and $\mathcal{N}(\mu_t, \tau_t^2)$ represents the normal random variable with mean μ_t and variance τ_t^2. It follows that

$$\mathbb{P}_\xi\{\|\xi - \mu\|_\infty \leq r\} \geq 1 - \sum_{t=1}^{T} c_t\mathbb{P}\{|\mathcal{N}(\mu_t, \tau_t^2) - \mu_t| \geq r\},\ \ \forall r \geq 0.$$

This is equivalent to concentration inequalities (4.1) with $\Xi_r = \{\xi : \|\xi - \mu\|_\infty \leq r\}$, $\epsilon(r) = \sum_{t=1}^{T} c_t\mathbb{P}\{|\mathcal{N}(\mu_t, \tau_t^2) - \mu_t| \geq r\}$, and $\mathcal{R} = [0, \infty)$.

Sub-Gaussian and independence: An equivalent definition of ξ_t being sub-Gaussian is that $\|\xi_t - \mu_t\|_{\psi_2} := \sup_{p \geq 1} p^{-1/2}\left(\mathbb{E}_{\mathbb{P}_\xi}|\xi_t - \mu_t|^p\right)^{1/p} < \infty$. In addition to the sub-Gaussian information, suppose that the components of ξ are independent and $s_t^2 := \mathbb{E}_{\mathbb{P}_\xi}[(\xi_t - \mu_t)^2]$. Then, the Hanson-Wright inequality (see, e.g., [120]) implies

$$\mathbb{P}_\xi\left\{(\xi - \mu)^\top\Sigma(\xi - \mu) \leq \sum_{t=1}^{T} s_t^2\Sigma_{t,t} + r\right\} \geq 1 - 2\exp\left[-c\min\left\{\frac{r^2}{\kappa^4\|\Sigma\|_{\text{F}}^2},\ \frac{r}{\kappa^2\|\Sigma\|}\right\}\right],$$

$\forall r \geq 0$, for any matrix $\Sigma \in \mathbb{R}^{T\times T}$, and a constant $c > 0$, where $\kappa := \max_{t\in[T]}\|\xi_t - \mu_t\|_{\psi_2}$, $\|\Sigma\|_{\text{F}} := \left(\sum_{r=1}^{T}\sum_{t=1}^{T}\Sigma_{r,t}^2\right)^{1/2}$ represents the Frobenius norm, and $\|\Sigma\| := \sup_{\|\xi\|_2\leq 1}\|\Sigma\xi\|_2$ represents the operator norm. In other words, sub-Gaussian and independence information infers concentration inequalities (4.1) with $\Xi_r = \{\xi : (\xi-\mu)^\top\Sigma(\xi-\mu) \leq \sum_{t=1}^{T} s_t^2\Sigma_{t,t}+r\}$, $\epsilon(r) = 2\exp\big[-c\min\big\{r^2/(\kappa^4\|\Sigma\|_{\text{F}}^2),\, r/(\kappa^2\|\Sigma\|)\big\}\big]$, and $\mathcal{R} = [0, \infty)$.

4.3 Upper Bounds of the Worst-Case Expectation

In this section, we derive upper bounds of the worst-case expectation in (DRO) for a general cost function $f(x, \xi)$. In particular, we consider two topologies of $\mathcal{R}$: (i) $\mathcal{R}$ is a discrete set of scalars and (ii) $\mathcal{R}$ is a continuous interval on $\mathbb{R}$. For notational brevity, we parameterize the support set $\Xi := \Xi_{\bar{r}}$ such that $\bar{r} \geq \sup\{\mathcal{R}\}$. We formalize the discrete case in the following assumption.

Assumption 4.3.1 (Discrete Topology)**.** $\mathcal{R} = \{r_0, \ldots, r_K\}$ *with* $r_0 < \cdots < r_K$*. In addition,* $\Xi_{r_k} \subsetneqq \Xi$ *and* $\Xi_{r_k} \setminus \Xi_{r_{k-1}} \neq \emptyset$ *for all* $k \in [K]$.

We note that Assumption 4.3.1 can be made with little loss of generality. Indeed, we can always redefine Ξ_r to be $\Xi_r \cap \Xi$ and/or enlarge Ξ infinitesimally to make Ξ_r a strict subset of Ξ. In addition, we can merge the duplicated confidence sets and their confidence bounds to ensure that $\{\Xi_{r_k}\}_{k=0}^K$ is strictly increasing under inclusion. We summarize the main result for the discrete topology in the following theorem.

Theorem 4.3.1. *Under Assumption 4.3.1, we have*

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] \;\leq\; \inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}\Big[H(x,p,\zeta)\Big],$$

where ζ *represents a* 1*-dimensional random variable,*

$$H(x,p,\zeta) \;:=\; \sup_{\xi\in\Xi_\zeta}\big\{f(x,\xi) - p^\top(A\xi - m)\big\}, \tag{4.3}$$

and $\mathbb{Q}_K$ *represents its probability distribution with the cumulative distribution function*

$$\mathbb{Q}_K\{\zeta \leq x\} \;=\; \begin{cases} 0, & \text{if } x < r_0, \\ 1-\epsilon(r_{k-1}), & \text{if } r_{k-1} \leq x < r_k \text{ for } k \in [K], \\ 1-\epsilon(r_K), & \text{if } r_K \leq x < \bar{r}, \\ 1, & \text{if } x \geq \bar{r}. \end{cases}$$

Proof. First, we formulate the distributionally robust expectation in (DRO) as the following optimization problem

$$\begin{aligned} z_P^K \;=\; \sup_{\mathbb{P}} \;& \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] && \text{(4.4a)} \\ \text{s.t.} \;& \mathbb{E}_{\mathbb{P}}[A\xi] = m, && \text{(4.4b)} \\ & \mathbb{E}_{\mathbb{P}}\big[-\mathbb{1}_{\Xi_{r_k}}(\xi)\big] \;\leq\; \epsilon(r_k) - 1, \;\; \forall k \in [0,K]_{\mathbb{Z}}, && \text{(4.4c)} \end{aligned}$$

$$\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}_{\Xi}(\xi)\big] \;=\; 1, \tag{4.4d}$$

whose dual is

$$z_D^K \;=\; \inf_{p\in\mathbb{R}^T,\, q\in\mathbb{R},\, g_k\geq 0} \; m^\top p + q - \sum_{k=0}^{K} \left[1-\epsilon(r_k)\right] g_k \tag{4.5a}$$

$$\text{s.t. } \; q + p^\top A\xi - \sum_{k=0}^{K} \mathbb{1}_{\Xi_{r_k}}(\xi) g_k \geq f(x,\xi), \;\; \forall \xi\in\Xi, \tag{4.5b}$$

where dual variables p are associated with primal constraints (4.4b), q is associated with primal constraint (4.4d), g_k is associated with primal constraint (4.4c) for each $k\in[0,K]_{\mathbb{Z}}$, and dual constraints (4.5b) are associated with primal variables $\mathbb{P}$. By construction, weak duality holds between problems (4.4a)–(4.4d) and (4.5a)–(4.5b), i.e., $z_P^K \leq z_D^K$.

Second, we categorize the value of r and recast constraints (4.5b) as follows:

$$\begin{cases} q + p^\top A\xi - \sum_{k=0}^{K} g_k \geq f(x,\xi), & \forall \xi\in\Xi_{r_0}, \\ q + p^\top A\xi - \sum_{k=j}^{K} g_k \geq f(x,\xi), & \forall \xi\in\Xi_{r_j}\setminus\Xi_{r_{j-1}},\; \forall j\in[K], \\ q + p^\top A\xi \geq f(x,\xi), & \forall \xi\in\Xi\setminus\Xi_{r_K}, \end{cases}$$

$$\Leftrightarrow \; \begin{cases} q - \sum_{k=0}^{K} g_k \geq \sup\limits_{\xi\in\Xi_{r_0}} \left\{f(x,\xi) - p^\top A\xi\right\}, \\ q - \sum_{k=j}^{K} g_k \geq \sup\limits_{\xi\in\Xi_{r_j}\setminus\Xi_{r_{j-1}}} \left\{f(x,\xi) - p^\top A\xi\right\}, \; \forall j\in[K], \\ q \geq \sup\limits_{\xi\in\Xi\setminus\Xi_{r_K}} \left\{f(x,\xi) - p^\top A\xi\right\}, \end{cases} \tag{4.6a}$$

$$\Leftrightarrow \; \begin{cases} q - \sum_{k=0}^{K} g_k \geq \sup\limits_{\xi\in\Xi_{r_0}} \left\{f(x,\xi) - p^\top A\xi\right\}, \\ q - \sum_{k=j}^{K} g_k \geq \sup\limits_{\xi\in\Xi_{r_j}} \left\{f(x,\xi) - p^\top A\xi\right\}, \; \forall j\in[K], \\ q \geq \sup\limits_{\xi\in\Xi} \left\{f(x,\xi) - p^\top A\xi\right\}, \end{cases} \tag{4.6b}$$

$$\Leftrightarrow \; q - \sum_{k=j}^{K} g_k \geq \bar{H}(r_j), \;\; \forall j\in[0,K+1]_{\mathbb{Z}},$$

where $\bar{H}(t) := \sup_{\xi\in\Xi_t}\Big\{f(x,\xi) - p^\top A\xi\Big\}$ for all $t\in\mathcal{R}$ and $\Xi_{r_{K+1}} := \Xi$. The equivalence between (4.6a) and (4.6b) is because $g_k\geq 0$ and so, for any $i,j\in[K]$ and $i<j$, we have $q-\sum_{k=j}^{K} g_k \geq q-\sum_{k=i}^{K} g_k \geq \sup_{\xi\in\Xi_{r_i}\setminus\Xi_{r_{i-1}}} \left\{f(x,\xi) - p^\top A\xi\right\}$. Similarly, $q-\sum_{k=j}^{K} g_k \geq$

$q - \sum_{k=0}^{K} g_k \geq \sup_{\xi \in \Xi_{r_0}} \left\{ f(x,\xi) - p^\top A \xi \right\}$. It follows that

$$\begin{aligned} q - \sum_{k=j}^{K} g_k \;&\geq\; \max_{i \in [j]} \left\{ \sup_{\xi \in \Xi_{r_i} \setminus \Xi_{r_{i-1}}} \left\{ f(x,\xi) - p^\top A \xi \right\}, \; \sup_{\xi \in \Xi_{r_0}} \left\{ f(x,\xi) - p^\top A \xi \right\} \right\} \\ &= \sup_{\xi \in \Xi_{r_j}} \left\{ f(x,\xi) - p^\top A \xi \right\}. \end{aligned}$$

We recast the dual problem (4.5a)–(4.5b) as

$$\inf_{p \in \mathbb{R}^T, \; q \in \mathbb{R}, \; g_k \geq 0} \; m^\top p + q - \sum_{k=0}^{K} [1 - \epsilon(r_k)] g_k \tag{4.6c}$$

$$\text{s.t.} \quad q - \sum_{k=j}^{K} g_k \geq \bar{H}(r_j), \;\; \forall j \in [0, K+1]_{\mathbb{Z}}, \tag{4.6d}$$

$$= \inf_{p \in \mathbb{R}^T, \; q \geq \bar{H}(\bar{r})} \; m^\top p + q - \sup_{g_k \geq 0} \; \sum_{k=0}^{K} [1 - \epsilon(r_k)] g_k \tag{4.6e}$$

$$\text{s.t.} \;\; \sum_{k=j}^{K} g_k \leq q - \bar{H}(r_j), \;\; \forall j \in [0, K]_{\mathbb{Z}}. \tag{4.6f}$$

Third, we solve the maximization problem embedded in the formulation (4.6e)–(4.6f). To this end, we conduct a change of variables by defining $G_j = \sum_{k=j}^{K} g_k$ for all $j \in [0, K]_{\mathbb{Z}}$. It follows that $g_K = G_K$ and $g_k = G_k - G_{k+1}$ for all $k \in [0, K-1]_{\mathbb{Z}}$, i.e., it is equivalent to work with either variables g_k or G_k in (4.6e)–(4.6f). Hence, we recast the maximization problem as:

$$\sup_{G_k \geq 0} \; [1 - \epsilon(r_0)] G_0 + \sum_{k=1}^{K} [\epsilon(r_{k-1}) - \epsilon(r_k)] G_k \tag{4.6g}$$

$$\text{s.t.} \;\; G_k \leq q - \bar{H}(r_k), \;\; \forall k \in [0, K]_{\mathbb{Z}}, \tag{4.6h}$$

$$G_0 \geq G_1 \geq \cdots \geq G_K \geq 0. \tag{4.6i}$$

Suppose that we relax constraints (4.6i). Then, $G_k = q - \bar{H}(r_k)$ for all $k = 0, \ldots, K$ is optimal to the above problem because $1 - \epsilon(r_0) \geq 0$ and $\epsilon(r_{k-1}) - \epsilon(r_k) \geq 0$ for all $k \in [K]$. But $q - \bar{H}(r_k)$ is nonnegative because $q \geq \bar{H}(\bar{r})$, and nonincreasing in k because $\bar{H}(t)$ is nondecreasing in t and $r_0 < \cdots < r_K$. Hence, constraints (4.6i) are automatically satisfied and so the optimal value of the problem (4.6g)–(4.6i) is $[1 - \epsilon(r_0)] \, [q - \bar{H}(r_0)] + \sum_{k=1}^{K} [\epsilon(r_{k-1}) - \epsilon(r_k)] \, [q - \bar{H}(r_k)]$.

Finally, we plug this optimal value into (4.6e) and recast the dual problem (4.5a)–(4.5b)

as

$$\begin{aligned}
&\inf_{p\in\mathbb{R}^T,\ q\geq\bar{H}(\overline{r})} \ m^\top p + q - [1-\epsilon(r_0)][q-\bar{H}(r_0)] - \sum_{k=1}^{K}[\epsilon(r_{k-1})-\epsilon(r_k)][q-\bar{H}(r_k)] \\
= &\inf_{p\in\mathbb{R}^T,\ q\geq\bar{H}(\overline{r})} \ m^\top p + \epsilon(r_K)q + [1-\epsilon(r_0)]\ \bar{H}(r_0) + \sum_{k=1}^{K}[\epsilon(r_{k-1})-\epsilon(r_k)]\bar{H}(r_k) \\
= &\inf_{p\in\mathbb{R}^T} \ m^\top p + [1-\epsilon(r_0)]\ \bar{H}(r_0) + \sum_{k=1}^{K}[\epsilon(r_{k-1})-\epsilon(r_k)]\bar{H}(r_k) + \epsilon(r_K)\bar{H}(\overline{r}) \\
= &\inf_{p\in\mathbb{R}^T} \ \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)].
\end{aligned}$$

Therefore, $z_P^K \leq z_D^K = \inf_{p\in\mathbb{R}^T}\{\mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)]\}$ and the proof is completed. □

Theorem 4.3.1 upper bounds the worst-case expectation in (DRO) by the optimal value of a stochastic program with an *unambiguous* probability distribution. As a result, a reliable upper bound of the worst-case expectation can be obtained as long as we are able to solve the unambiguous stochastic program. In Section 4.4, we shall show that the upper bound is actually sharp for a wide class of cost functions. The continuous case, however, is more complicated than the discrete one. This is largely because $\mathcal{R}$ now incorporates an uncountable number of elements. Accordingly, the ambiguity set $\mathcal{D}$ becomes (uncountably) infinitely constrained. Nevertheless, we identify the following mild condition under which a counterpart of Theorem 4.3.1 holds for the continuous case.

Assumption 4.3.2 (Continuous Topology)**.** *$\mathcal{R} = [r_L, r_U]$ with $\Xi_r \subsetneq \Xi$, $\Xi_r \setminus (\bigcup_{t<r}\Xi_t) \neq \emptyset$, and $\lim_{t\to r} D(\Xi_t,\Xi_r) = 0$ for all $r \in \mathcal{R}$, where $D(A,B) := \max\{\sup_{\xi\in A}\inf_{\xi'\in B}\|\xi-\xi'\|, \sup_{\xi'\in B}\inf_{\xi\in A}\|\xi-\xi'\|\}$ represents the Hausdorff distance between sets A and B. In addition, $\epsilon(r)$ is continuous on $\mathcal{R}$. Finally, for all x, $f(x,\xi)$ is piecewise continuous in ξ. That is, there exists a partition $\{P_n\}_{n=1}^N$ of Ξ, such that $P_m \cap P_n = \emptyset$ for all $m,n \in [N]$ and $m \neq n$, $\cup_{n=1}^N P_n = \Xi$, and $f(x,\xi)$ is continuous in ξ on P_n for all $n \in [N]$.*

Assumption 4.3.2 is mild because, like in the discrete topology, we can always enlarge Ξ infinitesimally to make it a strict superset of all ξ_r. In addition, all concentration inequalities discussed in Section 4.2 satisfy the assumptions on $\{\Xi_r\}_{r\in\mathcal{R}}$. We summarize the main result for the continuous topology in the following theorem.

Theorem 4.3.2. *Under Assumption 4.3.2, we have*

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] \ \leq \ \inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}}\Big[H(x,p,\zeta)\Big],$$

where ζ represents a 1-dimensional random variable with probability distribution $\mathbb{Q}$ and

cumulative distribution function

$$\mathbb{Q}\{\zeta \le x\} \;=\; \begin{cases} 0, & \textit{if } x < r_{L}, \\ 1-\epsilon(x), & \textit{if } r_{L} \le x \le r_{U}, \\ 1-\epsilon(r_{U}), & \textit{if } r_{U} < x < \overline{r}, \\ 1, & \textit{if } x \ge \overline{r}, \end{cases}$$

and $H(x,p,\zeta)$ *is defined in* (4.3).

Proof. First, we formulate the distributionally robust expectation in (DRO) as the following optimization problem

$$z_P \;=\; \sup_{\mathbb{P}} \; \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] \tag{4.7a}$$

$$\text{s.t. } \mathbb{E}_{\mathbb{P}}[A\xi] = m, \tag{4.7b}$$

$$\mathbb{E}_{\mathbb{P}}\Big[-\mathbb{1}_{\Xi_r}(\xi)\Big] \;\le\; \epsilon(r) - 1, \;\; \forall r \in [r_{\mathrm{L}}, r_{\mathrm{U}}], \tag{4.7c}$$

$$\mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}_{\Xi}(\xi)\Big] \;=\; 1, \tag{4.7d}$$

whose dual is

$$z_D \;=\; \inf_{\substack{p\in\mathbb{R}^T,\, q\in\mathbb{R},\\ g(r)\ge 0}} \; m^{\top}p + q - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} [1-\epsilon(r)]g(r)dr \tag{4.8a}$$

$$\text{s.t. } q + p^{\top}A\xi - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} \mathbb{1}_{\Xi_r}(\xi)g(r)dr \ge f(x,\xi), \;\; \forall \xi \in \Xi, \tag{4.8b}$$

where dual variables p are associated with primal constraints (4.7b), q is associated with primal constraint (4.7d), $g(r)$ is associated with primal constraint (4.7c) for each $r \in [r_{\mathrm{L}}, r_{\mathrm{U}}]$, and dual constraints (4.8b) are associated with primal variables $\mathbb{P}$. By construction, weak duality holds between problems (4.7a)–(4.7d) and (4.8a)–(4.8b), i.e., $z_P \le z_D$.

Second, we categorize the value of r and recast constraints (4.8b) as follows:

$$q + p^{\top}A\xi - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} \mathbb{1}_{\Xi_r}(\xi)g(r)dr \ge f(x,\xi), \;\; \forall \xi \in \Xi_{r_{\mathrm{L}}}, \tag{4.9a}$$

$$q + p^{\top}A\xi - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} \mathbb{1}_{\Xi_r}(\xi)g(r)dr \ge f(x,\xi), \;\; \forall r_{\mathrm{L}} \le t \le r_{\mathrm{U}}, \;\; \forall \xi \in \Xi_t \setminus \left(\bigcup_{s<t} \Xi_s\right), \tag{4.9b}$$

$$q + p^{\top}A\xi - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} \mathbb{1}_{\Xi_r}(\xi)g(r)dr \ge f(x,\xi), \;\; \forall \xi \in \Xi \setminus \Xi_{r_{\mathrm{U}}}. \tag{4.9c}$$

For inequalities (4.9a), as $\xi \in \Xi_{r_{\mathrm{L}}}$, we have $\mathbb{1}_{\Xi_r}(\xi) = 1$ for all $r \in [r_{\mathrm{L}}, r_{\mathrm{U}}]$ and so (4.9a)

reduces to

$$q - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} g(r)dr \geq \bar{H}(r_{\mathrm{L}}), \tag{4.9d}$$

where function $\bar{H}(t)$ is defined as $\bar{H}(t) := \sup_{\xi \in \Xi_t} \Big\{ f(x,\xi) - p^\top A\xi \Big\}$. Applying similar analyses on constraints (4.9b)–(4.9c) yields the following reformulation

$$q - \int_{t}^{r_{\mathrm{U}}} g(r)dr \geq \sup_{\xi \in \Xi_t \setminus (\bigcup_{s<t} \Xi_s)} \Big\{ f(x,\xi) - p^\top A\xi \Big\}, \ \ \forall r_{\mathrm{L}} \leq t \leq r_{\mathrm{U}}, \tag{4.9e}$$

$$q \geq \sup_{\xi \in \Xi \setminus \Xi_{r_{\mathrm{U}}}} \Big\{ f(x,\xi) - p^\top A\xi \Big\}, \tag{4.9f}$$

for (4.9b) and (4.9c), respectively. Now consider any $s, t \in [r_{\mathrm{L}}, r_{\mathrm{U}}]$ with $s \leq t$. As $g(r) \geq 0$ for all $r \in [r_{\mathrm{L}}, r_{\mathrm{U}}]$, we have $q - \int_t^{r_{\mathrm{U}}} g(r)dr \geq q - \int_s^{r_{\mathrm{U}}} g(r)dr$ and so $q - \int_t^{r_{\mathrm{U}}} g(r)dr \geq \sup_{\xi \in \Xi_s \setminus (\bigcup_{u<s} \Xi_u)} \big\{ f(x,\xi) - p^\top A\xi \big\}$ by (4.9e). It follows that $q - \int_t^{r_{\mathrm{U}}} g(r)dr \geq \sup_{\xi : \Xi_t \setminus \Xi_{r_{\mathrm{L}}}} \big\{ f(x,\xi) - p^\top A\xi \big\}$. In addition, $q - \int_t^{r_{\mathrm{U}}} g(r)dr \geq q - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} g(r)dr \geq \bar{H}(r_{\mathrm{L}})$ by (4.9d). Hence, we recast constraints (4.9e) as

$$q - \int_{t}^{r_{\mathrm{U}}} g(r)dr \geq \bar{H}(t), \ \ \forall t \in [r_{\mathrm{L}}, r_{\mathrm{U}}]. \tag{4.9g}$$

Furthermore, as $q \geq \bar{H}(r_{\mathrm{U}})$ by (4.9g) with $t = r_{\mathrm{U}}$, we recast constraints (4.9f) as

$$q \geq \bar{H}(\bar{r}). \tag{4.9h}$$

To sum up, we have recast the dual problem (4.8a)–(4.8b) as

$$\begin{aligned}
&\inf_{p \in \mathbb{R}^T,\ q \in \mathbb{R},\ g(r) \geq 0} \ m^\top p + q - \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} [1 - \epsilon(r)] g(r) dr \\
&\qquad\qquad \text{s.t.} \ \ \text{(4.9g)–(4.9h)}, \\
=\ &\inf_{p \in \mathbb{R}^T,\ q \geq \bar{H}(\bar{r})} \ m^\top p + q - \sup_{g(r) \geq 0} \ \int_{r_{\mathrm{L}}}^{r_{\mathrm{U}}} [1 - \epsilon(r)] g(r) dr && \text{(4.10a)} \\
&\qquad\qquad\qquad\qquad \text{s.t.} \ \ \int_{t}^{r_{\mathrm{U}}} g(r)dr \leq q - \bar{H}(t), \ \ \forall r_{\mathrm{L}} \leq t \leq r_{\mathrm{U}}. && \text{(4.10b)}
\end{aligned}$$

