Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Bartolomeo Stellato

Data-Driven Chance Constrained Optimization

Master Thesis

Automatic Control Laboratory

Swiss Federal Institute of Technology (ETH) Zürich

Supervision

Bart P. G. Van Parys
Xiaojing Zhang
Dr. Paul J. Goulart
Prof. Dr. John Lygeros

September 2014

To my parents

Nicola and Daniela

Contents

Abstract

Traditional optimization methods for decison-making under uncertainty assume perfect model information. In practice, however, such precise knowledge is rarely available. Thus, optimal decisions coming from these approaches can be very sensitive to perturbations and unreliable. Stochastic optimization programs take into account uncertainty but are intractable in general and need to be approximated. Of late, distributionally robust optimization methods have shown to be powerful tools to reformulate stochastic programs in a tractable way. Moreover, the recent advent of cheap sensing devices has caused the explosion of available historical data, usually referred to as “Big Data”. Thus, modern optimization techniques are shifting from traditional methods to data-driven approaches.

In this thesis, we derive data-driven tractable reformulations for stochastic optimization programs based on distributionally robust optimization. In the first part of this work we provide our theoretical contributions. New distributionally robust probability bounds are derived and used to reformulate uncertain optimization programs assuming limited information about the uncertainty. Then, we show how this information can be derived from historical data. In the second part of this work, we compare the developed methods to support vector machines in a machine learning setting and to randomized optimization and in a control context.

Acknowledgements

My thanks go to all the people who made it possible for me to write this thesis.

First of all, I would like to thank Bart P. G. Van Parys for his support, mentoring and patience throughout this project. I am also indebted to Xiaojing Zhang for his supervision and his guidance during these months. A further thanks to Dr. Sergio Grammatico for his fruitful suggestions in the theoretical part of this thesis.

My gratitude goes also to Dr. Paul J. Goulart for his interesting ideas on the research directions of this work. I would like to thank also to Professor John Lygeros who gave me the opportunity to work on this project.

Finally, a very special word of gratitude goes to my parents, for providing me the great privilege of studying at ETH Zürich. Without their lovely support I could not have achieved this degree, which I believe will be the first big step in my future career.

1 Introduction

In the last decades, with the advent of fast processors and recent mathematical tools, optimization has become one of the mainstream frameworks in many decision-making and control contexts. In particular, with the recent development of fast interior point methods for convex optimization, in particular semidefinite programming (e.g. Boyd and Vandenberghe [9],[47]), large problems can now be rapidly solved allowing us to apply these techniques on fast dynamical systems.

Convex optimization problems can be described as:

$$\begin{aligned} \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & g(x) \leq 0, \end{aligned} \tag{1.1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$, $c \in \mathbb{R}^n$, $g \colon \mathcal{X} \to \mathbb{R}^{n_g}$. The constraint g is a convex function and the cost function is linear without loss of generality, see [9].

Traditional approaches assume that problem (1.1) describes a perfectly known model governed by constraints g. Unfortunately, such precise information is rarely available in practice because systems are too complex and need to be approximated or because their parameters cannot always be exactly estimated. Indeed, it has long been known that solutions to problem (1.1) can exhibit high sensitivity to function g. Hence, the solution x^* might be highly infeasible and/or suboptimal in practice. This issue has been recently adressed in a rigorous way in [3].

Uncertain Optimization Problems Robust optimization is a computationally attractive method to deal with stochastic programs. Uncertainty $\xi \in \mathbb{R}^{n_\xi}$ is introduced in problem (1.1) so that the computed solution is optimal for any realization of ξ in set $\Xi \subseteq \mathbb{R}^{n_\xi}$:

$$\begin{aligned} \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & g(x, \xi) \leq 0, \quad \forall \xi \in \Xi, \end{aligned} \tag{1.2}$$

where, now, $g \colon \mathcal{X} \times \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_g}$. In the 1970s, Soyster in [44] and Falk in [25] started to discuss ways to model and solve robust programs. However, only in the 1990s with the results from Ben-Tal and Nemirovski (see [2]) and by El Ghaoui in [24], the interest of the optimization community focused on the issue of robustness. Unfortunately, robust programs are not as easy to solve as the original problem and are $\mathcal{NP}$-hard in general, [2]. In addition, set Ξ is usually unknown and thus, there is no rigorous way to define it if we do not know ξ precisely.

Another approach to deal with uncertainty is chance constrained optimization. This method assumes ξ has a probability distribution $\mathbb{P}$ and ensures that the optimal solution satisfies the constraints with high probability:

$$\begin{aligned} \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \mathbb{P}\left(g(x, \xi) \leq 0\right) \geq 1 - \epsilon, \end{aligned} \tag{1.3}$$

where ϵ is a small positive number. This technique has been investigated for decades starting from the 1950s by Dantzig in [19]. There is an extensive literature regarding chance constrained programs and their reformulations (see e.g. [6] and [31]). However, also chance constrained programs are computationally intractable for generic distributions. Indeed, Shapiro and Nemirovski [34] pointed out that computing the probability of a weighted sum of uniformly distributed variables being nonpositive is already $\mathcal{NP}$-hard. Thus, programs of the form (1.3) need to be approximated. Moreover, the true probability distribution $\mathbb{P}$ is unknown in general and it has to be estimated.

Distributionally robust optimization has gained a lot of popularity in the last decade because it allows us to reformulate chance constraints in a tractable way taking into account our limited knowledge about the distribution $\mathbb{P}$. A distributionally robust optimization problem can be written as:

$$\begin{aligned} \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \mathbb{P}\left(g(x,\xi) \leq 0\right) \geq 1-\epsilon, \quad \forall\, \mathbb{P} \in \mathcal{P}, \end{aligned} \tag{1.4}$$

where set $\mathcal{P}$ is the smallest set containing the true distribution given limited information. Several authors have recently shown that knowing only the first two moments of $\mathbb{P}$, it is possible to reformulate the chance constraints preserving computational tractability. Rockafellar and Uryasev in [37], then El Ghaoui in [23], derived distributionally robust reformulations under the framework of Conditional Value-at-Risk. More recently Joel and Melvyn in [28], developed tractable approximations for chance constrained linear programs. Moreover, Chen et al. in [17] used distributionally robust optimization to construct uncertainty sets with probabilistic guarantees in order to transform problem (1.4) in (1.2) and solve it efficiently. Of late, Zymler et al. [50] derived distributionally robust reformulations for multiple chance constraints with limited moment information about the distribution. In 2013, Van Parys et al. [46] applied within Conditional Value-at-Risk setting, distributionally robust optimization for control synthesis providing probabilistic guarantees. Thus, due to its computational attractiveness many research directions are currently converging to this field.

Probability Inequalities All the distributionally robust optimization methods are based on the worst-case probability distributions within a family $\mathcal{P}$. Typically $\mathcal{P}$ is chosen as as the set of all distributions sharing the first two moments because they can be easily estimated. The problem of finding probability bounds holding for all the distributions sharing mean and covariance has been studied since the second half 19th century. In 1867, Chebyshev proved in [16] the following inequality:

Theorem 1.0.1 (Chebyshev Inequality [16]). *Let $\xi \in \mathbb{R}$ be a random variable with mean μ and non-zero standard deviation σ. Then, the following holds*

$$\mathbb{P}\left(|\xi-\mu| > k\sigma\right) \leq \begin{cases} \dfrac{1}{k^2} & k > 1 \\ 1 & \text{otherwise.} \end{cases} \tag{1.5}$$

In this case, the considered region is the complement of a line segment in $\mathbb{R}$ centered at the origin of length $2k$. In 1910, Cantelli proved an extension of this inequality considering only one of the distribution tails:

Theorem 1.0.2 (Cantelli Inequality [15]). *Let $\xi \in \mathbb{R}$ be a random variable with mean μ and non-zero standard deviation σ. Then, the following holds*

$$\mathbb{P}\left(\xi - \mu > k\sigma\right) \leq \begin{cases} \dfrac{1}{1+k^2} & k > 0 \\ 1 - \dfrac{1}{1+k^2} & k < 0. \end{cases} \tag{1.6}$$

The considered region is now the half line $(k, +\infty)$. The attractiveness of these bounds is their distribution-free nature in the sense that they depend only on the set $\mathcal{P}$. On the other hand, the worst-case distributions for these inequalities are discrete with few atoms and are unlikely to be encountered in practice. Thus, the obtained bounds are in general pessimistic. In order to mitigate this over-pessimism, it is possible to assume additional information about the elements of $\mathcal{P}$. A structural property often encountered in practice is unimodality. Informally, a continuous distribution is unimodal with mode γ if its Probability Density Function (PDF) is non-increasing with increasing distance from the mode. In addition, most of the commonly studied distributions are unimodal: e.g. Gauss, Cauchy, Gamma, etc. In 1821, Gauss proved in [27] an inequality similar to the Chebyshev inequality (1.5) with the additional assumption of unimodality:

Theorem 1.0.3 (Gauss Inequality [27]). *Let $\xi \in \mathbb{R}$ be a random variable with mode equal to the mean μ and non-zero standard deviation σ. Then, the following holds*

$$\mathbb{P}\left(|\xi - \mu| > k\sigma\right) \leq \begin{cases} \dfrac{4}{9k^2} & k > \dfrac{2}{3} \\ 1 - \dfrac{k}{\sqrt{3}} & \text{otherwise}. \end{cases} \tag{1.7}$$

This result provides a much less pessimistic bound improving the Chebyshev inequality by a factor 4/9. We will show that the unimodality assumption can improve the chance constrained programs reformulations by giving less conservative solutions.

Data-driven methods Even if $\mathcal{P}$ includes limited distribution information, in practice we do not know it. However, with the growing availability of cheap sensing devices and big storage capacities, the amount of data available has increased exponentially. Hence, we can use these data to estimate the set $\mathcal{P}$ describing our knowledge about the distribution. In this direction, Delage and Ye in [20] provided a rigorous way to estimate distribution moments from data. Also Bertsimas et al. in [5] reformulated chance constraints starting from data and constructing uncertainty sets with probabilistic guarantees.

Recently, Calafiore and Campi in [12], later Campi and Garatti in [14] and then Calafiore in [11] developed the so called Scenario Approach to solve chance constrained programs from data. The Scenario Approach reformulates the problem (1.3) using the past N data samples $\xi^{(i)}$ of uncertainty ξ as

$$\begin{aligned} &\underset{x \in \mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && g(x, \xi^{(i)}) \leq 0, \quad i = 1, \dots, N. \end{aligned} \tag{1.8}$$

This approach is based on randomized optimization as the formulated program is itself random. Even though the number of samples is limited, it has been proven that if N is large enough, then, with high confidence, the solution of the random program (1.8) will satisfy the chance constraints in (1.3), see [11]. Even though this method is intuitive and easy to implement, it is computationally expensive because the N required is typically very large even for medium size problems. Thus, also the number of constraints to be satisfied becomes large leading to an intractable program.

1.1 Data-driven Tractable Reformulations

The main contribution of the present work is the reformulations of chance constrained linear programs focusing on:

- Tractability of the proposed methods,
- Conservatism of the obtained solutions,
- Data-driven reformulations.

We obtain an extension of Gauss inequality to unimodal distributions parametrized by a positive index α (see [21]). Then, we generalize this bound to multiple dimensions. These new inequalities are used to construct multidimensional ellipsoidal sets with probabilistic guarantees. In particular, we obtain the closed form solution for the Minimum Volume Ellipsoid (MVE) containing a minimum amount of probability mass for all the distributions in $\mathcal{P}$. Furthermore, the so called empirical Chebyshev inequality [38] is generalized in a multivariate settings providing a direct extension of the obtained probability bounds for plug-in estimates of distribution moments.

From the obtained results, we exactly reformulate single linear chance constraints as Second-Order Cone (SOC) constraints (see [9] for SOC details) using the unimodality assumption. We then develop two approximations of multiple chance constraints. The first one based on Bonferroni inequality deals with chance constraints individually. The second one is based on robust optimization: we construct multidimensional ellipsoidal uncertainty sets from the probabilistic results we derived, in order to solve robust programs with respect to all the uncertainty realizations within these sets. Afterwards the extension to data-driven estimation of $\mathcal{P}$ is studied providing rigorous approaches to estimate distribution moments.

In order to validate the theoretical results, we compare our uncertain programs reformulations in two different areas:

Machine Learning The problem of linear classification is reformulated with the unimodality assumption obtaining stochastic programs within the Minimax Probability Machine (MPM) [32] framework. These results are compared with the commonly used Support Vector Machines (SVMs).

Model Predictive Control A stochastic water reservoir management problem is described in a Model Predictive Control (MPC) [26] fashion. Then, the program is reformulated and solved with our methods and the unimodality assumption. The obtained solutions are benchmarked against the Scenario Approach (SA) in terms of performance, tractability and conservatism.

1.2 Organization

In the first part of this work we present the theoretical results behind the developed methods. Chapter 2 introduces the α-unimodality framework that will be used throughout this work. In Chapter 3, we derive the probability inequalities in multiple dimensions with α-unimodality assumption. Ellipsoidal uncertainty sets with probabilistic guarantees are computed from these inequalities. Then, the empirical Chebyshev inequality is generalized to multiple dimensions. Chapter 4 introduces chance constrained linear programs: single linear chance constraints reformulations using the obtained probabilistic bounds are studied and then extended to multiple dimensions. Furthermore, these approaches are generalized to data-driven moment estimation.

In the second part of this thesis we benchmark the developed approaches. In Chapter 5, our methods are test in a Machine Learning setting and in Chapter 6 we implement our contributions in a Control context.

1.3 Mathematical Preliminaries

In this section we introduce the basic notation and assumptions in order to avoid ambiguities throughout the different chapters. The complete notation can be found at page 89.

Notation We use $\mathbb{R}$ to denote the set of real numbers, $\mathbb{R}_+$ to denote the set of nonnegative real numbers, and $\mathbb{R}_{++}$ to denote the set of positive real number. The set of real n-vectors is denoted $\mathbb{R}^n$ and the set of real $m \times n$ matrices is denoted $\mathbb{R}^{m\times n}$. Given a set A, $|A|$ denotes its cardinality. We denote by $\mathbb{S}^n$, $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ the sets of all symmetric, positive semidefinite and positive definite matrices in $\mathbb{R}^{m\times n}$ respectively. The relation $X \succeq Y$ ($X \preceq Y$) indicates that $X - Y \in \mathbb{S}^n_+$ ($Y - X \in \mathbb{S}^n_+$). Given $A \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(A)$ designates its trace. Given two matrices $A, B^\top \in \mathbb{R}^{m\times n}$, their inner product is denoted by $\langle A, B\rangle = \mathrm{tr}(AB)$. For any $t \in \mathbb{R}$, $\lceil t \rceil$ indicates the smallest integer not less than t ,while $\lfloor t \rfloor$ indicates the largest integer not more than t. Given $a, b \in \mathbb{R}$, (a, b) and $[a, b]$ denote respectively the open and closed intervals between a and b. Furthermore, if $a, b \in \mathbb{R}^n$, (a, b) and $[a, b]$ denote respectively the open and closed line segments between them. The indicator function $\mathbf{1}_B$ of a set $B \subseteq \mathbb{R}^n$ is defined as: $\mathbf{1}_B(x) = 1$ if $x \in B$; $= 0$ otherwise.

Probability Framework In this paragraph we define some basic formal probability notions and their link to the notation we use in this work, for more details see [22]. We define the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, where: Ω is the space of elementary events, $\mathcal{F}$ is a σ-algebra of subsets of Ω, $\mathbb{Q}$ is a probability measure defined on the events of $\mathcal{F}$. Moreover, let $(\mathbb{R}^{n_\xi}, \mathcal{B}(\mathbb{R}^{n_\xi}))$ be a measurable space with $\mathcal{B}(\mathbb{R}^{n_\xi})$ being the Borel sigma algebra on $\mathbb{R}^{n_\xi}$. A random variable ξ is defined as a measurable function

$$\xi \colon \Omega \to \mathbb{R}^{n_\xi},$$

i.e. $\forall A \in \mathcal{B}(\mathbb{R}^{n_\xi})$, $\xi^{-1}(A) = \{\omega \in \Omega : \xi(\omega) \in A\} \in \mathcal{F}$. We associate to the random variable ξ, the measure $\mathbb{P}_\xi$ on $\mathbb{R}^{n_\xi}$ such that $\mathbb{P}_\xi(\mathbb{R}^{n_\xi}) = 1$, defined as:

$$\begin{aligned}\mathbb{P}_\xi \colon \mathcal{B}(\mathbb{R}^{n_\xi}) &\to [0, 1]\\ A &\mapsto \mathbb{Q}(\xi^{-1}(A)).\end{aligned}$$

We assume that Ω is rich enough to guarantee that, for all $A \in \mathcal{B}(\mathbb{R}^{n_\xi})$ and for all distributions $\mathbb{P}_\xi$ on $\mathbb{R}^{n_\xi}$, it is possible to find a set $\mathcal{C} \subseteq \Omega$ and random variable ξ that maps $\mathcal{C}$ to A.

Throughout this work, we will refer directly to $\xi \in \mathbb{R}^{n_\xi}$ as random variables and to $\mathbb{P}$ on $\mathbb{R}^{n_\xi}$ as their related distribution. Moreover, we will adopt the following notation for $\mathbb{P}$:

$$\mathbb{P}(A) = \mathbb{P}(\{\xi \in \mathbb{R}^{n_\xi} : \xi \in A\}) \coloneqq \mathbb{Q}(\{\omega \in \Omega : \xi(\omega) \in A\}).$$

The set of all Borel probability distributions on $\mathbb{R}^{n_\xi}$ is denoted by $\mathcal{P}_\infty$.

Part I

Theoretical Results

2 α-Unimodality

In this chapter the basic concepts of univariate and multivariate unimodality are introduced. The α-unimodality framework is described using the results by Dharmadhikari and Joag-Dev [21] and Van Parys et al. [45]. These definitions and theorems will be extensively used in the rest of this work, when deriving new probability bounds based on the distribution moments.

2.1 Unimodality

Unimodality is a natural property of many distributions commonly encountered in both theory and practice and it can often be justified by empirical or theoretical motivations.

Let $\xi \in \mathbb{R}$ be a random variable and $\mathbb{P}$ its distribution. Moreover, let the mapping $t \mapsto \mathbb{P}(\xi \leq t)$ be its Cumulative Distribution Function (CDF). The most basic definition of unimodality is the following:

Definition 2.1.1 (Univariate unimodality)**.** *A univariate distribution $\mathbb{P}$ is called unimodal with mode 0 if the mapping $t \mapsto \mathbb{P}(\xi \leq t)$ is convex for $t < 0$ and concave for $t > 0$.*

In the case of continuous unimodal distributions the Probability Density Function (PDF), defined as the mapping $t \mapsto \mathbb{P}(\xi = t)$, decreases as the distance with respect to the mode 0 increases. By consequence, the CDF is convex for $t < 0$ and concave for $t > 0$. Please note that, without loss of generality, the mode is located at the origin, which can always be enforced by a suitable coordinate transformation. In Figure 2.1 the case for the Gaussian distribution is presented.

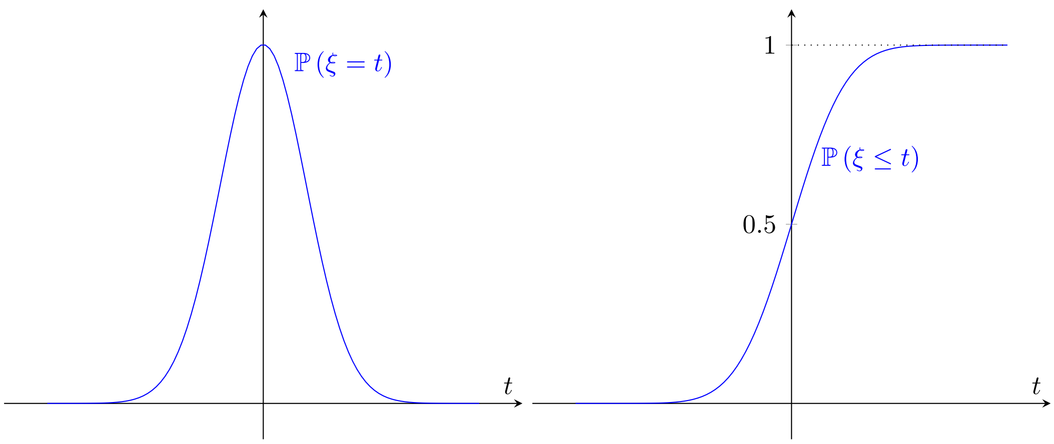

Figure 2.1: Example of unimodal distribution (Gaussian): PDF (left), CDF (right).

In the multivariate case, our definition of unimodality is based on the notion of star-shaped sets. Let now $\xi \in \mathbb{R}^{n_\xi}$ be a random vector.

Definition 2.1.2 (Star-shaped sets). *A set $B \subseteq \mathbb{R}^{n_\xi}$ is said to be star-shaped with center 0 if for every $x \in B$, the line segment $[0, x]$ is a subset of B.*

Even though there are several different notions of unimodality in multiple dimensions (linear, convex, log-concave, etc...), our work will be based on the *star-unimodality* definition:

Definition 2.1.3 (Star-unimodality). *A distribution $\mathbb{P} \in \mathcal{P}_\infty$ is called star-unimodal with mode 0 if it belongs to the weak closure of the convex hull of all uniform distributions on star-shaped sets with center 0. The set of all star-unimodal distribution with mode 0 is denoted as $\mathcal{P}_\star$.*

Please note that all definitions of unimodality in multiple dimensions are equivalent in the univariate case [21]. Moreover, the definition of star-unimodality is coherent with our intuitive idea of unimodality when $\mathbb{P}$ has a continuous PDF. In that case it is possible to prove that $\mathbb{P}$ is star unimodal if and only if its PDF is not-increasing along any ray emanating from the origin, which means that the mapping $t \mapsto \mathbb{P}(\xi = t\zeta)$ is non-increasing in $t \in (0, \infty)$ for all $\zeta \neq 0$.

2.2 Choquet Representations

In order to derive tractable reformulations of probability inequalities, we make use of Choquet theory [35] because it provides us with powerful tools for extreme point representations of convex compact subsets of general topological vector spaces. In this work, we are interested in extreme point representations of ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_\infty$.

In order to apply the results of Choquet theory, we must endow $\mathcal{P}_\infty$ with a topology such that its open, closed and compact subsets are well-defined. We assume that $\mathcal{P}_\infty$ is equipped with the weak convergence topology [35] which allows us to construct the Borel σ-algebra on $\mathcal{P}_\infty$: the smallest σ-algebra containing all open subsets of $\mathcal{P}_\infty$. Given an ambiguity set $\mathcal{P}$, we can now define its extreme distributions

Definition 2.2.1 (Extreme distributions). *A distribution $\mathbb{P} \in \mathcal{P}_\infty$ is said to be an extreme point of an ambiguity set $\mathcal{P} \subseteq \mathcal{P}_\infty$ if it is not representable as a strict convex combination of two distinct distributions in $\mathcal{P}$. The set of all extreme points of $\mathcal{P}$ is denoted as $\operatorname{ex}\mathcal{P}$.*

Given the previous definitions, it is possible to define the Choquet representation of set $\mathcal{P}$:

Definition 2.2.2 (Choquet representation). *A weakly closed convex ambiguity set $\mathcal{P} \subseteq \mathcal{P}_\infty$ is said to admit a Choquet representation if for every distribution $\mathbb{P} \in \mathcal{P}$ there exists a Borel probability measure $\mathbb{P}_m$ on $\operatorname{ex}\mathcal{P}$ with*

$$\mathbb{P}(\cdot) = \int_{\operatorname{ex}\mathcal{P}} e(\cdot)\mathbb{P}_m(\mathrm{d}e). \tag{2.1}$$

Please note that this is a generalization of the notion of convex combination: this representation expresses each $\mathbb{P} \in \mathcal{P}$ as a weighted average (mixture) of extreme elements of $\mathcal{P}$. Moreover, as $\mathbb{P}_m$ is a probability measure, the Choquet representation can be viewed as a generalized (infinite) convex combination and $\mathbb{P}_m$ will be refferred to as a mixture distribution. One of the main results of Choquet theory [35] ensures that every convex compact subset of $\mathcal{P}_\infty$ has a Choquet representation of type (2.1). It happens to be the case that convex subsets of $\mathcal{P}_\infty$ sometimes admit explicit Choquet representations even though they are not compact.

The Choquet representation can be reduced to the following definition if the extreme distributions of $\mathcal{P}$ admit a spacial parametrization:

Definition 2.2.3 (Spatial Parametrization). *The set of extreme distributions of a closed convex set $\mathcal{P} \subseteq \mathcal{P}_\infty$ admits a spatial parametrization if $\operatorname{ex}\mathcal{P} = \{e_x : x \in \mathbb{X}\}$ where $x \in \mathbb{R}^l$ parametrizes the extreme distributions of $\mathcal{P}$ and ranges over a closed convex set $\mathbb{X} \subseteq \mathbb{R}^l$ while the mapping $x \mapsto e_x(B)$ is a Borel-measurable function for any fixed Borel set $B \subseteq \mathcal{B}(\mathbb{R}^{n_\xi})$.*

As a consequence, if a convex closed set $\mathbb{P} \subseteq \mathbb{P}_\infty$ has a spatial parametrization, the Choquet representation of any of its elements $\mathbb{P}$, given a mixture distribution $\mathbb{P}_m$, reduces to:

$$\mathbb{P}(\cdot) = \int_{\mathbb{X}} e_x(\cdot)\mathbb{P}_m(\mathrm{d}x).$$

In the case of $\mathcal{P} = \mathcal{P}_\infty$, the extreme points are given by Dirac distributions and thus, the set of extreme distributions is given by $\operatorname{ex}\mathcal{P} = \{\delta_x : x \in \mathbb{R}^{n_\xi}\}$. Furthermore, in this case $\mathcal{P}$ admits a trivial Choquet representation because any $\mathbb{P} \in \mathcal{P}$ is representable as a mixture of Dirac distributions with $\mathbb{P}_m = \mathbb{P}$:

$$\mathbb{P}(\cdot) = \int_{\mathbb{R}^{n_\xi}} \delta_x(\cdot)\mathbb{P}(\mathrm{d}x).$$

2.3 Unimodality Characterization Using a Positive Index

In this section we introduce a generalized notion of unimodality parametrized by a positive index α and we define the Choquet representation of α-unimodal probability distribution sets.

Definition 2.3.1 (α-unimodality [21]). *For any fixed $\alpha > 0$, a random n_ξ-vector ξ is said to have an α-unimodal distribution about 0 if, for every bounded, nonnegative, Borel measurable function g on $\mathbb{R}^n$, the quantity*

$$t^\alpha \mathbb{E}[g(t\xi)] = t^\alpha \int_{\mathbb{R}^{n_\xi}} g(t\xi)\mathbb{P}(\mathrm{d}\xi) \tag{2.2}$$

is nondecreasing in $t \in (0,\infty)$. The set of all α-unimodal distribution is denoted by $\mathcal{P}_\alpha$.

From this definition it can be shown that $\mathcal{P}_\alpha$ is closed under weak convergence. An intuitive way of interpreting this definition comes again from continuous distributions. In the star-unimodal case, the PDF is non-increasing along rays emanating from the origin while in the case of an α-unimodal distribution the PDF can actually increase, but with a rate controlled by α. It is possible to prove [21] that if $\mathbb{P}$ is continuous, it is α-unimodal about 0 if and only if the mapping $t \mapsto t^{n_\xi-\alpha}\,\mathbb{P}\,(\xi = t\zeta)$ is non increasing in $t \in (0,\infty)$ for every fixed $\zeta \neq 0$. In other words, the PDF does not grow faster than $\|\xi\|^{\alpha-n_\xi}$. In the case when $\alpha = n_\xi$, the density is non-increasing along the rays emanating from the origin, which corresponds to the star-unimodality of Definition 2.1.3.

We will introduce the radial α-unimodal distributions that are of key importance in the bound computations.

Definition 2.3.2 (Radial α-unimodal distributions). *For any $\alpha > 0$ and $x \in \mathbb{R}^{n_\xi}$ we denote the radial distribution supported on the line segment $[0,x] \subset \mathbb{R}^{n_\xi}$ as*

$$\delta^\alpha_{[0,x]}([0,tx]) = t^\alpha, \qquad \forall t \in [0,1].$$

By applying Definition 2.3.1, it is possible to prove that $\delta^{\alpha}_{[0,x]} \in \mathcal{P}_{\alpha}$. Let us first note that, from the Definition 2.3.2:

$$\int_{\mathbb{R}^{n_\xi}} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi) = \int_{[0,x]} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi) = \int_0^1 \delta^{\alpha}_{[0,x]}(x\mathrm{d}u) = \int_0^1 \alpha u^{\alpha-1}\mathrm{d}u, \tag{2.3}$$

where in the second inequality we substituted $ux = \xi$ with $u \in [0,1]$. Then, from last expression and Equation (2.2), we can rewrite:

$$\begin{aligned} t^{\alpha} \int_{\mathbb{R}^{n_\xi}} g(t\xi)\mathbb{P}(\mathrm{d}\xi) &= t^{\alpha} \int_{[0,x]} g(t\xi)\delta^{\alpha}_{[0,x]}(d\xi) \\ &= t^{\alpha} \int_0^1 g(tux)\alpha u^{\alpha-1}\mathrm{d}u \\ &= t^{\alpha} \int_0^t g(px)\alpha p^{\alpha-1} t^{1-\alpha} \frac{\mathrm{d}p}{t} \\ &= \int_0^t g(px)\alpha p^{\alpha-1}\mathrm{d}p, \end{aligned}$$

where in the third equality we plugged in $p = tu \in [0,t]$. From the definition of $g(\cdot)$ and the sign of α, the last expression is non-decreasing in $t \in (0,\infty)$ and therefore $\delta^{\alpha}_{[0,x]} \in \mathcal{P}_{\alpha}$.

The first and second order moments of $\delta^{\alpha}_{[0,x]}$ can be computed from its definition.

