UNIVERSITÀ DEGLI STUDI DI PAVIA
DIPARTIMENTO DI MATEMATICA "FELICE CASORATI"

CORSO DI LAUREA MAGISTRALE IN MATEMATICA

On the exactness of Jabr-like models and distributionally robust stochastic optimal power flow.

RELATORE: STEFANO GUALANDI
CORRELATORE: AMBROGIO MARIA BERNARDELLI

LAUREANDO:
GABOR RICCARDI
Matricola: 512240

Anno Accademico
2023/2024

"We demand rigidly defined areas of doubt and uncertainty!"

- The Hitchhiker's Guide to the Galaxy
Douglas Adams

Abstract

This thesis delves into the relationship between the inherent graph structure of power networks and the OPF problem, with a focus on Jabr-like models. It examines the exactness of a Jabr equality relaxation in tree network and introduces "loop constraints" to achieve exactness in more general graph structures, offering insights into their solution spaces. Additionally, the thesis explores the realm of stochastic OPF. The objective is to determine power scheduling policies for the network, minimizing the mean cost over the worst case of power output prediction errors. The contribution of this thesis lie in the implementation of nonlinear constraints within the stochastic model. It also introduces a new approach for modeling affine inequality constraints as policy constraints based on the continuity norm. The model is then applied to a IEEE 118-bus test, with data sourced by CESI.

Contents

Introduction

In 1962, Carpentier introduced the Optimal Power Flow problem (OPF) as an extension of the problem of optimal economic dispatch of generation in electric power systems [Car79]. Carpentier's key contribution was the inclusion of the electric power flow equations in the set of equality constraints. Today, the defining feature of OPF remains the presence of power flow equations in the set of equality constraints. A power flow study determines the voltages, currents, and real and reactive power flows in an electric network under given load conditions. The goal of an Optimal Power Flow (OPF) problem is to minimize generation costs, losses, emissions, and constraint violations. However, the OPF problem, in its most realistic form, is complex: it is large-scale, non-smooth, non-convex, and nonlinear. It generally has multiple local minima and a corresponding feasibility problem that is strongly NP-hard [BV19]. The transportation of power from generating plants (generators) to users (loads) occurs by means of a network, called power grid, whose nodes and edges are called buses and lines, respectively. A bus may host at the same time both generators and loads. Electrical network may transport Direct Current (DC) or Alternating Current (AC). In direct current (DC), the electric charge (current) only flows in one direction. Electric charge in alternating current (AC), on the other hand, changes direction periodically. DC is generally used in small electronic devices while the transportation of power on any geographical area is typically based on AC. The complexity of the model makes it challenging to understand its relationship with the characteristics of its inherent graph structure of the network. In addition, with emerging technologies like renewable generation, batteries, and electric vehicles to tackle climate change, the problem further grows in complexity since OPF problems are needed to adapt to new stochastic variables in energy production and consumption. Future power networks will require more sophisticated methods for managing risks at both transmission and distribution levels.

Optimization of power system operation occurs via incremental planning: long-term planning procedures make high-level decisions on simplified models, while short-term procedures refine earlier decisions using detailed models but more limited decision spaces. This new class of problems is called stochastic OPF. Various approaches have been taken to tackle this problem,

many formulations assume that uncertain forecast errors follow a prescribed probability distribution [BCH14] however such assumptions can often fail in weather prediction. Another approach is to formulate a stochastic OPF as a minimization problem over a Wasserstein ball in the infinite-dimensional space of multivariate probability distributions [Guo+19b] which uses duality results in optimal transport developed in [EK18]. This thesis aims (1) to develop a deeper understanding of the relationship between the network's graph structure and the OPF problem focusing on the study of a novel family of Jabr-like models, and (2) to develop and implement an accurate model of stochastic distributionally robust OPF, utilizing data sourced from CESI, an Italian global consulting and testing company specializing in the energy sector.

The thesis is divided as follows. The first part of the thesis, presented in introductory sections, focuses on the fundamental theory required for the following chapters of the thesis:

1. **An Introduction to Graph Theory**. In this section, we review some basic definitions of Graph Theory and results on Cycle Basis which will be used Chapter 4. The main reference for this section is [Kav+09].

2. **Optimal transport**. The distributionally robust problem addressed in this thesis is ultimately a minimization problem over a Wasserstein ball in the infinite-dimensional space of multivariate probability distributions. In this section, we introduce the definition of the Wasserstein distance. The main sources for this part are [San15] and [YV14].

3. **Convex Analysis**. This section reviews foundational concepts of Convex Analysis, like Duality and Minimax theorems. The main source for this chapter is [Ber03].

4. **Data-driven Distributionally Robust Optimization** . We leverage the results of the previous chapters to elucidate the primary finite reduction result of distributionally robust optimization problems, as presented in [EK18]. These results will be then used in Chapter 5 on a distributionally robust stochastic optimal power flow problem.

5. **An introduction to OPF**. In this chapter, we introduce the OPF problem and its various formulations. The main references for this chapter are [FR16] and [Bie+20].

All new contributions of this thesis are contained within the second part, which comprises two chapters. Chapter 4 is the analysis of a class of Jabr-like models [Jab06]. Firstly, we prove that a non-convex Jabr equality relaxation is exact over a multisource radial network. We then introduce a new class of constraints, referred to as *loop constraints* and prove that the Jabr equality relaxation with the additional loop constraints is exact and that it is sufficient

to enforce this new constraint over a cycle basis of the network. We conclude this chapter with an exploration of the solution space of the loop constraints, where we discern a parametrization.

In the context of emerging technologies like renewable generation, batteries, and electric vehicles, there is a need for new techniques to solve OPF problems that account for stochastic variables in energy production and consumption. In Chapter 5 we formulate a general multistage stochastic OPF problem, based on [Guo+19b]. Our objective is to determine the policy that minimizes the worst-case expectation of the cost of implementing the policy while accounting for the cost of constraint violations over a fixed time horizon. The worst-case expectation is calculated over an ambiguity set $\hat{\mathcal{P}}$, which contains all distributions of the prediction errors that could have generated the training samples with high confidence. We apply this model on a relaxation of the Jabr model reviewed in Chapter 4. Affine inequality constraints can then be enforced directly on the policy, while the risk of violation of other kinds of constraints is modeled using CVar. The stochastic OPF problem is then reduced to a finite form using results reviewed in Chapter 2. To conclude, we apply the implementation of the reduced problem to a variant of the IEEE 118-bus test [Guo+19a], [Zim+11] where the wind power forecast errors are obtained from a data source provided by CESI.

Chapter 1

Backgrounds

1.1 Graph Theory

Part of this thesis will focus on the relationship between the feasible space of
the Optimal Power Flow problem and the graph associated with the electrical
network. Specifically, this graph uses the buses of the network as its vertices
and the transmission lines as its edges. In Chapter 4, we will assume that
the graph of the electrical network is a simple graph, implying that for every
pair of buses, there exists at most one transmission line connecting them.
This assumption typically holds true for transmission networks. We will also
assume that the number of nodes is finite.

Definition 1.1.1. *A simple graph G is an ordered pair $G = (V, E)$ where:*

- *V is a set of* vertices *or* nodes,

- *$E \subset \mathcal{P}_2(V)$ is a set of* edges *or* links

where $\mathcal{P}_2(V) := \{A \in \mathcal{P}(V) \mid |A| = 2\}$ and $\mathcal{P}(V)$ is the partition set of V.

We also note that in the modeling of the Optimal Power Flow, the electri-
cal network graph considered is the directed graph where each edge $(a, b) \in E$
denotes with a the bus to which the transformer on the line (if present) is
closest. A directed graph is a graph where the pairs of nodes defining the
edges are ordered.

Definition 1.1.2. *A simple directed graph D is an ordered pair $D = (V, A)$
where:*

- *V is a set of* vertices *or* nodes,

- *$A \subset V \times V$ is a set of* arcs.

For each arc $(i, j) \in V$ we will refer to the node i as predecessor *of node j
and to node j as the successor of node i.*

Definition 1.1.3 (subgraph)**.** *A subgraph $G' = (V', E')$ of $G = (V, E)$ is a graph such that $V' \subset V$ and $E' \subset E$. We denote with $G' \subset G$ when G' is a subgraph of G.*

We also give the following list of definitions which will be used. Let $G = (V, E)$ be a graph.

1. **Adjacency.** Two arcs that share a common node are called adjacent. Similarly, two nodes $i, j \in V$ are adjacent if $\{i, j\} \in E$.

2. **Walk.** A sequence of nodes $l = (v_0, v_1, \ldots, v_r)$ such that $\{v_{i-1}, v_i\} \in E$ for all $i = 1, ..., r$ is called a *walk* of *length r*. The walk is *closed* if and only if $v_0 = v_r$ and is open *otherwise*.

3. **Connected graph.** G is connected if and only if for every $i, j \in V, i \neq j$ there is a walk from i to j.

4. **Trail and Path.** A *trail* in G is a walk with no edge repeated. A *path* is a walk with no vertex repeated.

5. **Cycle.** A *cycle* is a closed walk with no vertex repeated (except for the first and last vertex). A cycle of length n is denoted by C_n. Cycles are uniquely identified by the set of edges composing them, so sometimes we will refer to a subset of edges $A \subset E$ as a cycle if there exists a cycle composed of the edges in A.

6. **Tree.** A *tree* is a graph which does not have any cycles.

7. **Spanning tree.** $(V', E') = T \subset G = (V, E)$ is a spanning tree of G if T is a tree and $V = V'$.

If an electrical network is not connected, we can consider its individual connected components independently. In the case of a connected graph, we have:

Observation 1.1.1. *Let $G = (V, E)$ be a graph with $|V| = n$ and $T = (V, E') \subset G$. The following statements are equivalent.*

1. *T is a spanning tree.*

2. *T does not have any cycles and $|E'| = n - 1$.*

3. *T is connected and $|E'| = n - 1$.*

4. *For each pair of nodes $i, j \in V$ there exists a unique path in T connecting i and j.*

Furthermore we have that:

Observation 1.1.2. *Given a spanning tree $T = (V, E') \subset G$ and two nodes $i, j \in V$ such that $\{i, j\} \notin E'$, the graph $G' = (V, E' \cup \{\{i, j\}\} \subset G$ has exactly one cycle.*

1.1.1 Cycle basis

When considering the sum of voltage angle differences along a cycle in a electrical network it will be natural to consider various vector space structures on the set of cycles of a graph. We begin with an example where sum of two cycles is defined as the symmetric difference (or exclusive disjunction) of the edges of the two cycles. For instance, consider two cycles C_1 and C_2 in a graph. The sum $C_1 + C_2$ will consist of edges that are present in either C_1 or C_2 but not both. This is a $\mathbb{Z}_2$-vector space and the defined operation can be graphically represented as follows:
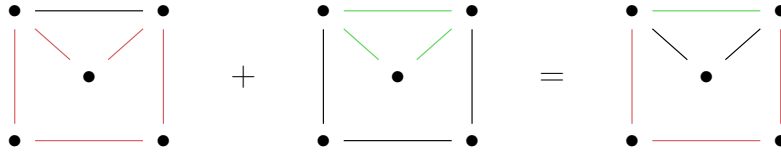
A natural question that arises is the following: how do we describe a basis for these vector spaces? Generalizing the previous example we obtain the following definition.

Definition 1.1.4 (Vector associated to a set of edges). *Given a directed graph $D = (V, A)$ and a field K, we can consider the K-vector space $\mathcal{V} := K^A$. For every set of arcs A, we can define the associated vector $v_A \in \mathcal{V}$ such that $(v_A)_{(i,j)} = 1$ if $(i,j) \in A$, $(v_A)_{(i,j)} = -1$ if $(j,i) \in A$, and 0 otherwise.*

Here, for $v \in \mathcal{V}$ and $e \in A$, v_e denotes the component of v indexed by e. Instead of v_e, we will also sometimes write $v(e)$ when $v = v_i$ belongs to an indexed family.

For a directed graph $D = (V, E)$, for each node $v \in V$, we define the sets $\delta^+(v) = \{(v, u) \mid (v, u) \in A\}$ and $\delta^-(v) = \{(u, v) \mid (u, v) \in A\}$. By construction, for every cycle c, we have

$$\sum_{e \in \delta^+(v)} v_c(e) = \sum_{e \in \delta^-(v)} v_c(e) \tag{1.1}$$

Such relation is referred to as *flow conservation*. Due to sum linearity, flow conservation is preserved when considering linear combinations of the vectors v_c where c represents cycles. Moreover, if A is a set of edges and v_A satisfies the flow conservation, then A represents a circuit. Hence, c_A is generated by cycles since a circuit can be expressed as a union of cycles. Where by *generated by cycles* we mean that $v_{C_A} \in \mathrm{Span}(v_c \mid c$ is a cycle). To summarize:

Observation 1.1.3. *Let A be a set of arcs, then $v_A \in \mathcal{V}$ is generated by cycles if and only if it satisfies the flow conservation.*

A subspace of $\mathcal{V}$ of particular interest is the *cycle space*.

Definition 1.1.5 (Cycle space)**.** *The* cycle space $\mathcal{C}_K(D)$ *of a graph* $D = (V, A)$ *is the subspace of vectors satisfying the flow conservation. The elements of the cycle space are called K-cycles. A* Cycle Basis $\mathcal{B}$ *is a set of cycles such that the set* $\{v_c \mid c \in \mathcal{B}\}$ *is a basis for the cycle space.*

Given an undirected graph G, we can always convert it into a directed graph D by arbitrarily choosing an orientation for each edge in the graph. We refer to D as an *orientation* of G. If D and D' are orientations of the same undirected graph G, their respective cycle spaces $\mathcal{C}_K(D)$ and $\mathcal{C}_K(D')$ are isomorphic. Specifically, if $v \in K^A$ represents a cycle in D, the corresponding cycle in D' can be identified by reversing the signs of the components $v(e)$ for which the edge e is oriented differently in D and D'. We deduce that the vector space is not influenced by the orientation D; instead, it is uniquely defined by the underlying undirected graph G. Thus, we can also denote this space as $\mathcal{C}_K(G)$. We will now summarize a few interesting proprieties of the cycle space.

As a side note, it might be tempting (as done by the author of this thesis) to define the vectors v_c for a cycle c in an undirected graph G as vectors that possess a value of 1 for the components $v_c(e)$ where e is an edge in c and 0 elsewhere. One might then try to describe the cycle space as the span of these v_c vectors. However, the resulting vector space is not always isomorphic to the cycle space $\mathcal{C}_K(D)$ (notably, they might differ in dimension). This discrepancy arises because the flow conservation property is not maintained with this naive definition.

Observation 1.1.4. *Let D be a directed connected graph and let T be any spanning tree of D. If v is a cycle that uses only edges in T, i.e., $v_e = 0$ for every $e \notin T$, then $v = 0$. Hence if v and v' are cycles with $v_e = v'_e$ for all $e \notin T$, then $v = v'$.*

Proof. Let $\pi(v)$ be the projection of $v \in K^A$ to K^T. $\pi(v)$ satisfies the flow conservation, therefore $\pi(v)$ is a cycle. But the cycles space for a tree is $\{0\}$ since it is generated by v_c where c is a cycle and there are no cycles in a tree. Therefore $\pi(v) = 0$ and hence $v = 0$. The second part of the observation follows from linearity. $\qquad\square$

Observation 1.1.5. *Let $\mathcal{B}$ be a set of K-cycles in G and let T be any spanning forest of D. For any cycle $c \in \mathcal{B}$, let $\pi(c)$ be its projection to K^N where $N := E \setminus T$. The cycles in $\mathcal{B}$ are linearly independent if and only if their projections to N are linearly independent.*

Proof. Since the projection is a linear operation, the linear dependence of cycles in $\mathcal{B}$ implies the linear dependence of their projections. Conversely,

suppose that $0 = \sum_{c \in \mathcal{B}} \lambda_c \pi(c) = \pi(\sum_{c \in \mathcal{B}} \lambda_c c)$ is a non trivial linear combination. This means that the linear combination $\sum_{c \in \mathcal{B}} \lambda_c c$ uses no edges outside T. But since trees have no cycles, this linear combination equals the zero vector, implying the cycles in $\mathcal{B}$ are linearly dependent. $\qquad\square$

We have the following method for determining a cycle basis of a graph:

Theorem 1.1.6. *Given a connected simple graph $G = (V, E)$, a spanning tree $T = (V, E') \subset G$ and a field K, the set $\mathcal{B} := \{v_{c_e} \mid e \notin E'\}$, where c_e corresponds to the only cycle contained in $(V, E_T \cup \{e\})$, is a cycle basis for G. In particular the dimension of the cycle space is $r := |E| - |V| + 1$, which is also referred to as the* cyclomatic number *or* cycle rank *of G.*

Proof. The projections of $\mathcal{B}$ to K^N, where $N := E \setminus E'$ is the canonical basis and thus by Observation 1.1.5, the elements of $\mathcal{B}$ are independent. Let $c \in K^E$ be a cycle and consider the cycle

$$\hat{c} := \sum_{e \in N} c_e v_{c_e}$$

Indeed, for any $e \in N$, we have $c(e) = \hat{c}(e)$, and hence $c - \hat{c}$ is a cycle using only edges of T. Thus $c - \hat{c} = 0$. $\qquad\square$

We note that we can also consider the *cycles space* as the subspace generated by the vector in the form v_c where c is a cycle of G.

1.2 Optimal Transport

In this section we summarize some results of Optimal Transport and Convex Analysis which will be used in the next chapter.

The history of optimal transport began in France with Gaspard Monge in 1781, who proposed a problem involving two collections of particles with unit mass, represented by densities f and g on $\mathbb{R}^d$. The objective was to find a map $T : \mathbb{R} \to \mathbb{R}$ that transports the first distribution to the second one, while minimizing a particular quantity: $M(\mathrm{T}) := \int_{\mathbb{R}^d} |\mathrm{T}(x) - x| \, dx$. In other words, the movement of each particle had to be chosen in such a way that it minimizes the average displacement. In its most general version, and in modern language the problem is stated as follows.

Problem 1.2.1 (Monge Problem)**.** *Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and a cost function $c : X \times Y \to [0, +\infty]$, the goal is to solve the following optimization problem:*

$$\inf \left\{ M(\mathrm{T}) := \int c(x, \mathrm{T}(x)) \, d\mu(x) \mid \mathrm{T} : X \to Y, \ \mathrm{T}_\# \mu = \nu \right\}. \quad \text{(MP)}$$

The measure $\mathrm{T}_{\#}\,\mu \in \mathcal{P}(Y)$ denotes the *pushforward measure* of μ by the map T and is defined through $(T_{\#}\mu)(A) := \mu(T^{-1}(A))$. A natural generalization appears from the work of Kantorovich:

Problem 1.2.2 (Kantorovich Problem). *Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and a cost function $c : X \times Y \to [0, +\infty]$, the goal is to solve the following optimization problem:*

$$\inf\left\{ \mathrm{K}(\gamma) := \int_{X \times Y} c\,d\gamma \mid \gamma \in \Pi(\mu, \nu) \right\}, \qquad \text{(KP)}$$

Where $\Pi(\mu, \nu) \subset \mathcal{P}(X \times Y)$ is the set of distributions having μ and ν as marginals, that is $(\pi_X)\gamma = \mu$ and $(\pi_Y)\gamma = \nu$ for all $\gamma \in \Pi(\mu, \nu)$, where π_X and π_Y are the two projections of $X \times Y$ onto X and Y respectively. The elements of $\Pi(\mu, \nu)$ are called transport plans.