Third, we define $G(t) = \int_t^{r_{\mathrm{U}}} g(r)dr$ for $t \in [r_{\mathrm{L}}, r_{\mathrm{U}}]$. Then, $G(t)$ is continuous, non-increasing in t on the interval $[r_{\mathrm{L}}, r_{\mathrm{U}}]$, and $G(r_{\mathrm{U}}) = 0$. Then, we recast the maximization

problem in (4.10a)–(4.10b) as

$$\sup_{G(r)\geq 0} \quad -\int_{r_{\rm L}}^{r_{\rm U}} [1-\epsilon(r)] dG(r) \tag{4.10c}$$

$$\text{s.t. } G(t) \leq q - \bar{H}(t), \ \ \forall r_{\rm L} \leq t \leq r_{\rm U}, \tag{4.10d}$$

$$G(r_{\rm U}) = 0, \tag{4.10e}$$

$$G(t) \text{ continuous and nonincreasing in } t \text{ on } [r_{\rm L}, r_{\rm U}]. \tag{4.10f}$$

We show that the optimal value of problem (4.10c)–(4.10f) is $[1-\epsilon(r_{\rm L})][q-\bar{H}(r_{\rm L})] + \int_{r_{\rm L}}^{r_{\rm U}} [q-\bar{H}(r)] d[1-\epsilon(r)]$. To this end, on the one hand, the objective function (4.10c) satisfies

$$\begin{aligned} -\int_{r_{\rm L}}^{r_{\rm U}} [1-\epsilon(r)]\, dG(r) \ &= [1-\epsilon(r_{\rm L})]G(r_{\rm L}) + \int_{r_{\rm L}}^{r_{\rm U}} G(r)\, d[1-\epsilon(r)] \\ &\leq [1-\epsilon(r_{\rm L})]\, [q-\bar{H}(r_{\rm L})] + \int_{r_{\rm L}}^{r_{\rm U}} [q-\bar{H}(r)]\, d[1-\epsilon(r)], \end{aligned}$$

where the equality follows from integration by parts and the inequality follows from constraint (4.10d). On the other hand, we construct a sequence of functions $\{G^n(r)\}_{n\in\mathbb{N}}$ such that

$$G^n(r) \ = \ \begin{cases} q-\bar{H}(r) & \text{if } r_{\rm L} \leq r \leq r_{\rm U} - \frac{1}{n}, \\ -n(r-r_{\rm U})\big[q-\bar{H}(r)\big] & \text{if } r_{\rm U} - \frac{1}{n} \leq r \leq r_{\rm U}. \end{cases}$$

We prove that $\{G^n(r)\}_{n\in\mathbb{N}}$ are a sequence of feasible solutions to problem (4.10c)–(4.10f) with their objective values converging to the claimed optimal value. To avoid clutter, we assume that $\bar{H}(t)$ is continuous on the interval $[r_{\rm L}, r_{\rm U}]$. For the general $\bar{H}(t)$, we can prove the same claim by modifying the above definition of $G^n(r)$ within the neighborhood of each discontinuous point of $\bar{H}(t)$ as that within the interval $[r_{\rm U} - \frac{1}{n}, r_{\rm U}]$. Under this assumption, when $n > 1/(r_{\rm U} - r_{\rm L})$, $G^n(r)$ is continuous, nonincreasing in r, $G^n(r_{\rm U}) = 0$, and $G^n(r) \leq q - \bar{H}(r)$ for all $r_{\rm L} \leq r \leq r_{\rm U}$. Hence, $G^n(r)$ is a feasible solution to problem (4.10c)–(4.10f). It follows that, for all n sufficiently large, the optimal value of this problem is bounded from below by

$$\begin{aligned} & -\int_{r_{\rm L}}^{r_{\rm U}} \big[1-\epsilon(r)\big]\, dG^n(r) \\ = \ & -\int_{r_{\rm L}}^{r_{\rm U}-1/n} \big[1-\epsilon(r)\big]\, d\big[q-\bar{H}(r)\big] - \int_{r_{\rm U}-1/n}^{r_{\rm U}} \big[1-\epsilon(r)\big]\, d\Big[-n(r-r_{\rm U})\big[q-\bar{H}(r)\big]\Big] \\ = \ & -\int_{r_{\rm L}}^{r_{\rm U}-1/n} \big[1-\epsilon(r)\big]\, d\big[q-\bar{H}(r)\big] + \Big[1-\epsilon\Big(r_{\rm U}-\frac{1}{n}\Big)\Big]\Big[q-\bar{H}\Big(r_{\rm U}-\frac{1}{n}\Big)\Big] \end{aligned}$$

$$+ \int_{r_{\rm U}-1/n}^{r_{\rm U}} n\big[q - \bar{H}(r)\big]\big(r_{\rm U} - r\big)\, d\big[1-\epsilon(r)\big], \tag{4.10g}$$

where the second equality follows from integration by parts. We analyze the convergence of the three terms in (4.10g) as $n \to \infty$. Specifically, $-\int_{r_{\rm L}}^{r_{\rm U}-1/n}\big[1-\epsilon(r)\big]\, d\big[q-\bar{H}(r)\big] \to -\int_{r_{\rm L}}^{r_{\rm U}}\big[1-\epsilon(r)\big]\, d\big[q-\bar{H}(r)\big]$ by the monotone convergence theorem, because of the monotonicity of functions $\mathbb{1}[r_{\rm L} \le r \le r_{\rm U} - 1/n](r)\ [1-\epsilon(r)]$ as n increases. In addition, $\big[1-\epsilon\big(r_{\rm U}-1/n\big)\big]\big[q-\bar{H}\big(r_{\rm U}-1/n\big)\big] \to \big[1-\epsilon\big(r_{\rm U}\big)\big]\big[q-\bar{H}\big(r_{\rm U}\big)\big]$ because of the left continuity of $\epsilon(r)$ and $\bar{H}(r)$ at $r_{\rm U}$. Furthermore, $\int_{r_{\rm U}-1/n}^{r_{\rm U}} n\big[q-\bar{H}(r)\big]\big(r_{\rm U}-r\big)\, d\big[1-\epsilon(r)\big] \to 0$ because

$$\begin{aligned}
&\left| \int_{r_{\rm U}-1/n}^{r_{\rm U}} n\big[q - \bar{H}(r)\big]\big(r_{\rm U} - r\big)\, d\big[1-\epsilon(r)\big] \right| \\
\le\ & \left(\sup_{r_{\rm U}-1/n \le r \le r_{\rm U}} \big|q - \bar{H}(r)\big| \right) \left(\sup_{r_{\rm U}-1/n \le r \le r_{\rm U}} \big|r_{\rm U} - r\big| \right) \left| \int_{r_{\rm U}-1/n}^{r_{\rm U}} n\, d\big[1-\epsilon(r)\big] \right| \\
=\ & \Big[q - \bar{H}\Big(r_{\rm U} - \frac{1}{n}\Big)\Big]\ \Big[\epsilon\Big(r_{\rm U} - \frac{1}{n}\Big) - \epsilon(r_{\rm U})\Big] \\
\to\ & 0
\end{aligned}$$

as $n \to \infty$, where the convergence follows from the left continuity of $\epsilon(r)$ at $r_{\rm U}$. It follows that

$$\begin{aligned}
& - \int_{r_{\rm L}}^{r_{\rm U}} \big[1-\epsilon(r)\big]\, dG^n(r) \\
\to\ & - \int_{r_{\rm L}}^{r_{\rm U}} \big[1-\epsilon(r)\big]\, d\big[q-\bar{H}(r)\big] + \big[1-\epsilon\big(r_{\rm U}\big)\big]\big[q-\bar{H}\big(r_{\rm U}\big)\big] \\
=\ & [1-\epsilon(r_{\rm L})]\ [q-\bar{H}(r_{\rm L})] + \int_{r_{\rm L}}^{r_{\rm U}} [q-\bar{H}(r)]\, d[1-\epsilon(r)].
\end{aligned}$$

Hence, the optimal value of problem (4.10c)–(4.10f) is $[1-\epsilon(r_{\rm L})]\ [q-\bar{H}(r_{\rm L})] + \int_{r_{\rm L}}^{r_{\rm U}}[q-\bar{H}(r)]\ d[1-\epsilon(r)]$. Finally, we plug this optimal value into (4.10a) and recast the dual problem (4.8a)–(4.8b) as

$$\begin{aligned}
& \inf_{p\in\mathbb{R}^T,\ q \ge \bar{H}(\overline{r})}\ m^\top p + q - [1-\epsilon(r_{\rm L})]\ [q-\bar{H}(r_{\rm L})] - \int_{r_{\rm L}}^{r_{\rm U}} [q-\bar{H}(r)]\, d[1-\epsilon(r)] \\
=\ & \inf_{p\in\mathbb{R}^T,\ q \ge \bar{H}(\overline{r})}\ m^\top p + \epsilon(r_{\rm U})q + [1-\epsilon(r_{\rm L})]\ \bar{H}(r_{\rm L}) + \int_{r_{\rm L}}^{r_{\rm U}} \bar{H}(r)\, d[1-\epsilon(r)] \\
=\ & \inf_{p\in\mathbb{R}^T}\ m^\top p + [1-\epsilon(r_{\rm L})]\ \bar{H}(r_{\rm L}) + \int_{r_{\rm L}}^{r_{\rm U}} \bar{H}(r)\, d[1-\epsilon(r)] + \epsilon(r_{\rm U})\bar{H}(\overline{r})
\end{aligned}$$

$$= \inf_{p \in \mathbb{R}^T} \ \mathbb{E}_{\mathbb{Q}}[H(x, p, \zeta)].$$

Therefore, $z_P \leq z_D = \inf_{p \in \mathbb{R}^T} \{\mathbb{E}_{\mathbb{Q}}[H(x, p, \zeta)]\}$ and the proof is completed. □

4.4 Sharpness of the Upper Bounds

In this section, we show that the upper bounds derived in Theorems 4.3.1–4.3.2 are actually sharp for a wide class of cost functions $f(x, \xi)$. To this end, we first formalize the sufficient conditions on the sharpness in the following assumption.

Assumption 4.4.1. *Ξ and $\{\Xi_r\}_{r \in \mathcal{R}}$ are convex and compact. In addition, for $I \in \mathbb{N}_+$, $f(x, \xi) = \max_{i \in [I]} \{f_i(x, \xi)\}$, where $-f_i(x, \xi)$ is proper, convex, and lower-semicontinuous in ξ for each $i \in [I]$.*

We note that all settings described in Section 4.2 satisfy Assumption 4.4.1 on the convexity and compactness of the confidence sets. In addition, Assumption 4.4.1 designates that $f(x, \xi)$ can be written as the pointwise maximum of a finite number of concave functions. We next describe three examples of $f(x, \xi)$ that satisfy this assumption.

Piecewise Linear Function. For all $i \in [I]$, define $f_i(x) := a_i(x)^\top \xi + b_i(x)$, where $a_i(x) : \mathbb{R}^n \to \mathbb{R}^T$ and $b_i(x) : \mathbb{R}^n \to \mathbb{R}$ represent affine functions of x. Then, such functions satisfy Assumption 4.4.1. These functions arise when modeling (i) the distributionally robust conditional Value-at-Risk (CVaR) constraints and (ii) two-stage adaptive DRO models.

Indicator Function. Consider a polyhedron $\{\xi : (A_i x + b_i)^\top \xi \leq c_i^\top x + d_i, \ \forall i \in [I]\}$. Then, the indicator function of its complement is

$$\mathbb{1}_{\left[\exists i \in [I]: (A_i x + b_i)^\top \xi > c_i^\top x + d_i\right]}(\xi) \ = \ \max \left\{ 0, \max_{i \in [I]} \left\{ 1 - \chi_{[(A_i x + b_i)^\top \xi > c_i^\top x + d_i]}(\xi) \right\} \right\}.$$

Then, such functions satisfy Assumption 4.4.1. These functions arise when quantifying the uncertainty of ξ by computing the worst-case probability of ξ locating outside of a polyhedron.

Single Concave Function. Suppose that $I = 1$ and $f(x, \xi)$ is concave in ξ. Then, such functions satisfy Assumption 4.4.1. These functions arise when modeling uncertain objective coefficients, e.g., $f(x, \xi) = \min_{y \in Y(x)} \xi^\top A y$, where $Y(x) \subseteq \mathbb{R}^m$ represents a feasible region of variables y that is a multifunction of variables x, $A \in \mathbb{R}^{n \times m}$ represents a given matrix, and $A^\top \xi$ represents the uncertain objective coefficients.

We are now ready to present the sharpness result for the discrete case in the following theorem.

Theorem 4.4.1. *Suppose that Assumptions 4.3.1 and 4.4.1 hold. Then,*

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] \;=\; \inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}\Big[H(x,p,\zeta)\Big],$$

where $H(x,p,\zeta)$ is defined in (4.3)*. In addition, $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]$ is attained by the distribution $\mathbb{P}^*_\xi \in \mathcal{D}$ with $\mathbb{P}^*_\xi\left\{\xi = \pi^*_{ij}\right\} = \alpha^*_{ij}$ for all $i \in [I]$ and $j \in [0,K+1]_{\mathbb{Z}}$, where (π^*,α^*) represents an optimal solution to the following optimization problem:*

$$\sup_{\alpha\geq 0,\pi} \quad \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} f_i(\pi_{ij}) \tag{4.11a}$$

$$\text{s.t.} \quad \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} = 1, \tag{4.11b}$$

$$\pi_{ij} \in \Xi_j, \;\; \forall i \in [I], \;\; \forall j \in [0,K+1]_{\mathbb{Z}}, \tag{4.11c}$$

$$\sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} A\pi_{ij} = m, \tag{4.11d}$$

$$\sum_{i=1}^{I}\sum_{j=0}^{k} \alpha_{ij} \geq 1-\epsilon(r_k), \;\; \forall k \in [0,K]_{\mathbb{Z}}, \tag{4.11e}$$

where we define $\Xi_{r_{K+1}} := \Xi$.

Proof. Recall, from the proof of Theorem 4.3.1, that $\inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)]$ equals the optimal value of the formulation (4.6c)–(4.6d), which we duplicate below:

$$\inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)] \;=\; \inf_{\substack{p\in\mathbb{R}^T,\; q\in\mathbb{R},\\ g_k\geq 0}} \; m^\top p + q - \sum_{k=0}^{K}[1-\epsilon(r_k)]g_k \tag{4.12a}$$

$$\text{s.t.} \quad q - \sum_{k=j}^{K} g_k \geq \sup_{\xi\in\Xi_{r_j}} \left\{f_i(x,\xi) - p^\top A\xi\right\}, \quad \forall i \in [I], \;\; \forall j \in [0,K+1]_{\mathbb{Z}}. \tag{4.12b}$$

In the following, we recast the formulation (4.12a)–(4.12b) as (4.11a)–(4.11e). To this end,

we recast the right-hand side of constraints (4.12b) as follows:

$$\begin{aligned}
&\sup_{\xi\in\Xi_{r_j}} \left\{ f_i(x,\xi) - p^\top A\xi \right\} \\
=\ &\sup_{\xi\in\mathbb{R}^T} \left\{ f_i(x,\xi) - p^\top A\xi - \chi_{\Xi_{r_j}}(\xi) \right\} && \text{(4.12c)} \\
=\ &\left(-f_i(x,\cdot) + \chi_{\Xi_{r_j}} \right)^* (-A^\top p) && \text{(4.12d)} \\
=\ &\text{cl}\left(\inf_{v_{ij}} \left\{ (-f_i)^*(x, -A^\top p - v_{ij}) + \sigma_{\Xi_{r_j}}(v_{ij}) \right\} \right) && \text{(4.12e)}
\end{aligned}$$

where equality (4.12c) follows from the definition of characteristic function $\chi_{\Xi_{r_j}}$, equality (4.12d) follows from the definition of conjugacy, and equality (4.12e) is because we rewrite the conjugate of the sum of $-f_i$ and $\chi_{\Xi_{r_j}}$ as the lower-semicontinuous closure of the infimal convolution of their conjugates. As the objective function (4.12a) is linear, we can relax the closure operation without loss of optimality. Plugging this representation back into (4.12b) yields that $\inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)]$ equals the following:

$$\inf_{\substack{p\in\mathbb{R}^T,\ q\in\mathbb{R},\\ g_k\geq 0,\ v_{ij}\in\mathbb{R}^T}} \quad m^\top p + q - \sum_{k=0}^{K} [1-\epsilon(r_k)] g_k \tag{4.12f}$$

$$\text{s.t. } q - \sum_{k=j}^{K} g_k \geq (-f_i)^*(x, -A^\top p - v_{ij}) + \sigma_{\Xi_{r_j}}(v_{ij}),\ \ \forall i \in [I],\ \forall j \in [0, K+1]_{\mathbb{Z}} \tag{4.12g}$$

$$\begin{aligned}
= \sup_{\alpha_{ij}\geq 0} \inf_{\substack{p\in\mathbb{R}^T,\ q\in\mathbb{R},\\ g_k\geq 0,\ v_{ij}\in\mathbb{R}^T}} \ & m^\top p + q - \sum_{k=0}^{K} [1-\epsilon(r_k)] g_k - \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} \Bigg[q - \sum_{k=j}^{K} g_k \\
& - (-f_i)^*(x, -A^\top p - v_{ij}) - \sigma_{\Xi_{r_j}}(v_{ij}) \Bigg]
\end{aligned}$$

$$= \sup_{\alpha_{ij}\geq 0} \inf_{p\in\mathbb{R}^T, v_{ij}\in\mathbb{R}^T} \ m^\top p + \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} \left[(-f_i)^*(x, -A^\top p - v_{ij}) + \sigma_{\Xi_{r_j}}(v_{ij}) \right] \tag{4.12h}$$

$$\text{s.t. } \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} = 1, \tag{4.12i}$$

$$\sum_{i=1}^{I}\sum_{j=0}^{k} \alpha_{ij} \geq 1 - \epsilon(r_k),\ \ \forall k \in [0, K]_{\mathbb{Z}}, \tag{4.12j}$$

where we take the Lagrangian dual of the formulation (4.12f)–(4.12g) with dual variables α_{ij} associated with constraints (4.12g), and dual constraints (4.12i) and (4.12j) are associ-

ated with primal variables q and g_k, respectively. The strong duality holds between formulations (4.12f)–(4.12g) and (4.12h)–(4.12j) because (4.12f)–(4.12g) is a convex program (note that $(-f_i)^*(x,\cdot)$ is convex in its second argument and $\sigma_{\Xi_{r_j}}$ is convex) and it satisfies the Slater's condition (indeed, we can arbitrarily increase the value of q to strictly satisfy constraints (4.12g)). For fixed $\alpha_{ij} \geq 0$, we proceed by representing the inner minimization problem in (4.12h) as follows:

$$\begin{aligned}
&\inf_{p\in\mathbb{R}^T, v_{ij}\in\mathbb{R}^T} \; m^\top p + \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij}\left[(-f_i)^*(x, -A^\top p - v_{ij}) + \sigma_{\Xi_{r_j}}(v_{ij})\right] \\
= &\inf_{p\in\mathbb{R}^T, v_{ij}\in\mathbb{R}^T} \; m^\top p + \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij}\left[(-f_i)^*(x, -A^\top p - v_{ij}) + \sup_{\pi_{ij}\in\Xi_{r_j}}\{v_{ij}^\top \pi_{ij}\}\right] && \text{(4.12k)} \\
= &\sup_{\pi_{ij}\in\Xi_{r_j}} \inf_{p\in\mathbb{R}^T, v_{ij}\in\mathbb{R}^T} \; m^\top p + \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij}\left[(-f_i)^*(x, -A^\top p - v_{ij}) + v_{ij}^\top \pi_{ij}\right] && \text{(4.12l)} \\
= &\sup_{\pi_{ij}\in\Xi_{r_j}} \inf_{p\in\mathbb{R}^T, w_{ij}\in\mathbb{R}^T} \; m^\top p + \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij}\left[(-f_i)^*(x, w_{ij}) + \left(-A^\top p - w_{ij}\right)^\top \pi_{ij}\right] && \text{(4.12m)} \\
= &\sup_{\pi_{ij}\in\Xi_{r_j}} \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} \inf_{w_{ij}\in\mathbb{R}^T}\left\{(-f_i)^*(x, w_{ij}) - w_{ij}^\top \pi_{ij}\right\} && \text{(4.12n)} \\
&\text{s.t.} \; \sum_{i=1}^{I}\sum_{j=0}^{K+1} \alpha_{ij} A\pi_{ij} = m, && \text{(4.12o)}
\end{aligned}$$

where equality (4.12k) follows from the definition of support function $\sigma_{\Xi_{r_j}}$. In equality (4.12l), we switch the order of $\inf_{p,v_{ij}}$ and $\sup_{\pi_{ij}}$ because (i) the objective function in (4.12k) is convex in variables (p, v_{ij}) and concave in variables π_{ij} and (ii) Ξ_{r_j} is convex and compact. In addition, equality (4.12m) follows from a change of variables through $w_{ij} := -A^\top p - v_{ij}$, equality (4.12n) is because $\alpha_{ij} \geq 0$ and the minimization over w_{ij} can be done separately for each $i \in [I]$ and $j \in [0, K+1]_{\mathbb{Z}}$, and the dual constraints (4.12o) are associated with variables p. Furthermore, we note that

$$\begin{aligned}
&\inf_{w_{ij}\in\mathbb{R}^T}\left\{(-f_i)^*(x, w_{ij}) - w_{ij}^\top \pi_{ij}\right\} \\
= &-\sup_{w_{ij}\in\mathbb{R}^T}\left\{\pi_{ij}^\top w_{ij} - (-f_i)^*(x, w_{ij})\right\} \\
= &\; f_i(x, \pi_{ij})
\end{aligned}$$

by the definition of biconjugacy and the assumption that $-f_i$ is proper, convex, and lower-

semicontinuous. Plugging (4.12n)–(4.12o) back into the objective function (4.12h) with the representation yields (4.11a)–(4.11e). Finally, we note that the probability distribution $\mathbb{P}^*_\xi$ associated with the optimal solution (π^*, α^*) belongs with ambiguity set $\mathcal{D}$ by constraints (4.11b)–(4.11e). It follows that $\inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)] \leq \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(x,\xi)]$ and so $\inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}_K}[H(x,p,\zeta)] = \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(x,\xi)]$ by Theorem 4.3.1. The proof is completed. □

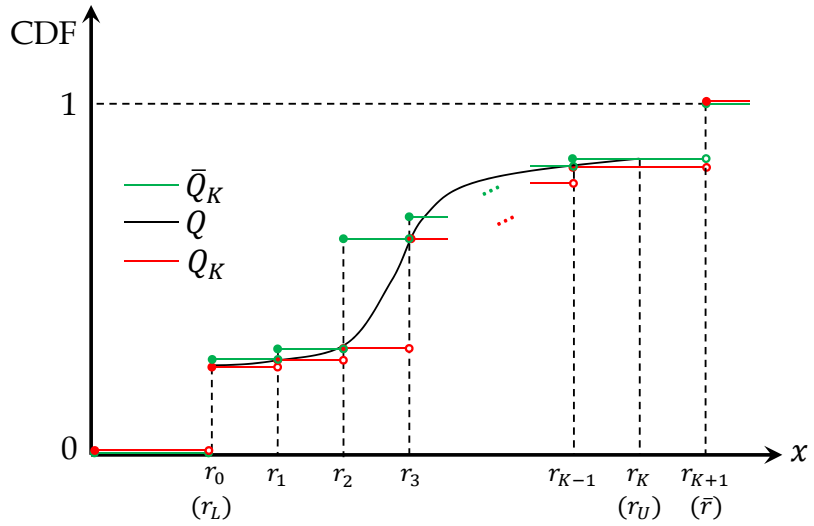


Figure 4.1: An example of the cumulative distribution functions (CDF) of $\overline{\mathbb{Q}}_K$, $\mathbb{Q}$, and $\mathbb{Q}_K$

Next, we present the sharpness result for the continuous case.

Theorem 4.4.2. *Suppose that Assumptions 4.3.2 and 4.4.1 hold. Then,*

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] = \inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}}\Big[H(x,p,\zeta)\Big],$$

where $H(x,p,\zeta)$ *is defined in* (4.3).

