Mathematical Programming manuscript No.
(will be inserted by the editor)

Generalized Gauss Inequalities via Semidefinite Programming

Bart P.G. Van Parys · Paul J. Goulart · Daniel Kuhn

Abstract A sharp upper bound on the probability of a random vector falling outside a polytope, based solely on the first and second moments of its distribution, can be computed efficiently using semidefinite programming. However, this Chebyshev-type bound tends to be overly conservative since it is determined by a discrete worst-case distribution. In this paper we obtain a less pessimistic Gauss-type bound by imposing the additional requirement that the random vector's distribution must be unimodal. We prove that this generalized Gauss bound still admits an exact and tractable semidefinite representation. Moreover, we demonstrate that both the Chebyshev and Gauss bounds can be obtained within a unified framework using a generalized notion of unimodality. We also offer new perspectives on the computational solution of generalized moment problems, since we use concepts from Choquet theory instead of traditional duality arguments to derive semidefinite representations for worst-case probability bounds.

1 Introduction

In classical probability theory, the Chebyshev inequality provides an upper bound on the tail probability of a univariate random variable based on limited moment information. The most common formulation of this inequality asserts that the probability that a random variable $\xi \in \mathbb{R}$ with distribution $\mathbb{P}$ differs from its mean by more than κ standard deviations is

$$\mathbb{P}(|\xi - \mu| \geq \kappa\sigma) \leq \begin{cases} \frac{1}{\kappa^2} & \text{if} \quad \kappa > 1, \\ 1 & \text{otherwise,} \end{cases} \tag{1}$$

where κ is a strictly positive constant, while μ and σ denote the mean and standard deviation of ξ under $\mathbb{P}$, respectively. The inequality (1) was discovered by Bienaymé [3] in 1853 and proved by Chebyshev [7] in 1867. An alternative proof was offered by Chebyshev's student Markov [20] in 1884. The popularity of this inequality arises largely from its distribution-free nature. It holds for any distribution $\mathbb{P}$ under which ξ has mean μ and variance σ^2, and therefore can be used to construct robust confidence intervals for ξ relying exclusively on first and second-order moment information. Moreover, the inequality is sharp in the sense that, for any fixed κ, there exists a distribution for which (1) holds as an equality.

Recent generalizations of the Chebyshev inequality (1) provide upper bounds on the probability of a multivariate random vector $\xi \in \mathbb{R}^n$ falling outside a prescribed confidence region $\Xi \subseteq \mathbb{R}^n$ if only a few low-order moments of ξ

Bart P.G. Van Parys
Automatic Control Laboratory, Swiss Federal Institute of Technology Zürich (ETHZ), Physikstrasse 3, 8092 Zürich, Switzerland
E-mail: bartvan@control.ee.ethz.ch

Paul J. Goulart
Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, United Kingdom
E-mail: paul.goulart@eng.ox.ac.uk

Daniel Kuhn
Risk Analytics and Optimization Chair, Swiss Federal Institute of Technology Lausanne (EPFL), Station 5, 1015 Lausanne, Switzerland
E-mail: daniel.kuhn@epfl.ch

are known. The best upper bound of this kind is given by the optimal value of the worst-case probability problem

$$\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\xi \notin \Xi), \tag{P}$$

where $\mathscr{P}$ represents a set of all distributions consistent with the available moment information. The problem (P) has a natural interpretation as a generalized moment problem. If all given moments can be expressed as expectations of polynomials and Ξ is described by polynomial inequalities, one can use a convex optimization approach by Bertsimas and Popescu [2] to approximate (P) by a hierarchy of increasingly accurate bounds, each of which is computed by solving a tractable semi-definite program (SDP). Vandenberghe et al. [37] later showed that (P) admits an exact reformulation as a single SDP whenever Ξ is described through linear and quadratic inequalities and $\mathscr{P}$ contains all distributions sharing the same first and second-order moments. The resulting generalized Chebyshev bounds are widely used across many different application domains, ranging from distributionally robust optimization [9] to chance-constrained programming [8, 39, 43], stochastic control [36], machine learning [16], signal processing [38], option pricing [1, 13, 18], portfolio selection and hedging [40, 44], decision theory [33] etc.

Under mild technical conditions, the solution of the equivalent SDP by Vandenberghe et al. [37] can be used to construct a *discrete* optimal distribution for (P) with at most $\frac{1}{2}(n+2)(n+1)$ discretization points, where n denotes the dimension of ξ. The existence of optimal discrete distributions has distinct computational benefits and can be viewed as the key enabling property that facilitates the SDP reformulation of (P). However, it also renders the corresponding Chebyshev bound rather pessimistic. Indeed, uncertainties encountered in real physical, technical or economic systems are unlikely to follow discrete distributions with few atoms. By accounting for such pathological distributions, problem (P) tends to overestimate the probability of the event $\xi \notin \Xi$.

In order to mitigate the over-conservatism of the Chebyshev bound, one could impose additional restrictions on $\mathscr{P}$ that complement the given moment information. A minimal structural property commonly encountered in practical situations is unimodality. Informally, a continuous probability distribution is unimodal if it has a center m, referred to as the *mode*, such that the probability density function is non-increasing with increasing distance from the mode. Note that most distributions commonly studied in probability theory are unimodal. So too are all stable distributions, which are ubiquitous in statistics as they represent the attractors for properly normed sums of independent and identically distributed random variables.

In 1821 Gauss [11] proved that the classical Chebyshev inequality (1) can be improved by a factor of $\frac{4}{9}$ when $\mathbb{P}$ is known to be unimodal with center $m = \mu$, that is,

$$\mathbb{P}(|\xi-\mu| \geq \kappa\sigma) \leq \begin{cases} \frac{4}{9\kappa^2} & \text{if} \quad \kappa > \frac{2}{\sqrt{3}}, \\ 1-\frac{\kappa}{\sqrt{3}} & \text{otherwise.} \end{cases} \tag{2}$$

The Gauss inequality (2) is again sharp but provides a much less pessimistic bound on the tail probability than the Chebyshev inequality (1) when ξ is known to have a unimodal distribution.

The purpose of this paper is to generalize the Gauss inequality (2) to multivariate distributions, providing a counterpart to the generalized Chebyshev inequality (1) described by Vandenberghe et al. [37]. More precisely, we seek an SDP reformulation of problem (P) for any polytopic confidence region Ξ assuming that $\mathscr{P}$ contains all *unimodal* distributions that share the same mode as well as the same first and second-order moments. Extensions of the univariate Gauss inequality involving generalized moments have previously been proposed by Sellke [30], while multivariate extensions have been investigated by Meaux et al. [21]. Popescu [27] uses elegant ideas from Choquet theory in conjunction with sums-of-squares polynomial techniques to derive *approximate* multivariate Gauss-type inequalities. However, to the best of our knowledge, until now no efficient algorithm is known to compute the underlying worst-case probabilities *exactly*. Families of distributions with unimodal marginals have also been used by Shapiro and Kleywegt [32] and Natarajan et al. [22] to mitigate the conservatism of minimax stochastic programming and distributionally robust optimization problems, respectively. However, these models seem not to admit straightforward generalizations to multivariate families of unimodal distributions.

Notation: We denote by $\mathbb{S}^n$ ($\mathbb{S}^n_+$) the set of all symmetric (positive semidefinite) matrices in $\mathbb{R}^{n\times n}$. For any $X, Y \in \mathbb{S}^n$ the relation $X \succeq Y$ ($X \preceq Y$) indicates that $X - Y \in \mathbb{S}^n_+$ ($Y - X \in \mathbb{S}^n_+$). Moreover, we denote by $[x, y] \subseteq \mathbb{R}^n$ the line segment connecting two points $x, y \in \mathbb{R}^n$. The indicator function $\mathbf{1}_B$ of a set $B \subseteq \mathbb{R}^n$ is defined as $\mathbf{1}_B(x) = 1$ if $x \in B$ and $1_B(x) = 0$ otherwise. Moreover, δ_B denotes the uniform distribution on B. If $B = \{x\}$ is a singleton, we use the shorthand δ_x instead of $\delta_{\{x\}}$ to denote the Dirac point distribution at x. The set of all Borel probability distributions on $\mathbb{R}^n$ is denoted by $\mathscr{P}_\infty$. For any $c \in \mathbb{R}$ we denote by $\lceil c \rceil$ the smallest integer not less than c.

2 Generalized Chebyshev and Gauss inequalities

We first formalize the problem statement and present our main results in a non-technical manner. The proofs are deferred to Section 4, which may be skipped by readers whose primary interest is in applications. Our aim is to derive sharp upper bounds on the probability of a multivariate random vector $\xi \in \mathbb{R}^n$ falling outside a prescribed confidence region $\Xi \subseteq \mathbb{R}^n$, given that the distribution $\mathbb{P}$ of ξ is only known to lie within a prescribed ambiguity set $\mathscr{P}$. We assume throughout that Ξ is an open polyhedron representable as a finite intersection of open half spaces,

$$\Xi = \left\{ \xi \in \mathbb{R}^n \ : \ a_i^\top \xi < b_i \ \forall i = 1, \ldots, k \right\}, \tag{3}$$

where $a_i \in \mathbb{R}^n$, $a_i \neq 0$, and $b_i \in \mathbb{R}$ for all $i = 1, \ldots, k$. Moreover, we will consider different ambiguity sets $\mathscr{P}$, reflecting different levels of information about the distribution of ξ. In the setting of the generalized Chebyshev bounds studied by Vandenberghe et al. [37], $\mathscr{P}$ is defined as

$$\mathscr{P}(\mu, S) = \left\{ \mathbb{P} \in \mathscr{P}_\infty \ : \ \int_{\mathbb{R}^n} \xi \ \mathbb{P}(\mathrm{d}\xi) = \mu, \quad \int_{\mathbb{R}^n} \xi \xi^\top \ \mathbb{P}(\mathrm{d}\xi) = S \right\}, \tag{4}$$

where $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$, while $\mathscr{P}_\infty$ represents the set of all distributions on $\mathbb{R}^n$. Thus, $\mathscr{P}(\mu, S)$ contains all distributions that share the same mean μ and second-order moment matrix S.

Theorem 1 (Generalized Chebyshev bounds [37]) *If Ξ is a polytope of the form* (3)*, the worst-case probability problem* (P) *with ambiguity set $\mathscr{P}(\mu, S)$ is equivalent to a tractable SDP:*

$$\begin{aligned} \sup_{\mathbb{P} \in \mathscr{P}(\mu,S)} \ \mathbb{P}(\xi \notin \Xi) = \max \ & \textstyle\sum_{i=1}^k \lambda_i \\ \text{s.t. } & z_i \in \mathbb{R}^n, \ Z_i \in \mathbb{S}^n, \ \lambda_i \in \mathbb{R} \quad \forall i = 1, \ldots, k \\ & a_i^\top z_i \geq b_i \lambda_i \quad \forall i = 1, \ldots, k \\ & \textstyle\sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \\ & \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 \quad \forall i = 1, \ldots, k. \end{aligned} \tag{5}$$

This result is powerful because the SDP reformulation of the worst-case probability problem is exact and can be solved in polynomial time using modern interior point methods [41]. Moreover, the equivalent SDP can conveniently be embedded into higher-level optimization problems such as distributionally robust stochastic programs [9]. Thus, for many practical purposes the generalized Chebyshev bound of Theorem 1 is as useful as a closed-form solution. On the negative side, however, the bound can be rather pessimistic as it is determined largely by discrete worst-case distributions that have only few discretization points. Vandenberghe et al. [37] describe this shortcoming as follows: *'In practical applications, the worst-case distribution will often be unrealistic, and the corresponding bound overly conservative.'*

The objective of this paper is to mitigate the conservatism of the generalized Chebyshev bounds by excluding the pathological discrete distributions from the ambiguity set $\mathscr{P}(\mu, S)$, without sacrificing the exactness and computational tractability of the SDP reformulation of the corresponding worst-case probability problem. This is achieved by requiring the distributions in $\mathscr{P}(\mu, S)$ to be unimodal. Unimodality is a natural property of many distributions and often enjoys strong theoretical and/or empirical justification. A huge variety of popular named distributions are unimodal. Indeed even when sticking to the first few letters of the alphabet, we have that the Bates, Beta (α, $\beta > 1$), Birnbaum-Saunders, Burr, Cauchy, Chi and Chi-squared distributions are all unimodal. A distribution is unimodal if it assigns a higher odds to large deviations from the mode than to smaller ones – a property which is often implicitly assumed to hold. The frequent appearance of unimodal distributions in both theory and practice is hence not surprising. In the remainder we adopt the following standard definitions of unimodality; see e.g. Dharmadhikari and Joag-Dev [10]:

Definition 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 multivariate case, the definition of unimodality is based on the notion of a star-shaped set:

Definition 2 (Star-shaped sets) A set $B \subseteq \mathbb{R}^n$ is said to be star-shaped with center 0 if for every $\xi \in B$ the line segment $[0,\xi]$ is a subset of B.

Definition 3 (Star-unimodality) A distribution $\mathbb{P} \in \mathscr{P}_\infty$ 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 distributions with mode 0 is denoted as $\mathscr{P}_\star$.

Definition 3 assumes without loss of generality that the mode of a star-unimodal distribution is located at the origin, which can always be enforced by applying a suitable coordinate translation. We remark that for multivariate distributions there exist several other notions of unimodality such as linear, convex or log-concave unimodality etc. While not equivalent for $n > 1$, all customary notions of unimodality, including the star unimodality of Definition 3, coincide with Definition 1 in the univariate case; see e.g. [10]. We also remark that the definition of star-unimodality is in line with our intuitive idea of unimodality when $\mathbb{P} \in \mathscr{P}_\infty$ has a continuous density function $f(\xi)$. In this case one can prove that $\mathbb{P}$ is star-unimodal iff $f(t\xi)$ is non-increasing in $t \in (0,\infty)$ for all $\xi \neq 0$, which means that the density function is non-increasing along any ray emanating from the origin [10]. Definition 3 extends this intuitive idea to a broader class of distributions that may have no density functions.