We observe that every function $\mathrm{T} : X \longrightarrow Y$ maps μ onto ν if and only if the distribution $\gamma := (\mathrm{Id}, \mathrm{T})_{\#}\mu$ belongs to $\Pi(\mu, \nu)$ and therefore the Kantorovich Problem 1.2.2 is a relaxation of the Monge Problem 1.2.1. This is because $(\pi_X)_{\#}(\mathrm{Id}, \mathrm{T})_{\#}\mu = (\pi_X \circ (Id, T))_{\#} = Id_{\#}\mu = \mu$ and $(\pi_Y)_{\#}(\mathrm{Id}, \mathrm{T})_{\#}\mu = T_{\#}\mu = \nu$. We can now define the Wasserstein metric between two probability distributions $\mathbb{Q}_1$ and $\mathbb{Q}_2$ on X such that $\int_X x\mathbb{Q}_i(dx) < \infty$. The Wasserstein metric coincides to the Kantorovich Problem (1.2.2) with $\mu = \mathbb{Q}_1$, $\nu = \mathbb{Q}_2$ and $c(x, y) = \|x - y\|$.

Definition 1.2.1 (Wasserstein Metric). *The Wasserstein metric $\mathrm{d_w} : \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{R}_+$ is defined by*

$$\mathrm{d_w}(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\gamma \in \Pi(\mathbb{Q}_1, \mathbb{Q}_2)} \int \|x - y\|\,d\gamma \qquad (1.2)$$

Where $\Pi(\mathbb{Q}_1, \mathbb{Q}_2)$ is the set of distributions having marginals $\mathbb{Q}_1$ and $\mathbb{Q}_2$ respectively.

1.3 Convex Analysis

In this chapter we review foundational concepts of Convex Analysis, such as Duality and Minimax theorems. The main reference for this chapter is [Ber03].

Definition 1.3.1. *A subset C of $\mathbb{R}^n$ is* convex *if*

$$\alpha x + (1 - \alpha)y \in C, \qquad \forall x, y \in C, \ \forall \alpha \in [0, 1].$$

A function $f : C \longrightarrow \mathbb{R}$ is convex if C is a convex subset of $\mathbb{R}^n$ and:

$$f(\alpha x + (1 - \alpha)y) \le \alpha f(x) + (1 - \alpha)f(y), \qquad \forall x, y \in C, \ \forall \alpha \in [0, 1].$$

Since not all convex functions defined on a convex set $C \subset \mathbb{R}^n$ can be extend to real-valued convex functions on $\mathbb{R}^n$ (for example the function sending x to $\frac{1}{x^2}$ for $x > 0$), it is often convenient, instead of restricting the domain of f to where it takes real values, to allow f to take infinite values and extend the domain to all $\mathbb{R}^n$. Such extended real-valued convex functions can be characterized using the notion of the *epigraph* of f, which is the set of points lying above the graph of the function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\infty, -\infty\}$, that is

$$\mathrm{epi}(f) := \{(x, w) \in \mathbb{R}^{n+1} \mid w \geq f(x)\}. \tag{1.3}$$

It is often important to exclude the degenerate case where f is identically equal to $+\infty$, and the case where it takes the value $-\infty$ in some point. This is only true iff $\mathrm{epi}(f)$ contains a vertical line. We will thus say that a function $f : X \longrightarrow \mathbb{R}$ is proper if $f(x) < \infty$ for at least one $x \in X$ and $f(x) > -\infty$ $\forall x \in X$, and that f is *improper* if it is not proper. The generalization of the definition of convex function to extended real valued functions is as follows

Definition 1.3.2. *Let C be a convex subset of $\mathbb{R}^n$. We say than an extended real-valued function $f : C \longrightarrow [-\infty, \infty]$ is* convex *if $\mathrm{epi}(f)$ is a convex subset of $\mathbb{R}^{n+1}$.*

If f takes only real values, then this definition coincides with the traditional definition of convexity for real-valued functions.

The following are some basic properties of convex functions:

- **Sup of convex functions.** If $\{f_\alpha\}_{\alpha \in A}$ is an arbitrary family of convex functions on the same domain $C \subseteq \mathbb{R}^n$, then $f(x) := \sup_{\alpha \in A} f_\alpha(x)$ is a convex function.

- **Sublevel Sets.** For any α, the sublevel set $\{x \mid f(x) \leq \alpha\}$ is convex.

We will now develop a notion of direction of recession of a convex set. This notion is important in the context of existence of solutions of closed convex optimization problems.

Definition 1.3.3. *Let $C \subset \mathbb{R}^n$ be a closed convex set. We say that a vector $d \in \mathbb{R}^n$ is a* direction of recession *of C if $x + \alpha d \in C$ for all $x \in C$ and $\alpha \geq 0$. The set of all directions of recession is a cone containing the origin and is called* recession cone *of C and it is denoted by R_C.*

The recession cone of an intersection of closed convex sets corresponds to the intersection of their individual recession cones:

Observation 1.3.1. *If $\{C_i\}_{i \in I}$ is an arbitrary family of closed convex sets of $\mathbb{R}^n$ and $C = \cap_{i \in I} C_i$ is not empty, then $R_C = \cap_{i \in I} C_i$*

Proof. Let $x \in C \subset C_i$ and $d \in R_C$, then for all $\lambda \geq 0$ we have $x + \lambda d \in C \subset C_i$ for all $i \in I$. By the Recession Cone Theorem 1.3.3 if follows that $d \in R_{C_i}$ and $R_C \subset \cap_{i \in I} R_{C_i}$. Viceversa, if $d \in \cap_{i \in I} R_{C_i}$, then for all $i \in I$ and x in $C \subset C_i$, $x + \lambda d \in C_i$ for $\lambda \geq 0$ thus $\cap_{i \in I} R_{C_i} \subset R_C$. $\square$

The recession cone characterizes the compactness of closed convex sets:

Observation 1.3.2. *A closed convex set $C \subset \mathbb{R}^n$ is compact if and only if the recession cone of C, R_C is equal to $\{0\}$.*

Proof. If $R_C \neq 0$, then C is not bounded and hence not compact. Conversely, if C is not bounded, let z_k be an unbounded sequence in C. Choose any $x \in C$ and consider the sequence $d_k := \frac{z_k - x}{\|z_k - x\|}$ and let d be a limit point of d_k. For all $\alpha \in [0, \|z_k - x\|]$, $x + \alpha d_k \in C$ by the convexity of C. Hence taking the limit as $k \to \infty$, we have $x + \alpha d \in C$, $\forall \alpha \geq 0$. So R_C is not empty. $\square$

Theorem 1.3.3 (Recession Cone Theorem). *Let C be a nonempty closed convex set.*

- *The recession cone R_C is closed and convex.*

- *A vector d belongs to R_C if and only if there exists a vector $x \in C$ such that $x + \lambda d \in C$ for all $\lambda \geq 0$.*

Proof. We first prove convexity of R_C. If $d, b \in R_C$, $\forall \alpha \in [0, 1]$, we check that $\alpha d + (1 - \alpha)b \in R_C$. Let $x \in C$, and $\lambda \geq 0$, then

$$\lambda(\alpha d + (1 - \alpha)b) + x = \alpha(\lambda d + x) + (1 - \alpha)(\lambda b + x)$$

is in C since $\lambda d + x$ and $\lambda b + x$ are in C. Hence, $\alpha d + (1 - \alpha)b \in R_C$, and R_C is convex.

We now show that R_C is closed. If $\{d_k\}$ is a sequence of points in R_C such that $\lim_{k \to \infty} d_k = d$, then

$$\forall \lambda \geq 0, \ \forall x \in C, \ \lim_{k \to \infty} (\lambda d_k + x) = \lambda d + x,$$

so $\lambda d + x \in C$ because C is closed by hypothesis and $\lambda d_k + x$ are in C for all k. Hence, d is in R_C and R_C is closed.

We now prove the second statement. Let $x, y \in C$ with x as in the statement and y arbitrary and $\lambda \geq 0$. We want to prove that $y + \lambda d \in C$. If $\lambda = 0$ then $y + \lambda d = y \in C$. Let $\lambda > 0$. For all integers $n > 0$, $x + nd \in C$. Let

$$z_n := \frac{\lambda}{\|(x + nd - y)\|}(x + nd - y) + y.$$

There exists $\bar{n}$ such that $\forall n \geq \bar{n}$

$$\frac{\lambda}{\|(x + nd - y)\|} \leq 1,$$

thus z_n is in C for all $n \geq \bar{n}$ since $x + nd$ and y are in C. Also,

$$\lim_{n \to \infty} z_n = \lambda d + y,$$

hence $\lambda d + y \in C$ because C is closed. Since y and λ were arbitrary, this proves that $d \in R_C$. $\qquad \square$

The Recession Cone Theorem has some interesting consequences when applied to the convex sets charasteristic of convex functions. In particular the recession cones of the sublevel sets of a convex function either all coincide or are all empty. Moreover, these sublevels are either all compact or none are compact.

Proposition 1.3.4. *Let* $f : \mathbb{R}^n \longrightarrow (-\infty, \infty]$ *be a closed proper convex function and consider the level sets* $V_\gamma = \{x \mid f(x) \leq \gamma\}$. *Then:*

1. *All the nonempty level sets* V_γ *have the same recession cone, denoted* F_f, *and given by:* $R_f := \{d \mid (d, 0) \in R_{epi(f)}\}$. *Where* $R_{epi(f)}$ *is the recession cone of the epigraph of* f.

2. *If one nonempty level set* V_γ *is compact, then all of the level set are compact.*

Proof. (1) Fix a γ such that V_γ is non empty. We show that $R_f \subset R_{V_\gamma}$. $x \in V_\gamma$ if and only if $f(x) \leq \gamma$ and $(x, \gamma) \in epi(f)$. If $d \in R_f$, Then $(x + \lambda d, \gamma) = (x, \gamma) + \lambda(d, 0) \in R_f$ because $(d, 0) \in R_{epi(f)}$ by definition of R_f. Vice versa, if $d \in R_{V_\gamma}$ then $x + \lambda \in V_\gamma$, then $f(x + \lambda d) \leq \gamma$ and therefore $(x, \gamma) + \lambda(d, 0) \in epi(f)$ for all $\lambda \geq 0$, since f is closed then $epi(f)$ is closed, so by the Recession Cone Theorem 1.3.3, $(d, 0) \in R_{epi(f)}$ and $R_{V_\gamma} \subset R_f$.

(2) A closed convex set C is compact if and only if its recession cone $R_C = 0$. All empty level sets are compact. If there exists a nonempty compact $V_{\hat{\gamma}}$, the recession cone $R_f = 0$ and thus also all the other nonempty V_γ are compact. $\qquad \square$

The recession cone theorem also implies that an improper closed convex function f is either constantly ∞ or $-\infty$, since if there exists a point such that $f(x) = -\infty$, then $-e_{n+1} \in R_{epif(f)}$ and thus $f(x) = \infty$ for all x.

1.3.1 Conjugate function

We will now review the concept of *conjugacy transformation* or *Legendre tansformation* of a function $f : \mathbb{R}^n \longrightarrow [-\infty, \infty]$, the idea here is to describe f in terms of the affine functions that are majorized by f.

Definition 1.3.4 (Conjugate function of f). *Consider a function* $f : \mathbb{R}^n \longrightarrow [-\infty, \infty]$. *The* conjugate function of f *is the function* $f^* : \mathbb{R}^n \longrightarrow [-\infty, \infty]$ *defined by*

$$f^*(y) := \sup_{x \in \mathbb{R}^n} \{x'y - f(x)\} \qquad y \in \mathbb{R}^n \qquad (1.4)$$

We note that regardless of the structure of f, the conjugate of f, since it is the supremum of affine functions, is a closed convex function. Furthermore a fundamental property of the coonjugate functions is that $f : \mathbb{R}^n \to \mathbb{R}$ is closed, proper and if and only if $f = (f^*)^*$, where f^* is the Legendre transform of f. An especially interesting operation based on minimization which behaves well with conjugacy is epi-addition.

Definition 1.3.5. *For functions $f, f : \mathbb{R}^n \to [-\infty, +\infty]$ the epi-sum of f and g is defined by*

$$(f \# g)(x) := \inf_{x_1 + x_2 = w} (f(x_1) + g(x_2)) = \inf_w (f(w) + g(x - w)). \qquad (1.5)$$

We note that if since $f \# g$ is the infimum of $f(w) + g(x - w)$, then if f and g are proper, closed and convex then so is $f \# g$.

Observation 1.3.5. *For all functions $f, g : \mathbb{R}^n \to [-\infty, +\infty]$ we have that $(f^* \# g^*)^* = f^{**} + g^{**}$. In particular if f and g are closed, proper and convex, then $(f^* \# g^*) = (f + g)^*$.*

Proof.

$$(f^* \# g^*)^*(x) = \sup_y [xy - (f^* \# g^*(y))] = \sup_y [xy - (f^* \# g^*(y))] =$$

$$= \sup_y [xy - \inf_{y_1 + y_2 = t} (f^*(y_1) + g^*(y_2))] =$$

$$= \sup_y [xy + \sup_{y_1 + y_2 = y} (-f^*(y_1) - g^*(y_2))] =$$

$$= \sup_{\substack{y \\ y_1 + y_2 = y}} [xy - f^*(y_1) - g^*(y_2))] =$$

$$= \sup_{y_1, y_2} [xy_1 + xy_2 - f^*(y_1) - g^*(y_2))] = f^{**}(x) + g^{**}(x).$$

$\square$

Example 1. *Let $\Xi \subset \mathbb{R}^N$, the characteristic functions χ_Ξ is define via $\chi_\Xi(x) = 0$ if $x \in \Xi$ and $\chi_\Xi(x) = \infty$ otherwise. Then the conjugate function of χ_Ξ is*

$$\delta_\Xi(y) := (\chi_\Xi)^*(y) = \sup_{x \in \mathbb{R}^N} x'y - \chi_\Xi = \sup_{x \in \mathbb{R}^N} x'y$$

Which is called the support function *of Ξ.*

1.3.2 Duality

Duality in optimization is often viewed as a manifestation of the fundamental description of a closed convex set as the intersection of all closed halfspaces

containing the set:

$$\begin{aligned}
\text{epi}(f) &= \{(x, w) \mid f(x) \le w\} \\
&= \{(x, w) \mid (f^*)^*(x) = \sup_{y \in \mathbb{R}^n} \{x'y - f^*(y)\} \le w\} \\
&= \cap_{y \in \mathbb{R}^n} \{(x, w) \mid x'y - f^*(y) \le w\}
\end{aligned}$$

Thus f^* defines the intersection of closed halfspaces which give $\text{epi}(f)$. Most essential characteristics of duality can be captured by the *min common* and *max crossing* (MC/MC) problems for short. The min common problem looks for the element in $M \subset \mathbb{R}^{n+1}$ with the smallest last component.

Problem 1.3.1 (Min Common Point Problem). *Given $M \subset \mathbb{R}^{n+1}$ find:*

$$\min w \tag{1.6}$$
$$s.t. \ (0, w) \in M. \tag{1.7}$$

Its optimal value is denoted by w^*. The Max Crossing Point Problem searches the point with the highest first component between the points on the $n+1$-axis that are contained in the union of upper halfspaces containing M and that are nonvertical (non-parallel to e_{n+1}). That is the halfspaces in the form $H_{\mu,q} := \{(u, w) \in \mathbb{R}^{n+1} \mid \mu u + w \ge q\}$ such that $q \le \mu'u + w \ \forall (u, w) \in M$ or equivalently $q \le \inf_{(u,w) \in M} \{w + \mu'u\}$. Thus we express the (MC) problem as:

Problem 1.3.2 (Max Crossing Point Problem). *Let $q : \mathbb{R}^n \to \mathbb{R}$, $q(u) := \inf_{(u,w) \in M} \{w + \mu'u\}$. Find:*

$$\max q(\mu) \tag{1.8}$$
$$s.t. \ \mu \in \mathbb{R}^n. \tag{1.9}$$

We also refer to this as the *dual problem* and denote its optimal value as $q*$. We note that since q is the infimum of affine functions we have:

Observation 1.3.6. *q is concave and upper semi-continuous.*

Theorem 1.3.7 (Weak Duality). *The following holds*

$$q^* \le w^*. \tag{1.10}$$

Proof. For every $(0, w) \in M$ and for all $\mu \in \mathbb{R}^n$ we have $w = w + \mu'0 \ge \inf_{(u,w) \in M} \{w + \mu'u\} = q(\mu)$. Taking the infimum over $(0, w) \in M$ on the left side we have $w^* \ge q(\mu)$ for all $\mu \in \mathbb{R}^n$, thus taking the supremum over $\mu \in \mathbb{R}^n$ over the right side we obtain the thesis. $\qquad \square$

Conditions for which the inequality (1.10) becomes an equality are called *MC/MC strong duality* conditions.

1.3.3 Nonvertical separating Hyperplanes

MC/MC strong duality theorems essentially need to show that the inequality $q^* \geq w^*$ holds. Because of the definition of the Max Crossing problem, this is essentially done by showing that the point $(0, w^* - \varepsilon)$ can always be separated by a nonvertical hyperplane from the set M. In this section we review some Nonvertical separating hyperplane results. We say that two sets C_1 and C_2 in a normed K-vector space X, where $K = \mathbb{C}$ of $K = \mathbb{R}$, are separated by a closed hyperplane if there exists a continuous linear function $f : X \longrightarrow \mathbb{R}$ such that for all $x \in C_1$ and $y \in C_2$ we have that $\langle f, x \rangle \leq \langle f, y \rangle$ where $\langle f, x \rangle := f(x)$. f is continuous if and only if $\|f\|_{X'} := sup_{x \in X} \langle f, x \rangle < \infty$. Furthermore $\| \cdot \|_{X'}$ is a norm on X' called *continuity norm*, where X' is the set of continuous linear functions form X to $\mathbb{R}$. If X is a finite dimensional vector space X' is isomorphic to X and and the continuity norm is also called *dual norm* of $\| \cdot \|$ and is denoted by $\| \cdot \|_*$. And closed hyperplane separation is equivalent to the existence of a vector $a' \neq 0$ such that for every $x \in C_1$, $y \in C_2$ we have $a'x \leq a'y$ or equivalently that $sup_{x \in C_1} a'x \leq \inf_{x \in C_2} a'x$. We say that C_1 and C_2 are *strictly separated by an hyperplane* if the inequality is strict. In this case X' is isomorphic to X and and the continuity norm is also called *dual norm* of $\| \cdot \|$ and is denoted by $\| \cdot \|_*$. An important general result for separating hyperplanes is the Geometric Hahn Banach Theorem which we recall here without proof.

Theorem 1.3.8 (Geometric Hahn Banach Theorem). *Let X be a K-vector space, where $K = \mathbb{C}$ or $K = \mathbb{R}$. $A, B \subset X$ are two non empty, disjoint convex sets.*

1. *If A is open then there exists a closed hyperplane separating A and B.*

2. *If A is closed and B is compact then there exists a closed hyperplane strictly separating A and B.*

An important consequence is the following characterization of closed convex sets.