Proof. For a fixed $K \in \mathbb{N}_+$ and the given $\mathcal{R}$ and $\mathcal{D}$, define a sequence of real numbers $\{r_k\}_{k=0}^K$ such that $r_k = (k/K)r_{\text{U}} + (1 - k/K)r_{\text{L}}$ for all $k = 0, \ldots, K$ (e.g., $r_0 = r_{\text{L}}$ and $r_K = r_{\text{U}}$). In addition, define a probability distribution $\overline{\mathbb{Q}}_K$ of random variable ζ such that

$$\overline{\mathbb{Q}}_K\{\zeta \leq x\} = \begin{cases} 0, & \text{if } x < r_0, \\ 1 - \epsilon(r_k), & \text{if } r_{k-1} \leq x < r_k \text{ for } k \in \{1, \ldots, K\}, \\ 1 - \epsilon(r_K), & \text{if } r_K \leq x < \overline{r}, \\ 1, & \text{if } x \geq \overline{r}. \end{cases}$$

Correspondingly, define an ambiguity set $\overline{\mathcal{D}}_K$ such that

$$\overline{\mathcal{D}}_K \;=\; \Big\{\mathbb{P}\in\mathcal{P}(\Omega):\; \mathbb{E}_{\mathbb{P}}[A\xi]=m,\; \mathbb{P}\big\{\xi\in\Sigma_{r_k}\big\}\geq 1-\epsilon(r_{k+1}),\; \forall k=0,\ldots,K-1\Big\}.$$

Note that $\overline{\mathbb{Q}}_K$ is different from the $\mathbb{Q}_K$ defined in Theorem 4.3.1. In fact, for the given $\mathcal{D}$ and $\mathbb{Q}$ defined in Theorem 4.3.2, $\mathbb{Q}_K$ stochastically dominates $\mathbb{Q}$, which further stochastically dominates $\overline{\mathbb{Q}}_K$ (see Figure 4.1 for an example). By Theorems 4.3.1 and 4.4.1, we have

$$\sup_{\mathbb{P}\in\overline{\mathcal{D}}_K}\mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] \;=\; \inf_{p\in\mathbb{R}^T}\mathbb{E}_{\overline{\mathbb{Q}}_K}\left[H(x,p,\zeta)\right]$$

because $\mathbb{Q}_K$ and the ambiguity set $\mathcal{D}'$ with respect to $\mathcal{R}'=\{r_0,\ldots,r_K\}$ respectively recover $\overline{\mathbb{Q}}_K$ and $\overline{\mathcal{D}}_K$ if we replace each $1-\epsilon(r_k)$ in the definition of such $\mathcal{D}'$ by $1-\epsilon(r_{k+1})$ for all $k=0,\ldots,K-1$ and keep $1-\epsilon(r_K)$ unchanged.

On the one hand, we note that $\overline{\mathcal{D}}_K\subseteq\mathcal{D}$ because $\mathbb{P}\big\{\xi\in\Xi_{r_k}\big\}\geq 1-\epsilon(r_{k+1})$ for all $k=0,\ldots,K-1$ implies that $\mathbb{P}\big\{\xi\in\Xi_r\big\}\geq 1-\epsilon(r)$ for all $r\in[r_{\text{L}},r_{\text{U}}]$. It follows that $\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\overline{\mathbb{Q}}_K}\left[H(x,p,\zeta)\right]\leq\sup_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]$ for all $K\in\mathbb{N}_+$. For all $x\in[r_{\text{L}},r_{\text{U}}]$, we have $1-\epsilon(x)\leq\overline{\mathbb{Q}}_K\{\zeta\leq x\}\leq 1-\epsilon\left(x+(r_{\text{U}}-r_{\text{L}})/K\right)$ because $1-\epsilon(x)$ is non-decreasing in x. Then, $\lim_{K\to\infty}\overline{\mathbb{Q}}_K\{\zeta\leq x\}=\mathbb{Q}\{\zeta\leq x\}$ for all $x\in[r_{\text{L}},r_{\text{U}}]$. In addition, $\overline{\mathbb{Q}}_K\{\zeta\leq x\}=\mathbb{Q}\{\zeta\leq x\}$ for all $x\in[0,r_{\text{L}})\cap[r_{\text{U}},\infty)$ by construction. It follows that $\lim_{K\to\infty}\overline{\mathbb{Q}}_K\{\zeta\leq x\}=\mathbb{Q}\{\zeta\leq x\}$ for all x where $\mathbb{Q}\{\zeta\leq x\}$ is continuous, or equivalently, the sequence of cumulative distribution functions $\{\overline{\mathbb{Q}}_K\{\zeta\leq\cdot\}\}_{K\in\mathbb{N}_+}$ weakly converges to $\mathbb{Q}\{\zeta\leq\cdot\}$. Hence, for any fixed $p\in\mathbb{R}^T$, $\lim_{K\to\infty}\mathbb{E}_{\overline{\mathbb{Q}}_K}\left[H(x,p,\zeta)\right]=\mathbb{E}_{\mathbb{Q}}\left[H(x,p,\zeta)\right]$ by the Helly-Bray Theorem. Furthermore, it can be shown that there exists a compact set $\bar{P}\subseteq\mathbb{R}^T$ such that $\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{P}_\xi}\left[H(x,p,\zeta)\right]=\inf_{p\in\bar{P}}\mathbb{E}_{\mathbb{P}_\xi}\left[H(x,p,\zeta)\right]$ for any probability distribution $\mathbb{P}_\xi\in\mathcal{P}(\Omega)$. It follows that $\lim_{K\to\infty}\Big\{\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\overline{\mathbb{Q}}_K}\left[H(x,p,\zeta)\right]\Big\}=\lim_{K\to\infty}\Big\{\inf_{p\in\bar{P}}\mathbb{E}_{\overline{\mathbb{Q}}_K}\left[H(x,p,\zeta)\right]\Big\}=\inf_{p\in\bar{P}}\mathbb{E}_{\mathbb{Q}}\left[H(x,p,\zeta)\right]=\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\left[H(x,p,\zeta)\right]$. Therefore, $\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\left[H(x,p,\zeta)\right]\leq\sup_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]$.

On the other hand, we have $\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\left[H(x,p,\zeta)\right]\geq\sup_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right]$ by Theorem 4.3.2. The proof is hence completed. □

Remark 4.4.3. *Theorem 4.4.2 formally confirms that the DRO model with shape information can be recast as a stochastic program with a one-dimensional random variable and an unambiguous probability distribution. This representation holds for a wide range of shape information under the continuous topology (see Assumption 4.3.2) and a wide range of cost functions (see Assumption 4.4.1). Notably, this reformulation is also computationally tractable. Indeed, for any the cost function $f(x,\xi)$ that is convex in x, the reformulation $\inf_{x,p}\mathbb{E}_{\mathbb{Q}}[H(x,p,\zeta)]$ is a convex stochastic program. This stochastic program can be effi-*

ciently solved by the sample average approximation algorithm because it only involves one single random variable (i.e., ζ).

We close this section with examples of applying Theorems 4.4.2 on cost functions that satisfy Assumption 4.4.1.

Piecewise Linear Function. Suppose that $A = I_{T\times T}$, $\Xi_r = \{\xi : \|\xi - m\| \le r\}$ and $f(x,\xi) = \max_{i\in[I]}\{f_i(x,\xi)\}$ with $f_i(x) := a_i(x)^\top \xi + b_i(x)$ for all $i \in [I]$, where $a_i(x) : \mathbb{R}^n \to \mathbb{R}^T$ and $b_i(x) : \mathbb{R}^n \to \mathbb{R}$ represent affine functions of x. Then,

$$\begin{aligned} H(x,p,\zeta) &= \max_{i\in[I]} \sup_{\xi:||\xi-m||\le\zeta} \left\{a_i(x)^\top \xi + b_i(x) - p^\top(\xi - m)\right\} \\ &= \max_{i\in[I]} \sup_{\xi:||\xi-m||\le\zeta} \left\{(a_i(x)-p)^\top(\xi - m) + a_i(x)^\top m + b_i(x)\right\} \\ &= \max_{i\in[I]} \left\{\zeta\,||a_i(x)-p||_* + a_i(x)^\top m + b_i(x)\right\}. \end{aligned}$$

It follows from Theorem 4.4.2 that

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[f(x,\xi)\right] = \inf_{p\in\mathbb{R}^T} \mathbb{E}_{\mathbb{Q}}\Big[\max_{i\in[I]}\Big\{\zeta\,||a_i(x)-p||_* + a_i(x)^\top m + b_i(x)\Big\}\Big].$$

Distributionally Robust CVaR Constraint. Suppose that $A = I_{T\times T}$ and $\Xi_r = \{\xi : \|\xi - m\| \le r\}$. Consider the distributionally robust CVaR constraint $\sup_{\mathbb{P}\in\mathcal{D}} \text{CVaR}^{\mathbb{P}}_{1-\epsilon}\left(a(x)^\top\xi\right) \le b(x)$, where $a(x) : \mathbb{R}^n \to \mathbb{R}^T$ and $b(x) : \mathbb{R}^n \to \mathbb{R}$ represent affine functions of x and $1-\epsilon$ represents the confidence level of CVaR. Then, by Theorem 4.4.2 and the definition of CVaR (see, e.g., [115]), we have

$$\begin{aligned} &\sup_{\mathbb{P}\in\mathcal{D}} \text{CVaR}^{\mathbb{P}}_{1-\epsilon}\left(a(x)^\top\xi\right) \\ &= \inf_{\beta\in\mathbb{R}}\left\{\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}\Big[\max\big\{a(x)^\top\xi - \beta,\ 0\big\}\Big]\right\} \\ &= \inf_{\beta\in\mathbb{R}}\left\{\beta + \frac{1}{\epsilon}\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\Big[\max\Big\{\zeta\|a(x)-p\|_* + a(x)^\top m - \beta,\ \zeta\|p\|_*\Big\}\Big]\right\} \\ &= \inf_{\substack{\beta\in\mathbb{R}\\ p\in\mathbb{R}^T}}\left\{\beta + \frac{1}{\epsilon}\,\mathbb{E}_{\mathbb{Q}}\Big[\max\Big\{\zeta\|a(x)-p\|_* + a(x)^\top m - \beta,\ \zeta\|p\|_*\Big\}\Big]\right\}. \end{aligned}$$

It follows that the distributionally robust CVaR constraint can be recast as the following expectation conic constraint:

$$\epsilon\beta + \mathbb{E}_{\mathbb{Q}}\Big[\max\Big\{\zeta\|a(x)-p\|_* + a(x)^\top m - \beta,\ \zeta\|p\|_*\Big\}\Big] \le \epsilon b(x).$$

Indicator Function. Suppose that $A = I_{T\times T}$ and $\Xi_r = \{\xi : \|\xi - m\|_\infty \leq r\}$. Given $x \in \mathbb{R}^n_+$, consider polyhedron $S = \{\xi : \xi \geq m + x\}$ and the uncertainty quantification problem $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\{\xi \notin S\}$. Then, it follows from Theorem 4.4.2 and the fact $\mathbb{1}_{\mathbb{R}^T\setminus S}(\xi) = \max\big\{0, \max_{t\in[T]}\{1 - \chi_{[\xi_t<m_t+x_t]}(\xi)\}\big\}$ that

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\{\xi \notin S\} \;=\; \inf_{p\in\mathbb{R}^T} \left\{ \mathbb{E}_{\mathbb{Q}}[\zeta]\; \|p\|_1 + \mathbb{E}_{\mathbb{Q}}\left[\max_{t\in[T]} \{1 - (\zeta - x_t)_+(p_t)_+\}\right]_+ \right\},$$

where $(x)_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. Hence, we can quantify the uncertainty of $\mathbb{P}\{\xi \notin S\}$ by solving a stochastic linear program.

Single Concave Function. Suppose that $I = 1$ and $f(x, \xi)$ is concave in ξ. The following proposition shows that it suffices to represent the uncertainty of random vector ξ only by its mean value μ. For example, consider a two-stage stochastic program $\min_{x\in X}\{c^\top x + \mathbb{E}_{\mathbb{P}}[f(x, \xi)]\}$ with uncertain objective coefficients, i.e., $f(x, \xi) = \min_{y\in Y(x)} \xi^\top Ay$, where $Y(x) := \{y : Tx + Wy \leq h\}$ represents a deterministic polyhedron that is nonempty for any $x \in X$. The following proposition implies that $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(x, \xi)] = \min_{y\in Y(x)} \mu^\top Ay$. Hence, we can solve the two-stage DRO model $\min_{x\in X}\{c^\top x + \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(x, \xi)]\}$ by solving a linear program *without* uncertainty.

Proposition 4.4.4. *Suppose that $A = I_{T\times T}$ and denote the mean value of ξ as $\mu := m$. If $f(x, \xi)$ is concave in ξ for any given x then*

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[f(x, \xi)] \;=\; f(x, \mu).$$

Proof. First, we show $\inf_{p\in\mathbb{R}^m}\; \mu^\top p + \mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x, \xi) - p^\top\xi\Big\}\right] \geq f(x, \mu)$. For all $p \in \mathbb{R}^m$ and for all $\zeta \in [r_L, \overline{r}]$, we have

$$\sup_{\xi\in\Xi_\zeta}\Big\{f(x, \xi) - p^\top\xi\Big\} \geq f(x, \mu) - p^\top\mu,$$

because $\xi = \mu$ is a feasible solution for this maximization problem. Then for all $p \in \mathbb{R}^\top$,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x, \xi) - p^\top\xi\Big\}\right] \geq f(x, \mu) - p^\top\mu.$$

Hence

$$\inf_{p\in\mathbb{R}^m}\ \mu^\top p + \mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x,\xi)-p^\top\xi\Big\}\right] \geq f(x,\mu).$$

To prove the other direction, we start with the following inequality

$$\sup_{\xi\in\Xi_\zeta}\Big\{f(x,\xi)-p^\top\xi\Big\} \leq \sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top\xi\Big\}, \quad \forall p\in\mathbb{R}^m,$$

which holds since the right-hand side is a relaxation of the left-hand side. Then for all $p\in\mathbb{R}^m$, we have

$$\mu^\top p+\mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x,\xi)-p^\top\xi\Big\}\right] \leq \mu^\top p+\sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top\xi\Big\} = \sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top(\xi-\mu)\Big\}.$$

Hence

$$\inf_{p\in\mathbb{R}^m}\ \mu^\top p + \mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x,\xi)-p^\top\xi\Big\}\right] \leq \inf_{p\in\mathbb{R}^m}\ \sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top(\xi-\mu)\Big\}.$$

Note that

$$\inf_{p\in\mathbb{R}^m}\ \sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top(\xi-\mu)\Big\}$$

is the Lagrange dual problem of

$$\begin{aligned}\sup_{\xi\in\mathbb{R}^m}\quad & f(x,\xi)\\ \text{s.t.}\quad & \xi=\mu,\end{aligned} \tag{4.13}$$

where strong duality holds since Slater's condition is satisfied. Thus we have

$$\inf_{p\in\mathbb{R}^m}\ \sup_{\xi\in\mathbb{R}^m}\Big\{f(x,\xi)-p^\top(\xi-\mu)\Big\} = f(x,\mu),$$

which gives

$$\inf_{p\in\mathbb{R}^m}\mu^\top p + \mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi\in\Xi_\zeta}\Big\{f(x,\xi)-p^\top\xi\Big\}\right] \leq f(x,\mu).$$

Thus, this proof is completed. □

4.5 Extensions

In this section, we extend the results to other important applications of DRO. In particular, we derive a sharp upper bound of the covariance matrix in Section 4.5.1 and a computationally tractable reformulation of two-stage adaptive DRO in Section 4.5.2.

4.5.1 Sharp Upper Bound of the Covariance Matrix

In this section, we extend the discussion to the case beyond Assumption 4.4.1, i.e., when $f(x,\xi)$ is convex in ξ. In particular, we derive a sharp upper bound of the covariance matrix $\mathbb{E}_{\mathbb{P}_\xi}[(\xi-\mu)(\xi-\mu)^\top]$ by using Theorem 4.3.2. We summarize this result in the following proposition.

Proposition 4.5.1. *Suppose that $A = I_{T\times T}$, $\mu := m$, and $\Xi_r = \{\xi : \|\Sigma^{-1/2}(\xi-\mu)\|_2 \le r\}$. Then,*

$$\mathbb{E}_{\mathbb{P}}\left[(\xi-\mu)(\xi-\mu)^\top\right] \preceq \mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma, \qquad \forall \mathbb{P}\in\mathcal{D}.$$

Furthermore, the upper bound $\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$ is sharp in the sense that, for any symmetric matrix $\Delta\in\mathbb{R}^{T\times T}$, $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top]\preceq\Delta$ for all $\mathbb{P}\in\mathcal{D}$ implies that $\Delta\succeq\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$.

Proof. First, we define function $f:\mathbb{R}^T\to\mathbb{R}$ such that

$$f(a) \;=\; \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[a^\top(\xi-\mu)(\xi-\mu)^\top a\right].$$

By Theorems 4.3.2, we have

$$\begin{aligned} f(a) \;&\le\; \inf_{p\in\mathbb{R}^T}\left\{\mu^\top p + \mathbb{E}_{\mathbb{Q}}\left[\max_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le\zeta}\left\{a^\top(\xi-\mu)(\xi-\mu)^\top a - p^\top\xi\right\}\right]\right\} \\ &=\; \inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le\zeta}\left\{a^\top(\xi-\mu)(\xi-\mu)^\top a - p^\top(\xi-\mu)\right\}\right]. \qquad (4.14a) \end{aligned}$$

Let $\bar{f}(a) := \inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le\zeta}\left\{a^\top(\xi-\mu)(\xi-\mu)^\top a - p^\top(\xi-\mu)\right\}\right]$.

Second, we show that $p=0$ is an optimal solution to the (outer-layer) minimization problem in (4.14a). On the one hand, for any $p\in\mathbb{R}^T$ and $\zeta>0$, we note that

$$\begin{aligned} &\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2\le\zeta}\left\{a^\top(\xi-\mu)(\xi-\mu)^\top a - p^\top(\xi-\mu)\right\} \\ =\;&\sup_{\phi:\|\phi\|_2\le\zeta}\left\{\left[(\Sigma^{\frac{1}{2}}a)^\top\phi\right]^2 - (\Sigma^{\frac{1}{2}}p)^\top\phi\right\} \qquad (4.14b) \end{aligned}$$

$$\geq \zeta^2(a^\top \Sigma a) + \frac{|p^\top \Sigma a|}{\|\Sigma^{\frac{1}{2}} a\|_2}\zeta \tag{4.14c}$$

$$\geq a^\top \left(\zeta^2 \Sigma\right) a, \tag{4.14d}$$

where equality (4.14b) follows from a change of variable by defining $\phi = \Sigma^{-\frac{1}{2}}(\xi - \mu)$, inequality (4.14c) holds because $\frac{\Sigma^{\frac{1}{2}} a}{\|\Sigma^{\frac{1}{2}} a\|_2}\zeta$ and $-\frac{\Sigma^{\frac{1}{2}} a}{\|\Sigma^{\frac{1}{2}} a\|_2}\zeta$ are both feasible (but possibly suboptimal) to the maximization problem in (4.14b). It follows that $\bar{f}(a) \geq a^\top \Big(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma\Big) a$ because inequality (4.14d) holds valid for any $p \in \mathbb{R}^T$ and $\zeta > 0$. On the other hand, this lower bound is attainable by letting $p = 0$ because

$$\begin{aligned}
& \mathbb{E}_{\mathbb{Q}}\left[\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2 \leq \zeta} \left\{a^\top(\xi-\mu)(\xi-\mu)^\top a\right\}\right] \\
= \; & \mathbb{E}_{\mathbb{Q}}\left[\sup_{\phi:\|\phi\|_2 \leq \zeta} \left\{\left[(\Sigma^{\frac{1}{2}} a)^\top \phi\right]^2\right\}\right] \\
= \; & \mathbb{E}_{\mathbb{Q}}\left[\zeta^2 \left(a^\top \Sigma a\right)\right].
\end{aligned}$$

It follows that $\bar{f}(a) \leq a^\top \Big(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma\Big) a$, hence $\bar{f}(a) = a^\top \Big(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma\Big) a$. Then for all $a \in \mathbb{R}^T$, $f(a) \leq \bar{f}(a)$ implies $\sup_{\mathbb{P}_\xi \in \mathcal{D}} a^\top \mathbb{E}_{\mathbb{P}_\xi}[(\xi-\mu)(\xi-\mu)^\top] a \leq a^\top \Big(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma\Big) a$. Therefore, $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top] \preceq \mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$ for all $\mathbb{P} \in \mathcal{D}$.

Third, we show that $f(a) \geq \bar{f}(a)$. We prove by constructing a $\mathbb{P}^* \in \mathcal{D}$ such that

$$\mathbb{E}_{\mathbb{P}^*}\left[a^\top(\xi-\mu)(\xi-\mu)^\top a\right] = \mathbb{E}_{\mathbb{Q}}\left[\zeta^2 \left(a^\top \Sigma a\right)\right].$$

Let $p_\zeta(x)$ be the corresponding probability density function of $\mathbb{Q}\{\zeta \leq x\}$. And $\forall x \in [r_L, r_U] \cup \{\bar{r}\}$, let $\xi_x^+ = \frac{\Sigma a}{\|\Sigma^{\frac{1}{2}} a\|_2} x + \mu$ and $\xi_x^- = -\frac{\Sigma a}{\|\Sigma^{\frac{1}{2}} a\|_2} x + \mu$. Note that ξ_x^+, ξ_x^- correspond to the optimal solution of $\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2 \leq x} \left\{a^\top(\xi-\mu)(\xi-\mu)^\top a\right\}$. We define $p_\xi(y)$ based on $p_\zeta(x)$, i.e., $p_\xi(\xi_x^+) = p_\xi(\xi_x^-) = \frac{1}{2} p_\zeta(x)$, $\forall x \in [r_L, r_U] \cup \{\bar{r}\}$ and $p_\xi(y) = 0$ for all the other $y \in \Xi$. Then, $\int_{y \in \Xi} p_\xi(y) dy = 1$. Let $\mathbb{P}^*$ be the cumulative distribution function corresponding to $p_\xi(y)$. Then, $\mathbb{P}^* \in \mathcal{P}(\Xi)$. Because $\xi_x^+ + \xi_x^- = \mu, \forall x$, $\mathbb{E}_{\mathbb{P}^*}[\xi] = \mu$. By the definition of $\mathbb{Q}$ and $p_\xi(y)$, we have $\mathbb{P}^*\{\xi \in \Xi_r\} \geq 1 - \epsilon(r), \forall r \in \mathcal{R}$. Hence $\mathbb{P}^* \in \mathcal{D}$. It remains to check $\mathbb{E}_{\mathbb{P}^*}\left[a^\top(\xi-\mu)(\xi-\mu)^\top a\right] = \mathbb{E}_{\mathbb{Q}}\left[\zeta^2 \left(a^\top \Sigma a\right)\right]$. This holds because $\forall x \in [r_L, r_U] \cup \{\bar{r}\}$, we set ξ as the two optimal solutions ξ_x^+, ξ_x^- of $\sup_{\xi:\|\Sigma^{-\frac{1}{2}}(\xi-\mu)\|_2 \leq x} \left\{a^\top(\xi-\mu)(\xi-\mu)^\top a\right\}$, with $p_\xi(\xi_x^+) = p_\xi(\xi_x^-) = \frac{1}{2} p_\zeta(x)$ and the optimal value is $x^2 \left(a^\top \Sigma a\right)$. It follows that $f(a) \geq \bar{f}(a)$ and hence $f(a) = \bar{f}(a)$.

