This is a pre-print of an article published by *Mathematical Programming*. The final authenticated version is available online at: http://doi.org/10.1007/s10107-017-1220-x

# Distributionally robust expectation inequalities for structured distributions

Bart P.G. Van Parys, Paul J. Goulart, Manfred Morari

**Abstract**

Quantifying the risk of unfortunate events occurring, despite limited distributional information, is a basic problem underlying many practical questions. Indeed, quantifying constraint violation probabilities in distributionally robust programming or judging the risk of financial positions can both be seen to involve risk quantification, notwithstanding distributional ambiguity. In this work we discuss worst-case probability and conditional value-at-risk (CVaR) problems, where the distributional information is limited to second-order moment information in conjunction with structural information such as unimodality and monotonicity of the distributions involved. We indicate how *exact and tractable* convex reformulations can be obtained using standard tools from Choquet and duality theory. Our reformulations can be embedded conveniently into higher-level problems such as distributionally robust programs. We make our theoretical results concrete with a stock portfolio pricing problem and an insurance risk aggregation example.

***Keywords***— optimal inequalities, extreme distributions, convex optimisation, CVaR

## 1 Introduction

In a wide range of applications, one is faced with the problem of quantifying the expected cost $L(\xi)$ of a random variable $\xi$ with distribution $\mathbb{P}$. Common problems include determining the expected profit of a stock portfolio with uncertain stock returns [3,9], or quantifying the symbol error rate in a noisy communication channel [27]. When the distribution $\mathbb{P}$ of the random vector $\xi$ is known, computing $\mathbb{E}_{\mathbb{P}}\{L(\xi)\}$ typically reduces to the evaluation of a (high dimensional) integral. The evaluation of a high dimensional integral is in general a computationally formidable task [11].

Furthermore, in practice it is often the case that the information available concerning the distribution $\mathbb{P}$ is limited. This means that the distribution of $\xi$ is ambiguous and only known to belong to some *ambiguity set* $\mathcal{P}$ containing all distributions consistent with the known partial information concerning $\mathbb{P}$. We are thus limited to providing an upper bound on the expected cost $\mathbb{E}_{\mathbb{P}}\{L(\xi)\}$ holding uniformly for all distributions $\mathbb{P}$ in the ambiguity set $\mathcal{P}$. Hence when faced with limited information on the distribution of $\xi$, the least upper bound on the expected cost is $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\{L(\xi)\}$.

Unfortunately, such worst-case expectation bounds or inequalities are generally unavailable in closed form, except in special cases where one can resort to classical bounds such as the Chebyshev or Gauss inequalities [21]. On the other hand, tractable reformulations based on convex programming are known [27, 29] for the case where $\mathcal{P}$ consists of all distributions sharing a known mean and variance. Thanks to modern interior point algorithms [13], these convex programming reformulations provide a de facto closed form solution to the worst-case expectation problem. The resulting inequalities are widely used across many different disciplines such as distributionally robust optimisation [4] and control [25, 26] or portfolio selection and hedging [28, 30].

The main downside of these inequalities stems from the fact that the ambiguity set $\mathcal{P}$, consisting of all distributions sharing a known mean and variance, contains distributions that are not realistic in many applications and that consequently render the inequalities overly pessimistic. Indeed, the distributions achieving the worst-case expectation bound generically have discrete support with a finite number of discretisation points. Fortunately, recent work has demonstrated that this pessimism can be partially mitigated by restricting the ambiguity set $\mathcal{P}$ to contain only distributions satisfying additional structural requirements [18, 24]. In this paper, we will therefore consider the following worst-case expectation problem with second-order moment information:

$$\begin{array}{lll} B_{\mathrm{wc}}\left(L, \mathcal{P}_s, \mu, S\right) := \sup\limits_{\mathbb{P}\in\mathcal{P}_s} & \mathbb{E}_{\mathbb{P}}\{L(\xi)\} & \\ \text{s.t.} & \mathbb{P} \in \mathcal{P}(\mu, S), & \end{array} \tag{$P_{\mathrm{wc}}$}$$