Lemma 2.3.1 ([45]). *For any $\alpha > 0$ and $x \in \mathbb{R}^{n_\xi}$, the mean value and the second-orded moment matrix of the radial distribution $\delta^{\alpha}_{[0,x]}$ are given by $\frac{\alpha}{\alpha+1}x$ and $\frac{\alpha}{\alpha+2}xx^\top$ respectively.*

Proof. Given $x \in \mathbb{R}^{n_\xi}$, let y be the corresponding vector on a coordinate axis in $\mathbb{R}^{n_\xi}$ from which we obtain x after the coordinate transformation $R \in \mathbb{R}^{n_\xi \times n_\xi}$, i.e. $x = Ry$. The first order moment can be computed, using (2.3), as:

$$\begin{aligned} \mu_{1,\delta^{\alpha}_{[0,x]}} &= \int_{\mathbb{R}^{n_\xi}} \xi \delta^{\alpha}_{[0,x]}(d\xi) = \int_{[0,Ry]} \xi \delta^{\alpha}_{[0,Ry]}(\mathrm{d}\xi) = \\ &= \int_0^1 uRy\alpha u^{\alpha-1}\mathrm{d}u = Ry\alpha \int_0^1 u^{\alpha}\mathrm{d}u = \frac{\alpha}{\alpha+1}Ry = \frac{\alpha}{\alpha+1}x. \end{aligned}$$

The second order moment can be obtained in an analogous way:

$$\begin{aligned} \mu_{2,\delta^{\alpha}_{[0,x]}} &= \int_{\mathbb{R}^{n_\xi}} \xi\xi^\top \delta^{\alpha}_{[0,x]}(d\xi) = \int_{[0,Ry]} \xi\xi^\top \delta^{\alpha}_{[0,Ry]}(\mathrm{d}\xi) = \\ &= \int_0^1 uRyy^\top R^\top u\alpha u^{\alpha-1}\mathrm{d}u = Ryy^\top R^\top \alpha \int_0^1 u^{\alpha+1}\mathrm{d}u \\ &= \frac{\alpha}{\alpha+2}Ryy^\top R^\top = \frac{\alpha}{\alpha+2}xx^\top. \end{aligned}$$

■

The main reason to use the radial α-unimodal distributions comes from the following theorem:

Theorem 2.3.1. *For every $\mathbb{P} \in \mathcal{P}_\alpha$ there exists a unique mixture distribution $\mathbb{P}_m \in \mathcal{P}_\infty$ with*

$$\mathbb{P}(\cdot) = \int_{\mathbb{R}^{n_\xi}} \delta^\alpha_{[0,x]}(\cdot)\mathbb{P}_m(\mathrm{d}x).$$

Proof. The proof can be found in [21, Theorem 3.5]. ■

Corollary 2.3.1. *The radial distributions $\delta^\alpha_{[0,x]}$, $x \in \mathbb{R}^{n_\xi}$ are extremal in $\mathcal{P}_\alpha$.*

Proof. ([45]) If $\delta^\alpha_{[0,x]}$ is not extremal in $\mathcal{P}_\alpha$ for some $x \in \mathbb{R}^n$, there exist $\mathbb{P}^1, \mathbb{P}^2 \in \mathcal{P}_\alpha$ with $\mathbb{P}^1 \neq \mathbb{P}^2$, and $\lambda \in (0,1)$ with $\mathbb{P} = \lambda\,\mathbb{P}^1 + (1-\lambda)\,\mathbb{P}^2$. Thus, the mixture distribution of $\delta^\alpha_{[0,x]}$ can be represented as $\lambda\,\mathbb{P}^1_m + (1-\lambda)\,\mathbb{P}^2_m$, where $\mathbb{P}^1_m$ and $\mathbb{P}^2_m$ are the unique mixture distributions corresponding to $\mathbb{P}^1$ and $\mathbb{P}^2$ respectively. Nonetheless, the unique mixture distribution of $\delta^\alpha_{[0,x]}$ is the Dirac distribution concentrating the unit mass at x, and it cannot be representable as a strict convex combination of to distinct mixture distributions. By consequence, $\delta^\alpha_{[0,x]}$ must be extremal in $\mathcal{P}_\alpha$. ■

Please note that the ambiguity sets $\mathcal{P}_\alpha$ benefit from the nesting property: $\mathcal{P}_\alpha \subseteq \mathcal{P}_\beta$ if and only if $0 < \alpha \leq \beta \leq \infty$. Therefore, it is possible to define the α-unimodality index of a generic ambiguity set $\mathcal{P}$ as the smallest α such that $\mathcal{P} \subseteq \mathcal{P}_\alpha$. In Figure 2.2 there is example of nested α-unimodal sets.

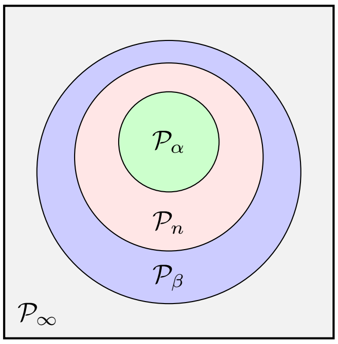

Figure 2.2: Nested α-unimodal ambiguity sets for $\alpha < n < \beta < \infty$.

It can be shown [21] that, given x, as $\alpha \to \infty$ the radial distributions $\delta^\alpha_{[0,x]}$ converge weakly to Dirac distribution δ_x. As every distribution $\mathbb{P} \in \mathcal{P}_\infty$ is representable as a mixture of Dirac distributions with $\mathbb{P}_m = \mathbb{P}$, the weak closure of $\cup_{\alpha>0}\mathcal{P}_\alpha$ corresponds to $\mathcal{P}_\infty$. For this reason, when $\alpha \to \infty$, the probability bounds we are going to derive for α-unimodal distributions, will correspond to the ones for all Borel measurable distributions in $\mathcal{P}_\infty$ (generalized Chebyshev inequalities). On the other hand, when $\alpha \to 0$, the radial distributions weakly converge to δ_0. This means that the "most" α-unimodal distribution is a Dirac distribution in the mode. In Figure 2.3 there is an example of radial distributions in one dimension.

As previously said, when $\alpha = n_\xi$, our definition of α-unimodality coincides with star-unimodality. Moreover, as the dimension $n_\xi \to \infty$, $\mathcal{P}_{n_\xi} = \mathcal{P}_\infty$, see [45]. In other words, according to our definitions, all distributions become n_ξ-unimodal as the dimension of the probability space tends to ∞. An intuitive explaination of this result follows from the observation that in high-dimensional

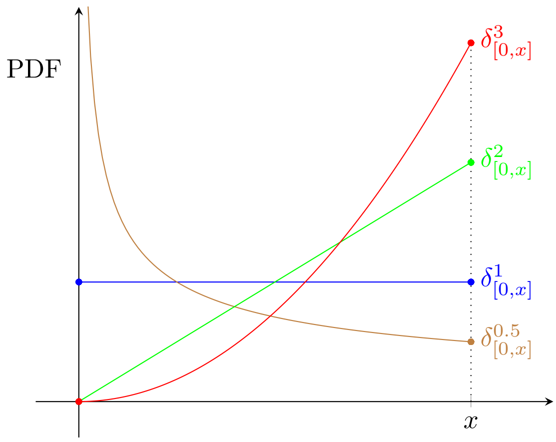

Figure 2.3: Example of four radial α-unimodal distributions in one dimension ($n_\xi = 1$). The n_ξ-unimodal distribution is a uniform one on $[0, x]$. It is clear that, as α increases, the distributions δ_x while as $\alpha \to 0$, the distributions converge to the Dirac one in 0.

star-shaped sets, most of the volume is concentrated in a thin layer near their surface and, for this reason, all the radial distributions $\delta^\alpha_{[0,x]}$ converge weakly to δ_x as n_ξ grows. In Figure 2.4 there is an example of multivariate radial distributions for $n_\xi = 2$.

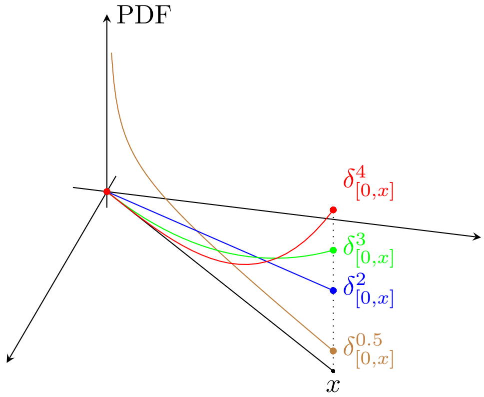

Figure 2.4: Example of multivariate radial α-unimodal distributions. Compared to Figure 2.3 where ($n_\xi = 1$), in this case $n_\xi = 2$ and the n_ξ-unimodal radial distribution density increases from the mode until the vector x with a shape that is already "closer" to a Dirac distribution in x. Furthermore, it is clear that as $\alpha \to \infty$, $\delta^\alpha_{[0,x]}$ distributions converge to a Dirac distribution in x while, when $\alpha \to 0$, they converge to δ_0.

3 Probability Bounds and Ellipsoidal Uncertainty Sets

In this chapter we derive probability bounds for univariate distributions given the first two moments and the α-unimodality index introduced in Chapter 2. Then, we show that it is possible to compute explicitly the minimum volume ellipsoid containing a certain amount of probability mass given only the first two moments and the α-unimodality index of the distribution. Afterwards, this explicit result will be used to generalize the univariate inequalities into a multivariate and data-driven setting.

3.1 Moment Problems

The problem of finding bounds for a set of probability distributions given a finite number of moments, can be classified inside the general class of "moment problems". Henceforth, we describe the generalized moment problems and provide two powerful methods by Popescu [36] and Van Parys et al. [45] to reformulate and solve them.

Generalized Moment Problems Let us define a convex closed probability measure set $\mathcal{P} \subseteq \mathcal{P}_\infty$, $m+1$ measurable moment functions $f_0(\xi), \dots, f_m(\xi)$ and $m+1$ feasible moment sequences $\mu_0, \dots, \mu_m$. Moreover, we define $f_0(\xi) = \mathbf{1}_\Xi$ as the indicator function of the closed measurable convex set $\Xi \subseteq \mathbb{R}^{n_\xi}$ and $\mu_0 = 1$. The generalized moment problem can be expressed as:

$$\begin{aligned} MP_P: \quad & \underset{\mathbb{P}\in\mathcal{P}}{\text{maximize}} && \int_{\mathbb{R}^{n_\xi}} f_0(\xi)\mathbb{P}(d\xi) = \int_\Xi \mathbb{P}(d\xi) = \mathbb{P}(\Xi) \\ & \text{subject to:} && \int_{\mathbb{R}^{n_\xi}} f_i(\xi)\mathbb{P}(d\xi) = \mu_i \qquad \forall i = 1, \dots, m. \end{aligned}$$

Please note that the total probability mass constraint $\int_{\mathbb{R}^{n_\xi}} \mathbb{P}(\mathrm{d}\xi) = \mu_0 = 1$ is implicitly satisfied by all the probability distributions in $\mathcal{P}$. This program is a semi-infinite Linear Program (LP) (see [9]) with finitely many moment constraints but an infinite-dimensional feasible set. The dual of this problem takes the form [36]:

$$\begin{aligned} MP_D: \quad & \underset{\lambda_0,\dots,\lambda_m}{\text{minimize}} && \sum_{i=0}^{m} \lambda_i^T \mu_i \\ & \text{subject to:} && \int_{\mathbb{R}^{n_\xi}} \lambda_0 + \sum_{i=1}^{m} \lambda_i f_i(\xi) - f_0(\xi)\, \mathbb{P}(d\xi) \geq 0 \qquad \forall \mathbb{P} \in \mathcal{P}. \end{aligned}$$

It constitutes a semi-infinite LP with finitely many decision variables and infinitely many constraints parametrized by the distributions $\mathbb{P} \in \mathcal{P}$. Strong duality holds under mild regularity conditions [42].

If the extremal distributions $e_x \in \operatorname{ex}\mathcal{P}$ of the ambiguity set $\mathcal{P}$ admit a spatial parametrization over a semi-algebraic set $\mathbb{X}$, then the semi-infinite dual LP is equivalent to:

$$\begin{aligned} MP_{D_{par}}: \quad & \underset{\lambda_0,\dots,\lambda_m}{\text{minimize}} && \sum_{i=0}^{m} \lambda_i^T \mu_i \\ & \text{subject to:} && \int_{\mathbb{R}^{n_\xi}} \lambda_0 + \sum_{i=1}^{m} \lambda_i^\top f_i(\xi) - f_0(\xi)\, e_x(d\xi) \geq 0 \qquad \forall x \in \mathbb{X}, \end{aligned} \tag{3.1}$$

for the derivation see [36, Lemma 3.1].

α-Unimodal Moment Problems In the moment problems encountered in this work, the first two moments ($m = 2$) of the probability distributions are assumed to be known: the mean value $\mu \in \mathbb{R}^{n_\xi}$ and the second-order moment matrix $\Sigma + \mu\mu^\top \in \mathbb{S}_+^{n_\xi}$, where $\Sigma \in \mathbb{S}_+^{n_\xi}$ is the covariance matrix of ξ. The mode is assumed to be at the origin without loss of generality because this condition can always be enforced by a suitable coordinate transformation. Moreover, the families of α-unimodal probability distributions will be considered. The ambiguity set:

$$\mathcal{P}_\alpha(\mu, \Sigma) \coloneqq \mathcal{P}(\mu, \Sigma) \cap \mathcal{P}_\alpha \tag{3.2}$$

will denote all α-unimodal distributions sharing known μ and Σ. By restricting the ambiguity set to this class of functions, the moment problem MP_P becomes

$$\begin{aligned} MP_{P_\alpha}: \quad & \underset{\mathbb{P}\in\mathcal{P}_\alpha}{\text{maximize}} && \int_{\mathbb{R}^{n_\xi}} f_0(\xi)\mathbb{P}(d\xi) \\ & \text{subject to:} && \int_{\mathbb{R}^{n_\xi}} f_1(\xi)\mathbb{P}(d\xi) = \mu \\ & && \int_{\mathbb{R}^{n_\xi}} f_2(\xi)\mathbb{P}(d\xi) = \Sigma + \mu\mu^\top, \end{aligned} \tag{3.3}$$

while, by considering the radial α-unimodal distributions $\delta^\alpha_{[0,x]}$ from Definition 2.3.2 as extremal distribution in $\operatorname{ex}\mathcal{P}_\alpha$, the parametrized dual problem $MP_{D_{par}}$ can be written as:

$$\begin{aligned} MP_{D_{par,\alpha}}: \quad & \text{minimize} && \langle \lambda_2, \Sigma + \mu\mu^\top \rangle + \lambda_1\mu + \lambda_0 \\ & \text{subject to:} && \lambda_0 \in \mathbb{R},\ \lambda_1 \in \mathbb{R}^{n_\xi},\ \lambda_2 \in \mathbb{S}^{n_\xi} \\ & && \int_{\mathbb{R}^{n_\xi}} \lambda_0 + \lambda_1^\top f_1(\xi) + \langle \lambda_2, f_2(\xi)\rangle - f_0(\xi)\, \delta^\alpha_{[0,x]}(d\xi) \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}, \end{aligned} \tag{3.4}$$

for more details see [42]. Please note that the extremal distributions $\delta^\alpha_{[0,x]}$ admit a spatial parametrization according to Definition 2.2.3 over the set $\mathbb{R}^{n_\xi}$.

Solutions Methods In [36], Popescu showed that the parametric dual problem (3.1) can be solved efficiently for many choices of moment functions and ambiguity sets $\mathcal{P}$ with the corresponding set $\mathbb{X}$, because in these cases the integral in the constraint evaluates to a piecewise polynomial in x. Consequently, the semi-infinite program can be exactly reformulated as a linear matrix inequality (LMI) if $\mathbb{X}$ is one-dimensional or approximated as a hierarchy of increasingly tight LMIs by using sum of squares techniques if $\mathbb{X}$ is multidimensional [33]. Using α-unimodality assumption, this result can be applied also to problem (3.4) with the extreme distributions $\delta^{\alpha}_{[0,x]}$. Thus, the dual problems (3.1) and (3.4) can be systematically reduced to a tractable Semidefinite Program (SDP) that can be solved efficiently with interior point methods [9]. On the other hand, in [45] Van Parys et al. derived an exact SDP reformulation of primal problem (3.3) when the set Ξ is a polyhedron. Both these two reformulations provide efficient tools to construct generalized Chebyshev-like inequalities for unidimensional and multidimensional sets.

3.2 Multivariate Generalized Gauss Inequality

In this section primal and dual reformulations of the α-unimodal moment problem will be adapted for specific types of sets. In particular, the main objective of these derivations will be to bound the maximum probability of the complement of a set: i.e. given a set Φ, the maximum probability of $\Xi = \Phi^c$ has to be computed. These results will give a multidimensional generalization of Gauss inequality.

3.2.1 Dual Program Reformulation

The most common approach to solve moment problems is to compute the dual in order to deal with a finite number of variables. As mentioned before, Popescu in [36] proved that for convex classes of distributions it is possible to reformulate these dual problems as SDPs.

The main idea behind solving the dual problem is to minimize an objective function dependent on given moments and subject to the fact that it is an upper bound to the probability of set Ξ for every distribution of the class $\mathcal{P}_\alpha$ taken into account. Consider the function $\psi\colon \mathbb{R}^{n_\xi} \to \mathbb{R}_+$ defined as:

$$\psi(\xi) = \xi^\top P \xi + q^\top \xi + r,$$

where $P \in \mathbb{S}^{n_\xi}, q \in \mathbb{R}^{n_\xi}$ and $r \in \mathbb{R}$. Independently from the α-unimodality index, the expected value of ψ with respect to all the distributions $\mathbb{P} \in \mathcal{P}(\mu, \Sigma)$ having same mean and covariance is:

$$\begin{aligned} \mathbb{E}\left(\psi(\xi)\right) &= \mathbb{E}\left(\xi^\top P \xi + q^\top \xi + r\right) \\ &= \mathbb{E}\left(\operatorname{tr}\left(P \xi \xi^\top\right)\right) + \mathbb{E}\left(q^\top \xi\right) + r \\ &= \operatorname{tr}\left(P(\Sigma + \mu\mu^\top)\right) + q^\top \mu + r \\ &= \langle P, \Sigma + \mu\mu^\top \rangle + q^\top \mu + r. \end{aligned} \tag{3.5}$$

It is immediate to see that, last expression is equivalent to the objective function of problem (3.4) with $\lambda_0 = r, \lambda_1 = q$ and $\lambda_2 = P$. By setting $f_0 = \mathbf{1}_\Xi$, it is possible to rewrite (3.4) as

$$\begin{aligned} &\text{minimize} && \langle P, \Sigma + \mu\mu^\top \rangle + q^\top \mu + r \\ &\text{subject to:} && P \in \mathbb{S}^{n_\xi},\ q \in \mathbb{R}^{n_\xi},\ r \in \mathbb{R} \\ & && \int_{\mathbb{R}^{n_\xi}} r + q^\top \xi + \xi^\top P \xi - \mathbf{1}_\Xi(\xi)\ \delta^{\alpha}_{[0,x]}(d\xi) \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}. \end{aligned}$$

Please note that the known moments affect only the objective function while the α-unimodality index changes the constraints giving different optima according to its value. By using the moments of $\delta^{\alpha}_{[0,x]}$ in (2.3.1), the first part of the integral can be rewritten using the same algebraic manipulations as in (3.5), as

$$\int_{\mathbb{R}^{n_\xi}} r + q^\top \xi + \xi^\top P \xi \ \delta^{\alpha}_{[0,x]}(d\xi) = \mathbb{E}_{\delta^{\alpha}_{[0,x]}}\left(\psi(\xi)\right) = \frac{\alpha}{\alpha+2}\operatorname{tr}\left(xx^\top P\right) + \frac{\alpha}{\alpha+1} q^\top x + r$$
$$= \frac{\alpha}{\alpha+2} x^\top P x + \frac{\alpha}{\alpha+1} q^\top x + r.$$

where last equality comes from the fact that $\operatorname{tr}\left(xx^\top P\right) = \operatorname{tr}\left(x^\top P x\right) = x^\top P x$. From last equation, problem (3.4) becomes:

$$\begin{aligned} &\text{minimize} && \langle P, \Sigma + \mu\mu^\top \rangle + q^\top \mu + r \\ &\text{subject to:} && P \in \mathbb{S}^{n_\xi},\ q \in \mathbb{R}^{n_\xi},\ r \in \mathbb{R} \\ &&& \frac{\alpha}{\alpha+2} x^\top P x + \frac{\alpha}{\alpha+1} q^\top x + r - \int_{\mathbb{R}^{n_\xi}} \mathbf{1}_{\Xi}(\xi)\ \delta^{\alpha}_{[0,x]}(d\xi) \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}. \end{aligned} \tag{3.6}$$

Depending on the chosen set Ξ, the integral in (3.6) can be computed in different ways often leading to computationally tractable problems.

Let us now consider the case when the set Ξ is the complement of a polyhedron Φ defined by k hyperplanes:

$$\Phi = \left\{\xi \in \mathbb{R}^n : a_i^\top \xi \leq b_i,\ \forall i \in \{1,\dots,k\}\right\}, \quad \textit{and} \quad \Xi = \Phi^{\mathsf{c}}. \tag{3.7}$$

If $x \notin \Xi$, the integral in (3.6) is 0 while if $x \in \Xi$ it can be rewritten using Equation (2.3) as follows:

$$\int_{\mathbb{R}^{n_\xi}} \mathbf{1}_{\Xi}(\xi)\delta^{\alpha}_{[0,x]}(\mathrm{d}\xi) = \int_0^1 \mathbf{1}_{\Xi}(xu)\alpha u^{\alpha-1}\mathrm{d}u = \int_{\frac{b_i}{a_i^\top x}}^1 \alpha u^{\alpha-1}\mathrm{d}u = 1 - \left(\frac{b_i}{a_i^\top x}\right)^{\alpha}, \quad \forall i \in \{1,\dots,k\}. \tag{3.8}$$

Problem (3.6) becomes:

$$\begin{aligned} &\text{minimize} && \langle P, S \rangle + q^\top \mu + r \\ &\text{subject to:} && P \in \mathbb{S}^{n_\xi},\ q \in \mathbb{R}^{n_\xi},\ r \in \mathbb{R} \\ &&& \frac{\alpha}{\alpha+2} x^\top P x + \frac{\alpha}{\alpha+1} q^\top x + r - 1 + \left(\frac{b_i}{a_i^\top x}\right)^{\alpha} \geq 0 \quad \forall i \in \{1,\dots,k\}, \quad \forall x \in \Xi \\ &&& \frac{\alpha}{\alpha+2} x^\top P x + \frac{\alpha}{\alpha+1} q^\top x + r \geq 0, \quad \forall x \in \mathbb{R}^{n_\xi}. \end{aligned} \tag{3.9}$$

Please note that if $x \notin \Xi$, the integral in (3.8) becomes negative. In order to express a constraint equivalent to the one in (3.8) for a fixed x, we have to decouple it in two ones: one that has to be valid whenever the integral in (3.6) is null and another one that becomes tighter if $x \in \Xi$. Thus, for a fixed x, the inequality in problem (3.6) is decoupled in $k+1$ ones: one for each face of the polyhedron plus last one to ensure nonnegativity of the integral in (3.6). Please note that if $\alpha = n_\xi$, from what said in Chapter 2, the problem coincides to the Generalized Gauss bound in n_ξ-dimensions.

It is possible to approximate the inequalities in (3.9) to increasingly tight Linear Matrix Inequality (LMI) using sum-of-squares techniques [33]. However, there are more specific instances, when the solution can be simplified and computed exactly without any approximation.

Generalized Chebyshev Inequalities When $\alpha = \infty$, the problem can be simplified and exactly reformulated as an SDP. In this case the ambiguity set is $\mathcal{P}_\infty(\mu, \Sigma)$ (all Borel measurable distributions on $\mathbb{R}^n$ with mean μ and covariance Σ). By consequence, the unimodality information does not give any advantage and the probability bound corresponds to a generalization of Chebyshev inequality in more dimensions. Problem (3.9) becomes

$$\begin{array}{ll} \text{minimize} & \langle P, S\rangle + q^\top \mu + r \\ \text{subject to:} & P \in \mathbb{S}^{n_\xi},\ q \in \mathbb{R}^{n_\xi},\ r \in \mathbb{R} \\ & x^\top P x + q^\top x + r - 1 \geq 0 \quad \forall x \in \Xi \\ & x^\top P x + q^\top x + r \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}. \end{array} \tag{3.10}$$

The first set of parametric inequalities in x can be translated into this condition:

$$a_i^\top x \geq b_i \Rightarrow x^\top P x + q^\top x + r - 1 \geq 0 \qquad \forall i \in \{1, \ldots, k\} \quad \forall x \in \mathbb{R}^{n_\xi}.$$

By applying the S-procedure (Appendix A.2), it can be rewritten as: there exist $\tau_1, \ldots, \tau_k \geq 0$ such that:

$$\begin{bmatrix} P & q/2 \\ q^\top/2 & r-1 \end{bmatrix} \succeq \tau_i \begin{bmatrix} 0 & a_i/2 \\ a_i^\top/2 & -b_i \end{bmatrix} \qquad \tau_i \geq 0 \quad \forall i \in \{1, \ldots, k\}.$$

The second set of parametric inequalities in x can be also reformulated as an LMI:

$$\begin{bmatrix} P & q/2 \\ q^\top/2 & r \end{bmatrix} \succeq 0.$$

Thus, problem (3.10) can be finally reformulated as:

$$\begin{array}{ll} \text{minimize} & \langle P, S\rangle + q^\top \mu + r \\ \text{subject to:} & P \in \mathbb{S}^{n_\xi}, q \in \mathbb{R}^{n_\xi}, r \in \mathbb{R} \\ & \tau_i \geq 0 \quad \forall i \in \{1, \ldots, k\} \\ & \begin{bmatrix} P & q/2 \\ q^\top/2 & r-1 \end{bmatrix} \succeq \tau_i \begin{bmatrix} 0 & a_i/2 \\ a_i^\top/2 & -b_i \end{bmatrix} \quad \forall i \in \{1, \ldots, k\} \\ & \begin{bmatrix} P & q/2 \\ q^\top/2 & r \end{bmatrix} \succeq 0. \end{array}$$

In [48], Vandenberghe et al. derived the same reformulation without using the α-unimodality framework. In this work it will be shown that for many distributions encountered in theory and practice, this type of bound, even though computationally tractable, leads to pessimistic results.

3.2.2 Primal Program Reformulation

As previously mentioned, another method to deal with moment problems is to directly reformulate the primal problem. Recently Van Parys et al. [45] derived the following SDP reformulation when the set Ξ is a polyhedron defined as in (3.7) and $\alpha \in \mathbb{N}$:

$$\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{k} (\lambda_i - t_{i,0}) \\
\text{subject to:} \quad & z_i \in \mathbb{R}^{n_\xi},\ Z_i \in \mathbb{S}^{n_\xi},\ \lambda_i \in \mathbb{R},\ t_i \in \mathbb{R}^{l+1} \quad \forall i \in \{1,\dots,k\} \\
& \begin{bmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{bmatrix} \succeq 0, \quad a_i^\top z_i \geq 0, \quad t_i \geq 0 \quad \forall i \in \{1,\dots,k\} \\
& \sum_{i=1}^{k} \begin{bmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{bmatrix} \\
& \left\| \begin{bmatrix} 2\lambda_i b_i \\ t_{i,l} b_i - a_i^\top z_i \end{bmatrix} \right\|_2 \leq t_{i,l} b_i + a_i^\top z_i \quad \forall i \in \{1,\dots,k\} \\
& \left\| \begin{bmatrix} 2t_{i,j+1} \\ t_{i,j} - \lambda_i \end{bmatrix} \right\|_2 \leq t_{i,j} + \lambda_i \quad \forall j \in E, \quad \forall i \in \{1,\dots,k\} \\
& \left\| \begin{bmatrix} 2t_{i,j+1} \\ t_{i,j} - t_{i,l} \end{bmatrix} \right\|_2 \leq t_{i,j} + t_{i,l} \quad \forall j \in O, \quad \forall i \in \{1,\dots,k\},
\end{aligned} \tag{3.11}$$

where $l = \lceil \log_2 \alpha \rceil$, $E = \{ j \in \{0,\dots,l-1\} : \lceil \alpha/2^j \rceil \text{ is even} \}$ and $O = \{ j \in \{0,\dots,l-1\} : \lceil \alpha/2^j \rceil \text{ is odd} \}$. For more details and the complete derivation, please see [45]. This reformulation will be used in this work to compute exact probability bounds without resorting to Sum-Of-Squares (SOS) approximations when it is not possible to derive an explicit solution to the dualized moment problem.

3.3 Univariate Generalized Gauss Inequality

In this section the univariate case where $\xi \in \mathbb{R}$ will be studied. Assuming that the mean $\mu \in \mathbb{R}$ is equal to the mode in 0 and that the variance is $\sigma^2 \in \mathbb{R}$, it is possible to derive explicit bounds generalizing the Gauss inequality to the α-unimodality framework.

3.3.1 Double-Sided Problem

The generalized Gauss inequality for α-unimodal distributions is defined in the following theorem:

Theorem 3.3.1 (α-unimodal Gauss Inequality). *Let $\xi \in \mathbb{R}$ be an α-unimodal random variable with given mode equal to its mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$. Then, for every $k \in \mathbb{R}_+$:*

$$\mathbb{P}(|\xi - \mu| > k\sigma) \leq \begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{k^2} & k > \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}} \\ 1 - k^\alpha \left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{2}} & 0 \leq k \leq \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}} \end{cases}. \tag{3.12}$$

Proof. Given $k \in \mathbb{R}_+$, the set Ξ of which we need to bound the probability, is now defined as the complement of the line segment of length $2k$ centered in 0:

$$\Xi := \{\xi \in \mathbb{R} : |\xi| > k\} = \{\xi \in \mathbb{R} : \xi > k \wedge -\xi > k\}.$$

The multivariate problem (3.9) can be reduced to:

$$\begin{aligned}
\text{minimize} \quad & p\sigma^2 + r \\
\text{subject to:} \quad & p, q, r \in \mathbb{R} \\
& \frac{\alpha}{\alpha+2} px^2 + \frac{\alpha}{\alpha+1} qx + r - 1 + \left(\frac{k}{x}\right)^{\alpha} \geq 0 \quad \forall x \in \mathbb{R}_{\geq k} \\
& \frac{\alpha}{\alpha+2} px^2 + \frac{\alpha}{\alpha+1} qx + r - 1 + \left(\frac{k}{-x}\right)^{\alpha} \geq 0 \quad \forall x \in \mathbb{R}_{\leq -k} \\
& \frac{\alpha}{\alpha+2} px^2 + \frac{\alpha}{\alpha+1} qx + r \geq 0, \quad \forall x \notin \Xi.
\end{aligned}$$

As the set Ξ is symmetric, also the function we are trying to minimize will be symmetric. Thus we can set $q = 0$. Moreover, we notice that the elements $(k/x)^\alpha$ when $x \geq k$ and $(k/-x)^\alpha$ when $x \leq -k$ can be translated into a single constraint. Hence, we can be rewrite the problem as follows:

$$\begin{aligned}
\text{minimize} \quad & p\sigma^2 + r \\
\text{subject to:} \quad & p, r \in \mathbb{R} \\
& \frac{\alpha}{\alpha+2} px^2 + r - 1 + \left(\frac{k}{\|x\|}\right)^{\alpha} \geq 0 \quad \forall x \in \mathbb{R} \\
& \frac{\alpha}{\alpha+2} px^2 + r \geq 0, \quad \forall x \notin \mathbb{R}.
\end{aligned}$$

As the problem does not depend on the sign of x, but only on the norm, it is possible to define the new variable $t = \|x\| \in \mathbb{R}_+$ and rewrite it as:

$$\begin{aligned}
\text{minimize} \quad & p\sigma^2 + r \\
\text{subject to:} \quad & p, r \in \mathbb{R} \\
& \frac{\alpha}{\alpha+2} pt^2 + r - 1 + \left(\frac{k}{t}\right)^{\alpha} \geq 0 \quad \forall i \in \{1, \ldots, k\}, \quad \forall t \in \mathbb{R}_+ \\
& \frac{\alpha}{\alpha+2} pt^2 + r \geq 0, \quad \forall t \in \mathbb{R}_+.
\end{aligned} \tag{3.13}$$

The second inequality can be rewritten as $p \geq -\frac{\alpha+2}{\alpha}\frac{r}{t^2}$ where, together with the first inequality, the right-hand side is always a negative number that goes to $-\infty$ as $t \to 0$. Thus, it is always satisfied if $p \geq 0$. In the same way, the first inequality implies that $r \geq 0$.