Finally, we prove by contradiction that the upper bound $\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$ is sharp. Suppose that

there exists a symmetric matrix $\Delta \in \mathbb{R}^{T\times T}$ such that $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top] \preceq \Delta$ for all $\mathbb{P}\in\mathcal{D}$ and $\Delta - \mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$ is not positive semidefinite. Then, there exists a vector $\hat{a}\in\mathbb{R}^T$ such that $\hat{a}^\top(\Delta-\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma)\hat{a} < 0$, i.e., $\hat{a}^\top\Delta\hat{a} < \hat{a}^\top(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma)\hat{a}$. But it is shown above that $f(\hat{a}) = \sup_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}[\hat{a}^\top(\xi-\mu)(\xi-\mu)^\top\hat{a}] = \hat{a}^\top(\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma)\hat{a}$ and it follows that there exists a probability distribution $\hat{\mathbb{P}}\in\mathcal{D}$ such that $\mathbb{E}_{\hat{\mathbb{P}}}[\hat{a}^\top(\xi-\mu)(\xi-\mu)^\top\hat{a}] > \hat{a}^\top\Delta\hat{a}$. This contradicts the assumption that $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top]\preceq\Delta$ for all $\mathbb{P}\in\mathcal{D}$. Thus, the upper bound $\mathbb{E}_{\mathbb{Q}}[\zeta^2]\Sigma$ is sharp. □

Remark 4.5.2. *Proposition 4.5.1 indicates that the shape information may be strong enough to subsume the covariance information. In particular, suppose that one knows $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top]\preceq\Sigma$ from the covariance information. But if $\mathbb{E}_{\mathbb{Q}}[\zeta^2] < 1-\epsilon$ for an $\epsilon\in(0,1)$, then Proposition 4.5.1 implies that $\mathbb{E}_{\mathbb{P}}[(\xi-\mu)(\xi-\mu)^\top]\preceq(1-\epsilon)\Sigma$, which already implies the covariance bound. From the perspective of computation, we may compute $\mathbb{E}_{\mathbb{Q}}[\zeta^2]$ before incorporating the covariance information into $\mathcal{D}$ to avoid unnecessary computational burden.*

4.5.2 Two-Stage Adaptive DRO with Fixed Recourse

In this section, we study a special case of shape information based on the ℓ_1-norm and show that it leads to a computationally tractable reformulation of the two-stage adaptive DRO model with fixed recourse. We summarize this result in the following theorem.

Theorem 4.5.3. *Consider an ℓ_1-norm based ambiguity set*

$$\mathcal{D}_1 := \Big\{\mathbb{P}\in\mathcal{P}(\Xi):\ \mathbb{E}_{\mathbb{P}}[A\xi]=m,\ \mathbb{P}\big\{\|\Sigma^{-1/2}(\xi-\mu)\|_1\le r\big\}\ge 1-\epsilon(r),\ \forall r\in[r_L,\ r_U]\Big\},$$

where the support set $\Xi := \{\xi\in\mathbb{R}^T : \|\Sigma^{-1/2}(\xi-\mu)\|_1\le\bar{r}\}$ and $\Sigma\in\mathbb{R}^{T\times T}$ represent a positive definite matrix. Then, the two-stage adaptive DRO model

$$\min_{x\in X}\left\{c^\top x + \sup_{\mathbb{P}\in\mathcal{D}_1}\mathbb{E}_{\mathbb{P}}[Q(x,\xi)]\right\}$$

with

$$\begin{aligned} Q(x,\xi) := \min_{y\ge 0}\ & q^\top y \\ \text{s.t. }\ & T(\xi)x + Wy \ge h(\xi), \end{aligned}$$

where $T(\xi)$ and $h(\xi)$ respectively represent the technology matrix and the right-hand side

that are affine in ξ, is equivalent to

$$\inf_{p\in\mathbb{R}^T,\, x\in X}\left\{c^\top x+(m-A\mu)^\top p+\mathbb{E}_{\mathbb{Q}}\left[\max_{t\in[T]}\left\{Q(x,\mu\pm\zeta\Sigma^{1/2}e_t)\mp\zeta e_t^\top\Sigma^{1/2}A^\top p\right\}\right]\right\},$$

where e_t represents the t^{th} standard basis vector in $\mathbb{R}^T$ and probability distribution $\mathbb{Q}$ is defined in Theorem 4.3.2.

Proof. Following Theorem 4.4.2, we have

$$\sup_{\mathbb{P}\in\mathcal{D}_1}\mathbb{E}_{\mathbb{P}}[Q(x,\xi)]=\inf_{p\in\mathbb{R}^T}\mathbb{E}_{\mathbb{Q}}[H(x,p,\zeta)], \tag{4.15a}$$

where

$$\begin{aligned}
H(x,p,\zeta) &= \sup_{\xi:\|\Sigma^{-1/2}(\xi-\mu)\|_1\le\zeta}\left\{Q(x,\xi)-p^\top(A\xi-m)\right\}\\
&= \sup_{\alpha:\|\alpha\|_1\le\zeta}\left\{Q(x,\mu+\Sigma^{1/2}\alpha)-p^\top(A\Sigma^{1/2}\alpha+A\mu-m)\right\} &\text{(4.15b)}\\
&= \max_{t\in[T]}\left\{Q(x,\mu\pm\zeta\Sigma^{1/2}e_t)-p^\top(\pm\zeta A\Sigma^{1/2}e_t+A\mu-m)\right\}, &\text{(4.15c)}
\end{aligned}$$

where equality (4.15b) follows from a change of variable $\alpha\equiv\Sigma^{-1/2}(\xi-\mu)$. In addition, as $Q(x,\mu+\Sigma^{1/2}\alpha)$ is convex in α, the supremum in (4.15b) is attained at one of the extreme points of the ℓ_1-ball $\{\alpha:\|\alpha\|_1\le\zeta\}$, which are $\{\pm\zeta e_t\}_{t=1}^T$. Then, equality (4.15c) follows from enumerating these extreme points. The conclusion follows from plugging the representation $H(x,p,\zeta)$ back into (4.15a). The proof is completed. □

4.6 Case Study

To demonstrate the proposed DRO model with shape information, we conduct numerical experiments on the appointment scheduling problem (ASP) and the risk-constrained optimal power flow problem (RC-OPF). All computations are implemented by CPLEX 12.6 on a 64-bit Windows 7 PC with Intel(R) Xeon(R) E3-1241 processor, running at 3.50 GHz and 32 GB memory. All instances of ASP and RC-OPF instances can be solved within 3 and 200 seconds, respectively.

In all DRO models with shape information, we assume $A=I_{T\times T}$, $\Xi_r=\{\xi:\|\xi-m\|\le r\}$ for all $r\in\mathcal{R}$ in the ambiguity set, and $\Xi=\{\xi:\|\xi-m\|\le\bar{r}\}$ with $\bar{r}\ge\sup\mathcal{R}$, where $\|\cdot\|$ represents a norm on $\mathbb{R}^n$ to be specified later.

4.6.1 Appointment Scheduling

We consider an ambiguous ASP model described in [95], where a single server needs to schedule a set of appointments with ambiguous service durations. Due to the uncertainty, an appointments may start after the scheduled time due to the late completion of the previous appointment (waiting) and the server may work overtime due to late completion of the last appointment (overtime). The object is to minimize a weighted sum of the waiting and overtime penalty by optimally deciding the appointment times. Specifically, we consider scheduling n appointments within a time limit T. For all $i \in [n]$, we let $x := [x_1, \ldots, x_n]^\top$ and $w := [w_1, \ldots, w_n]^\top$ represent the scheduled appointment durations and waiting times, respectively, and w_{n+1} the overtime. In addition, we let c_i represent the unit penalty cost for waiting/overtime w_i, $\forall i \in [n+1]$. Random vector $\xi := [\xi_1, \ldots, \xi_n]^\top$ represents the service durations. Given x and ξ, the waiting times and the overtime can be computed by solving the following linear program:

$$f_1(x, \xi) := \min_{w} \quad \sum_{i=1}^{n+1} c_i w_i \tag{4.16a}$$

$$\text{s.t.} \quad w_{i+1} \geq w_i + \xi_i - x_i, \ \forall i \in [n], \tag{4.16b}$$

$$w_1 = 0, \quad w_i \geq 0, \ \forall i \in [2, n+1]_{\mathbb{Z}}. \tag{4.16c}$$

The objective function (4.16a) minimizes the total penalty cost from a system perspective. However, this could result in uneven waiting times (or overtime), e.g., some appointments have much longer waiting times than others. To promote a more fair appointment schedule, we also consider minimizing the maximum penalty in the following linear program:

$$f_2(x, \xi) := \min_{w} \quad \max \left\{ \max_{i \in [n]} \{c_i w_i\}, \ \gamma c_{n+1} w_{n+1} \right\} \tag{4.17a}$$

$$\text{s.t.} \quad \text{(4.16b)–(4.16c)}, \tag{4.17b}$$

where parameter γ trades off between the maximum waiting time and the overtime. In this experiment, we set $\gamma = 3$. To address the ambiguity of ξ, we consider a distributionally robust appointment scheduling (DR-ASP) model with moment and shape information. Specifically, we formulate

$$\min_{x \geq 0} \quad \sup_{\mathbb{P} \in \mathcal{D}} \ \mathbb{E}_{\mathbb{P}}[f(x, \xi)] \tag{4.18a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i \leq T, \tag{4.18b}$$

where constraint (4.18b) designates that all appointments are scheduled before the time limit. We derive tractable reformulation of the (DR-ASP) model. To that end, we specify the norm in $\mathcal{D}$.

Definition 4.6.1. *For given $c^+, c^- \in \mathbb{R}^T_+$, let $\|\xi\|_{\mathsf{G}} = \max_{t\in[T]}\{\max\{c_t^+\xi_t,\ -c_t^-\xi_t\}\}$.*

Throughout this experiment, we adopt $\|\cdot\|_{\mathsf{G}}$ in $\mathcal{D}$ for modeling shape information. We note that if $c^+ = c^- = e$ then $\|\cdot\|_{\mathsf{G}} = \|\cdot\|_\infty$. On the other hand, if $c_t^+ \neq c_t^-$ for some $t \in [T]$ then $\|\cdot\|_{\mathsf{G}}$ can capture the skewness of the distribution of ξ_t. We also note that the dual norm of $\|\cdot\|_{\mathsf{G}}$ is $\|\cdot\|_{\mathsf{G}*}$ such that $\|\xi\|_{\mathsf{G}*} = \sum_{t=1}^T \max\{\xi_j/c_t^+,\ -\xi_t/c_t^-\}$. The next theorem shows that (DR-ASP) with both objective functions f_1 and f_2 admits a tractable reformulation.

Theorem 4.6.2. *Under norm $\|\cdot\|_{\mathsf{G}}$ and objective function f_1, the (DR-ASP) model* (4.18a)–(4.18b) *is equivalent to stochastic linear program $\min_{p\in\mathbb{R}^n,\ x\in\mathbb{R}^n_+:(4.18b)} \mathbb{E}_{\mathbb{Q}}[J_1(x,p,\zeta)]$, where*

$$
\begin{aligned}
J_1(x,p,\zeta) \ =\ \min_{\lambda,\phi}\ & \sum_{j=1}^n \lambda_j \\
\text{s.t.}\ & \sum_{j=k}^{\min\{\ell,n\}} \lambda_j \geq \sum_{j=k}^{\min\{\ell,n\}} [\zeta\phi_{j\ell} + (\mu_j - x_j)\pi_{j\ell}],\ \ \forall k\in[n],\ \forall \ell\in[k,n+1]_{\mathbb{Z}}, \\
& c_k^-\phi_{k\ell} - p_k \geq -\pi_{k\ell},\ \ \forall k\in[n],\ \forall \ell\in[k,n+1]_{\mathbb{Z}}, \\
& c_k^+\phi_{k\ell} + p_k \geq \pi_{k\ell},\ \ \forall k\in[n],\ \forall \ell\in[k,n+1]_{\mathbb{Z}},
\end{aligned}
$$

where $\pi_{k\ell} = \sum_{j=k+1}^{\ell} c_j$ for all $k\in[n]$ and $\ell\in[k,n+1]_{\mathbb{Z}}$. In addition, under norm $\|\cdot\|_{\mathsf{G}}$ and objective function f_2, the (DR-ASP) model (4.18a)–(4.18b) *is equivalent to stochastic linear program $\min_{p\in\mathbb{R}^n,\ x\in\mathbb{R}^n_+:(4.18b)} \mathbb{E}_{\mathbb{Q}}[J_2(x,p,\zeta)]$, where*

$$
\begin{aligned}
J_2(x,p,\zeta) \ =\ \min_{\theta,\varphi}\ & \theta \\
\text{s.t.}\ & \theta \geq \zeta\sum_{k=1}^n \varphi_{ijk} - b_{ij}(x) + a_{ij}^\top\mu,\ \ \forall i\in[2,n+1]_{\mathbb{Z}},\ \forall j\in[i], \\
& c_k^-\varphi_{ijk} - p_k \geq -a_{ijk},\ \ \forall i\in[2,n+1]_{\mathbb{Z}},\ \forall j\in[i],\ \forall k\in[n], \\
& c_k^+\varphi_{ijk} + p_k \geq a_{ijk},\ \ \forall i\in[2,n+1]_{\mathbb{Z}},\ \forall j\in[i],\ \forall k\in[n],
\end{aligned}
$$

$$
a_{ijk} \ =\ \begin{cases} c_i,\ \textit{if } k\in[j,i-1]_{\mathbb{Z}} \\ 0,\ \textit{otherwise} \end{cases} \qquad \textit{and } b_{ij}(x) \ =\ c_i\sum_{k=j}^{i-1} x_k.
$$

Proof. First, for objective function f_1, we take the dual of formulation (4.16a)–(4.16c) to obtain

$$f_1(x,\xi) = \max_{y} \; y^\top \xi - y^\top x \tag{4.19a}$$

$$\text{s.t. } y_i - y_{i-1} \geq -c_i, \; \forall 2 \leq i \leq n, \tag{4.19b}$$

$$y_n \leq c_{n+1}, \tag{4.19c}$$

$$y_i \geq 0, \; \forall 1 \leq i \leq n. \tag{4.19d}$$

Let $\mathcal{Y}$ denote the feasible region of (4.19) and $\mathcal{Y}^{(xp)}$ the set of extreme points of $\mathcal{Y}$. Because (4.19) is a linear program, it can be equivalently written as

$$f_1(x,\xi) = \max_{y^{(i)} \in \mathcal{Y}^{(xp)}} \left\{ \xi^\top y^{(i)} - x^\top y^{(i)} \right\},$$

that is, $f_1(x,\xi)$ is convex and piecewise linear. It then follows from Theorem 4.4.2 that

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[f_1(x,\xi)] &= \min_{p \in \mathbb{R}^n} \mathbb{E}_{\mathbb{Q}} \left[\max_{y^{(i)} \in \mathcal{Y}^{(xp)}} \left\{ \zeta \|y^{(i)} - p\|_{\mathbb{G}*} + (\mu - x)^\top y^{(i)} \right\} \right] \\ &= \min_{p \in \mathbb{R}^n} \mathbb{E}_{\mathbb{Q}_K} \left[\max_{y \in \mathcal{Y}} \left\{ \zeta \|y - p\|_{\mathbb{G}*} + (\mu - x)^\top y \right\} \right]. \end{aligned} \tag{4.20}$$

The second equality holds because $\zeta\|y-p\|_{G*}+(\mu-x)^\top y$ is convex on y. By an essentially identical proof of the Proposition 2 in [95], we recast the embedded maximization problem in (4.20) as

$$\min_{\lambda} \; \sum_{j=1}^{n+1} \lambda_j \tag{4.21a}$$

$$\text{s.t. } \sum_{j=k}^{\min\{\ell,n\}} \lambda_j \geq \sum_{j=k}^{\min\{\ell,n\}} \left[\zeta \max \left\{ -\frac{\pi_{j\ell} - p_j}{c_j^-}, \frac{\pi_{j\ell} - p_j}{c_j^+} \right\} + (\mu_j - x_j)\pi_{j\ell} \right],$$

$$\forall k \in [n], \; \forall \ell \in [k, n+1]_{\mathbb{Z}}. \tag{4.21b}$$

The proof is completed by introducing auxiliary variables ϕ to linearize the maximum terms on the right-hand side of constraints (4.21b) and plugging formulation (4.21) back into (4.20).

Second, for objective function f_2, we represent variables w_i, $i \in [2, n+1]_{\mathbb{Z}}$ as follows

by constraints (4.16b)–(4.16c):

$$w_i = \max_{j\in[i]} \left\{ \sum_{k=j}^{i-1} (\xi_k - x_k) \right\}.$$

Accordingly, we rewrite $f_2(x,\xi)$ as

$$\begin{aligned} f_2(x,\xi) &= \max_{i\in[2,n+1]_{\mathbb{Z}}} \left\{ c_i \max_{j\in[i]} \left\{ \sum_{k=j}^{i-1} (\xi_k - x_k) \right\} \right\} \\ &= \max_{\substack{i\in[2,n+1]_{\mathbb{Z}} \\ j\in[i]}} \left\{ c_i \sum_{k=j}^{i-1} (\xi_k - x_k) \right\} \\ &= \max_{\substack{i\in[2,n+1]_{\mathbb{Z}} \\ j\in[i]}} \left\{ a_{ij}^\top \xi - b_{ij}(x) \right\}. \end{aligned}$$

As $f_2(x,\xi)$ is convex and piecewise linear, we have by Theorem 4.4.2 that

$$\sup_{\mathbb{P}} \mathbb{E}_{\mathbb{P}\in\mathcal{D}}[f_2(x,\xi)] = \min_{p\in\mathbb{R}^n} \mathbb{E}_{\mathbb{Q}_K} \left[\max_{\substack{i\in[2,n+1]_{\mathbb{Z}} \\ j\in[i]}} \left\{ \zeta \|a_{ij} - p\|_{\mathbb{G}*} + a_{ij}^\top \mu - b_{ij}(x) \right\} \right]. \tag{4.22}$$

The proof is completed by introducing auxiliary variables θ and φ to linearize formulation (4.22). □

To infer the shape information in $\mathcal{D}$, we perform regression analysis on a set of N historical data $\{\xi^i\}_{i=1}^N$ randomly generated from a log-normal distribution.[2] We consider two types of functions, polynomial and exponential, in this analysis. For the polynomial functions, we assume that

$$\mathbb{P}\{\xi \in \Xi_r\} \;=\; \beta_0 + \sum_{j=1}^{d} \beta_j r^j + \varepsilon,$$

where d represents the degree of the polynomial function, $\{\beta_j\}_{j=0}^d$ represent the polynomial coefficients to be fitted, and ε represents the unobserved random error. To perform the regression, we surrogate $\mathbb{P}\{\xi \in \Xi_r\}$ with its empirical estimate $(1/N)\sum_{i=1}^N \mathbb{1}_{\Xi_r}(\xi^i)$ (in this case, we interpret the unobserved random error ε as the difference between $\mathbb{P}\{\xi \in \Xi_r\}$ and its empirical estimate) and obtain $\{\beta_j\}_{j=0}^d$ via the least square estimation. Regression to the exponential function can be done similarly. In this study, we consider three polynomial

[2]Note that the log-normality is not assumed in the (DR-ASP) model, and it is only used to generate training and testing data.

functions with $d = 1, 2, 3$ and denote them as P1, P2, and P3, respectively. Additionally, we denote EXP as the regression to the exponential function. Furthermore, for each function we obtain two set of coefficients as detailed below.

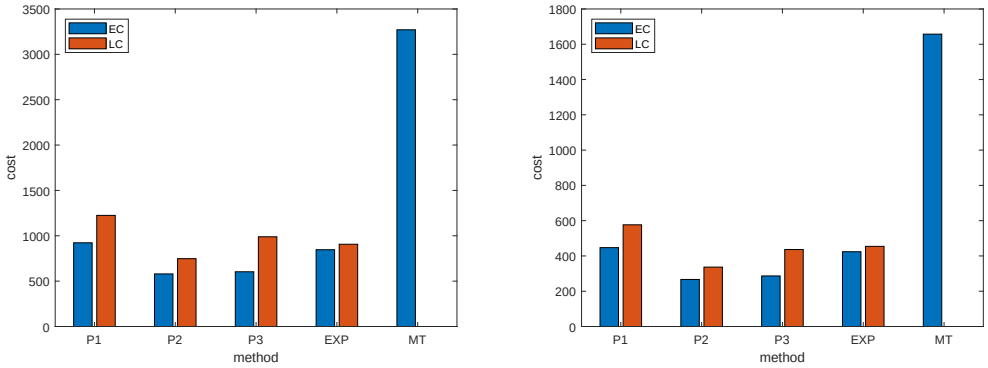
- EC: the coefficients that attain the least square error, e.g., P1-EC denotes the degree-one polynomial function with coefficients attaining the least square error.
- LC: the coefficients pertaining to the 95% lower confidence band, e.g., EXP-LC denotes the exponential function with coefficients pertaining to the 95% lower confidence band.

As a benchmark, we also consider the model with no shape information, denoted as MT. That is, MT only considers the mean and support information in the ambiguity set.

We report the results in Table 4.1 and Figures 4.2–4.4. In these results, we use TWT to represent the DR-ASP model with cost function f_1 and MWT for the DR-ASP model with f_2, respectively. For each model, 1,000 samples are generated from each function to solve the problem by using sample average approximation.

Table 4.1: Comparisons on the (DR-ASP) out-of-sample performance

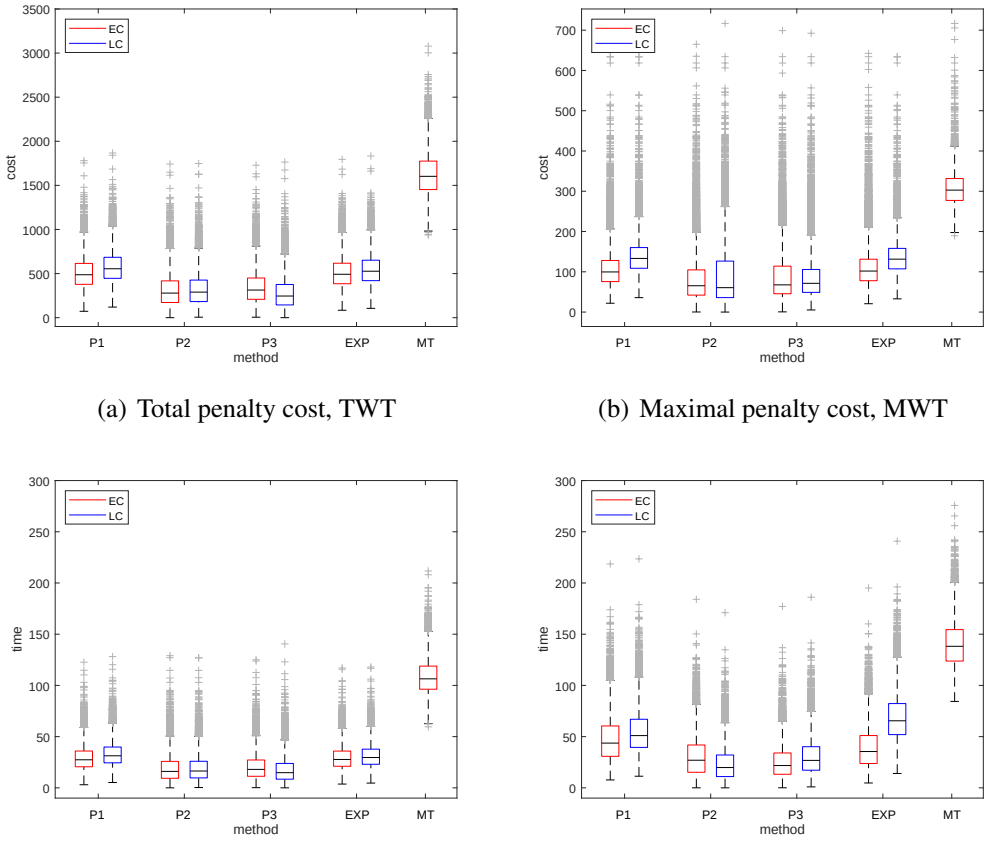
		TWT			MWT		
		Objective value	Out-of-sample test		Objective value	Out-of-sample test	
			Cost	Ind. waiting		Cost	Ind. waiting
P1	EC	922.66	509.56	29.21	447.05	108.71	48.00
	LC	1224.75	577.32	33.06	576.43	139.38	55.29
P2	EC	579.60	316.10	19.27	266.71	88.73	31.01
	LC	747.78	324.56	19.40	336.76	91.77	23.59
P3	EC	603.26	345.78	20.59	286.62	92.41	25.84
	LC	989.06	281.28	17.75	436.30	91.22	30.70
EXP	EC	846.29	513.07	29.36	423.79	115.02	39.63
	LC	906.71	547.22	31.22	454.29	137.51	69.12
MT		3270.66	1624.84	108.29	1657.35	306.68	140.70



(a) TWT (b) MWT

Figure 4.2: Objective value comparison

From Table 4.1 (see column Optimal Value) and Figure 4.2, we observe that, for both TWT and MWT, MT yields the largest optimal values from the corresponding (DR-ASP) model. In fact, the MT optimal values often double or triple the corresponding (DR-ASP) models with shape information. This indicates that the shape information can help to significantly strengthen the ambiguity set and reduce the conservativeness of the DRO model. This is further confirmed by only looking at the results generated from the (DR-ASP) models with shape information. For example, the polynomial function with degree two or three yields a smaller optimal value than that with degree one, and adopting EC in the shape information leads to a smaller optimal value than adopting LC. This indicates that the ambiguity set and the (DR-ASP) model are sensitive to the shape information. This highlights the importance of incorporating shape information in DRO models.