We now generalize the univariate Gauss bound (2) to multivariate distributions, providing a counterpart to the generalization of the univariate Chebyshev bound (1) supplied by Theorem 1. To this end, we introduce the ambiguity set

$$\mathscr{P}_\star(\mu,S) = \mathscr{P}(\mu,S) \cap \mathscr{P}_\star \tag{6}$$

which contains all star-unimodal distributions with mode 0 sharing the same mean μ and second-order moment matrix S. We emphasize that the unimodality requirement eliminates all discrete distributions from $\mathscr{P}_\star(\mu,S)$ with the exception that the distributions in $\mathscr{P}_\star(\mu,S)$ may assign a nonzero probability to the scenario $\xi = 0$. The following theorem asserts that the worst-case probability problem (P) with ambiguity set $\mathscr{P} = \mathscr{P}_\star(\mu,S)$ still admits a tractable SDP reformulation:

Theorem 2 (Generalized Gauss bounds) *If Ξ is a polytope of the form* (3) *with $0 \in \Xi$, the worst-case probability problem* (P) *with ambiguity set $\mathscr{P}_\star(\mu,S)$ is equivalent to a tractable SDP,*

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathscr{P}_\star(\mu,S)} \mathbb{P}(\xi \notin \Xi) = \max\ & \textstyle\sum_{i=1}^k (\lambda_i - t_{i,0}) \\
\text{s.t. } & z_i \in \mathbb{R}^n,\ Z_i \in \mathbb{S}^n,\ \lambda_i \in \mathbb{R},\ t_i \in \mathbb{R}^{\ell+1} \quad \forall i = 1,\dots,k \\
& \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0,\ \ a_i^\top z_i \geq 0,\ \ t_i \geq 0 \quad \forall i = 1,\dots,k \\
& \textstyle\sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} \frac{n+2}{n} S & \frac{n+1}{n}\mu \\ \frac{n+1}{n}\mu^\top & 1 \end{pmatrix} \\
& \left\| \begin{pmatrix} 2\lambda_i b_i \\ t_{i,\ell} b_i - a_i^\top z_i \end{pmatrix} \right\|_2 \leq t_{i,\ell} b_i + a_i^\top z_i \quad \forall i = 1,\dots,k \\
& \left\| \begin{pmatrix} 2t_{i,j+1} \\ t_{i,j} - \lambda_i \end{pmatrix} \right\|_2 \leq t_{i,j} + \lambda_i \quad \forall j \in E, \quad \forall i = 1,\dots,k \\
& \left\| \begin{pmatrix} 2t_{i,j+1} \\ t_{i,j} - t_{i,\ell} \end{pmatrix} \right\|_2 \leq t_{i,j} + t_{i,\ell} \quad \forall j \in O, \quad \forall i = 1,\dots,k,
\end{aligned} \tag{7}$$

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

In the remainder of the paper we prove Theorems 1 and 2 within a unified framework. In sharp contrast to most of the existing literature, we will not attempt to derive tractable reformulations of the semi-infinite linear program dual to (P). Instead, we will establish an SDP reformulation of (P) by exploiting directly a Choquet representation of the underlying ambiguity set. Moreover, we will demonstrate that both Theorems 1 and 2 emerge as special cases of a more general result pertaining to a generalized concept of unimodality.

The rest of the paper develops as follows. In Section 3 we review some basic ideas of Choquet theory and discuss their relevance for deriving tractable reformulations of generalized moment problems. Section 4 introduces the notion of α-unimodality and presents a generalized Gauss-type bound for α-unimodal distributions. In Sections 5 and 6 we demonstrate that the n-variate Chebyshev and Gauss bounds of Theorems 1 and 2 can be recovered from the generalized Gauss-type bound for α-unimodal distributions when setting $\alpha = \infty$ and $\alpha = n$, respectively. Section 7 discusses various extensions, and Section 8 reports numerical results.

3 Choquet representations

A classical result in convex analysis due to Minkowski asserts that any compact convex set $C \subseteq \mathbb{R}^n$ coincides with the convex hull of its extreme points. Choquet theory seeks similar extreme point representations for convex compact subsets of general topological vector spaces. A comprehensive introduction to Choquet theory is given by Phelps [25]. In this paper we are interested in extreme point representations for ambiguity sets $\mathscr{P} \subseteq \mathscr{P}_\infty$ because they will enable us to derive tractable reformulations of worst-case probability problems of the type encountered in Section 2.

Definition 4 (Extreme distributions) A distribution $\mathbb{P} \in \mathscr{P}_\infty$ is said to be an extreme point of an ambiguity set $\mathscr{P} \subseteq \mathscr{P}_\infty$ if it is not representable as a strict convex combination of two distinct distributions in $\mathscr{P}$. The set of all extreme points of $\mathscr{P}$ is denoted as $\operatorname{ex} \mathscr{P}$.

In order to apply the results of Choquet theory, we must endow $\mathscr{P}_\infty$ with a topology so that open, closed and compact subsets of $\mathscr{P}_\infty$ are well-defined. In the following we assume that $\mathscr{P}_\infty$ is equipped with the topology of weak convergence [4], which allows us to construct the Borel σ-algebra on $\mathscr{P}_\infty$ in the usual way as the smallest σ-algebra containing all open subsets of $\mathscr{P}_\infty$. We are now ready to define the concept of a Choquet representation, which is central to our subsequent analysis and generalizes the notion of a convex combination:

Definition 5 (Choquet representation) A weakly closed and convex ambiguity set $\mathscr{P} \subseteq \mathscr{P}_\infty$ is said to admit a Choquet representation if for every $\mathbb{P} \in \mathscr{P}$ there exists a Borel probability measure $\mathbb{m}$ on $\operatorname{ex} \mathscr{P}$ with

$$\mathbb{P}(\cdot) = \int_{\operatorname{ex} \mathscr{P}} e(\cdot)\, \mathbb{m}(\mathrm{d}e)\,. \tag{8}$$

Note that the Choquet representation (8) expresses each $\mathbb{P} \in \mathscr{P}$ as a weighted average (or mixture) of the extreme points of $\mathscr{P}$. As $\mathbb{m}$ constitutes a probability measure, (8) can be viewed as a generalized (infinite) convex combination. We will henceforth refer to $\mathbb{m}$ as a mixture distribution. A classical result of Choquet theory [25, Chapter 3] guarantees that every convex compact subset of $\mathscr{P}_\infty$ has a Choquet representation of the type (8). We emphasize, however, that convex subsets of $\mathscr{P}_\infty$ sometimes admit a Choquet representation even though they fail to be compact. Indeed, in order to prove our main theorem in Section 4, we will exploit the fact that many interesting non-compact ambiguity sets of unimodal distributions admit *explicit* Choquet representations and that their extreme distributions admit a spatial parameterization in the sense of the following definition.

Definition 6 (Spatial parameterization) We say that the set of extreme distributions of a convex closed set $\mathscr{P} \subseteq \mathscr{P}_\infty$ admits a spatial parameterization if $\operatorname{ex} \mathscr{P} = \{e_x \,:\, x \in \mathbb{X}\}$, where $x \in \mathbb{R}^\ell$ parameterizes the extreme distributions of $\mathscr{P}$ and ranges over a closed convex set $\mathbb{X} \subseteq \mathbb{R}^\ell$, while the mapping $x \mapsto e_x(B)$ is a Borel-measurable function for any fixed Borel set $B \subseteq \mathscr{B}(\mathbb{R}^n)$.

If a convex closed set $\mathscr{P} \subseteq \mathscr{P}_\infty$ has a spatial parameterization, then the Choquet representation of any $\mathbb{P} \in \mathscr{P}$ reduces to $\mathbb{P}(\cdot) = \int_{\mathbb{X}} e_x(\cdot)\, \mathbb{m}(\mathrm{d}x)$ for some mixture distribution $\mathbb{m}$ on $\mathbb{X}$.

Example 1 The extreme points of the ambiguity set $\mathscr{P}_\infty$ are given by the Dirac distributions. Thus, the set of extreme distributions $\operatorname{ex} \mathscr{P}_\infty = \{\delta_x \,:\, x \in \mathbb{R}^n\}$ is isomorphic to $\mathbb{R}^n$. As any $\mathbb{P} \in \mathscr{P}_\infty$ is representable as a mixture of Dirac distributions with mixture distribution $\mathbb{m} = \mathbb{P}$, i.e., $\mathbb{P}(\cdot) = \int_{\mathbb{R}^n} \delta_x(\cdot)\, \mathbb{P}(\mathrm{d}x)$, we conclude that $\mathscr{P}_\infty$ admits a (trivial) Choquet representation.

In the remainder we describe a powerful method by Popescu [27] that uses Choquet representations to construct tractable approximations for generalized moment problems of the form

$$\begin{aligned} \sup_{\mathbb{P}\in\mathscr{P}} \quad & \int_{\mathbb{R}^n} f_0(\xi)\,\mathbb{P}(\mathrm{d}\xi) \\ \text{s.t.} \quad & \int_{\mathbb{R}^n} f_i(\xi)\,\mathbb{P}(\mathrm{d}\xi) = \mu_i \quad \forall i = 1,\dots,m, \end{aligned} \tag{GM}$$

where $\mathscr{P} \subseteq \mathscr{P}_\infty$ is a convex closed set of probability distributions. Here we assume that the moment functions $f_i(\xi)$, $i = 1,\dots,m$, are measurable, and $f_0(\xi)$ is integrable with respect to all $\mathbb{P} \in \mathscr{P}$ satisfying the moment constraints in (GM). Note that (GM) encapsulates the worst-case probability problem (P) as a special case if we define $f_0(\xi)$ as the indicator function of $\mathbb{R}^n \setminus \Xi$, while the moment functions $f_i(\xi)$, $i = 1,\dots,\frac{1}{2}(n+2)(n+1)$, are set to the (distinct) linear and quadratic monomials in ξ, and $\mathscr{P}$ is identified with $\mathscr{P}_\infty$ or a structural ambiguity set such as $\mathscr{P}_\star$. The generalized moment problem (GM) constitutes a semi-infinite linear program (LP) with finitely many moment constraints but an infinite-dimensional feasible set.

Problem (GM) can be assigned the following dual problem, which is reminiscent of the dual of a finite-dimensional LP in standard form:

$$\inf_{\lambda\in\mathbb{R}^{m+1}} \left\{ \lambda_0 + \sum_{i=1}^m \lambda_i \mu_i \;:\; \int_{\mathbb{R}^n} \lambda_0 + \sum_{i=1}^m \lambda_i f_i(\xi) - f_0(\xi)\,\mathbb{P}(\mathrm{d}\xi) \geq 0 \quad \forall \mathbb{P} \in \mathscr{P} \right\} \tag{D}$$

An explicit derivation of (D) is provided in [27, Section 3.1]. Note that (D) constitutes a semi-infinite LP with finitely many decision variables but infinitely many constraints parameterized in the distributions $\mathbb{P} \in \mathscr{P}$. Strong duality holds under a mild regularity condition [15, 31]. If the extreme distributions of the ambiguity set $\mathscr{P}$ admit a spatial parameterization in the sense of Definition 6 and $\mathbb{X}$ is semi-algebraic, then the semi-infinite constraint in (D) is equivalent to

$$\int_{\mathbb{R}^n} \lambda_0 + \sum_{i=1}^m \lambda_i f_i(\xi) - f_0(\xi)\, e_x(\mathrm{d}\xi) \geq 0 \quad \forall x \in \mathbb{X}, \tag{9}$$

see [27, Lemma 3.1]. It turns out that the parametric integral in (9) evaluates to a piecewise polynomial in x for many natural choices of the moment functions $f_i(\xi)$, $i = 0,\dots,m$, and the ambiguity set $\mathscr{P}$ with corresponding extreme distributions e_x, $x \in \mathbb{X}$. In this case the semi-infinite constraint (9) requires a piecewise polynomial to be non-negative on a semi-algebraic set and can thus be reformulated as a linear matrix inequality (LMI) (if $\mathbb{X}$ is one-dimensional) or approximated by a hierarchy of increasingly tight LMIs (if $\mathbb{X}$ is multidimensional) by using sum-of-squares techniques [17]. Thus, the dual problem (D) can be reduced systematically to a tractable SDP. Popescu [27] has used this general approach to derive efficiently computable, albeit approximate, Chebyshev and Gauss-type bounds for several structured classes of distributions.

In this paper we adopt a complementary approach and use Choquet representations directly to construct tractable reformulations of (GM). This approach results in *exact* SDP reformulations of many interesting worst-case probability problems—even in the multivariate case. Moreover, the resulting SDPs are smaller than those constructed via (inexact) sum-of-squares methods. Our direct method contrasts sharply with the mainstream approach aiming to derive tractable reformulations of the dual problem (D), which is adopted almost exclusively throughout the extensive body of literature focusing on computational solutions of generalized moment problems. A notable exception is [33], where the primal problem (GM) is approximated by a finite dimensional LP by gridding the sample space $\mathbb{R}^n$. However, this gridding approach is only applicable to low-dimensional problems due to the notorious *curse of dimensionality*.

4 α-Unimodality

In this section we will argue that the Chebyshev inequality of Theorem 1 and the Gauss inequality of Theorem 2 can be analyzed in a unified framework using a generalized notion of unimodality. Specifically, we will demonstrate that the worst-case probability problems underlying the Chebyshev and Gauss inequalities can be embedded into a parametric family of optimization problems, each of which admits an exact and tractable SDP reformulation.

Definition 7 **(α-Unimodality [10])** For any fixed $\alpha > 0$, a distribution $\mathbb{P} \in \mathscr{P}_\infty$ is called α-unimodal with mode 0 if $t^\alpha \mathbb{P}(B/t)$ is non-decreasing in $t \in (0,\infty)$ for every Borel set $B \in \mathscr{B}(\mathbb{R}^n)$. The set of all α-unimodal distributions with mode 0 is denoted as $\mathscr{P}_\alpha$.