Let us analyze the first inequality more deeply: the left-hand side has a two local minima in $t^{\alpha-1} = 0$ and in $t = \sqrt{\frac{1-r}{p}}$, both in $\mathbb{R}_+$. The case when $t^{\alpha-1} = 0$ corresponds $t = 0$ when $\alpha \neq 1$ and it is not satisfied when $\alpha = 1$. When $t = 0$, the inequalities become independent from the optimization variables and, thus, the problem becomes unbounded below. We will, then, focus on the case when $t = \sqrt{\frac{1-r}{p}}$ to analyze the minimum of the left-hand side of the first inequality that has to be always greater than 0. We can rewrite the convex feasible region of our program as a system of three inequalities: one to ensure that the minimum of the left-hand side of first inequality in (3.13) is always greater than 0 together with the two ones ensuring the second inequality in (3.13) always holds:

$$p \geq \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{k^2}(1-r)^{\frac{\alpha+2}{\alpha}} \quad \wedge \quad r \geq 0 \quad \wedge \quad p \geq 0. \tag{3.14}$$

In order to obtain the minimum of this convex program with linear objective function we need to find a point where the closure of the feasible region is tangent to the level set of the cost function or, if there are no points with this property, the vertex of the feasible region that minimizes the objective function. In Figure 3.1 the feasible region is plotted for both cases. The negative

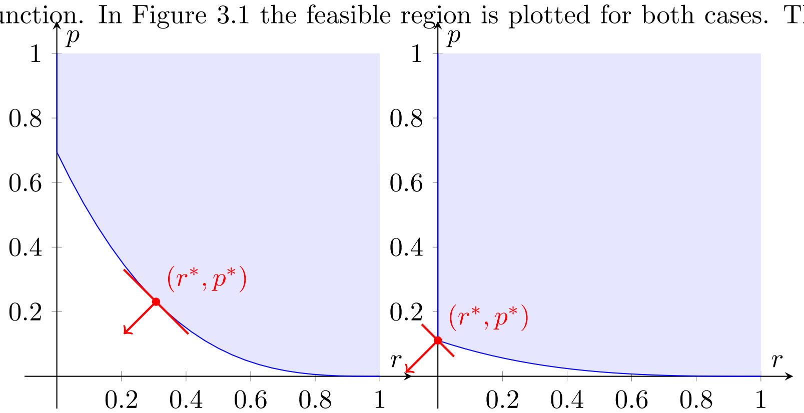

Figure 3.1: Optimal solutions for two different values of k and with $\sigma^2 = 1$ and $\alpha = 1$: on the left $k = 0.8 \leq 2/\sqrt{3}$ and on the right $k = 2 > 2/\sqrt{3}$.

gradient of the cost function is given by $c = [-1, -\sigma^2]^T$. As we do not accept $\sigma^2 = 0$ nor $\sigma^2 = \infty$, the straight lines $r = 0$ and $p = 0$ cannot be tangent to any level set. For $r \leq 1$, the curve

$$p = \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{k^2} (1-r)^{\frac{\alpha+2}{\alpha}} \tag{3.15}$$

is positive and has always a negative derivative given by

$$\frac{\partial p}{\partial r} = -2^{\frac{2}{\alpha}} \alpha^{-1} (\alpha+2)^{1-2/\alpha} \frac{1}{k^2} (1-r)^{2/\alpha} \leq 0 \qquad \forall r \in [0,1]\,.$$

Moreover, if $r \leq 1$, the derivative is always increasing as the second derivative of (3.15) is

$$\frac{\partial^2 p}{\partial r^2} = 2^{1+\frac{2}{\alpha}} \alpha^{-2} (\alpha+2)^{1-\frac{2}{\alpha}} \frac{1}{k^2} (1-r)^{\frac{2}{\alpha}-1} \geq 0, \qquad \forall r \in [0,1]\,.$$

By consequence, the derivative of (3.15) has a minimum inside the feasible region in $r = 0$. If the minimum is greater than $-\frac{1}{\sigma^2}$, then there is no point of the curve tangent to the level sets for $p, r \geq 0$. Thus, the minimum is achieved at the vertex:

$$r^* = 0, \quad p^* = \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{k^2}.$$

If the minimum of the derivative is lower than $-\frac{1}{\sigma^2}$ we have

$$-2^{\frac{2}{\alpha}} \alpha^{-1} (\alpha+2)^{1-2/\alpha} \frac{1}{k^2} \leq -\frac{1}{\sigma^2} \Longrightarrow k \leq 2^{\frac{1}{\alpha}} \alpha^{-\frac{1}{2}} (\alpha+2)^{\frac{\alpha-2}{2\alpha}} (\sigma^2)^{\frac{1}{2}}.$$

Then, the tangent point is reached when

$$-2^{\frac{2}{\alpha}} \alpha^{-1} (\alpha+2)^{1-2/\alpha} \frac{1}{k^2} (1-r)^{2/\alpha} = -\frac{1}{\sigma^2},$$

that gives

$$r^* = 1 - \frac{1}{2}\frac{k^\alpha}{(\sigma^2)^{\frac{\alpha}{2}}}\alpha^{\frac{\alpha}{2}}(\alpha+2)^{1-\frac{\alpha}{2}}, \quad p^* = \frac{1}{2}\frac{k^\alpha}{(\sigma^2)^{1+\frac{\alpha}{2}}}\alpha^{1+\frac{\alpha}{2}}(\alpha+2)^{-\frac{\alpha}{2}}.$$

The optimum in the two cases is given by:

$$\begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}}\frac{\sigma^2}{k^2} & \text{if } k > \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}\sigma \\ 1 - \frac{k^\alpha}{(\sigma^2)^{\frac{\alpha}{2}}}\left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{2}} & \text{if } k \le \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}\sigma \end{cases}.$$

After assuming that the random variable has arbitrary mean μ and subtracting it, we can finally obtain the explicit bound for α-unimodal distributions:

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\sigma^2)} \mathbb{P}(|\xi-\mu|>k) = \begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}}\frac{\sigma^2}{k^2} & k > \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}\sigma \\ 1 - \left(\frac{k}{\sigma}\right)^\alpha\left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{2}} & 0 \le k \le \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}\sigma \end{cases}. \quad (3.16)$$

■

Remark *The two regions switch continuously when*

$$k = \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}\sigma \qquad \textit{with} \qquad \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\sigma^2)} \mathbb{P}(|\xi-\mu|>k) = \frac{\alpha}{\alpha+2}.$$

Proposition 3.3.1. *If $\alpha = n_\xi = 1$, the bound defined in Theorem 3.3.1 corresponds to the Gauss inequality* (1.7) *while, for $\alpha \to \infty$, it corresponds to the Chebyshev inequality* (1.5).

Proof. The case when $\alpha = n_\xi = 1$ can be directly verified from Equation (1.7). As $\alpha \to \infty$, the feasible region (3.14) becomes

$$p \ge \frac{1}{k^2}(1-r) \quad \wedge \quad r \ge 0 \quad \wedge \quad p \ge 0.$$

The first inequality in this case corresponds to an halfspace and the line $p = \frac{1}{k^2}(1-r)$ passes through the points $(0, \frac{1}{k^2})$ and $(1,0)$. It is still convenient to discuss the different values the constant derivative of this line

$$\frac{\partial p}{\partial r} = -\frac{1}{k^2} \le 0,$$

could assume and compare them to the one of the level sets $-\frac{1}{\sigma^2}$. We have now three cases:

$$\begin{array}{l} -\frac{1}{k^2} < -\frac{1}{\sigma^2} \to k < \sqrt{\sigma^2} \Longrightarrow \text{Optimal point } (r^*,p^*) = (1,0) \\ -\frac{1}{k^2} > -\frac{1}{\sigma^2} \to k > \sqrt{\sigma^2} \Longrightarrow \text{Optimal point } (r^*,p^*) = (0,1/k^2) \\ -\frac{1}{k^2} = -\frac{1}{\sigma^2} \to k = \sqrt{\sigma^2} \Longrightarrow \text{Optimal point } (r^*,p^*) \in [(1,0),(0,1/\sigma^2)] \end{array}.$$

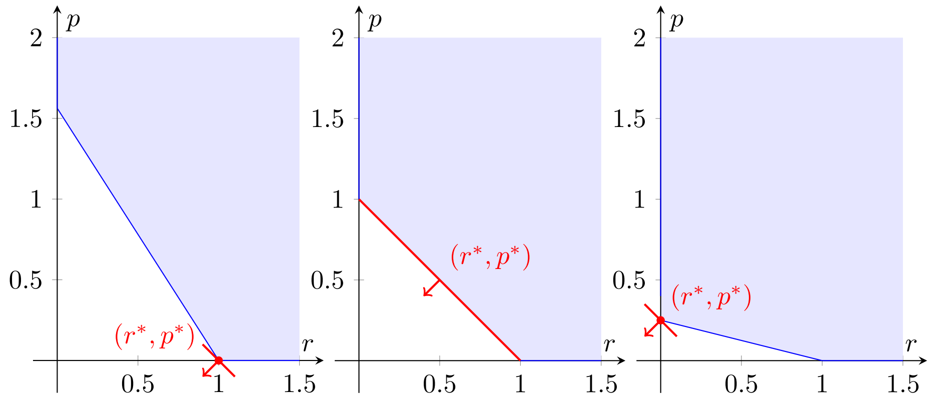

Figure 3.2: Optimal solutions for $\sigma^2 = 1$, $\alpha = \infty$ and $k = 0.9$ (left), $k = 1$ (center), $k = 2$ (right).

The three possible solutions are displayed in Figure 3.2. We can, then, write the bound when $\alpha \to \infty$ and extremal distributions become Dirac couples at $\pm x$:

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty(\mu,\sigma^2)} \mathbb{P}(|\xi - \mu| > k) = \begin{cases} \frac{\sigma^2}{k^2} & k > \sigma \\ 1 & 0 \leq k \leq \sigma \end{cases} \tag{3.17}$$

■

In [21, Theorem 3.10] Dharmadhikari and Joag-Dev already derived a similar but more conservative result than the bound in Theorem 3.3.1, having only one region:

Theorem 3.3.2 (α-unimodal Gauss Inequality with One Region ([21, Theorem 3.10])). *Let $\xi \in \mathbb{R}$ be an α-unimodal random variable with given mode equal to its mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$. Then, for every $k \in \mathbb{R}_+$:*

$$\mathbb{P}(|\xi - \mu| > k) \leq \left(\frac{2}{\alpha + 2}\right)^{\frac{2}{\alpha}} \frac{\sigma^2}{k^2} \quad k \in \mathbb{R}_+. \tag{3.18}$$

In Figure 3.3 the univariate α-unimodal bounds are plotted for different values of α togther with the result from [21].

3.3.2 One-Sided Formulation

Unfortunately, in the one-sided case, problem (3.9) cannot be simplified in such a way to compute an explicit solution as the double-sided inequality (3.16). For this reason, the one-sided version of the inequality will be numerically computed in the remainder of this work using the primal reformulation (3.11) adapted for the following set Ξ:

$$\Xi := \{\xi \in \mathbb{R} : \xi > k\}, \qquad k \in \mathbb{R}_+.$$

Now, Ξ is the half-line (k, ∞). Using the recent results in [45], we can compute the bound on the probability of Ξ as:

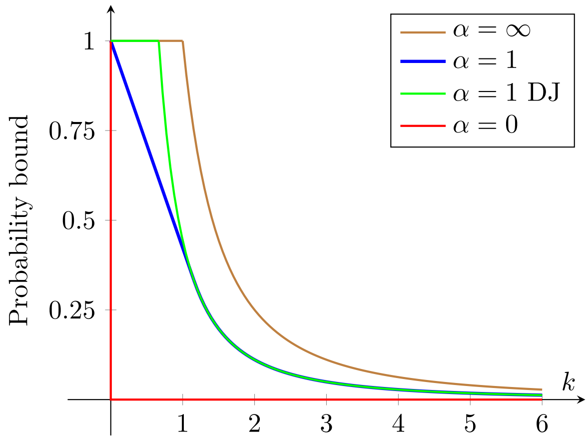

Figure 3.3: Generalized univariate α-unimodal double sided probability bound for different values of α together with the previous bound (DJ) from Dharmadhikari and Joag-Dev in [21].

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}_\infty(\mu,\sigma^2)} \mathbb{P}(\xi-\mu>k) = \text{maximize} \quad & \lambda - t_0 \\
\text{subject to:} \quad & z\in\mathbb{R},\ Z\in\mathbb{R}_+,\ \lambda\in\mathbb{R},\ t\in\mathbb{R}^{l+1} \\
& \begin{bmatrix} Z & z \\ z & \lambda \end{bmatrix} \succeq 0, \quad z\geq 0, \quad t\geq 0 \\
& \begin{bmatrix} Z & z \\ z & \lambda \end{bmatrix} \preceq \begin{bmatrix} \frac{\alpha+2}{\alpha}\sigma^2 & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{bmatrix} \\
& \left\| \begin{bmatrix} 2\lambda k \\ t_l k - z \end{bmatrix} \right\|_2 \leq t_l k + z \\
& \left\| \begin{bmatrix} 2t_{j+1} \\ t_j - \lambda \end{bmatrix} \right\|_2 \leq t_j + \lambda \quad \forall j\in E \\
& \left\| \begin{bmatrix} 2t_{j+1} \\ t_j - t_l \end{bmatrix} \right\|_2 \leq t_j + t_l \quad \forall j\in O,
\end{aligned} \tag{3.19}$$

where $l = \lceil \log_2 \alpha \rceil$, $E = \big\{ j\in\{0,\dots,l-1\} : \lceil \alpha/2^j \rceil \text{ is even} \big\}$ and $O = \{ j\in\{0,\dots,l-1\} : \lceil \alpha/2^j \rceil \text{ is odd} \big\}$.

If the distribution is symmetric with respect to its mean, the one-sided bound could still be explicitly computed because it corresponds to half of the double-sided one (3.16).

In Figure 3.4 there is a comparison between the numerically computed bounds and the explicit ones for symmetric distributions. It is clear from the plot that the added knowledge about symmetry always improves the one-sided bound. Furthermore, when $k<0$ the bound becomes 1 because it is always possible to construct a probability distribution with given α, $\mu=0$ and σ^2 that is entirely contained in the half line $[k,\infty)$.

Moreover, when $\alpha=\infty$, the one-sided bound for arbitrary distributions converges to the Cantelli

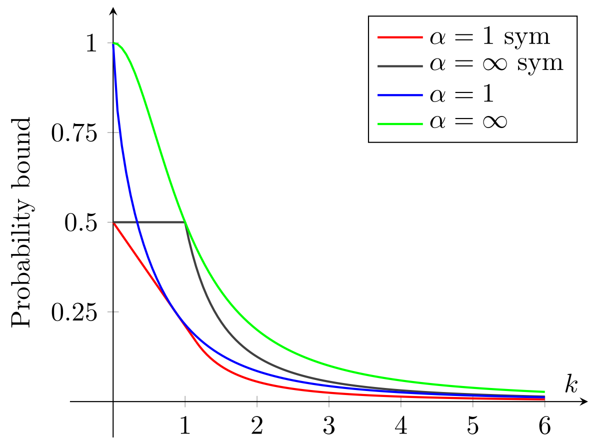

Figure 3.4: Univariate α-unimodal one-sided probability bounds for $\alpha = 1$ and $\alpha = \infty$ and for symmetric and generic distributions.

inequality (1.6) and can be explicitly computed:

$$\sup_{\mathbb{P}\in\mathcal{P}_\infty(\mu,\sigma^2)} \mathbb{P}(\xi - \mu > k) = \frac{\sigma^2}{\sigma^2 + k^2}.$$

3.3.3 Inverse Bounds

In the next chapters it will be necessary to use bounds for standardized distributions with zero mean μ equal to the mode and unit variance $\sigma^2 = 1$.

From Equation (3.16) we define the parametric function $f_{\alpha,ds}\colon \mathbb{R}_+ \mapsto [0,1]$ representing the double-sided bound as

$$f_{\alpha,ds}(k) := \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,1)} \mathbb{P}(|\xi| > k) = \begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{k^2} & k > \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}} \\ 1 - k^\alpha \left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{2}} & 0 \le k \le \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}} \end{cases}. \tag{3.20}$$

Its inverse $f_{\alpha,ds}^{-1}\colon [0,1] \mapsto \mathbb{R}_+$ can be computed explicitly as:

$$\begin{aligned} f_{\alpha,ds}^{-1}(\epsilon) &:= \inf_{k\in\mathbb{R}_+} \left\{ \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,1)} \mathbb{P}(|\xi| > k) \le \epsilon \right\} \\ &= \begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{1}{\epsilon}\right)^{\frac{1}{2}} & 0 \le \epsilon < \frac{\alpha}{\alpha+2} \\ \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}} (1-\epsilon)^{\frac{1}{\alpha}} & \frac{\alpha}{\alpha+2} \le \epsilon \le 1 \end{cases}. \end{aligned} \tag{3.21}$$

The parametric function $f_{\alpha,os}\colon \mathbb{R}_+ \mapsto [0,1]$ for the one-sided bound will be denoted as:

$$f_{\alpha,os}(k) \coloneqq \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,1)} \mathbb{P}(\xi > k), \tag{3.22}$$

and, from what previously discussed in Section 3.3.2, it can be computed numerically by solving problem (3.19). By consequence, its inverse $f_{\alpha,ds}^{-1}\colon [0,1] \mapsto \mathbb{R}_+$ defined as

$$f_{\alpha,os}^{-1}(\epsilon) \coloneqq \inf_{k\in\mathbb{R}_+} \left\{ \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,1)} \mathbb{P}(\xi > k) \leq \epsilon \right\}, \tag{3.23}$$

has to be computed by numerically inverting $f_{\alpha,os}$. It is interesting to see that the limit of the inverse function

$$\lim_{\alpha\to\infty} f_{\alpha,os}^{-1}(\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon}},$$

corresponds to the inverse of Cantelli's bound [15] (one-sided Chebyshev inequality).

In Figure 3.5 the curves $f_{1,os}^{-1}(\epsilon)$ and $f_{\infty,os}^{-1}(\epsilon)$ are plotted.

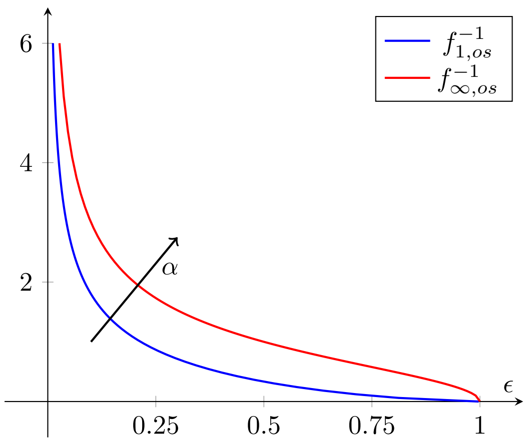

Figure 3.5: Inverse one-sided bounds for univariate α-unimodal distributions with $\alpha = 1$ and $\alpha = \infty$

3.4 Minimum Volume Ellipsoid Containing a Certain Amount of Probability Mass

From what was introduced at the beginning of this work, a commonly used technique to address uncertain optimization problems is to construct sets in which the uncertainty realizes with some probabilistic guarantees and robustify the optimization with respect to all the elements of these sets. In this section, we will derive an explicit formulation to derive the Minimum Volume Ellipsoid (MVE) in $\mathbb{R}^{n_\xi}$ containing at least a $1-\epsilon$ amount of probability mass for all the distributions in our ambiguity set.

Let $\xi \in \mathbb{R}^{n_\xi}$ be a random vector and let its distribution $\mathbb{P} \in \mathcal{P}_\alpha(\mu, \Sigma)$, with $\mu \in \mathbb{R}^{n_\xi}$ be its mean (assumed to be equal to the mode) and $\Sigma \in \mathbb{S}_+^{n_\xi}$ be its covariance matrix.

The problem can be written as

$$\begin{aligned} \underset{\mathcal{E}}{\text{minimize}} \quad & \text{Vol}(\mathcal{E}) \\ \text{subject to:} \quad & \inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}(\mathcal{E}) \geq 1 - \epsilon. \end{aligned} \tag{3.24}$$

It is possible to express the constraint using the complement of $\mathcal{E}$ by inverting the inequality sign as follows:

$$\begin{aligned} \underset{\mathcal{E}}{\text{maximize}} \quad & \text{Vol}(\mathcal{E}) \\ \text{subject to:} \quad & \sup_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^c\right) \leq \epsilon. \end{aligned} \tag{3.25}$$

We are now going to derive a tractable reformulation for the constraint in (3.25) by first centering the ellipsoid at the mean μ and then by imposing a constraint on the matrix describing its shape.

Assumption 3.4.1 (MVE Centering for Symmetric Distributions)**.** *Given the ambiguity set of all symmetric distributions $\mathcal{P}_{\alpha,sym}(\mu, \Sigma)$ on $\mathbb{R}^{n_\xi}$ with mean $\mu \in \mathbb{R}^{n_\xi}$ equal to the mode and covariance matrix $\Sigma \in \mathbb{S}_+^n$, then the MVE containing at least $1 - \epsilon$ of probability mass for all distributions inside the ambiguity set, is centered at μ.*

This assumption is justified from the class of considered distributions: it is reasonable to assume that no offset from μ would increase the probabilistic guarantees of $\mathcal{E}$ for the worst-case symmetric distributions of the ambiguity set $\mathcal{P}_{\alpha,sym}(\mu, \Sigma)$.

The case of asymmetric distributions is described by the following lemma:

Lemma 3.4.1 (MVE Centering)**.** *Given a ambiguity set of distributions on $\mathcal{P}_\alpha(\mu, \Sigma)$ on $\mathbb{R}^{n_\xi}$ with mean $\mu \in \mathbb{R}^{n_\xi}$ equal to the mode and covariance matrix $\Sigma \in \mathbb{S}_+^n$, then the MVE containing at least $1 - \epsilon$ of probability mass for all distributions inside the ambiguity set, is centered at μ.*

Proof. Let us choose ellipsoid $\mathcal{E}_{sym}^*$ being the MVE of the ambiguity set $\mathcal{P}_{\alpha,sym}(\mu, \Sigma)$ such that:

$$\inf_{\mathbb{P} \in \mathcal{P}_{\alpha,sym}(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}_{sym}^*\right) \geq 1 - \epsilon. \tag{3.26}$$

From Assumption 3.4.1, $\mathcal{E}_{sym}^*$ has to be centered at the mode. It can be shown that for every set of asymmetric radial extreme distributions, it is possible to construct a symmetric one by taking half of their length and their symmetric version with respect to the mode and preserving the same mean and covariance. Thus, as the infimum in Equation (3.26) holds over all distributions in $\mathcal{P}_{\alpha,sym}(\mu, \Sigma)$, it holds also over all distributions in $\mathcal{P}_\alpha(\mu, \Sigma)$, i.e.:

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}_{sym}^*\right) \geq 1 - \epsilon.$$

We define $\mathcal{E}^*$ being the MVE for set $\mathcal{P}_\alpha(\mu, \Sigma)$. From problem (3.24), we know that:

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^*\right) \geq 1 - \epsilon.$$

Moreover, as $\mathcal{E}^*$ is the MVE of set $\mathcal{P}_\alpha(\mu, \Sigma)$ and as $\mathcal{E}_{sym}^*$ satisfies Equation (3.4), it holds also that

$$\text{Vol}\left(\mathcal{E}^*\right) \leq \text{Vol}\left(\mathcal{E}_{sym}^*\right).$$

Since $\mathcal{P}_{\alpha,sym}(\mu,\Sigma) \subset \mathcal{P}_{\alpha}(\mu,\Sigma)$, we have that:

$$\inf_{\mathbb{P}\in\mathcal{P}_{\alpha,sym}(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^*\right) \geq \inf_{\mathbb{P}\in\mathcal{P}_{\alpha}(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^*\right) \geq 1-\epsilon.$$

Thus, as $\mathcal{E}^*_{sym}$ is the MVE for set $\mathcal{P}_{\alpha,sym}(\mu,\Sigma)$, we need to have

$$\mathrm{Vol}\left(\mathcal{E}^*\right) \geq \mathrm{Vol}\left(\mathcal{E}^*_{sym}\right).$$

By consequence $\mathcal{E}^*$ and $\mathcal{E}^*_{sym}$ have the same volume. In addition, from the problem construction, two ellipsoids with the same volume and the same probabilistic guarantees have to be symmetric with respect to a line passing through the mode. Since $\mathcal{E}^*_{sym}$ is centered at the mode, also $\mathcal{E}^*$ has to be centered at that point. ■

In the case when the ellipsoid $\mathcal{E}$ is centered at the mean μ, the constraint can be exactly reformulated as an LMI:

Theorem 3.4.1. *Given an ambiguity set $\mathcal{P}_{\alpha}(\mu,\Sigma)$ (all the α-unimodal distributions with mean μ and covariance Σ) and a μ-centered ellipsoid $\mathcal{E}$ defined as $\mathcal{E} = \left\{(x-\mu)^\top C(x-\mu) \leq 1\right\}$, then the probabilistic constraint*

$$\sup_{\mathbb{P}\in\mathcal{P}_{\alpha}(\mu,\Sigma)} \mathbb{P}(\mathcal{E}^c) \leq \epsilon$$

is equivalent to the following linear matrix inequality (LMI) condition:

$$\exists C \in \mathbb{S}^n_+ : \left(f^{-1}_{\alpha,ds}(\epsilon)\right)^2 \langle C, \Sigma\rangle \leq 1,$$

where $f^{-1}_{\alpha,ds} : [0,1] \to \mathbb{R}_+$ is defined in Equation (3.21).

Proof. To simplify the derivation we assume, without loss of generality, that $\mu = 0$. This condition can always be enforced by a suitable coordinate transformation. Hence, by using parametrization (B.8), the MVE can be described by its shape matrix $C \in \mathbb{S}^{n_\xi}_+$ alone:

$$\mathcal{E} = \left\{\xi \in \mathbb{R}^{n_\xi} \mid \xi^\top C\xi \leq 1\right\}.$$

Moreover, the objective function can be rearranged[1] and the optimization problem rewritten as follows:

$$\begin{aligned} &\text{maximize} && \log\det C \\ &\text{subject to:} && C \in \mathbb{S}^{n_\xi}_+ \\ & && \sup_{\mathbb{P}\in\mathcal{P}_{\alpha}(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^{\mathsf{c}}\right) \leq \epsilon. \end{aligned} \tag{3.27}$$

Let us now focus on computing the supremum inside the constraint of problem (3.27). It can be seen as a moment problem (3.6) where we consider $\Xi = \mathcal{E}^{\mathsf{c}}$. As the set $\mathcal{E}^{\mathsf{c}}$ is symmetric around the mode, also the worst-case distributions can be assumed to be symmetric. Thus, we can set $q = 0$ and rewrite (3.6) as:

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}_{\alpha}(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^{\mathsf{c}}\right) = &\text{minimize} && \langle P, \Sigma\rangle + r \\ &\text{subject to:} && P \in \mathbb{S}^{n_\xi},\ r \in \mathbb{R} \\ & && \frac{\alpha}{\alpha+2} x^\top P x + r - \int_{\mathbb{R}^{n_\xi}} \mathbf{1}_{\mathcal{E}^{\mathsf{c}}}(\xi)\ \delta^{\alpha}_{[0,x]}(d\xi) \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}. \end{aligned} \tag{3.28}$$

[1] $\log\sqrt{\det C^{-1}} = -\frac{1}{2}\log\det C$

We first analyze the problem as $\mathcal{E}$ was a unit sphere $\mathcal{S}$ and then we make a coordinate transformation to relate P to $\mathcal{E}$. The transformation is the following: if y belongs to a sphere $\mathcal{S}$ and x to the transformed ellipsoid $\mathcal{E}$ we have that:

$$y^\top I y \leq 1 \quad \xrightarrow[y=C^{1/2}x]{} \quad x^\top C x \leq 1$$

When the indicator function inside problem (3.28) is related to the complement of an unit sphere $\mathcal{S}^\mathsf{c}$, the worst case probability distribution can be taken in the direction of smallest growth of P corresponding to its smallest eigenvalue λ_{min}. If the inequality holds in this case, it will hold in all the other cases being the value of the integral larger and the inequality satisfied.