(a) Total penalty cost, TWT

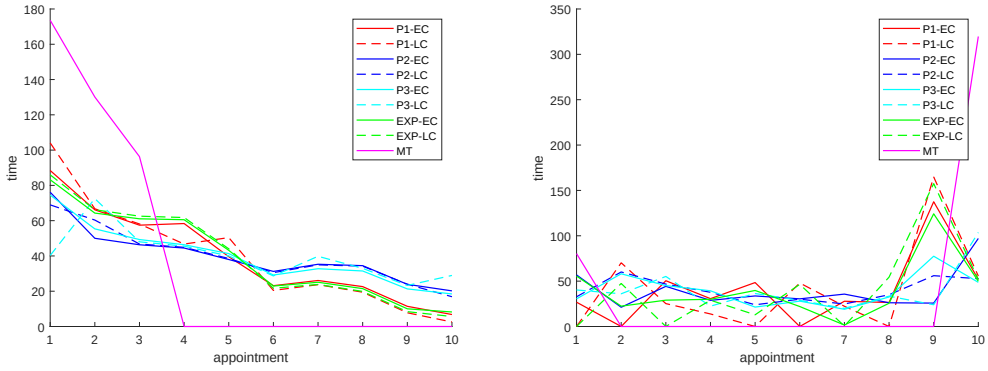
(b) Maximal penalty cost, MWT

(c) Individual waiting time, TWT

(d) Individual waiting time, MWT

Figure 4.3: Comparisons on the out-of-sample performance

Figure 4.3 and Table 4.1 (see column Out-of-Sample) report the out-of-sample performance of TWT, MWT, and MT models under various settings. Observations we make from the comparisons of optimal values carry on in the comparison of out-of-sample performance. Notably, we observe that all three polynomial models (i.e., Pd with $d = 1, 2, 3$) outperform the exponential model (EXP). This indicates that polynomial functions are better fit for learning the shape information in the (DR-ASP) model. More importantly, we observe that the degree-two polynomial function yields the best performance in nearly all aspects (see Figure 4.3(a) for total penalty cost, Figure 4.3(b) for maximal penalty cost, Figure 4.3(c) for individual waiting time in TWT, and Figure 4.3(d) for individual waiting time in MWT) among all the approaches we test. This indicates that the concentration inequalities (4.1) may be best fitted when $\epsilon(r)$ is a quadratic function. Decreasing the degree to one (i.e., using linear regression) may lose certain shape information, and increasing it to three (i.e., using cubic regression) may not yield much additional benefit.



(a) TWT (b) MWT

Figure 4.4: Optimal appointment schedule comparison

Figure 4.4 report the optimal appointment schedules generated in various models. First, we observe that the time slot assignment in MT is significantly more variable than that in other DR-ASP models with shape information. In particular, there are a good portion of zero-length time slots in the MT schedules (both in TWT and in MWT), indicating batch arrivals (i.e., multiple appointments are scheduled to arrive at the same time). This indicates that the MT schedule may be impacted by certain extreme scenarios of service durations (e.g., short service durations for multiple appointments in a row). The incorporation of shape information mitigates the impacts from such unlikely scenarios. Second, by comparing Figures 4.4(a) and 4.4(b), we observe that the optimal schedules in TWT generally assign longer time slots for the earlier appointments, while those in MWT assign the time slots more evenly, possibly with longer time slots for the last two appointments. This indicates that TWT puts more emphasis on avoiding waiting time accumulation earlier in the day to minimize the total penalty cost. In contrast, MWT treats most appointments equally and tends to assign longer time slots at the end of the day to avoid long overtime.

4.6.2 Risk-Constrained Optimal Power Flow

We consider a RC-OPF model that aims to minimize the total power production and reserve cost under safety constraints, including the transmission line capacity limits and reserve capacity limits [86]. In face of uncertainty (e.g., wind power), risk constraints are often incorporated to make sure that these safety constraints are satisfied with high probability (see e.g., [132]). In this experiment, we employ the conditional Value-at-Risk (CVaR) to protect the transmission line capacity limits and reserve capacity limits in a multi-period optimal power flow model. Intuitively, the CVaR of a random variable at confidence level

Table 4.2: Comparisons on the (RC-OPF) out-of-sample performance

		ℓ_∞-norm						ℓ_2-norm					
		$\epsilon = 0.2$		$\epsilon = 0.6$		$\epsilon = 1$		$\epsilon = 0.2$		$\epsilon = 0.6$		$\epsilon = 1$	
		cost	violation	cost	violation	cost	violation	cost	violation	cost	violation	cost	violation
P1	FD	11431	0.00%	11181	0.00%	9802	0.56%	9101	0.00%	8940	0.00%	7953	0.00%
	LD	10899	0.00%	10736	0.00%	9661	0.56%	8854	0.00%	8814	0.00%	8099	0.00%
P2	FD	11239	0.00%	9109	0.00%	7530	2.78%	8209	0.00%	9102	0.00%	6958	0.93%
	LD	10786	0.00%	9098	0.00%	7578	3.70%	7869	0.19%	8831	0.00%	6753	2.41%
P3	FD	11031	0.00%	10191	0.00%	8852	1.48%	8619	0.00%	8731	0.00%	7837	0.00%
	LD	10829	0.00%	9912	0.00%	8750	2.04%	8337	0.00%	8839	0.00%	7427	1.67%
EXP	FD	11405	0.00%	10555	0.00%	9497	0.37%	8781	0.00%	9062	0.00%	8210	0.00%
	LD	10901	0.00%	10409	0.00%	9486	0.56%	8656	0.00%	8867	0.00%	7850	0.00%
MT	FD	11406	0.00%	11406	0.00%	11406	0.00%	8795	0.00%	8795	0.00%	8795	0.00%
	LD	10875	0.00%	10875	0.00%	10875	0.00%	8681	0.00%	8681	0.00%	8681	0.00%

$1-\epsilon$ evaluates the conditional expectation at the upper ϵ-tail of the distribution of this random variable. A power system operator can adjust the confidence level to reflect her risk preference. For example, a risk-neutral operator sets $\epsilon = 1$, in which case the CVaR reduces to the expected value of the random variable. On the other hand, a risk-averse operator sets ϵ to be close zero, in which case the CVaR focuses on scenarios with smaller likelihood but larger impacts. For a group of constraints $\{a_i(x)^\top \xi + b_i(x) \leq 0,\ \forall i \in [I]\}$ subject to ambiguous uncertainty ξ, we consider the following distributionally robust CVaR constraint:

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\text{-CVaR}_\epsilon \left(\max_{i\in[I]} \left\{ a_i(x)^\top \xi + b_i(x) \right\} \right) \leq 0, \tag{4.23}$$

where the ambiguity set $\mathcal{D}$ incorporates shape information. By the definition of CVaR (see, e.g., [114]), we can recast (4.23) as

$$\exists \beta \in \mathbb{R}: \quad \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}} \left[\max_i \left\{ a_i(x)^\top \xi + b_i(x) - \beta \right\} \right]_+ \leq 0. \tag{4.24}$$

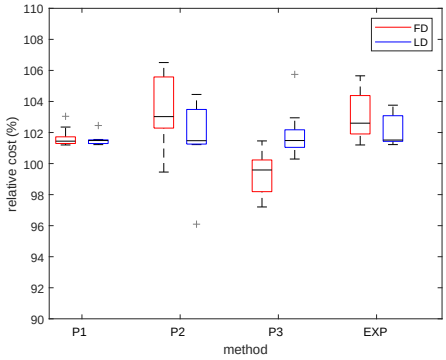
Then, we follow Theorem 4.4.2 to recast (4.24) as expected value constraint. As a consequence, we can solve the resulting RC-OPF formulation by the sample average approximation algorithm as described in [146].

We conduct numerical experiments of the RC-OPF model on the IEEE 30-bus system, which contains 30 nodes and 41 transmission lines (see `http://www.ee.washington.edu/research/pstca/` for the data set). We place nine thermal generators and three wind farms in this system. The three wind farms are based on three onshore sites analyzed by the National Renewable Energy Laboratory (NREL) [41, 40, 90, 79]. The random variables ξ represent wind forecasting errors. We consider the following two training data sets:

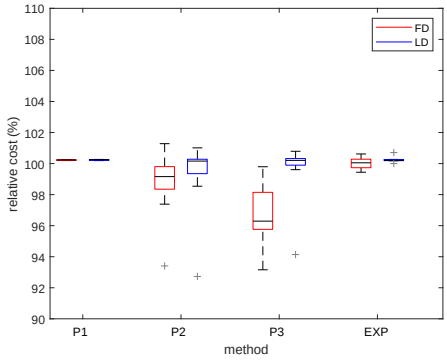
- FD: 60 days of wind forecasting error data per year for six years (2007–2012).
- LD: 30 days of wind forecasting error data per year for three years (2010–2012).

Intuitively, FD allows a larger data set for the ambiguity set to learn from, while LD allows a shorter learning time window but possibly with high data quality. In addition, we follow the same regression approach described in Section 4.6.1 to calibrate the shape information in $\mathcal{D}$ using P1, P2, P3, and EXP. In each calibration model, we consider two types of norms: the ℓ_∞-norm and the ℓ_2-norm. Moreover, the ϵ value in the CVaR constraints are set to be 1, 0.6, and 0.2. Finally, after obtaining an optimal solution of the RC-OPF model in each experiment setting, we conduct out-of-sample evaluation of the optimal solution based on the wind forecasting errors in 60 days in 2013. We evaluate the out-of-sample performance based on the following two criteria:

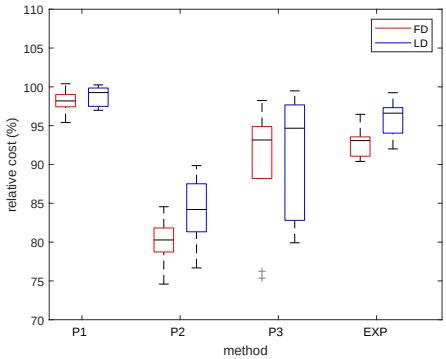
1. Cost: the average out-of-sample total cost in the 60 days, including the power production cost, reserve cost, and the penalty cost due to constraint violation.
2. Violation: the average out-of-sample frequency of violating any safety constraints in the 60 days.



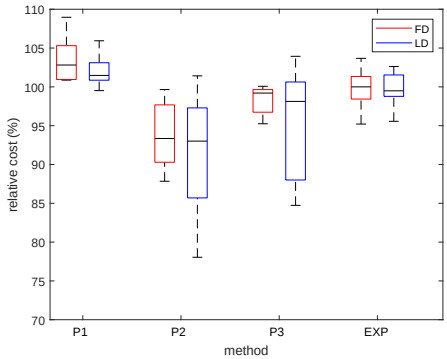
(a) ℓ_∞ with $\epsilon = 0.2$



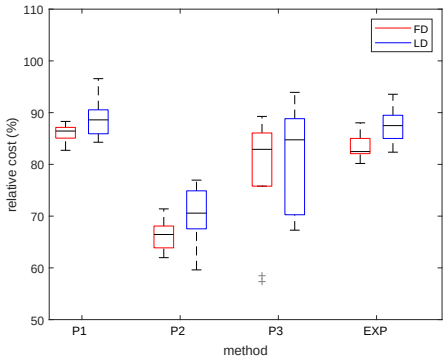
(b) ℓ_2 with $\epsilon = 0.2$



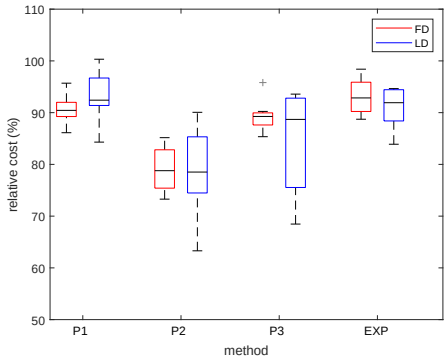
(c) ℓ_∞ with $\epsilon = 0.6$



(d) ℓ_2 with $\epsilon = 0.6$

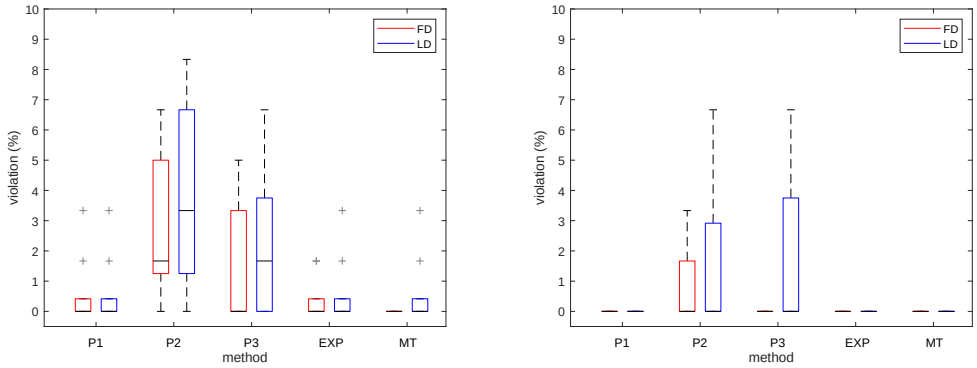


(e) ℓ_∞ with $\epsilon = 1$



(f) ℓ_2 with $\epsilon = 1$

Figure 4.5: Out-of-sample cost comparison



(a) ℓ_∞ with $\epsilon = 1$ (b) ℓ_2 with $\epsilon = 1$

Figure 4.6: Out-of-sample violation comparison (the violation frequency almost always becomes zero for all approaches when $\epsilon = 0.2$ or 0.6)

We report the results in Table 4.2 and Figures 4.5–4.6. Note that we only report the out-of-sample violation when $\epsilon = 1$ in Figure 4.6 because the violation frequency almost always becomes zero for all approaches when $\epsilon = 0.2$ or 0.6. First, Figures 4.5–4.6 show that, as ϵ increases from 0.2 to 1, the out-of-sample cost of all approaches decrease and the out-of-sample violation of all approaches increase. This demonstrates how the CVaR confidence level affects the risk-averseness of the RC-OPF solution. Second, by comparing MT with other DRO models with shape information, we observe that MT always yields the lowest out-of-sample violation, confirming that MT provides the most conservative RC-OPF solutions because its ambiguity set does not incorporate any shape information. However, we also observe that lower violation frequency does not always result in higher average cost. For example, when the risk-averseness is high (e.g., when $\epsilon = 0.2$ under the ℓ_∞ norm), most DRO models with shape information yield higher out-of-sample cost than MT does. Intuitively, this indicates that the high risk-averseness of CVaR offsets the conservativeness of the MT ambiguity set and the RC-OPF solutions of all approaches become similarly conservative. In this case, as MT leads to a deterministic formulation and all other DRO models lead to expectation constraints after reformulation, the RC-OPF solutions in the DRO models with shape information can become even more conservative due to the sample average approximation algorithm. Third, from Table 4.2, we observe that the average cost generated under the ℓ_2-norm are always lower than those generated under the ℓ_∞-norm. Meanwhile, the violation frequency generated under the ℓ_2 norm are lower. This suggests that the shape information characterized under the ℓ_2-norm is more effective than that under the ℓ_∞-norm. Fourth, the LD training data set tends to give RC-OPF solutions with lower average costs and higher violation frequencies. Note that LD contains the more

recent data. Hence, if the power system operator could tolerate slightly higher risk, she can ignore some older data on purpose in exchange of lower average costs. Finally, P2 almost always yields the lowest out-of-sample average cost, and lowest violation frequencies except when $\epsilon = 1$, among all tested approaches. This indicates that the concentration inequalities (4.1) may be best fitted when $\epsilon(r)$ is a quadratic function.

4.7 Conclusion

In this chapter, we investigated a DRO framework that models shape information in a computationally tractable manner. A new class of ambiguity sets with shape information were proposed in work. In particular, we incorporated concentration inequalities can be derived from various shape information into the ambiguity set, potentially making it infinitely constrained. We showed that the resulting DRO model can be conservatively approximated by a stochastic program with respect to an (unambiguous) probability distribution. This facilitates efficient solution algorithms (e.g., sample average approximation) for DRO models with shape information. In addition, we showed that, under mild conditions, this approximation is tight for a wide class of DRO models. We further extended the framework to give tractable reformulation of the two-stage adaptive DRO models with fixed recourse if the ambiguity set is defined based on the ℓ_1 norm. Finally, we demonstrated the theoretical results via computational case studies on the appointment scheduling problem and the risk-constrained optimal power flow problem.

CHAPTER 5

Adaptive Robust Transmission Expansion Planning with Prioritization

5.1 Introduction

The transmission grid is a critical infrastructure that connects large-scale electricity generation with the distribution system that further serves industrial and household electricity loads. The transmission expansion planning (TEP) aims to decide an optimal augmentation plan of an existing transmission grid to serve the forthcoming electric load while satisfying the security constraints. Augmentation decisions include installing new transmission lines and adding new circuits to existing lines. The investment on the transmission network is capital-intensive and has a long-term impact on the power system [94]. For instance, the European Network of Transmission System Operators for Electricity had a total budget of over EUR 100 billion for their 2012–2022 electrical grid investments [29]. The TEP problem has been investigated over decades (see, e.g., review articles [84, 85]). Although TEP decisions can be made based on the ac power flow model (see, e.g., [112, 66, 71]), the dc power flow is commonly applied largely because it leads to more tractable formulations (see, e.g., [122, 8, 69]). The TEP problem based on dc power flow can be formulated as a mixed-integer nonlinear program. Various methods have been applied to solve the TEP problem, such as linear programming [140], nonlinear programming [159, 5], mixed-integer programming [7], game theory [48], hierarchical decomposition [118], and dynamic programming [43]. There are also approaches based on Meta-heuristic (see e.g., [128, 111, 6]).

The electricity load uncertainty due to, e.g., demographic shift, has been long existing in the TEP literature (see, e.g., [117, 168]). During the last decade, renewable energy (e.g., wind and solar) has taken a growing share in the energy mix [110]. For example, the U.S. Department of Energy has analyzed a scenario that wind power contributes 20% to

the total energy utilization by 2030 [91]. Other uncertainty sources in the TEP problem include the availability of power system components [117] and the electricity markets [157]. These uncertainties make the TEP problem very challenging because the current design of transmission grid may not well serve the load and renewable energy to be realized in the future. In addition, planning under uncertainty requires to find an expansion plan that optimizes a specific metric (e.g., expected cost, worst-case cost, etc.) over a huge set of future scenarios, which is significantly more complicated than planning for a deterministic future. Stochastic programming has been studied for more than 60 years (see, e.g., [36, 96, 76, 54, 77, 62, 106, 22]). Stochastic TEP (see, e.g., [4, 92, 28]) minimizes the expected total cost of the future transmission grid while satisfying the security constraints. As this approach often models the uncertainties by using a set of scenarios, it brings two challenges: first, the true probability distribution of the uncertain parameters is usually unavailable given the long planning horizon; and second, if a large number of scenarios are generated to well represent the uncertainties, the resulting stochastic program becomes computationally intractable. Robust optimization (RO), on the other hand, only requires the knowledge of support (termed an uncertainty set) of the uncertain parameters [12]. For a given uncertainty set, robust TEP minimizes the cost of the future transmission grid with regard to the worst-case scenario within the uncertainty set. Although the uncertainty set usually consists of an infinite number of scenarios, the worst-case scenario can be computed by solving an optimization model. This saves the effort of scenario enumeration and leads to better computational efficacy of robust TEP. Recently, robust TEP has received growing attention in the literature [70, 32, 101].

To compute the worst-case cost, robust TEP needs to compute the optimal power flow after the uncertainties are realized, leading to a two-stage (adaptive) RO model. In the first stage, a set of line/circuit are installed; and in the second stage, after the uncertainties (e.g., budget, load, and renewable energy) are realized, an additional set of line/circuit are adaptively installed and the power is re-dispatched to yield an optimal power flow based on the augmented grid. Unfortunately, the adaptive installation result in a RO model with mixed-integer recourse, which is computationally prohibitive. Existing literature addresses this challenge in two ways. The first stream of literature waives the option of adaptive installation and accordingly robust TEP reduces to a two-stage RO model with continuous recourse (see, e.g., [69]). We can then apply the column-and-constraint generation algorithm [160] to efficiently solve robust TEP. The second stream of literature designs decision rules to determine the adaptive installation as a function of the first-stage installation and the realized uncertainties (see, e.g., [60, 16]). In this work, we adopt the decision rule of prioritization [80, 81]. Prioritization was initially proposed to solve uncertain resource-constrained

activity-selection problems. This decision rule involves optimally placing activities into a priority list before the uncertainty is realized, and making an activity selection following the priority list after the uncertainty realization. Prioritization is consistent with the practice in many industrial and governmental decision-making processes [24, 123] and can provide more interpretable solutions.

The remainder of this chapter is organized as follows. We describe the mathematical formulations of the proposed robust TEP with prioritization (TEP-P) model in Section 5.2. In Section 5.3, we present computationally tractable reformulations of the proposed model and a tailored column-and-constraint generation algorithm. Section 5.4 validates the proposed approach via case studies based on the real-world wind power data. Section 5.5 summarizes this study. In addition, we list the nomenclature in Section 5.6.

5.2 Mathematical Formulation

We first present the deterministic TEP model. Then we generalize it to stochastic and robust TEP-P models.

5.2.1 Deterministic TEP

If the installation budget, load, and renewable energy is deterministic and known, then the TEP model [69] can be formulated as follows:

$$\min_{n^1,f,g,h,r,\theta} \quad \sum_{b\in\mathcal{B}}(c_b^g g_b + c_b^r r_b) \tag{5.1a}$$

$$\text{s.t.} \quad Af + g + h + r = \tilde{d}, \tag{5.1b}$$

$$f_{ij} = \gamma_{ij}\left(n_{ij}^0 + \sum_{k\in\mathcal{N}_{ij}^1} n_k^1\right)(\theta_i - \theta_j),\ \forall (i,j)\in\mathcal{L}, \tag{5.1c}$$

$$|f_{ij}| \le \left(n_{ij}^0 + \sum_{k\in\mathcal{N}_{ij}^1} n_k^1\right)\bar{f}_{ij},\ \forall (i,j)\in\mathcal{L}, \tag{5.1d}$$

$$\mathbf{0} \le g \le \bar{g}, \tag{5.1e}$$

$$\mathbf{0} \le h \le \tilde{h}, \tag{5.1f}$$

$$\mathbf{0} \le r \le \tilde{d}, \tag{5.1g}$$

$$-\frac{\pi}{2}\mathbf{1} \le \theta \le \frac{\pi}{2}\mathbf{1}, \tag{5.1h}$$

$$\sum_{(i,j)\in\mathcal{L}} \sum_{k\in\mathcal{N}_{ij}^1} c_k n_k^1 \leq B^1, \tag{5.1i}$$

$$n^1 \in \{0,1\}^{\mathcal{N}^1}, \tag{5.1j}$$

where $\mathbf{1}$ (resp. $\mathbf{0}$) represents an all-one (resp. all-zero) vector of proper dimension, $\mathcal{B}$ is the set of buses, $\mathcal{L}$ is the set of lines, $\mathcal{N}_{ij}^1$ is the set of circuits can be installed to line (i,j) now, $\mathcal{N}^1 = \bigcup_{(i,j)\in\mathcal{L}} \mathcal{N}_{ij}^1$, c_b^g (resp. c_b^r) is the unit power generation (resp. load curtailment penalty) cost at bus b, A is the node-branch incidence matrix of the transmission grid, d are the loads, γ_{ij} is the susceptance of of each circuit on line (i,j), n_{ij}^0 is the initial number of circuits on line (i,j), $\bar{f}_{ij}$ is the power flow capacity of each circuit on line (i,j), $\bar{g}$ is the thermal generation capacity, $\tilde{h}$ is the available renewable energy, B^1 is the current budget for circuits installation, c_k is installation cost of circuit k, variables g represent power generation amounts from thermal generators, variables h represent power generation amounts from renewable generators, variables r represent load curtailment amounts, variable f_{ij} represents the power flow on line (i,j), binary variable n_k^1 represents whether circuit $k \in \mathcal{N}^1$ is installed now, and variable θ_b represents the phase angle at bus b.