Corollary 1.3.8.1. *If $C \subset \mathbb{R}^n$ is a closed convex set iff intersection of closed halfspaces containing C is equal to C.*

Proof. If C is the intersection of closed halfspaces then it is closed and convex. Viceversa, lets denote with $\hat{C}$ the intersection of closed halfspaces containing C. $C \subset \hat{C}$ is always true. If C is closed and convex, let $x \notin C$, by Hahn Banach there exists a closed hyperplane H strictly separating x and C. Thus $x \notin \hat{C}$ for all $x \notin C$, therefore $\hat{C} \subset C$. $\square$

Theorem 1.3.9 (Nonvertical Hyperplane Theorem). *Let C be a nonempty convex subset of $\mathbb{R}^{n+1}$ that contains no vertical lines. Let the vectors in $\mathbb{R}^{n+1}$ be denoted by (u, w), where $u \in \mathbb{R}^n$ and $w \in \mathbb{R}$. Then:*

1. *C is contained in a closed halfspace corresponding to a novertical hyperplane, i.e., there exist a vector $\mu \in \mathbb{R}^n$, a scalar $\beta \neq 0$, and a scalar γ such that*

$$\mu'u + \beta w \geq \gamma, \qquad \forall (u,w) \in C. \tag{1.11}$$

2. *If $(\bar{u}, \bar{w})$ does not belong to $cl(C)$, the closure of C, there exists a nonvertical hyperplane strictly separating $(\bar{u}, \bar{w})$ and C.*

Proof. 1) We prove the first statement by contradiction. If every hyperplane containing C in one of its closed halfspace is vertical, then since $C \subset cl(C)$, the same must be true for all the halfspaces containing $cl(C)$. Therefore all the closed halfspaces containing $cl(C)$ have e_{n+1} in their recession cones. Since the intersection of the closed halfspaces containing $cl(X)$ is equal to $cl(X)$, for the Observation 1.3.1 follows that $e_{n+1} \in R_{cl(C)} = R_{cl(C)}$ which contradicts the hypothesis that C does not contain a vertical line.

2) If $(\bar{u}, \bar{w}) \notin cl(C)$, by the Geometric Hahn Banach Theorem there exists an hyperplane strictly separating $(\bar{u}, \bar{w})$ and $cl(C)$. If this hyperplane is nonvertical, since $C \subset cl(C)$, we are done. Otherwise, we have a nonzero vector $\bar{\mu}$ and a scalar $\bar{\gamma}$ such that for all $(u,w) \in cl(C)$

$$\bar{\mu}'u > \bar{\gamma} > \bar{\mu}'\bar{u}. \tag{1.12}$$

The concept is to make a slight adjustment to the orientation of the hyperplane, ensuring it is not vertical, while still maintaining the separation. Since C does not contain a vertical line, neither does $cl(C)$, as seen in part (1) of the proof. Hence, by part (1), there exists a nonvertical hyperplane containing $cl(C)$ in one of its closed halfspaces, so that for some (μ, β) and γ, with $\beta \neq 0$, we have $\mu'u + \beta w \geq \gamma$ for all $(u,w) \in cl(C)$. By multiplying this relation with an $\varepsilon > 0$ and combining with (1.12) we obtain for all $(u,w) \in cl(C)$

$$(\bar{u} + \varepsilon\mu)'u + \varepsilon\beta w > \bar{\gamma} + \varepsilon\gamma, \ \forall \varepsilon > 0.$$

Since $\bar{\gamma} > \bar{\mu}'\bar{u}$ there is a small enough ε such that $\gamma + \varepsilon\gamma > (\bar{m}u + \varepsilon\mu)'\bar{u} + \varepsilon\beta\bar{w}$.

From the above two relations we have

$$(\bar{\mu} + \varepsilon\mu)'u + \varepsilon\beta w > (\bar{\mu} + \varepsilon\mu)'\bar{u} + \varepsilon\beta\bar{w} \ \forall (u,w) \in cl(C).$$

Thus the hyperplane defined by $(\bar{\mu} + \varepsilon\mu, \varepsilon\beta)$ strictly separates $(\bar{u}, \bar{w})$ and $cl(C)$ and therefore also C. □

Theorem 1.3.10 (MC/MC Strong Duality). *Let w^* be given such that either $w^* < \infty$, or $w^* = \infty$ and M contains no vertical lines. If the set*

$$\bar{M} = M + \{(0,w) \mid w \geq 0\}$$

is convex, then $q^* = w^*$ if and only if for every sequence $\{(u_k, w_k)\} \subset M$ with $u_k \to 0$, it holds that $w^* \leq \liminf_{k \to \infty} w_k$.

Proof. Assume that $w^* \leq \liminf_{k\to\infty} w_k$ for all $\{(u_k, w_k)\}$ with $u_k \to 0$. Consider the case $w^* = -\infty$ then by the weak duality theorem also $q^* = -\infty$, so the conclusion trivially follows. Consider next the case where $w^* = \infty$ and M contains no vertical lines. Since $\liminf_{k\to\infty} w_k = w^* = \infty$ for all $\{(u_k, w_k)\}$ with $u_k \to 0$, it follows that the vertical axis of $\mathbb{R}^{n+1}$ contains non closure points of $\bar{M}$. Hence by the Nonvertical Hyperplane Theorem, for any vector $(0, w) \in \mathbb{R}^{n+1}$, there exists a nonvertical hyperplane strictly separating $(0, w)$ and $\bar{M}$. The crossing point of this hyperplane with the vertical axis lies between w and q^*, so $w < q^*$ for all $w \in \mathbb{R}$ which implies $q^* = w^* = \infty$.

Consider finally the case where w^* is a real number. We first show by contradiction that for any $\varepsilon > 0$, $(0, w^* - \varepsilon) \notin \operatorname{cl}(\bar{M})$, where $\operatorname{cl}(\bar{M})$ is the closure of $\bar{M}$. If $(0, w^* - \varepsilon)$ were a closure point of $\bar{M}$ for some $\varepsilon > 0$, let $\{(u_k, \bar{w}_k)\} \subset \bar{M}$ be a sequence converging to $(0, w^* - \varepsilon)$. In view of the definition of $\bar{M}$, this implies the existence of another sequence $\{(u_k, w_k)\} \subset M$ with $u_k \to 0$ and $w_k \leq \bar{w}_k$ for all k, so that $\liminf_{k\to\infty} w_k \leq w^* - \varepsilon$, which contradicts the assumption $w^* \leq \liminf_{k\to\infty} w_k$. We next argue by contradiction that $\bar{M}$ does not contain any vertical lines. If this were not so, by convexity of $\bar{M}$ and the Recession Cone Theorem 1.3.3, $(0, -1)$ would be a direction of recession of $\operatorname{cl}(\bar{M})$. Because $(0, w^*) \in \bar{M}$, it follows that the entire half-line $\{(0, w - \varepsilon) \mid \varepsilon \geq 0\}$ belongs to $\operatorname{cl}(\bar{M})$, contradicting what was shown earlier.

Since $\bar{M}$ does not contain any vertical lines and the vector $(0, w^* - \varepsilon)$ does not belong to $cl(\bar{M})$ for any $\varepsilon > 0$, it follows from Non Vertical Hyperplane Theorem that there exists a nonvertical hyperplane strictly separating $(0, w^* - \varepsilon)$ and $\bar{M}$. This hyperplane crosses the $(n+1)$-st axis at a unique vector $(0, q)$, which must lie between $(0, w^* - \varepsilon)$ and $(0, w^*)$, i.e. $w^* - \varepsilon < q \leq w^*$. Furthermore $q \leq q^*$, thus together with weak duality we have:

$$w^* - \varepsilon \leq q^* \leq w^*$$

Since ε can be arbitrarily small, it follows that $q^* = w^*$. Conversely, assume $q^* = w^*$ and consider a sequence $(u_k, w_k) \subset M$ such that $u_k \to 0$. Indeed we have

$$q(\mu) = \inf_{(u,w)\in M} w + \mu' u \leq w_k + \mu' u_k \tag{1.13}$$

and by taking the limit as $k \to \infty$, we obtain

$$w^* = q^* = \sup_{\mu \mathbb{R}^n} q(\mu) \leq \liminf_{k\to\infty} w_k$$

$\square$

To illustrate an application of the MC/MC Strong Duality Theorem 1.3.10, we will prove the *LP strong duality theorem*.

Theorem 1.3.11 (LP strong duality theorem). *Consider the following Primal Linear problem:*

$$\inf_{x \in \mathbb{R}^n} c'x$$
$$\text{s.t. } Ax \le b \tag{1.14}$$

and the following dual problem

$$\sup_{p} \mu'b$$
$$\text{s.t. } \mu \ge 0 \tag{1.15}$$
$$\mu'A = c'$$

Then if the primal problem has an optimal solution then the dual has an optimal solution too, and the two optima are equal.

Proof. The proof consists in showing that with an appropriate selection of the set M, the assumption can be stated as a MC/MC problem satistying the assumptions of the MC/CM Strong Duality Theorem 1.3.10. Let $C := \{x \in \mathbb{R}^n \mid Ax \le 0\}$ be the feasible set of the LP Problem 1.14, we consider the "perturbed" sets $C_u := \{x \in \mathbb{R}^n \mid Ax - b \le u\}$ and the function

$$F(x, u) := \begin{cases} c'x & \text{if } x \in C_u \\ \infty & \text{otherwise} \end{cases} \tag{1.16}$$

and the partial minimization function defined as $p(u) := \inf_{x \in \mathbb{R}^n} F(x, u) = \inf_{Ax-b \le u} F(x, u)$. We consider as $M = epi(p) \subset \mathbb{R}^m$. Then $w^* = p(0)$, furthermore since $F(x, 0) = c'x$ then $p(0)$ also coincides with the optimal value of the LP Problem 1.14. While the objective function of the Max Crossing Problem 1.3.2,

$$q(\mu) = \inf_{u \in \mathbb{R}^m} p(u) + \mu'u = \begin{cases} \inf_{x \in \mathbb{R}^n} c'x + \mu'(b - Ax) & \text{if} \mu \ge 0 \\ -\infty & \text{otherwise} \end{cases}$$

Thus, since $c'x + \mu'(b - Ax) = (c' - \mu'A) + p\mu'b$:

$$q^* = sup_{\mu \ge 0} \inf_{x \in \mathbb{R}^n} (c' - \mu'A)x + \mu'b = \tag{1.17}$$

$$= sup_{\mu \ge 0, \mu'A=c'} \inf_{x \in \mathbb{R}^n} (c' - \mu'A)x + \mu'b = \tag{1.18}$$

$$= \sup_{\mu \ge 0, \mu'b=A} \mu'b \tag{1.19}$$

Thus q^* coincides with the optimal value of the Dual LP 1.15. We only need to show that the assumption of the MC/MC Strong Duality Theorem 1.3.10 hold. Since by assumption exists an optimal real solution to the LP Problem and such optimal is equal to w^*, then $w^* \le \infty$. The set $\bar{M}$ equals to $epi(p)$ and thus is convex. p is lower semicontinous since it's the infimum of affine functions. If $\{(u_k, w_k)\} \subset M$ with $u_k \to 0$, then $w_k \ge p(u_k)$ and $w^* \le \liminf_{k \to \infty} p(u_k) \le \liminf_{k \to \infty} w_k$. Thus by MC/MC Strong Duality theorem we have $w^* = q^*$. $\square$

1.3.4 Minimax theorems

A core inequality often used when formulating optimization problems in an equivalent form, and in particular the main result of finite reduction in Kuhn et al. [EK18] which will be used in this thesis, is the following.

Observation 1.3.12 (Max-min inequality). *For any function $\Phi : Z \times W \longrightarrow \mathbb{R}$ we have:*

$$\sup_{z \in Z} \inf_{w \in W} \Phi(z, w) \leq \inf_{w \in W} \sup_{z \in Z} \Phi(z, w). \tag{1.20}$$

Proof. For all $w \in W$ and $z \in Z$, since $\Phi(z, w) \leq \sup_{z \in Z} \Phi(z, w)$, we have

$$\inf_{w \in W} \Phi(z, w) \leq \inf_{w \in W} \sup_{z \in Z} \Phi(z, w)$$

and thus

$$\sup_{z \in Z} \inf_{w \in W} \Phi(z, w) \leq \inf_{w \in W} \sup_{z \in Z} \Phi(z, w).$$

$\square$

We are interested in when (1.20) becomes an equality. Results giving such conditions are referred to as *Minimax Theorems*.

Proposition 1.3.13. *Let $F : \mathbb{R}^{n+m} \longrightarrow (-\infty, \infty]$ be a closed proper convex function, and consider the function $f : \mathbb{R}^n \longrightarrow (-\infty, \infty]$ given by $f(x) := \inf_{z \in \mathbb{R}^m} F(x, z)$. If for some $\bar{x} \in \mathbb{R}^n$ and $\bar{\gamma} \in \mathbb{R}$ the set $\{z \mid F(\hat{x}, z) \leq \gamma\}$ is nonempty and compact, then f is closed proper convex. Furthermore, for each $x \in dom(f)$, the set of minima in the definition of $f(x)$ is nonempty and compact.*

Proof. We first show that f is convex since F is convex. If $epi(f) = \emptyset$, i.e. $f(x) = +\infty \ \forall x \in X$ then f is convex. Assume $epi(f) \neq \emptyset$ and let $(\bar{x}, \bar{w})$ and $(\hat{x}, \hat{w})$ be two points, not necessarily distinct, in $epi(f)$. Then $f(\bar{x}) < \infty$, $f(\hat{x}) < \infty$. And there exist sequences $\{\hat{z}_k\}$ and $\{\bar{z}_k\}$ such that $F(\bar{x}, \bar{z}_k) \to f(\bar{x})$ and $F(\hat{x}, \hat{z}_k) \to f(\hat{x})$. Then:

$$f(\alpha \bar{x} + (1 - \alpha)\hat{x}) \leq F(\alpha \bar{x} + (1 - \alpha)\hat{y}, \alpha \bar{z}_k + (1 - \alpha)\hat{z}_k)$$
$$\leq \alpha F(\bar{x}, \bar{z}_k) + (1 - \alpha)F(\hat{x}, \hat{z}_k)$$

By taking the limit $k \to \infty$ we obtain $f(\alpha \bar{x} + (1 - \alpha)\hat{x}) \leq \alpha f(\bar{x}) + (1 - \alpha)f(\hat{x})$ and thus f is convex. $\square$

We will also assume the following:

Assumption 1. *X and Z are nonempy convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ and $\Phi : X \times Z \longrightarrow \mathbb{R}$ is a function such that $\Phi(\cdot, z) : X \longrightarrow \mathbb{R}$ is convex and closed for each $z \in Z$, and $-\Phi(x, \cdot) : Z \longrightarrow \mathbb{R}$ is convex and closed for each $x \in X$.*

Theorem 1.3.14 (Minimax 1). *Let $p : \mathbb{R}^m \to \mathbb{R}$ be the function*

$$p(u) := \inf_{x \in X} \sup_{z \in Z} \{\Phi(x,z) - u'z\} \ \forall u \in \mathbb{R}^m.$$

If p satisfies either $p(0) < \infty$, or else $p(0) = \infty$ and $p(u) > -\infty$ for all $u \in \mathbb{R}^m$. Then:

$$\sup_{z \in Z} \inf_{x \in X} \Phi(x,z) = \inf_{x \in X} \sup_{z \in Z} \Phi(x,z)$$

if and only if p is lower semicontinuous at $u = 0$

Proof. The proof consists of showing that with an appropriate selection of the set M, the assumptions are equivalent to the corresponding assumptions of the MC/CM strong duality theorem 1.3.10. We choose M as the epigraph of p, $M = \bar{M} = \{(u,w) \mid u \in \mathbb{R}^m, \ p(u) \leq w\}$, which is convex since p is convex. Also, since $w^* = p(0) = \inf_{x \in X} \sup_{z \in Z} \Phi(x,z)$, we have that $p(0) < \infty$, or else $p(0) = \infty$ and $p(u) > -\infty$ for all u. Finally, the condition of lower semicontinuity at $u = 0$ is $p(0) \leq \liminf_{k \to \infty} p(u_k)$ for all sequences $\{u_k\}$ with $u_k \to 0$. This is equivalent to $w^* \leq \liminf_{k \to \infty} w_k$ for every sequence $\{(u_k, w_k)\} \subset M$. Thus, by the conclusion of the MC/CM strong duality theorem, the lower semicontinuity at $u = 0$ holds if and only if $q^* = w^*$, which is in turn equivalent to the minimax equality. $\square$

We consider the function of partial maximization $t : \mathbb{R}^n \mapsto (-\infty, \infty]$ given by:

$$t(x) = \begin{cases} \sup_{z \in Z} \Phi(x,z) & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

Theorem 1.3.15 (Minimax 2). *Assume that t is proper and that the level sets $\{x \mid t(x) \leq \gamma\}$, $\gamma \in \mathbb{R}$, are compact. Then:*

$$\sup_{z \in Z} \inf_{x \in X} \Phi(x,y) = \inf_{x \in X} \sup_{z \in Z} \Phi(x,y). \tag{1.21}$$

Proof. We consider the function $F : Z \times \mathbb{R}^m \longrightarrow \mathbb{R}$ given by:

$$F(x,u) = \begin{cases} \sup_{z \in Z} \{\Phi(x,z) - u'z\} & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

F is convex because it is the sup of closed convex functions. We note that $t(x) = F(x,0)$, so since t is proper there exist a point $x \in X$ such that $F(x,0) = t(x) < +\infty$. If F is improper then F has only values in $\{+\infty, -\infty\}$. Therefore the same is true for t which gives a contradiction since t is proper. It follows from Proposition 1.3.13, since the level sets of t are compact by hypothesis, that $p(u) := \inf_{x \in \mathbb{R}^n} F(x,u)$ is closed, proper and that $p(0)$ is finite. Finally the infimum over X in the right-hand side of the minimax equality is attained as the set of minima of t, which is not empty and compact since t is proper and has compact levels sets. $\square$

Chapter 2

Data-Driven Distributionally Robust Optimization

In this chapter, we present a summary of the results discussed in Mohajerin Esfahani et al. [EK18].

Stochastic programming is a powerful modeling paradigm for addressing optimization problems under uncertainty. In a typical single-stage stochastic program, the objective is to determine a decision or policy $x \in \Pi$, where Π is a subset of a finite dimensional $\mathbb{R}$ vector space, that minimizes the expected cost $\mathbb{E}^{\mathbb{P}}[h(x, \xi)]$, where the expectation is taken with respect to the distribution $\mathbb{P}$ of the random vector $\xi \in \mathbb{R}^m$. However, in practice, the true distribution $\mathbb{P}$ is often unknown and needs to be inferred from data.

When we only have a limited number of samples, minimizing the cost over the inferred distribution can lead to poor out-of-sample performance. To address this issue, an alternative approach is to consider the worst-case expected cost over a family of plausible distributions $\mathcal{P}$ for the random variable ξ. This formulation results in a Distributionally Robust Optimization problem of the form $\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[h(x, \xi)]$. The design of the ambiguity set $\mathcal{P}$ is crucial in this framework. Ideally, the ambiguity set should have a high confidence of containing the true data-generating distribution. At the same time, it should exclude pathological distributions that could lead to overly conservative decisions. Furthermore, it is desirable for the ambiguity set to allow for a tractable reformulation of the problem from a computational perspective, enabling efficient solution approaches. A possibility is to define the ambiguity set as a ball centered on the inferred from data probability $\hat{\mathbb{P}}$ in the space of probability distributions by using a probability distance function such as the Wasserstein metric. The Wasserstain distance between $\mathbb{Q}_1$ and $\mathbb{Q}_2$ arises from Optimal Transport Theory and can be viewed as the minimum trasportation cost for moving the probability distribution mass from $\mathbb{Q}_1$ to $\mathbb{Q}_2$.

In the remainder of this thesis, we will consider as the ambiguity set

the Wasserstein ball centered on the empirical distribution $\hat{\mathbb{P}}_N := \frac{1}{N} \sum_{i1}^{N_s} \delta_{\hat{\xi}_i}$ with radius ε. Where $\hat{\xi}_i$ for $i = 1, \ldots, N_s$ are referred to as *data samples* and are N_s independent realizations of the random variable ξ having probability distribution $\mathbb{P}$.

$$\mathbb{B}_\varepsilon(\hat{\mathbb{P}}_{N_s}) := \{\mathbb{Q} \in \mathcal{M}(\Xi) \mid \mathrm{d_w}(\mathbb{P}, \mathbb{Q}) \leq \varepsilon\}. \tag{2.1}$$

where $\mathcal{M}(\Xi)$ is the set of probability distributions over the ambiguity set $\Xi \subset \mathbb{R}^{N_\xi}$, and $\mathbb{P} \in \mathcal{M}(\Xi)$. Therefore the distributionally robust optimization problem we are trying to solve is the following:

$$\inf_{x \in \Pi} \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[h(x, \xi)] \tag{2.2}$$

The associated *worst-case expectation problem*

$$\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[h(x, \xi)] \tag{2.3}$$

constitutes an infinite-dimensional optimization problem over probability distributions and is thus intractable. We will now review the results in [EK18] showing that (2.3) can be re-expressed as a finite-dimensional convex program under the following assumptions on the uncertainty set Ξ and the cost function $h : \mathbb{R}^{N_\xi} \times \mathbb{R}^n \longrightarrow [-\infty, \infty]$.