If $x \in \mathcal{S}$, the integral in (3.28) is 0 while if $x \in \mathcal{S}^\mathsf{c}$ it can be rewritten using Equation (2.3) as follows:

$$\int_{\mathbb{R}^{n_\xi}} \mathbf{1}_{\mathcal{S}^\mathsf{c}}(\xi)\delta^\alpha_{[0,x]}(\mathrm{d}\xi) = \int_0^1 \mathbf{1}_{\mathcal{S}^\mathsf{c}}(xu)\alpha u^{\alpha-1}\mathrm{d}u = \int_{\frac{1}{\|x\|_2}}^1 \alpha u^{\alpha-1}\mathrm{d}u = 1 - \left(\frac{1}{\|x\|_2}\right)^\alpha. \tag{3.29}$$

The constraints in (3.28) can be then rewritten as

$$\frac{\alpha}{\alpha+2}x^\top P x + r - 1 + \left(\frac{1}{\|x\|_2}\right)^\alpha \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}$$

Now we choose x in the direction of the slowest growth of P and write it as $x = \|x\|_2 \, d$ where $d = x/\|x\|_2$ is the directional unit vector. Then,

$$x^\top P x = x^\top (Px) = x^\top (\lambda_{min} x) = \|x\|_2 \underbrace{d^\top d}_{1} \lambda_{min} \|x\|_2 = \lambda_{min} \|x\|_2^2 .$$

The constraints now become:

$$\frac{\alpha}{\alpha+2}\lambda_{min}\|x\|_2^2 + r - 1 + \left(\frac{1}{\|x\|_2}\right)^\alpha \geq 0 \quad \forall x \in \mathbb{R}^{n_\xi}$$

As the problem depends only on the norm of x, it is convenient to define $t = \|x\|_2 \in \mathbb{R}_+$ to parametrize the chance constraints. By introducing λ_{min} as a decision variable and by ensuring that the minimum eigenvalue of P has to be larger than λ_{min} we can rewrite the moment problem as follows:

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}(\mathcal{E}^\mathsf{c}) = \text{minimize} \quad & \langle P, \Sigma\rangle + r \\
\text{subject to:} \quad & P \in \mathbb{S}^{n_\xi},\ r \in \mathbb{R},\ \lambda_{min} \in \mathbb{R}_+ \\
& \frac{\alpha}{\alpha+2}\lambda_{min}t^2 + r - 1 + \left(\frac{1}{t}\right)^\alpha \geq 0 \quad \forall t \in \mathbb{R}_+ \\
& P \succeq \lambda_{min} I
\end{aligned} \tag{3.30}$$

We need to transform the coordinates from sphere $\mathcal{S}$ to ellipse $\mathcal{E}$. From Section 3.2.1, the cost function can be seen as the expected value $\forall\, \mathbb{P} \in \mathcal{P}_\alpha(0,\Sigma)$:

$$\mathbb{E}\left(\xi^\top P \xi + r\right) = \langle P, \Sigma\rangle + r$$

If we apply the coordinate transformation defined in (3.4) by defining $\zeta = C^{1/2}\xi$, we can rewrite the cost function as:

$$\mathbb{E}\left(\zeta^\top \left(C^{1/2}\right)^\top P C^{1/2}\zeta + r\right) = \left\langle \left(C^{1/2}\right)^\top P C^{1/2}, \Sigma \right\rangle + r$$

By applying the same transformation to the generalized inequality we obtain:

$$(C^{1/2})^\top P C^{1/2} \succeq \lambda_{min} C.$$

Then, the moment problem (3.30) can be rearranged as:

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^c\right) = \text{minimize} \quad & \left\langle \left(C^{1/2}\right)^\top P C^{1/2}, \Sigma \right\rangle + r \\ \text{subject to:} \quad & P \in \mathbb{S}^{n_\xi},\ r \in \mathbb{R},\ \lambda_{min} \in \mathbb{R}_+ \\ & \frac{\alpha}{\alpha+2}\lambda_{min} t^2 + r - 1 + \left(\frac{1}{t}\right)^\alpha \geq 0 \quad \forall t \in \mathbb{R}_+ \\ & \left(C^{1/2}\right)^\top P C^{1/2} \succeq \lambda_{min} C. \end{aligned}$$

Finally, by defining $\tilde{P} \coloneqq \frac{1}{\lambda_{min}} \left(C^{1/2}\right)^\top P C^{1/2}$, the problem becomes:

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^c\right) = \text{minimize} \quad & \lambda_{min} \left\langle \tilde{P}, \Sigma \right\rangle + r \\ \text{subject to:} \quad & \tilde{P} \in \mathbb{S}^{n_\xi},\ r \in \mathbb{R},\ \lambda_{min} \in \mathbb{R}_+ \\ & \frac{\alpha}{\alpha+2}\lambda_{min} t^2 + r - 1 + \left(\frac{1}{t}\right)^\alpha \geq 0 \quad \forall t \in \mathbb{R}_+ \\ & \tilde{P} \succeq C. \end{aligned}$$

The two constraints act separately on $\tilde{P}$ and on the other optimization variables λ_{min} and r. Moreover, as $\lambda_{min}, r, \langle \tilde{P}, \Sigma \rangle \geq 0$, it is possible to optimize separately over $\tilde{P}$ and then over the other two variables. The problem

$$\begin{aligned} \text{minimize} \quad & \left\langle \tilde{P}, \Sigma \right\rangle \\ \text{subject to:} \quad & \tilde{P} \succeq C, \end{aligned}$$

clearly admits an unique solution in $\mathbb{P}^* = C$. Thus, by solving this optimization in advance, the moment problem can be rewritten as

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^c\right) = \text{minimize} \quad & \lambda_{min} \langle C, \Sigma \rangle + r \\ \text{subject to:} \quad & r \in \mathbb{R},\ \lambda_{min} \in \mathbb{R}_+ \\ & \frac{\alpha}{\alpha+2}\lambda_{min} t^2 + r - 1 + \left(\frac{1}{t}\right)^\alpha \geq 0 \quad \forall t \in \mathbb{R}_+ \end{aligned}$$

As $\lambda_{min}, r \in \mathbb{R}_+$, this problem is equivalent (3.13) when $\sigma^2 = \langle C, \Sigma \rangle$. By consequence, we have reduced the multidimensional problem into a unidimensional one. By defining $g_\alpha : \mathbb{R}_+ \mapsto [0, 1]$ as

$$g_\alpha(x) \coloneqq \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,x)} \mathbb{P}\left(|\xi| \geq 1\right),$$

with $\xi \in \mathbb{R}$, we can write

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^{\mathsf{c}}\right) = \sup_{\mathbb{P}\in\mathcal{P}_\alpha(0,\langle C,\Sigma\rangle)} \mathbb{P}\left(|\xi|\geq 1\right) = g_\alpha(\langle C,\Sigma\rangle)$$
$$= \begin{cases} \left(\dfrac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \langle C,\Sigma\rangle & \langle C,\Sigma\rangle \leq \left(\dfrac{\alpha+2}{2}\right)^{\frac{2}{\alpha}} \left(\dfrac{\alpha}{\alpha+2}\right) \\ 1-\left(\dfrac{\alpha}{\alpha+2}\dfrac{1}{\langle C,\Sigma\rangle}\right)^{\frac{\alpha}{2}} & \langle C,\Sigma\rangle \geq \left(\dfrac{\alpha+2}{2}\right)^{\frac{2}{\alpha}} \left(\dfrac{\alpha}{\alpha+2}\right) \end{cases}.$$

Using this result it is possible to rewrite the constraint

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}^{\mathsf{c}}\right) \leq \epsilon \qquad \Longleftrightarrow \qquad g_\alpha(\langle C,\Sigma\rangle) \leq \epsilon,$$

or, equivalently, as:

$$\langle C,\Sigma\rangle \leq g_\alpha^{-1}(\epsilon),$$

where $g_\alpha^{-1} : [0,1] \mapsto \mathbb{R}_+$ is defined as

$$g_\alpha^{-1}(\epsilon) = \begin{cases} \left(\dfrac{\alpha+2}{2}\right)^{\frac{2}{\alpha}} \epsilon & 0 \leq \epsilon < \dfrac{\alpha}{\alpha+2} \\ \left(\dfrac{1}{1-\epsilon}\right)^{\frac{2}{\alpha}} \dfrac{\alpha}{\alpha+2} & \epsilon \geq \dfrac{\alpha}{\alpha+2} \end{cases}.$$

From the definition of $f_{\alpha,ds}^{-1}$ in Equation (3.21) as the inverse of the double-sided bound in one dimension, we notice that:

$$g_\alpha^{-1}(\epsilon) = \frac{1}{\left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2}.$$

Finally, by using a coordinate transformation to shift μ to a different value than 0, we complete the proof. ■

Thus, from Lemma 3.4.1 it is possible to directly center $\mathcal{E}$ in μ. In addition, from Theorem 3.4.1, the problem of finding the MVE in (3.25) can be rewritten as the following SDP over the shape matrix C:

$$\begin{aligned} \text{maximize} \quad & \log\det C \\ \text{subject to:} \quad & C \in \mathbb{S}_+^{n_\xi} \\ & \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C,\Sigma\rangle \leq 1. \end{aligned}$$

The solution can be computed explicitly from the KKT conditions. By defining the Lagrange multiplier $\lambda \in \mathbb{R}$ associated to the only constraint, we can write the Lagrangian function as:

$$L(C,\lambda) = -\log\det C + \lambda\left(\left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C,\Sigma\rangle - 1\right)$$

The KKT conditions at the optimum C^*, λ^* are:

$$\begin{cases} \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C,\Sigma\rangle - 1 \leq 0 & \textit{(Primal feasibility)} \\ \lambda^* \geq 0 & \textit{(Dual feasibility)} \\ \lambda^*\left(\left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C,\Sigma\rangle - 1\right) = 0 & \textit{(Complementary Slackness)} \\ \nabla_C L(C^*,\lambda^*) = 0 & \textit{(Stationarity)} \end{cases}$$

α	Volume
0.5	34.31
1	37.23
2	41.89
5	50.76
100	77.44
∞	83.76

Table 3.1: Volumes of the optimal ellipsoids for different values of α.

The last condition can be expressed as:

$$-(C^*)^{-1} + \lambda^* \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \Sigma = 0 \quad \Rightarrow \quad C^* = \frac{\Sigma^{-1}}{\lambda^* \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2},$$

see [9, Section A.4.1]. By plugging C^* into the *Complementary Slackness* condition we get:

$$\lambda^* \left(\left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C^*, \Sigma\rangle - 1\right) = 0 \quad \Rightarrow \quad \lambda^* = n_\xi.$$

Using Lemma 3.4.1, Theorem 3.4.1 and the KKT conditions, we can finally write the explicit form of the MVE in the following Theorem:

Theorem 3.4.2 (Minimum Volume Ellipsoid). *Given the ambiguity set $\mathcal{P}_\alpha(\mu, \Sigma)$ defining all α-unimodal distributions with mode equal to mean $\mu \in \mathbb{R}^{n_\xi}$ and covariance $\Sigma \in \mathbb{S}_+^{n_\xi}$, the Minimum Volume Ellipsoid (MVE) $\mathcal{E}_{MVE}^\alpha$ such that:*

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}_{MVE}^\alpha\right) \geq 1 - \epsilon,$$

is centered in μ and can be written explicitly as:

$$\mathcal{E}_{MVE}^\alpha = \{\xi \in \mathbf{R}^{n_\xi} : \ (\xi - x_c)^\top C(\xi - x_c) \leq 1\},$$

where

$$x_c = \mu, \qquad C = \frac{\Sigma^{-1}}{n \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2}$$

and $f_{\alpha,ds}^{-1}(\epsilon)$ is defined in Equation (3.21).

We show this result for different values of α in the following example:

Example *Let us assume the mode is equal to the zero mean and that the covariance matrix of the ambiguity set is:*

$$\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}.$$

The MVEs for different values of α are shown in Figure 3.6. In Table 3.1 are shown the values of the volume of the optimal ellipsoids.

It is evident that the volume tends to 0 as α tends to 0 as the extreme distributions of the ambiguity set $\mathcal{P}_\alpha(0, \Sigma)$ converge to a Dirac in the origin. Moreover, the volume of the ellipsoid for $\alpha = n = 2$ is circa half of the one for $\alpha = \infty$. Thus, there is a good improvement on the optimal ellipsoids for random vectors of small dimensions.

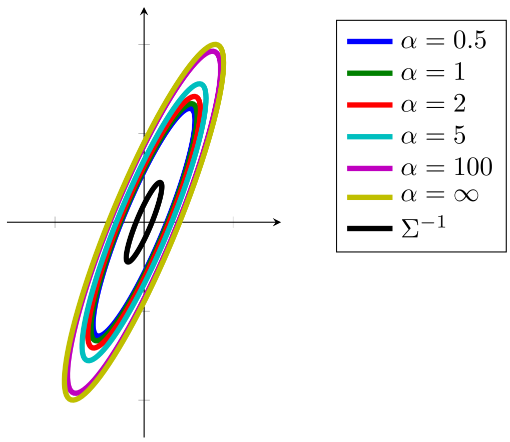

Figure 3.6: Minimum volume ellipsoids for different α, mean in the origin and covariance matrix Σ.

It is interesting to see what happens when we deal with univariate random variables. The minimum volume ellipsoid in that case becomes the minimum length segment centered at the mean that contains at least $1-\epsilon$ probability mass. The mean becomes a scalar μ ($= 0$ $w.l.o.g.$) while the variance σ^2. The matrix C is then:

$$C = \frac{1}{\left(f^{-1}_{\alpha,ds}(\epsilon)\sigma\right)^2} = \frac{1}{d^2}$$

where d is the minimum distance from μ. Thus, $d = f^{-1}_{\alpha,ds}(\epsilon)\sigma$ because it corresponds to the right part of the inequality of the Gauss bound when the random variable is scaled by its standard deviation. d is then the minimum distance because $f^{-1}_{\alpha,ds}(\epsilon)$ is the inverse of the double sided Gauss bound in one dimension for zero mean and standardized variance.

3.5 Multivariate Gauss Inequality over Ellipsoids

In this section the MVE results will be used to generalize the Gauss inequality from Theorem 1.0.3 in multiple dimensions. From Theorem 3.4.2 it is possible to state the following inequality:

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left((\xi-\mu)^\top\Sigma^{-1}(\xi-\mu) > n_\xi\left(f^{-1}_{\alpha,ds}(\epsilon)\right)^2\right) \leq \epsilon. \tag{3.31}$$

If we define $\lambda^2 = n_\xi\left(f^{-1}_{\alpha,ds}(\epsilon)\right)^2$, by Equation (3.21), we can rewrite ϵ as follows

$$\lambda^2 = n_\xi\left(f^{-1}_{\alpha,ds}(\epsilon)\right)^2 \iff \frac{\lambda}{\sqrt{n_\xi}} = f^{-1}_{\alpha,ds}(\epsilon) \iff \epsilon = f_{\alpha,ds}\left(\frac{\lambda}{\sqrt{n_\xi}}\right),$$

where $f_{\alpha,ds}$ is one-dimensional Gauss bounds for distributions with 0 mean and variance 1 defined in Equation (3.20). Equation (3.31) can be rewritten as:

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left((\xi-\mu)^\top\Sigma^{-1}(\xi-\mu) > \lambda^2\right) \leq f_{\alpha,ds}\left(\frac{\lambda}{\sqrt{n_\xi}}\right). \tag{3.32}$$

From this result it is immediate to derive the generalization of Gauss inequality for α-unimodal distributions in multiple dimensions:

Theorem 3.5.1 (Mutivariate Gauss Inequality over Ellipsoids). *Let $\xi \in \mathbb{R}^{n_\xi}$ be an α-unimodal random variable with mean $\mu \in \mathbb{R}^{n_\xi}$ and covariance $\Sigma \in \mathbb{S}_+^{n_\xi}$. Then, for every $\lambda^2 \in \mathbb{R}_+$, the following holds:*

$$\mathbb{P}\left((\xi-\mu)^\top \Sigma^{-1}(\xi-\mu) > \lambda^2\right) \le f_{\alpha,ds}\left(\frac{\lambda}{\sqrt{n_\xi}}\right), \tag{3.33}$$

where $f_{\alpha,ds}$ is defined in Equation (3.20).

Please note that, as $\alpha \to \infty$, the right-hand side of (3.33) tends to the Chebyshev bound for distributions with 0 mean and variance 1:

$$\lim_{\alpha\to\infty} f_{\alpha,ds}\left(\frac{\lambda}{\sqrt{n_\xi}}\right) = \begin{cases} \frac{n_\xi}{\lambda^2} & \frac{\lambda}{\sqrt{n_\xi}} > 1 \\ 1 & 0 \le \frac{\lambda}{\sqrt{n_\xi}} \le 1 \end{cases} = \min\left(1, \frac{n_\xi}{\lambda^2}\right). \tag{3.34}$$

3.6 Multivariate Ellipsoidal Sampled Chebyshev Inequality

It has been shown [38] that it is possible to define a Chebyshev-like inequality in one dimension based on the sampled mean and covariance with only the weak exchangeability assumption on the distribution. In this section we will extend the result to a multivariate setting using the Euclidean norm.

Theorem 3.6.1. *Let $\xi \in \mathbb{R}^{n_\xi}$ be a random variable and let $N \in \mathbb{Z}_{\ge n_\xi}$. Given $\xi^{(1)}, \dots, \xi^{(N)}, \xi^{(N+1)} \in \mathbb{R}^{N_\xi}$ i.i.d. samples of ξ, define the empirical mean and covariance of the first n samples as:*

$$\hat{\xi} := \frac{1}{N}\sum_{i=1}^{N} \xi^{(i)}, \qquad \hat{\Sigma} := \frac{1}{N-1}\sum_{i=1}^{N}(\xi^{(i)}-\hat{\xi})(\xi^{(i)}-\hat{\xi})^\top.$$

If we assume that $\hat{\Sigma}$ is nonsingular, then for all $\lambda \in \mathbb{R}_{++}$ it holds that:

$$\mathbb{P}^{N+1}\left(\left\{\xi^{(1)}, \dots, \xi^{(N+1)}\right\} \in \mathbb{R}^{n_\xi(N+1)} : (\xi^{(N+1)}-\hat{\xi})^\top \hat{\Sigma}^{-1}(\xi^{(N+1)}-\hat{\xi}) > \lambda^2\right) \le h(N, n_\xi, \lambda^2), \tag{3.35}$$

where

$$h(N, n_\xi, \lambda^2) = \min\left(1, \frac{n_\xi(N^2-1+N\lambda^2)}{N^2\lambda^2}\right).$$

Remark *As $N \to \infty$, the right-hand side of the inequality tends to*

$$\min\left(1, \frac{n_\xi}{\lambda^2}\right),$$

that corresponds to the Multivariate Gauss inequality over Ellipsoids when $\alpha \to \infty$ in Equation (3.34).

Remark *The assumption that $\hat{\Sigma}$ is nonsingular can be justified in practice. In the case when ξ is a continuous random variable, the true covariance $\Sigma \in \mathbb{S}_+^{n_\xi}$ is nonsingular and the number of samples is greater than n_ξ, it can be shown that the probability of having a singular $\hat{\Sigma}$ is null.*

3.6.1 Proof of the Inequality

The proof is based on the following preliminary result.

Lemma 3.6.1. *Let $k \in \mathbb{R}_{++}$ and $N \in \mathbb{Z}_{\geq 2}$. Consider a set of vectors $U := \{u_1, \ldots, u_N\}$ with $u_i \in \mathbb{R}^{n_\xi}$ for all $i \in \{1, \ldots, N\}$. Moreover, let us assume that*

$$\sum_{i=1}^{N} u_i = 0_{n_\xi}, \qquad \sum_{i=1}^{N} u_i u_i^\top = N I_{n_\xi \times n_\xi},$$

and define the set

$$J := \{i \in \{1, \ldots, N\} \ : \ \|u_i\|_2 > k\}.$$

Then we have

$$|J| \leq \left\lfloor \frac{n_\xi N}{k^2} \right\rfloor. \tag{3.36}$$

Proof. We first note that

$$\|u_i\|_2 > k \quad \iff \quad u_i^\top u_i > k^2.$$

If we sum both sides of last inequality over all the elements of J we get:

$$k^2 |J| < \sum_{i \in J} u_i^\top u_i \leq \sum_{i=1}^{N} u_i^\top u_i = \mathrm{tr}\left(\sum_{i=1}^{N} u_i u_i^\top\right) = n_\xi N$$

and the result follows immediately. ∎

We now define the empirical sample mean and the biased covariance matrix based of all $N+1$ samples as:

$$\hat{\xi}^* := \frac{1}{N+1} \sum_{i=1}^{N+1} \xi_i, \qquad \hat{\Sigma}^* := \frac{1}{N+1} \sum_{i=1}^{N+1} (\xi^{(i)} - \hat{\xi})(\xi^{(i)} - \hat{\xi})^\top.$$

Proposition 3.6.1. *The following holds:*

$$\xi^{(N+1)} - \hat{\xi}^* = \frac{N}{N+1}(\xi^{(N+1)} - \hat{\xi}) \tag{3.37}$$

$$\hat{\Sigma}^* = \frac{1}{(N+1)^2}\left[(N^2-1)\hat{\Sigma} + N(\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top\right]. \tag{3.38}$$

Proof. The first equality can be obtained from the following algebraic steps:

$$\begin{aligned}
\xi^{(N+1)} - \hat{\xi}^* &= \xi^{(N+1)} - \frac{1}{N+1}\left(\sum_{i=1}^{N}\{\xi^{(i)}\} + \xi^{(N+1)}\right) \\
&= \frac{N}{N+1}\xi^{(N+1)} - \frac{1}{N+1}\sum_{i=1}^{N}\xi^{(i)} \\
&= \frac{N}{N+1}\left(\xi^{(N+1)} - \hat{\xi}\right).
\end{aligned}$$

The second inequality can be proven as follows

$$
\begin{aligned}
\hat{\Sigma}^* &= \frac{1}{N+1}\sum_{i=1}^{N+1}(\xi^{(i)} - \hat{\xi}^*)(\xi^{(i)} - \hat{\xi}^*)^\top \\
&= \frac{1}{N+1}\left(\sum_{i=1}^{N+1}\left\{\xi^{(i)}\xi^{(i)\top}\right\} - \hat{\xi}^*\sum_{i=1}^{N+1}\left\{\xi^{(i)\top}\right\} - \sum_{i=1}^{N+1}\left\{\xi^{(i)}\right\}\hat{\xi}^{*\top} + (N+1)\hat{\xi}^*\hat{\xi}^{*\top}\right) \\
&= \frac{1}{N+1}\left(\sum_{i=1}^{N+1}\left\{\xi^{(i)}\xi^{(i)\top}\right\} - (N+1)\hat{\xi}^*\hat{\xi}^{*\top} - (N+1)\hat{\xi}^*\hat{\xi}^{*\top} + (N+1)\hat{\xi}^*\hat{\xi}^{*\top}\right) \\
&= \frac{1}{N+1}\left(\sum_{i=1}^{N+1}\left\{\xi^{(i)}\xi^{(i)\top}\right\} - (N+1)\hat{\xi}^*\hat{\xi}^{*\top}\right) \\
&= \frac{1}{N+1}\left(\sum_{i=1}^{N}\{\xi^{(i)}\xi^{(i)\top}\} + \xi^{(N+1)}\xi^{(N+1)\top}\right. \\
&\qquad\left. - \frac{1}{N+1}\left(\sum_{i=1}^{N}\{\xi^{(i)}\} + \xi^{(N+1)}\right)\left(\sum_{i=1}^{N}\{\xi^{(i)}\} + \xi^{(N+1)}\right)^\top\right) \\
&= \frac{1}{N+1}\left(\sum_{i=1}^{N}\{\xi^{(i)}\xi^{(i)\top}\} + \xi^{(N+1)}\xi^{(N+1)\top}\right. \\
&\qquad\left. - \frac{1}{N+1}\left(N^2\hat{\xi}\hat{\xi}^\top + \xi^{(N+1)}\xi^{(N+1)\top} + N\xi^{(N+1)}\hat{\xi}^\top + N\hat{\xi}\xi^{(N+1)\top}\right)\right),
\end{aligned}
$$

we now add and subtract the quantity $\frac{N}{N-1}\hat{\xi}\hat{\xi}^\top$ and multiply and divide by $N-1$:

$$
\begin{aligned}
&= \frac{N-1}{N+1}\Bigg(\underbrace{\frac{1}{N-1}\sum_{i=1}^{N}\{\xi^{(i)}\xi^{(i)\top}\} - \frac{N}{N-1}\hat{\xi}\hat{\xi}^\top}_{\hat{\Sigma}} + \frac{N}{N-1}\hat{\xi}\hat{\xi}^\top + \frac{1}{N-1}\xi^{(N+1)}\xi^{(N+1)\top} \\
&\qquad - \frac{1}{(N+1)(N-1)}\left(N^2\hat{\xi}\hat{\xi}^\top + \xi^{(N+1)}\xi^{(N+1)\top} + N\xi^{(N+1)}\hat{\xi}^\top + N\hat{\xi}\xi^{(N+1)\top}\right)\Bigg) \\
&= \frac{N-1}{N+1}\hat{\Sigma} + \frac{N}{N+1}\hat{\xi}\hat{\xi}^\top + \frac{1}{N+1}\xi^{(N+1)}\xi^{(N+1)\top} - \frac{N^2}{(N+1)^2}\hat{\xi}\hat{\xi}^\top \\
&\qquad - \frac{1}{(N+1)^2}\xi^{(N+1)}\xi^{(N+1)\top} - \frac{N}{(N+1)^2}\xi^{(N+1)}\hat{\xi}^\top - \frac{N}{(N+1)^2}\hat{\xi}\xi^{(N+1)\top} \\
&= \frac{1}{(N+1)^2}\Big((N^2-1)\hat{\Sigma} + N(N+1)\hat{\xi}\hat{\xi}^\top + (N+1)\xi^{(N+1)}\xi^{(N+1)\top} - N^2\hat{\xi}\hat{\xi}^\top \\
&\qquad - \xi^{(N+1)}\xi^{(N+1)\top} - N\xi^{(N+1)}\hat{\xi}^\top - N\hat{\xi}\xi^{(N+1)\top}\Big) \\
&= \frac{1}{(N+1)^2}\left((N^2-1)\hat{\Sigma} + N\hat{\xi}\hat{\xi}^\top + N\xi^{(N+1)}\xi^{(N+1)\top} - N\xi^{(N+1)}\hat{\xi}^\top - N\hat{\xi}\xi^{(N+1)\top}\right) \\
&= \frac{1}{(N+1)^2}\left((N^2-1)\hat{\Sigma} + N(\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top\right).
\end{aligned}
$$

■

Since we assumed that $\hat{\Sigma} \succ 0$, it follows from Equation (3.38) that $\hat{\Sigma}^* \succ 0$. We next normalize each of the $N+1$ samples $\xi^{(i)}$ as follows

$$u_i\left(\xi^{(1)}, \ldots, \xi^{(N+1)}\right) := \left(\hat{\Sigma}^*\right)^{-1/2}\left(\xi^{(i)} - \hat{\xi}^*\right) \qquad \forall i \in \{1, \ldots, N+1\} \tag{3.39}$$

so that we get

$$\sum_{i=1}^{N+1} u_i\left(\xi^{(1)}, \ldots, \xi^{(N+1)}\right) = 0_d,$$
$$\sum_{i=1}^{N+1} u_i\left(\xi^{(1)}, \ldots, \xi^{(N+1)}\right) u_i\left(\xi^{(1)}, \ldots, \xi^{(N+1)}\right)^\top = (N+1) I_{n_\xi \times n_\xi}.$$

Let us write u_i instead of $u_i(\xi^{(1)}, \ldots, \xi^{(N+1)})$ for notational convenience. Please notice that as we assumed $\xi^{(i)}$ to be i.i.d., also u_i will be i.i.d. with $i \in \{1, \ldots, N+1\}$.

We now partition the space $\mathbb{R}^{n_\xi(N+1)}$ into different $\tilde{U}\left(\tilde{u}_{(1)}, \ldots, \tilde{u}_{(N+1)}\right)$ each one having fixed ordered vectors $\left\{\tilde{u}_{(1)}, \ldots, \tilde{u}_{(N+1)}\right\}$, with $\tilde{u}_{(i)} \in \mathbb{R}^{n_\xi}$ for all $i \in \{1, \ldots, N+1\}$, such that $\left\|\tilde{u}_{(1)}\right\|_2 \leq \cdots \leq \left\|\tilde{u}_{(N+1)}\right\|_2$. Each one of these partitions represents all possible vector sets $\{u_1, \ldots, u_{N+1}\}$ that can be reordered as $\left\{u_{(1)} = \tilde{u}_{(1)}, \ldots, u_{(N+1)} = \tilde{u}_{(N+1)}\right\}$, see Figure 3.7.

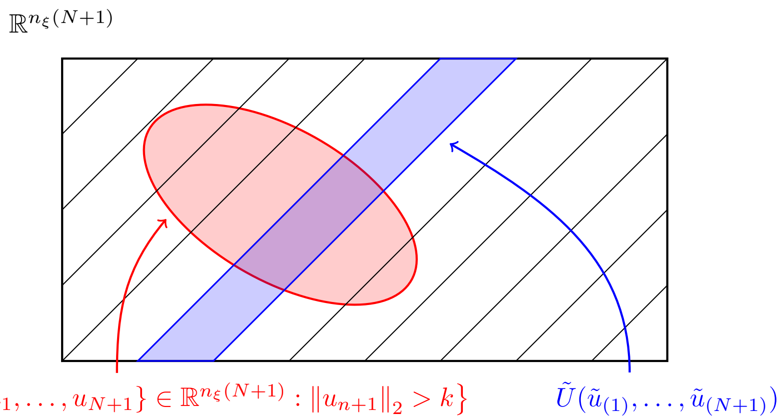

Figure 3.7: The partitioning is intuitively described by a finite number of subsets of $\mathbb{R}^{n_\xi(N+1)}$. The goal of our computations is to bound the probability of set $\left\{\{u_1, \ldots, u_{N+1}\} \in \mathbb{R}^{n_\xi(N+1)} : \|u_{N+1}\|_2 > k\right\}$. The conditioning corresponds to constraining the space into one of the partitions drawn as the oblique stripes, e.g. the blue region. The intersection of the blue and the red regions corresponds to the set of which we compute the conditioned probability.

Let $\left\{\bar{u}_{(1)}, \ldots, \bar{u}_{(N+1)}\right\} \in \mathbb{R}^{n_\xi(N+1)}$ and the corresponding $\overline{U}\left(\bar{u}_{(1)}, \ldots, \bar{u}_{(N+1)}\right)$ be a fixed partition of $\mathbb{R}^{n_\xi}$. We define the distribution $\mathbb{P}_{\overline{U}}^{N+1}$ on $\mathbb{R}^{n_\xi(N+1)}$ such that:

$$\mathbb{P}_{\overline{U}}^{N+1}\left(\left\{y \in \left\{\bar{u}_{(1)}, \ldots, \bar{u}_{(N+1)}\right\} : y = t\right\}\right) := \frac{\left|\{i \in \{1, \ldots, N+1\} : \bar{u}_{(i)} = t\}\right|}{N+1}.$$

As all the u_i with $i \in \{1, \dots, N+1\}$ are exchangeable, the probability of a set $\{u_1, \dots, u_{N+1}\}$ such that $\|u_{N+1}\|_2 > k$, given the partition $\overline{U}\left(\bar{u}_{(1)}, \dots, \bar{u}_{(N+1)}\right)$, is equal to the probability of having one of the elements of $\left\{\bar{u}_{(1)}, \dots, \bar{u}_{(N+1)}\right\}$ that is greater than k, i.e.:

$$\begin{aligned}
&\mathbb{P}^{N+1}\left(\{u_1, \dots, u_{N+1}\} \in \mathbb{R}^{n_\xi(N+1)} : \|u_{N+1}\|_2 > k \;\middle|\; \overline{U}\left(\bar{u}_{(1)}, \dots, \bar{u}_{(N+1)}\right)\right) \\
&\quad = \mathbb{P}^{N+1}_{\overline{U}}\left(y \in \left\{\bar{u}_{(1)}, \dots, \bar{u}_{(N+1)}\right\} : \|y\|_2 > k\right) \\
&\quad = \frac{1}{N+1}\left|\left\{i \in \{1, \dots, N+1\} : \left\|\bar{u}_{(i)}\right\|_2 > k\right\}\right| \\
&\quad \leq \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor,
\end{aligned} \tag{3.40}$$

where last inequality comes from Lemma 3.6.1.