The goal of formulation (5.1a)–(5.1j) is to find an optimal transmission grid augmentation to meet the load with the minimum total cost, including power generation costs and load curtailment penalty costs. Constraints (5.1b) describe the nodal power balance and constraints (5.1c) are the dc line flow equations. Constraints (5.1d)–(5.1h) specify the bounds of power flow, thermal power generation amount, renewable energy, load curtailment, and phase angle, respectively. Constraint (5.1i) describes the TEP budget limit.

5.2.2 TEP with Prioritization

In an uncertain environment, the TEP budget can be divided into two parts: the budget B^1 available now and the (uncertain) budget $\tilde{B}^2$ to be provided in the future. Let $\mathcal{N}_{ij}^2$ be the set of circuits that can be installed to line (i,j) in the future and $\mathcal{N}^2 = \bigcup_{(i,j)\in\mathcal{L}} \mathcal{N}_{ij}^2$. Accordingly, we use variables n_k^1, $k \in \mathcal{N}^1$ to denote the installation plan in the present and n_k^2, $k \in \mathcal{N}^2$ to denote the plan in the future. In practice, the renewable generation $\tilde{h}$, load $\tilde{d}$, as well as the future budget $\tilde{B}^2$ are uncertain. Recently, [80] proposed a prioritization decision rule to deal with the uncertainty in resource-constrained activity-selection problems. The concept of prioritization is consistent with the widely-used decision-making process in practice [24, 123]. In this approach, we rank all the activities in a priority list, and commit to each activity based on its priority and the realized uncertainty. Prioritization requires that an activity can be selected only if all the activities with higher priority have been se-

lected. To apply the prioritization decision rule, we introduce binary decision variables $s_{kk'}$ to model the priority order between n_k^2 and $n_{k'}^2$. Specifically, $s_{kk'}$ equals 1 if n_k^2 is ranked at least as high as $n_{k'}^2$ and it equals 0 otherwise.

5.2.2.1 Stochastic TEP-P

In (5.2), we present the stochastic TEP-P model. Let Ω be the set of scenarios, where $|\Omega| < \infty$. For each $\omega \in \Omega$, we let $(\tilde{B}^{2\omega}, \tilde{d}^{\omega}, \tilde{h}^{\omega})$ denote the random variable realizations with probability q^{ω} and $(f^{\omega}, g^{\omega}, h^{\omega}, r^{\omega}, \theta^{\omega}, n^{2\omega})$ be the corresponding decision variables for scenario ω.

$$\min_{\substack{n^1,s,f^{\omega}\\ g^{\omega},h^{\omega},r^{\omega}\\ \theta^{\omega},n^{2\omega}}} \sum_{\omega\in\Omega} q^{\omega}\left(\sum_{b\in\mathcal{B}} c_b^g g_b^{\omega} + c_b^r r_b^{\omega}\right) \tag{5.2a}$$

$$\text{s.t. } Af^{\omega} + g^{\omega} + h^{\omega} + r^{\omega} = \tilde{d}^{\omega},\ \ \forall\omega\in\Omega, \tag{5.2b}$$

$$\gamma_{ij}\left(n_{ij}^0 + \sum_{k\in\mathcal{N}_{ij}^1} n_k^1 + \sum_{k\in\mathcal{N}_{ij}^2} n_k^{2\omega}\right)(\theta_i^{\omega} - \theta_j^{\omega}) = f_{ij}^{\omega},\ \ \forall(i,j)\in\mathcal{L},\ \forall\omega\in\Omega, \tag{5.2c}$$

$$|f_{ij}^{\omega}| \le \left(n_{ij}^0 + \sum_{k\in\mathcal{N}_{ij}^1} n_k^1 + \sum_{k\in\mathcal{N}_{ij}^2} n_k^{2\omega}\right)\bar{f}_{ij},\ \ \forall(i,j)\in\mathcal{L},\ \forall\omega\in\Omega, \tag{5.2d}$$

$$\mathbf{0} \le g^{\omega} \le \bar{g},\ \ \forall\omega\in\Omega, \tag{5.2e}$$

$$\mathbf{0} \le h^{\omega} \le \tilde{h}^{\omega},\ \ \forall\omega\in\Omega, \tag{5.2f}$$

$$\mathbf{0} \le r^{\omega} \le \tilde{d}^{\omega},\ \ \forall\omega\in\Omega, \tag{5.2g}$$

$$-\frac{\pi}{2}\mathbf{1} \le \theta^{\omega} \le \frac{\pi}{2}\mathbf{1},\ \ \forall\omega\in\Omega, \tag{5.2h}$$

$$s_{kk'} + s_{k'k} \ge 1,\ \ \forall k,k'\in\mathcal{N}^2, k\ne k', \tag{5.2i}$$

$$s_{kk''} \ge s_{kk'} + s_{k'k''} - 1,\ \ \forall k,k',k''\in\mathcal{N}^2, k\ne k'\ne k'', \tag{5.2j}$$

$$s \in \{0,1\}^{\mathcal{N}^2\times\mathcal{N}^2}, \tag{5.2k}$$

$$n_k^{2\omega} \ge n_{k'}^{2\omega} + s_{kk'} - 1,\ \ \forall k,k'\in\mathcal{N}^2, k\ne k',\ \forall\omega\in\Omega, \tag{5.2l}$$

$$\sum_{k\in\mathcal{N}^2} c_k n_k^{2\omega} \le \tilde{B}^{2\omega},\ \ \forall\omega\in\Omega, \tag{5.2m}$$

$$n^{2\omega} \in \{0,1\}^{\mathcal{N}^2},\ \forall\omega\in\Omega, \tag{5.2n}$$

$$\sum_{k\in\mathcal{N}^1} c_k n_k^1 \le B^1, \tag{5.2o}$$

$$n^1 \in \{0,1\}^{\mathcal{N}^1}. \tag{5.2p}$$

The objective function (5.2a) consists of expected generation costs and expected load curtailment penalty costs. Constraints (5.2b)–(5.2h) correspond to constraints (5.1b)–(5.1h) under each scenario ω. For each pair of circuits to be installed, constraints (5.2i)–(5.2k) ensure that either they have the same priority or one has higher priority than the other. Constraints (5.2l) describe the priority rule, i.e., a circuit can be installed only if all the circuits with higher rank have been installed. Constraints (5.2m) describe the future TEP budget limit and Constraints (5.2o) describe the present TEP budget limit. Note that formulation (5.2) is a mixed-integer nonlinear program due to the bilinear terms in constraints (5.2c). By introducing the McCormick inequalities [98], this model can be recast as a mixed-integer linear program (MILP) that can be solved by commercial solvers, e.g., CPLEX.

5.2.2.2 Robust TEP-P

In this section, we propose a two-stage adaptive robust formulation with prioritization to deal with the TEP problem under uncertainty. In this model, n^1 and s are the first-stage decision variables, parameters $(\tilde{B}^2, \tilde{d}, \tilde{h})$ are random variables and $(f, g, h, r, \theta, n^2)$ are second-stage decision variables. In the future installation plan, we assume, in each scenario, the budget will be used as much as possible. Then, constraints (5.2l)–(5.2m) can be replaced by constraints (5.3).

$$\tilde{B}^2 - \sum_{\substack{k' \in \mathcal{N}^2 \\ k' \neq k}} c_{k'} s_{k'k} - c_k \leq M n_k^2, \ \forall k \in \mathcal{N}^2, \tag{5.3a}$$

$$\sum_{\substack{k' \in \mathcal{N}^2 \\ k' \neq k}} c_{k'} s_{k'k} + c_k - \tilde{B}^2 \leq M(1 - n_k^2), \ \forall k \in \mathcal{N}^2. \tag{5.3b}$$

Without loss of generality, we can assume $\tilde{B}^2 \neq \sum_{k' \in \mathcal{N}^2,\ k' \neq k} c_{k'} s_{k'k} + c_k$ by perturbing $\tilde{B}^2$. We can treat n^2 as random variables because n^2 can be fully determined by the priority list s and the realization of the random variable $\tilde{B}^2$. After incorporating the priority list s, we assume the uncertainty set $\mathcal{U}(s)$ is of the form (5.4).

$$\mathcal{U}(s) := \left\{ \left(\tilde{B}^2, \tilde{d}, \tilde{h}, n^2\right) : \underline{d} \leq \tilde{d} \leq \overline{d}, \right. \tag{5.4a}$$

$$d^L \leq \sum_{b \in \mathcal{B}} \tilde{d}_b \leq d^U, \tag{5.4b}$$

$$\underline{h} \leq \tilde{h} \leq \overline{h}, \tag{5.4c}$$

$$h^L \leq \sum_{b \in \mathcal{B}} \tilde{h}_b \leq h^U, \tag{5.4d}$$

$$\underline{\psi}(\tilde{d},\tilde{h}) \le \tilde{B}^2 \le \overline{\psi}(\tilde{d},\tilde{h}), \tag{5.4e}$$

$$n^2 \in \{0,1\}^{\mathcal{N}^2}, \tag{5.4f}$$

$$\text{Constraints (5.3a) and (5.3b)}\Big\} \tag{5.4g}$$

where constraints (5.4a) and (5.4c) describe the lower and upper bounds of the renewable energy and load at each bus, respectively. Constraints (5.4b) and (5.4d) describe the total renewable energy and load bounds throughout all buses, respectively. Constraint (5.4e) set lower and upper bounds of the budget, which linearly depend on the total renewable energy and load. $\underline{\psi}(\cdot), \overline{\psi}(\cdot)$ represent budget upper and lower bounds as functions of the realized renewable energy and load, respectively. Constraints (5.4f)–(5.4g) ensure that the circuits are installed according to the priority list and the future budget. Based on $\mathcal{U}(s)$, we propose a two-stage adaptive robust TEP-P (5.5) to address the load, renewable energy, and future budget uncertainty.

$$\min_{n^1,s} \quad \max_{(\tilde{B}^2,\tilde{d},\tilde{h},n^2)\in\mathcal{U}(s)} \quad \min_{\substack{(f,g,h,r,\theta)\in \\ \Psi(\tilde{B}^2,\tilde{d},\tilde{h},n^1,n^2)}} \sum_{b\in\mathcal{BS}} (c_b^g g_b + c_b^r r_b) \tag{5.5a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{N}^1} c_k n_k^1 \le B^1, \tag{5.5b}$$

$$n_k^1 \in \{0,1\} \ \ \forall k \in \mathcal{N}^1. \tag{5.5c}$$

$$s_{kk'} + s_{k'k} \ge 1, \ \ \forall k,k' \in \mathcal{N}^2, k \ne k', \tag{5.5d}$$

$$s_{kk''} \ge s_{kk'} + s_{k'k''} - 1, \ \forall k,k',k'' \in \mathcal{N}^2, k \ne k' \ne k'', \tag{5.5e}$$

$$s \in \{0,1\}^{\mathcal{N}^2\times\mathcal{N}^2}, \tag{5.5f}$$

where

$$\Psi\left(\tilde{B}^2,\tilde{d},\tilde{h},n^1,n^2\right) = \Big\{(f,g,h,r,\theta):$$

$$Af + g + h + r = \tilde{d}, \tag{5.5g}$$

$$f_{ij} = \gamma_{ij}\left(n_{ij}^0 + \sum_{k\in\mathcal{N}_{ij}^1} n_k^1 + \sum_{k\in\mathcal{N}_{ij}^2} n_k^2\right)(\theta_i - \theta_j), \ \forall (i,j)\in\mathcal{L}, \tag{5.5h}$$

$$|\gamma_{ij}(\theta_i-\theta_j)| \le \bar{f}_{ij}, \ \forall (i,j)\in\mathcal{L} : n_{ij}^0 \ne 0, \tag{5.5i}$$

$$|\gamma_{ij}(\theta_i-\theta_j)| \le \bar{f}_{ij} + (1-n_k^1)(M-\bar{f}_{ij}), \ \forall k\in\mathcal{N}_{ij}^1, \forall (i,j)\in\mathcal{L} : n_{ij}^0 = 0, \tag{5.5j}$$

$$|\gamma_{ij}(\theta_i-\theta_j)| \le \bar{f}_{ij} + (1-n_k^2)(M-\bar{f}_{ij}), \ \forall k\in\mathcal{N}_{ij}^2, \forall (i,j)\in\mathcal{L} : n_{ij}^0 = 0, \tag{5.5k}$$

$$\mathbf{0} \le g \le \bar{g}, \tag{5.5l}$$

$$\mathbf{0} \le h \le \tilde{h}, \tag{5.5m}$$

$$\mathbf{0} \leq r \leq \tilde{d}, \tag{5.5n}$$

$$-\frac{\pi}{2}\mathbf{1} \leq \theta \leq \frac{\pi}{2}\mathbf{1}\Big\}. \tag{5.5o}$$

The objective function (5.5a) consists of power generation costs and load curtailment penalty costs. In the first stage, the present circuit installation decisions n^1 and prioritization decisions s are made. In the second stage, the generation and transmission decisions are made for each scenario. In this formulation, the load $\tilde{d}$, renewable energy $\tilde{h}$, future budget $\tilde{B}^2$, and circuit installation n^2 are subject to uncertainty and described by the uncertainty set $\mathcal{U}(s)$. Note that, in each scenario, the uncertain future circuit installation decisions n^2 will be determined by the future budget and priority list. The objective is to minimize the total generation costs and load curtailment penalty costs under the worst-case scenario. Constraints (5.5d)–(5.5e) describe the priority of each pair of circuits, i.e., either they have the same priority or one has higher priority than the other. Constraints (5.5g)–(5.5o) are the stochastic counterpart of constraints (5.1b)–(5.1h). Note that constraints (5.5i)–(5.5k) are equivalent to

$$|f_{ij}| \leq \left(n^0_{ij} + \sum_{k \in \mathcal{N}^1_{ij}} n^1_k + \sum_{k \in \mathcal{N}^2_{ij}} n^2_k \right) \bar{f}_{ij}, \forall (i,j) \in \mathcal{L}.$$

We use constraints (5.5i)–(5.5k) because they can reduce the number of bilinear terms in the reformulation.

5.3 Solution Methodology

To simplify the notation, we will use an abstract form of the robust TEP-P model (5.5). Let E_1, E_2, F_0, H_1, H_2 be constant matrices and $c, \lambda, \alpha_0, \beta_0$ be constant vectors. Let $F_1(\cdot), F_2(\cdot)$ be matrices and $\alpha_1(\cdot), \alpha_2(\cdot), \beta_1(\cdot)$ be vectors, all of which are affine functions of the input variables. Also, let y represent all second-stage variables corresponding to (f, g, h, r, θ) and ξ represent the random variables $(\tilde{B}^2, \tilde{d}, \tilde{h})$. Then, we present an abstract form of (5.5) as follows.

$$\min_{n^1, s} \quad \max_{(\xi, n^2) \in \mathcal{U}(s)} \min_{y \in \Psi(\xi, n^1, n^2)} c^\top y \tag{5.6a}$$

$$\text{s.t.} \quad E_1 n^1 + E_2 s \geq \lambda, \tag{5.6b}$$

$$n^1 \in \{0,1\}^{\mathcal{N}^1}, s \in \{0,1\}^{\mathcal{N}^2 \times \mathcal{N}^2}, \tag{5.6c}$$

where

$$\Psi\left(\xi, n^1, n^2\right) = \left\{y : \left[F_0 + F_1(n^1) + F_2(n^2)\right] y \geq \alpha_0 + \alpha_1(n^1) + \alpha_2(\xi, n^2)\right\}, \quad (5.6d)$$

and

$$\mathcal{U}(s) = \Big\{ \left(\xi, n^2\right) : H_1\xi + H_2 n^2 \geq \beta_0 + \beta_1(s), \quad (5.6e)$$

$$n^2 \in \{0,1\}^{\mathcal{N}^2}\Big\}. \quad (5.6f)$$

The proposed formulation has a min-max-min objective function and cannot be solved directly by commercial solvers. Therefore, efficient algorithms need to be developed. In Section 5.3.2, we design a tailored column-and-constraint generation algorithm to solve formulation (5.6).

5.3.1 Second-Stage Problem Reformulation

We view n^2 as random variables and the inner minimization problem of formulation (5.6) becomes a linear program. By dualizing it and combining two maximization operators, we recast the second-stage problem as follows. Strong duality holds because the inner problem is a linear program.

$$\Pi(n^1, s) = \max_{\eta, \xi, n^2} \ \left[\alpha_0 + \alpha_1(n^1) + \alpha_2(\xi, n^2)\right]^\top \eta \quad (5.7a)$$

$$\left[F_0 + F_1(n^1) + F_2(n^2)\right]^\top \eta = c, \quad (5.7b)$$

$$\eta \geq 0, \quad (5.7c)$$

$$H_1\xi + H_2 n^2 \geq \beta_0 + \beta_1(s), \quad (5.7d)$$

$$n^2 \in \{0,1\}^{\mathcal{N}^2}, \quad (5.7e)$$

where dual variables η are associated with constraints (5.6d) and dual constraints (5.7b) are associated with primal variables y. We apply the following reformulations to linearize the bilinear terms in the objective function (5.7a) and constraint (5.7b). Note that there are two types of bilinear terms:

$n_k^2\eta_j$: It is the product of an integer variable and a continuous variable. We directly apply the McCormick inequalities [98] to linearize this term. Formally, we can define auxiliary variable $z_{kj}^{n\eta}$ to replace $n_k^2\eta_j$ and add the following constraints:

$$z_{kj}^{n\eta} \leq \eta_j, \quad (5.8a)$$

$$z_{kj}^{n\eta} \leq M n_k^2, \tag{5.8b}$$
$$z_{kj}^{n\eta} \geq \eta_j - M(1 - n_k^2), \tag{5.8c}$$
$$z_{kj}^{n\eta} \geq 0. \tag{5.8d}$$

where M is a sufficiently large value that upper bounds η_j.

$\xi_i\eta_j$: It is the product of two continuous variables. We first apply the binary expansion approach [158] on dual variables η_j by adding the following auxiliary variables and constraints.

$$\eta_j \approx \sum_{t=-T_L}^{T_U} 2^t z_{j,t}^{\eta},$$
$$z_{j,t}^{\eta} \in \{0,1\}, \forall t \in [-T_L, T_U] \cap \mathbb{Z}.$$

With the binary expansion, the new bilinear term $\xi_i z_{j,t}^{\eta}$ becomes the product of an integer variable and a continuous variable. We apply similar McCormick inequalities as (5.8) to linearize these terms.

Therefore, the maximization problem (5.7) can be reformulated as a MILP that can readily be solved by commercial solvers.

5.3.2 Column-and-Constraint Generation Algorithm

In this section, we design a tailored column-and-constraint generation algorithm to solve the robust TEP-P model. We solve the following master problem iteratively by adding new primal constraints to cut off the suboptimal solutions. Note that we do not need feasibility cuts due to the relatively complete recourse of formulation (5.5).

Master problem The master problem is

$$\min_{s,n^1,\Theta,y^v,n^{2v}} \quad \Theta \tag{5.9a}$$
$$\text{s.t.} \quad E_1 n^1 + E_2 s \geq \lambda, \tag{5.9b}$$
$$n_k^1 \in \{0,1\}^{\mathcal{N}^1}, s \in \{0,1\}^{\mathcal{N}^2 \times \mathcal{N}^2}, \tag{5.9c}$$
$$\Theta \geq c^\top y^v, \ \forall v, \tag{5.9d}$$
$$\left[F_0 + F_1(n^1) + F_2(n^{2v})\right] y^v \geq \alpha_0 + \alpha_1\left(n^1\right) + \alpha_2\left(\hat{\xi}^v, n^{2v}\right), \ \forall v, \tag{5.9e}$$
$$H_1 \hat{\xi}^v + H_2 n^{2v} \geq \beta_0 + \beta_1(s), \ \forall v, \tag{5.9f}$$

$$n^{2v} \in \{0,1\}^{\mathcal{N}^2},\ \forall v, \tag{5.9g}$$

where constraints (5.9d)–(5.9g) are optimality cuts in primal form and v represents the index of the optimality cuts (formally described below) and also the iteration number. Note that the optimality cuts contain bilinear terms $n_k^1 y_j^v$ and $n_k^{2v} y_j^v$ in (5.9e). Therefore, we once again employ the McCormick inequalities [98] to linearize these terms.

Optimality cuts In the master problem (5.9), Θ represents the optimal value of the second-stage problem (5.7). Suppose in iteration v, by solving the master problem (5.9), we obtain an incumbent optimal solution $(\hat{\Theta}, \hat{s}, \hat{n}^1)$. We substitute $\hat{s}$ and $\hat{n}^1$ into the second-stage problem (5.7) to obtain the optimal objective value $\Pi(\hat{n}^1, \hat{s})$ and the corresponding worst-case scenario $\hat{\xi}^v$. If $\hat{\Theta} < \Pi(\hat{n}^1, \hat{s})$ then the solution $(\hat{\Theta}, \hat{s}, \hat{n}^1)$ is suboptimal with regard to $\hat{\xi}^v$. Therefore, we add an optimality cut (5.9d)–(5.9g) in the primal form based on $\hat{\xi}^v$.s

5.4 Case Study

To test the proposed two-stage adaptive robust TEP-P model, we conduct computational case studies based on the Garver 6-bus system [119] and the southern Brazilian 46-bus system [57]. All instances are solved by CPLEX 12.8 solver on a 64-bit Windows 10 machine with Intel(R) Core(TM) i7-6600U processor, running at 2.60 GHz with 8 GB memory.

5.4.1 Experiment System

5.4.1.1 6-bus system

The Garver 6-bus system contains six buses and nine transmission lines for adding new circuits [119], which was used in [49] originally. In this system, the total load is 760 MW and the detailed data is shown in Table 5.1 and Table 5.2. We place three wind farms in this system, i.e., in the bus 3, 4 and 6. The data of these three wind farms is given by three onshore sites from the National Renewable Energy Laboratory (NREL) [41, 40, 90, 79].

Table 5.1: 6-bus generation and load data

Generator index	Bus index	$\bar{g}$ (MW)	d (MW)	c^g (\$)	c^r (\$)
1	1	50	80	2.02	3.27
2	2	0	240	0.60	3.89
3	3	165	40	2.15	3.07
4	4	0	160	2.01	0.90
5	5	0	240	2.10	0.62
6	6	545	0	1.60	3.19

Table 5.2: 6-bus transmission line data

Line index	i-j	n_{ij}^0	Reactance (p.u.)	$\bar{f}_{ij}$ (MW)	c (10^3\$)
1	1-2	0	0.4	100	6
2	1-4	1	0.6	80	4
3	1-5	0	0.2	100	3
4	2-3	1	0.2	100	5
5	2-4	0	0.4	100	5.5
6	3-5	0	0.2	100	3.5
7	6-2	0	0.3	100	4.5
8	6-4	0	0.3	100	6.5
9	6-5	0	0.61	78	7

5.4.1.2 46-bus system

The 46-bus system is a medium-sized system representing the southern part of the Brazilian interconnected network [57]. It contains 46 buses and 71 transmission lines for adding new circuits. In this system, the total load is 6880 MW. All the relevant data can be found in [57]. We place 6 wind farms in this system. The information of these 6 wind farms is based on 6 onshore sites from NREL.