Using standard measure-theoretic arguments, one can show that $\mathbb{P} \in \mathscr{P}_\infty$ is α-unimodal iff

$$t^\alpha \int_{\mathbb{R}^n} g(\xi t)\, \mathbb{P}(\mathrm{d}\xi)$$

is non-decreasing in $t \in (0,\infty)$ for every nonnegative, bounded and continuous function $g(\xi)$. From this equivalent characterization of α-unimodality it is evident that $\mathscr{P}_\alpha$ is closed under weak convergence. To develop an intuitive understanding of Definition 7, it is instructive to study the special case of continuous distributions. Recall from Section 2 that a continuous star-unimodal distribution has a density function that is non-increasing along rays emanating from the origin. The density function of a continuous α-unimodal distribution may in fact increase along rays, but the rate of increase is controlled by α. We have that a distribution $\mathbb{P} \in \mathscr{P}_\infty$ with a continuous density function $f(\xi)$ is α-unimodal about 0 iff $t^{n-\alpha} f(t\xi)$ is non-increasing in $t \in (0,\infty)$ for every fixed $\xi \neq 0$. This implies that if an α-unimodal distribution on $\mathbb{R}^n$ has a continuous density function $f(\xi)$, then $f(\xi)$ does not grow faster than $\|\xi\|^{\alpha-n}$. In particular, for $\alpha = n$ the density is non-increasing along rays emanating from the origin and is thus star-unimodal in the sense of Definition 3.

Definition 8 **(Radial α-unimodal distributions)** For any $\alpha > 0$ and $x \in \mathbb{R}^n$ we denote by $\delta^\alpha_{[0,x]}$ the radial distribution supported on the line segment $[0,x] \subset \mathbb{R}^n$ with the property that

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

One can confirm that $\delta^\alpha_{[0,x]} \in \mathscr{P}_\alpha$ by direct application of Definition 7. Indeed, we have

$$t^\alpha \delta^\alpha_{[0,x]}(B/t) = t^\alpha \int_{\mathbb{R}^n} \mathbf{1}_B(\xi t)\, \delta^\alpha_{[0,x]}(\mathrm{d}\xi) = t^\alpha \int_0^1 \mathbf{1}_B(xtu)\, \alpha u^{\alpha-1} \mathrm{d}u = \int_0^t \mathbf{1}_B(xu)\, \alpha u^{\alpha-1} \mathrm{d}u\,,$$

and the last expression is manifestly non-decreasing in $t \in (0,\infty)$. Alternatively, one can express the radial distributions $\delta^\alpha_{[0,x]}$ as weak limits of continuous distributions that are readily identified as members of $\mathscr{P}_\alpha$. As $\mathscr{P}_\alpha$ is closed under weak convergence, one can again conclude that $\delta^\alpha_{[0,x]} \in \mathscr{P}_\alpha$. For example, denote by $\mathbb{P}_\theta$ the uniform distribution on the intersection of the closed ball $\mathbb{B}_{\|x\|}(0) = \{\xi \in \mathbb{R}^n : \|\xi\| \leq \|x\|\}$ and the second-order cone

$$\mathbb{K}(x,\theta) = \left\{ \xi \in \mathbb{R}^n : \frac{x^\top \xi}{\|x\|} \leq \tan(\theta) \left\| \left(\mathbb{I} - \frac{xx^\top}{\|x\|^2} \right) \xi \right\| \right\}$$

with principal axis x and opening angle $\theta \in (0,\pi/2)$. As both $\mathbb{B}_{\|x\|}(0)$ and $\mathbb{K}(x,\theta)$ are star-shaped, $\mathbb{P}_\theta$ is star-unimodal. Using standard arguments, one can show that $\mathbb{P}_\theta$ converges weakly to $\delta^n_{[0,x]}$ as θ tends to 0, which confirms the (maybe surprising) result that $\delta^n_{[0,x]}$ is star-unimodal.

The importance of the radial distributions $\delta^\alpha_{[0,x]}$ is highlighted in the following theorem:

Theorem 3 *For every $\mathbb{P} \in \mathscr{P}_\alpha$ there exists a unique mixture distribution $\mathbb{m} \in \mathscr{P}_\infty$ with*

$$\mathbb{P}(\cdot) = \int_{\mathbb{R}^n} \delta^\alpha_{[0,x]}(\cdot)\, \mathbb{m}(\mathrm{d}x)\,. \tag{10}$$

Proof See [10, Theorem 3.5]. □

Theorem 3 asserts that every α-unimodal distribution admits a unique Choquet representation, i.e. it can be expressed as a mixture of radial distributions $\delta^\alpha_{[0,x]}$, $x \in \mathbb{R}^n$. Thus, $\mathscr{P}_\alpha$ is generated by the simple family of radial α-unimodal distributions.

Corollary 1 *The radial distributions $\delta^\alpha_{[0,x]}$, $x \in \mathbb{R}^n$, are extremal in $\mathscr{P}_\alpha$.*

Proof If $\delta^{\alpha}_{[0,x]}$ is not extremal in $\mathscr{P}_\alpha$ for some $x \in \mathbb{R}^n$, there exist $\mathbb{P}^1, \mathbb{P}^2 \in \mathscr{P}_\alpha$, $\mathbb{P}^1 \neq \mathbb{P}^2$, 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 $\lambda \mathbb{m}^1 + (1-\lambda)\mathbb{m}^2$, where $\mathbb{m}^1$ and $\mathbb{m}^2$ are the unique mixture distributions corresponding to $\mathbb{P}^1$ and $\mathbb{P}^2$, respectively. However, the unique mixture distribution of $\delta^{\alpha}_{[0,x]}$ is the Dirac distribution concentrating unit mass at x, and this distribution is not representable as a strict convex combination of two distinct mixture distributions. Thus, $\delta^{\alpha}_{[0,x]}$ must be extremal in $\mathscr{P}_\alpha$. □

It is easy to verify that the radial distributions $\delta^{\alpha}_{[0,x]}$ converge weakly to the Dirac distribution δ_x as α tends to ∞. As every distribution $\mathbb{P} \in \mathscr{P}_\infty$ is trivially representable as a mixture of Dirac distributions with $\mathbb{m} = \mathbb{P}$ (see Example 1), this allows us to conclude that the weak closure of $\cup_{\alpha>0}\mathscr{P}_\alpha$ coincides with $\mathscr{P}_\infty$. The ambiguity sets $\mathscr{P}_\alpha$ enjoy the nesting property $\mathscr{P}_\alpha \subseteq \mathscr{P}_\beta$ whenever $0 < \alpha \leq \beta \leq \infty$. This nesting enables us to define the α-unimodality of a generic ambiguity set $\mathscr{P}$ as the smallest number α with $\mathscr{P} \subseteq \mathscr{P}_\alpha$. A listing of α-unimodality values for commonly used ambiguity sets is provided in Appendix A.

In the remainder of this section we will consider the ambiguity set

$$\mathscr{P}_\alpha(\mu, S) = \mathscr{P}(\mu, S) \cap \mathscr{P}_\alpha\,,$$

which contains all α-unimodal distributions sharing a known mean value $\mu \in \mathbb{R}^n$ and second-order moment matrix $S \in \mathbb{S}^n$. We will address first the question of whether this ambiguity set is non-empty, and we will demonstrate that it can be answered efficiently by checking an LMI. This result relies critically on Theorem 3 and the following lemma.

Lemma 1 *For any $\alpha > 0$ and $x \in \mathbb{R}^n$, the mean value and second-order 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 It is easy to verify that the formulas for the mean value and the second-order moment matrix are correct if x represents a point on a coordinate axis in $\mathbb{R}^n$. The general formulas then follow from the relations $\int_{\mathbb{R}^n} \xi\, \delta^{\alpha}_{[0,Rx]}(\mathrm{d}\xi) = R\int_{\mathbb{R}^n} \xi\, \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)$ and $\int_{\mathbb{R}^n} \xi\xi^\top \delta^{\alpha}_{[0,Rx]}(\mathrm{d}\xi) = R\int_{\mathbb{R}^n} \xi\xi^\top \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)R^\top$ for arbitrary orthogonal coordinate transformations $R \in \mathbb{R}^{n\times n}$. □

Proposition 1 *The set $\mathscr{P}_\alpha(\mu, S)$ is non-empty iff*

$$\begin{pmatrix} \frac{\alpha+2}{\alpha}S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} \succeq 0\,. \tag{11}$$

Proof By Theorem 3, any $\mathbb{P} \in \mathscr{P}_\alpha(\mu, S)$ has a Choquet representation of the form (10) for some mixture distribution $\mathbb{m} \in \mathscr{P}_\infty$. Thus, we have

$$\begin{aligned}
\begin{pmatrix} \frac{\alpha+2}{\alpha}S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} &= \int_{\mathbb{R}^n} \begin{pmatrix} \frac{\alpha+2}{\alpha}\xi\xi^\top & \frac{\alpha+1}{\alpha}\xi \\ \frac{\alpha+1}{\alpha}\xi^\top & 1 \end{pmatrix} \mathbb{P}(\mathrm{d}\xi) \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \begin{pmatrix} \frac{\alpha+2}{\alpha}\xi\xi^\top & \frac{\alpha+1}{\alpha}\xi \\ \frac{\alpha+1}{\alpha}\xi^\top & 1 \end{pmatrix} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)\, \mathbb{m}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^n} \begin{pmatrix} xx^\top & x \\ x^\top & 1 \end{pmatrix} \mathbb{m}(\mathrm{d}x) \;\succeq 0\,,
\end{aligned}$$

where the first equality holds because of the moment conditions in the definition of $\mathscr{P}_\alpha(\mu, S)$, the third equality follows from Lemma 1, and the matrix inequality holds because the matrix of first and second-order moments of the mixture distribution $\mathbb{m}$ must be positive semidefinite.

Conversely, if (11) holds, there exists a normal distribution $\mathbb{m}$ with mean value $\frac{\alpha+1}{\alpha}\mu$ and second-order moment matrix $\frac{\alpha+2}{\alpha}S$. Then, $\mathbb{P}(\cdot) = \int_{\mathbb{R}^n} \delta^{\alpha}_{[0,x]}(\cdot)\, \mathbb{m}(\mathrm{d}x)$ is a mixture of α-unimodal distributions and is therefore itself α-unimodal. By construction, it further satisfies

$$\int_{\mathbb{R}^n} \begin{pmatrix} \xi\xi^\top & \xi \\ \xi^\top & 1 \end{pmatrix} \mathbb{P}(\mathrm{d}\xi) = \int_{\mathbb{R}^n} \begin{pmatrix} \frac{\alpha}{\alpha+2}xx^\top & \frac{\alpha}{\alpha+1}x \\ \frac{\alpha}{\alpha+1}x^\top & 1 \end{pmatrix} \mathbb{m}(\mathrm{d}x) = \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix}.$$

Thus, $\mathscr{P}_\alpha(\mu, S)$ is non-empty. □

Defining the covariance matrix in the usual way as $\Sigma = S - \mu\mu^\top$, a standard Schur complement argument shows that (11) is equivalent to

$$\Sigma \succeq \frac{1}{\alpha(\alpha+2)}\mu\mu^\top,$$

which confirms our intuition that condition (11) becomes less restrictive as α grows. Moreover, we also observe that Proposition 1 remains valid for $\alpha \to \infty$, in which case (11) reduces to the standard requirement that the covariance matrix must be positive semidefinite. Finally, we remark that (11) is naturally interpreted as the matrix of first and second-order moments of the mixture distribution $\mathbb{m}$ corresponding to any $\mathbb{P} \in \mathscr{P}_\alpha(\mu, S)$.

4.1 An α-unimodal bound

We consider now an open polytope Ξ of the form (3), and will assume throughout this section that $0 \in \Xi$. Since Ξ is convex, this requirement ensures that Ξ is star-shaped about 0, and it implies that $b_i > 0$ for all $i = 1, \ldots, k$. We will now investigate the worst-case probability of the event $\xi \notin \Xi$ over all distributions from within $\mathscr{P}_\alpha(\mu, S)$,

$$B_\alpha(\mu, S) = \sup_{\mathbb{P} \in \mathscr{P}_\alpha(\mu, S)} \mathbb{P}(\xi \notin \Xi), \tag{P_α}$$

and we will prove that $B_\alpha(\mu, S)$ can be computed efficiently by solving a tractable SDP. This proof will be constructive in the sense that it will indicate how an extremal distribution for problem (P_α) can be derived explicitly from an optimizer of its SDP reformulation.

Theorem 4 **(α-Unimodal bound)** *If $0 \in \Xi$ and $\alpha > 0$, then problem (P_α) is equivalent to*

$$\begin{aligned} B_\alpha(\mu, S) = \max\ & \textstyle\sum_{i=1}^k (\lambda_i - \tau_i) \\ \text{s.t. } & z_i \in \mathbb{R}^n,\ Z_i \in \mathbb{S}^n,\ \lambda_i \in \mathbb{R},\ \tau_i \in \mathbb{R} \quad \forall i = 1, \ldots, k \\ & \textstyle\sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} \\ & \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 \quad \forall i = 1, \ldots, k \\ & a^\top z_i \geq 0,\ \tau_i \geq 0,\ \tau_i (a_i^\top z_i)^\alpha \geq \lambda_i^{\alpha+1} b_i^\alpha \quad \forall i = 1, \ldots, k. \end{aligned} \tag{SDP_α}$$

Remark 1 Note that problem (SDP_α) fails to be an SDP in standard form due to the nonlinearity of its last constraint. However, in Lemmas 2 and 3 below we will show that this constraint is in fact second-order cone representable under the mild additional assumption that α is rational and not smaller than 1. In this case, (SDP_α) is thus equivalent to a tractable SDP.