In order to remove the conditioning, we integrate the distribution over all sets $\tilde{U}(\tilde{u}_{(1)}, \dots, \tilde{u}_{(N+1)})$:

$$\begin{aligned}
&\mathbb{P}^{N+1}\left(\{u_1, \dots, u_{N+1}\} \in \mathbb{R}^{n_\xi(N+1)} : \|u_{N+1}\|_2 > k\right) \\
&\quad \leq \int_{\left\{\tilde{u}_{(1)}, \dots, \tilde{u}_{(N+1)}\right\} \in \mathbb{R}^{n_\xi(N+1)}} \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor \mathbb{P}^{N+1}\left(\tilde{U}\left(\mathrm{d}\tilde{u}_1, \dots, \mathrm{d}\tilde{u}_{N+1}\right)\right) \\
&\quad = \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor \underbrace{\int_{\left\{\tilde{u}_{(1)}, \dots, \tilde{u}_{(N+1)}\right\} \in \mathbb{R}^{n_\xi(N+1)}} \mathbb{P}^{N+1}\left(\tilde{U}\left(\mathrm{d}\tilde{u}_1, \dots, \mathrm{d}\tilde{u}_{N+1}\right)\right)}_{1} \\
&\quad = \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor,
\end{aligned}$$

where in the inequality we used Equation (3.40). We are going to rearrange the following inequality

$$\mathbb{P}^{N+1}\left(\{u_1, \dots, u_{N+1}\} \in \mathbb{R}^{n_\xi(N+1)} : \|u_{N+1}\|_2 > k\right) \leq \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor, \tag{3.41}$$

so that we get the one in Theorem 3.6.1. Let us rewrite the complement of the argument inside the probability operator by using Equation (3.39):

$$\begin{aligned}
\|u_{N+1}\|_2 \leq k &\iff u_{N+1}^\top u_{N+1} \leq k^2 \\
&\iff (\xi^{(N+1)} - \hat{\xi}^*)^\top (\hat{\Sigma}^*)^{-1} (\xi^{(N+1)} - \hat{\xi}^*) \leq k^2
\end{aligned}$$

By applying the Schur complement (Appendix A.1) last inequality, together with the assumption of positive definiteness of $\hat{\Sigma}^*$, we get

$$\begin{bmatrix} \hat{\Sigma}^* & \xi^{(N+1)} - \hat{\xi}^* \\ \left(\xi^{(N+1)} - \hat{\xi}^*\right)^\top & k^2 \end{bmatrix} \succeq 0.$$

This means that all the leading principal minors of the matrix are positive semidefinite:

$$\hat{\Sigma}^* \succeq 0 \qquad \wedge \qquad \hat{\Sigma}^* k^2 - (\xi^{(N+1)} - \hat{\xi}^*)(\xi^{(N+1)} - \hat{\xi}^*)^\top \succeq 0.$$

It is possible to rewrite these condition involving $\hat{\xi}$ and $\hat{\Sigma}$ based only on the first N samples by plugging in Equations (3.37) and (3.38) as follows:

$$\begin{aligned}
&\hat{\Sigma}^* k^2 - (\xi^{(N+1)} - \hat{\xi}^*)(\xi^{(N+1)} - \hat{\xi}^*)^\top \succeq 0 \\
&\iff \frac{1}{(N+1)^2}\left[(N^2-1)\hat{\Sigma} + N(\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top\right] k^2 \\
&\qquad\qquad - (\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top \frac{N^2}{(N+1)^2} \succeq 0 \\
&\iff \hat{\Sigma}\frac{(N^2-1)k^2}{N(N-k^2)} - (\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top \succeq 0.
\end{aligned}$$

By noting that $\hat{\Sigma} \succeq 0$, these two conditions

$$\hat{\Sigma} \succeq 0 \qquad \wedge \qquad \hat{\Sigma}\frac{(N^2-1)k^2}{N(N-k^2)} - (\xi^{(N+1)} - \hat{\xi})(\xi^{(N+1)} - \hat{\xi})^\top \succeq 0,$$

correspond to this matrix inequality:

$$\begin{bmatrix} \hat{\Sigma} & \xi^{(N+1)} - \hat{\xi} \\ (\xi^{(N+1)} - \hat{\xi})^\top & \frac{(N^2-1)k^2}{N(N-k^2)} \end{bmatrix} \succeq 0.$$

By using again the Schur complement, the condition can be translated to:

$$\hat{\Sigma} \succeq 0 \qquad \wedge \qquad (\xi^{(N+1)} - \hat{\xi})^\top \hat{\Sigma}^{-1} (\xi^{(N+1)} - \hat{\xi}) \leq \frac{(N^2-1)k^2}{N(N-k^2)}.$$

Thus, we have obtained the following equivalent conditions:

$$\begin{aligned}
\|u_{N+1}\|_2 \leq k &\iff (\xi^{(N+1)} - \hat{\xi}^*)^\top (\hat{\Sigma}^*)^{-1} (\xi^{(N+1)} - \hat{\xi}^*) \leq k^2 \\
&\iff (\xi^{(N+1)} - \hat{\xi})^\top \hat{\Sigma}^{-1} (\xi^{(N+1)} - \hat{\xi}) \leq \frac{(N^2-1)k^2}{N(N-k^2)},
\end{aligned}$$

from which it is immediate to obtain the complementary ones

$$\begin{aligned}
\|u_{N+1}\|_2 > k &\iff (\xi^{(N+1)} - \hat{\xi}^*)^\top (\hat{\Sigma}^*)^{-1} (\xi^{(N+1)} - \hat{\xi}^*) > k^2 \\
&\iff (\xi^{(N+1)} - \hat{\xi})^\top \hat{\Sigma}^{-1} (\xi^{(N+1)} - \hat{\xi}) > \frac{(N^2-1)k^2}{N(N-k^2)}.
\end{aligned}$$

We can now rewrite Equation (3.41) as:

$$\mathbb{P}^{N+1}\left(\left\{\xi^{(1)}, \ldots, \xi^{(N+1)}\right\} : (\xi^{(N+1)} - \hat{\xi})^\top \hat{\Sigma}^{-1} (\xi^{(N+1)} - \hat{\xi}) > \frac{(N^2-1)k^2}{N(N-k^2)}\right) \leq \frac{1}{N+1}\left\lfloor \frac{n_\xi(N+1)}{k^2} \right\rfloor. \tag{3.42}$$

Let us define λ such that

$$\lambda^2 = \frac{(N^2-1)k^2}{N(N-k^2)} \qquad \text{so that} \qquad k^2 = \frac{N^2\lambda^2}{N^2 - 1 + N\lambda^2}.$$

We finally plug λ^2 in Equation (3.42), so that (3.42) becomes:

$$\begin{aligned}
&\mathbb{P}^{N+1}\left(\left\{\xi^{(1)},\dots,\xi^{(N+1)}\right\}\in\mathbb{R}^{n_\xi(N+1)}:(\xi^{(N+1)}-\hat{\xi})^\top\hat{\Sigma}^{-1}(\xi^{(N+1)}-\hat{\xi})>\lambda^2\right)\\
&\quad\le\frac{1}{N+1}\left\lfloor\frac{n_\xi(N+1)(N^2-1+N\lambda^2)}{N^2\lambda^2}\right\rfloor\\
&\quad\le\frac{n_\xi(N^2-1+N\lambda^2)}{N^2\lambda^2}.
\end{aligned}$$

■

4 Chance Constrained Linear Programs

In this chapter uncertain optimization programs of the form (1.3) will be reformulated in a tractable way using the theoretical results from Chapter 3. We will study two ways to approach this problem: the first one based on distrbutionally robust optimization and the second one based on randomized optimization.

4.1 Distributionally Robust Approach

Distributionally robust optimization deals with chance constraints by considering probability distributions lying in an uncertainty set describing our knowledge about the uncertainty. In this section we will consider the ambiguity set $\mathcal{P}_\alpha(\mu, \Sigma)$ defined in (3.2). The distributionally robust optimization problem is in the form of (1.4) with $\mathcal{P}_\alpha(\mu, \Sigma)$:

$$\begin{aligned} &\underset{x\in\mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && \mathbb{P}\left(g(x,\xi) \leq 0\right) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{P}_\alpha(\mu, \Sigma). \end{aligned} \tag{4.1}$$

In this work, linear chance constraints with will be discussed focusing on the case when the uncertainty ξ affects linearly the coefficients.

4.1.1 Single Linear Chance Constraint

Single linear chance constraints have already been exactly reformulated as SOC constraints using distributionally robust optimization: e.g. by Calafiore and Ghaoui [13] or by Cinquemani et al. [18]. However, these reformulations do not take into account the unimodality of the distributions. We will reformulate distributionally robust single chance constraints in an exact way considering distributions in $\mathcal{P}_\alpha(\mu, \Sigma)$. The following problem will be analyzed:

$$\begin{aligned} &\underset{x\in\mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && \mathbb{P}\left(a(\xi)^\top x \leq b(\xi)\right) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{P}_\alpha(\mu, \Sigma). \end{aligned} \tag{4.2}$$

$\xi \in \Xi \subseteq \mathbb{R}^{n_\xi}$ is assumed to enter linearly in $a(\xi)$ and $b(\xi)$. Thus, the coefficients can be written as

$$\begin{aligned} a(\xi) &= a_0 + \sum_{j=1}^{n_\xi} a_j \xi_j = a_0 + \hat{A}\xi \\ b(\xi) &= b_0 + \sum_{j=1}^{n_\xi} b_j \xi_j = b_0 + \hat{b}^\top \xi, \end{aligned} \tag{4.3}$$

where $a_0 \in \mathbb{R}^{n_x}$, $b_0 \in \mathbb{R}$, $a_j \in \mathbb{R}^{n_x}$ $j = \{1, \dots, n_\xi\}$ and $b_j \in \mathbb{R}$ $j = \{1, \dots, n_\xi\}$. Hence, $\hat{b} \in \mathbb{R}^{n_\xi}$ and $\hat{A} \in \mathbb{R}^{n_x \times n_\xi}$. Please note that a_0 and b_0 are the nominal values of the coefficients while $\hat{A}$ and $\hat{b}$ describe the perturbation given by the uncertainty.

The single chance constraint can be written by using (4.3) and the infimum over all the probabilities in the ambiguity set in the following way:

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\left(a_0 + \hat{A}\xi\right)^\top x \le b_0 + \hat{b}^\top \xi\right) \ge 1 - \epsilon.$$

It is possible to rearrange the terms in the following way:

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\left(\hat{A}^\top x - \hat{b}\right)^\top \xi \le b_0 - a_0^\top x\right) \ge 1 - \epsilon.$$

It is clear that if the uncertainty is null, the inequality becomes the nominal one $a_0^\top x \le b_0$. We define

$$\tilde{a}(x) = \hat{A}^\top x - \hat{b}, \qquad \tilde{b}(x) = b_0 - a_0^\top x,$$

and rewrite the constraint as

$$\inf_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\tilde{a}(x)^\top \xi \le \tilde{b}(x)\right) \ge 1 - \epsilon.$$

By bounding the complement of the argument inside the constrant, it can be rewritten as:

$$\sup_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\tilde{a}(x)^\top \xi > \tilde{b}(x)\right) \le \epsilon.$$

We now focus on random variable $\tilde{a}(x)^\top \xi$. By noting that

$$\mathbb{E}\left(\tilde{a}(x)^\top \xi\right) = \tilde{a}(x)^\top \mu, \qquad \mathbb{V}\text{ar}\left(\tilde{a}(x)^\top \xi\right) = \left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2^2,$$

the constraint can be reformulated as:

$$\sup_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\frac{\tilde{a}(x)^\top \xi - \tilde{a}(x)^\top \mu}{\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2} > \frac{\tilde{b}(x) - \tilde{a}(x)^\top \mu}{\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2}\right) \le \epsilon. \tag{4.4}$$

The left-hand side of the inequality inside the probability measure, is the standardized version (0 mean and variance 1) of random variable $\tilde{a}(x)^\top \xi$. From the definition of the one-sided probability bound for 0 mean variance 1 variables in Equation (3.22), we know that:

$$\sup_{\mathbb{P} \in \mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\frac{\tilde{a}(x)^\top \xi - \tilde{a}(x)^\top \mu}{\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2} > \frac{\tilde{b}(x) - \tilde{a}(x)^\top \mu}{\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2}\right) = f_{\alpha,os}\left(\frac{\tilde{b}(x) - \tilde{a}(x)^\top \mu}{\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2}\right).$$

After applying the inverse $f_{\alpha,os}^{-1}$ defined in (3.23) to Equation (4.4), the single linear chance constraint can be reformulated as a SOC constraint:

$$\tilde{b}(x) - \tilde{a}(x)^\top \mu \ge f_{\alpha,os}^{-1}(\epsilon)\left\|\Sigma^{1/2}\tilde{a}(x)\right\|_2,$$

where the inequality sign changed because $f_{\alpha,os}^{-1}$ is a monotonically decreasing function.

Finally, problem (4.2) can be exactly reformulated as an Second-Order Cone Program (SOCP):

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \tilde{b}(x) - \tilde{a}(x)^\top \mu \geq f^{-1}_{\alpha,os}(\epsilon) \left\| \Sigma^{1/2}\tilde{a}(x) \right\|_2 . \end{aligned}$$

By comparing the inequality we get by neglecting the uncertainty[1]:

$$\tilde{b}(x) \geq 0,$$

to the one in the SOCP, we can see that there is a margin introduced by two elements: the first related to the mean μ and the second related to the covariance Σ together with the function $f^{-1}_{\alpha,os}(\epsilon)$. The latter adapts depending on the unimodality index α and on the probability level ϵ. Please see Figure 3.5 for the behavior $f^{-1}_{\alpha,os}(\epsilon)$. As ϵ tends to 0, the margin tends to infinity and the feasible region shrinks. On the other hand, assuming low α means considering distributions that are more concentrated around the mode and reducing the margin.

4.1.2 Multiple Linear Chance Constraints

Unfortunately multiple linear chance constraints cannot be in general reformulated in an exact tractable way. Therefore, we will show two tractable approximations: the first based on Bonferroni inequality and the second one based on robust optimization with respect to uncertainty ellipsoids. The following problem with n_g joint chance constraints will be discussed:

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \mathbb{P}\left(\bigcap_{i=1}^{n_g} \left(\tilde{a}_i(x)^\top \xi \leq \tilde{b}_i(x) \right) \right) \geq 1 - \epsilon \quad \forall\, \mathbb{P} \in \mathcal{P}_\alpha(\mu, \Sigma). \end{aligned} \tag{4.5}$$

4.1.2.1 Bonferroni Approximation

Problem (4.5) can be rewritten using the union of the complements of all chance constraints:

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \mathbb{P}\left(\bigcup_{i=1}^{n_g} \left(\tilde{a}_i(x)^\top \xi > \tilde{b}_i(x) \right) \right) \leq \epsilon \quad \forall\, \mathbb{P} \in \mathcal{P}_\alpha(\mu, \Sigma). \end{aligned} \tag{4.6}$$

From Bonferroni inequality we have that:

$$\mathbb{P}\left(\bigcup_{i=1}^{n_g} \left(\tilde{a}_i(x)^\top \xi > \tilde{b}_i(x) \right) \right) \leq \sum_{i=1}^{n_g} \mathbb{P}\left(\tilde{a}_i(x)^\top \xi > \tilde{b}_i(x) \right).$$

Thus, for any vector $\boldsymbol{\epsilon} \in \mathbb{R}^{n_g}$ such that $\mathbf{1}^\top \boldsymbol{\epsilon} \leq \epsilon$, the system of distributionally robust individual chance constraints

$$\sup_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\tilde{a}(x)^\top \xi > \tilde{b}(x) \right) \leq \epsilon_i, \quad i = 1, \ldots, n_g,$$

[1] Setting $\mu = 0 \in \mathbb{R}^{n_\xi}$ and $\Sigma = 0 \in \mathbb{R}^{n_\xi \times n_\xi}$

or equivalently

$$\inf_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\tilde{a}(x)^\top \xi \leq \tilde{b}(x)\right) \geq 1-\epsilon_i, \quad i=1,\dots,n_g,$$

corresponds to a conservative approximation of the constraint in problem (4.6). In this work, we will choose for simplicity $\epsilon_i = \epsilon/n_g$. It can be shown that even if the vector $\boldsymbol{\epsilon}$ is chosen optimally, this approximation can still be conservative [50].

Finally, using Bonferroni inequality, we can reformulate each single chance constraint in the same way as in Section 4.1.1 and rewrite problem (4.6) as an SOCP as follows:

$$\begin{aligned} &\underset{x\in\mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && \tilde{b}_i(x) - \tilde{a}_i(x)^\top \mu \geq f^{-1}_{\alpha,os}(\epsilon/n_g)\left\|\Sigma^{1/2}\tilde{a}_i(x)\right\|_2, \qquad i=1,\dots,n_g. \end{aligned} \tag{4.7}$$

We will denote this reformulation as Bonferroni Approximation (BA). Also in this case, the margin depends on the mean μ and on a second term given by a 2-norm multiplying the inverse function $f^{-1}_{\alpha,os}$. Differently from the single chance constraint case, the argument of $f^{-1}_{\alpha,os}$ is reduced by the number of joint chance constraints. From Figure 3.5 it is clear that as the argument of $f^{-1}_{\alpha,os}$ tends to 0 the margin increases, shrinking the feasible region. By consequence, more joint chance constraints will increase the conservatism of this approximation.

4.1.2.2 Ellipsoid Approximation

Another way to deal with joint chance constraints is to reformulate them as robust ones with respect to uncertainty sets having distributionally robust probabilistic guarantees.

We will consider ellipsoidal uncertainty sets $\mathcal{E}$ defined using parametrization (B.5) in the Appendix:

$$\mathcal{E} = \{x_c + Bu \mid \|u\|_2 \leq 1\}.$$

$x_c \in \Xi \subseteq \mathbb{R}^{n_\xi}$ is the center and $B \in \mathbb{S}^{n_\xi}_+$ is the shape matrix.

If the set $\mathcal{E}$ has probabilistic guarantess with respect to all the distributions in $\mathcal{P}_\alpha(\mu,\Sigma)$, i.e.

$$\inf_{\mathbb{P}\in\mathcal{P}_\alpha(\mu,\Sigma)} \mathbb{P}\left(\mathcal{E}\right) \geq 1-\epsilon, \tag{4.8}$$

then the solution to the robust program

$$\begin{aligned} &\underset{x\in\mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && \tilde{a}_i(x)^\top \xi \leq \tilde{b}_i(x), \quad \forall \xi\in\mathcal{E} \quad i=1,\dots,n_g, \end{aligned}$$

will have at least the same probabilistic guarantees as the ones of the set $\mathcal{E}$. In other words, the optimal solution will have the probabilistic guarantees of problem (4.5). It is well known that robust linear programs can be reformulated as SOCP, see [2]. The reformulation can be obtained by imposing

$$\begin{aligned} &\sup_{\xi\in\mathcal{E}} \left(\tilde{a}_i(x)^\top \xi\right) \leq \tilde{b}_i(x), \quad i=1,\dots,n_g, \\ \Longleftrightarrow\ &\tilde{a}_i(x)^\top x_c + \sup_{\|u\|_2\leq 1} \left(\tilde{a}_i(x)^\top Bu\right) \leq \tilde{b}_i(x), \quad i=1,\dots,n_g, \\ \Longleftrightarrow\ &\tilde{a}_i(x)^\top x_c + \|B\tilde{a}_i(x)\|_2 \leq \tilde{b}_i(x), \quad i=1,\dots,n_g, \\ \Longleftrightarrow\ &\tilde{b}_i(x) - \tilde{a}_i(x)^\top x_c \geq \|B\tilde{a}_i(x)\|_2, \quad i=1,\dots,n_g. \end{aligned} \tag{4.9}$$

In Appendix B, we derived the relationship between different ellipsoid parametrization. Parametrization (B.5) used here corresponds to parametrization (B.1) when $B = C^{-1/2}$. Thus, by adopting uncertainty set

$$\mathcal{E} = \left\{\xi \in \Xi : (\xi - x_c)^\top C(\xi - x_c) \le 1\right\},$$

and using Equation (4.9), we can reformulate the robust LP as the following SOCP

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \tilde{b}_i(x) - \tilde{a}_i(x)^\top x_c \ge \left\| C^{-1/2}\tilde{a}_i(x) \right\|_2, \quad i = 1, \ldots, n_g. \end{aligned} \tag{4.10}$$

The degree of conservatism introduced by this approximation critically depends on the choice of the ellipsoid $\mathcal{E}$. Whatever ellipsoid is chosen, it must satisfy probabilistic guarantees (4.8) in order to have a feasible solution x^*.

Minimum Volume Ellipsoid Reformulation A reasonable ellipsoid $\mathcal{E}$ we can use is the Minimum Volume Ellipsoid (MVE). In Theorem 3.4.2 we derived the following closed form solution for the MVE containing at least $1 - \epsilon$ of probability mass:

$$x_c = \mu, \qquad C = \frac{\Sigma^{-1}}{n\left(f_{\alpha,ds}^{-1}\epsilon)\right)^2}.$$

By plugging last equation in problem (4.10), we obtain the Minimum Volume Ellipsoid Approximation (MVEA):

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \tilde{b}_i(x) - \tilde{a}_i(x)^\top \mu \ge \sqrt{n_\xi} f_{\alpha,ds}^{-1}(\epsilon) \left\| \Sigma^{1/2}\tilde{a}_i(x) \right\|_2, \quad i = 1, \ldots, n_g. \end{aligned} \tag{4.11}$$

Comparison Bonferroni and Minimum Volume Ellipsoid Approximations So far we have seen two different ways to deal with multiple joint linear chance constraints: Bonferroni Approximation (4.7) and Minimum Volume Ellipsoid Approximation (4.11). Their structure and dependence on μ, Σ is the same:

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & \tilde{b}_i(x) - \tilde{a}_i(x)^\top \mu \ge \rho \left\| \Sigma^{1/2}\tilde{a}_i(x) \right\|_2, \quad i = 1, \ldots, n_g, \end{aligned} \tag{4.12}$$

and the only difference is the term ρ multiplying the 2-norm.

The coefficients for the two methods are displayed in Table 4.1. The BA depends on the number of joint chance constraints n_g and becomes more conservative as n_g increases. On the other hand, the MVEA depends on the square root of the dimension of the uncertainty n_ξ and becomes more conservative when n_ξ is large. Thus, it preferable to choose one or the other method depending on the structure of the chance constrained program.

In Figure 4.1 the the MVEA coefficient with $n_\xi = 3$ is compared with the BA coefficient with $n_g = 2, 3, 4$. In addition, in Figure 4.2 the BA coefficient with $n_g = 3$ us compared with the MVEA coefficient with $n_\xi = 2, 3, 4$. From the graphs it is clear that the BA is better than the MVEA if and only if $n_\xi \ge n_g$.

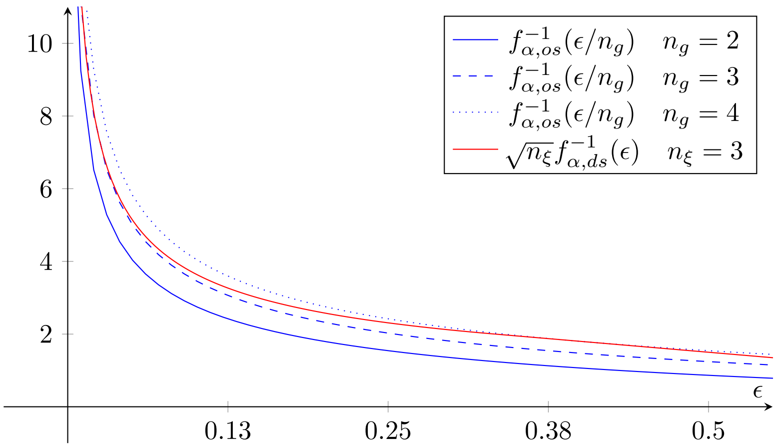

Figure 4.1: Comparison of MVEA and BA coefficients with $n_g = 2, 3, 4$ and $n_\xi = 3$.

BA	MVE
$f^{-1}_{\alpha,os}(\epsilon/n_g)$	$\sqrt{n_\xi} f^{-1}_{\alpha,ds}(\epsilon)$

Table 4.1: BA and MVEA coefficient ρ dependence comparison.

Iterative Ellipsoid Approach Another way to define the ellipsoid inside problem (4.10) having distributionally robust probabilistic guarantees, is to use Theorem 3.4.1 stating that an ellipsoid $\mathcal{E}$ centered at the mean μ with shape matrix C such that:

$$\left(f^{-1}_{\alpha,ds}(\epsilon)\right)^2 \langle C, \Sigma \rangle \leq 1, \tag{4.13}$$

satisfies Equation (4.8). Thus, it is possible to iteratively solve problem (4.10) by first optimizing with a fixed ellipsoid $\mathcal{E}$ and then by optimizing with respect to the shape matrix C to find a new one giving a better optimum while enforcing the previous solution to be feasible.

We will denote this procedure as Iterative Ellipsoid Approximation (IEA). It is described in Algorithm 1 and works as follows. First we solve the robust LP (4.10) with the MVE as a first guess for the ellipsoid obtaining the solution x^*_{cur}, the optimal cost ϕ^*_{cur} and the set $\mathcal{A}$ of active constraints indeces. Then, while the current optimal cost ϕ^*_{cur} and the previous one ϕ^*_{prev} differ for more than ϵ, we perform the iterations. We save x^*_{cur} and ϕ^*_{cur} inside the previous optimization results x^*_{prev} and ϕ^*_{prev}. Then, we reshape the matrix C by ensuring that the solution x^*_{prev} satisfies the constraints. This is done by squaring the constraints of (4.9) in x^*_{prev}:

$$\left(\tilde{b}_i(x^*_{prev}) - \tilde{a}_i(x^*_{prev})^\top \mu\right)^2 \geq \tilde{a}_i(x^*_{prev})^\top C^{-1} \tilde{a}_i(x^*_{prev}), \quad i = 1, \ldots, n_g. \tag{4.14}$$

Please note that (4.14) and the constraints in (4.9) in x^*_{prev} are equivalent if and only if

$$\tilde{b}_i(x^*_{prev}) - \tilde{a}_i(x^*_{prev})^\top \mu \geq 0.$$

This is implied by the fact that x^*_{prev} satisfies problem (4.9). Afterwards, we minimize the right-hand side of (4.14) for the active constraints in $\mathcal{A}$ while ensuring that x^*_{prev} is always

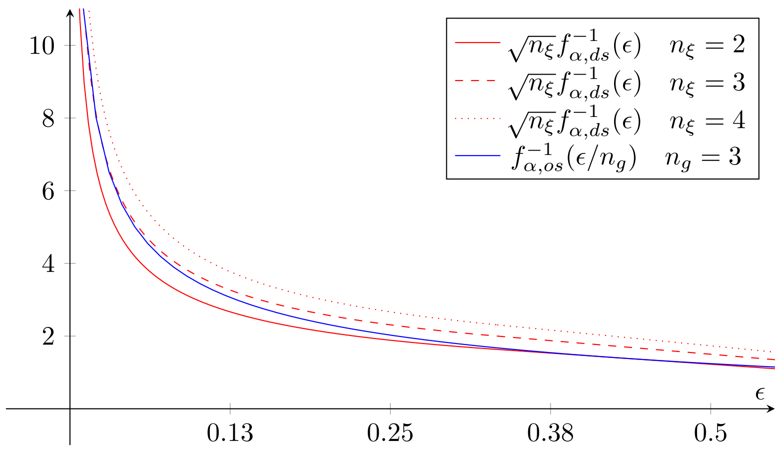

Figure 4.2: Comparison of MVEA and BA coefficients with $n_g = 3$ and $n_\xi = 2, 3, 4$.

feasible for the new shape matrix and that (4.13) holds. It can be done by solving the following SDP:

$$
\begin{aligned}
\text{minimize} \quad & t \\
\text{subject to:} \quad & C \in \mathbb{S}_{++}^{n_\xi},\ t \in \mathbb{R} \\
& \begin{bmatrix} C & \tilde{a}_i(x^*_{prev}) \\ \tilde{a}_i(x^*_{prev})^\top & \left(\tilde{b}_i(x^*_{prev}) - \tilde{a}_i(x^*_{prev})^\top \mu\right)^2 \end{bmatrix} \succeq 0 \quad i = 1, \dots, n_g \\
& \begin{bmatrix} C & \tilde{a}_j(x^*_{prev}) \\ \tilde{a}_j(x^*_{prev})^\top & t \end{bmatrix} \succeq 0 \quad \forall j \in \mathcal{A} \\
& \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \langle C, \Sigma \rangle \le 1,
\end{aligned}
\tag{4.15}
$$

obtaining C^*. Then, we solve again the robust LP obtaining a new x^*_{cur}, ϕ^*_{cur} and a new set $\mathcal{A}$ of active constraints indices. We repeat this procedure until the solution converges and $\left|\phi^*_{cur} - \phi^*_{prev}\right| < \epsilon$.

Algorithm 1 Iterative Ellipsoid

Initialization: Solve robust LP (4.10) with MVE obtaining x^*_{cur}, ϕ^*_{cur} and set $\mathcal{A}$.
while $\left|\phi^*_{cur} - \phi^*_{prev}\right| \ge \epsilon$ **do**
  $x^*_{prev} \leftarrow x^*_{cur}$ and $\phi^*_{prev} \leftarrow \phi^*_{cur}$
  Reshape matrix C by solving SDP (4.15) obtaining C^*
  Solve robust LP (4.10) with $x_c = \mu$ and $C = C^*$ obtaining new x^*_{cur}, ϕ^*_{cur} and set $\mathcal{A}$.
end while
$x^* \leftarrow x^*_{cur}$

This iterative procedure has shown to empirically converge to an ellipsoid $\mathcal{E}$ having the same guarantees as the MVE but giving a better optimal solution for the optimization problem. As one can intuitively see, this approach is computationally more expensive than BA and MVE because we need to solve several SOCPs (4.10) and also SDPs (4.15).

4.1.3 Moment Uncertainty

In practical problems, limited information about the ambiguity set $\mathcal{P}_\alpha(\mu, \Sigma)$ is available. Usually, we have to gather this knowledge from data and construct the ambiguity set accordingly. The most intuitive method to obtain μ and Σ is to compute the estimated ones. Based on N samples $\xi^{(1)}, \dots, \xi^{(N)}$ of random variable $\xi \in \Xi \subseteq \mathbb{R}^{n_\xi}$, empirical mean and covariance are defined:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \xi^{(i)}, \qquad \hat{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N} \left(\xi^{(i)} - \hat{\mu}\right)\left(\xi^{(i)} - \hat{\mu}\right)^\top .$$

Unfortunately, plugging $\hat{\mu}$ and $\hat{\Sigma}$ inside the ambiguity set $\mathcal{P}_\alpha(\mu, \Sigma)$ will not give reliable probabilistic guarantees, especially if they are estimated from few data samples. We will introduce a more rigorous approach to use the estimates while taking into account estimation errors.

Shawe-Taylor and Cristianini in [43] extended the Hoeffding's inequality to random vectors and matrices. The first lemma below describes random vectors behavior

Lemma 4.1.1 ([43, Theorem 3]). *Let $\xi^{(1)}, \dots, \xi^{(N)} \in \Xi \subseteq \mathbb{R}^{n_\xi}$ be i.i.d. random samples of ξ and let*

$$R = \sup_{\xi \in \Xi} \|\xi\|_2 , \qquad \beta \in (0,1).$$

Then, with confidence at least $1-\beta$, the following holds:

$$\|\mu - \hat{\mu}\|_2 \le r_1, \qquad r_1 = \frac{R}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{1}{\beta}}\right).$$

The next lemma provides a similar result for the empirical covariance.