Assumption 2 (Convexity). *The uncertainty set $\Xi \subset \mathbb{R}^{N_\xi}$ is convex and closed, and the loss function $h(\xi, x) = \max_{k \leq K} \ell_k(\xi, x)$ is a pointwise maximum of more elementary measurable functions $\ell_k : \mathbb{R}^{N_\xi} \times \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$. The negative constituent functions $-\ell_k$ are proper, convex and lower semicontinuous with respect to ξ for all $k \leq K$. Moreover, we assume ℓ_k is not identically $-\infty$ on ξ for all $k \leq K$.*

Given that $\hat{\mathbb{P}}_{N_s}$ follows a discrete probability distribution, the probability measures within $\Pi(\mathbb{P}_{N_s}, \mathbb{Q})$ can be expressed as:

Observation 2.0.1. *Let $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{M}(\Xi)$ with $\mathbb{P}_1 = \sum_{i=1}^n \lambda_i \delta_{\hat{\xi}_i}$. Then $\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)$ can be reformulated as $\gamma = \sum_{i=1}^n \lambda_i \delta_{\hat{\xi}_t} \otimes \mathbb{Q}_i$. Where, $\mathbb{Q}_i$ is in $\mathcal{M}(\Xi)$ and $\mathbb{Q} = \sum_{i=1}^n \lambda_i \mathbb{Q}_i$.*

Proof. Let $\mathbb{Q}_i(A) := \frac{\gamma(\{\xi_i\} \times A)}{\lambda_i}$ for all $A \in \mathcal{B}(\Xi)$. $\mathbb{Q}_i$ is in $\mathcal{M}(\Xi)$ and for all $B \in \mathcal{M}$ we have $\gamma(A \times B) = \gamma(A \setminus \{\xi_1, ..., \xi_n\} \times B \cup (A \cap \{\xi_1, ..., \xi_n\}) \times B) = \sum_{\xi_i \in A} \gamma(\xi_i \otimes B) = \sum_{i=1}^n \lambda_i \delta_{\xi_i} \otimes \mathbb{Q}_i(A \times B)$. Thus, γ and $\sum_{i=1}^n \lambda_i \delta_{\xi_i} \otimes \mathbb{Q}_i$ coincide on a π-set generating $\mathcal{M}(\Xi^2)$ and therefore coincide. $\square$

Theorem 2.0.2 (Convex reduction). *If the convexity Assumption 2 holds, then for any $\varepsilon \geq 0$ the worst-case expectation $\sup_{\mathbb{Q} \in B(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}(\ell)$ equals the*

optimal value of the finite convex program:

$$\inf_{\lambda, s_i, z_{ik}, \nu_{ik}} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^{N} s_i \tag{2.4}$$

$$\text{s.t. } [-\ell_k]^*(z_{ik} - \nu_{ik}) + \sigma_\Xi(\nu_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \;\; \forall i \leq N, \; \forall k \leq K \tag{2.5}$$

$$\|z_{ik}\|_* \leq \lambda \tag{2.6}$$

where $[-\ell_k]^*$ denotes the conjugate of $-\ell_k$ and $\|\cdot\|_*$ is the dual norm of z_{ik}.
χ_Ξ is the characteristic function of Ξ and σ_Ξ is its support function.

Proof. By using definition (1.2.1) we have:

$$\sup_{\mathbb{Q} \in B_\varepsilon(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}(\ell) = \begin{cases} \sup_{\gamma, \mathbb{Q}} \int_\Xi \ell(x) \mathbb{Q}(dx) \\ \text{s.t. } \int_{\Xi^2} \|x - y\| \gamma(dx, dy) \leq \varepsilon, \\ \qquad \gamma \in \Pi(\hat{\mathbb{P}}_N, \mathbb{Q}). \end{cases} \tag{2.7}$$

Using the previous observation with $\mathbb{P}_1 := \hat{\mathbb{P}}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\xi}_i}$, we have that for all $\gamma \in \Pi(\hat{\mathbb{P}}_N, \mathbb{Q})$ there exist $\mathbb{Q}_1, \ldots, \mathbb{Q}_N$ such that $\mathbb{Q} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{Q}_i$ and $\gamma = \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\xi}_i} \otimes \mathbb{Q}_i$. Therefore,

$$\int_\Xi \ell d\mathbb{Q} = \frac{1}{N} \sum_{i=1}^{N} \int_\Xi \ell d\mathbb{Q}_i,$$

$$\int_{\Xi^2} \|x - y\| \, \gamma(dx, dy) = \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi^2} \|x - y\| \, (\delta_{\hat{\xi}_i} \otimes \mathbb{Q}_i)(dx, dy) =$$

$$= \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi^2} \|\hat{\xi}_i - y\| \, \mathbb{Q}_i(dy)$$

We obtain that (2.7) equals to:

$$\begin{cases} \sup_{\mathbb{Q}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^{N} \int_\Xi \ell d\mathbb{Q}_i \\ \text{s.t. } \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi^2} \|\hat{\xi}_i - y\| \, \mathbb{Q}_i(dy) \leq \varepsilon. \end{cases} \tag{2.8}$$

Using a duality argument, we have that:

$$\sup_{\mathbb{Q}\in B_\varepsilon(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}(\ell) = \sup_{\mathbb{Q}_i\in\mathcal{M}(\Xi)} \inf_{\lambda\geq 0} \frac{1}{N}\sum_{i=1}^{N}\int_\Xi \ell\, d\mathbb{Q}_i$$

$$+ \lambda\left(\varepsilon - \frac{1}{N}\sum_{i=1}^{N}\int_\Xi \|y-\hat{\xi}_i\|\mathbb{Q}(dy)\right) \tag{2.9}$$

$$\leq \inf_{\lambda\geq 0} \sup_{\mathbb{Q}_i\in\mathcal{M}(\Xi)} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\int_\Xi (\ell(y)-\lambda\|y-\hat{\xi}_i\|)\mathbb{Q}_i(dy) \tag{2.10}$$

where (2.10) follows from the Max-min inequality (1.20). Since Ξ is compact there exists $\hat{y}\in\Xi$ such that $\sup_{y\in\Xi}(\ell(y)-\lambda\|y-\hat{\xi}_i\|) = \ell(\hat{y})-\lambda\|\hat{y}-\hat{\xi}_i\|$. Passing the sup inside the summation, since $\int_\Xi(\ell(y)-\lambda\|y-\hat{\xi}_i\|)\mathbb{Q}_i(dy) \leq \ell(\hat{y})-\lambda\|\hat{y}-\hat{\xi}_i\| = \int_\Xi(\ell(y)-\lambda\|y-\hat{\xi}_i\|)\delta_{\hat{y}}(dy)$, and $\delta_{\hat{y}}\in\mathcal{M}(\Xi)$, we have that (2.10) equals to:

$$\inf_{\lambda\geq 0} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\sup_{y\in\Xi}(\ell(y)-\lambda\|y-\hat{\xi}_i\|). \tag{2.11}$$

Introducing the epigraphical variables s_i, $i\leq N$, allows us to reformulate (2.11) as

$$\begin{cases} \inf_{\lambda,s_i} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \\ \text{s.t. } \sup_{y\in Xi}\left(\ell(y)-\lambda\|y-\hat{\xi}_i\|\right) \leq s_i \quad \forall i\leq N \\ \lambda\geq 0 \end{cases} \tag{2.12}$$

$$= \begin{cases} \inf_{\lambda,s_i} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \\ \text{s.t. } \sup_{y\in Xi}\left(\ell_k(y)-\max_{\|z_{ik}\|_*\leq\lambda}\langle z_{ik}, y-\hat{\xi}_i\rangle\right) \leq s_i \quad \forall i\leq N \\ \lambda\geq 0 \end{cases} \tag{2.13}$$

$$\leq \begin{cases} \inf_{\lambda,s_i} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \\ \text{s.t. } \min_{\|z_{ik}\|_*\leq\lambda}\sup_{y\in\Xi}\left(\ell_k(y)-\langle z_{ik}, y-\hat{\xi}_i\rangle\right) \leq s_i \quad \forall i\leq N \\ \lambda\geq 0 \end{cases} \tag{2.14}$$

In the formulation (2.13), we leverage the definition of the dual norm and the decomposibility of $\ell(\xi)$ into its components ℓ_k, $k \leq K$. Then we obtain (2.14) by converting the maximization to a minimization and using the Max-min inequality (1.20). This transition leads to a restriction of the feasible set and consequently in an eleveted infimum value. The resulting upper bound (2.14) can re re-expressed as

$$
\begin{cases}
\inf_{\lambda, s_i, z_{ik}} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^{N} s_i \\
\quad \text{s.t.} \sup_{y \in \Xi} \left(\ell_k(y) - \langle z_{ik}, y \rangle \right) + \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \quad \forall i \leq N, \ \forall k \leq K \\
\quad \|z_{ik}\|_* \leq \lambda
\end{cases}
\tag{2.15}
$$

$$
=
\begin{cases}
\inf_{\lambda, s_i, z_{ik}} \lambda \varepsilon + \frac{1}{N} \sum_{i=1}^{N} s_i \\
\quad \text{s.t.} \left[-\ell_k + \chi_\Xi \right]^* (z_{ik}) - \langle z_{ik}, \hat{\xi}_i \rangle \leq s_i \quad \forall i \leq N, \ \forall k \leq K \\
\quad \|z_{ik}\|_* \leq \lambda
\end{cases}
\tag{2.16}
$$

where (2.16) follows from the definition of conjugacy, and the substitution of z_{ik} with $-z_{ik}$. Next, we show that Assumption 2 reduces the inequalities (2.10) and (2.14) to equalities. Under Assumption 2 the inequality (2.10) is in fact an equality for any $\varepsilon > 0$ by virtue of the Strong Duality Theorem for moment problems (Prop 3.4 in [Sha00]). It can also be easily shown the the equality holds also for when $\varepsilon = 0$. The inequality (2.14) also reduces to an equality under Assumption 2 thanks to the Minimax Theorem (Theorem 1.3.15). Assumption 2 further implies that the function $-\ell_k + \chi_\Xi$ is proper, convex and lower semicontinuous. Properness holds because ℓ_k is not identically $-\infty$ on Ξ. By Observation 1.3.5 we have that:

$$
\left[-\ell_k + \chi_\Xi \right]^* (z_{ik}) = \inf_{\nu_{ik}} \left(\left[-\ell_k \right]^* (z_{ik} - \nu_{ik}) + (\chi_\Xi)^* (\nu_{ik}) \right)
\tag{2.17}
$$

$$
= \inf_{\nu_{ik}} \left(\left[-\ell_k \right]^* (z_{ik} - \nu_{ik}) + \sigma_\Xi(\nu_{ik}) \right).
\tag{2.18}
$$

Thus (2.16) is equivalent to (2.4) and this concludes the proof. $\qquad \square$

Corollary 2.0.2.1 (Piecewise affine loss fuctions)**.** *Suppose that the uncertainty set is a polytope, that is $\Xi = \{ \xi \in \mathbb{R}^{N_\xi} \mid H\xi \leq d \}$ where H is a matrix and d is a vector of appropriate dimensions. Moreover, consider the affine functions $a_k(\xi) := \langle a_k, \xi \rangle + b_k$ for all $k \leq K$. If $\ell(\xi) = \max_{k \leq K}(a_k(\xi) + b_k)$,*

then the worst-case expectation (2.3) evaluates to:

$$\inf_{\lambda,s_i,\gamma_{ik}} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \tag{2.19}$$

$$s.t. \; b_k + \langle a_k, \hat{\xi}_i \rangle + \langle \gamma_{ik}, d - H\hat{\xi}_i \rangle \le s_i \qquad \forall i \le N, \forall k \le K, \tag{2.20}$$

$$\|H^T\gamma_{ik} - a_k\|_* \le \lambda \qquad \forall i \le N, \forall k \le K, \tag{2.21}$$

$$\gamma_{ik} \ge 0 \qquad \forall i \le N, \forall k \le K. \tag{2.22}$$

Proof. The assertion is a consequence of Theorem 2.0.2, which applies because ℓ is the supremum of a finite number of affine function and thus Assumption 2 holds. By definition of the conjugacy operator we have:

$$[-\ell_k]^*(z) = [-a_k]^*(z) = \sup_y \left(\langle z, y \rangle + \langle a_k, y \rangle + b_k \right) = \begin{cases} b_k & \text{if } z = -a_k, \\ \infty & \text{otherwise} \end{cases}$$

and if $H = 0$ and $d = 0$ (i.e. $\Xi = \mathbb{R}^{N_\xi}$) then $\delta_\Xi(\nu) = 0$ iff $\nu = 0$ and equals $+\infty$ otherwise. Thus expression (2.4) becomes:

$$\inf_{\lambda,s_i,\gamma_{ik}} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \tag{2.23}$$

$$s.t. \; b_k + \langle a_k, \hat{\xi}_i \rangle \le s_i \qquad \forall i \le N, \forall k \le K \tag{2.24}$$

$$\|a_k\|_* \le \lambda \qquad \forall k \le K \tag{2.25}$$

$$\gamma_{ik} \ge 0 \qquad \forall i \le N, \forall k \le K \tag{2.26}$$

which corresponds to (2.19) with $H = 0$ and $d = 0$. If $H \ne 0$ or $d \ne 0$ then we have

$$\sigma_\Xi(\nu) = \begin{cases} \sup_y & \langle \nu, y \rangle \\ \text{s.t.} & Hy \le d \end{cases} = \begin{cases} \inf_{\gamma \ge 0} & \langle \gamma, d \rangle \\ \text{s.t.} & H^T\gamma = \nu, \end{cases}$$

where the last equality follows from the LP Strong Duality theorem (1.3.10), which holds as the uncertainty set is non-empty. The Corollary then follows by substituting the above expressions into (2.4). $\qquad\square$

Chapter 3

OPF problem formulation

Optimal Power Flow (OPF) is the set of optimization problems in electric power systems engineering which seek to optimize the operation of an electrical power system subject to the physical constraints imposed by electrical laws and engineering limits. We refer to [FR16] and [Bie+20] for this chapter's introduction to Optimization of power systems.

Electric power systems may be modeled as a network of buses (nodes) interconnected by branches (arcs) that represent transmission lines, transformers, and similar power systems equipment. Buses represent points of interconnection between power systems equipment and branches represent paths for the flow of electrical current. The objective of electric power system is the transfer of electrical energy from generation (supply) buses to load (demand) buses.

For the OPF model construction it is convenient to model the network as directed graph $(\mathbf{B}, \mathbf{L})$ where $\mathbf{B}$ is the set of Buses and and $\mathbf{L} \subset \mathbf{B} \times \mathbf{B}$ is the set of branches of the network and for each adjacent buses k, m both (k, m) and (m, k) are in $\mathbf{L}$. So the line l adjacent to k, m is modeled by two edges in the arc $\{(k, m), (m, k)\}$. L can be partitioned in L_0 and L_1 with $|L_0| = |L_1|$ where every line l, adjacent to the buses k, m and with a transformer at k, is oriented so that $(k, m) \in L_0$ and $(m, k) \in L_1$. We also consider a set $\mathcal{G}$ of generators, partitioned into (possibly empty) subsets $\mathcal{G}_k$ for every bus $k \in \mathbf{B}$.

Phasor representation

In Alternating Current Optimal Power Flow (ACOPF) all voltages and currents are sinusoids with fixed magnitude, frequency, and phase shift. Under these conditions the time domain differential equations governing the voltages and currents of the system can be transformed into a set of complex algebraic equations. The frequency f of a signal is fixed therefore it is not considered in such transformation. The *phasor* of the sinusodial function $c \sin(2\pi f t + \gamma)$ is the complex value $c e^{j\gamma}$. We note that the imaginary part

of the phasor corresponds to the sinusoidal function at $t = 0$ and therefore the phasor univocally represents the sinusoidal function. We denote the phase current on the line l adjacent to the buses k, m by I_{km}, voltage at bus k by $V_k = |V_k|e^{i\delta_k}$, where $|V_k|$ is the *voltage magnitude* and δ_k is the *voltage angle* or *voltage phase* at bus k. Situations where the magnitudes are zero typically represent exceptional cases like faults or open circuits. In normal system analysis and operation, voltages are assumed to be nonzero, allowing voltage angles to be well-defined. The power injected in l at k is denoted by $S_{km} = P_{km} + iQ_{km} \in \mathbb{C}$, where the real and imaginary parts P_{km} and Q_{km} are called *real power* and *reactive power* respectively. Kirchoff's current law remains valid for AC circuit analysis.

Definition 3.0.1 (Kirchoff's current law). *For all buses $k \in \mathbf{B}$ we have:*

$$\sum_{e \in \delta^+(k)} I(e) = \sum_{e \in \delta^-(k)} I(e). \tag{3.1}$$

We note that considering the vector $I \in \mathbb{C}^{\mathbf{L}}$ having as components for all $(k, m) \in \mathbf{L}$ $I(a, b) = I_{ab}$, the Kirchoff's current law corresponds to the flow conservation (1.1) introduced in Chapter 1.1, thus I is a $\mathbb{C} - cycle$. Meanwhile Ohm's law takes the following form:

Definition 3.0.2 (Ohm's law for phasors). *For all $(k, m) \in \mathbf{L}$ we have:*

$$S_{km} = V_k I_{km}^* \tag{3.2}$$

where I_{km}^* denotes the complex conjugate value of I_{km}.

All feasible solutions to the OPF problem must satisfy equations (3.1) and (3.2). Equivalently, we can replace (3.1) with the power flow equations, which are obtained by taking the conjugate of the sums in (3.1), multiplying by V_k, and then applying (3.2).

Definition 3.0.3 (Power flow equations). *For all buses $k \in \mathbf{B}$ we have:*

$$\sum_{e \in \delta^+(k)} S(e) = \sum_{e \in \delta^-(k)} S(e) \tag{3.3}$$

where for all $e = (k, m) \in \mathbf{L}$, $S(e) = S_{km}$.

If the line l hosts a transformer *close to the k end*, then (k, m) and (m, k) have different current injected power values, for this reason for each line l we consider the vector $\mathbf{I}_l = \mathbf{I}_{km} = \mathbf{I}_{mk} \in \mathbb{C}^2$ where the first component $(\mathbf{I}_l)_1$ corresponds to I_{km}, that is the power injected from the node to which the transformer is closer, into the line l, while the second component $(\mathbf{I}_l)_2$ corresponds to I_{mk}. The relationship between current and voltage for the two antiparallel arcs is expressed by a vectorial equation in two components.

Definition 3.0.4 (Admittance matrix). *The admittance matrix of the line l, adjacent to the buses k, m is the matrix $\mathbf{Y} \in \mathbb{C}^{2 \times 2}$ such that*

$$\mathbf{I}_{km} = \mathbf{Y}_{km} \mathbf{V}_{km} \tag{3.4}$$

where $\mathbf{V} \in \mathbb{C}^2$ is the vector having as first component V_k and second component V_m.