5.4.2 Experiment Setting

In this case study, we consider uncertainty in renewable energy $\tilde{h}$ and the future installation budget $\tilde{B}^2$. The uncertainty set $\mathcal{U}(s)$ is constructed based on the historical data from NREL Wind Integration National Dataset Toolkit [41, 40, 90, 79] and the circuit installation costs. We collect the historical data of wind energy from Jan. 18 to Feb. 14 in 2007–2011 as the

training data and wind energy from the same period in 2012–2013 as the test data. There are hence 140 training samples and 56 testing samples. In the uncertainty set $\mathcal{U}(s)$, we consider the upper and lower bounds of the wind power on each farm as well as the upper and lower bounds of the total wind energy of all 3 (or 6) wind farms, and the upper and lower bounds of the future TEP budget. To test our proposed two-stage robust adaptive TEP-P model, we conduct out-of-sample tests on the following approaches. We refer to them as:

- RP: two-stage robust TEP-P approach (5.5).
- RG: two-stage robust TEP approach. In this approach, we approximately solve the robust TEP model by considering a finite subset of scenarios in the uncertainty set.
- SP: two-stage stochastic TEP-P approach (5.2). Ω contains all the 140 historical data samples and the probability distribution is assumed to be a uniform distribution over the samples.
- SG: two-stage stochastic TEP approach. Similar to RG, we solve stochastic TEP by sample average approximation (SAA).
- RM: randomized approach. We generate uniformly random present and future installation plans.

We perform the out-of-sample tests based on the following two different wind data sets.

- Real data: wind data obtained from NREL Wind Integration National Dataset Toolkit.
- Misspecified data: real data plus a 10% increase, which mimics a potentially accelerated adoption of wind energy at the end of the planning horizon.

In addition, we evaluate the performance of the approaches based on the following two criteria:

- Cost: the average total costs, including power generation costs and load curtailment penalty costs.
- Time: the average CPU seconds.

All approaches are tested under different penetration levels of renewable energy, from 10% to 20%. For example, if the penetration level is 20%, the data will be scaled such that the median of historical data samples of the total wind energy from all the sites can provide 20% of the total electricity load. We also conduct a sensitivity analysis on the budget levels for RP and RG.

5.4.3 Results

5.4.3.1 6-bus system

Comparisons For the 6-bus system, we compare the above five approaches. We use the 140 historical data samples to construct the uncertainty set for RP and RG, 500 data samples generated from the obtained uncertainty set to approximately solve RG, and 30 historical data samples to solve SG using SAA. The first-stage budget (i.e., B^1) is set to be 10, and the second-stage budget (i.e., $\tilde{B}^2$) is randomly sampled from Unif$[10, 40]$. We report the out-of-sample performance of the above five approaches on both real data samples and the corresponding misspecified data samples.

Table 5.3 and Figure 5.1 compare the average total costs among the five approaches on the two data sets. In Figure 5.1, the x-coordinate represents the penetration level and the y-coordinate represents the cost. From these results, we first observe that the worst-case performance of RP and RG is better than the others, while the average performance of SP and SG is better than the others. RM performs the worst among the all the approaches in all experiments. These results are consistent with the metric used in each approach. Second, except for the worst-case cost of SP, both the worst-case and average costs become lower as the penetration level increases. This is because as we incorporate more renewable energy, the cost-effectiveness of using renewable energy becomes more significant. Since the object of SP is to optimize the average performance, it may lead to some bad worst-case out-of-sample performance. Third, the total cost difference between RP and RG is very small, which suggests that the prioritization decision rules can provide near-optimal decisions. Fourth, the results demonstrate that the performance of the RG and SG is exactly the same. The reason is that the second-stage problems of these two approaches are both fully adaptive. As a result, they just need to optimize the first-stage installation plan. Since the 6-bus system only contains three candidate circuits in the first stage, it is likely that these two approaches yield the same optimal solution. Finally, these approaches perform similarly under both real and misspecified data. This indicates that the proposed approach for transmission expansion planning can handle mild misspecification of wind power data.

Table 5.4 reports the CPU seconds of four approaches. The CPU time of RP is around 1.5 seconds, which is dramatically less than those of the other approaches. For the other three approaches, since they are all solved with a finite number of scenarios, their CPU seconds highly depend on the number of scenarios considered in each approach. This result shows that SG is the most time-consuming.

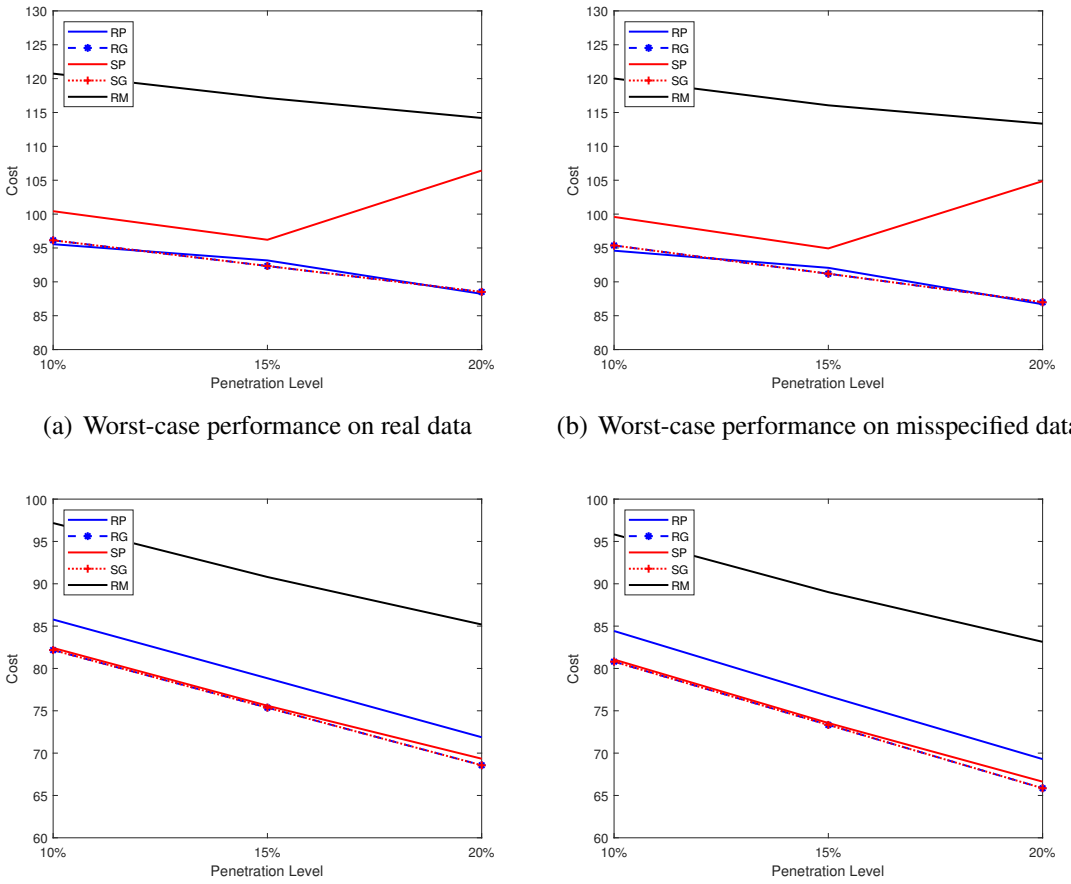
Table 5.3: Cost comparison on the 6-bus system

	Worst-case performance on real data			Worst-case performance on misspecified data		
Penetration level	10%	15%	20%	10%	15%	20%
RP	95.56	93.16	88.23	94.60	92.07	86.69
RG	96.13	92.34	88.51	95.37	91.20	86.98
SP	100.43	96.21	106.43	99.59	94.93	104.87
SG	96.13	92.34	88.51	95.37	91.20	86.98
RM	120.74	117.14	114.19	120.02	116.06	113.35
	Average performance on real data			Average performance on misspecified data		
Penetration level	10%	15%	20%	10%	15%	20%
RP	85.78	78.84	71.88	84.42	76.75	69.29
RG	82.17	75.37	68.57	80.81	73.33	65.85
SP	82.42	75.60	69.34	81.05	73.56	66.63
SG	82.17	75.37	68.57	80.81	73.33	65.85
RM	97.17	90.80	85.19	95.84	89.01	83.14

Table 5.4: CPU seconds comparison on the 6-bus system

Penetration level	10%	15%	20%
RP	0.9	2.5	1.1
RG	682.2	960.0	769.9
SP	99.4	178.6	184.0
SG	301.1	627.0	3485.7

Sensitivity analysis Table 5.5 and Figure 5.2 display the sensitivity analysis of RP and RG on different budget levels. In this experiment, the penetration level is 20%. We compare the out-of-sample performance of these two approaches on eight different budget levels and the two data sets. Each budget level contains five units, from 0–5 to 35–40. The result first shows that both worst-case and average costs are decreasing as the budget level increases and they all decrease dramatically in the lower budget levels and become more smooth beyond the budget level 10–15. Second, RP and RG have quite similar costs. Their worst-case costs are even identical on some budget levels, e.g., 5–10, 30–35, and 35–40. As compared with the results for the 10–40 budget level reported in Table 5.3, Table 5.5 also shows that the difference between these two approaches become smaller when the budget level is lower. This is because the full adaptivity in the second-stage formulation becomes less impactful with a smaller budget.



(a) Worst-case performance on real data

(b) Worst-case performance on misspecified data

(c) Average performance on real data

(d) Average performance on misspecified data

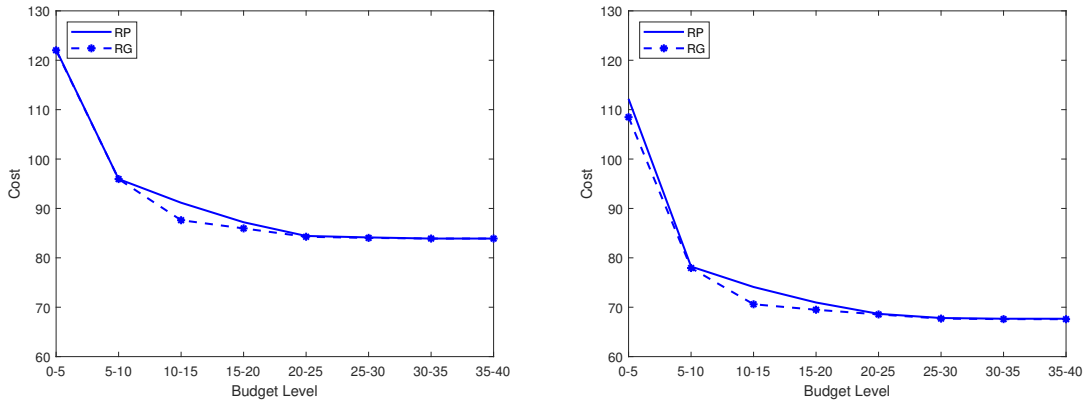
Figure 5.1: Cost comparison on the 6-bus system

5.4.3.2 46-bus system

Comparisons In the 46-bus system, we compare three approaches, i.e., RP, RG, and RM, because the other two approaches become too time-consuming to solve. We still use 140 historical data samples to construct the uncertainty set for RP and RG, and 100 data samples to approximately solve RG. We set the first-stage budget to be 100 and the second-stage budget to be Unif$[50, 150]$. We report the out-of-sample performance of these three approaches on the 56 real data samples in Table 5.6 and Figure 5.3. From these results, we make similar observations as in the 6-bus system. The approximated RG has the lowest objective values on all penetration levels and under both evaluation criteria, and RM performs the worst. The difference between RP and RG remains very small, which suggests the prioritization decision rule remains highly effective in a medium-sized system. RP has an excellent computational time (see Table 5.7). We omit the out-of-sample performance of these approaches on the misspecified data samples and the sensitivity analysis because

Table 5.5: Sensitivity analysis on the 6-bus system based on real data

	Budget level	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Worst-case	RP	122.17	95.95	91.15	87.21	84.43	84.13	83.90	83.90
performance	RG	122.02	95.95	87.62	85.96	84.27	84.04	83.90	83.90
Average	RP	112.18	78.23	74.10	70.96	68.66	67.82	67.67	67.67
performance	RG	108.48	77.94	70.60	69.49	68.54	67.70	67.59	67.58



(a) Worst-case performance on real data

(b) Average performance on real data

Figure 5.2: Sensitivity analysis on the 6-bus system based on real data

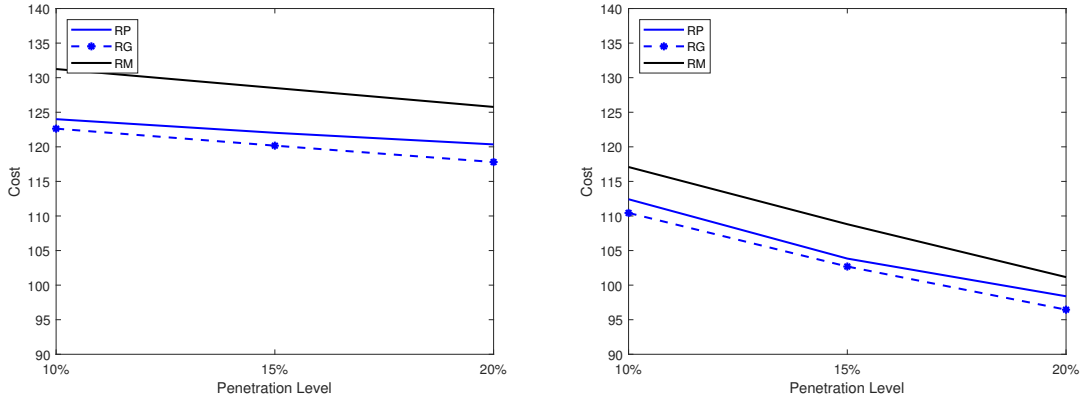
the results are similar as those in the 6-bus system.

Table 5.6: Cost comparison on the 46-bus system based on real data

	Worst-case performance			Average performance		
Penetration level	10%	15%	20%	10%	15%	20%
RP	124.00	122.03	120.35	112.42	103.84	98.39
RG	122.63	120.17	117.80	110.44	102.69	96.45
RM	131.25	128.51	125.77	117.08	108.81	101.16

5.5 Conclusion

In this chapter, we investigated a two-stage adaptive robust TEP model with prioritization. The core idea of prioritization is to rank the candidate circuits in the first-stage formulation, and after uncertainty is realized, to expand the transmission grid from the circuits with top prioritization in the second-stage formulation. It leads to more tractable formulations as well as more interpretable decisions. A tailored column-and-constraint generation algorithm was designed to solve the proposed model. We tested the robust TEP-P model on



(a) Worst-case performance on real data

(b) Average performance on real data

Figure 5.3: Cost comparison on the 46-bus system based on real data

Table 5.7: CPU seconds comparison on the 46-bus system

Penetration level	10%	15%	20%
RP	8.9	208.1	18.1
RG	5750.0	2595.5	7903.9

the Garver 6-bus system and the southern Brazilian 46-bus system. The numerical experiments demonstrated that the proposed approach performs near-optimally with significantly less computational time.

5.6 Nomenclature

A. Sets and Indices

$\mathcal{B}$ Set of buses.

$\mathcal{L}$ Set of lines.

$\mathcal{N}_{ij}^1, \mathcal{N}_{ij}^2$ Set of circuits that can be installed to line (i, j) now and in the future, respectively.

$\mathcal{N}^1, \mathcal{N}^2$ $\bigcup_{(i,j)\in\mathcal{L}} \mathcal{N}_{ij}^1$ and $\bigcup_{(i,j)\in\mathcal{L}} \mathcal{N}_{ij}^2$, respectively.

Ω Set of scenarios.

B. Parameters

A Node–branch incidence matrix of the transmission grid.

B^1 Current budget for circuit installation ($).

c_k Installation cost of circuit k ($).

c_b^r Unit load curtailment penalty cost of bus b ($/MW).

c_b^g Unit power generation cost at bus b ($/MW).

γ_{ij} Susceptance of each circuit on line (i, j).

$\bar{f}_{ij}$ Power flow capacity of each circuit on line (i, j) (MW).

$\bar{g}_b$ Thermal generation capacity of bus b (MW).

n_{ij}^0 Initial number of circuits on line (i, j).

$\underline{\psi}(\cdot), \overline{\psi}(\cdot)$ Budget upper and lower bounds as functions of the realized renewable energy and load, respectively.

C.Random Variables

$\tilde{h}_b$ Renewable energy at bus b (MW).

$\tilde{d}_b$ Load at bus b (MW).

$\tilde{B}^2$ Budget for circuit installation in the future ($).

D. Decision Variables

f_{ij} Power flow on line (i, j) (MW).

g_b Power generation amount from thermal generators at bus b (MW).

h_b Power generation amount from renewable generators at bus b (MW).

n_k^1 Binary variable such that $n_k^1 = 1$ if circuit $k \in \mathcal{N}^1$ is installed now and $n_k^1 = 0$ otherwise.

n_k^2 Binary variable such that $n_k^2 = 1$ if circuit $k \in \mathcal{N}^2$ is installed in the future and $n_k^2 = 0$ otherwise.

r_b Load curtailment amount at bus b (MW).

θ_b Phase angle at bus b.

$s_{kk'}$ Binary variable to denote the priority order between circuit k and k' such that $s_{kk'} = 1$ if circuit k is ranked at least as high as k' and $s_{kk'} = 0$ otherwise.

CHAPTER 6

Conclusions

In this dissertation, we study data-driven optimization methods for power system operations by modeling uncertainty directly based on the historical data. More specifically, we statistically infer key characteristics of the (ambiguous) probability distribution (e.g., support, mean, mean absolute deviation, unimodality, etc.) from the historical data and construct an ambiguity set consisting of all probability distributions that match the inferred characteristics. Then, we make distributionally robust decisions that hedge against the worst-case distributions within the ambiguity set. Efficient solution approaches are proposed for these data-driven distributionally robust optimization models. In addition, we apply the proposed models in power system operations, including optimal power flow, unit commitment, and transmission expansion planning.

In Chapter 2, We proposed a DRO approach for unit commitment and reserve procurement in power systems with renewable energy integration. Moment information obtained from historical data was considered in the ambiguity set and the resulting model was sovled by an algorithm based on the generalized linear decision rule. The numerical experiments on real-world wind energy data indicated that the proposed DRO approach can enhance the system flexibility and capability of accommodating renewable energy, and a real-time economic re-dispatch can help fully utilize the system flexibility. In Charter 3, we proposed an ambiguity set containing both moment and probability discrepancy information, i.e., a Wasserstein-moment ambiguity set, for the multi-period distributionally robust chance-constrained optimal power flow problem. A tractable convex conservative approximation was derived based on worst-case conditional value-at-risk. Compared with benchmark approaches, the proposed ambiguity set can better balance cost-effectiveness and reliability. In Chapter 4, we proposed a general framework of DRO models to incorporate shape information in a computationally tractable manner. A new class of ambiguity sets with shape information were proposed in work. In particular, we incorporated concentration inequalities can be derived from various shape information into the ambiguity set and showed that

the resulting DRO model can be conservatively approximated by a stochastic program with respect to an (unambiguous) probability distribution. In addition, we showed that, under mild conditions, this approximation is tight for a wide class of DRO models. Computational case studies on the appointment scheduling problem and the risk-constrained optimal power flow problem are conducted to demonstrate the theoretical results. In Chapter 5, we investigated a two-stage adaptive robust TEP model with prioritization. The prioritization decision rule led to more tractable formulations as well as more interpretable decisions. A tailored column-and-constraint generation algorithm was designed to solve the proposed model. As compared with the fully adaptive model, the numerical experiments demonstrated that the proposed approach can provide near-optimal solution with significantly less computational time.

Possible future directions include:

- Investigating other distributional information to strengthen the ambiguity set while leading to tractable and efficient reformulations. For example, one interesting open question here is whether DRO with ambiguity set containing both probability discrepancy and concentration inequalities admits tractable reformulation. Study of decision rules is another possible direction. Decision rules can significantly reduce the computational burden. An efficient decision rule with good performance will be of practical interest.

- Extending the proposed models to handle dynamic decision making under uncertainty. For example, a transmission network is often managed over a long time horizon, and hence expanding the network can naturally be modeled as a multistage optimization problem. Multistage optimization problem is considerably more difficult than two-stage problem, especially when ambiguous distribution is taken into account. A multistage DRO inventory model with moment-based ambiguity set was investigated in [154]. Its result heavily relies on the property of newsvendor problem. It will be very interesting to obtain more general results for the moment-based and other ambiguity sets.

- Studying data-driven optimization methods with applications beyond power system operations, e.g., healthcare operations, disaster risk management, and transportation. One important problem in healthcare operations is appointment scheduling. Many DRO appoaches have been studied for this problem (see e.g., [82, 95, 164]). DRO approach with shape information was applied on single server appointment scheduling in Chapter 4. We would also like to investigate multi-server appointment scheduling in the future. Data-driven optimization models have also been proposed to improve

the performance of transportation systems, such as vehicle balancing [99], vehicle routing [27] and transportation network design [133]. Applying the proposed models in different categories of applications could lead to interesting results.