Lemma 4.1.2 ([43, Corollary 6]). *Let $\xi^{(1)}, \dots, \xi^{(N)} \in \Xi \subseteq \mathbb{R}^{n_\xi}$ be i.i.d. random samples of ξ and let*

$$R = \sup_{\xi \in \Xi} \|\xi\|_2 , \qquad \beta \in (0,1).$$

Then, provided that $N > \left(2 + \sqrt{2\ln\frac{2}{\beta}}\right)^2$, it holds with confidence at $1-\beta$ that:

$$\left\|\Sigma - \hat{\Sigma}\right\|_F \le r_2, \qquad r_2 = \frac{2R^2}{\sqrt{N}}\left(2 + \sqrt{2\ln\frac{2}{\beta}}\right).$$

Combining Lemma 4.1.1 and Lemma 4.1.2 (both with $\beta/2$ instead of β), Calafiore and El Ghaoui [13] formulated a theorem defining the regions for the true mean and covariance centered at the estimated ones with an overall confidence at least $1-\beta$.

Theorem 4.1.1 ([13, Theorem 4.1]). *Let $\xi^{(1)}, \dots, \xi^{(N)} \in \Xi \subseteq \mathbb{R}^{n_\xi}$ be i.i.d. random samples of ξ such that*

$$R = \sup_{\xi \in \Xi} \|\xi\|_2$$

Then, provided that $N > \left(2 + \sqrt{2 \ln \frac{4}{\beta}}\right)^2$*, the following holds with confidence at least* $1 - \beta$*:*

$$\begin{aligned} \|\mu - \hat{\mu}\|_2 &\leq r_1 & r_1 &= \frac{R}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{2}{\beta}}\right) \\ \left\|\Sigma - \hat{\Sigma}\right\|_F &\leq r_2 & r_2 &= \frac{2R^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{4}{\beta}}\right). \end{aligned}$$

As the radii depend on the square root of the logarithm of β, it is possible to set β to a a very low value and without increasing too much the confidence regions.

4.1.3.1 Reformulations Using Moments Robustification

These confidence regions can be included inside the SOCP reformulations obtained in the previous sections in order to have a data-driven reformulation of the chance constrained problem (4.1).

BA and MVEA Robustification We edit the constraints in the general SOCP program (4.12) coming from both BA and MVE reformulations in analogous way to what Calafiore and Ghaoui did in [13] without unimodality and for a single linear chance constraint. From Theorem 4.1.1, it holds with confidence at least $1 - \beta$ that

$$\mu = \hat{\mu} + d, \qquad d \in \mathbb{R}^{n_\xi} : \|d\|_2 \leq r_1, \tag{4.16}$$

and that

$$\Sigma = \hat{\Sigma} + \Delta, \qquad \Delta \in \mathbb{R}^{n_\xi \times n_\xi} : \|\Delta\|_F \leq r_2. \tag{4.17}$$

By plugging Equations (4.16) and (4.17) inside the i-th constraint of problem (4.12) we get:

$$\rho \left\|(\Sigma + \Delta)^{1/2} \tilde{a}_i(x)\right\|_2 + \tilde{a}_i(x)^\top (\hat{\mu} + d) - \tilde{b}_i(x) \leq 0.$$

We can majorize the left-hand side as follows:

$$\begin{aligned} &\rho \left\|(\Sigma + \Delta)^{1/2} \tilde{a}_i(x)\right\|_2 + \tilde{a}_i(x)^\top (\hat{\mu} + d) - \tilde{b}_i(x) \\ &= \rho \sqrt{\tilde{a}_i(x)^\top \Sigma \tilde{a}_i(x) + \operatorname{tr}\left(\Delta \tilde{a}_i(x) \tilde{a}_i(x)^\top\right)} + \tilde{a}_i(x)^\top (\hat{\mu} + d) - \tilde{b}_i(x) \\ &\leq \rho \sqrt{\tilde{a}_i(x)^\top \Sigma \tilde{a}_i(x) + \|\Delta\|_F \|\tilde{a}_i(x) \tilde{a}_i(x)^\top\|_F} + \tilde{a}_i(x)^\top (\hat{\mu} + d) - \tilde{b}_i(x) \\ &\leq \rho \sqrt{\tilde{a}_i(x)^\top (\Sigma + r_2 I) \tilde{a}_i(x)} + \tilde{a}_i(x)^\top \hat{\mu} + r_1 \|\tilde{a}_i(x)\|_2 - \tilde{b}_i(x) \\ &\leq \rho \left\|(\Sigma + r_2 I)^{1/2} \tilde{a}_i(x)\right\|_2 + \tilde{a}_i(x)^\top \hat{\mu} + r_1 \|\tilde{a}_i(x)\|_2 - \tilde{b}_i(x). \end{aligned} \tag{4.18}$$

Thus, problem (4.12) can be robustified against moment uncertainty as a modified SOCP:

$$\begin{aligned} &\underset{x \in \mathcal{X}}{\text{minimize}} && c^\top x \\ &\text{subject to:} && \tilde{b}_i(x) - \tilde{a}_i(x)^\top \mu - r_1 \|\tilde{a}_i(x)\|_2 \geq \rho \left\|(\Sigma + r_2 I)^{1/2} \tilde{a}_i(x)\right\|_2, \quad i = 1, \ldots, n_g. \end{aligned} \tag{4.19}$$

The complexity does not increase as the program keeps the same structure.

With confidence $1-\beta$, a solution x^* to program (4.19) will satisfy the chance constraints in (4.1). Hence, this reformulation allows us to use the BA and the MVEA in a data-driven setting.

IEA Robustification In order to use the IEA empirical estimates of mean and covariance we have do adapt the iterative algorithm. In particular the ellipsoid reshaping in problem (4.15) has to be adapted in order to take uncertainty into account. Condition (4.13) needed to have probabilistic guarantees on ellipsoid $\mathcal{E}$, has to be edited using Equation (4.17) by majorizing the left-hand side as follows

$$\begin{aligned}&\left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \left\langle C, \left(\hat{\Sigma}+\Delta\right)\right\rangle \\ &\qquad = \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \left(\langle C,\hat{\Sigma}\rangle + \langle C,\Delta\rangle\right) \\ &\qquad \leq \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \left(\langle C,\hat{\Sigma}\rangle + \|C\|_F \|\Delta\|_F\right) \\ &\qquad \leq \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \left(\langle C,\hat{\Sigma}\rangle + r_2 \|C\|_F\right).\end{aligned} \tag{4.20}$$

Thus, by using Equation (4.20) and the mean robustificaiton in Equation (4.19), we can rewrite problem (4.13) as:

$$\begin{aligned}\text{minimize}\quad & t \\ \text{subject to:}\quad & C \in \mathbb{S}_{++}^{n_\xi},\ t \in \mathbb{R} \\ & \begin{bmatrix} C & \tilde{a}_i(x^*_{prev}) \\ \tilde{a}_i(x^*_{prev})^\top & \left(\tilde{b}_i(x^*_{prev}) - \tilde{a}_i(x^*_{prev})^\top \hat{\mu} - r_1 \left\|\tilde{a}_i(x^*_{prev})\right\|_2\right)^2 \end{bmatrix} \succeq 0 \quad i=1,\dots,n_g \\ & \begin{bmatrix} C & \tilde{a}_j(x^*_{prev}) \\ \tilde{a}_j(x^*_{prev})^\top & t \end{bmatrix} \succeq 0 \quad \forall j \in \mathcal{A} \\ & \left(f_{\alpha,ds}^{-1}(\epsilon)\right)^2 \left(\langle C,\hat{\Sigma}\rangle + r_2 \|C\|_F\right) \leq 1\end{aligned} \tag{4.21}$$

After reshaping the ellipsoid obtaining C^*, the SOCP required to get the new solution x^*_{cur} and cost ϕ^*_{cur} needs to be robustified only with respect to the estimated mean $\hat{\mu}$, and thus can be written as:

$$\begin{aligned}\underset{x\in\mathcal{X}}{\text{minimize}}\quad & c^\top x \\ \text{subject to:}\quad & \tilde{b}_i(x) - \tilde{a}_i(x)^\top \hat{\mu} - r_1 \|\tilde{a}_i(x)\|_2 \geq \rho \left\|C^{*-1/2}\tilde{a}_i(x)\right\|_2, \quad i=1,\dots,n_g.\end{aligned} \tag{4.22}$$

In Algorithm 2 the modified iterative procedure is described using last changes in (4.20), (4.21) and (4.22).

With confidence $1-\beta$, a solution x^* to program (4.19) will satisfy the chance constraints in (4.1) with probability $1-\epsilon$.

4.1.3.2 Multivariate Sampled Chebyshev Approach

An alternative way to deal with the errors estimating μ and Σ comes from the empirical Chebyshev inequality in multiple dimensions proven in Theorem 3.6.1. By using $N+1$ samples

Algorithm 2 Robustified Iterative Ellipsoid

Initialization: Solve (4.19) with MVE obtaining x^*_{cur}, ϕ^*_{cur} and set $\mathcal{A}$.
while $\left|\phi^*_{cur} - \phi^*_{prev}\right| \geq \epsilon$ **do**
  $x^*_{prev} \leftarrow x^*_{cur}$ and $\phi^*_{prev} \leftarrow \phi^*_{cur}$
  Reshape matrix C by solving SDP (4.21) obtaining C^*
  Solve (4.22) with $x_c = \mu$ and $C = C^*$ obtaining new x^*_{cur}, ϕ^*_{cur} and set $\mathcal{A}$.
end while
$x^* \leftarrow x^*_{cur}$

$\xi^{(1)}, \ldots, \xi^{(N)}, \xi^{(N+1)} \in \Xi \subseteq \mathbb{R}^{n_\xi}$ of ξ, it is possible to exploit that result in order to construct an ellipsoid $\mathcal{E}(\xi^{(1)}, \ldots, \xi^{(N)})$ depending on the first N samples such that:

$$\mathbb{P}^{N+1}\left(\xi^{N+1} \in \mathcal{E}\left(\xi^{(1)}, \ldots, \xi^{(N)}\right)\right) \geq 1 - \epsilon. \tag{4.23}$$

Using Equation (3.35), we define $\mathcal{E}$ as

$$\mathcal{E}(\xi^{(1)}, \ldots, \xi^{(N)}) = \left\{(\xi - \hat{\mu})^\top \frac{\hat{\Sigma}^{-1}}{\lambda^2}(\xi - \hat{\mu}) \leq 1\right\}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the estimated mean and covariance from the first N samples and λ^2 is a scaling parameter that determines the probability in (4.23). We choose λ by enforcing the right-hand side of (3.35) being lower than a predefined ϵ.

$$\frac{n_\xi(N^2 - 1 + N\lambda^2)}{N^2\lambda^2} \leq \epsilon \implies \lambda^2 \geq \frac{n_\xi(N^2 - 1)}{N(\epsilon N - n_\xi)}.$$

Then, by selecting N in order to have a positive λ^2 and the minimum λ to have the smallest ellipsoid with probabilistic bound ϵ, we obtain:

$$N > \frac{n_\xi}{\epsilon}, \qquad \lambda = \sqrt{\frac{n_\xi(N^2 - 1)}{N(\epsilon N - n_\xi)}}. \tag{4.24}$$

Then, by plugging the defined ellipsoid inside problem (4.10) we get the following data-driven reformulation still as SOCP:

$$\begin{array}{ll} \underset{x \in \mathcal{X}}{\text{minimize}} & c^\top x \\ \text{subject to:} & \tilde{b}_i(x) - \tilde{a}_i(x)^\top \hat{\mu} \geq \lambda \left\|\hat{\Sigma}^{1/2}\tilde{a}_i(x)\right\|_2, \quad i = 1, \ldots, n_g. \end{array} \tag{4.25}$$

We denote this approach as the Multivariate Sampled Chebyshev Approximation (MSCA). The probabilistic guarantees of solution x^* to the problem above are now different from the ones obtained from Theorem 4.1.1. In this setting there is no β. The bound ϵ means the following: the probability of drawing $N+1$ samples, constructing $\mathcal{E}(\xi^{(1)}, \ldots, \xi^{(N)})$ based on the first N ones, solving problem (4.25) obtaining x^* that violates the constraints defined by $\xi^{(N+1)}$ in (1.3), is lower than ϵ.

This kind of guarantees are useful when we care only about the $(N+1)$-th sample, e.g. when performing receding horizon optimization. If we optimize at each time stage taking care of only the next sample and then we optimize again, we do not need two layers of probability describing all the future samples but just the $N+1$-th. In the Applications part of this work, we will show how this method performs in an MPC example where we optimize at each time stage.

4.2 Scenario Approach

Recently developed by Calafiore and Campi in [12] and then by Campi and Garatti in [14], the SA has shown to be an intuitive and effective way to deal with chance constrained programs of the form (1.3). Instead of solving the untractable problem (1.3), the SA defines a random program based on the N i.i.d. samples $\xi^{(1)}, \ldots, \xi^{(N)} \in \Xi \subseteq \mathbb{R}^{n_\xi}$ of ξ.

$$\begin{aligned} RP_N : \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to:} \quad & g(x, \xi^{(i)}) \leq 0, \quad i = 1, \ldots, N. \end{aligned} \tag{4.26}$$

Although obtained based only on a finite number of samples, the solution x_N^* to the random program comes with precise guarantees about its feasibility for problem (1.3). There are few assumptions on problem (1.3):

Assumption 4.2.1. *Set $\mathcal{X}$ is convex and constraint g is convex in x. Moreover* (4.26) *admits an unique feasible solution with probability one.*

Here there are some definitions we will use to define the probabilistic guarantees of x^*.

Definition 4.2.1 (Violation probability [11]). *The violation probability of an element $x \in \mathbb{R}^n$ is the probability that there exists an element $\delta \in \Delta$ for which the constraints are not satisfied.*

$$V(x) = \mathbb{P}\left(\xi \in \mathbb{R}^{n_\xi} : g(x, \xi) > 0\right)$$

As we are dealing with a random program, its solution x_N^* is a random variable. In addition, the violation probability of the solution $V(x_N^*)$ is also a random variable.

Definition 4.2.2 (Support constraint [11]). *A constraint $\xi^{(r)}$ with $r \in \{1, \ldots, N\}$ is a support constraint for RP_N if its removal changes the solution of RP_N.*

Definition 4.2.3 (Helly's dimension [11]). *Helly's dimension of RP_N is the least integer ζ such that* $\operatorname{ess\,sup}_{\xi \in \Xi} |\mathrm{Sc}\,(RP_N)| \leq \zeta$ *holds for any finite $N \geq 1$.*

From [11, Lemma 2.2 and 2.3], $\zeta \leq n$ if RP_N is feasible with probability one, whereas $\zeta \leq n+1$ in all other cases.

The probabilistic guarantees of solution x_N^* can be described in the following Theorem

Theorem 4.2.1 ([11, Theorem 3.3]). *Consider problem* (4.26) *with $N \geq \zeta$. If Assumption 4.2.1 holds, then*

$$\mathbb{P}^N \left(\xi^{(1)}, \ldots, \xi^{(N)} \in \Xi^N : \mathbb{P}\left(V\left(x_N^*\right)\right) > \epsilon \right) \leq \sum_{j=1}^{\zeta-1} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j}.$$

Hence, we can bound the right-hand side by β as follows:

$$\sum_{j=1}^{\zeta-1} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j} \leq \beta. \tag{4.27}$$

If N, ϵ and β satisfy Equation (4.27), then with confidence $1-\beta$, the solution x_N^* of problem RP_N will satisfy the chance constraints of problem (1.3) with at probability $1-\epsilon$. By consequence, this theorem gives two layers probabilistic guarantees on x_N^*.

Equation (4.27) can be numerically inverted, or by using the Chernoff bound on the lower binomial tail, can be rewritten as:

$$N \geq \frac{\frac{e}{e-1}}{\epsilon}\left(\zeta - 1 + \ln\frac{1}{\beta}\right). \tag{4.28}$$

Since the minimum number of samples depends logarithmically on β^{-1}, we can choose a very high confidence without increasing the required samples too much. Unfortunately, this bound still depends linearly on ϵ^{-1}, and it can become computationally expensive when ϵ is chosen small.

It is also possible to give one layer probabilistic guarantees on solution x_N^*.

Theorem 4.2.2 ([10, Theorem 2.1]). *Consider problem* (4.26) *with* $N \geq \zeta$. *If Assumption 4.2.1 holds, then*

$$\mathbb{P}^{N+1}\left(\xi^{(1)}, \ldots, \xi^{(N)}, \xi^{(N+1)} \in \Xi^{N+1} : \mathbb{P}\left(g(x_N^*, \xi^{(N+1)}) > 0\right)\right) \leq \frac{\zeta}{N+1}.$$

Hence, we can bound the right hand side by ϵ as follows:

$$\frac{\zeta}{N+1} \leq \epsilon \implies N \geq \frac{\zeta}{\epsilon} - 1. \tag{4.29}$$

If ϵ and N satisfy Equation (4.29), then the probability of drawing $N+1$ samples, solving RP_N based on the first N ones obtaining x_N^* that violates the constraint defined by $\xi^{(N+1)}$ is lower than ϵ.

As discussed in the previous Section, this kind of guarantees are useful when dealing with receding horizon optimization problems. In the Applications part of this work, we will show how this method performs in a MPC control example.

Helly's dimension ζ can be explicitly bounded by exploiting the problem structure. In this work, as we assume affine dependence of the constraints with respect to the uncertainty, we will make use of the following recent result by Zhang et al.,

Proposition 4.2.1 ([49, Proposition 3]). *If problem* (1.3) *consists of* n_g *joint chance constraints with linear dependence on the uncertainty* $\xi \in \Xi \subseteq \mathbb{R}^{n_\xi}$*, the Helly's dimension can be bounded by:*

$$\zeta \leq n_g(n_\xi + 1).$$

When dealing with decision variables of higher dimension than the uncertainty one, this result can greatly reduce the number of samples required in order to have probabilistic guarantees. Please note that there are also other results to limit the Helly's dimension that in some situations can give better results than Proposition 4.2.1 such as the s-rank, see [41]. We will use Proposition 4.2.1 in the applications discussed in this work.

Part II

Applications

5 Machine Learning

The problem of choosing a linear discriminant that minimizes the misclassification probability of future data has been studied for decades in Machine Learning literature [7]. Traditional approaches constructing optimal linear discriminants are based on SVMs: they build an optimization problem from labeled data points trying to classify correctly as many of them as possible [7]. Unfortunately, this technique does not compute probabilistic guarantees of correctly classifying future data points. Lanckriet et al. [32] developed another approach to construct linear discriminants giving probabilistic guarantees based on distributionally robust optimization. We will discuss how this method has been improved and reformulated in several ways and how the probabilistic guarantees could benefit from the bounds proven in Chapter 3.

5.1 Minimax Probability Machines (MPMs)

The approach defined by Lanckriet et al. [32] constructs Minimax Probability Machines (MPMs). In this section the minimax formulation of linear classifiers is presented and rearranged using our α-unimodality results.

Let x and y be random vectors in a binary classification problem modeling data from two classes distributed respectively with mean and covariance matrix $(\bar{x}, \Sigma_x)$ and $(\bar{y}, \Sigma_y)$, having $x, \bar{x} \in \mathcal{X} \subseteq \mathbb{R}^n$, $y, \bar{y} \in \mathcal{Y} \subseteq \mathbb{R}^n$ and $\Sigma_x, \Sigma_y \in \mathbb{S}^n_{++}$. Let us define $\mathcal{P}_\alpha(\mu, \Sigma)$ from Equation (3.2) as the set of α-unimodal distributions on $\mathbb{R}^n$ having mean μ and covariance matrix Σ. For simplicity, in this section the two distributions are assumed to have the same α.

The goal of this optimization is to obtain an hyperplane $\mathcal{H}(a, b) = \{z \,|\, a^\top z = b\}$ where $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$. Given the minimum probability of correct classification $\gamma \in [0, 1)$, the MPM can be modeled as

$$
\begin{aligned}
MPM: \quad & \text{maximize} && \gamma \\
& \text{subject to:} && \gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\
& && \inf_{\mathbb{P} \in \mathcal{P}_\alpha(\bar{x}, \Sigma_x)} \mathbb{P}(a^\top x \geq b) \geq \gamma \\
& && \inf_{\mathbb{P} \in \mathcal{P}_\alpha(\bar{y}, \Sigma_y)} \mathbb{P}(a^\top y \leq b) \geq \gamma.
\end{aligned}
\tag{5.1}
$$

We can rewrite the problem as

$$\begin{aligned} MPM' : \quad & \text{maximize} \quad \gamma \\ & \text{subject to:} \quad \gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ & \qquad \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\bar{x},\Sigma_x)} \mathbb{P}(-a^\top x > -b) \le 1-\gamma \\ & \qquad \sup_{\mathbb{P}\in\mathcal{P}_\alpha(\bar{y},\Sigma_y)} \mathbb{P}(a^\top y > b) \le 1-\gamma, \end{aligned} \tag{5.2}$$

where $1-\gamma$ corresponds to the worst-case misclassification probability. Optimization problems MPM and MPM' are clearly chance constrained programs with two individual chance constraints. Thus, by exploiting the results obtained in Section 4.1.1, it is possible to rewrite MPM' as a SOCP in the following way:

$$\begin{aligned} & \text{maximize} \quad \gamma \\ & \text{subject to:} \quad \gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ & \qquad -b + a^\top \bar{x} \ge f^{-1}_{\alpha,os}(1-\gamma) \left\| \Sigma_x^{1/2} a \right\|_2 \\ & \qquad b - a^\top \bar{y} \ge f^{-1}_{\alpha,os}(1-\gamma) \left\| \Sigma_y^{1/2} a \right\|_2, \end{aligned} \tag{5.3}$$

where $f^{-1}_{\alpha,os}$ has been defined in Equation (3.23) and introduces a margin influencing the 2-norm in each one of the SOC constraints. In Figure 5.1 the margin behavior with respect to γ is plotted as a monotonically increasing function. From what said in Section 4.1.1, assuming "small" α is equivalent to assuming that the distributions are more concentrated around the mode and reduces the margin in the SOCP reformulation. When $\alpha \to \infty$, the inverse function limit is:

$$\lim_{\alpha\to\infty} f^{-1}_{\alpha,os}(1-\gamma) = \sqrt{\frac{\gamma}{1-\gamma}}.$$

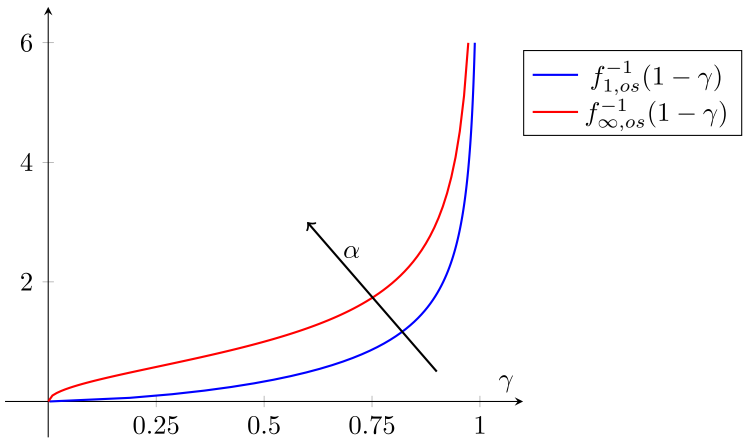

Figure 5.1: Inverse one-sided bounds behavior with respect to α and $1-\gamma$

Since $f^{-1}_{\alpha,os}(1-\gamma)$ is monotonically increasing in γ, we can define a new optimization variable

$$\omega := f^{-1}_{\alpha,os}(1-\gamma), \tag{5.4}$$

and rewrite our problem as

$$\begin{aligned} \text{maximize} \quad & \omega \\ \text{subject to:} \quad & \omega \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ & -b + a^\top \bar{x} \geq \omega \left\| \Sigma_x^{1/2} a \right\|_2 \\ & b - a^\top \bar{y} \geq \omega \left\| \Sigma_y^{1/2} a \right\|_2 . \end{aligned}$$

Given the optimum ω^*, a^* and b^*, the worst-case correct classification probability can be calculated as:

$$\gamma^* = 1 - f_{\alpha,os}(\omega^*), \tag{5.5}$$

where $f_{\alpha,os}$ is defined in (3.22). As the objective function is linear, the inequalities become tight at the optimum, i.e.

$$b^* = a^{*\top} \bar{x} - \omega^* \left\| \Sigma_x^{1/2} a^* \right\|_2 = a^{*\top} \bar{y} + \omega^* \left\| \Sigma_y^{1/2} a^* \right\|_2 . \tag{5.6}$$

By consequence, it is possible to neglect b and rewrite the optimization problem by summing the two constraints:

$$\begin{aligned} \text{maximize} \quad & \omega \\ \text{subject to:} \quad & \omega \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top (\bar{x} - \bar{y}) \geq \omega \left(\left\| \Sigma_x^{1/2} a \right\|_2 + \left\| \Sigma_y^{1/2} a \right\|_2 \right) . \end{aligned}$$

If $\bar{x} = \bar{y}$, then $\omega^* = 0$. By consequence also $\gamma^* = 0$ and it is impossible to have a meaningful solution. We will assume for the rest of this Chapter that $\bar{x} \neq \bar{y}$.

In the case when $\bar{x} \neq \bar{y}$, we notice that the constraint is positively homogeneous in a: if a satisfies it, also qa does with $q \in \mathbb{R}_+$. Moreover, the constraint implies that $a^\top(\bar{x} - \bar{y}) \geq 0$ from the definition of norm. Hence, we can impose $a^\top(\bar{x} - \bar{y}) = 1$ without loss of generality. This implies $a \neq 0$ and $\left\| \Sigma_x^{1/2} a \right\|_2 + \left\| \Sigma_y^{1/2} a \right\|_2 \neq 0$. Thus, we can write the optimization problem as:

$$\begin{aligned} \text{maximize} \quad & \omega \\ \text{subject to:} \quad & \omega \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\ & \omega \leq \frac{1}{\left\| \Sigma_x^{1/2} a \right\|_2 + \left\| \Sigma_y^{1/2} a \right\|_2} \\ & a^\top (\bar{x} - \bar{y}) = 1. \end{aligned}$$

Also in this case the cost function is linear. We can eliminate variable ω and rewrite the problem as follows:

$$\begin{aligned} \text{minimize} \quad & \left\| \Sigma_x^{1/2} a \right\|_2 + \left\| \Sigma_y^{1/2} a \right\|_2 \\ \text{subject to:} \quad & a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top (\bar{x} - \bar{y}) = 1. \end{aligned} \tag{5.7}$$

For positive definite Σ_x and Σ_y the problem is strictly convex and feasible for $\bar{x} \neq \bar{y}$. Thus, the optimal point is unique.

Given the optimum a^*, it is possible to obtain the optimal values of the decision variables of the initial problem (5.1) by computing:

$$\omega^* = \frac{1}{\left\|\Sigma_x^{1/2} a\right\|_2 + \left\|\Sigma_y^{1/2} a\right\|_2}$$

and using Equations (5.5) and (5.6).

It is interesting to notice that the solution to problem (5.7) does not change with α. This is due to the fact that we enforced γ being the same for both classes. Hence, even though we get different probabilistic guarantees while assuming different α, the optimal hyperplane is always the same independently from how unimodal the measures are.

5.2 Biased Minimax Probability Machines (BMPMs)

When one of the two classes is less important, it could be useful to reformulate the optimization problem with two different probability guarantees for the two classes trying to maximize only one of them and having a predefined lower bound on the less important one. This problem has been analyzed by Huang et al. in [29] without using unimodality assumption. In this section we will deal with Biased Minimax Probability Machines (BMPMs) using α-unimodality:

$$\begin{aligned} BMPM: \quad & \text{maximize} && \gamma \\ & \text{subject to:} && \gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ &&& \inf_{\mathbb{P} \in \mathcal{P}_{\alpha_x}(\bar{x}, \Sigma_x)} \mathbb{P}(a^\top x \geq b) \geq \gamma \\ &&& \inf_{\mathbb{P} \in \mathcal{P}_{\alpha_y}(\bar{y}, \Sigma_y)} \mathbb{P}(a^\top y \leq b) \geq \delta_0, \end{aligned} \tag{5.8}$$

where δ_0 is the fixed minimum probability of correct classification for class y. Please note that now the two distributions are assumed to have two different α: α_x and α_y. Analogously to what has been done in the previous Section, the chance constrained program can be formulated as a SOCP:

$$\begin{aligned} & \text{maximize} && \gamma \\ & \text{subject to:} && \gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ &&& -b + a^\top \bar{x} \geq f^{-1}_{\alpha_x, os}(1-\gamma) \left\|\Sigma_x^{1/2} a\right\|_2 \\ &&& b - a^\top \bar{y} \geq f^{-1}_{\alpha_y, os}(1-\delta_0) \left\|\Sigma_y^{1/2} a\right\|_2, \end{aligned} \tag{5.9}$$

As $f^{-1}_{\alpha_x,os}(1-\gamma)$ is monotonically increasing in γ it is possible to define the optimization variable $\omega_\gamma := f^{-1}_{\alpha_x,os}(1-\gamma)$ as we did in Equation (5.4). Please note that $f^{-1}_{\alpha_y,os}(1-\delta_0)$ is fixed by the values of α_y and δ_0. For notational convenience, we will define also $\omega_{\delta_0} := f^{-1}_{\alpha_y,os}(1-\delta_0)$. Moreover, as the objective function is linear, it is possible to sum the two constraints and remove variable b

using Equation (5.6). Then, the problem can be rewritten as:

$$\begin{aligned} \text{maximize} \quad & \omega_\gamma \\ \text{subject to:} \quad & \omega_\gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top(\bar{x} - \bar{y}) \geq \omega_\gamma \left\| \Sigma_x^{1/2} a \right\|_2 + \omega_{\delta_0} \left\| \Sigma_y^{1/2} a \right\|_2 . \end{aligned}$$

In the same way as in Section 5.1, as the constraint is positively homogeneous in a, we can rewrite the problem as

$$\begin{aligned} \text{maximize} \quad & \omega_\gamma \\ \text{subject to:} \quad & \omega_\gamma \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\ & 1 \geq \omega_\gamma \left\| \Sigma_x^{1/2} a \right\|_2 + \omega_{\delta_0} \left\| \Sigma_y^{1/2} a \right\|_2 \\ & a^\top(\bar{x} - \bar{y}) = 1. \end{aligned}$$

Since Σ_x is assumed to be always positive definite, it is possible to rewrite the first constraint as:

$$\omega_\gamma \leq \frac{1 - \omega_{\delta_0} \left\| \Sigma_y^{1/2} a \right\|_2}{\left\| \Sigma_x^{1/2} a \right\|_2}.$$

As the objective function is linear in ω_γ, the inequality becomes tight at the optimum and it is possible to optimize only over a by rewriting the problem as a Fractional Program (FP):

$$\begin{aligned} \text{maximize} \quad & \frac{1 - \omega_{\delta_0} \left\| \Sigma_y^{1/2} a \right\|_2}{\left\| \Sigma_x^{1/2} a \right\|_2} \\ \text{subject to:} \quad & a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top(\bar{x} - \bar{y}) = 1. \end{aligned} \tag{5.10}$$

Since we assumed that $\bar{x} \neq \bar{y}$, we can consider $a \in \mathbb{R}^n$ while enforcing $a \neq 0$ with the equality constraint. Hence, this FP is concave and, thus, every local optimum is a global optimum, see Schaible [39]. In order to solve it, we transform it into an equivalent concave program. According to [40, Equation (7)], by defining the transformation

$$p := \frac{1}{\left\| \Sigma_x^{1/2} a \right\|_2} a, \quad t := \frac{1}{\left\| \Sigma_x^{1/2} a \right\|_2},$$

the solution p^* and t^* of

$$\begin{aligned} \text{maximize} \quad & t - \omega_{\delta_0} \left\| \Sigma_y^{1/2} p \right\|_2 \\ \text{subject to:} \quad & p \in \mathbb{R}^n \setminus \{0\}, t \in \mathbb{R} \\ & p^\top(\bar{x} - \bar{y}) = t \\ & \left\| \Sigma_x^{1/2} p \right\|_2 \leq 1 \\ & t > 0, \end{aligned}$$

gives an optimal solution to (5.10), with:

$$a^* = \frac{p^*}{t^*}. \tag{5.11}$$

The optimal value of ω_γ is given by:

$$\omega_\gamma^* = \frac{1 - \omega_{\delta_0} \left\| \Sigma_y^{1/2} a^* \right\|_2}{\left\| \Sigma_x^{1/2} a^* \right\|_2}.$$

Finally, the optimal values of b^* and γ^* can be computed from ω_γ^* using Equations (5.6) and (5.5) respectively.