Proof (Proof of Theorem 4) We first show that any feasible solution for (SDP_α) can be used to construct a feasible solution $\mathbb{P} \in \mathscr{P}_\alpha(\mu, S)$ achieving the same objective value in (P_α). Let $\{z_i, Z_i, \lambda_i, \tau_i\}_{i=1}^k$ be feasible in (SDP_α) and set $x_i = z_i/\lambda_i$ if $\lambda_i > 0$; $= 0$ otherwise. Moreover, assume without loss of generality that $Z_i = z_i z_i^\top/\lambda_i$ if $\lambda_i > 0$; $= 0$ if $\lambda_i = 0$. This choice preserves feasibility of $\{z_i, Z_i, \lambda_i, \tau_i\}_{i=1}^k$ and has no effect on its objective value in (SDP_α). Next, define

$$\begin{pmatrix} Z_0 & z_0 \\ z_0^\top & \lambda_0 \end{pmatrix} = \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} - \sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0, \tag{12}$$

which is positive semidefinite due to the first constraint in (SDP_α). Assume now that $\lambda_0 > 0$ and define $\mu_0 = \frac{\alpha}{\alpha+1} z_0/\lambda_0$ and $S_0 = \frac{\alpha}{\alpha+2} Z_0/\lambda_0$. Proposition 1 then guarantees the existence of an α-unimodal distribution $\mathbb{P}_0 \in \mathscr{P}_\alpha(\mu_0, S_0)$, which allows us to construct $\mathbb{P}(\cdot) = \lambda_0 \mathbb{P}_0(\cdot) + \sum_{i=1}^k \lambda_i \delta^\alpha_{[0,x_i]}(\cdot)$. Note that $\mathbb{P}$ is α-unimodal as it is a convex combination of α-unimodal distributions. The first and second moments of $\mathbb{P}$ are given by

$$\begin{aligned} \int_{\mathbb{R}^n} \begin{pmatrix} \xi \\ 1 \end{pmatrix} \begin{pmatrix} \xi \\ 1 \end{pmatrix}^\top \mathbb{P}(\mathrm{d}\xi) &= \lambda_0 \begin{pmatrix} S_0 & \mu_0 \\ \mu_0^\top & 1 \end{pmatrix} + \sum_{i=1}^k \lambda_i \begin{pmatrix} \frac{\alpha}{\alpha+2} x_i x_i^\top & \frac{\alpha}{\alpha+1} x_i \\ \frac{\alpha}{\alpha+1} x_i^\top & 1 \end{pmatrix} \\ &= \sum_{i=0}^k \begin{pmatrix} \frac{\alpha}{\alpha+2} Z_i & \frac{\alpha}{\alpha+1} z_i \\ \frac{\alpha}{\alpha+1} z_i^\top & \lambda_i \end{pmatrix} = \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix}, \end{aligned}$$

where the first equality follows from Lemma 1, while the last equality holds due to (12). We conclude that $\mathbb{P} \in \mathscr{P}_\alpha(\mu, S)$.

For $\lambda_i > 0$, we can estimate the probability of the event $\xi \notin \Xi$ under the distribution $\delta^\alpha_{[0,x_i]}$ as

$$\delta^\alpha_{[0,x_i]}(\xi \notin \Xi) \geq \delta^\alpha_{[0,x_i]}(a_i^\top \xi \geq b_i) \geq \delta^\alpha_{[0,x_i]}([x_i \sqrt[\alpha]{\tau_i/\lambda_i}, x_i]) \geq 1 - \frac{\tau_i}{\lambda_i}\,,$$

where the first inequality is due to the definition of Ξ, while the second inequality exploits the definition of $\delta^\alpha_{[0,x_i]}$ and the fact that $a_i^\top x_i \sqrt[\alpha]{\tau_i/\lambda_i} \geq b_i$, which follows from the constraints in the last line of (SDP_α). For $\lambda_i = 0$, we have the trivial estimate $\delta^\alpha_{[0,x_i]}(\xi \notin \Xi) \geq 0$. Thus, we find

$$\mathbb{P}(\xi \notin \Xi) = \lambda_0 \mathbb{P}_0(\xi \notin \Xi) + \sum_{i=1}^k \lambda_i \, \delta^\alpha_{[0,x_i]}(\xi \notin \Xi) \geq \sum_{i=1}^k (\lambda_i - \tau_i)\,.$$

In summary, $\mathbb{P}$ is feasible in the worst-case probability problem (P_α) with an objective value that is at least as large as that of $\{z_i, Z_i, \lambda_i, \tau_i\}_{i=1}^k$ in (SDP_α). If $\lambda_0 = 0$, we can set $\hat{z}_i = (1-\varepsilon) z_i$, $\hat{Z}_i = (1-\varepsilon) Z_i$, $\hat{\lambda}_i = (1-\varepsilon)\lambda_i$ and $\hat{\tau}_i = (1-\varepsilon)\tau_i$, $i = 1, \ldots, k$, for some $\varepsilon \in (0,1)$. By repeating the above arguments for $\lambda_0 > 0$, we can use $\{\hat{z}_i, \hat{Z}_i, \hat{\lambda}_i, \hat{\tau}_i\}_{i=1}^k$ to construct a feasible solution of (P_α) with an objective value of at least $(1-\varepsilon)\sum_{i=1}^k(\lambda_i - \tau_i)$. As ε tends to zero, we obtain a sequence of distributions feasible in (P_α) whose objective values asymptotically approach that of $\{z_i, Z_i, \lambda_i, \tau_i\}_{i=1}^k$ in (SDP_α). We conclude that (SDP_α) provides a lower bound on (P_α).

Next, we prove that (P_α) also provides a lower bound on (SDP_α) and that any feasible solution for (P_α) gives rise to a feasible solution for (SDP_α) with the same objective value. To this end, we set

$$\Xi_i = \left\{ \xi \in \mathbb{R}^n : a_i^\top \xi \geq b_i,\ \frac{a_i^\top \xi}{b_i} \geq \frac{a_j^\top \xi}{b_j}\ \forall j < i,\ \frac{a_i^\top \xi}{b_i} > \frac{a_j^\top \xi}{b_j}\ \forall j > i \right\} \quad \forall i = 1, \ldots, k.$$

Note that the Ξ_i are pairwise disjoint and that their union coincides with the complement of Ξ. Consider now a distribution $\mathbb{P}(\cdot) = \int_{\mathbb{R}^n} \delta^\alpha_{[0,x]}(\cdot)\, \mathbb{m}(\mathrm{d}x)$ that is feasible in (P_α), where $\mathbb{m} \in \mathscr{P}_\infty$ is the underlying mixture distribution whose existence is guaranteed by Theorem 3. Next, define

$$\begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} = \int_{\Xi_i} \begin{pmatrix} x \\ 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix}^\top \mathbb{m}(\mathrm{d}x) \quad \text{and} \quad \tau_i = \int_{\Xi_i} \left[\int_{\Xi} \delta^\alpha_{[0,x]}(\mathrm{d}\xi) \right] \mathbb{m}(\mathrm{d}x) \tag{13}$$

for $i = 1, \ldots, k$. By construction, we have

$$\begin{aligned}
\sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} &= \sum_{i=1}^k \int_{\Xi_i} \begin{pmatrix} x \\ 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix}^\top \mathbb{m}(\mathrm{d}x) \preceq \int_{\mathbb{R}^n} \begin{pmatrix} x \\ 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix}^\top \mathbb{m}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \begin{pmatrix} \frac{\alpha+2}{\alpha} \xi\xi^\top & \frac{\alpha+1}{\alpha}\xi \\ \frac{\alpha+1}{\alpha}\xi^\top & 1 \end{pmatrix} \delta_{[0,x]}(\mathrm{d}\xi) \right] \mathbb{m}(\mathrm{d}x) = \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix},
\end{aligned}$$

where the inequality holds because the Ξ_i are non-overlapping and cover $\mathbb{R}^n$ only partly, while the equalities in the second line follow from Lemma 1 and the fact that $\mathbb{P}$ is an element of $\mathscr{P}_\alpha(\mu, S)$, respectively. We also find

$$\begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0, \quad a_i^\top z_i \geq 0, \quad \tau_i \geq 0 \quad \forall i = 1, \ldots, k,$$

where the second inequality exploits the fact that $a_i^\top x \geq b_i > 0$ for all $x \in \Xi_i$. Moreover, by the construction of Ξ_i, we have

$$\int_{\Xi} \delta^\alpha_{[0,x]}(\mathrm{d}\xi) = \begin{cases} \left(\frac{b_i}{a_i^\top x}\right)^\alpha & \text{if } x \in \Xi_i,\ i = 1, \ldots, k, \\ 1 & \text{if } x \in \Xi, \end{cases}$$

which implies

$$\begin{aligned}
& \tau_i \ = \ \int_{\Xi_i} \left(\frac{b_i}{a_i^\top x}\right)^\alpha \mathbb{m}(\mathrm{d}x) \ = \ b_i^\alpha \int_{\Xi_i} \left(a_i^\top x\right)^{-\alpha} \mathbb{m}(\mathrm{d}x) \\
\implies \ & \left(\int_{\Xi_i} a_i^\top x \, \mathbb{m}(\mathrm{d}x) \right)^\alpha \tau_i \ \geq \ \int_{\Xi_i} \mathbb{m}(\mathrm{d}x) \left(b_i \int_{\Xi_i} \mathbb{m}(\mathrm{d}x) \right)^\alpha \\
\implies \ & (a_i^\top z_i)^\alpha \, \tau_i \ \geq \ \lambda_i \, (b_i \lambda_i)^\alpha\,.
\end{aligned}$$

Here, the second line holds due to Jensen's inequality, which is applicable as the mapping $t \mapsto t^{-\alpha}$ is convex on the positive real axis for $\alpha > 0$. Thus, the $\{z_i, Z_i, \lambda_i, \tau_i\}_{i=1}^k$ constructed in (13) are feasible in (SDP_α). Their objective value in (SDP_α) can be represented as

$$\begin{aligned}
\sum_{i=1}^k (\lambda_i - \tau_i) &= \sum_{i=1}^k \left(\int_{\Xi_i} \mathbb{m}(\mathrm{d}x) - \int_{\Xi_i} \left[\int_{\Xi} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)) \right] \mathbb{m}(\mathrm{d}x) \right) \\
&= \sum_{i=1}^k \int_{\Xi_i} \left[\int_{\mathbb{R}^n \setminus \Xi} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)) \right] \mathbb{m}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n \setminus \Xi} \delta^{\alpha}_{[0,x]}(\mathrm{d}\xi)) \right] \mathbb{m}(\mathrm{d}x) = \mathbb{P}(\xi \notin \Xi)
\end{aligned}$$

and thus coincides with the objective value of $\mathbb{P}$ in (P_α). □

It remains to be shown that (SDP_α) is equivalent to a tractable SDP whenever α is a rational number not smaller than 1.

Lemma 2 (Second-order cone representation for rational α) *Suppose that $b > 0$ and $\alpha = p/q$ for $p, q \in \mathbb{N}$ with $p \geq q$. If the linear constraints $\lambda \geq 0$, $\tau \geq 0$ and $a^\top z \geq 0$ hold, then the nonlinear constraint $(a^\top z)^\alpha \tau \geq \lambda^{\alpha+1} b^\alpha$ has an equivalent representation in terms of $\mathscr{O}(p)$ second-order constraints involving z, τ, λ and $\mathscr{O}(p)$ auxiliary variables.*

Proof We have

$$\begin{aligned}
(a^\top z)^{\frac{p}{q}} \tau \geq \lambda^{\frac{p+q}{q}} b^{\frac{p}{q}} &\iff (a^\top z)^p \tau^q \geq \lambda^{p+q} b^p \\
&\iff \exists t \geq 0 \;:\; (a^\top z)^p t^p \geq \lambda^{2p} b^p,\; \tau^q \lambda^{p-q} \geq t^p \\
&\iff \exists t \geq 0 \;:\; (a^\top z) t \geq \lambda^2 b,\; \tau^q \lambda^{p-q} t^{2^\ell - p} \geq t^{2^\ell},
\end{aligned}$$

where $\ell = \lceil \log_2 p \rceil$. Both constraints in the last line of the above expression are second-order cone representable. Indeed, the first (hyperbolic) constraint is equivalent to

$$\left\| \begin{pmatrix} 2\lambda b \\ tb - a^\top z \end{pmatrix} \right\|_2 \leq tb + a^\top z,$$

while the second constraint can be reformulated as $(\tau^q \lambda^{p-q} t^{2^\ell - p})^{1/2^\ell} \geq t$ and thus requires the geometric mean of 2^ℓ nonnegative variables to be non-inferior to t. By [23, § 6.2.3.5], this requirement can be re-expressed in terms of $\mathscr{O}(2^\ell)$ second-order cone constraints involving $\mathscr{O}(2^\ell)$ auxiliary variables. □

Lemma 2 establishes that (SDP_α) has a tractable reformulation for any rational $\alpha \geq 1$ by exploiting a well-known second-order cone representation for geometric means. When α is integer, one can construct a more efficient reformulation involving fewer second-order cone constraints and auxiliary variables. The following lemma derives this reformulation explicitly.

Lemma 3 (Second-order cone representation for integer α) *Suppose that $b > 0$ and $\alpha \in \mathbb{N}$. If the linear constraints $\lambda \geq 0$, $\tau \geq 0$ and $a^\top z \geq 0$ hold, then the nonlinear constraint $(a^\top z)^\alpha \tau \geq \lambda^{\alpha+1} b^\alpha$ is equivalent to*

$$\exists t \in \mathbb{R}^{\ell+1} : \begin{cases} t \geq 0,\;\; t_0 = \tau,\;\; \left\| \begin{pmatrix} 2\lambda b \\ t_\ell b - a^\top z \end{pmatrix} \right\|_2 \leq t_\ell b + a^\top z, \\ \left\| \begin{pmatrix} 2t_{j+1} \\ t_j - \lambda \end{pmatrix} \right\|_2 \leq t_j + \lambda \;\; \forall j \in E,\;\; \left\| \begin{pmatrix} 2t_{j+1} \\ t_j - t_\ell \end{pmatrix} \right\|_2 \leq t_j + t_\ell \;\; \forall j \in O, \end{cases}$$

where $\ell = \lceil \log_2(\alpha) \rceil$, $E = \{j \in \{0, \ldots, \ell - 1\} : \lceil \alpha / 2^j \rceil$ is even$\}$ and $O = \{j \in \{0, \ldots, \ell - 1\} : \lceil \alpha / 2^j \rceil$ is odd$\}$.