By applying (3.2) we obtain the equivalent formulation of (3.4) for power:

$$\mathbf{S}_{km} = \operatorname{diag} V_{km} (\mathbf{Y}_{km} \mathbf{V}_{km})^*. \tag{3.5}$$

The structure of $\mathbf{Y}_{km}$ depends upon various physical characteristics of the line and the transformer, a general formulation of the admittance matrix is the following:

$$\mathbf{Y}_{km} = \begin{pmatrix} (\frac{1}{r_{km}+ix_{km}} + i\frac{b_{km}}{2})/\tau_{km}^2 & -\frac{1}{(r_{km}+ix_{km})\tau e^{-i\nu_{km}}} \\ -\frac{1}{(r_{km}+ix_{km})\tau e^{i\nu_{km}}} & \frac{1}{r_{km}+ix_{km}} + i\frac{b_{km}}{2} \end{pmatrix} \tag{3.6}$$

Where $N_{km} = \tau_{km} e^{i\nu_{km}}$ is the ratio of the AC transformer installed on the line $(k, m) \in \mathbf{L}$. If there is no transformer on the line, then $N_{km} = 1$. The series impedance of the line, which can be seen as analogous to the resistance in a direct current circuit, is $y_{km} = r_{km} + ix_{km}$. Lastly, the term b_{km} is the line charging susceptance due to the interaction of the line with the ground. From this formulation, it is apparent that $\mathbf{Y}_{km}$ is symmetric whenever there is no transformer installed on the line. We note that while $I_{km} \neq I_{mk}$ in general, we have that $\mathbf{I}_l = \mathbf{I}_{km} = \mathbf{I}_{mk}$. Whenever a variable does not depend on the orientation of the arc, we will only consider L_0 for that variable, which is analogous to considering the undirected graph associated with the network. Furthermore, $I_{km} = -I_{mk}$ whenever the line charging susceptance b_{km} is zero. In this case, the alternating current of the line behaves like direct current.

3.1 Polar formulation of the ACOPF problem

The objective function is a real-valued polynomial of degree d of the power generated at $g \in \mathcal{G}$ with coefficients $c_0, \dots, c_d$. We consider the following parameters:

- At each arc $(k, m) \in L_0$, the voltage phase difference $\theta_{km} := \delta_k - \delta_m$ is constrained to lie in a given real range $[\theta_{km}^{min}, \theta_{km}^{max}]$.

- A chosen bus index $r \in \mathbf{B}$ is designated as "reference", which entails having voltage phase $\delta_r = 0$.

- The power load at bus $k \in \mathbf{B}$ is denoted by $S_k^L = P_k^L + iQ_k^L \in \mathbb{C}$ where P_k^L, Q_k^L are respectively the real and reactive power demand.

- $\mathcal{G}_k$ is the set of generators installed at bus $l \in \mathbf{B}$. The real and reactive power generated at $g \in \mathcal{G}_k$, $S_g^G = P_g^G + Q_g^G$ are constrained to be respectively in the intervals $[P_k^{min}, P_k^{max}]$ and $[Q_K^{min}, Q_K^{max}]$.

- At each arc $(k, m) \in \mathbf{L}$ the power flow magnitude $|S_{km}|^2$ may have a maximum carrying capacity $U_{km} \in \mathbb{R}_+$.

Equations (3.1) and (3.2) hold whenever there are neither loads, nor generators at bus k. The general power flow equations can be written as:

$$\sum_{(k,m)\in L} S_{ba} + S_b^L = -A_k^* |V_k|^2 + \sum_{g \in \mathcal{G}} S_g^G \tag{3.7}$$

where $-A_k^* |V_k|^2$ represents the shunt admittance due to the leakage of current from the branch to the reference node k.

Considering the cartesian representation of the admittance matrix $\mathbf{Y} = G + iB$ and including the constraints mentioned in this section we obtain the following ACOPF formulation denoted as *ACOPF Polar Formulation*.

$$\inf_{\substack{P_g^G, Q_g^G, \delta_k, \\ |V_k|, S_{km}}} \sum_{g \in \mathcal{G}} F_g(P_g^G, Q_g^G) \tag{3.8}$$

s.t.

Voltage to Power Flow constraints: $\forall km \in \mathbf{L}$

$$S_{km} = (G_{kk} - iB_{kk})|V_k|^2 + (G_{km} - iB_{km})|V_k||V_m| \cdot (\cos(\theta_{km}) + i\sin(\theta_{km})) \tag{3.9}$$

Power balance constraints: $\forall k \in \mathbf{B}$

$$\sum_{km \in L} S_{km} + P_k^L + iQ_k^L = \sum_{g \in \mathcal{G}(k)} P_g^G + i \sum_{g \in \mathcal{G}(k)} Q_g^G \tag{3.10}$$

Power flow, Voltage, and Power generation limits:

$$|S_{km}|^2 \leq U_{km} \tag{3.11}$$

$$V_k^{min^2} \leq |V_k| \leq V_k^{max^2} \tag{3.12}$$

$$P_g^{min} \leq P_g^G \leq P_g^{max} \tag{3.13}$$

$$\theta_{km}^{min} \leq \theta_{km} \leq \theta_{km}^{max}. \tag{3.14}$$

3.1.1 DC optimal power flow

DC optimal power flow (OPF) is a linear approximation of the AC OPF problem, as defined in (3.8). Historically, the DC formulation has been the primary method for solving practical OPF problems. However, it has limitations, especially in the case of heavily loaded transmission lines or large differences in voltage which can lead to significant error [BA09]. In particular it has been shown in [Bak21] that a DC OPF solution is never

a feasible solution to the AC OPF problem (3.8). These limitations render it unsuitable for studying power grid resilience, where instances stressed systems conditions are particularly important. The name "DC optimal power flow" is derived from the fact that its equations resemble the power flow equations in a direct current (DC) network. However, it's important to note that DC OPF is a modeling approach used for alternating current networks.

The DC optimal power flow is based on the following assumptions.

Assumption 3. *1. All system branch resistances are approximately zero.*

2. *The difference of voltage angles between adjacent buses is small, therefore* $\sin(\theta_{km}) \cong \theta_{km}$ *and* $\cos(\theta_{km}) \cong 1$.

3. *The system bus voltages are approximately equal to 1.0 p.u.*

4. *line charging susceptance equal to zero and the voltage ratios are equal to one.*

Based on assumption (4), the structure of the admittance matrix $\mathbf{Y}$ seen in (3.6) simplifies significantly, allowing us to reformulate Equation 3.9 as follows:

$$\begin{cases} P_{km} = \dfrac{1}{r_{km}^2 + x_{km}^2}[r_{km}(V_k^2 - V_k V_m \cos\theta_{km}) + x_{km}V_k V_m \sin(\theta_{km})] \\ Q_{km} = \dfrac{1}{x_{km}^2 + x_{km}^2}[r_{km}(V_k^2 - V_k V_m \cos\theta_{km}) + r_{km}V_k V_m \sin(\theta_{km})]. \end{cases}$$

Since the resistance r_{km} of each branch in negligible, $r_{km} \cong 0$ we obtain:

$$\begin{cases} P_{km} \cong \dfrac{1}{x_{km}}(V_k V_m \sin(\theta_{km}) \\ Q_{km} \cong \dfrac{1}{x_{km}}(V_k^2 - V_k V_m \cos(\theta_{km})). \end{cases}$$

To conclude, since bus voltage magnitudes are approximated by one per unit everywhere and voltage angle differences are very small we obtain the equation in its final form:

$$P_{km} = \frac{\theta_{km}}{x_{km}}. \tag{3.15}$$

While all the other equations remain the same as in (3.8). The DC-OPF formulation simplifies the original problem significantly by eliminating several variables and constraints, resulting in a set of linear constraints. The linearity of the DC OPF formulation makes it a suitable starting point for obtaining a tractable Stochastic OPF model, as was done in [Guo+19b]. In this thesis, we will pursue a different approach. We will construct our Stochastic OPF model by starting from a relaxation of the polar formulation (3.8). This represents an initial step toward developing a more realistic stochastic model, particularly in a context where accurate modeling of extreme load conditions caused by severe weather events is of increasing importance.

Chapter 4

A bridge between Graph Theory and OPF

We consider a known variant of the Jabr relaxation model [Jab06] that can be derived from the polar ACOPF formulation. Let $\theta_{km} := \delta_k - \delta_m$, we replace $|V_k||V_m|\cos(\theta_{km})$ and $|V_k||V_m|\sin(\theta_{km})$ with two new variables c_{km} and s_{km}, respectively:

$$\inf_{\substack{P_g^G,Q_g^G,c_{km}, \\ s_{km},S_{km},P_{km},Q_{km}}} \sum_{g\in\mathcal{G}} F_g(P_g^G) \tag{4.1}$$

Subject to: $\forall km \in \mathbf{L}$

$$c_{km}^2 + s_{mk}^2 = c_{kk}c_{mm} \tag{4.2}$$

$$P_{km} = G_{kk}c_{kk} + G_{km}c_{km} + B_{km}s_{km} \tag{4.3}$$

$$Q_{km} = -B_{kk}c_{kk} - B_{km}c_{km} + G_{km}s_{km} \tag{4.4}$$

$$S_{km} = P_{km} + jQ_{km} \tag{4.5}$$

Power balance constraints: $\forall k \in \mathbf{B}$

$$\sum_{km\in L} S_{km} + P_k^L + iQ_k^L = \sum_{g\in\mathcal{G}(k)} P_g^G + i\sum_{g\in\mathcal{G}(k)} Q_g^G \tag{4.6}$$

Power flow, Voltage, and Power generation limits:

$$P_{km}^2 + Q_{km}^2 \leq U_{km} \tag{4.7}$$

$$V_k^{\min^2} \leq c_{kk} \leq V_k^{\max^2} \tag{4.8}$$

$$P_g^{\min} \leq P_g^G \leq P_g^{\max} \tag{4.9}$$

$$c_{kk} \geq 0 \tag{4.10}$$

$$c_{km} = c_{mk}, \ s_{km} = -s_{mk}. \tag{4.11}$$

Where since $|V_k|^2 = |V_k|^2\cos(\theta_{kk})$, we replaced $|V_k|^2$ with c_{kk}, and similarly, $s_{kk} = 0$. Since $|V_k^2| \geq 0$, $\cos(\theta_{km}) = \cos(\theta_{mk})$ and $\sin(\theta_{km}) = -\sin(\theta_{mk})$ we

also added the constraints: $c_{kk} \geq 0$, $c_{km} = c_{mk}$ and $s_{km} = -s_{mk}$. Lastly, considering that $(|V_k||V_k|\cos(\theta_{km}))^2 + (|V_k||V_m|\sin(\delta_k - \delta_m))^2 = |V_k|^2|V_m|^2$, we introduced the following constraints: $c_{km}^2 + c_{mk}^2 = c_{kk}c_{mm}$. We refer to this relaxation as *Jabr equality ACOPF relaxation*. We note that since S_{km}, P_{km}, Q_{km} are dummy variables of the problem we will only refer to $P_g^G, Q_g^G, s_{km}, c_{km}$ when talking about a solution of (4.1). Analogously we say $P_g^G, Q_g^G, \sigma_k, |V_k|$ when referring to solutions of (3.8). This relaxation is in general non exact but in [Jab06] the corresponding feasibility problem was shown to be exact when the graph is a tree (multisource radial network). Here we show that the corresponding optimization problem is also exact for multisource radial networks.

Proposition 4.0.1. *If* $(\mathbf{B}, \mathbf{L})$ *is a multisource radial network, then the Jabr equality ACOPF relaxation* (4.1) *is exact.*

Proof. Let $P_g^G, Q_g^G, c_{km}, s_{km}$ be an optimal solution to the Jabr equality ACOPF relaxation (4.1). Let $|V_k| := \sqrt{c_{kk}}$. Since $\frac{c_{km}^2}{c_{kk}c_{mm}} + \frac{s_{km}^2}{c_{kk}c_{mm}} = 1$ then there exist $\theta_{k,m}$ such that $\frac{c_{km}^2}{c_{kk}c_{mm}} = \cos(\theta_{km})$ and $\frac{s_{km}^2}{c_{kk}c_{mm}} = \sin(\theta_{km})$. For the reference bus $r \in \mathbf{B}$ we have $\delta_r = 0 \mod 2\pi$. We need to define the voltage angle δ_k for every $k \in \mathbf{B}$ such that

$$\theta_{km} = \delta_k - \delta_m \mod 2\pi \quad \forall km \in \mathbf{L}. \tag{4.12}$$

For each $k \in \mathbf{B}$ let $(k_1, ..., k_n)$ be a path in the graph where $k_1 = r$ and $k_n = k$. We define $\delta_h := \sum_{i=1}^{i=n-1} \theta_{k_{i+1}k_i} \mod 2\pi$. The various voltage angles are well defined because there are no loops in the graph and therefore there exist only one path connecting k and h. Furthermore, with this definition, (4.12) holds by construction. This is a feasible solution for the ACOPF problem in polar form (3.8) because the magnitude constraints (4.7), (4.8), (4.9) and the Power Balance constrains (4.6), (4.5) remain the same. While, since with this definition of $|V_k|$ and δ_k we have that $c_{km} = |V_K||V_m|\cos\theta_{km}$ and $s_{km} = |V_K||V_m|\sin\theta_{km}$, then equations (4.3) and (4.4) imply that the Power Flow equation in the Polar ACOPF model (3.9) holds. Since the Jabr equality ACOPF (4.1) is a relaxation of the Polar ACOPF model (3.8), $|V_k|$ and δ_k are also an optimal solution of (3.8). $\qquad\square$

As done in the proof, given a solution to the Jabr relaxation we can always recover $|V_k|, \theta_{km}$ such that: $c_{kk} = |V_k|^2$, $c_{km} = |V_k||V_m|\cos\theta_{km}$ and $s_{km} = |V_k||V_m|\sin\theta_{km}$. In general, the model is exact whenever at each node $k \in \mathbf{B}$ we can define δ_k such that for all $(k, m) = l \in \mathbf{L}$ we have: $\theta_{km} = \delta_k - \delta_m \mod 2\pi$. This is possible whenever the following loop condition holds for every loop $(k_1, ..., k_n, k_{n+1} = k_1)$ in the graph:

$$\sum_{i=1}^n \theta_{k_{i+1}k_i} = 0 \mod 2\pi. \tag{4.13}$$

Let δ_h be defined as in the proof of Proposition 4.0.1. Then, if $(k_1 = r, ..., k_n = k)$ and $(h_1 = r, ..., h_n = k)$ are two paths from r to k, we have $\delta_h = \sum_{i=1}^{i=n-1} \theta_{k_{i+1}k_i} = \sum_{i=1}^{i=n-1} \theta_{h_{i+1}h_i} \mod 2\pi$ for the loop condition, as illustrated in Fig 4.1. So δ_h is well defined and we obtain an optimal feasible solution for the Polar ACOPF model 3.8.
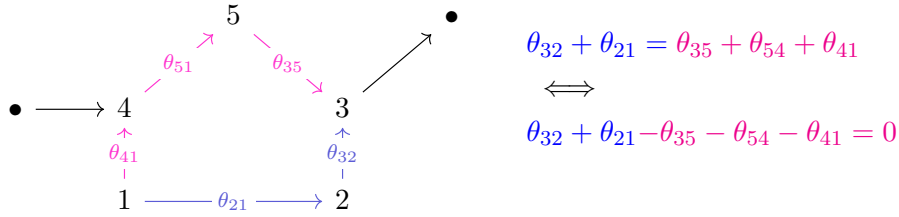

Figure 4.1: Representation of the loop condition (4.13)

We summarize this in the following observation

Observation 4.0.2. *Let $(c_{km}, s_{km}, P_g^G, Q_g^G)$ be an optimal feasible solution for the Jabr equality ACOPF relaxation (4.1) and $\theta_{ij}, |V_i|, \delta_i$ for all $\{i, j\} \in \mathbf{L}$ be as in the proof of Proposition 4.0.1. If for every loop $(k_1, ..., k_n, k_1)$ in the graph we have that*

$$\sum_{i=1}^{n} \theta_{k_{i+1}k_i} = 0 \mod 2\pi \tag{4.14}$$

holds, where $k_{n+1} = k_1$. Then, δ_k is well defined for every $k \in \mathbf{B}$ and $(|V_k|, \delta_k, P_g^G, Q_g^G)$ is an optimal feasible solution for the OPF problem 3.8.

We can check that this condition holds directly on the variables k_{km} and s_{km} since $\sum_{i=1}^{n} \theta_{k_{i+1}k_i} = 0 \mod 2\pi \iff \cos(\theta_{k_1k_n} + \sum_{i=1}^{n-1} \theta_{k_{i+1}k_i}) = 1$ and thanks to the following generalized sum-to-product formula.

Proposition 4.0.3 (Generalized sum-to-product formula)**.** *The following is true:*

$$\sin\left(\sum_{i=1}^{n} \theta_i\right) = \sum_{k=0}^{\lfloor n/2 \rfloor} \sum_{\substack{A \subset [n] \\ |A|=2k+1}} (-1)^k \prod_{h \in A} \sin(\theta_h) \prod_{h \in A^c} \cos(\theta_h), \tag{4.15}$$

$$\cos\left(\sum_{i=1}^{n} \theta_i\right) = \sum_{k=0}^{\lfloor n/2 \rfloor} \sum_{\substack{A \subset [n] \\ |A|=2k}} (-1)^k \prod_{h \in A} \sin(\theta_h) \prod_{h \in A^c} \cos(\theta_h). \tag{4.16}$$

Proof. Let $\Im(x)$ and $\Re$ denote the imaginary and real part of $x \in \mathbb{C}$. Using Euler's identity and since $\Im(i^k)$ is zero when k is even, we obtain:

$$\sin\left(\sum_{j=1}^{n} \theta_j\right) = \Im(\exp(i \sum_{j=1}^{n} \theta_j))$$

$$= \Im(\prod_{j=1}^{n} \exp(i\theta_j))$$

$$= \Im(\prod_{j=1}^{n} (\cos(\theta_j) + i\sin(\theta_j)))$$

$$= \Im(\sum_{A \subset [n]} \prod_{h \in A} (i\sin(\theta_h)) \prod_{h \in A^c} (\cos(\theta_h)))$$

$$= \Im(\sum_{A \subset [n]} i^{|A|} \prod_{h \in A} (\sin(\theta_h)) \prod_{h \in A^c} (\cos(\theta_h)))$$

$$= \sum_{k=0}^{\lfloor n/2 \rfloor} \sum_{\substack{A \subset [n] \\ |A|=2k+1}} (-1)^k \prod_{h \in A} \sin(\theta_h) \prod_{h \in A^c} \cos(\theta_h).$$

$\square$

We note that for the constraint (3.12), the voltage $|V_k|$ is non zero at each bus $k \in \mathbf{B}$. By multiplying (4.16), applied to $\sum_{i=1}^{n} \theta_{k_{i+1}k_i}$, by $\prod_{h=1}^{n} |V_{k_h}|^2$ and using the definition of the variables c and s we obtain:

Observation 4.0.4 (Loop constraint). *Let $(k_0, ..., k_n, k_1)$ be a loop in the graph. Then $\sum_{i=1}^{n} \theta_{k_{i+1}k_i} = 0 \mod 2\pi$, where $n + 1 = 1$, if the voltages $|V_i| \neq 0$ for each bus and the following equation holds:*

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \sum_{\substack{A \subset [n] \\ |A|=2k}} (-1)^k \prod_{h \in A} s_{k_h k_{h+1}} \prod_{h \in A^c} c_{k_h k_{h+1}} = \prod_{k=1}^{n} c_{k_i, k_i}. \qquad (4.17)$$

For the Observation 4.0.2, if the loop constraint holds for every loop in the network, then an optimal solution to the Jabr equality Equation (4.1) is an optimal solution to the OPF problem Equation (3.8). However, in real-life applications, the number of loops in a graph can be very large for large networks, making this result impractical. Therefore, in the next part of this chapter, we aim to restrict the number of loops on which we enforce the loop constraints in order to achieve exactness.