5.3 Minimum Error Minimax Probability Machines (MEMPMs)

Assuming the worst-case accuracies of the two classes to be the same as in the standard MPM or fixing one probability guarantee as in the BMPM, does not assure to minimize the worst-case error rate for future data. The problem of constructing the distributionally robust classifier that minimizes the overall misclassification error has been formulated in [30] as Minimum Error Minimax Probability Machine (MEMPM):

$$\begin{aligned} MEMPM: \quad & \text{maximize} \quad && \theta\gamma + (1-\theta)\delta \\ & \text{subject to:} && \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ & && \inf_{\mathbb{P} \in \mathcal{P}_{\alpha_x}(\bar{x}, \Sigma_x)} \mathbb{P}(a^\top x \geq b) \geq \gamma \\ & && \inf_{\mathbb{P} \in \mathcal{P}_{\alpha_y}(\bar{y}, \Sigma_y)} \mathbb{P}(a^\top y \leq b) \geq \delta, \end{aligned} \tag{5.12}$$

where $\theta \in [0,1]$ is the prior probability of class x and $1-\theta$ is the prior probability of class y.

An interesting interpretation of the problem comes from the objective function. Maximizing $\theta\gamma + (1-\theta)\delta$ can be viewed as maximizing the worst case accuracy for future data. If we change, without loss of generality, the objective function by inverting the sign, adding 1 and by adding and subtracting θ, we get the following equivalent formulation

$$\text{minimize} \quad \theta(1-\gamma) + (1-\theta)(1-\delta).$$

By noting that $1-\gamma$ is the probability of miscalsification of class x and that $1-\delta$ is the one of class y, it is clear that the solution of problem (5.12) minimizes the maximum Bayes error constructing the Bayes optimal hyperplane [7] in the worst-case scenario.

In Figure 5.2 the MPM and MEMPM classifiers are compared when the ambiguity sets for class x and y contain only one distribution each (the drawn ones) and when prior probabilities are equal, i.e. $\theta = 1-\theta = 1/2$. In the MPM case, the integrals of the two tails representing $1-\gamma$ and $1-\delta$ are equal because in the problem formulation we enforce $\gamma = \delta$. On the other hand, after solving MEMPM the separating hyperplane falls in the point where the two densities intersect because the prior probabilities are equal. This is also the most intuitive place where we would put the classifier by looking at the two distributions.

Analogously to what discussed in the previous Section, we now reformulate the problem in a tractable way using SOC constraints. The chance constrained program can be rewritten

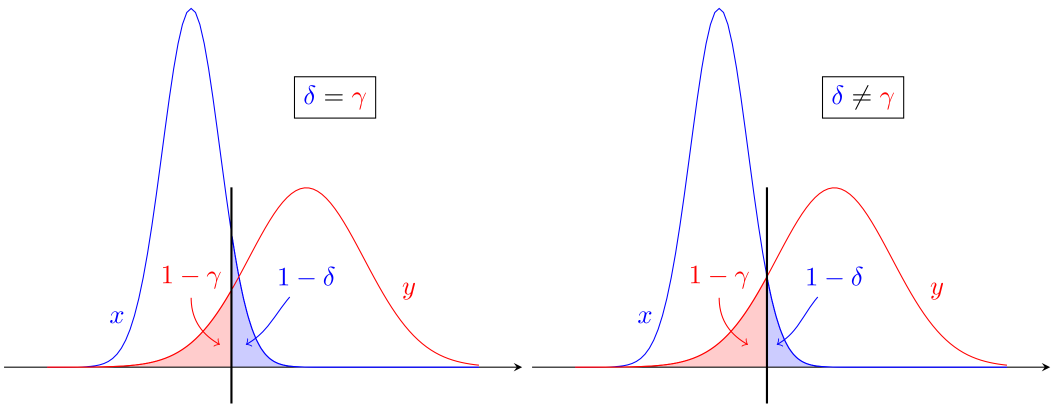

Figure 5.2: Comparison MPM (left) and MEMPM (right) classifiers in one dimension when prior probabilities are equal and the ambiguity sets for class x and y contain only one distribution each. The black lines represent the optimal classifiers.

as:

$$\begin{aligned} \text{maximize} \quad & \theta\gamma + (1-\theta)\delta \\ \text{subject to:} \quad & \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\ & -b + a^\top \bar{x} \geq f^{-1}_{\alpha_x,os}(1-\gamma) \left\| \Sigma_x^{1/2} a \right\|_2 \\ & b - a^\top \bar{y} \geq f^{-1}_{\alpha_y,os}(1-\delta) \left\| \Sigma_y^{1/2} a \right\|_2 . \end{aligned} \tag{5.13}$$

As $f^{-1}_{\alpha_x,os}(1-\gamma)$ and $f^{-1}_{\alpha_y,os}(1-\delta)$ are monotonically increasing in γ and δ respectively, it is possible to define two new variables

$$\omega_\gamma := f^{-1}_{\alpha_x,os}(1-\gamma), \quad \text{and} \quad \omega_\delta := f^{-1}_{\alpha_y,os}(1-\delta) \tag{5.14}$$

such that:

$$\gamma = 1 - f_{\alpha_x,os}(\omega_\gamma), \qquad \delta = 1 - f_{\alpha_y,os}(\omega_\delta). \tag{5.15}$$

We simplify b in the same way as we did deriving the MPM reformulation, by adding the constraints. Moreover, as in Section 5.1, we notice that the obtained constraint is homogeneous in a and rewrite the problem as:

$$\begin{aligned} \text{maximize} \quad & \theta\gamma + (1-\theta)\delta \\ \text{subject to:} \quad & \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\ & 1 \geq \omega_\gamma \left\| \Sigma_x^{1/2} a \right\|_2 + \omega_\delta \left\| \Sigma_y^{1/2} a \right\|_2 \\ & a^\top (\bar{x} - \bar{y}) = 1. \end{aligned} \tag{5.16}$$

In the following lemma we show that the solution is attained at the boundary of the feasible region described by the two constraints.

Lemma 5.3.1 ([30, Lemma 3]). *The maximum value of $\theta\gamma + (1-\theta)\delta$ under the constraints of problem* (5.16) *is achieved when:*

$$1 = \omega_\gamma \left\| \Sigma_x^{1/2} a \right\|_2 + \omega_\delta \left\| \Sigma_y^{1/2} a \right\|_2 .$$

Proof. Assume that the maximum is achieved when

$$1 > \omega_\gamma \left\| \Sigma_x^{1/2} a \right\|_2 + \omega_\delta \left\| \Sigma_y^{1/2} a \right\|_2 .$$

A new solution constructed by increasing γ (or ω_γ) by a small positive amount while mantaining δ and a fixed can be constructed satisfying the constraints providing a higher cost function value. ∎

Using last lemma, we can write

$$\omega_\gamma = \frac{1 - \omega_\delta \left\| \Sigma_y^{1/2} a \right\|_2}{\left\| \Sigma_x^{1/2} a \right\|_2}.$$

In addition, if we fix $\delta = \bar{\delta}$, the optimization problem can be transformed to:

$$\begin{aligned} \text{maximize} \quad & \frac{1 - \omega_{\bar{\delta}} \left\| \Sigma_y^{1/2} a \right\|_2}{\left\| \Sigma_x^{1/2} a \right\|_2} \\ \text{subject to:} \quad & a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top (\bar{x} - \bar{y}) = 1, \end{aligned} \tag{5.17}$$

that is equivalent to the BMPM Fractional Program (5.10). Hence, it can be converted to an equivalent concave program and solved efficiently. We can, then, update δ in order to find the optimal one. We denote as $\gamma_{\bar{\delta}}^*$ the optimal value we obtain from problem (5.17) and the function

$$\begin{aligned} g\colon [0,1] &\to [0,1] \\ \bar{\delta} &\mapsto \theta \gamma_{\bar{\delta}}^* + (1-\theta)\bar{\theta}, \end{aligned} \tag{5.18}$$

mapping from the fixed $\bar{\delta}$ to the relative MEMPM optimal cost. Hence, finding the optimal δ corresponds to performing a line search over the function $g(\delta)$ that, instead of being explicitly available, has to be computed from problem (5.18).

Although there are many ways to solve the line search problem, we will refer to the Quadratic Interpolation (QI) method, see [4]. QI finds the maximum point by updating a three-point pattern $\delta_1, \delta_2, \delta_3$ repeatedly. The new δ is denoted by δ_{new} and computed using the quadratic interpolation of the three-point pattern. At each iteration a new three point pattern is constructed using δ_{new} and two of $\delta_1, \delta_2, \delta_3$. It can be shown that this method converges superlinearly to a local optimum point [4]. Moreover, as discussed in [30], although MEMPM does not guarantee concavity in general, empirically it is often concave. Thus, the local optimum achieved using QI will often be global optimum.

After finding the optimal values for δ^* and γ^*, it is immediate to compute a^* from (5.11). Then, b^* can be explicitly computed from (5.6).

5.4 Moment Uncertainty

Empirical estimates of mean and covariance could happen to be inaccurate because of wrong data or limited data. In this case, the solution using plug-in estimates, could be unreliable and

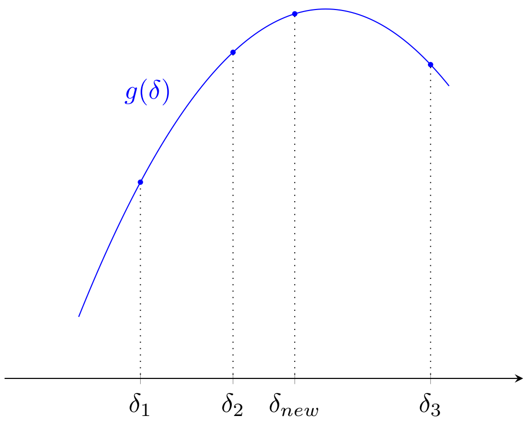

Figure 5.3: Example of QI line search method. The three-point pattern for the next iteration is $\delta_2, \delta_{new}, \delta_3$.

the worst-case accuracy lower than the computed bounds. We will reformulate the MEMPM using results from Section 4.1.3.

We will denote as $\hat{x}$ and $\hat{y}$ the empirical estimates of the mean of class x, y respectively and $\hat{\Sigma}_x$ and $\hat{\Sigma}_y$ the ones of covariance matrix. By assuming the support of classes x and y to be bounded by two balls of radiuses R_x and R_y respectively, we can reformulate Theorem 4.1.1 for two classes as follows:

Theorem 5.4.1. *Let $x^{(1)}, \ldots, x^{(N_x)} \in \mathcal{X} \subseteq \mathbb{R}^n$ be i.i.d. random samples of x and $y^{(1)}, \ldots, y^{(N_y)} \in \mathcal{Y} \subseteq \mathbb{R}^n$ be i.i.d. random samples of y such that*

$$R_x = \sup_{x \in \mathcal{X}} \|x\|_2, \qquad R_y = \sup_{y \in \mathcal{Y}} \|y\|_2$$

Then, provided that $N_x, N_y > \left(2 + \sqrt{2 \ln \frac{8}{\beta}}\right)^2$, the following holds with confidence at least $1 - \beta$:

$$\begin{aligned}
\|\bar{x} - \hat{x}\|_2 &\leq r_{1,x} & r_{1,x} &= \frac{R_x}{\sqrt{N_x}}\left(2 + \sqrt{2 \ln \frac{4}{\beta}}\right) \\
\left\|\Sigma_x - \hat{\Sigma}_x\right\|_F &\leq r_{2,x} & r_{2,x} &= \frac{2R_x^2}{\sqrt{N_x}}\left(2 + \sqrt{2 \ln \frac{8}{\beta}}\right) \\
\|\bar{y} - \hat{y}\|_2 &\leq r_{1,y} & r_{1,y} &= \frac{R_x}{\sqrt{N_y}}\left(2 + \sqrt{2 \ln \frac{4}{\beta}}\right) \\
\left\|\Sigma_y - \hat{\Sigma}_y\right\|_F &\leq r_{2,y} & r_{2,y} &= \frac{2R_y^2}{\sqrt{N_y}}\left(2 + \sqrt{2 \ln \frac{8}{\beta}}\right).
\end{aligned}$$

Using the same algebraic manipulations as in Equation (4.18), we can robustify problem (5.13) as:

$$\begin{aligned}
&\text{maximize} && \theta\gamma + (1-\theta)\delta \\
&\text{subject to:} && \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\},\ b \in \mathbb{R} \\
& && -b + a^\top \hat{x} - r_{1,x} \|a\|_2 \geq f^{-1}_{\alpha_x,os}(1-\gamma) \left\| \left(\hat{\Sigma}_x + r_{2,x} I\right)^{1/2} a \right\|_2 \\
& && b - a^\top \hat{y} - r_{1,y} \|a\|_2 \geq f^{-1}_{\alpha_y,os}(1-\delta) \left\| \left(\hat{\Sigma}_y + r_{2,y} I\right)^{1/2} a \right\|_2 .
\end{aligned} \tag{5.19}$$

As previously discussed, we can define ω_γ and ω_δ as in Equation (5.14) and add the constraints to neglect b obtaining:

$$\begin{aligned}
&\text{maximize} && \theta\gamma + (1-\theta)\delta \\
&\text{subject to:} && \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\
& && a^\top (\hat{x} - \hat{y}) - \|a\|_2 (r_{1,x} + r_{1,y}) \geq \omega_\gamma \left\| \left(\hat{\Sigma}_x + r_{2,x} I\right)^{1/2} a \right\|_2 + \omega_\delta \left\| \left(\hat{\Sigma}_y + r_{2,y} I\right)^{1/2} a \right\|_2 .
\end{aligned} \tag{5.20}$$

Optimal b can be computed once we get the optimal solution $a^*, \omega^*_\gamma, \omega^*_\delta$ as:

$$\begin{aligned}
b^* &= a^{*\top} \hat{y} + r_{1,y} \|a^*\|_2 + \omega_\delta \left\| \left(\hat{\Sigma}_y + r_{2,y} I\right)^{1/2} a^* \right\|_2 \\
&= a^{*\top} \hat{x} - r_{1,x} \|a^*\|_2 - \omega_\gamma \left\| \left(\hat{\Sigma}_x + r_{2,x} I\right)^{1/2} a^* \right\|_2 .
\end{aligned} \tag{5.21}$$

Since the constraint is still positively homogeneous in a and the problem is equivalent to

$$\begin{aligned}
&\text{maximize} && \theta\gamma + (1-\theta)\delta \\
&\text{subject to:} && \gamma \in \mathbb{R}_+,\ \delta \in \mathbb{R}_+,\ a \in \mathbb{R}^n \setminus \{0\} \\
& && 1 \geq \omega_\gamma \left\| \left(\hat{\Sigma}_x + r_{2,x} I\right)^{1/2} a \right\|_2 + \omega_\delta \left\| \left(\hat{\Sigma}_y + r_{2,y} I\right)^{1/2} a \right\|_2 \\
& && a^\top (\hat{x} - \hat{y}) - \|a\|_2 (r_{1,x} + r_{1,y}) = 1
\end{aligned} \tag{5.22}$$

From Lemma 5.3.1, the solution is attained at the boundary of the feasible region where the inequality is tight. Thus, by noting that ω_γ is a positively increasing function of γ, by choosing

$$\omega_\gamma = \frac{1 - \omega_\delta \left\| \left(\hat{\Sigma}_y + r_{2,y} I\right)^{1/2} a \right\|_2}{\left\| \left(\hat{\Sigma}_x + r_{2,x} I\right)^{1/2} a \right\|_2},$$

and by fixing $\delta = \bar{\delta}$, the optimization problem becomes

$$\begin{array}{ll} \text{maximize} & \dfrac{1 - \omega_\delta \left\| \left(\hat{\Sigma}_y + r_{2,y} I \right)^{1/2} a \right\|_2}{\left\| \left(\hat{\Sigma}_x + r_{2,x} I \right)^{1/2} a \right\|_2} \\ \text{subject to:} & a \in \mathbb{R}^n \setminus \{0\} \\ & a^\top (\hat{x} - \hat{y}) - \|a\|_2 (r_{1,x} + r_{1,y}) = 1, \end{array} \tag{5.23}$$

which is a FP similar to the BMPM. Since we the probability that $\hat{x} = \hat{y}$ is null, we can consider $a \in \mathbb{R}^n$ while enforcing $a \neq 0$ with the equality constraint. Thus, it is a concave FP and, thus, every local optimum is a global optimum, see Schaible [39]. It can be rewritten as a concave program using similar transformations to the ones in Section 5.2. According to [40, Equation (7)], by defining the transformation

$$p := \frac{1}{\left\| (\Sigma_x + r_{2,x} I)^{1/2} a \right\|_2} a, \quad t := \frac{1}{\left\| (\Sigma_x + r_{2,x} I)^{1/2} a \right\|_2},$$

the solution p^* and t^* of

$$\begin{array}{ll} \text{maximize} & t - \omega_{\bar{\delta}} \left\| \left(\hat{\Sigma}_y + r_{2,y} I \right)^{1/2} p \right\|_2 \\ \text{subject to:} & p \in \mathbb{R}^n \setminus \{0\}, t \in \mathbb{R} \\ & p^\top (\hat{x} - \hat{y}) - \|p\|_2 (r_{1,x} + r_{1,y}) = t \\ & \left\| \left(\hat{\Sigma}_x + r_{2,x} I \right)^{1/2} p \right\|_2 \leq 1 \\ & t > 0, \end{array}$$

gives an optimal solution to (5.23), with:

$$a^* = \frac{p^*}{t^*}.$$

The optimal value of ω_γ is given by:

$$\omega_\gamma^* = \frac{1 - \omega_{\bar{\delta}} \left\| \left(\hat{\Sigma}_y + r_{2,y} I \right)^{1/2} a^* \right\|_2}{\left\| \left(\hat{\Sigma}_x + r_{2,x} I \right)^{1/2} a^* \right\|_2}.$$

The optimal values of b^* and γ^* can be computed from ω_γ^* using Equations (5.21) and (5.15) respectively.

Then, as we did in Section 5.3, we can update iteratively δ using a line search QI algorithm until we find the optimal one giving the best worst-case accurancy.

5.5 Benchmarks

We evaluate the MPM and the MEMPM algorithms against the standard SVM classifier on two UCI machine learning repository datasets: Ionosphere and Prima diabetes. The prior probabilities θ and $1-\theta$ are estimated from the proportion of data labeled as one class over the other one. The used mean and covariance matrix of each class is the plug-in estimate from data. The α-unimodality index is chosen equal to the dimension of the data vectors for both classes.

Each dataset was randomly partitioned into 90% training and 10% test sets. The reported results are averaged over 50 random partitions. In Tables 5.1 and 5.2 there is a comparison of the results using each algorithm with the estimated Test Set Accuracy (TSA).

	MEMPM					
	γ	δ	TSA_x	TSA_y	$\theta\gamma+(1-\theta)\delta$	TSA
Ionosphere	$44.98 \pm 0.4\%$	$94.73 \pm 0.2\%$	$67.77 \pm 0.4\%$	$99.02 \pm 0.3\%$	$76.89 \pm 0.1\%$	$87.64 \pm 0.3\%$
Prima diabetes	$18.39 \pm 0.3\%$	$66.20 \pm 0.2\%$	$57.77 \pm 1.3\%$	$86.35 \pm 0.4\%$	$49.53 \pm 0.0\%$	$76.25 \pm 0.6\%$

Table 5.1: MEMPM classifier benchmarks.

The MPM approach has the worst performance together with the worst-case accuracy estimate γ for both datasets because it enforces the same γ for both classes. On the other hand, the MEMPM approach produces worst-case accuracy bounds much closer to the TSA for each class. However, in the Prima diabetes dataset, the worst-case estimates are still far from the empirical ones. This depends on the chosen α unimodality index which is probably lower than the one we assumed.

We notice that, for these benchmark data, using plug-in estimates of mean and covariance matrix without the robust formulation is successful because the TSA is always higher than the estimated one for both classes. However, this does not always happen in general and having poor mean and covariance estimates can produce a TSA lower than the estimated bound.

	MPM		SVM
	γ	TSA	TSA
Ionosphere	$63.19 \pm 0.1\%$	$84.16 \pm 1.2\%$	$87.46 \pm 0.5\%$
Prima diabetes	$32.13 \pm 0.0\%$	$75.89 \pm 0.3\%$	$77.11 \pm 0.1\%$

Table 5.2: MPM and SVM classifiers benchmarks.

Finally, it is important to remark that this MEMPM implementation improves the worst-case accuracy estimates for both classes compared to the ones in [30]. This is due to the fact that Huang et al. did not take into account unimodality of the distributions ($\alpha=\infty$).

6 Control

Another interesting application of optimization problems reformulations in Chapter 4 is the control of dynamical systems. In this Chapter we will compare the derived joint chance constrained programs reformulations with the Scenario Approach in both open and closed loop control schemes in water reservoir management setting.

6.1 Water Reservoir Problem

The model has been introduced by Andrieu et al. in [1] and modified by Zymler in [50] in a distributionally robust optimization setting. We extend this model to multiple customers.

The system consists of a single water reservoir with m clients, see Figure 6.1. The water released by the water reservoir at each time stage k is defined as $u_k \in \mathbb{R}^m$ and is used to produce electrical energy that is sold to each customer. The water level inside the reservoir is $x_k \in \mathbb{R}_+$. The system's uncertainty comes from precipitations $w_k \in \mathbb{R}$ that increase the water level. The model dynamics can be written as

$$x_{k+1} = Ax_k + Bu_k + Fw_k, \quad A = 1,\ B = -\mathbf{1}^\top,\ F = 1.$$

We will keep notation with A, B and F because the reformulation we will derive could be applied also to more complex systems by directly changing the matrices describing the dynamics. The initial water level is denoted by $x(0)$.

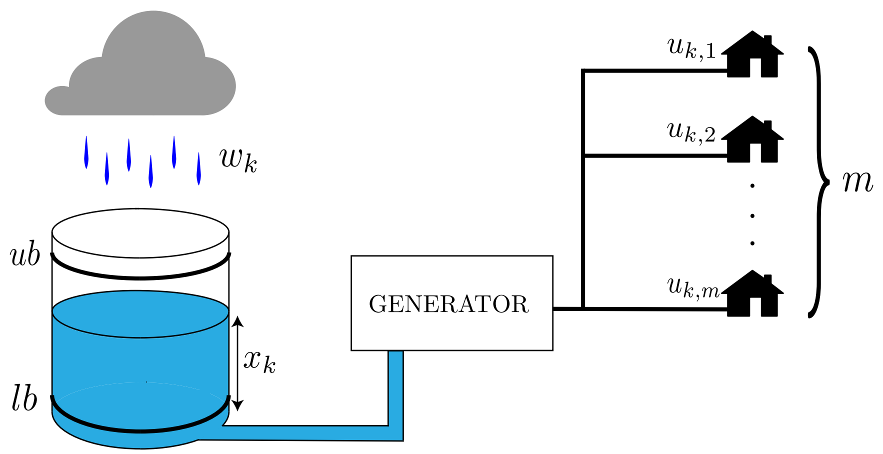

Figure 6.1: Water Reservoir

The control problem can be described in a Model Predictive Control (MPC) fashion as follows:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{k=0}^{T-1} c_k^\top u_k \\
\text{subject to:} \quad & u_k \in \mathbb{R}^m, \quad k = 1, \dots, T-1 \\
& u_k \geq u_{min}\mathbf{1}, \quad k = 0, \dots, T-1 \\
& x_{k+1} = Ax_k + Bu_k + Fw_k, \quad k = 0, \dots, T-1 \\
& x_0 = x(0) \\
& \mathbb{P}\begin{pmatrix} x_k \leq ub \\ x_k \geq lb \end{pmatrix} \geq 1 - \epsilon, \quad k = 1, \dots, T
\end{aligned} \tag{6.1}
$$

The objective is to maximize the profit by selling energy over the period T while ensuring that each customer gets at least u_{min} energy at each stage k. Moreover, the water level has to be within the upper bound ub and the lower bound lb at each time stage in the horizon with probability at least $1-\epsilon$.

It is possible to rewrite the problem in vector form in order to make the notation easier. We define

$$
\boldsymbol{x} = \begin{bmatrix} x_1, \dots, x_T \end{bmatrix}^\top, \; \boldsymbol{u} = \begin{bmatrix} u_0, \dots, u_{T-1} \end{bmatrix}^\top, \; \boldsymbol{w} = \begin{bmatrix} w_0, \dots, w_{T-1} \end{bmatrix}^\top, \; \boldsymbol{c} = \begin{bmatrix} c_0, \dots, c_{T-1} \end{bmatrix}^\top,
$$

as the state, the decision, the uncertainty and the cost vectors respectively. We also introduce matrices

$$
\mathcal{A} = \begin{bmatrix} A \\ \vdots \\ A^T \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B & 0 & & \dots & 0 \\ AB & B & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ & & & & 0 \\ A^{T-1}B & A^{T-2}B & & \dots & B \end{bmatrix}
$$

$$
\mathcal{F} = \begin{bmatrix} F & 0 & & \dots & 0 \\ AF & F & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ & & & & 0 \\ A^{T-1}F & A^{T-2}F & & \dots & F \end{bmatrix}
$$

system dynamics, can be trewritten as:

$$
\boldsymbol{x} = \mathcal{A}x_0 + \mathcal{B}\boldsymbol{u} + \mathcal{F}\boldsymbol{w},
$$

The single state x_k in the horizon can be described using intermediate vectors

$$
\boldsymbol{x}_k = \begin{bmatrix} x_1, \dots, x_k \end{bmatrix}^\top, \; \boldsymbol{u}_k = \begin{bmatrix} u_0, \dots, u_{k-1} \end{bmatrix}^\top, \; \boldsymbol{w}_k = \begin{bmatrix} w_0, \dots, w_{k-1} \end{bmatrix}^\top,
$$

and the rows of matrices $\mathcal{A}, \mathcal{B}$ and $\mathcal{F}$ defined as follows:

$$
\mathcal{A}_k = A^k, \quad \mathcal{B}_k = \begin{bmatrix} A^{k-1}B & A^{k-2}B & \dots & B \end{bmatrix}, \quad \mathcal{F}_k = \begin{bmatrix} A^{k-1}F & A^{k-2}F & \dots & F \end{bmatrix}.
$$

Then, we can write x_k as:

$$
x_k = \mathcal{A}_k x_0 + \mathcal{B}_k \boldsymbol{u}_k + \mathcal{F}_k \boldsymbol{w}_k. \tag{6.2}
$$

Hence, by plugging (6.2) inside problem (6.1) and by using the vector formulations explained above, (6.1) can be rewritten as:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}^\top \boldsymbol{u} \\ \text{subject to:} \quad & \boldsymbol{u} \in \mathbb{R}^{mT} \\ & \boldsymbol{u} \geq u_{min}\mathbf{1} \\ & \mathbb{P}\begin{pmatrix} \mathcal{F}_k \boldsymbol{w}_k \geq ub - \mathcal{A}_k x(0) - \mathcal{B}_k \boldsymbol{u}_k \\ -\mathcal{F}_k \boldsymbol{w}_k \geq -lb + \mathcal{A}_k x(0) + \mathcal{B}_k \boldsymbol{u}_k \end{pmatrix} \geq 1 - \epsilon, \quad k = 1, \ldots, T. \end{aligned} \tag{6.3}$$

This is a LP with $T-1$ different chance constraints, each of them including two joint chance constraints. Thus, it can be analyzed with the methods described in Chapter 4.

6.2 Reformulations

Each chance constraint at time k depends on the uncertainty vector $\boldsymbol{w}_k \in \mathbb{R}^k$. Hence, if the horizon increases, also the dimension of the uncertainty inside the chance constraints related to the last stages in the horizon T increases. We will denote by $\mu_k \in \mathbb{R}^k$ and $\Sigma_k \in \mathbb{S}_+^k$ the mean and the covariance of $\boldsymbol{w}_k$. We have N_k samples of each $\boldsymbol{w}_k$: $\left\{\boldsymbol{w}_k^{(1)}, \ldots, \boldsymbol{w}_k^{(N)}\right\}$.

If we assume the uncertainty w_k acting on the single stage bounded by a ball in $\mathbb{R}$ of radius R, we can bound each uncertainty vector $\boldsymbol{w}_k$ by a ball in $\mathbb{R}^k$ of radius $R_k = \sqrt{k}R$. Then, the robustifications of Bonferroni Approximation (BA) and Minimum Volume Ellipsoid Approximation (MVEA) in Section 4.1.3.1 can be applied to problem (6.3) obtaining the following LP:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}^\top \boldsymbol{u} \\ \text{subject to:} \quad & \boldsymbol{u} \in \mathbb{R}^{mT} \\ & \boldsymbol{u} \geq u_{min}\mathbf{1} \\ & ub - \mathcal{A}_k x(0) - \mathcal{B}_k \boldsymbol{u}_k - \mathcal{F}_k \mu_k - r_{1,k} \left\|\mathcal{F}_k\right\|_2 \geq \rho \left\|(\Sigma_k + r_{2,k} I)^{1/2} \mathcal{F}_k\right\|_2 \quad k = 1, \ldots, T \\ & -lb + \mathcal{A}_k x(0) + \mathcal{B}_k \boldsymbol{u}_k + \mathcal{F}_k \mu_k - r_{1,k} \left\|\mathcal{F}_k\right\|_2 \geq \rho \left\|(\Sigma + r_{2,k} I)^{1/2} \mathcal{F}_k\right\|_2 \quad k = 1, \ldots, T, \end{aligned} \tag{6.4}$$

where $r_{1,k}$ and $r_{2,k}$ are the radiuses defined in Theorem 4.1.1 with respect to the empirical mean $\hat{\mu}_k$ and covariance $\hat{\Sigma}_k$ of uncertainty vector $\boldsymbol{w}_k$. Please note that, at each time stage k, the minimum number of samples of $\boldsymbol{w}_k$ required by Theorem (4.1.1) is:

$$N_k > \left(2 + \sqrt{2 \ln \frac{4}{\beta}}\right)^2.$$

This formulation is simpler than the SOCP we would expect from Equation (4.12) because in this case $\tilde{a}(x)$ defined in Chapter 4 is constant. This happens because the control input has no feedback and acts in open-loop. Hence, the optimization is even faster.