where the ambiguity set $\mathcal{P}(\mu, S)$ is defined as the collection of all distributions sharing a known

mean and variance $\mathcal{P}(\mu, S) = \left\{ \mathbb{P} \in \mathcal{P}_n \;\middle|\; \int x\, \mathbb{P}(\mathrm{d}x) = \mu,\ \int xx^\top\, \mathbb{P}(\mathrm{d}x) = S \right\}$. The set $\mathcal{P}_s$ will be used to characterize any further structural information about the distributions $\mathbb{P}$ considered, e.g. symmetry, unimodality or monotonicity. When $\mathcal{P}_s$ is taken to be the standard probability simplex $\mathcal{P}_n$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then the worst-case expectation problem reduces to the standard generalised moment problem discussed in [27,29]. The principal aim of this paper is to provide a unified approach to the situations under which problem $(P_{\mathrm{wc}})$ admits a tractable reformulation, specifically for those situations in which $\mathcal{P}_s$ is more richly structured.

### 1.1 Conditional value-at-risk

A closely related and popular alternative to the expected cost of $L(\xi)$ is its expected shortfall or conditional value-at-risk (CVaR).

**Definition 1.1** (CVaR)**.** *For any measurable loss function $L : \mathbb{R}^n \to \mathbb{R}$, probability distribution $\mathbb{P}$ and tolerance $\epsilon \in (0, 1)$, the CVaR of the random loss $L(\xi)$ at level $\epsilon$ with respect to $\mathbb{P}$ is defined as*

$$\mathbb{P}\text{-CVaR}_\epsilon\left(L(\xi)\right) := \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left\{ (L(\xi) - \beta)^+ \right\} \right\}. \tag{1}$$

Rockafellar and Uryasev [19] have shown that the set of optimal solutions for $\beta$ in (1) is a closed interval whose left endpoint is given by the $1 - \epsilon$ quantile of $L(\xi)$. Moreover, it can be shown that if the random loss $L(\xi)$ follows a continuous distribution, then CVaR coincides with the conditional expectation of $L(\xi)$ above its $1 - \epsilon$ quantile. This observation originally motivated the term *conditional value-at-risk*.

While CVaR is an interesting risk measure, it nevertheless still requires that the distribution of $\xi$ be known. As in the worst-case expectation problem, we therefore consider instead the following worst-case CVaR problem:

$$\begin{array}{ll} B_{\mathrm{CVaR}} := \sup\limits_{\mathbb{P} \in \mathcal{P}_s} & \mathbb{P}\text{-CVaR}_\epsilon\left(L(\xi)\right) \\ \text{s.t.} & \mathbb{P} \in \mathcal{P}(\mu, S). \end{array}$$

However, from a computational point of view the CVaR problem can be reduced to a worst-case

expectation problem. Defining $\mathcal{L}(\beta, \mathbb{P}) := \beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left\{(L(\xi) - \beta)^+\right\}$ and recalling the definition (1), our worst-case CVaR problem becomes

$$\begin{aligned} B_{\text{CVaR}} &= \sup_{\mathbb{P}\in\mathcal{P}} \inf_{\beta} \ \mathcal{L}(\beta, \mathbb{P}) = \inf_{\beta} \sup_{\mathbb{P}\in\mathcal{P}} \ \mathcal{L}(\beta, \mathbb{P}) \\ &= \inf_{\beta} \left\{ \beta + \sup_{\mathbb{P}\in\mathcal{P}} \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left\{(L(\xi) - \beta)^+\right\} \right\}. \end{aligned}$$