When considering an eventual relaxations to the loop constraints it may be useful to consider the parametrization of the loop constraints solution space alone.

Proposition 4.0.5. *Let* $(\mathbf{B}, \mathbf{L})$ *be a network. If the loop constraint 4.0.4 holds for every element of a cycle basis, then it holds for every loop in the network. In particular the number of loop constraints corresponds to the cyclomatic number* $|\mathbf{L}| - |\mathbf{B}| + 1$.

Proof. Fix an orientation for $\mathbf{L}$, which we will still denote as $\mathbf{L} \subset \mathbf{B} \times \mathbf{B}$ for simplicity. We can now consider the vector $\underline{\theta} \in \mathbb{R}^{\mathbf{L}}$ where for every $e = (h, k) \in \mathbf{L}$, $\underline{\theta}(e) \coloneqq \theta_{kh} = -\theta_{hk}$. Given a loop $L = (k_1, ..., k_{n-1}, k_n = k_0)$, we define the loop constraint vector associated with L as $v_L \in \mathbb{R}^{\mathbf{L}}$, where we define $v_L(e) \coloneqq 0$ if e is not an edge in the path L. For every $e \in \mathbf{L}$ we have $v_L(e) \coloneqq \theta_{k_{i+1}k_i}$ if there exists an i such that $(k_i, k_{i+1}) = e$ or $(k_{i+1}, k_i) = e$. We note that in the first case $v_L(e) = \underline{\theta}_e$ while in the latter (that is, if e runs contrary to the cycle) $v_L(e) = -\underline{\theta}_e$. Therefore multiplying the loop vector u_L associated to L as in Definition 1.1.4 in Chapter 1.1 by $\bar{\theta}$ component wise, we obtain v_L, that is: $v_L = D u_L$ where $D = \mathrm{diag}(\underline{\theta})$ is the diagonal matrix having the components of $\bar{\theta}$ as elements of the diagonal. This means that D maps the Cycle Space onto the loop constraint vectors.

The loop constraint (4.0.4) is equivalent to the conditions: $\mathbb{1}' v_L = 0$ mod 2π for every loop L. Let $\mathcal{B} = \{b_1, \ldots, b_k\}$ be the cycle basis over which the loop constraint holds. For Theorem 1.1.6 there exists $\lambda_i \in \mathbb{R}, i = 1, \ldots, k$, such that $u_L = \sum_{i=1}^{k} \lambda_i b_i$. Therefore, $\mathbb{1}' v_L = \sum_{i=1}^{k} \mathbb{1}' D b_i = 0$ mod 2π, $\mathbb{1}' D b_i = 0$ mod 2π because the loop constraint holds for every cycle of the cycle basis by hypothesis and $D b_i$ is the loop constraint vector associated to the cycle defining b_i. $\square$

As shown in Chapter 1.1, one approach to construct a Cycle Basis is to start with a spanning tree $T = (\mathbf{B}, E) \subset (\mathbf{B}.\mathbf{L})$. By adding an arc in $\mathbf{B} \setminus E$, we can generate an element for the Cycle Basis. Repeating this process for all arcs in $\mathbf{B} \setminus E$, we can obtain all the cycles, which together form a Cycle Basis. This method also provides a way to obtain all admissible solutions of the loop constraints, parameterized by the values of c_{hk} and s_{hk} in $\mathbb{R}$, along with an additional parameter λ_{hk} for all $hk \in \mathbf{B} \setminus E$.

Proposition 4.0.6. *Let* $(\mathbf{B}, \mathbf{L})$ *be a network and* $T = (\mathbf{B}, E) \subset (\mathbf{B}, \mathbf{L})$ *a spanning tree inducing the cycle basis* $\mathcal{B}$. *Let* $N \coloneqq \mathbf{L} \setminus E$. *Then by fixing* $c_h, s_h, c_{kk} \in \mathbb{R}$ *for every* $h \in E$ *and* $k \in \mathbf{B}$, *satisying the Jabr equality, there exists unique* $s_e, c_e \in \mathbb{R}$ *for all* $e \in N$ *such that* (s_h, c_h, c_{kk}) *for all* $e \in \mathbf{L}, k \in \mathbf{B}$ *is a solution to the loop constraints and to the Jabr equality constraints. Furthermore the values of* c_e, s_e *for* $e \in N$ *are given by the following relation:*

$$c_e + is_e = \frac{\prod_{k \in nodes(C_e)} c_{kk}}{\prod_{l \in edges(C_e) \setminus \{e\}} c_l + is_l}. \tag{4.18}$$

Proof. We note that the Jabr constraint is satified for ever $(k, h) \in E$ if and only if $\mathcal{I}(kh) \coloneqq c_{kh} + is_{kh} = \sqrt{c_{kk}c_{hh}} e^{i\theta_{kh}}$ for some $\theta_{kh} \in \mathbb{R}/2\pi\mathbb{R}$. For

every $e \in N$ we search for $\mathcal{I}(e) \in \mathbb{C}$ such that the loop and Jabr constraints hold. With this notation the loop constraint $\sum_{l \in edges(C_e)} \theta_l = 0 \mod 2\pi$ for loop C_e is equivalent to: $\prod_{l \in edges(C_e)} \mathcal{I}(l) = \prod_{k \in nodes(C_e)} c_{kk}$. Thus $\mathcal{I}(e) := \frac{\prod_{k \in nodes(C_e)} c_{kk}}{\prod_{l \in edges(C_e) \backslash \{e\}} \mathcal{I}(l)}$. By construction of the cycle basis $\mathcal{C}$ for a fixed $e \in N$, $\mathcal{I}(e)$ appears only in the loop constraint associated to C_e. Thus all the loops constraints hold. We need to check that the Jabr constraints hold. That is that $c_e^2 + s_e^2 = |\mathcal{I}(e)| = c_{hh}c_{kk}$ for h, k such that $(h, k) = e$. $|\mathcal{I}(e)| = \frac{\prod_{k \in nodes(C_e)} c_{kk}}{\prod_{l \in edges(C_e) \backslash \{e\}} |\mathcal{I}(l)|} = \frac{\prod_{k \in nodes(C_e)} c_{kk}}{\prod_{l \in edges(C_e) \backslash \{e\}} \sqrt{c_{l_1 l_1} c_{l_2 l_2}}} = \sqrt{c_{hh}c_{kk}}$. Where $l = (l_1, l_1)$ for every $l \in C_e$. □

We conclude with a note regarding the parametrization of the solution space of the loop constraints alone, which could prove valuable when contemplating a potential relaxation of these constraints in the future.

Proposition 4.0.7. *Let* $(\mathbf{B}, \mathbf{L})$ *be a network and* $T = (\mathbf{B}, E) \subset (\mathbf{B}, \mathbf{L})$ *a spanning tree inducing the Cycle Basis* $\mathcal{B}$. *Let* $N := \mathbf{L} \setminus E$ *Then fixing* (c_h, s_h, c_{kk}) *for every* $h \in E$ *ad* $k \in \mathbf{B}$ *and:*

$$\binom{s_e}{c_e} = \frac{\prod_{i=1}^{n_e} c_{k_i^e k_i^e}}{\mathscr{S}_e^2 + \mathscr{C}_e^2} \binom{\mathscr{S}_e}{\mathscr{C}_e} + \lambda_e \binom{-\mathscr{C}_e}{\mathscr{S}_e} \quad \forall e \in N. \qquad (4.19)$$

Where

$$\mathscr{S}_e = \sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e) \backslash \{e\} \\ |A| = 2k-1}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h \qquad (4.20)$$

and

$$\mathscr{C}_e = \sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e) \backslash \{e\} \\ |A| = 2k}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h. \qquad (4.21)$$

Then with this definition (s_h, c_h, c_{kk}) $\forall e \in \mathbf{L}$ $\forall k \in \mathbf{B}$ *is a solution to the loop constraint and all the solutions to the loop constraint equations are in this form.*

Proof. Let $N := \mathbf{L} \setminus E$. We search for the solutions, satisfying the constraints of the ACOPF Jabr relaxation 4.1, of the system of equations:

$$\sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e) \\ |A| = 2k}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h = \prod_{i=1}^{n_b} c_{k_{i+1}^e k_{i+1}^e} \quad \forall e \in N \qquad (4.22)$$

Where $[n] = \{1, 2, \ldots, n\}$ and for every $e \in N$, $C_e = (k_1^e, \ldots, k_{n_e}^e) \in \mathcal{B}$ is the only cycle contained in $(\mathbf{B}, E \cup \{e\})$. And by $edges(C_e)$ we denoted the oriented set of edges defining C_e. We note that that each pair (c_e, s_e)

appears exactly once in the system of equation for every $e \in N$, and that fixing the other variables (c_h, s_h) for all $h \in E$ and c_{kk} for all $k \in \mathbf{B}$, the system of equations becomes an independent system of equation in the variable (c_e, s_e) $\forall e \in N$ in particular equation (4.22) is equal to:

$$\prod_{i=1}^{n_b} c_{k_{i+1}^e k_{i+1}^e} = \sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e) \\ |A|=2k}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h = \tag{4.23}$$

$$\left(\sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e)\backslash\{e\} \\ |A|=2k-1}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h \right) s_e + \tag{4.24}$$

$$\left(\sum_{k=0}^{\lfloor n_e/2 \rfloor} \sum_{\substack{A \subset edges(C_e)\backslash\{e\} \\ |A|=2k}} (-1)^k \prod_{h \in A} s_h \prod_{h \in A^c} c_h \right) c_e \tag{4.25}$$

Denoting the quantities in round brackets multiplying s_e and c_e by $\mathscr{S}_e$ and $\mathscr{C}_e$ respectively we have:

$$\mathscr{S}_e s_e + \mathscr{C}_e c_e = \prod_{i=1}^{n_e} c_{k_i^e k_i^e}. \tag{4.26}$$

Thus, since the equations are independent, the solutions to the system of equations 4.22 are in the form:

$$\begin{pmatrix} s_e \\ c_e \end{pmatrix} = \begin{pmatrix} \bar{s}_e \\ \bar{c}_e \end{pmatrix} + \lambda_e \begin{pmatrix} -\mathscr{C}_e \\ \mathscr{S}_e \end{pmatrix} \tag{4.27}$$

Where $\begin{pmatrix} \bar{s}_e \\ \bar{c}_e \end{pmatrix}$ is a particular solution to (4.26). In particular since we have that $(\mathscr{S}, \mathscr{C}) \neq 0$ we can take $\begin{pmatrix} \bar{s}_e \\ \bar{c}_e \end{pmatrix} = \frac{\prod_{i=1}^{n_e} c_{k_i^e k_i^e}}{\mathscr{S}_e^2 + \mathscr{C}_e^2} \begin{pmatrix} \mathscr{S}_e \\ \mathscr{C}_e \end{pmatrix}$. $\qquad \square$

Chapter 5

A Stochastic Distributionally Robust Jabr OPF model

In the previous chapters, the ACOPF formulations that we considered were deterministic, and are generally solved in real-time scenarios. In day-ahead models, they need to take into account errors in the prediction of loads and of power production of renewable energy sources. The distribution of the error of predictions can vary greatly depending upon the time of the year. Consequently, when determining the distribution of predictive errors, it is practical to use data that is temporally proximate, often resulting in a limited dataset. Given this constraint, it is crucial for the adopted policy to showcase robust out-of-sample performances. This is translated into trying to find the policy which minimizes the cost over the worst case of plausible prediction errors. In this chapter following [Guo+19b]. we first formulate a general stochastic OPF problem as a distributionally robust stochastic optimal control problem. We then construct a version of this model starting from a relaxed version of the Jabr OPF model (4.1) and then apply the finite reduction results reviewed in Chapter 2. We conclude this chapter with the numerical results of our implementation applied to a variant of the IEEE 118-bus test system.

A general Distributionally Robust OPF formulation

The DRO OPF formulation we consider will be in the following form. Let $x_t \in \mathbb{R}^n$ denote a state vector at time t that includes the internal states of all devices in the network, and let $\xi_t \in \mathbb{R}^{N_\xi}$ denote a random vector in a probability space $(\Omega, \mathcal{F}, \mathbb{P}_t)$ that includes forecast errors of all uncertainties in the network.

$$\inf_{\pi \in \Pi} \sup_{\mathbb{P} \in \mathcal{P}} \ \mathbb{E}^{\mathbb{P}} \sum_{t=0}^{T} h_t(x_t, \xi_t) \tag{5.1}$$

Subject to:

$$x_t = \pi(x_0, ..., x_{t-1}, \xi_t, \mathcal{D}_t) \tag{5.2}$$

$$x_t \in \mathcal{X}_t. \tag{5.3}$$

The set $\mathcal{X}_t$ denotes the set of feasible state of the network while Π is the set of admissable policies. Some constraints may be modeled deterministically as contraints on set Π and others may be included as risk terms in the objective function.

5.1 System model

We now aim to build a model satistying the *Convexity Assumption* (Assumption 2) in order to apply Kuhn's finite reduction result (Theorem 2.0.2). Since many of the constraints in the OPF problem (3.8) are not concave, we first consider the following concave *Jabr like* relaxation, where by concave constraint we mean a constraint in the form $f(x) \geq 0$ with f a convex function.

$$\inf_{\substack{P_g^G, Q_g^G, c_{km}, \\ s_{km}, S_{km}}} \sum_{g \in \mathcal{G}} F_g(P_g^G) \tag{5.4}$$

Subject to:

$$c_{km}^2 + s_{km}^2 \geq c_{kk} c_{mm} \tag{5.5}$$

$$P_{km} + P_{km} \geq 0 \tag{5.6}$$

$$P_{km} = G_{kk} c_{kk} + G_{km} c_{km} + B_{km} s_{km} \tag{5.7}$$

$$Q_{km} = -B_{kk} c_{kk} - B_{km} c_{km} + G_{km} s_{km} \tag{5.8}$$

$$S_{km} = P_{km} + j Q_{km} \tag{5.9}$$

Power balance constraints:

$$\sum_{km \in L} S_{km} + P_k^L + i Q_k^L = \sum_{g \in \mathcal{G}(k)} P_g^G + i \sum_{g \in \mathcal{G}(k)} Q_g^G \tag{5.10}$$

Power flow, Voltage, and Power generation limits:

$$P_{km}^2 + S_{km}^2 \leq U_{km} \tag{5.11}$$

$$V_k^{\min^2} \leq |c_{kk}| \leq V_k^{\max^2} \tag{5.12}$$

$$P_g^{\min} \leq P_g^G \leq P_g^{\max} \tag{5.13}$$

$$Q_g^{\min} \leq Q_g^G \leq Q_g^{\max}. \tag{5.14}$$

5.1.1 Parameters and decision variables

For each variable in the model, we introduce an additional variable for each time step: $c_{tkm}, s_{tkm}, P_{tk}^G, Q_{tk}^G, S_{tkm}$. We use the term "stochastic variables" to refer to the variables in the model that depend on the prediction errors ξ_t. These variables can be classified into two categories: those that directly depend on the prediction errors and those that depend indirectly through the policy π.

The first category includes variables like the power generation S_k^t at each renewable generator $k \in \mathcal{G}_r$, which are directly influenced by the uncertainty introduced by the prediction errors.

The second category comprises variables related to traditional generators and the values of c and s. These variables are influenced by the policy π, which acts as a correction based on the observed prediction errors. The policy adjusts the planned energy production of the network to account for the uncertainty introduced by the prediction errors.

By incorporating both direct and indirect dependencies on the prediction errors, the stochastic variables provide a comprehensive representation of the system's behavior under uncertainty.

1. **Renewable generators.** For each renewable generator $k \in \mathbb{R}$, the generated power is decomposed as $P_{tk}^G = P_{tk}^{Gn} + P_{tk}^{Ge}(\xi_t)$ and $Q_{tk}^G = Q_{tk}^{Gn} + Q_{tk}^{Ge}(\xi_t)$, where P_{tk}^{Gn}, Q_{tk}^{Gn} represent the predicted power generation at bus t and bus k, and $P_{tk}^{Ge}(\xi_t), Q_{tk}^{Ge}(\xi_t)$ are stochastic variables that correspond to the coordinates of the random vector ξ_t representing the error predictions at time t.

2. **Traditional generators.** Let $P_t^G, Q_t^G \in \mathbb{R}^{\mathcal{CG}}$ be the vectors representing the real and reactive power generation at time t for each traditional generator. We model the time interdependencies through affine policies as follows: $P_{tk}^G = P_{tk}^{Ge}\xi_{\mathbf{t}} + P_t^{Gn}$, where $P_t^{Gn} \in \mathbb{R}^t$ can be interpreted as a nominal schedule and $P_t^{Ge}k \in \mathbb{R}^{N_\xi}$ represents the reserve policy.

3. **c and s variables.** The variables c_{tkm} are defined as follows. $c_{tkm} = c_{tkm}^e\xi_t + c_{tkm}^n$ for all $t = 1, \ldots, T$ and $\{k, m\} \in \mathbf{L}$ if $k \neq m$ or $k = m \in \mathbf{B}$. $e_{tkm}^n \in \mathbb{R}$ represents the nominal schedule and $c_{tkm}^e \in \mathbb{R}^{N_\xi}$ represents the adjustment policy. Analogously we define s_{tkm} as $s_{tkm} := s_{tkm}^e\xi_t + s_{tkm}^n$.

The affine policy function is represented by the list:

$$\pi = (P_h^{G_n}, P_h^{G_e}, Q_h^{G_n}, c_{tkm}^e, c_{tkm}^n, s_{tkm}^e, s_{tkm}^n)$$

for all $t = 1, \ldots, T$, $h \in \mathcal{G}_t$, and $k, m \in \mathbf{L}$ if $k \neq m$ or $k = m \in \mathbf{B}$.