Coefficient ρ is the one defined in Table 4.1 depending on the approximation used: BA or MVEA. Please note that this formulation can be used within the IEA. The solution $\boldsymbol{u}$ to problem (6.4) will satisfy each of the chance constraints in (6.3) with probability at least $1-\epsilon$ with confidence $1-\beta$.

Alternatively, by using the Multivariate Sampled Chebyshev Approximation (MSCA) from Section 4.1.3.2, it is possible to reformulate problem (6.3) using directly plug-in estimates $\hat{\mu}_k$ and $\hat{\Sigma}_k$ obtaining:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}^\top \boldsymbol{u} \\ \text{subject to:} \quad & \boldsymbol{u} \in \mathbb{R}^{mT} \\ & \boldsymbol{u} \geq u_{min}\mathbf{1} \\ & ub - \mathcal{A}_k x(0) - \mathcal{B}_k \boldsymbol{u}_k - \mathcal{F}_k \hat{\mu}_k \geq \lambda_k \left\| \hat{\Sigma}_k^{1/2} \mathcal{F}_k \right\|_2, \ k = 1, \ldots, T \\ & -lb + \mathcal{A}_k x(0) + \mathcal{B}_k \boldsymbol{u}_k + \mathcal{F}_k \hat{\mu}_k \geq \lambda_k \left\| \hat{\Sigma}^{1/2} \mathcal{F}_k \right\|_2, \ k = 1, \ldots, T, \end{aligned} \tag{6.5}$$

where, from Equation (4.24)

$$\lambda_k = \sqrt{\frac{k(N^2-1)}{N(\epsilon N - k)}}, \quad k = 1, \ldots, T,$$

and the minimum number of samples of $\boldsymbol{w}_k$ required at each stage is:

$$N_k > \frac{k}{\epsilon}, \quad k = 1, \ldots, T.$$

Also in this case, we get a LP instead of a SOCP because the input does not include any feedback. The optimization is, thus, faster. The solution to problem (6.5) will have the following probabilistic guarantees: optimal $\boldsymbol{u}$ will violate the chance constraint of problem (6.3) at each stage k defined by the $N+1$-th sample $\boldsymbol{w}_k^{(N+1)}$, with probability lower than ϵ.

6.3 Benchmarks

In this section we compare the algorithms in simulation. The chosen uncertainty w_k is an uniform distribution with mean 2 and support length 0.1. We set $ub = 9$ and $lb = 1$.

The tests are executed on a Macbook Pro with Intel 2.8GHz i7 processor and 16GB of RAM. We used solver MOSEK interfaced with MATLAB 2014a. The obtained computation times include only the solver execution and sampling times, but not the problem formulation.

6.3.1 Open-loop

We compare in open loop four different data-driven approaches with $\epsilon = 0.1$, $\beta = 0.01$ and the same number of samples N_k per stage:

- Bonferroni Approximation (BA) assuming $\alpha = \infty$
- Bonferroni Approximation (BA) assuming $\alpha = 1$
- Minimum Volume Ellipsoid Approximation (MVEA) assuming $\alpha = 1$

- Iterative Ellipsoid Approximation (IEA) assuming $\alpha = 1$
- Scenario Approach (SA) with two layers of probability

The results are averaged 10^2 times and the behavior is displayed in in Figure 6.2. The SA water level is the closest one to the boundaries while the others are more conservative keeping the water level far from the constraints.

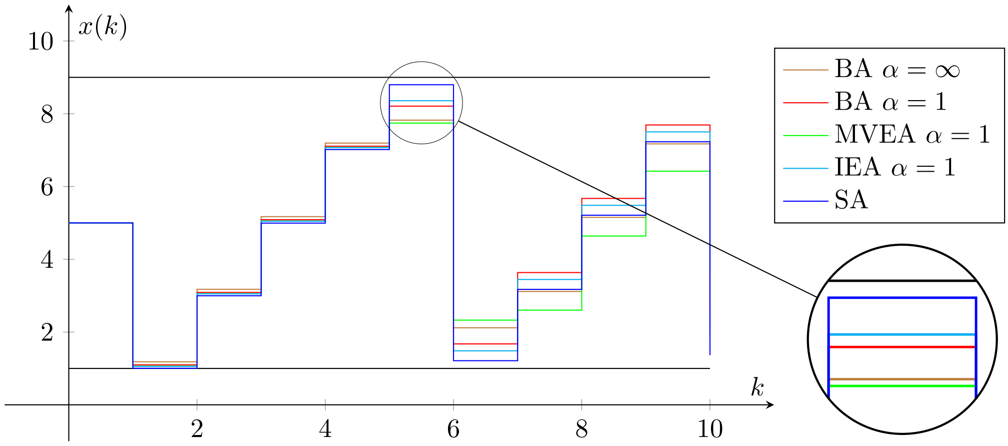

Figure 6.2: Water Reservoir open-loop behaviors

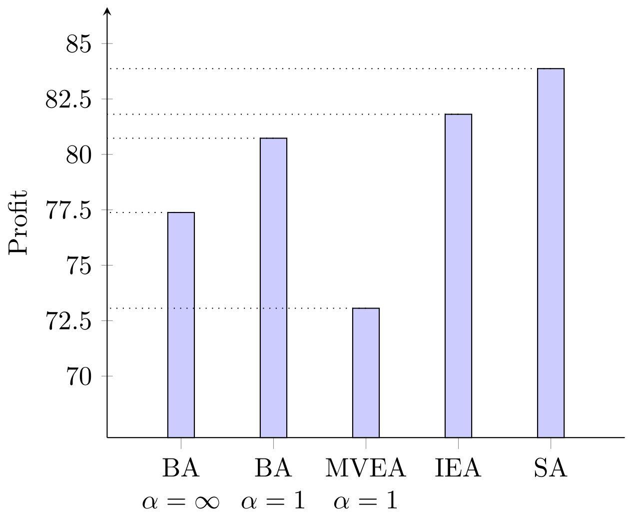

Figure 6.3: Water Reservoir open-loop profits

The obtained profits are displayed in Figure 6.3. The BA provide good results because the number of joint chance constraints per each stage are only $n_g = 2$. Thus, from what discussed in Section 4.1.2, it performs better than the MVEA. Moreover, BA with $\alpha = 1$ improves over the case when $\alpha = \infty$ because of the unimodality assumption. The MVEA not only provides worse performances with respect to the BA, but it provides the worst performances between all the compared methods. This happens because it depends on the dimension of the uncertainty $\boldsymbol{\omega}_k$ that grows along the horizon until dimension T. The IEA, while still optimizing over the ellipsoid, provides better performances than the BA and the MVEA because of the iterative reshaping explained in Section 4.1.2. The SA shows best performances in terms of profits.

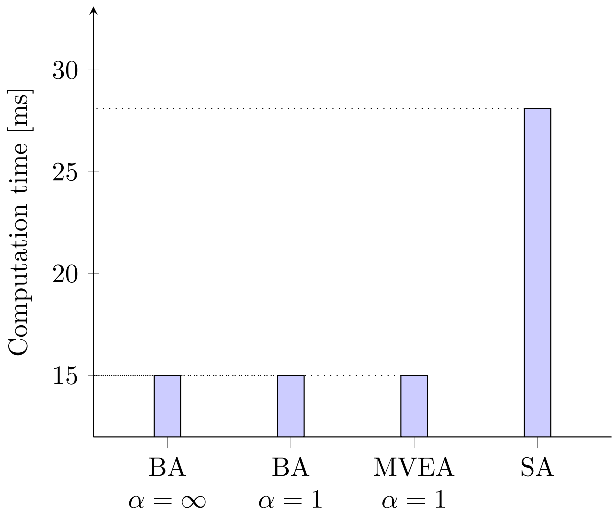

Figure 6.4: Water Reservoir open-loop computation times

Computation times are displayed in Figure 6.4. Since both the BA and MVEA solve a LP, on average their computation time is the same. Please note that these methods can all be solved within less than 20 ms and thus, they could be applied to systems with fast dynamics like quadrocopters or autonomous race cars. The SA shows higher computation times due to the fact that it solves a big LP with N_k constraints per stage. The IEA is not shown in the table because it provides the worst computation time of 5 sec and it cannot be displayed correctly in scale with the other methods. Please note that optimizing the solution times for these algorithms is beyond the scope of this work, and these computation times could be greatly improved for each algorithm.

6.3.2 Closed-loop

All the approaches of the previous section could be compared also in closed loop, but their profits would then be influenced by the closed-loop interaction of the controller with system dynamics. For this reason, even though important, this comparison is not fair.

In this section we will compare only two probabilistic methods we presented with one layer of probability $\epsilon = 0.1$ and the same samples N_k at each stage:

- Multivariate Sampled Chebyshev Approximation (MSCA)
- Scenario Approach (SA)

The results are averaged 10^2 times and the behavior displayed in Figure 6.5. Compared to the open-loop results in Figure 6.2, these benchmarks show a water level behavior much closer to the boundaries because the information about the current position is updated at each stage k and the input computed acccordingly.

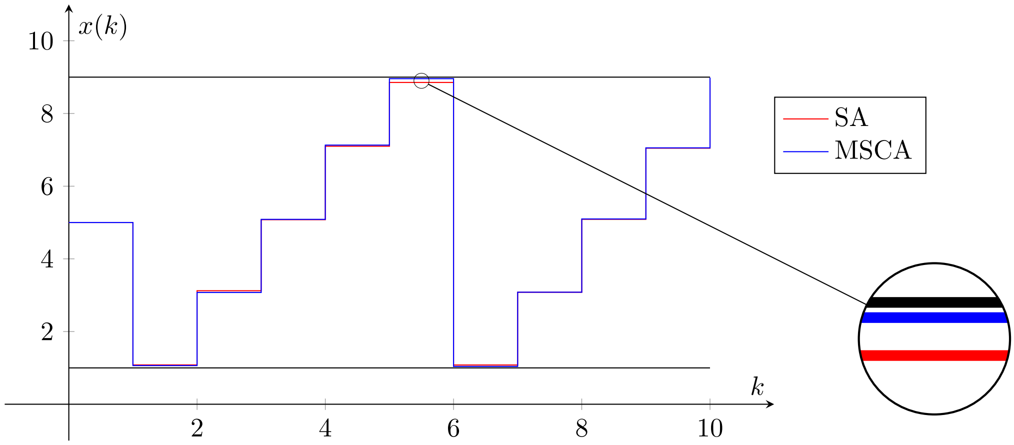

Figure 6.5: Water Reservoir closed-loop results

The profits for both methods are shown in Figure 6.6. The MSCA outperforms the SA. It is important to underline, however, that at each stage k, the SA has a better open-loop performance that decreases when closing the loop. This is one of the cases where a more conservative approach in open-loop (the MVEA) performs in a better way when the loop is closed. Please note this is does not happen always in general.

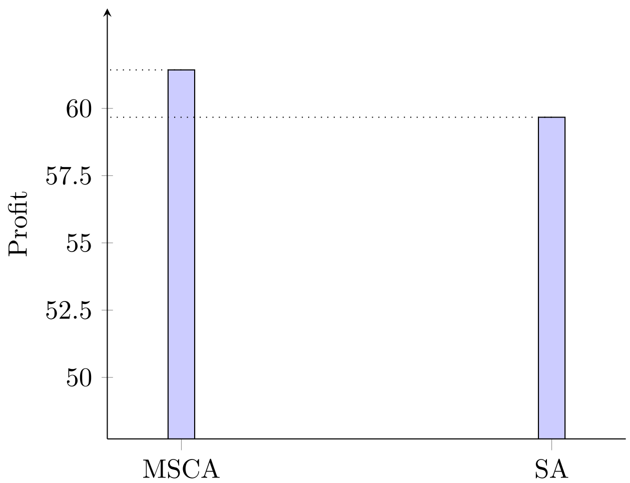

Figure 6.6: Water Reservoir closed-loop profits

The computation times are shown in Figure 6.7. Since the samples required for these approaches with one layer of probability are lower than the ones for the two layers, the computation times are

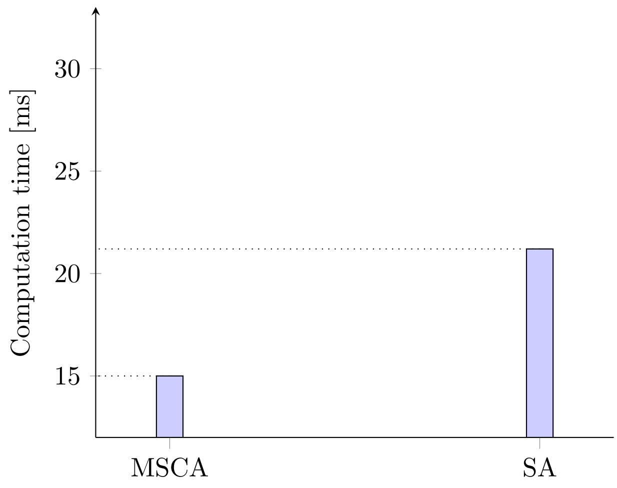

Figure 6.7: Water Reservoir closed-loop computation times

reduced. The MSCA does not change complexity, so on average it requires the same computation time as the other methods in Figure 6.4. The SA, on the other hand, reduce the computational effort to 21.2 ms.

7 Conclusions

We presented different methods to reformulate uncertain optimization programs with chance constraints. By extending Gauss inequalities first to the α-unimodality framework and then to multiple dimensions, we derived linear chance constraints reformulations. We exactly reformulated distributionally robust single linear chance constraints as Second-Order Cone (SOC) ones using α-unimodality and then, we extended this result to multiple chance constraints using two approaches. The first one is based on Bonferroni inequality dealing with each chance constraint independently and the second one makes use of ellipsoidal uncertainty sets having probabilistic guarantees. Moreover, we generalized these results to data-driven settings using two approaches. The first method is based on an extension of Hoeffding's inequality for vectors and matrices in order to describe and include in the SOCPs our degree of knowledge of the distribution moments. The second method makes use of a generalization of the empirical Chebyshev inequality in multiple dimensions.

Afterwards, these methods are compared against state of the art approaches in two different research fields.

We applied single linear chance constrained reformulations to the Machine Learning problem of linear binary classifications in two real world datasets and compared the solution to the standard Support Vector Machines (SVMs). The results showed a very similar Test Set Accuracy (TSA) on cross-validation data between the two methods. In addition, from the structure of our optimization problem, the obtained results come with probabilistic guarantees of worst-case misclassification and do not require any additional cross-validation test as usually done with SVMs.

Multiple linear chance constraints reformulations were applied to a water reservoir management example formulated as an MPC control problem. The results are compared to the Scenario Approach (SA) in both open and closed-loop. In open-loop, even though the SA showed better performance in term of cost function, our methods gave slightly more conservative solution while keeping the computation time within 20 ms per optimization. Moreover, the results from the closed-loop comparison not only showed better computational performances, but also a better cost function outperforming the SA.

7.1 Directions for Future Work

There are several ways to improve and generalize our results.

Our methods assumed to know the α-unimodality index being usually equal to the dimension of the uncertainty. Even though this case corresponds to the intuitive notion of unimodality in multiple dimensions, in practice distributions might have a smaller α that would lead to less pessimistic chance constrained approximations. It would be interesting to develop algorithms to estimate the minimum α from the available historical data and to include this information in our reformulations.

In addition, the empirical Chebyshev inequality extension we derived does not require any assumption on the distribution. Even though this result is remarkably general, the unimodality assumption would definitely improve the related bounds. Thus, combining an empirical estimate of α to the empirical Chebyshev inequality taking into account unimodality, would give different data-driven results that could outperform the proposed multivariate empirical Chebyshev inequality.

Finally, other kinds of uncertainty sets (e.g. boxes) with probabilistic guarantees could be analyzed and compared to the ellipsoidal ones used in this work. This would give us the freedom to choose between several kinds of sets using the ones performing better in the considered optimization problem.

Appendices

A Mathematical tools for Linear Matrix Inequalities

A.1 Schur Complement

The Schur complements arise in several contexts, and appear in many formulas and theorems in a wide range of scientific fields. The basic definition is the following (see [9]):

Definition A.1.1 (Schur Complement). *Consider a matrix $X \in \mathbb{S}^n$ partitioned as:*

$$X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}, \tag{A.1}$$

where $A \in \mathbb{S}^k$. If A is invertible, the matrix

$$S = C - B^\top A^{-1} B \tag{A.2}$$

is called the Schur complement of A in X.

The Schur complement is useful to determine the conditions of positive definiteness or semidefiniteness of block matrix X. In particular:

$$X \succ 0 \iff A \succ 0 \text{ and } S \succ 0, \tag{A.3}$$

and

$$A \succ 0 \implies X \succeq 0 \iff S \succeq 0. \tag{A.4}$$

A.2 S-procedure

In this work there are often constraints enforcing a quadratic function (or quadratic form) to be negative whenever other quadratic functions (or quadratic forms) are all negative. This condition can be usually casted into a LMI, see [8]:

Theorem A.2.1 (S-procedure). *Let $\Phi_1, \dots, \Phi_p$ be quadratic functions of the variable $\zeta \in \mathbb{R}^n$:*

$$\Phi_i(\zeta) \coloneqq \zeta^\top A_i \zeta + 2b_i^\top \zeta + c_i, \quad i = 0, \dots, p, \tag{A.5}$$

where $A_i = A_i^\top$. Then, the condition

$$\Phi_0(\zeta) \geq 0 \quad \forall \zeta \qquad \textit{such that} \qquad \Phi_i(\zeta) \geq 0, \quad i = 1, \dots, p, \tag{A.6}$$

holds if

$$\exists \tau_1, \dots, \tau_p \geq 0 \quad \textit{such that} \quad \forall \zeta, \quad \Phi_0(\zeta) - \sum_{i=1}^{p} \tau_i \Phi_i(\zeta) \geq 0. \tag{A.7}$$

Using LMI formulation, condition (A.7) can be rewritten as:

$$\begin{bmatrix} A_0 & b_0 \\ b_0^\top & c_0 \end{bmatrix} - \sum_{i=1}^{p} \tau_i \begin{bmatrix} A_i & b_i \\ b_i^\top & c_i \end{bmatrix} \geq 0, \quad i = 1, \ldots, p. \tag{A.8}$$

Remark *If* $p = 1$ *and* $\exists\, \zeta_0$ *such that* $\Phi_1(\zeta_0) > 0$*, then conditions* (A.6) *and* (A.7) *are equivalent.*

B Ellipsoid Representations

Ellipsoids are convex sets that are widely used in robust optimization. We parametrize them in four different ways as done in [9].

The first is the most common one:

$$\mathcal{E}_1 = \{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\} \tag{B.1}$$

where $P = P^\top \succ 0$. The vector $x_c \in \mathbb{R}^n$ is the center of the ellipsoid while the matrix P determines how far the ellipsoid extends in every direction from x_c. The lengths of the semi-axes of $\mathcal{E}$ are given by the square roots of the eigenvalues of P, while the relative eigenvectors determine the directions. The volume of the ellipsoid is computed, from this parametrization, as:

$$\mathrm{Vol}(\mathcal{E}_1) = \frac{4}{3}\pi\sqrt{\det P} \tag{B.2}$$

The second parametrization is obtained by rewriting the following inequality

$$\begin{aligned}(x - x_c)^\top P^{-1}(x - x_c) \leq 1 &\iff \|P^{-1/2}(x - x_c)\|_2 \leq 1 \\ &\iff \|\underbrace{P^{-1/2}}_{A}x - \underbrace{P^{-1/2}x_c}_{b}\|_2 \leq 1 \\ &\iff \|Ax - b\|_2 \leq 1\end{aligned}$$

Hence, the ellipsoid parametrization becomes:

$$\mathcal{E}_2 = \{x \mid \|Ax - b\|_2 \leq 1\} \tag{B.3}$$

where $A^\top = A = P^{-1/2} \succ 0$. From the definition of A, we know that $\det P = (\det A^{-1})^2$. Thus, the volume of the ellipsoid computed with this parametrization is:

$$\mathrm{Vol}(\mathcal{E}_2) = \frac{4}{3}\pi \det A^{-1}. \tag{B.4}$$

The third ellipsoid parametrization can be obtained by rearranging the inequality used for the previous ones, in another way:

$$(x - x_c)^\top P^{-1}(x - x_c) \leq 1 \iff \|\underbrace{P^{-1/2}(x - x_c)}_{u}\|_2 \leq 1 \iff \|u\|_2 \leq 1,$$

where

$$u = P^{-1/2}(x - x_c) \quad \text{and then} \quad x = x_c + \underbrace{P^{1/2}}_{B}u = x_c + Bu.$$

The shape matrix is now $B = P^{1/2} \succ 0$. The parametrization, in this case, is:

$$\mathcal{E}_3 = \{x_c + Bu \mid \|u\|_2 \leq 1\}. \tag{B.5}$$

From the definition of B, we have that $\det P = (\det B)^2$. Thus, the volume with respect to this parametrization is

$$\mathrm{Vol}(\mathcal{E}_3) = \frac{4}{3}\pi \det B. \tag{B.6}$$

The fourth parametrization is useful when it is necessary to describe the ellipsoid with convex quadratic inequalities. By manipulating representation (B.3) we get:

$$\|Ax-b\|_2 \leq 1 \iff x^\top \underbrace{A^\top A}_{C} x + 2\underbrace{(-A^\top b)^\top}_{d^\top} x + \underbrace{b^\top b - 1}_{e} \leq 0 \iff x^\top Cx + 2d^\top x + e \leq 0, \tag{B.7}$$

where $C = A^\top A = P^{-1} \succ 0$. The parametrization, in this case is:

$$\mathcal{E}_4 = \{x \mid x^\top Cx + 2d^\top x + e \leq 0\} \tag{B.8}$$

From the definition of C, $\det P = \det C^{-1}$. In this way the ellipsoid is represented as the level sets of a quadratic function. The volume, then, becomes:

$$\mathrm{Vol}(\mathcal{E}_4) = \frac{4}{3}\pi\sqrt{\det C^{-1}}. \tag{B.9}$$

Notation

Probability

Ω	space of elementary events
$\mathcal{F}$	σ-algebra of subsets of Ω
$\mathbb{Q}$	probability measure on $(\Omega, \mathcal{F})$
$\mathbb{P}$	probability distribution $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$
$\mathbb{E}$	expected value operator
$\mathbb{V}\mathrm{ar}$	variance operator
δ_x	Dirac distribution in x

Sets

$\mathbb{N}$	the natural numbers
$\mathbb{Z}$	the integer numbers
$\mathbb{R}$	the real numbers
$\mathbb{R}_+$	the nonnegative real numbers
$\mathbb{R}_{++}$	the positive real numbers
$\mathbb{R}^n$	the vectors with real numbers components with dimension $n \times 1$
$\mathbb{R}^{n\times n}$	the matrices with real numbers components with dimension $n \times n$
$\mathbb{S}^n$	the symmetric matrices in $\mathbb{R}^{n\times n}$
$\mathbb{S}^n_+$	the symmetric positive semidefinite matrices in $\mathbb{R}^{n\times n}$
$\mathbb{S}^n_{++}$	the symmetric positive definite matrices in $\mathbb{R}^{n\times n}$
$\|A\|$	cardinality of set A
$\mathbf{1}_A$	indicator function of set A
A^c	complement of set A: $\mathbb{R}^n \setminus A$
(a, b)	open line segment between a and b, with $a, b \in \mathbb{R}^n$
$[a, b]$	closed line segment between a and b, with $a, b \in \mathbb{R}^n$
$\mathrm{Vol}(A)$	volume of set A

Inequalities

$A \leq B$	element-wise inequality between A and B

$A < B$	strict element-wise inequality between A and B
$A \preceq B$	conic inequality between symmetric matrices: $B - A \succeq 0$
$A \prec B$	strict conic inequality between symmetric matrices: $B - A \succ 0$

Vectors and Matrices

I_n	identity matrix in $\mathbb{R}^{n \times n}$
$\mathbf{1}$	vector of ones of appropriate dimension
$A^\top$	transpose of the matrix A
$\text{tr}(A)$	trace of the matrix A
A_i	i^{th} row of the matrix A
$\langle A, B \rangle$	inner product between matrices A and B defined as $\langle A, B \rangle \coloneqq \text{tr}\,(AB)$

Other notation

$\lceil t \rceil$	smallest integer not less than t
$\lfloor t \rfloor$	largest integer not more than t

Acronyms

Probability Theory

PDF	Probability density function
CDF	Cumulative distribution function
MVE	Minimum Volume Ellipsoid

Optimization Programs

LP	Linear Program(ming)
QP	Quadratic Program(ming)
SOC	Second-Order Cone
SOCP	Second-Order Cone Program(ming)
FP	Fractional Program(ming)
SDP	Semidefinite Program(ming)
LMI	Linear Matrix inequality
SOS	Sum-Of-Squares

Uncertain Programs Approximations

BA	Bonferroni Approximation
MVEA	Minimum Volume Ellipsoid Approximation
IEA	Iterative Ellipsoid Approximation
MSCA	Multivariate Sampled Chebyshev Approximation

Others

SA	Scenario Approach
SVM	Support Vector Machine
MPM	Minimax Probability Machine
BMPM	Biased Minimax Probability Machine
MEMPM	Minimum Error Minimax Probability Machine
QI	Quadratic Interpolation
TSA	Test Set Accuracy
MPC	Model Predictive Control

Bibliography

[1] L. Andrieu, R. Henrion, and W. Römisch. A model for dynamic chance constraints in hydro power reservoir management. *European Journal of Operational Research*, 207(2):579–589, 2010.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of Linear Programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.

[4] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

[5] D. Bertsimas and N. Kallus. Data-Driven Robust Optimization. *Operations Research (submitted)*, 2013.

[6] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer, 2011.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

[9] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[10] G. C. Calafiore. On the Expected Probability of Constraint Violation inSampled Convex Programs. *Journal of Optimization Theory and Applications*, 143(2):405–412, June 2009.

[11] G. C. Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.

[12] G. C. Calafiore and M. C. Campi. The Scenario Approach to Robust Control Design. *IEEE Transactions on Automatic Control*, 51(5):742–753, May 2006.

[13] G. C. Calafiore and L. El Ghaoui. On Distributionally Robust Chance-Constrained Linear Programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, Dec. 2006.

[14] M. Campi and S. Garatti. The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[15] F. P. Cantelli. Intorno ad un teorema fondamentale della teoria del rischio. *Bollettino dell'Associazione degli Attuari Italiani*, pages 1–23, 1910.

[16] P. Chebyshev. Des valeurs moyennes. *Journal de Mathématiques pures et Appliquées*, 12(2):177–184, 1867.

[17] X. Chen, M. Sim, and P. Sun. A Robust Optimization Perspective on Stochastic Programming. *Operations Research*, 55(6):1058–1071, Dec. 2007.

[18] E. Cinquemani, M. Agarwal, D. Chatterjee, and J. Lygeros. Convexity and convex approximations of discrete-time stochastic control problems with constraints. *Automatica*, 47(9):2082–2087, 2011.

[19] G. B. Dantzig. Linear Programming under Uncertainty. *Management Science*, 1(3-4):197–206, Apr. 1955.

[20] E. Delage and Y. Ye. Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems. *Operations Research*, 58(3):595–612, June 2010.

[21] S. Dharmadhikari and K. Joag-Dev. *Unimodality, convexity, and applications*. 1988.

[22] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, second edition, 2002.

[23] L. El Ghaoui, M. Oks, and F. Oustry. Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach. *Operations Research*, 51(4):543–556, Aug. 2003.

[24] L. El Ghaoui, F. Oustry, and H. Lebret. Robust Solutions to Uncertain Semidefinite Programs. *SIAM Journal on Optimization*, 9(1):33–52, Jan. 1998.

[25] J. E. Falk. Technical NoteExact Solutions of Inexact Linear Programs. *Operations Research*, 24(4):783–787, Aug. 1976.

[26] C. E. García, D. M. Prett, and M. Morari. Model predictive control: Theory and practice - A survey. *Automatica*, 25(3):335–348, May 1989.

[27] C. F. Gauss. Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Pars Prior. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, 33:321–327, 1821.

[28] J. Goh and M. Sim. Distributionally Robust Optimization and Its Tractable Approximations. *Operations Research*, 58(4-part-1):902–917, Apr. 2010.

[29] K. Huang, H. Yang, I. King, and M. R. Lyu. Learning Classifiers from Imbalanced Data Based on Biased Minimax Probability Machine. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2:558–563, 2004.

[30] K. Huang, H. Yang, I. King, M. R. Lyu, and L. Chan. The Minimum Error Minimax Probability Machine. *The Journal of Machine Learning Research*, 5:1253–1286, 2004.

[31] P. Kali and S. W. Wallace. *Stochastic programming*. Wiley, Chichester, 1994.

[32] G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan. A Robust Minimax Approach to Classification. *The Journal of Machine Learning Research*, 3:555–582, 2002.

[33] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[34] A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal on Optimization*, 17(4):969–996, Dec. 2006.

[35] R. R. Phelps. *Lectures on Choquet's Theorem*, volume 1757. Springer, 2001.

[36] I. Popescu. A Semidefinite Programming Approach To Optimal Moment Bounds for Convex Classes of Distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.

[37] R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, pages 21–41, 2000.

[38] J. G. Saw, M. C. K. Yang, and T. C. Mo. Chebyshev Inequality With Estimated Mean and Variance. *The American Statistician*, 38(2):130–132, 1984.

[39] S. Schaible. Fractional programming. *Zeitschrift für Operations Research*, 27(1):39–54, 1983.

[40] S. Schaible and J. Shi. Recent developements in fractional programming: single ratio and max-min case. *Nonlinear analysis and convex analysis*, (2):1–11, 2004.

[41] G. Schildbach, L. Fagiano, and M. Morari. Randomized Solutions to Convex Programs with Multiple Chance Constraints. *SIAM Journal on Optimization*, 23(4):2479–2501, Jan. 2013.

[42] A. Shapiro. On duality theory of conic linear problems. *Nonlinear Optimization and its Applications*, 57:135-155, 2001.

[43] J. Shawe-Taylor and N. Cristianini. Estimating the moments of a random vector with applications. *Proceedings of GRETSI Conference*, I:47–52, 2003.

[44] A. L. Soyster. Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming. *Operations Research*, 21(5):1154–1157, 1973.

[45] B. P. G. Van Parys, P. J. Goulart, and D. Kuhn. Generalized Gauss Inequalities via Semidefinite Programming. *optimization-online.org*, 2014.

[46] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari. Distributionally robust control of constrained stochastic systems. *optimization-online.org*, 2013.

[47] L. Vandenberghe and S. Boyd. Semidefinite Programming. *SIAM Review*, 38(1):49–95, Mar. 1996.

[48] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev Bounds via Semidefinite Programming. *SIAM Review*, 49(1):52–64, Jan. 2007.

[49] X. Zhang, S. Grammatico, G. Schildbach, P. J. Goulart, and J. Lygeros. On the Sample Size of Random Convex Programs with Structured Dependence on the Uncertainty. *Automatica (submitted)*, 2014.

[50] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, Nov. 2011.