Proof We have

$$(a^\top z)^\alpha \tau \ge \lambda^{\alpha+1} b^\alpha \iff \exists s \ge 0 \,:\, (a^\top z)^\alpha s^\alpha \ge (\lambda^2 b)^\alpha,\;\; \tau \lambda^{\alpha-1} \ge s^\alpha,$$
$$\iff \exists s \ge 0 \,:\, a^\top z s \ge \lambda^2 b,\;\; \tau\lambda^{\alpha-1} \ge s^\alpha. \tag{14}$$

The first inequality in (14) is a hyperbolic constraint equivalent to

$$\left\| \begin{pmatrix} 2\lambda b \\ sb - a^\top z \end{pmatrix} \right\|_2 \le sb + a^\top z$$

and implies that $\lambda = 0$ whenever $s = 0$. Next, we show that the second inequality in (14) can be decomposed into ℓ hyperbolic constraints. To this end, we observe that

$$t_j \lambda^{\lceil \alpha/2^j \rceil - 1} \ge s^{\lceil \alpha/2^j \rceil} \iff \exists t_{j+1} \ge 0 \,:\, \lambda^{\lceil \alpha/2^j \rceil - 2} t_{j+1}^2 \ge s^{\lceil \alpha/2^j \rceil},\; t_j \lambda \ge t_{j+1}^2$$
$$\iff \exists t_{j+1} \ge 0 \,:\, \lambda^{\lceil \alpha/2^{j+1} \rceil - 1} t_{j+1} \ge s^{\lceil \alpha/2^{j+1} \rceil},\; t_j \lambda \ge t_{j+1}^2$$

for all $j \in E$ and $t_j \ge 0$, while

$$t_j \lambda^{\lceil \alpha/2^j \rceil - 1} \ge s^{\lceil \alpha/2^j \rceil} \iff \exists t_{j+1} \ge 0 \,:\, \lambda^{\lceil \alpha/2^j \rceil - 1} t_{j+1}^2 \ge s^{\lceil \alpha/2^j \rceil + 1},\; t_j s \ge t_{j+1}^2$$
$$\iff \exists t_{j+1} \ge 0 \,:\, \lambda^{\lceil \alpha/2^{j+1} \rceil - 1} t_{j+1} \ge s^{\lceil \alpha/2^{j+1} \rceil},\; t_j s \ge t_{j+1}^2$$

for all $j \in O$ and $t_j \ge 0$. Applying the above equivalences iteratively for $j = 0, \ldots, \ell - 1$, we find

$$\tau \lambda^{\alpha-1} \ge s^\alpha \iff \exists t \in \mathbb{R}^{\ell+1} \,:\, t \ge 0,\; t_0 = \tau,\; t_j \lambda \ge t_{j+1}^2 \;\; \forall j \in E,\; t_j s \ge t_{j+1}^2 \;\; \forall j \in O$$
$$\iff \exists t \in \mathbb{R}^{\ell+1} \,:\, t \ge 0,\; t_0 = \tau,$$
$$\left\| \begin{pmatrix} 2t_{j+1} \\ t_j - \lambda \end{pmatrix} \right\|_2 \le t_j + \lambda \;\; \forall j \in E,\;\; \left\| \begin{pmatrix} 2t_{j+1} \\ t_j - s \end{pmatrix} \right\|_2 \le t_j + s \;\; \forall j \in O.$$

The claim now follows as we can set $s = t_\ell$ without loss of generality. □

In order for the worst-case probability problem (P_α) to be of practical value we need to establish that its optimal value depends continuously on the distributional parameters μ and S [14].

Proposition 2 (Well-posedness of problem (P_α)) *The optimal value function $B_\alpha(\mu, S)$ of problem (P_α) is concave and continuous on the set of all $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$ with*

$$\begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} \succ 0\,.$$

Proof Concavity of $B_\alpha(\mu, S)$ is a direct consequence of [28, Proposition 2.22] and Theorem 4. By [28, Theorem 2.35], $B_\alpha(\mu, S)$ is thus continuous on the interior of its domain. The claim now follows from the characterization of the domain of $B_\alpha(\mu, S)$ in Proposition 1. □

4.2 Worst-case distributions

In addition to identifying a tractable reformulation of problem (P_α), it is also of interest to identify a worst-case distribution $\mathbb{P}^\star$ with $\mathbb{P}^\star(\xi \notin \Xi) = B_\alpha(\mu, S)$ if it exist. The fundamental theorem of linear programming states that if the supremum in problem (P_α) is attained, then it is attained in particular by a mixture distribution $\mathbb{P}^\star$ consisting of at most $(n+2)(n+1)/2$ extreme distributions of $\mathscr{P}_\alpha$, where $(n+2)(n+1)/2$ is the number of distinct moment constraints. The proof of the fundamental theorem is well known [17, 27, 29], and its finite dimensional LP counterpart bears the same name.

The proof of Theorem 4 suggests an explicit construction of a worst-case distribution as the convex combination of at most $2n + k$ radial distributions. With the help of any maximizer $\{\lambda_i^\star, z_i^\star\}_{i=1}^k$ of problem (SDP_α) satisfying

$0 < \lambda_0^\star = 1 - \sum_{i=1}^k \lambda_i^\star$, a worst-case distribution in the form $\mathbb{P}^\star = \lambda_0^\star \mathbb{P}_0 + \sum_{i=1}^k \lambda_i^\star \delta^\alpha_{[0,x_i^\star]}$ can be constructed where $x_i^\star = z_i^\star/\lambda_i^\star$ if $\lambda_i^\star > 0$; $= 0$ otherwise, and where $\mathbb{P}_0 \in \mathscr{P}_\alpha(\frac{\alpha}{\alpha+1}\mu_0, \frac{\alpha}{\alpha+2}S_0)$ with

$$\lambda_0^\star \begin{pmatrix} S_0 & \mu_0 \\ \mu_0^\top & 1 \end{pmatrix} = \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} - \sum_{i=1}^k \lambda_i^\star \begin{pmatrix} x_i^\star x_i^{\star\top} & x_i^\star \\ x_i^{\star\top} & 1 \end{pmatrix} \succeq 0.$$

Such a distribution $\mathbb{P}_0$ can always be found as the convex combination of at most $2n$ radial distributions. Indeed, the positive semidefinite covariance matrix $S_0 - \mu_0\mu_0^\top$ can be factored as $S_0 - \mu_0\mu_0^\top = \sum_{i=1}^r w_i w_i^\top$ where $r \le n$ is the rank of the covariance matrix. It can now be readily verified that the distribution

$$\mathbb{P}_0 = \sum_{i=1}^r \frac{1}{2r} \delta^\alpha_{[0,\mu_0+\sqrt{r}w_i]} + \sum_{i=1}^r \frac{1}{2r} \delta^\alpha_{[0,\mu_0-\sqrt{r}w_i]},$$

satisfies $\mathbb{P}_0 \in \mathscr{P}_\alpha(\frac{\alpha}{\alpha+1}\mu_0, \frac{\alpha}{\alpha+2}S_0)$. When compared to the fundamental theorem, it is clear that the worst-case distribution $\mathbb{P}^\star$ constructed from a solution of problem (SDP_α) is not necessarily maximally sparse.

We remark that the reformulation offered in Theorem 4 is exact even though no worst-case distribution $\mathbb{P}$ may exist. The nonexistence of a worst-case distribution in problem (P_α) occurs only when $\lambda_0^\star = 0$ in its reformulation (SDP_α). In that case, any maximizer of (SDP_α) can be used to construct a sequence of distributions $\{\mathbb{P}_t\}$, $\mathbb{P}_t \in \mathscr{P}_\alpha(\mu, S)$ with the property

$$\lim_{t\to\infty} \mathbb{P}_t(\xi \notin \Xi) = B_\alpha(\mu, S).$$

5 Generalized Chebyshev bounds

Vandenberghe et al. [37] derive the generalized Chebyshev bound of Theorem 1 by interpreting the worst-case probability problem

$$B(\mu, S) = \sup_{\mathbb{P}\in\mathscr{P}(\mu,S)} \mathbb{P}(\xi \notin \Xi) \tag{15}$$

as a generalized moment problem of the type (GM). They then reformulate its dual (D) as an SDP by using the celebrated $\mathscr{S}$-lemma [26] to convert the semi-infinite constraint in (D) to an LMI. The new techniques of Section 4 now offer an alternative route to prove this result. Recalling that the Chebyshev ambiguity set $\mathscr{P}(\mu, S)$ is non-empty iff the covariance matrix $S - \mu\mu^\top$ is positive semidefinite and that $\mathscr{P}_\infty$ admits a Choquet representation in terms of Dirac distributions, the proof of Theorem 4 can easily be adapted to a proof of Theorem 1 with minor modification. We omit the details for brevity, and demonstrate instead that the classical Chebyshev bound (1) arises indeed as a special case of Theorem 1.

Example 2 (Classical Chebyshev bound) To derive the classical Chebyshev bound (1), assume without loss of generality that $\mu = 0$, and define the confidence region Ξ as

$$\Xi = \{\xi \in \mathbb{R} : -\xi < \kappa\sigma, \quad \xi < \kappa\sigma\}.$$

The worst-case probability of the event $\xi \notin \Xi$ then coincides with the optimal value of the SDP of Theorem 1 and its dual, which are given by

$$\begin{array}{ll} \max \sum_{i=1}^2 \lambda_i & = \min \operatorname{Tr}\left\{ \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \right\} \\ \text{s.t. } \lambda_i, z_i, Z_i \in \mathbb{R} \ \ \forall i = 1,2 & \quad \text{s.t. } P, q, r \in \mathbb{R}, \ \tau_i \in \mathbb{R} \ \ \forall i = 1,2 \\ \quad a_i^\top z_i \ge b_i \lambda_i \ \ \forall i = 1,2 & \quad\quad \tau_i \ge 0 \ \ \forall i = 1,2 \\ \quad \sum_{i=1}^2 \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} & \quad\quad \begin{pmatrix} P & q \\ q^\top & r-1 \end{pmatrix} \succeq \tau_i \begin{pmatrix} 0 & \frac{a_i}{2} \\ \frac{a_i^\top}{2} & -b_i \end{pmatrix} \quad \forall i = 1,2 \\ \quad \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 & \quad\quad \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \succeq 0. \end{array}$$

A pair of optimal primal and dual solutions is provided in the following table. Note that the dual solution serves as a certificate of optimality for the primal solution.

Primal solution	**Dual solution**
$\lambda_1 = \lambda_2 = \begin{cases} \frac{1}{2\kappa^2} & \text{if } \kappa > 1 \\ \frac{1}{2} & \text{otherwise} \end{cases}$	$P = \begin{cases} \frac{1}{\sigma^2\kappa^2} & \text{if } \kappa > 1 \\ 0 & \text{otherwise} \end{cases}$
$z_1 = -z_2 = \begin{cases} \frac{\sigma}{2} & \text{if } \kappa > 1 \\ \frac{\kappa\sigma}{2} & \text{otherwise} \end{cases}$	$q = 0$ $r = \begin{cases} 0 & \text{if } \kappa > 1 \\ 1 & \text{otherwise} \end{cases}$
$Z_1 = Z_2 = \begin{cases} \frac{\sigma^2}{2} & \text{if } \kappa > 1 \\ \frac{\kappa^2\sigma^2}{2} & \text{otherwise} \end{cases}$	$\tau_1 = \tau_2 = \begin{cases} \frac{2}{\sigma\kappa} & \text{if } \kappa > 1 \\ 0 & \text{otherwise} \end{cases}$

The worst-case probability is thus given by $\lambda_1 + \lambda_2 = \min\left(\frac{1}{\kappa^2}, 1\right)$. Hence, Theorem 1 is a generalization of the classical Chebyshev bound (1).

In the remainder of this section we will formalize our intuition that the generalized Chebyshev bound $B(\mu,S)$ constitutes a special case of the α-unimodal bound $B_\alpha(\mu,S)$ when α tends to infinity. The next proposition establishes well-posedness of the Chebyshev bound, which is needed to prove this asymptotic result.

Proposition 3 (Well-posedness of the generalized Chebyshev bound) *The value function $B(\mu,S)$ of problem* (15) *is concave and continuous on the set of all $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$ with*

$$\begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \succ 0.$$

Proof The proof largely parallels that of Proposition 2 and is omitted. □

Note that the function $B(\mu,S)$ can be discontinuous on the boundary of its domain when the covariance matrix $S - \mu\mu^\top$ is positive semidefinite but has at least one zero eigenvalue. Since the confidence region Ξ constitutes an open polytope, there exists a converging sequence $(\xi_i)_{i\in\mathbb{N}}$ in Ξ whose limit $\xi = \lim_{i\to\infty}\xi_i$ is not contained in Ξ. Defining $\mu_i = \xi_i$ and $S_i = \xi_i\xi_i^\top$ for all $i \in \mathbb{N}$, it is clear that $\lim_{i\to\infty}(\mu_i,S_i) = (\xi,\xi\xi^\top)$. Since $\mathscr{P}(\mu_i,S_i) = \{\delta_{\xi_i}\}$ and $\xi_i \in \Xi$ for all $i \in \mathbb{N}$, we conclude that $\lim_{i\to\infty}B(\mu_i,S_i) = 0$. However, we also have $B(\xi,\xi\xi^\top) = 1$ because $\mathscr{P}(\mu,S) = \{\delta_\xi\}$ and $\xi \notin \Xi$. Thus, $B(\mu,\Sigma)$ is discontinuous at $(\xi,\xi\xi^\top)$.

We are now ready to prove that the Chebyshev bound $B(\mu,S)$ is *de facto* embedded into the family of all α-unimodal bounds $B_\alpha(\mu,S)$ for $\alpha > 0$.