BIBLIOGRAPHY

[1] NREL Eastern Wind Dataset. http://www.nrel.gov/electricity/transmission/eastern_wind_dataset.html. Accessed: 2015-10-30

[2] NREL National Solar Radiation Database. http://rredc.nrel.gov/solar/old_data/nsrdb/. Accessed: 2015-10-30

[3] Ahmed, S., Papageorgiou, D.J.: Probabilistic set covering with correlations. Operations Research **61**(2), 438–452 (2013)

[4] Akbari, T., Rahimikian, A., Kazemi, A.: A multi-stage stochastic transmission expansion planning method. Energy conversion and Management **52**(8-9), 2844–2853 (2011)

[5] Al-Hamouz, Z.M., Al-Faraj, A.S.: Transmission-expansion planning based on anonlinear programming algorithm. Applied energy **76**(1-3), 169–177 (2003)

[6] Al-Saba, T., El-Amin, I.: The application of artificial intelligent tools to the transmission expansion problem. Electric Power Systems Research **62**(2), 117–126 (2002)

[7] Alguacil, N., Motto, A.L., Conejo, A.J.: Transmission expansion planning: A mixed-integer lp approach. IEEE Transactions on Power Systems **18**(3), 1070–1077 (2003)

[8] Alizadeh, B., Jadid, S.: Reliability constrained coordination of generation and transmission expansion planning in power systems using mixed integer programming. IET generation, transmission & distribution **5**(9), 948–960 (2011)

[9] Baringo, L., Conejo, A.J.: Offering strategy via robust optimization. IEEE Transactions on Power Systems **26**(3), 1418–1425 (2011)

[10] Barth, R., Brand, H., Meibom, P., Weber, C.: A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power. International Conference on Probabilistic Methods Applied to Power Systems pp. 1–8 (2006)

[11] Bayraksan, G., Love, D.K.: Data-driven stochastic programming using phi-divergences. In: INFORMS TutORials in Operations Research, pp. 1–19. INFORMS (2015)

[12] Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust optimization, vol. 28. Princeton University Press (2009)

[13] Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A.: Adjustable robust solutions of uncertain linear programs. Mathematical Programming **99**(2), 351–376 (2004)

[14] Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., Rennen, G.: Robust solutions of optimization problems affected by uncertain probabilities. Management Science **59**(2), 341–357 (2013)

[15] Bertsekas, D.P.: Convex optimization theory. Athena Scientific Belmont (2009)

[16] Bertsimas, D., Caramanis, C.: Adaptability via sampling. In: 2007 46th IEEE Conference on Decision and Control, pp. 4717–4722. IEEE (2007)

[17] Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.P.: Models for minimax stochastic linear optimization problems with risk aversion. Mathematics of Operations Research **35**(3), 580–602 (2010)

[18] Bertsimas, D., Litvinov, E., Sun, X., Zhao, J., Zheng, T.: Adaptive robust optimization for the security constrained unit commitment problem. IEEE Transactions on Power Systems **28**(1), 52–63 (2013)

[19] Bertsimas, D., Sim, M.: The price of robustness. Operations Research **52**(1), 35–53 (2004)

[20] Bertsimas, D., Sim, M., Zhang, M.: A practicable framework for distributionally robust linear optimization. Technical report, Optimization Online (2013). http://www.optimization-online.org/DB_FILE/2013/07/3954.pdf

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

[22] Birge, J.R., Louveaux, F.: Introduction to stochastic programming. Springer Science & Business Media (2011)

[23] Bouffard, F., Galiana, F.: Stochastic security for operations planning with significant wind power generation. In: Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century. Pittsburgh, PA (2008)

[24] Brown, G., Carlyle, M., Salmerón, J., Wood, K.: Defending critical infrastructure. Interfaces **36**(6), 530–544 (2006)

[25] Calafiore, G., El Ghaoui, L.: On distributionally robust chance-constrained linear programs. Journal of Optimization Theory and Applications **130**(1), 1–22 (2006)

[26] Campi, M.C., Garatti, S., Prandini, M.: The scenario approach for systems and control design. Annual Reviews in Control **33**(2), 149–157 (2009)

[27] Carlsson, J.G., Behroozi, M., Mihic, K.: Wasserstein distance and the distributionally robust tsp. Operations Research **66**(6), 1603–1624 (2018)

[28] Carrión, M., Arroyo, J.M., Alguacil, N.: Vulnerability-constrained transmission expansion planning: A stochastic programming approach. IEEE Transactions on Power Systems **22**(4), 1436–1445 (2007)

[29] Chaniotis, D.: Security of supply, sustainability, competitiveness: a european approach. ENTSO-E (2012)

[30] Charnes, A., Cooper, W., Symonds, G.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. Management Science **4**(3), 235–263 (1958)

[31] Charnetski, J.R.: Scheduling operating room surgical procedures with early and late completion penalty costs. Journal of Operations Management **5**(1), 91–102 (1984)

[32] Chen, B., Wang, J., Wang, L., He, Y., Wang, Z.: Robust optimization for transmission expansion planning: Minimax cost vs. minimax regret. IEEE Transactions on Power Systems **29**(6), 3069–3077 (2014)

[33] Chen, X., Zhang, Y.: Uncertain linear programs: Extended affinely adjustable robust counterparts. Operations Research **57**(6), 1469–1482 (2009)

[34] Chen, Y., Guo, Q., Sun, H., Li, Z., Wu, W., Li, Z.: A distributionally robust optimization model for unit commitment based on kullback–leibler divergence. IEEE Transactions on Power Systems **33**(5), 5147–5160 (2018)

[35] Cheng, J., Delage, E., Lisser, A.: Distributionally robust stochastic knapsack problem. SIAM Journal on Optimization **24**(3), 1485–1506 (2014)

[36] Dantzig, G.B.: Linear programming under uncertainty. Management Science **1**(3-4), 197–206 (1955)

[37] Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations Research **58**(3), 595–612 (2010)

[38] Dharmadhikari, S., Joag-Dev, K.: Unimodality, Convexity, and Applications. Elsevier (1988)

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

[40] Draxl, C., Clifton, A., Hodge, B.M., McCaa, J.: The wind integration national dataset (wind) toolkit. Applied Energy **151**, 355–366 (2015)

[41] Draxl, C., Hodge, B., Clifton, A., McCaa, J.: Overview and meteorological validation of the wind integration national dataset toolkit. NREL/TP-5000-61740. Golden (CO): National Renewable Energy Laboratory (forthcoming), Tech. Rep. (2015)

[42] Duan, C., Fang, W., Jiang, L., Yao, L., Liu, J.: Distributionally robust chance-constrained approximate ac-opf with wasserstein metric. IEEE Transactions on Power Systems **33**(5), 4924–4936 (2018)

[43] Dusonchet, Y., El-Abiad, A.: Transmission planning using discrete dynamic optimizing. IEEE Transactions on Power apparatus and Systems (4), 1358–1371 (1973)

[44] El Ghaoui, L., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. Operations Research **51**(4), 543–556 (2003)

[45] Erdoğan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. Mathematical Programming **107**(1-2), 37–61 (2006)

[46] Erdogan, S.A., Denton, B.T.: Surgery planning and scheduling. Wiley encyclopedia of operations research and management science (2010)

[47] Esfahani, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming **171**(1-2), 115–166 (2018)

[48] Garcés, L.P., Conejo, A.J., García-Bertrand, R., Romero, R.: A bilevel approach to transmission expansion planning within a market environment. IEEE Transactions on Power Systems **24**(3), 1513–1522 (2009)

[49] Garver, L.L.: Transmission network estimation using linear programming. IEEE Transactions on Power Apparatus and Systems (7), 1688–1697 (1970)

[50] Ghaoui, L.E., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. Operations research **51**(4), 543–556 (2003)

[51] Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International statistical review **70**(3), 419–435 (2002)

[52] Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Operations research **58**(4-part-1), 902–917 (2010)

[53] Gourtani, A., Xu, H., Pozo, D., Nguyen, T.D.: Robust unit commitment with n-1 security criteria. Mathematical Methods of Operations Research **83**(3), 373–408 (2016)

[54] Grinold, R.C.: Infinite horizon stochastic programs. SIAM journal on control and optimization **24**(6), 1246–1260 (1986)

[55] Guan, Y., Wang, J.: Uncertainty sets for robust unit commitment. IEEE Transactions on Power Systems **3**(29), 1439–1440 (2014)

[56] Guédon, O., Nayar, P., Tkocz, T.: Concentration inequalities and geometry of convex bodies. Analytical and probabilistic methods in the geometry of convex bodies **2**, 9–86 (2014)

[57] Haffner, S., Monticelli, A., Garcia, A., Mantovani, J., Romero, R.: Branch and bound algorithm for transmission system expansion planning using a transportation model. IEE Proceedings-Generation, Transmission and Distribution **147**(3), 149–156 (2000)

[58] Hanasusanto, G.A.: Decision making under uncertainty: Robust and data-driven approaches. Ph.D. thesis, Imperial College London (2015)

[59] Hanasusanto, G.A., Kuhn, D.: Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. Operations Research **66**(3), 849–869 (2018)

[60] Hanasusanto, G.A., Kuhn, D., Wiesemann, W.: K-adaptability in two-stage robust binary programming. Operations Research **63**(4), 877–891 (2015)

[61] Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W.: Ambiguous joint chance constraints under mean and dispersion information. Operations Research **65**(3), 751–767 (2017)

[62] Heitsch, H., Römisch, W.: Scenario tree modeling for multistage stochastic programs. Mathematical Programming **118**(2), 371–406 (2009)

[63] Hodge, B.M., Florita, A., Orwig, K., Lew, D., Milligan, M.: A comparison of wind power and load forecasting error distributions. In: 2012 World Renewable Energy Forum, pp. 13–17 (2012)

[64] Hodge, B.M., Lew, D., Milligan, M., Holttinen, H., Sillanpää, S., Gómez-Lázaro, E., Scharff, R., Söder, L., Larsén, X.G.L., Giebel, G., Flynn, D., Dobschinski, J.: Wind power forecasting error distributions: An international comparison. In: 11th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants Conference (2012)

[65] Hodge, B.M., Milligan, M.: Wind power forecasting error distributions over multiple timescales. In: 2011 IEEE power and energy society general meeting, pp. 1–8. IEEE (2011)

[66] Hooshmand, R.A., Hemmati, R., Parastegari, M.: Combination of ac transmission expansion planning and reactive power planning in the restructured power system. Energy Conversion and Management **55**, 26–35 (2012)

[67] Hu, Z., Hong, L.J.: Kullback-leibler divergence constrained distributionally robust optimization. Available at Optimization Online (2013)

[68] International Energy Agency: 2014 Key World Energy Statistics. Tech. rep. (2014). URL http://www.iea.org/publications/freepublications/publication/KeyWorld2014.pdf

[69] Jabr, R.: Robust transmission network expansion planning with uncertain renewable generation and loads. IEEE Transactions on Power Systems **28**(4), 4558–4567 (2013)

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

[71] Jabr, R.A.: Optimization of ac transmission system planning. IEEE Transactions on Power Systems **28**(3), 2779–2787 (2013)

[72] Jiang, R., Conejo, A.J., Wang, J.: Day-ahead scheduling: Reserve determination and valuation. In: H. Chen (ed.) Power Grid Operations in a Market Environment: Economic Efficiency and Risk Mitigation. Wiley (2015)

[73] Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Mathematical Programming **158**(1-2), 291–327 (2016)

[74] Jiang, R., Guan, Y., Watson, J.P.: Risk-averse stochastic unit commitment with incomplete information. IIE Transactions **48**(9), 838–854 (2016)

[75] Jiang, R., Wang, J., Guan, Y.: Robust unit commitment with wind power and pumped storage hydro. IEEE Transactions on Power Systems **27**(2), 800–810 (2011)

[76] Kall, P.: Computational methods for solving two-stage stochastic linear programming problems. Zeitschrift für angewandte Mathematik und Physik ZAMP **30**(2), 261–271 (1979)

[77] Kall, P., Mayer, J.: Slp-ior: An interactive model management system for stochastic linear programs. Mathematical Programming **75**(2), 221–240 (1996)

[78] Kantorovich, L.V., Rubinshtein, G.: On a space of totally additive functions. Vestnik Leningradskogo Universiteta **13**, 52–59 (1958)

[79] King, J., Clifton, A., Hodge, B.M.: Validation of power output for the WIND Toolkit. National Renewable Energy Laboratory (2014)

[80] Koc, A., Morton, D.P., Popova, E., Hess, S.M., Kee, E., Richards, D.: Prioritizing project selection. The Engineering Economist **54**(4), 267–297 (2009)

[81] Koç, A., Morton, D.P.: Prioritization via stochastic optimization. Management Science **61**(3), 586–603 (2014)

[82] Kong, Q., Lee, C.Y., Teo, C.P., Zheng, Z.: Scheduling arrivals to a stochastic service delivery system using copositive cones. Operations Research **61**(3), 711–726 (2013)

[83] Kuhn, D., Wiesemann, W., Georghiou, A.: Primal and dual linear decision rules in stochastic and robust optimization. Mathematical Programming **130**(1), 177–209 (2011)

[84] Latorre, G., Cruz, R.D., Areiza, J.M., Villegas, A.: Classification of publications and models on transmission expansion planning. IEEE Transactions on Power Systems **18**(2), 938–946 (2003)

[85] Lee, C., Ng, S.K., Zhong, J., Wu, F.F.: Transmission expansion planning from past to future. In: 2006 IEEE PES Power Systems Conference and Exposition, pp. 257–265. IEEE (2006)

[86] Li, B., Jiang, R., Mathieu, J.L.: Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In: Decision and Control (CDC), 2016 IEEE 55th Conference on, pp. 2425–2430. IEEE (2016)

[87] Li, B., Jiang, R., Mathieu, J.L.: Ambiguous risk constraints with moment and unimodality information. Mathematical Programming **173**(1-2), 151–192 (2019)

[88] Li, B., Mathieu, J.L.: Analytical reformulation of chance-constrained optimal power flow with uncertain load control. In: PowerTech, 2015 IEEE Eindhoven, pp. 1–6. IEEE (2015)

[89] Li, B., Mathieu, J.L., Jiang, R.: Distributionally robust chance constrained optimal power flow assuming log-concave distributions. In: 2018 Power Systems Computation Conference (PSCC), pp. 1–7. IEEE (2018)

[90] Lieberman-Cribbin, W., Draxl, C., Clifton, A.: A guide to using the WIND Toolkit validation code. National Renewable Energy Laboratory (2014)

[91] Lindenberg, S.: 20% Wind Energy By 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply. Tech. Rep. DOE/GO-102008-2567, U.S. Department of Energy (2008)

[92] Lopez, J.A., Ponnambalam, K., Quintana, V.H.: Generation and transmission expansion under risk using stochastic programming. IEEE Transactions on Power Systems **22**(3), 1369–1378 (2007)

[93] Lorca, A., Sun, X.A., Litvinov, E., Zheng, T.: Multistage adaptive robust optimization for the unit commitment problem. Operations Research **64**(1), 32–51 (2016)

[94] Lumbreras, S., Ramos, A.: The new challenges to transmission expansion planning. survey of recent practice and literature review. Electric Power Systems Research **134**, 19–29 (2016)

[95] Mak, H.Y., Rong, Y., Zhang, J.: Appointment scheduling with limited distributional information. Management Science **61**(2), 316–334 (2014)

[96] Mangasarian, O., Rosen, J.: Inequalities for stochastic nonlinear programming problems. Operations Research **12**(1), 143–154 (1964)

[97] Margellos, K., Goulart, P., Lygeros, J.: On the road between robust optimization and the scenario approach for chance constrained optimization problems. IEEE Transactions on Automatic Control **59**(8), 2258–2263 (2014)

[98] McCormick, G.: Computability of global solutions to factorable nonconvex programs: Part I–Convex underestimating problems. Mathematical Programming **10**(1), 147–175 (1976)

[99] Miao, F., Han, S., Hendawi, A.M., Khalefa, M.E., Stankovic, J.A., Pappas, G.J.: Data-driven distributionally robust vehicle balancing using dynamic region partitions. In: Proceedings of the 8th International Conference on Cyber-Physical Systems, pp. 261–271. ACM (2017)

[100] Miller, B.L., Wagner, H.M.: Chance constrained programming with joint constraints. Operations Research **13**(6), 930–945 (1965)

[101] Moreira, A., Street, A., Arroyo, J.M.: An adjustable robust optimization approach for contingency-constrained transmission expansion planning. IEEE Transactions on Power Systems **30**(4), 2013–2022 (2015)

[102] Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM Journal on Optimization **17**(4), 969–996 (2006)

[103] Padhy, N.P.: Unit commitment-a bibliographical survey. IEEE Transactions on power systems **19**(2), 1196–1205 (2004)

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

[105] Pflug, G., Wozabal, D.: Ambiguity in portfolio selection. Quantitative Finance **7**(4), 435–442 (2007)

[106] Philpott, A.B., De Matos, V.L.: Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. European Journal of Operational Research **218**(2), 470–483 (2012)

[107] Popescu, I.: A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. Mathematics of Operations Research **30**(3), 632–657 (2005)

[108] Popescu, I.: Robust mean-covariance solutions for stochastic optimization. Operations Research **55**(1), 98–112 (2007)

[109] Prékopa, A.: On probabilistic constrained programming. In: Proceedings of the Princeton Symposium on Mathematical Programming, pp. 113–138. Citeseer (1970)

[110] Renewable Energy Policy Network for the 21st Century: Renewables 2014: Global Status Report. Tech. rep. (2014). URL http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/GSR2014_full%20report_low%20res.pdf

[111] Rezende, L.S., da Silva, A.M.L., de Mello Honório, L.: Artificial immune system applied to the multi-stage transmission expansion planning. In: International Conference on Artificial Immune Systems, pp. 178–191. Springer (2009)

[112] Rider, M., Garcia, A., Romero, R.: Power system transmission network expansion planning using ac model. IET Generation, Transmission & Distribution **1**(5), 731–742 (2007)

[113] Roald, L., Oldewurtel, F., Krause, T., Andersson, G.: Analytical reformulation of security constrained optimal power flow with probabilistic constraints. IEEE (2013)

[114] Rockafellar, R., Uryasev, S.: Optimization of conditional value-at-risk. Journal of Risk **2**, 21–42 (2000)

[115] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. Journal of Banking & Finance **26**(7), 1443–1471 (2002)

[116] Rogers, J., Fink, S., Porter, K.: Examples of wind energy curtailment practices. NREL, Subcontract Report NREL/SR-550 **48737** (2010)

[117] Roh, J.H., Shahidehpour, M., Wu, L.: Market-based generation and transmission planning with uncertainties. IEEE Transactions on Power Systems **24**(3), 1587–1598 (2009)

[118] Romero, R., Monticelli, A.: A hierarchical decomposition approach for transmission network expansion planning. IEEE transactions on power systems **9**(1), 373–380 (1994)

[119] Romero, R., Monticelli, A., Garcia, A., Haffner, S.: Test systems and mathematical models for transmission network expansion planning. IEE Proceedings-Generation, Transmission and Distribution **149**(1), 27–36 (2002)

[120] Rudelson, M., Vershynin, R., et al.: Hanson-wright inequality and sub-gaussian concentration. Electronic Communications in Probability **18** (2013)

[121] Ruiz, P.A., Philbrick, C.R., Zak, E., Cheung, K.W., Sauer, P.W.: Uncertainty management in the unit commitment problem. IEEE Transactions on Power Systems **24**(2), 642–651 (2009)

[122] Samarakoon, H., Shrestha, R., Fujiwara, O.: A mixed integer linear programming model for transmission expansion planning with generation location selection. International journal of electrical power & energy systems **23**(4), 285–293 (2001)

[123] Savage, S., Scholtes, S., Zweidler, D.: Probability management. OR MS TODAY **33**(1), 20 (2006)

[124] Scarf, H.: A min-max solution of an inventory problem. In: K. Arrow, S. Karlin, H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production, pp. 201–209. Stanford University Press (1958)

[125] Shahidehpour, M., Yamin, H., Li, Z.: Market Operations in Electric Power Systems. Wiley (2002)

[126] Shapiro, A.: On duality theory of conic linear problems. In: Semi-infinite programming, pp. 135–165. Springer (2001)

[127] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, vol. 9. SIAM (2009)

[128] da Silva, A.M.L., Rezende, L.S., da Fonseca Manso, L.A., de Resende, L.C.: Reliability worth applied to transmission expansion planning based on ant colony system. International Journal of Electrical Power & Energy Systems **32**(10), 1077–1084 (2010)

[129] Smith, J.E., Winkler, R.L.: The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Management Science **52**(3), 311–322 (2006)

[130] Soriano, A.: Comparison of two scheduling systems. Operations Research **14**(3), 388–397 (1966)

[131] Stellato, B.: Data-driven chance constrained optimization. Master's thesis, ETH Z´urich (2014)

[132] Summers, T., Warrington, J., Morari, M., Lygeros, J.: Stochastic optimal power flow based on conditional value at risk and distributional robustness. International Journal of Electrical Power & Energy Systems **72**, 116–125 (2015)

[133] Sun, H., Gao, Z., Szeto, W., Long, J., Zhao, F.: A distributionally robust joint chance constrained optimization model for the dynamic network design problem under demand uncertainty. Networks and Spatial Economics **14**(3-4), 409–433 (2014)

[134] Takriti, S., Birge, J., Long, E.: A stochastic model for the unit commitment problem. IEEE Transactions on Power Systems **11**(3), 1497–1508 (1996)

[135] Tuohy, A., Meibom, P., Denny, E., O'Malley, M.: Unit commitment for systems with significant wind penetration. IEEE Transactions on Power Systems **24**(2), 592–601 (2009)

[136] U.S. Department of Energy, Energy Efficiency & Renewable Energy Laboratory: 2013 Renewable Energy Data Book (2013)

[137] Van Parys, B.P., Goulart, P.J., Kuhn, D.: Generalized Gauss inequalities via semidefinite programming. Mathematical Programming **156**(1-2), 271–302 (2016)

[138] Van Parys, B.P., Goulart, P.J., Morari, M.: Distributionally robust expectation inequalities for structured distributions. Mathematical Programming **173**(1-2), 251–280 (2019)

[139] Vandenberghe, L., Boyd, S., Comanor, K.: Generalized Chebyshev bounds via semidefinite programming. SIAM Review **49**(1), 52–64 (2007)

[140] Villasana, R., Garver, L., Salon, S.: Transmission network planning using linear programming. IEEE transactions on power apparatus and systems (2), 349–356 (1985)

[141] Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G.: A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration. IEEE Transactions on Power Systems **28**(4), 3885–3896 (2013)

[142] Vrakopoulou, M., Mathieu, J.L., Andersson, G.: Stochastic optimal power flow with uncertain reserves from demand response. In: System Sciences (HICSS), 2014 47th Hawaii International Conference on, pp. 2353–2362. IEEE (2014)

[143] Wagner, M.: Stochastic 0–1 linear programming under limited distributional information. Operations Research Letters **36**(2), 150–156 (2008)

[144] Wang, J., Shahidehpour, M., Li, Z.: Contingency-constrained reserve requirements in joint energy and ancillary services auction. IEEE Transactions on Power Systems **24**(3), 1457–1468 (2009)

[145] Wang, P.P.: Static and dynamic scheduling of customer arrivals to a single-server system. Naval Research Logistics (NRL) **40**(3), 345–360 (1993)

[146] Wang, W., Ahmed, S.: Sample average approximation of expected value constrained stochastic programs. Operations Research Letters **36**(5), 515–519 (2008)

[147] Warrington, J., Goulart, P.J., Mariéthoz, S., Morari, M.: Robust reserve operation in power systems using affine policies. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pp. 1111–1117. IEEE (2012)

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

[149] Warrington, J., Hohl, C., Goulart, P.J., Morari, M.: Optimal unit commitment accounting for robust affine reserve policies. In: 2014 American Control Conference, pp. 5049–5055. IEEE (2014)

[150] Whittle, B., Yu, J., Teng, S., Mickey, J.: Reliability unit commitment in ERCOT. In: Proceedings of the IEEE Power Engineering Society General Meeting. Montreal, Qubec (2006)

[151] Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Operations Research **62**(6), 1358–1376 (2014)

[152] Wozabal, D.: A framework for optimization under ambiguity. Annals of Operations Research **193**(1), 21–47 (2012)

[153] Xie, W.: On distributionally robust chance constrained programs with wasserstein distance. arXiv preprint arXiv:1806.07418 (2018)

[154] Xin, L., Goldberg, D.A., Shapiro, A.: Distributionally robust multistage inventory models with moment constraints. Tech. rep., Georgia Institute of Technology, Atlanta, GA (2013)

[155] Xiong, P., Jirutitijaroen, P., Singh, C.: A distributionally robust optimization model for unit commitment considering uncertain wind power generation. IEEE Transactions on Power Systems **32**(1), 39–49 (2016)

[156] Xu, Y., Ralphs, T.K., Ladányi, L., Saltzman, M.J.: Computational experience with a software framework for parallel integer programming. INFORMS Journal on Computing **21**(3), 383–397 (2009)

[157] Yang, N., Wen, F.: A chance constrained programming approach to transmission system expansion planning. Electric Power Systems Research **75**(2-3), 171–177 (2005)

[158] Ye, H., Wang, J., Li, Z.: MIP reformulation for max-min problems in two-stage robust SCUC. IEEE Transactions on Power Systems **32**(2), 1237–1247 (2017)

[159] Youssef, H., Hackam, R.: New transmission planning model. IEEE Transactions on Power Systems **4**(1), 9–18 (1989)

[160] Zeng, B., Zhao, L.: Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters **41**(5), 457–461 (2013)

[161] Zhang, H., Li, P.: Probabilistic analysis for optimal power flow under uncertainty. IET Generation, Transmission & Distribution **4**(5), 553–561 (2010)

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

[163] Zhang, J., Hodge, B.M., Florita, A., Lu, S., Hamann, H.F., Banunarayanan, V.: Metrics for evaluating the accuracy of solar power forecasting. In: 3rd International workshop on integration of solar power into power systems, London, England (2013)

[164] Zhang, Y., Shen, S., Mathieu, J.L.: Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. IEEE Transactions on Power Systems **32**(2), 1378–1388 (2017)

[165] Zhao, C., Guan, Y.: Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics. Available on Optimization Online (2015)

[166] Zhao, C., Jiang, R.: Distributionally robust contingency-constrained unit commitment. IEEE Transactions on Power Systems **33**(1), 94–102 (2017)

[167] Zhao, C., Wang, J., Watson, J.P., Guan, Y.: Multi-stage robust unit commitment considering wind and demand response uncertainties. IEEE Transactions on Power Systems **28**(3), 2708–2717 (2013)

[168] Zhao, J.H., Dong, Z.Y., Lindsay, P., Wong, K.P.: Flexible transmission expansion planning with uncertainties in an electricity market. IEEE Transactions on Power Systems **24**(1), 479–488 (2009)

[169] Zhao, L., Zeng, B.: Robust unit commitment problem with demand response and wind energy. In: Power and Energy Society General Meeting, 2012 IEEE, pp. 1–8. IEEE (2012)

[170] Zheng, Q.P., Wang, J., Liu, A.L.: Stochastic optimization for unit commitment - A review. IEEE Transactions on Power Systems **30**(4), 1913–1924 (2015)

[171] Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. Mathematical Programming **137**(1-2), 167–198 (2013)

[172] Zymler, S., Kuhn, D., Rustem, B.: Worst-case value at risk of nonlinear portfolios. Management Science **59**(1), 172–188 (2013)