5.1.2 Contraints definition

All the constraints in (5.4) must hold for each time $t = 1, \ldots, T$. Thus, considering time dependency and combining (5.7), (5.8), (5.10), (5.9), we have:

$$
\begin{aligned}
P_{tk}^{G_n} + P_{tk}^{G_e} \xi_t - P_{tk}^d &= \sum_{km \in \delta(k)} P_{tkm} \\
&= \sum_{km \in \delta(k)} (G_{kk} c_{tkk} + G_{km} c_{tkm} + B_{km} s_{tkm}) \\
&= m_k^P \begin{pmatrix} c_t \\ s_t \end{pmatrix},
\end{aligned}
\tag{5.15}
$$

where $c_t \in \mathbb{R}^{\mathbf{B} \sqcup \mathbf{L}}$ is the vector with coordinates $(c_t)_k := c_{tkk} \; \forall k \in \mathbf{B}$ and $(c_t)_{kh} := c_{tkh} \; \forall kh \in \mathbf{L}$, $s_t \in \mathbb{R}^{\mathbf{L}}$ is the vector with coordinates $(s_t)_{kh} := s_{tkh} \; \forall kh \in \mathbf{L}$, and $m_k^P \in \mathbb{R}^{\mathbf{B} \times \mathbf{L} \times \mathbf{L}'}$ with $L' := L$, is the vector inducing the second equation with coordinates for all $h \in \mathbf{B}$, $(m_k^P)_h = G_{hh}$, for all $kh \in \delta(k) \subset \mathbf{L}$, $(m_k^P)_{kh} = G_{km}$, and for all $kh \in \delta(k) \subset \mathbf{L}'$, $(m_k^P)_{kh} = B_{kh}$. Since $c_t = c_t^n + c_t^e \xi_t$, where $c_t^e \in \mathbb{R}^{(\mathbf{B} \sqcup \mathbf{L}) \times N_\xi}$ is the matrix with rows $(c_t^e)_h := c_{th}^e \; \forall h \in \mathbf{B}$ and $(c_t^e)_{hk} = c_{thk}^e \; \forall hk \in \mathbf{L}$, and $s_t \in \mathbb{R}^{\mathbf{L} \times N_\xi}$ is analogously defined by stacking $s_{thk}^e \; \forall hk \in \mathbf{L}$. Thus we obtain:

$$
P_{tk}^{G_n} + P_{tk}^{G_e} \xi_t - P_{tk}^d - m_k^P \begin{pmatrix} c_t^n \\ s_t^n \end{pmatrix} - m_k^P \begin{pmatrix} c_t^e \\ s_t^e \end{pmatrix} \xi_t = 0.
$$

Since this constraint must hold for all possible realizations of ξ_t and therefore for all $\xi_t \in \Xi$, we obtain the following equivalent constraints for all $k \in \mathbf{B}$:

$$
P_{tk}^{G_n} - P_{tk}^d - m_k^P \begin{pmatrix} c_t^n \\ s_t^n \end{pmatrix} = 0,
\tag{5.16}
$$

$$
P_{tk}^{G_e} - m_k^P \begin{pmatrix} c_t^e \\ s_t^e \end{pmatrix} = 0.
\tag{5.17}
$$

The first constraint means that the nominal values c_t^n must be a solution of (5.4) substituting the variables of renewable power generation with their prediction at time t, while the second is a constraint on the components $P_{tk}^{G_e}, c_t^e$ and s_t^e of the policy π. We apply the same procedure to the reactive power in the equation (5.10) and thus obtain the following analogous constraints for all $k \in \mathbf{B}$:

$$
Q_{tk}^{G_n} - Q_{tk}^d - m_k^Q \begin{pmatrix} c_t^n \\ s_t^n \end{pmatrix} = 0,
\tag{5.18}
$$

$$
Q_{tk}^{G_e} - m_k^Q \begin{pmatrix} c_t^e \\ s_t^e \end{pmatrix} = 0.
\tag{5.19}
$$

5.1.3 Inequality affine constraints

In [Guo+19a], bounds on the wind power forecast errors, denoted as ξ_τ, were not explicitly set. Specifically the ambiguity set Ξ extended over the entirety of $\mathbb{R}^{N_\xi}$. However, in this section, we consider the potential advantages of viewing the set Ξ as a compact polytope. Given the inherent limits of wind energy production and the consequent bounds on its forecasting error, it is practical to see Ξ as finite. This perspective provides a distinct advantage: the ambiguity set $\mathcal{B}_\varepsilon(\hat{\mathbb{P}}_N)$ excludes implausible extreme distributions. For instance, scenarios where forecast error surpasses the maximal potential output of wind generators are disregarded.

Additionally, when the polytope is defined as $\Xi := \{y \in \mathbb{R}^{N_{xi}} \mid Hy \le d\}$, has its symmetry point at $0 \in \mathbb{R}^{N_{xi}}$, and when the matrix $H \in \mathbb{R}^{2N_{xi} \times N_{xi}}$ is full rank, the affine inequality constraints given in (5.4) can be expressed as constraints on the $\|\cdot\|_1$-norm of the policy π. To see this, consider an affine inequality constraint in the form:

$$A(\pi)\xi + b(\pi) \le M, \quad \forall \xi \in \Xi, \tag{5.20}$$

with the condition that $m \le b(\pi) \le M$. Rewriting (5.20), as

$$A(\pi)\xi \le M - b(\pi), \quad \forall \xi \in \Xi.$$

Let $r := M - b(\pi)$, then (5.20) is satisfied iff $A(\pi)\xi \in B_r(0)$ for all ε in Ξ, given 0 is a symmetry point of Ξ. If H is a $2N_{xi} \times N_{xi}$ full-rank matrix, a linear isomorphism $L : \mathbb{R}^{N_{xi}} \to \mathbb{R}^{N_{xi}}$ exists such that the $\|\cdot\|_\infty$-norm ball maps to Ξ. Hence, $A(\pi)\xi$ belongs to $B_r(0)$ $\forall \xi \in \Xi$ iff $A(\pi)Lx \in B_r(0)$ $\forall x \in B_r^{\|\cdot\|_\infty}(0)$. This holds true only when the function $A(\pi)L$ from $(\mathbb{R}^{N_{xi}}, \|\cdot\|_\infty)$ to $\mathbb{R}$ has a continuity norm $\|\|\|_1$ no greater than r. Thus, constraint (5.20) reduces to:

$$\|A(\pi)L\|_1 \le r = |M - b(\pi)|, \tag{5.21}$$

translating to a set of finite affine constraints on the policy π. Specifically, we apply this to constraints on voltage and power generation magnitudes. In our model we consider the ambiguity set to be in the form $\Xi = [-E_1, E_1] \times [-E_2, E_2] \times [-E_3, E_3]$. In this case L is the linear isomorphism associated to the 3×3 diagonal matrix having as entries E_1, E_2, E_3. We consider the constraint on the power generation at the generator $g \in \mathcal{G}$:

$$P_g^{\min} \le P_g^G = P_g^{Gn} + P_g^{Ge}\xi \le P_g^{\max} \quad \forall \xi \in \Xi,$$

which is a constraint in the form (5.20) and thus we obtain the following equivalent constraints:

$$\|P_g^{Ge}L\|_1 \le P_g^{Gn} - P_g^{\min}, \tag{5.22}$$

$$\|P_g^{Ge}L\|_1 \le P_g^{\max} - P_g^{Gn}. \tag{5.23}$$

Similarly, for the voltage magnitude constraints in MW, where $\bar{b} \in \mathbf{B}$ is the reference bus, we have for each bus $b \in \mathbf{B}$:

$$V_b^{\min^2} c_{\overline{bb}} \leq c_{bb} \leq V_b^{\max^2} c_{\overline{bb}},$$

thus

$$V_b^{\min^2}(c_{\overline{bb}}^n + c_{\overline{bb}}^n \xi) \leq c_{bb}^n + c_{bb}^e \xi \leq V_b^{\max^2}(c_{\overline{bb}}^n + c_{\overline{bb}}^n \xi) \quad \forall \xi \in \Xi,$$

which is equivalent to:

$$\|(c_{bb}^e - V_b^{\max^2} c_{\overline{bb}}^e)L\|_1 \leq V_b^{\max^2} c_{\overline{bb}}^n - c_{bb}^n, \tag{5.24}$$

$$\|(c_{bb}^e - V_b^{\min^2} c_{\overline{bb}}^e)L\|_1 \leq c_{bb}^n - V_b^{\max^2} c_{\overline{bb}}^n. \tag{5.25}$$

We observe from these equations that policies lean towards conservative values when b (the nominal value) approaches the constraint boundary, ensuring non-violation, while adopting a larger continuity norm otherwise. We note that a similar reasoning can be easily done with the $\|\cdot\|_p$-norm of the policy, with $p \in [1, +\infty]$, instead of the $\|\cdot\|_1$-norm.

5.1.4 Conditional Value at Risk

We will consider constraint violations of nonlinear constraints in terms of a loss function $C(\pi, \xi)$. We can treat $C(\pi, \xi)$ as a random variable. If we designate α as an acceptable loss, $\mathbb{P}(C(\pi, \xi) > \alpha)$ represents the probability of violating the acceptable loss α with the policy π. Viceversa we can also consider, given a risk level $\beta \in [0, 1]$, the Value at Risk (VaR) $\alpha_\beta(\pi) = \inf\{\alpha \in \mathbb{R} \mid \mathbb{P}(C(\pi, \xi) > \alpha) \leq \beta\}$ which corresponds to the minimum loss α with a probability smaller than β of being exceeded. The Conditional Value at Risk with risk level β is defined $\phi_\beta(\pi) = \frac{1}{\beta} \int_{C(\pi, \xi) \geq \alpha_\beta(x)} f(\pi, \xi) d\mathbb{P}$ and represents the average loss exceeding $\alpha_\beta(\pi)$. By Theorem 1 in [RU00] we can also express the conditional value at risk in the following form:

Theorem 5.1.1. *The CVaR with risk level β can be determined form the formula:*

$$\Phi_\beta = \inf_{\tau \in \mathbb{R}} \mathbb{E}^{\mathbb{P}}[\tau + \frac{1}{\alpha}[f(\pi, \xi) - \tau]^+]. \tag{5.26}$$

5.1.5 Objective Function Definition

The objective function of the DRO OPF model (5.29) comprises a weighted sum of an operational cost function and a constraint violation risk function:

$$h_t = J_{\text{Cost}}^t + \rho J_{\text{Risk}}^t.$$

Where $\rho \in \mathbb{R}_+$ is a weight that quantifies the network operator's risk aversion. The operational cost is assumed to be linear or convex quadratic:

$$J_{\text{Cost}}^t = \sum_{g \in \mathcal{G}} c_{1,g}(P_g^G)^2 + c_{2,g}P_g^G + c_{3,g}$$

which captures nominal and reserve costs of responding to wind energy forecast errors.

Affine inequality constraints such as (5.6) and concave constraints such as (5.5) cannot be expressed as constraints on the policy π as done for the affine equality constraints of the model. We group each of these constraints in the set $\mathcal{C}_t$. We model the constraints in $\mathcal{C}_t$ as components of the constraint violation risk function using the Conditional Value at Risk (CVaR) [RU00]. Each constraint c in $\mathcal{C}_t$ can be written in the form $\mathrm{C}_{\text{tc}}(\pi, \xi_t) \leq 0$ with C_{tc} affine (concave) with respect to ξ_t if the corresponding constraint c is affine (concave). The CVaR with risk level β of each individual constraint in the set $\mathcal{C}_t$ is given by:

$$\inf_{\sigma_{tc}} \frac{1}{\sigma_{tp}} \mathbb{E}_{\xi_t} \{ [\mathrm{C}_{\text{tc}}(\pi, \xi_t) + \sigma_{tc}]_+ - \sigma_{tp}\beta \}. \tag{5.27}$$

Thus, we obtain the following distributionally robust total CVaR objective:

$$\hat{J}_{\text{Risk}}^t = \inf_{\sigma_{tp}} \sum_{t=0}^{T} \sum_{c \in \mathcal{C}_t} \sup_{\mathcal{Q}_t \in \hat{\mathcal{P}}_t^{N_s}} \frac{1}{\sigma_{tp}} \mathbb{E}^{\mathcal{Q}_t} [\{ [\mathrm{C}_{\text{tc}}(\pi, \xi_t) + \sigma_{tc}]_+ - \sigma_{tp}\beta \}]. \tag{5.28}$$

We observe that the functions inside the expectations can be expressed as the maximum of affine or concave functions respect to ξ_t:

$$[\mathrm{C}_{\text{tc}}(\pi, \xi_t) + \sigma_{tc}]_+ - \sigma_{tp}\beta = \max\{\mathrm{C}_{\text{tc}}(\pi, \xi_t) + (1-\beta)\sigma_{tc}, -\sigma_{tc}\beta\}\}$$

for all $c \in \mathcal{C}_t$.

5.1.6 Pre-reduction model

We obtain the following Distributionally Robust Optimal power flow model with optimization variables the components defining the affine function $\pi = (P_h^{G_n}, P_h^{G_e}, , Q_h^{G_n}, c_{tkm}^e, c_{tkm}^n, s_{tkm}^e, s_{tkm}^n)$ for all $t = 1, \dots, T$, $h \in \mathcal{G}_t$, and $k, m \in \mathbf{L}$ if $k \neq m$ and for all $k = m \in \mathbf{B}$.

$$\inf_{\pi \in \Pi_{\text{aff}}, \sigma_c^t} \sum_{t=0}^{T} \left\{ \mathbb{E}\left[\hat{J}_{\text{Cost}}^t \right] \right.$$
$$\left. + \rho \sum_{c \in \mathcal{C}_t} \sup_{\mathbb{Q}_t \in \hat{\mathcal{P}}_t^{N_s}} \mathbb{E}^{\mathbb{Q}_t}\left[\max\{\mathrm{C}_{\text{tc}}(\pi, \xi_t) + (1 - \beta)\sigma_{tc}, -\sigma_{tc}\beta\} \right] \right] \right\}$$

(5.29)

s.t. $\forall k \in \mathbf{B}, \ t = 1, ..., T$:

$$P_{tk}^{G_n} - P_{tk}^d - m_k^P \begin{pmatrix} c_t^n \\ s_t^n \end{pmatrix} = 0, \tag{5.30}$$

$$P_{tk}^{G_e} - m_k^P \begin{pmatrix} c_t^e \\ s_t^e \end{pmatrix} = 0 \tag{5.31}$$

$$Q_{tk}^{G_n} - Q_{tk}^d - m_k^Q \begin{pmatrix} c_t^n \\ s_t^n \end{pmatrix} = 0, \tag{5.32}$$

$$Q_{tk}^{G_e} - m_k^Q \begin{pmatrix} c_t^e \\ s_t^e \end{pmatrix} = 0 \tag{5.33}$$

$$\|P_{tg}^{Ge} L\|_\infty \le P_{tg}^{Gn} - P_g^{\min} \tag{5.34}$$

$$\|P_{tg}^{Ge} L\|_\infty \le P_g^{\max} - P_{tg}^{Gn} \tag{5.35}$$

$$\|(c_{tbb}^e - V_b^{\max^2} c_{\overline{tbb}}^e)L\|_\infty \le V_b^{\max^2} c_{\overline{tbb}}^n - c_{tbb}^n \tag{5.36}$$

$$\|(c_{tbb}^e - V_b^{\min^2} c_{\overline{tbb}}^e)L\|_\infty \le c_{tbb}^n - V_b^{\max^2} c_{\overline{tbb}}^n \tag{5.37}$$

where $\mathbb{E}[\hat{J}_{\text{Cost}}^t]$ represents the mean cost over the empirical distribution $\hat{P}^{N_s}$.

5.1.7 Finite reduction

Problem 5.29 is an optimization problem over an infinite dimensional vector field $\hat{\mathcal{P}}_t^{N_s}$, to make make the problem tractable, we use results from Kuhn et al. [EK18] reviewed in Chapter 2 to reformulate

$$\sup_{\mathbb{Q}_t \in \hat{\mathcal{P}}_t^{N_s}} \mathbb{E}^{\mathbb{Q}_t}\left[\rho \max\{\mathrm{C}_{\text{tc}}(\pi, \xi_t) + (1 - \beta)\sigma_{tc}, -\sigma_{tc}\beta\} \right] \tag{5.38}$$

as a finite dimensional convex program. Whenever $\mathrm{C}_{\text{tc}}(\pi, \xi_t)$ is affine, we can utilize the reduction result for piecewise affine functions Corollary 2.19, and whenever $\mathrm{C}_{\text{tc}}(\pi, \xi_t)$ is concave with respect to ξ_t, we will employ Theorem 2.4.

Let us consider the first case. Let $c \in \mathcal{C}_t$ be an affine constraint then we can write $\mathrm{C}_{\text{tc}}(\pi, \xi_t) = A_{tc}(\pi)\xi_t + b_{tc}(\pi)$ for some $A_{tc}(\pi) \in \mathbb{R}^{N_\xi}$ and $b_{tc}(\pi) \in \mathbb{R}$ both depending linearly on π. Thus applying (2.0.2.1) we obtain that (5.38)

is equal to the infimum:

$$\inf_{\lambda_t^c, s_{ti}^c, \gamma_{tik}^c} \lambda_t^c + \frac{1}{N_s} \sum_{i=1}^{N_s} s_{ti}^c$$

$$\begin{aligned}
\text{s.t.} \quad & \rho(b_{tc}(\pi) + (1-\beta)\sigma_{tc} + \langle A_{tc}, \hat{\xi}_t \rangle + \langle \gamma_{ti1}^c, d - H\hat{\xi}_i \rangle) \leq s_{ti}^c \\
& \rho(-\sigma_{tc}\beta + \langle \gamma_{ti2}^c, d.H\hat{\xi}_i \rangle) \leq s_{ti}^c \\
& \|H^T \gamma_{ti1} - \rho A_t c(\pi)\|_\infty \leq \lambda_t^c \\
& \|H^T \gamma_{ti1}\|_\infty \leq \lambda_t^c \\
& \gamma_{tik} \geq 0 \\
& \forall i \leq N_s, \ k = 1, 2,
\end{aligned} \tag{5.39}$$

Let us now consider the case where $C_{tc}(\pi, \xi_t)$, of a constraint $c \in \mathcal{C}_t$, is concave with respect to ξ_t and π, then the maximum in (5.38) is the maximum of a concave and an affine function. Therefore, using Theorem 2.4, problem (5.38) is equal to the following infimum:

$$\inf_{\substack{\lambda_t^c, s_{ti}^c, \\ \gamma_{tik}^c}} \lambda_t^c + \frac{1}{N_s} \sum_{i=1}^{N_s} s_{ti}^c \tag{5.40}$$

$$\text{s.t.}$$

$$\begin{aligned}
& \rho\left([-C_{tc}(\pi, \xi_t) + (\beta - 1)\sigma_{tc}]^* (z_{i1}^c - \nu_{ti1}^c) + \sigma_\Xi(\nu_{ti1}) - \langle z_{ti1}^c, \hat{\xi}_i \rangle \right) \leq s_{ti}^c \\
& \rho\left([-\delta_{tc}\beta]^*(z_{i2}^c - \nu_{ti2}^c) + \sigma_\Xi(\nu_{ti1}) - \langle z_{ti1}^c, \hat{\xi}_i \rangle \right) \leq s_{ti}^c \\
& \|z_{ik}^c\|_\infty \leq \lambda_t^c \\
& \forall i \leq N_s, \ k = 1, 2.
\end{aligned}$$

We observe that the infimum corresponds to the optimal value of a convex problem since $[-\ell_{tk}]^*$ is convex. Finally, by substituting (5.39) and (5.40) into (5.29), we obtain the equivalent finite dimensional convex reformulation of Problem 5.29.