Proposition 4 (Embedding of the Chebyshev bound) *For any $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$ with $S \succ \mu\mu^\top$ we have $\lim_{\alpha\to\infty}B_\alpha(\mu,S) = B(\mu,S)$.*

Proof Select any $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$ with $S \succ \mu\mu^\top$. It is clear that $\lim_{\alpha\to\infty}B_\alpha(\mu,S) \le B(\mu,S)$ as (P_α) constitutes a restriction of (15) for all $\alpha > 0$. In order to prove the converse inequality, we need the following relation between the extremal distributions of $\mathscr{P}_\alpha$ and $\mathscr{P}_\infty$:

$$\left(1 - \tfrac{1}{\alpha}\right)\delta_{x/\alpha^{\frac{1}{\alpha}}}(\xi \notin \Xi) \;\le\; \delta^\alpha_{[0,x]}(\xi \notin \Xi) \tag{16}$$

For $x/\alpha^{\frac{1}{\alpha}} \in \Xi$ the left hand side vanishes and (16) is trivially satisfied. For $x/\alpha^{\frac{1}{\alpha}} \notin \Xi$ and Ξ star-shaped, an elementary calculation shows that $\delta^\alpha_{[0,x]}(\xi \notin \Xi) \ge \delta^\alpha_{[0,x]}([x/\alpha^{\frac{1}{\alpha}},x]) = 1 - \frac{1}{\alpha}$ for all $\alpha \ge 1$. Thus, (16) holds because $\delta_{x/\alpha^{\frac{1}{\alpha}}}(\xi \notin \Xi) \le 1$. Taking mixtures with $\mathbb{m} \in \mathscr{P}\left(\frac{\alpha+1}{\alpha}\mu, \frac{\alpha+2}{\alpha}S\right)$ on both sides of (16) yields

$$\left(1 - \tfrac{1}{\alpha}\right)\int_{\mathbb{R}^n}\delta_{x/\alpha^{\frac{1}{\alpha}}}(\xi \notin \Xi)\,\mathbb{m}(\mathrm{d}x) \;\le\; \int_{\mathbb{R}^n}\delta^\alpha_{[0,x]}(\xi \notin \Xi)\,\mathbb{m}(\mathrm{d}x)\,. \tag{17}$$

By using elementary manipulations in conjunction with Lemma 1, one can show that

$$\int_{\mathbb{R}^n} \delta_{x/\alpha^{\frac{1}{\alpha}}}(\cdot)\,\mathbb{m}(\mathrm{d}x) \in \mathscr{P}\left(\alpha^{-\frac{1}{\alpha}}\,\tfrac{\alpha+1}{\alpha}\mu, \alpha^{-\frac{2}{\alpha}}\,\tfrac{\alpha+2}{\alpha}S\right) \quad \text{and} \quad \int_{\mathbb{R}^n} \delta^{\alpha}_{[0,x]}(\cdot)\,\mathbb{m}(\mathrm{d}x) \in \mathscr{P}_\alpha(\mu,S)\,.$$

Maximizing both sides of (17) over all mixture distributions $\mathbb{m} \in \mathscr{P}\left(\frac{\alpha+1}{\alpha}\mu, \frac{\alpha+2}{\alpha}S\right)$ thus yields

$$\left(1-\tfrac{1}{\alpha}\right) B\left(\alpha^{-\frac{1}{\alpha}}\,\tfrac{\alpha+1}{\alpha}\mu, \alpha^{-\frac{2}{\alpha}}\,\tfrac{\alpha+2}{\alpha}S\right) \leq B_\alpha(\mu,S)\,.$$

Since $\lim_{\alpha\to\infty}\alpha^{-\frac{1}{\alpha}}\frac{\alpha+1}{\alpha} = \lim_{\alpha\to\infty}\alpha^{-\frac{2}{\alpha}}\frac{\alpha+1}{\alpha} = 1$ and $B(\mu,S)$ is continuous whenever $S \succ \mu\mu^\top$ (see Proposition 3), we conclude that $B(\mu,S) \leq \lim_{\alpha\to\infty} B_\alpha(\mu,S)$. □

In addition to generalizing the univariate Chebyshev bound (1), the multivariate Chebyshev bound $B(\mu,S)$ also generalizes Cantelli's classical one sided inequality and the bivariate Birnbaum-Raymond-Zuckerman inequality [5]. By virtue of Proposition 4, all of these classical inequalities can now be seen as special instances of the general problem P_α.

6 Generalized Gauss bounds

In Section 2 we defined the generalized Gauss bound for star-unimodal distributions as

$$B_\star(\mu,S) = \sup_{\mathbb{P}\in\mathscr{P}_\star(\mu,S)} \mathbb{P}(\xi \notin \Xi)\,. \tag{18}$$

From Definition 7 and the subsequent discussion we know that the ambiguity set $\mathscr{P}_\star$ coincides with $\mathscr{P}_\alpha$ for $\alpha = n$; see also [10]. We thus conclude that $B_\star(\mu,S) = B_n(\mu,S)$ for all $\mu \in \mathbb{R}^n$ and $S \in \mathbb{S}^n$, which in turn implies that Theorem 2 is actually a straightforward corollary of Theorem 4 in conjunction with Lemma 3. We demonstrate now that the classical Gauss bound (2) arises indeed as a special case of Theorem 2.

Example 3 (Classical Gauss bound) To derive the classical Gauss bound (2), assume without loss of generality that $\mu = 0$, and define the confidence region Ξ as

$$\Xi = \{\xi \in \mathbb{R} \,:\, -\xi < \kappa\sigma, \quad \xi < \kappa\sigma\}\,.$$

The worst-case probability of the event $\xi \notin \Xi$ then coincides with the optimal value of the SDP of Theorem 2 and its dual, which are given by

$$\begin{array}{ll}
\max \sum_{i=1}^{2} \lambda_i - \tau_i & = \min \operatorname{Tr}\left\{\begin{pmatrix} 3S & 2\mu \\ 2\mu^\top & 1\end{pmatrix}\begin{pmatrix} P & q \\ q^\top & r\end{pmatrix}\right\} \\
\text{s.t. } \lambda_i,\tau_i,z_i,Z_i \in \mathbb{R} \;\; \forall i=1,2 & \text{s.t. } P,q,r \in \mathbb{R},\; \Lambda_i \in \mathbb{S}^2 \;\; \forall i=1,2 \\
\quad \begin{pmatrix} \tau_i b_i & \lambda_i b_i \\ \lambda_i b_i & a_i^\top z_i \end{pmatrix} \succeq 0 \;\; \forall i=1,2 & \quad \begin{pmatrix} P & q \\ q^\top & r-1 \end{pmatrix} \succeq \begin{pmatrix} 0 & \frac{a_i}{2}\Lambda_{i,2,2} \\ \frac{a_i^\top}{2}\Lambda_{i,2,2} & 2b_i\Lambda_{i,1,2} \end{pmatrix} \\
\quad \sum_{i=1}^{2}\begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} 3S & 2\mu \\ 2\mu^\top & 1 \end{pmatrix} & \quad \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \succeq 0 \\
\quad \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 & \quad b_i\Lambda_{i,1,1} \leq 1,\;\; \Lambda_i \succeq 0 \;\; \forall i=1,2.
\end{array}$$

A pair of optimal primal and dual solutions is provided in the following table. Note that the dual solution serves as a certificate of optimality for the primal solution.

Primal Solution	**Dual Solution**
$\lambda_1 = \lambda_2 = \begin{cases} \frac{2}{3\kappa^2} & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ \frac{1}{2} & \text{otherwise} \end{cases}$ $\tau_1 = \tau_2 = \begin{cases} \frac{4}{9\kappa^2} & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ \frac{\kappa}{2\sqrt{3}} & \text{otherwise} \end{cases}$ $z_1 = -z_2 = \begin{cases} \frac{\sigma}{\kappa} & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ \frac{\sqrt{3}\sigma}{2} & \text{otherwise} \end{cases}$ $Z_1 = Z_2 = \frac{3\sigma^2}{2}$	$P = \begin{cases} \frac{4}{27\sigma^2\kappa^2} & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ \frac{\kappa}{6\sqrt{3}\sigma^2} & \text{otherwise} \end{cases}$ $q = 0$ $r = \begin{cases} 0 & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ 1 - \frac{\sqrt{3}\kappa}{2} & \text{otherwise} \end{cases}$ $\Lambda_1 = \Lambda_2 = \begin{cases} \frac{1}{\sigma\kappa}\begin{pmatrix} 1 & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{9} \end{pmatrix} & \text{if } \kappa > \frac{2}{\sqrt{3}} \\ \frac{1}{\sigma\kappa}\begin{pmatrix} 1 & -\frac{\kappa}{\sqrt{3}} \\ -\frac{\kappa}{\sqrt{3}} & \frac{\kappa^2}{3} \end{pmatrix} & \text{otherwise} \end{cases}$

The worst-case probability is thus given by $(\lambda_1 - \tau_1) + (\lambda_2 - \tau_2) = \frac{4}{9\kappa^2}$ when $\kappa > \frac{2}{\sqrt{3}}$; $= 1 - \frac{\kappa}{\sqrt{3}}$ otherwise. Hence, Theorem 2 is a generalization of the classical Gauss bound (2).

In the univariate case, the Gauss bound tightens the Chebyshev bound by a factor of $4/9$. However, the tightening offered by unimodality is less pronounced in higher dimensions since $B_\star(\mu, S) = B_n(\mu, S)$ converges to the Chebyshev bound $B(\mu, S)$ when n tends to infinity. To make this point more concrete, we now provide *analytic* probability inequalities for the quantity of interest $\mathbb{P}(\max_i |\xi_i - \mu_i| \geq \kappa\sigma)$, where it is only known that $\mathbb{P} \in \mathscr{P}_\alpha$, $\mathbb{E}_\mathbb{P}[\xi] = \mu = m$ and $\mathbb{E}_\mathbb{P}[\xi\xi^\top] = \sigma^2/n\ \mathbb{I}_n + \mu\mu^\top$. This problem can be seen to constitute a multivariate generalization of the classical Chebyshev (1) and Gauss (2) inequalities, but is itself a particular case of the general problem (P_α). We have indeed $\mathbb{P}(\max_i |\xi_i - \mu_i| \geq \kappa\sigma) \leq \sup_{\mathbb{P} \in \mathscr{P}_\alpha(0, \mathbb{I}_n/n)} \mathbb{P}(\max_i |\xi_i| \geq \kappa)$. This last particular instance of the problem (P_α) however admits an analytic solution which we state here without proof [34].

Lemma 4 *We have the inequality*

$$\mathbb{P}(\max_i |\xi_i - \mu_i| \geq \kappa\sigma) \leq \begin{cases} \left(\frac{2}{\alpha+2}\right)^{\frac{2}{\alpha}} \frac{1}{\kappa^2} & \text{if } \kappa > \left(\frac{2}{\alpha+2}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha+2}{\alpha}\right)^{\frac{1}{2}}, \\ 1 - \left(\frac{\alpha}{\alpha+2}\right)^{\frac{\alpha}{2}} \kappa^\alpha & \text{otherwise} \end{cases}$$

whenever $\mathbb{P} \in \mathscr{P}_\alpha$, $\mathbb{E}_\mathbb{P}[\xi] = \mu = m$ *and* $\mathbb{E}_\mathbb{P}[\xi\xi^\top] = \sigma^2/n\ \mathbb{I}_n + \mu\mu^\top$.

The corresponding Chebyshev bound can be obtained by letting $\alpha \to \infty$ and yields $\mathbb{P}(\max_i |\xi_i - \mu_i| \geq \kappa\sigma) \leq 1/\kappa^2$ if $\kappa > 1$; $= 1$ otherwise. When compared to the corresponding Gauss bound found as a particular case $\alpha = n$ of Lemma 4, we see that the factor by which the Chebyshev bound is improved upon is indeed $4/9 = (2/(n+2))^{2/n}$, $n = 1$ for univariate problems. However, for higher dimensional problems the returns diminish as

$$\left(\frac{2}{n+2}\right)^{\frac{2}{n}} = 1 - \frac{2\log\left(\frac{n}{2}\right)}{n} + \mathscr{O}\left(\frac{1}{n^2}\right), \quad \text{and} \quad \lim_{n\to\infty} \left(\frac{2}{n+2}\right)^{\frac{2}{n}} = 1.$$

An intuitive explanation for this seemingly surprising result follows from the observation that most of the volume of a high-dimensional star-shaped set is concentrated in a thin layer near its surface. Thus, the radial distributions $\delta^n_{[0,x]}$ converge weakly to the Dirac distributions δ_x as n grows, which implies via Theorem 3 that all distributions are approximately star-unimodal in high dimensions.

7 Discussion and extensions

We now demonstrate that the tools developed in this paper can be used to solve a wide range of diverse worst-case probability problems that may be relevant for applications in distributionally robust optimization, chance constrained programming or other areas of optimization under uncertainty. The corresponding ambiguity sets are more general than $\mathscr{P}_\alpha(\mu,S)$, which contains merely α-unimodal distributions with precisely known first and second-order moments.

Moment ambiguity: The worst-case probability bound $B_\alpha(\mu,S)$ requires full and accurate information about the mean and covariance matrix of the random vector ξ. In practice, however, these statistics must typically be estimated from noisy historical data and are therefore themselves subject to ambiguity. Assume therefore that the first and second-order moments are known only to belong to an SDP-representable confidence set $\mathscr{M}$ [23], that is,

$$\begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \in \mathscr{M} \subseteq \mathbb{S}_+^{n+1}. \tag{19}$$

Then, the worst-case probability problem (P) with ambiguity set $\mathscr{P} = \cup_{(\mu,S)} \mathscr{P}_\alpha(\mu,S)$, where the union is taken over all μ and S satisfying condition (19), can be reformulated as a tractable SDP. This is an immediate consequence of the identity

$$\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\xi \notin \Xi) = \max_{\mu,S} \; B_\alpha(\mu,S) \quad \text{s.t.} \; \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \in \mathscr{M}.$$

Theorem 4 and its corollaries imply that the optimization problem on the right hand side of the above expression admits a tractable SDP reformulation. Hence, the α-unimodal bound $B_\alpha(\mu,S)$ can be generalized to handle ambiguity in both the mean and the covariance matrix.

Support information: Suppose that, in addition to being α-unimodal, the distribution of ξ is known to be supported on a convex closed polytope representable as

$$\mathscr{C} = \left\{ x \in \mathbb{R}^n \;:\; c_i^\top x - d_i \le 0 \;\forall i = 1,\dots,l \right\}. \tag{20}$$

In this case, we could use an ambiguity set of the form $\mathscr{P} = \mathscr{P}_\alpha(\mu,S) \cap \mathscr{P}(\mathscr{C})$, where $P(\mathscr{C}) = \{\mathbb{P} \in \mathscr{P}_\infty \;:\; \mathbb{P}(\mathscr{C}) = 1\}$. Unfortunately, even checking whether the ambiguity set $\mathscr{P}$ is non-empty is NP-hard in general. Indeed, if $\mathscr{C}$ is given by the nonnegative orthant $\mathbb{R}_+^n$, it can be seen from the Choquet representation of $\mathscr{P}_\alpha(\mu,S)$ that checking whether $\mathscr{P}$ is non-empty is equivalent to checking whether the matrix (11) is completely positive, which is hard for any $\alpha > 0$. A tractable alternative ambiguity set is given by $\mathscr{P} = \cup_{S' \preceq S} \mathscr{P}_\alpha(\mu,S') \cap \mathscr{P}(\mathscr{C})$. The resulting generalized moment problem treats S as an upper bound (in the positive semidefinite sense) on the second-order moment matrix of ξ, as suggested in [9]. This relaxation is justified by the observation that the worst-case distribution in (P) tends to be maximally spread out and thus typically attains the upper bound imposed on its second-order moments. In all other cases, the relaxation results in a conservative estimate for the worst-case probability of the event $\xi \notin \Xi$. However, the relaxed problem always admits an exact reformulation as an SDP. This result is formalized in the following proposition. The proof is omitted because it requires no new ideas.

Proposition 5 (Support information) *The worst-case probability problem* (P) *with ambiguity set* $\mathscr{P} = \cup_{S' \preceq S} \mathscr{P}_\alpha(\mu,S') \cap \mathscr{P}(\mathscr{C})$*, where and* Ξ *and* $\mathscr{C}$ *are defined as in* (3) *and* (20)*, respectively, can be reformulated as*

$$\begin{aligned}
\max \quad & \sum_{i=1}^k (\lambda_i - \tau_i) \\
\text{s.t.} \quad & z_i \in \mathbb{R}^n,\; Z_i \in \mathbb{S}^n,\; \lambda_i \in \mathbb{R},\; \tau_i \in \mathbb{R} \quad \forall i = 1,\dots,k \\
& \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0 \quad \forall i = 1,\dots,k \\
& c_j z_i^\top \le d_j \lambda_i \quad \forall i = 1,\dots,k,\; \forall j = 1,\dots,l \\
& \sum_{i=1}^k \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \preceq \begin{pmatrix} \frac{\alpha+2}{\alpha} S & \frac{\alpha+1}{\alpha}\mu \\ \frac{\alpha+1}{\alpha}\mu^\top & 1 \end{pmatrix} \\
& \left(a_i^\top z_i\right)^\alpha \tau_i \ge \lambda_i^{\alpha+1} b_i^\alpha,\; a^\top z_i \ge 0,\; \tau_i \ge 0 \;\forall i = 1,\dots,k,
\end{aligned}$$

which is equivalent to a tractable SDP for any rational $\alpha \geq 1$.

Multimodality: Finally, assume that the distribution $\mathbb{P}$ of ξ is known to be α-multimodal around given centers $c_m \in \Xi$, $m = 1, \ldots, M$. This means that $\mathbb{P} = \sum_{m=1}^{M} p_m \mathbb{P}_m$, where p_m represents the probability of mode m and $\mathbb{P}_m \in \mathscr{P}_\alpha(c_m)$ the conditional distribution of ξ in mode m, while $\mathscr{P}_\alpha(c_m)$ denotes the set of all α-unimodal distributions with center c_m, $m = 1, \ldots, M$. The modal probabilities must be nonnegative and add up to 1 but may otherwise be uncertain. In the following, we thus assume that $p = (p_1, \ldots, p_M)$ is only known to be contained in an SDP-representable uncertainty set $\mathscr{U}$ contained in the probability simplex in $\mathbb{R}^M$. If $\mathbb{P}$ has known first and second-order moments as usual, then it is an element of the ambiguity set

$$\mathscr{P} = \left\{ \sum_{m=1}^{M} p_m \mathbb{P}_m \ : \ p \in \mathscr{U}, \ \mathbb{P}_m \in \mathscr{P}_\alpha(c_m) \ \forall m = 1, \ldots, M \right\} \bigcap \mathscr{P}(\mu, S). \tag{21}$$

Checking whether $\mathscr{P}$ is non-empty is equivalent to checking the feasibility of an LMI:

Proposition 6 *The ambiguity set $\mathscr{P}$ defined in* (21) *is non-empty iff there exist $Z_m \in \mathbb{S}^n$, $z_m \in \mathbb{R}^n$ and $\lambda_m \in \mathbb{R}$, $m = 1, \ldots, M$, satisfying*

$$\begin{aligned}
&\begin{pmatrix} Z_m & z_m \\ z_m^\top & \lambda_m \end{pmatrix} \succeq 0 \quad \forall m = 1, \ldots, M, \quad (\lambda_1, \ldots, \lambda_M) \in \mathscr{U}, \\
&\begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} = \sum_{m=1}^{M} \begin{pmatrix} \frac{\alpha}{\alpha+2} Z_m + \frac{\alpha}{\alpha+1} \left(z_m c_m^\top + c_m z_m^\top \right) + \lambda_m c_m c_m^\top & \frac{\alpha}{\alpha+1} z_m + \lambda_m c_m \\ \frac{\alpha}{\alpha+1} z_m^\top + \lambda_m c_m^\top & \lambda_m \end{pmatrix}.
\end{aligned}$$

It can easily be verified that this result reduces to the result of Proposition 1 in the special case when there is only a single mode centered at the origin. Thus, its proof is omitted for brevity of exposition. Proposition 6 constitutes a key ingredient for the proof of the next proposition, which establishes that the worst-case probability problem (P) with an ambiguity set of the form (21) can be formulated once again as a tractable SDP. The proof is omitted because it requires no new ideas.

Proposition 7 **(Multimodality)** *The worst-case probability problem* (P)*, where the confidence region Ξ and the ambiguity set $\mathscr{P}$ are defined as in* (3) *and* (21)*, respectively, can be reformulated as*

$$\begin{aligned}
\max \quad & \sum_{m=1}^{M} \sum_{i=1}^{k} (\lambda_{i,m} - \tau_{i,m}) \\
\text{s.t.} \quad & z_{i,m} \in \mathbb{R}^n, \ Z_{i,m} \in \mathbb{S}^n, \ \lambda_{i,m} \in \mathbb{R}, \ \tau_{i,m} \in \mathbb{R} \quad \forall i = 1, \ldots, k, \ \forall m = 1, \ldots, M \\
& \sum_{i=0}^{k} (\lambda_{i,1}, \ldots, \lambda_{i,M}) \in \mathscr{U}, \quad \begin{pmatrix} Z_{i,m} & z_{i,m} \\ z_{i,m}^\top & \lambda_{i,m} \end{pmatrix} \succeq 0 \quad \forall i = 0, \ldots, k, \ m = 1, \ldots, M \\
& \sum_{m=1}^{M} \sum_{i=0}^{k} \begin{pmatrix} \frac{\alpha}{\alpha+2} Z_{i,m} + \frac{\alpha}{\alpha+1} \left(z_{i,m} c_m^\top + c_m z_{i,m}^\top \right) + \lambda_{i,m} c_m c_m^\top & \frac{\alpha}{\alpha+1} z_{i,m} + \lambda_{i,m} c_m \\ \frac{\alpha}{\alpha+1} z_i^\top + \lambda_{i,m} c_m^\top & \lambda_{i,m} \end{pmatrix} \\
& \qquad = \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \\
& \left(a_i^\top z_{i,m} \right)^\alpha \tau_{i,m} \geq \lambda_{i,m}^{\alpha+1} \left(b_i - a_i^\top c_m \right)^\alpha \ \forall i = 1, \ldots, k, \ \forall m = 1, \ldots, M \\
& a^\top z_{i,m} \geq 0, \ \tau_{i,m} \geq 0 \ \forall i = 1, \ldots, k, \ \forall m = 1, \ldots, M,
\end{aligned}$$

which is equivalent to a tractable SDP for any rational $\alpha \geq 1$.

We remark that our techniques can be used to derive many other worst-case probability bounds involving α-unimodal distributions by, e.g., restricting $\mathscr{P}$ to contain only symmetric distributions or by combining any of the results discussed in this section. These further generalizations are omitted for the sake of brevity, and we consider instead a practical application of the new worst-case probability bounds derived in this paper.

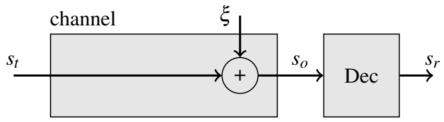

Fig. 1 Upon transmitting the symbol s_t, a noisy output $s_o = s_t + \xi$ is received and decoded using a maximum likelihood decoder into the symbol s_r.

8 Digital communication limits

We use the generalized Gauss bounds presented in this paper to estimate the probability of correct signal detection in a digital communication example inspired by [6]. All SDP problems are implemented in `Matlab` via the `YALMIP` interface and solved using `SDPT3`.

Consider a set of c possible symbols or signals $\mathscr{S} = \{s_1, \ldots, s_c\} \subseteq \mathbb{R}^2$, which is termed the signal constellation. The signals are transmitted over a noisy communication channel and perturbed by additive noise. A transmitted signal s_t thus results in an output $s_o = s_t + \xi$, where $\xi \in \mathbb{R}^2$ follows a star-unimodal distribution with zero mean and covariance matrix $\sigma^2 \mathbb{I}_2$. A *minimum distance detector*[1] then decodes the output, that is, it determines the symbol $s_r \in \mathscr{S}$ that is closest in Euclidian distance to the output s_o. Note that the detector is uniquely defined by the Voronoi diagram implied by the signal constellation $\mathscr{S}$ as shown in Figure 2(a).

The quantity of interest is the average probability of correct symbol transmission

$$p = \frac{1}{c}\sum_{i=1}^{c} \mathbb{P}(s_i + \xi \in C_i) = 1 - \frac{1}{c}\sum_{i=1}^{c} \mathbb{P}(s_i + \xi \notin C_i),$$

where C_i is the (polytopic) set of outputs that are decoded as s_i. The generalized Chebyshev bound of Theorem 1 and the generalized Gauss bound of Theorem 2 both provide efficiently computable lower bounds on p, which are plotted in Figure 2(b) as a function of the Channel noise power σ. Note that the generalized Gauss bound is substantially tighter because the Chebyshev bound disregards the star-unimodality of the channel noise. For the sake of comparison, Figure 2(b) also shows the probability of correct detection when the noise ξ is assumed to be normal or block uniformly distributed. We say that a random variable is block uniformly distributed if it is distributed uniformly on a square in $\mathbb{R}^n$. The latter reference probability was computed using numerical integration. Using the procedure described in Section 4.2, we are able to explicitly construct a worst-case distribution for $\mathbb{P}(s_1 + \xi \notin C_1)$. In Figures 3(a) and 3(b) the support of these worst-case distributions for respectively the Chebyshev and Gauss bound are shown in case the channel noise is $\sigma = 1$. We used `Matlab` on a PC[2] operated by `Debian GNU/Linux 7 (wheezy)` in combination with `YALMIP` [19] and `SDPT3` [35] to solve the resulting convex SDPs. The Chebyshev and Gauss bounds depicted in Figure 2(b) each required the solution of 700 SDPs, with seven SDPs per channel noise level. On our computing hardware it took on average 1.0 s and 2.0 s to solve each of the 700 SDPs for the Chebyshev and Gauss bounds, respectively, to an accuracy of six significant digits.

9 Conclusion

We have generalized the classical Gauss inequality to multivariate probability distributions by reformulating the corresponding worst-case probability problems as tractable SDPs. It was shown that the Chebyshev and Gauss bounds are intimately related and can be obtained from the solutions of generalized α-unimodal moment problems for $\alpha = \infty$ and $\alpha = n$, respectively. The paper also provides a new perspective on the computational solution of generalized moment problems. Instead of duality arguments, which are employed by the overwhelming majority of the existing literature, we used Choquet representations of the relevant ambiguity sets to derive tractable reformulations. We expect that the methods developed in this work might be extendible to different types of structured ambiguity sets (e.g. for symmetric distributions) and could thus spur a renewed interest in generalized moment problems.

[1] If the noise is Gaussian, then minimum distance decoding is the same as maximum likelihood decoding.

[2] An `Intel(R) Core(TM) Xeon(R) CPU E5540 @ 2.53GHz` machine.

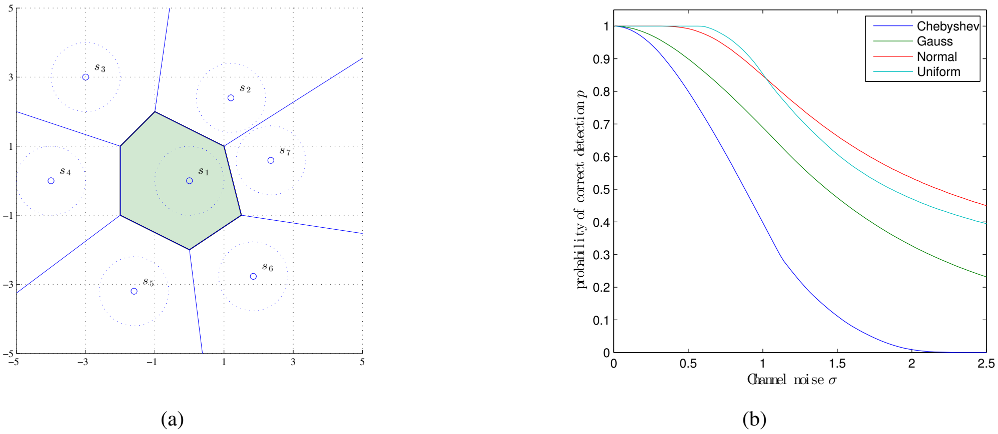

Fig. 2 Figure 2(a) depicts the signal constellation $\mathscr{S}$. The distribution of the outputs is visualized by the dashed circles, while the detector is visualized by its Voronoi diagram. For example, the green polytope represents the set of outputs s_o which are decoded as s_1. Figure 2(b) shows the lower bounds on the correct detection probabilities as predicted by the Chebyshev and Gauss inequalities. The exact detection probability for normal and block uniform distributed noise is shown for the sake of comparison.