$$\inf_{\substack{\pi\in\Pi_{\text{aff}},\sigma^t_c \\ \lambda^c_t,s^c_{ti}, \\ \gamma^c_{tik}}} \sum_{t=0}^{T} \left\{ \mathbb{E}\left[\hat{J}^t_{\text{Cost}}\right] + \rho \sum_{c\in\mathcal{C}_t} \lambda^c_t + \frac{1}{N_s} \sum_{i=1}^{N_s} s^c_{ti} \right.$$

$$\text{s.t. } \forall k \in \mathbf{B}, \ t = 1,...,T :$$

$$P^{G_n}_{tk} - P^d_{tk} - m^P_k \begin{pmatrix} c^n_t \\ s^n_t \end{pmatrix} = 0,$$

$$P^{G_e}_{tk} - m^P_k \begin{pmatrix} c^e_t \\ s^e_t \end{pmatrix} = 0$$

$$Q^{G_n}_{tk} - Q^d_{tk} - m^Q_k \begin{pmatrix} c^n_t \\ s^n_t \end{pmatrix} = 0,$$

$$Q^{G_e}_{tk} - m^Q_k \begin{pmatrix} c^e_t \\ s^e_t \end{pmatrix} = 0$$

$$\|P^{G_e}_{tg} L\|_\infty \le P^{G_n}_{tg} - P^{\min}_g$$

$$\|P^{G_e}_{tg} L\|_\infty \le P^{\max}_g - P^{G_n}_{tg}$$

$$\|(c^e_{tbb} - V^{\max^2}_b c^e_{\overline{tbb}})L\|_\infty \le V^{\max^2}_b c^n_{\overline{tbb}} - c^n_{tbb}$$

$$\|(c^e_{tbb} - V^{\min^2}_b c^e_{\overline{tbb}})L\|_\infty \le c^n_{tbb} - V^{\max^2}_b c^n_{\overline{tbb}}$$

$$\forall c \in \mathcal{C}_t \text{ concave constraints:}$$

$$\rho\left([-\text{C}_{\text{tc}}(\pi,\xi_t) + (\beta-1)\sigma_{tc}]^*(z^c_{i1} - \nu^c_{ti1}) + \sigma_\Xi(\nu_{ti1}) - \langle z^c_{ti1}, \hat{\xi}_i\rangle\right) \le s^c_{ti}$$

$$\rho\left([-\delta_{tc}\beta]^*(z^c_{i2} - \nu^c_{ti2}) + \sigma_\Xi(\nu_{ti1}) - \langle z^c_{ti1}, \hat{\xi}_i\rangle\right) \le s^c_{ti}$$

$$\|z^c_{ik}\|_\infty \le \lambda^c_t$$

$$\forall i \le N_s, \ k = 1,2.$$

$$\forall c \in \mathcal{C}_t \text{ affine constraints:}$$

$$\rho(b_{tc}(\pi) + (1-\beta)\sigma_{tc} + \langle A_{tc}, \hat{\xi}_t\rangle + \langle\gamma^c_{ti1}, d - H\hat{\xi}_i\rangle) \le s^c_{ti}$$

$$\rho(-\sigma_{tc}\beta + \langle\gamma^c_{ti2}, d.H\hat{\xi}_i\rangle) \le s^c_{ti}$$

$$\|H^T \gamma_{ti1} - \rho A_t c(\pi)\|_\infty \le \lambda^c_t$$

$$\|H^T \gamma_{ti1}\|_\infty \le \lambda^c_t$$

$$\gamma_{tik} \ge 0.$$

Example 2. *As a concrete example, we illustrate the modeling of the power loss inequality constraint* (5.6) *within our DRO OPF framework. This involves expanding the expressions of* P_{tkh} *and* P_{thk} *as follows.*

$$0 \geq - P_{tkh} - P_{thk} =$$
$$= - G_{kk}c_{kk} - G_{kh}c_{tkh} - B_{kh}s_{tkh} - G_{hh}c_{thh} - G_{hk}c_{thk} - B_{hk}s_{thk} =$$
$$= - (G_{hk} + G_{kh})c_{tkh} - (B_{kh} - B_{hk})s_{tkh} - G_{kk}c_{tkk} - G_{hh}c_{thh}$$
$$= - (G_{hk} + G_{kh})(c^n_{tkh} + c^e_{tkh}\xi_t) - (B_{kh} - B_{hk})(s^n_{tkh} + s^e_{tkh}\xi_t) +$$
$$- G_{kk}(c^n_{tkk} + c^e_{tkk}\xi_t) - G_{hh}(c^n_{thh} + c^e_{thh}\xi_t) = C_{t\bar{c}}(\pi, \xi_t).$$

Thus the power loss inequality constraint can be written in the form $\mathrm{C}_{\mathrm{tc}}(\pi, \xi_t) = A_{tc}(\pi)\xi_t + b_{tc}(\pi) \leq 0$ *with* $A_{tc}(\pi)$ *an* $b_{tc}(\pi)$ *defined as follows:*

$$A_{tc}(\pi) = - \left[(G_{hk} + G_{kh})c^e_{tkh} + (B_{kh} - B_{hk})s^e_{tkh} - G_{kk}c^e_{tkk} - G_{hh}c^e_{thh}\right],$$
$$b_{tc}(\pi) = - \left[(G_{hk} + G_{kh})c^n_{tkh} + (B_{kh} - B_{hk})s^n_{tkh} - G_{kk}c^n_{tkk} - G_{hh}c^n_{thh}\right].$$

Therefore, by following the passages at the beginning of this section, the CVaR of violating the power loss constraint, can be modeled in the form (5.39).

5.2 Numerical results

In this section, we apply the model developed in the previous section to a variant of the IEEE 118-bus test system, as referenced in [Guo+19a] and [Zim+11]. This modified version, depicted in Fig. 5.1, integrates three wind farms connected to buses #1, #9, and #26. Consequently, the conventional generators originally connected to bus #1 and bus #26 have been removed. The implementation details discussed in this chapter, along with the codes for models (4.1) and (5.4), are available for download at `https://github.com/frulcino/Jabr-OPF-models`.

The wind power forecast errors were obtained from a data source provided by CESI. The process involves fitting a Weibull distribution to represent the probability distribution of wind power production. The predicted power output is then set as the mean of this distribution. Error samples are generated by drawing random samples from the fitted distribution and then computing the discrepancy with the predicted power generation.

To align our findings with existing literature [Guo+19a], we adjusted our forecasting errors to have a zero mean. Additionally, standard deviations were set at $200MW$ for wind farms #1 and #9, and $300MW$ for wind farm #26. We focused on the single-period stochastic optimization problem following the approach in [Guo+19a].

The dataset comprises 152 observations and is divided into 13 samples to train the model and 139 test samples for measuring the out-of-sample performance of the solution. Therefore $N_s = 13$ and $N_\Xi = 3$ since there are three windfarms whose power generation is predicted. The optimization task was completed in 100 seconds for each loop on a laptop equipped with an Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz and 8GB of RAM.
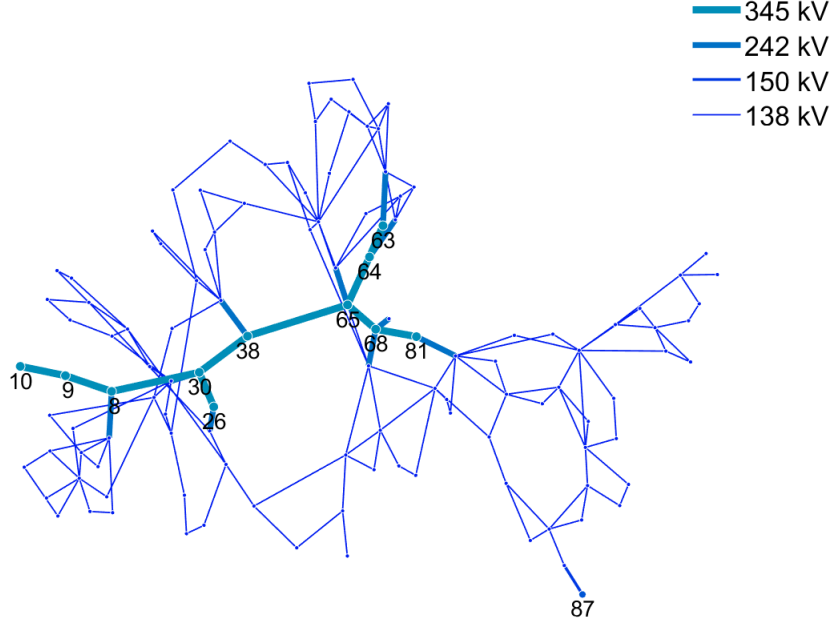
Figure 5.1: A modified IEEE 118-bus test system.

Figure 5.2 illustrates the variation in the average operational cost based on different values of the risk aversion parameter ρ and the radius ε of the Wasserstein metric. An evident trend is the rise in average nominal cost as ρ increases. This indicates a greater focus on avoiding constraint violations across all potential realizations of ξ. While this strategy ensures higher security, it does come at the cost of increased operational expenses.

Figure 5.3 showcases the nominal power loss in some key lines of the network, a discernible pattern is the decreasing constraint violation as ρ increases for all lines except for line 15. This happens because a overall lower cost is achieved by allowing higher risk of violating the power loss constraint in line 15. In Figure 5.4, we observe how out of sample performance varies with ε. In particular, 5.4 shows the average number of power loss constraint violations of the test samples. We observe that large ε ensures fewer constraints violations. Inequality affine constraints on power generation and voltage magnitudes, implemented as policy constraints as described in Section 5.1.3, are never violated for the tests samples.

Lastly, Figure 5.5 shows the relationship between nominal real power output and the risk aversion coefficient ρ. A discernible pattern is the increasing conservatism in policy as ρ elevates. This ensures respect for magnitude constraints, especially when confronted with extreme realizations of the predictions error ξ. However the change in power generation in some buses,
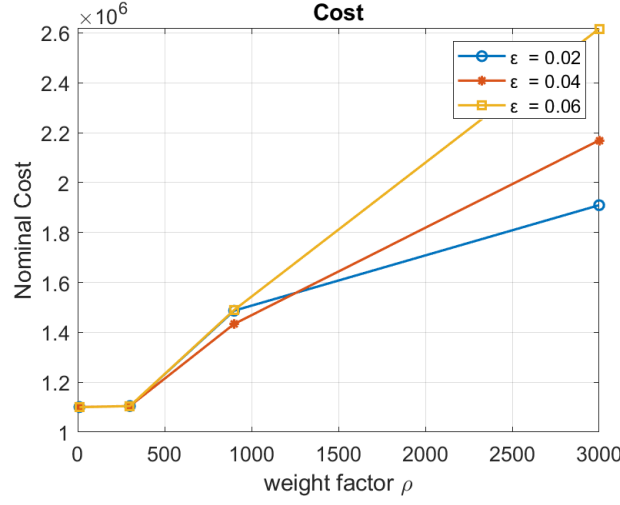
Figure 5.2: Comparison of the average operational cost for various values of risk aversion ρ and Wasserstein metric radius ε.
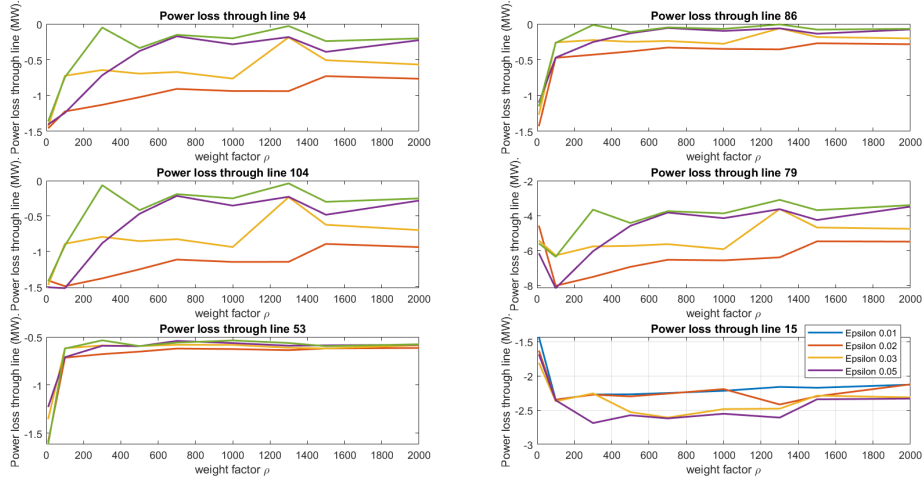
Figure 5.3: Comparison of the power loss in six lines of the network for various values of risk aversion parameter ρ and the Wasserstain metric radius ε.

as in bus 15 and bus 31, is not monotonic. This due to the nonlinearities of the OPF model, which make the behavior of the solution with respect to the changes of various parameters somewhat impredictable.
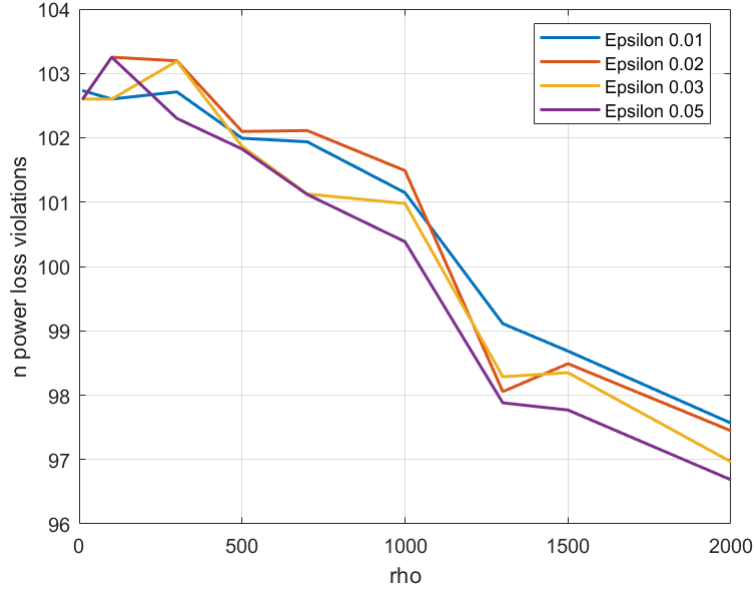
Figure 5.4: Comparison of the number of out of samples constraints violation for various values of risk aversion parameter ρ and the Wasserstain metric radius ε.

Figure 5.5: Comparison of output real powers of selected generators for different risk levels ρ and Wasserstein metric values ε.

Chapter 6

Conclusions

We have showcased the efficacy and adaptability of our proposed data-driven, distributionally robust stochastic OPF methodology. Unlike many existing models, ours directly integrates error training datasets, eliminating the need to presuppose a forecast error distribution. This approach allows us to use distributionally robust optimization techniques, leading to enhanced out-of-sample performance. Importantly, our underlying deterministic model operates with fewer assumptions compared to the conventional DC OPF. This ensures resilience, especially when faced with extreme weather events or fluctuating demand scenarios in the network. We introduced a novel method of modeling affine inequality constraints as continuity constraints on the policy, allowing us to impose fewer constraints using Conditional Value at Risk (CVaR). The key distinction between these two approaches is that the former is 'hard,' enforcing the constraint for every possible realization of error predictions but is only applicable to affine constraints, while the latter can be applied more generally to concave constraints but it does not guarantee compliance for every possible realization of error predictions. However, it is worth noting that our model tends to be more computationally demanding than the DCOPF-based stochastic optimal power flow model presented in [Guo+19a], due to the increased number of variables and constraints derived from the AC OPF formulation. To reduce the number of variables and constraints, one approach is to identify the critical lines in the network, by running a relaxation of the model without the line constraints and by identifying where these constraints are violated, and apply the CVaR constraints exclusively to these lines. From simulations results, we observed that larger ε may lead to large constraint violation when the sum of the CVars has a lower overall risk by allowing higher risk of violation of specific constraints. This may addressed by considering the risk of joint constraint violation instead of modeling each constraint individually. In our model we considered Ξ to be in the form $[-E_1, E_1] \times [-E_2, E_2] \times [-E_3, E_3]$. Although this would work well if the components of ξ were independent, this is rarely the case for wind

power prediction, and this results in overly conservative policies. Therefore a better approach, could be to consider the best polytope fitting the considered samples. Moreover, limiting our consideration to affine policies in a nonlinear problem like the OPF overlooks numerous admissible policies that are nonlinear but have the potential to be more optimal. A potential avenue for future research could involve exploring piecewise affine policies instead.

In Chapter 4, we integrated the loop constraints 4.0.4 into the Jabr model (4.1). This integration aimed at achieving exactness in non-radial networks. These constraints are in general non convex but a linear relaxation of these constraints can be implemented by using McCormick envelops. Furthermore, the calculation of the loop constraint becomes increasingly difficult with the length of the loop. Since there is not a unique Cycle Base, this could be addressed by choosing the Cycle Base with the smaller maximum length of the cycles in the Cycle Base. This is similar to the minimum weight cycle base problem [Kav+09]. Proposition 4.0.6 offers a parametrization of the solution space of the loop constraints, the behavior of this parametrization in conjunction with the other constraints of the OPF problem remains uncertain. A promising avenue for future research would be to discern the conditions under which the image of this parametrization is included in or equals to the solution space of the OPF problem 3.8.

Bibliography

[Bak21]   Kyri Baker. "Solutions of DC OPF are Never AC Feasible". In: (Oct. 2021), pp. 264–268. DOI: 10.1145/3447555.3464875.

[Ber03]   Dimitri Bertsekas. *Convex Analysis and Optimization*. Athena Scientific, Jan. 2003. ISBN: 1886529450.

[BCH14]   Daniel Bienstock, Michael Chertkov, and Sean Harnett. "Chance-Constrained Optimal Power Flow: Risk-Aware Network Control under Uncertainty". In: *SIAM Review* 56.3 (Apr. 2014), pp. 461–495. DOI: 10.48550/arXiv.1209.5779.

[Bie+20]   Daniel Bienstock, Mauro Escobar, Claudio Gentile, and Leo Liberti. "Mathematical Programming formulations for the Alternating Current Optimal Power Flow problem". In: *4OR* 18.3 (July 2020), pp. 249–292. DOI: 10.1007/s10288-020-00455-w.

[BV19]   Daniel Bienstock and Abhinav Verma. "Strong NP-hardness of AC power flows feasibility". In: *Operations Research Letters* 47.6 (2019), pp. 494–501. DOI: 10.1016/j.orl.2019.08.009.

[BA09]   Jorge Jardim Brian Stott and Ongun Alsac. "DC Power Flow Revisited". In: *IEEE Transactions on Power Systems* 24.3 (May 2009), pp. 1290–1300. DOI: 10.1109/TPWRS.2009.2021235.

[Car79]   Jean Carpentier. "Optimal power flows". In: *International Journal of Electrical Power & Energy Systems* 1.1 (1979), pp. 3–15. DOI: 10.1016/0142-0615(79)90026-7.

[EK18]   Peyman Mohajerin Esfahani and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations". In: *Mathematical Programming* 171.1 (Sept. 2018), pp. 115–166. DOI: 10.1007/s10107-017-1172-1.

[FR16]   Stephen Frank and Steffen Rebennack. "An introduction to optimal power flow: Theory, formulation, and examples". In: *IIE Transactions* 48.12 (2016), pp. 1172–1197. DOI: 10.1080/0740817X.2016.1189626.

[Guo+19a] Yi Guo, Kyri Baker, Emiliano Dall'Agnese, Zechun Hu, and Tyler Holt Summers. "Data-Based Distributionally Robust Stochastic Optimal Power Flow—Part II: Case Studies". In: *IEEE Transactions on Power Systems* 34.2 (2019), pp. 1493–1503. DOI: `10.1109/TPWRS.2018.2878380`.

[Guo+19b] Yi Guo, Kyri Baker, Emiliano Dall'Anese, Zechun Hu, and Tyler Holt Summers. "Data-Based Distributionally Robust Stochastic Optimal Power Flow—Part I: Methodologies". In: *IEEE Transactions on Power Systems* 34.2 (2019), pp. 1483–1492. DOI: `10.1109/TPWRS.2018.2878385`.

[Jab06] Rabih A. Jabr. "Radial distribution load flow using conic programming". In: *IEEE Transactions on Power Systems* 21.3 (2006), pp. 1458–1459. DOI: `10.1109/TPWRS.2006.879234`.

[Kav+09] Telikepalli Kavitha, Christian Liebchen, Kurt Mehlhorn, Torsten Ueckerdt Dimitrios Michail Romeo Rizzi, and Katharina A. Zweig. "Cycle bases in graphs characterization, algorithms, complexity, and applications". In: *Computer Science Review* 3.4 (Sept. 2009), pp. 199–243. DOI: `10.1016/j.cosrev.2009.08.001`.

[RU00] Roger Tyrrell Rockafellar and Stanislav Uryasev. "Optimization of conditional value-at risk". In: *Journal of Risk* 3 (Jan. 2000), pp. 21–41. DOI: `10.21314/JOR.2000.038`.

[San15] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians. Calculus of Variations, PDEs and Modeling.* Birkhäuser, 2015. ISBN: 3319208284.

[Sha00] Alexander Shapiro. "On Duality Theory of Conic Linear Problems". In: (Dec. 2000). DOI: `10.1007/978-1-4757-3403-4_7`.

[YV14] Hervé Pajot Yann Ollivier and Cédric Villani. "Optimal Transport: Theory and Applications". In: 413 (2014). DOI: `10.1017/CBO9781107297296`.

[Zim+11] Ray Daniel Zimmerman, Murillo-Sánchez, Carlos Edmundo, and Thomas Robert John. "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education". In: *IEEE Transactions on Power Systems* 26.1 (2011), pp. 12–19. DOI: `10.1109/TPWRS.2010.2051168`.

Acknowledgements