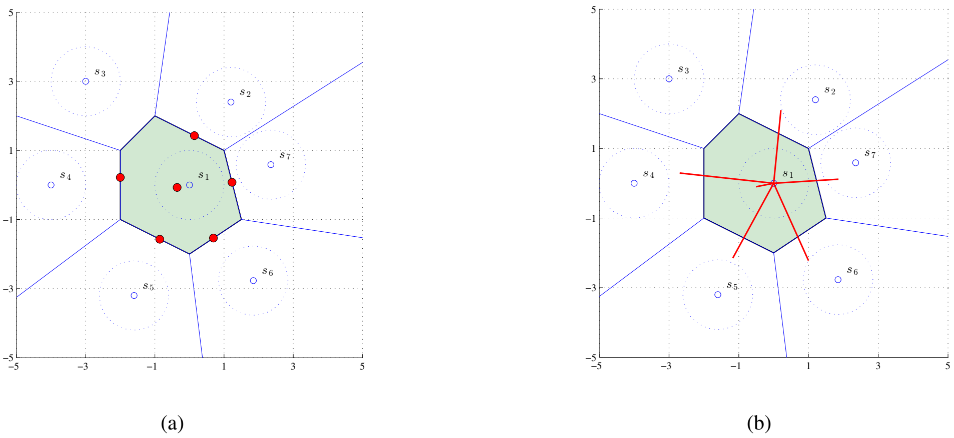

Fig. 3 Figures 3(a) and 3(b) depict the worst-case distributions for $\mathbb{P}(s_1 + \xi \notin C_1)$ for the Chebyshev and Gauss bounds, respectively, in case of the channel noise power $\sigma = 1$.

A Unimodality of ambiguity sets

In Section 4 the notion of α-unimodality was mainly used as a theoretical tool for bridging the gap between the Chebyshev and Gauss inequalities. The purpose of this section is to familiarize the reader with some of the properties and practical applications of α-unimodal ambiguity sets. This section borrows heavily from the standard reference on unimodal distributions [10] and the technical report [24] where α-unimodality was first proposed.

Since some of the properties of α-unimodal ambiguity sets depend not only on α but also on the dimension n, we make this dependence now explicit and denote the set of all α-unimodal distributions supported on $\mathbb{R}^n$ as $\mathscr{P}_{\alpha,n}$. Taking $B = \mathbb{R}^n$ in Definition 7, it is clear that $\mathscr{P}_{\alpha,n} = \emptyset$ for all $\alpha < 0$. Similarly, taking B a neighborhood around the origin shows that $\mathscr{P}_{0,n} = \{\delta_0\}$ justifying the condition $\alpha > 0$ required in Definition 7. As the α-unimodal ambiguity sets enjoy the nesting property $\mathscr{P}_{\alpha,n} \subseteq \mathscr{P}_{\beta,n}$ whenever $\alpha \leq \beta \leq \infty$, we may define the α-unimodality value of a generic ambiguity set $\mathscr{P}$ as the smallest α for which $\mathscr{P} \subseteq \mathscr{P}_{\alpha,n}$. In [24, Lemma 1] it is shown that the infimum over α is always achieved. Table 1 reports the α-unimodality values for some common distributions (that is, singleton ambiguity sets).

The notion of α-unimodality was originally introduced after the discovery that the projections of star-unimodal distributions need not be star-unimodal.[3] Indeed, a situation in which star-unimodality is not preserved under a projection is visualized in Figure 4. A correct projection property for α-unimodal distributions is given in the following theorem.

[3] Historically, in their famous monograph on the limit distributions of sums of independent random variables, Gnedenko and Kolmogorov [12] used a false theorem owing to Lapin stating a projection property for unimodal distributions. Chung highlighted the mistake in his English translation of the monograph.

$\mathbb{P}$	$\inf\{\alpha \,:\, \mathbb{P} \in \mathscr{P}_{\alpha,n}\}$
Dirac distribution δ_0	0
Normal distribution $N(0,\Sigma)$	n
Uniform distribution on star-shaped set	n
Dirac distribution δ_a with $a \neq 0$	∞

Table 1 α-unimodality values of some common distributions.

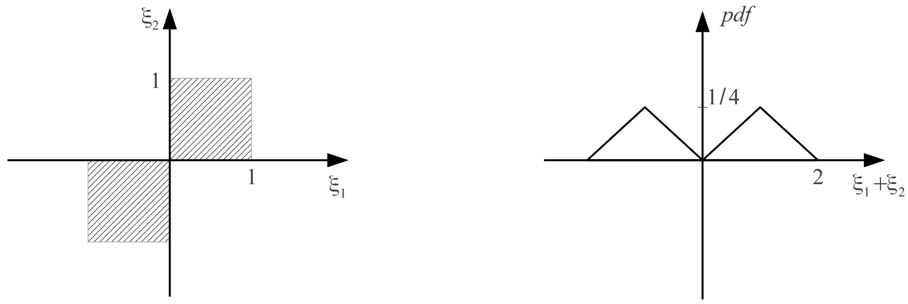

Fig. 4 The univariate distribution visualized in the right panel of the figure is not star-unimodal despite being the marginal projection of the uniform distribution on the star-shaped set $\{\xi \in \mathbb{R}^2 \,:\, 0 \le \xi_1 \le 1,\, 0 \le \xi_2 \le 1\} \cup \{\xi \in \mathbb{R}^2 \,:\, -1 \le \xi_1 \le 0,\, -1 \le \xi_2 \le 0\}$, shown in the left panel, under the linear transformation $A : \mathbb{R}^2 \to \mathbb{R}$, $(\xi_1, \xi_2) \mapsto \xi_1 + \xi_2$.

Theorem 5 (Projection property [10]) *If the random vector $\xi \in \mathbb{R}^n$ has a distribution $\mathbb{P} \in \mathscr{P}_{\alpha,n}$ for some $0 < \alpha \le \infty$, and A is a linear transformation mapping $\mathbb{R}^n$ to $\mathbb{R}^m$, then the distribution of $A\xi$ belongs to $\mathscr{P}_{\alpha,m}$.*

The projection property of Theorem 5 has great theoretical and practical value because it justifies, for instance, the identity

$$\sup_{\mathbb{P}\in\mathscr{P}_{\alpha,n}(\mu,S)} \mathbb{P}(A\xi \notin \Xi) \quad = \quad \sup_{\mathbb{P}\in\mathscr{P}_{\alpha,m}(A\mu,ASA^\top)} \mathbb{P}(\xi \notin \Xi)\,,$$

which can be useful for dimensionality reduction. See [42] for concrete practical applications of this identity. Additionally, Theorem 5 allows us to interpret the elements of $\mathscr{P}_{\alpha,n}$ for $\alpha \in \mathbb{N}$ as projections of star-unimodal distributions on $\mathbb{R}^\alpha$.

We can now envisage a further generalization of the definition of α-unimodality. Specifically, we can denote a distribution $\mathbb{P}$ as ϕ-unimodal if $\phi(t)\mathbb{P}(B/t)$ is non-decreasing in $t \in (0,\infty)$ for every Borel set $B \in \mathscr{B}(\mathbb{R}^n)$, while the set of all ϕ-unimodal distributions on $\mathbb{R}^n$ is denoted by $\mathscr{P}_{\phi,n}$. This is indeed a natural generalization as it reduces to the definition of α-unimodality for $\phi(t) = t^\alpha$. However, as shown in [24], the notions of ϕ-unimodality and α-unimodality are in fact equivalent in the sense that $\mathscr{P}_{\alpha,n} = \mathscr{P}_{\phi,n}$ for

$$\alpha = \inf_{t>0} \frac{1}{t\phi(t+0)} \frac{\mathrm{d}\phi(t)}{\mathrm{d}t},$$

where $\frac{\mathrm{d}\phi(t)}{\mathrm{d}t}$ represents the lower right Dini derivative of $\phi(t)$.

References

1. Bertsimas, D., Popescu, I.: On the relation between option and stock prices: A convex optimization approach. Operations Research **50**(2), 358–374 (2002)
2. Bertsimas, D., Popescu, I.: Optimal inequalities in probability theory: A convex optimization approach. SIAM Journal on Optimization **15**(3), 780–804 (2005)
3. Bienaymé, I.J.: Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. Comptes Rendus de l'Académie des Sciences **37**, 159–184 (1853)
4. Billingsley, P.: Convergence of Probability Measures, *Wiley Series in Probability and Statistics*, vol. 493. John Wiley & Sons (2009)
5. Birnbaum, Z., Raymond, J., Zuckerman, H.: A generalization of Tshebyshev's inequality to two dimensions. The Annals of Mathematical Statistics **18**(1), 70–79 (1947)
6. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
7. Chebyshev, P.: Des valeurs moyennes. Journal de Mathématiques Pures et Appliquées **12**(2), 177–184 (1867)
8. Cheng, J., Delage, E., Lisser, A.: Distributionally robust stochastic knapsack problem. Tech. rep., HEC Montréal (2013)
9. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations Research **58**(3), 595–612 (2010)
10. Dharmadhikari, S., Joag-Dev, K.: Unimodality, Convexity, and Applications, *Probability and Mathematical Statistics*, vol. 27. Academic Press (1988)
11. Gauss, C.: Theoria combinationis observationum erroribus minimis obnoxiae, pars prior. Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores **33**, 321–327 (1821)
12. Gnedenko, B., Kolmogorov, A., Chung, K., Doob, J.: Limit Distributions for Sums of Independent Random Variables, *Addison-Wesley Series in Statistics*, vol. 233. Addison-Wesley (1968)

13. Grundy, B.: Option prices and the underlying asset's return distribution. The Journal of Finance **46**(3), 1045–1069 (1991)
14. Hadamard, J.: Sur les problèmes aux dérivées partielles et leur signification physique. Princeton University Bulletin **13**, 49–52 (1902)
15. Isii, K.: On sharpness of Tchebycheff-type inequalities. Annals of the Institute of Statistical Mathematics **14**(1), 185–197 (1962)
16. Lanckriet, G., El Ghaoui, L., Bhattacharyya, C., Jordan, M.: A robust minimax approach to classification. The Journal of Machine Learning Research **3**, 555–582 (2003)
17. Lasserre, J.: Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization **11**(3), 796–817 (2001)
18. Lo, A.: Semi-parametric upper bounds for option prices and expected payoffs. Journal of Financial Economics **19**(2), 373–387 (1987)
19. Löfberg, J.: Yalmip: A toolbox for modeling and optimization in matlab. In: Proceedings of the CACSD Conference, pp. 284–289 (2004)
20. Markov, A.: On certain applications of algebraic continued fractions. Ph.D. thesis, St. Petersburg (1884)
21. Meaux, L., Seaman Jr, J., Boullion, T.: Calculation of multivariate Chebyshev-type inequalities. Computers & Mathematics with Applications **20**(12), 55–60 (1990)
22. Natarajan, K., Song, M., Teo, C.P.: Persistency model and its applications in choice modeling. Management Science **55**(3), 453–469 (2009)
23. Nesterov, Y., Nemirovskii, A.: Interior-Point Polynomial Algorithms in Convex Programming, *Studies in Applied and Numerical Mathematics*, vol. 13. SIAM (1994)
24. Olshen, R., Savage, L.: A Generalized Unimodality. Tech. Rep. 143, Stanford University (1969)
25. Phelps, R.: Lectures on Choquet's Theorem, *Lecture Notes in Mathematics*, vol. 1757. Springer (2001)
26. Pólik, I., Terlaky, T.: A survey of the S-lemma. SIAM Review **49**(3), 371–418 (2007)
27. Popescu, I.: A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. Mathematics of Operations Research **30**(3), 632–657 (2005)
28. Rockafellar, R., Wets, R.B.: Variational Analysis, *Grundlehren der mathematischen Wissenschaften*, vol. 317. Springer (1998)
29. Rogosinski, W.: Moments of non-negative mass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **245**(1240), 1–27 (1958)
30. Sellke, T.: Generalized Gauss-Chebyshev inequalities for unimodal distributions. Metrika **43**(1), 107–121 (1996)
31. Shapiro, A.: On duality theory of conic linear problems. Nonconvex Optimization and its Applications **57**, 135–155 (2001)
32. Shapiro, A., Kleywegt, A.: Minimax analysis of stochastic problems. Optimization Methods and Software **17**(3), 523–542 (2002)
33. Smith, J.: Generalized Chebychev inequalities: Theory and applications in decision analysis. Operations Research **43**(5), 807–825 (1995)
34. Stellato, B.: Data-driven chance constrained optimization (2014). DOI 10.3929/ethz-a-010266857
35. Tütüncü, R., Toh, K., Todd, M.: Solving semidefinite-quadratic-linear programs using SDPT3. Mathematical programming **95**(2), 189–217 (2003)
36. Van Parys, B., Kuhn, D., Goulart, P., Morari, M.: Distributionally robust control of constrained stochastic systems. Tech. rep., ETH Zürich (2013)
37. Vandenberghe, L., Boyd, S., Comanor, K.: Generalized Chebyshev bounds via semidefinite programming. SIAM Review **49**(1), 52–64 (2007)
38. Vorobyov, S., Chen, H., Gershman, A.: On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints. IEEE Transactions on Signal Processing **56**(11), 5719–5724 (2008)
39. Xu, H., Caramanis, C., Mannor, S.: Optimization under probabilistic envelope constraints. Operations Research **60**(3), 682–699 (2012)
40. Yamada, Y., Primbs, J.: Value-at-risk estimation for dynamic hedging. International Journal of Theoretical and Applied Finance **5**(4), 333–354 (2002)
41. Ye, Y.: Interior Point Algorithms: Theory and Analysis, *Wiley Series in Discrete Mathematics and Optimization*, vol. 44. John Wiley & Sons (2011)
42. Yu, Y.L., Li, Y., Schuurmans, D., Szepesvári, C.: A general projection property for distribution families. In: Advances in Neural Information Processing Systems, pp. 2232–2240 (2009)
43. Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. Mathematical Programming **137**(1-2), 167–198 (2013)
44. Zymler, S., Kuhn, D., Rustem, B.: Worst-case value-at-risk of non-linear portfolios. Management Science **59**(1), 172–188 (2013)