Since $\mathcal{L}(\beta, \mathbb{P})$ is convex in $\beta$ and linear in $\mathbb{P}$, the interchange of the supremum and infimum operations is justified when the ambiguity set $\mathcal{P} = \mathcal{P}_s \cap \mathcal{P}(\mu, S)$ is weakly closed by virtue of a stochastic saddle point theorem due to [23]. The worst-case expectation problem can now be seen to constitute an inner problem in the worst-case CVaR problem. Since the optimal $\beta^\star$ is known to lie in a closed interval [19] and $\sup_{\mathbb{P}\in\mathcal{P}} \ \mathcal{L}(\beta, \mathbb{P})$ is convex in $\beta$, computing a solution to the worst-case CVaR problem reduces to solving a sequence of worst-case expectation problems. For instance, the golden section search can be used to optimise $\sup_{\mathbb{P}\in\mathcal{P}} \ \mathcal{L}(\beta, \mathbb{P})$ only requiring a polynomial number of evaluations $\sup_{\mathbb{P}\in\mathcal{P}} \ \mathbb{E}_{\mathbb{P}}\left\{(L(\xi) - \beta)^+\right\}$ [8]. Hence in what follows, we will deal with the more general worst-case expectation problem ($P_{\text{wc}}$) directly.

### 1.2 Outline of the paper

In Section 2, we describe the worst-case expectation bound $B_{\text{wc}}(L, \mathcal{P}_n, \mu, S)$ over the standard simplex $\mathcal{P}_n$. We then show how the more general expectation bound $B_{\text{wc}}(L, \mathcal{P}_s, \mu, S)$ over the restricted ambiguity set $\mathcal{P}_s \subseteq \mathcal{P}_n$ can be reduced to an equivalent expectation bound $B_{\text{wc}}(L_s, \mathcal{P}_n, \mu_s, S_s)$ over the standard simplex $\mathcal{P}_n$ using an integral or Choquet star representation of $\mathcal{P}_s$. In Sections 3, we show that two important classes of structured distributions – namely unimodal and monotone distributions – admit such Choquet star representations. In Section 4 we make the abstract results concrete for the case of unimodal and monotone distributions, respectively, for both worst case probability and expectation inequalities. Section 5 illustrates the results on an stock portfolio problem and an insurance risk aggregation problem.

The main results presented in this paper, from a practitioners point of view, are summarised in Table 1. We will focus mainly on indicator functions of polytopic sets $\Xi$ which arise in worst-case probability inequalities and piecewise affine functions which arise when dealing with convex cost functions $L : \mathbb{R}^n \to \mathbb{R}_+$.

| $B_{\text{wc}}(L, \mathcal{P}_s, \mu, S)$ | Probability inequalities $L(x) = \mathbb{1}_{\mathbb{R}^n \setminus \Xi}(x)$ | Expectation inequalities $L(x) = \max_{i \in \mathcal{I}} \ a_i^\top x - b_i$ |
|---|---|---|
| Standard simplex $\mathcal{P}_n$ | [27] or Example 4.1 | [29] or Example 4.2 |
| Choquet star simplex | Section 2.3 | Section 2.4 |
| Unimodal $\mathcal{U}_\alpha$ | Cor. 4.1 or [24] | Cor. 4.2 |
| Monotone $\mathcal{M}_\gamma$ | Cor. 4.3 | Cor. 4.4 |

Table 1: Optimal inequalities described by problem ($P_{\text{wc}}$) and discussed in this paper.

### 1.3 Notation

We denote by $\mathbb{I}_n$ the identity matrix in $\mathbb{R}^{n \times n}$ and by $\mathbb{S}_+^n$ and $\mathbb{S}_{++}^n$ the sets of all positive semidefinite and positive definite symmetric matrices in $\mathbb{R}^{n \times n}$, respectively. For any matrix $A \in \mathbb{R}^{n \times n}$ we denote its pseudo-inverse with $A^\dagger$. The beta function, or Euler integral of the first kind, $B : \mathbb{R}_{++}^2 \to \mathbb{R}_{++}$ is defined as the integral

$$B(u, v) := \int_0^1 \lambda^{u-1} \cdot (1 - \lambda)^{v-1} \, \mathrm{d}\lambda.$$

The gamma function, or Euler integral of the second kind, $\Gamma : \mathbb{R} \to \mathbb{R}$ is defined as the integral

$$\Gamma(t) := \int_0^\infty \lambda^{t-1} \cdot \mathrm{e}^{-\lambda} \, \mathrm{d}\lambda.$$

For any set $S \subseteq \mathbb{R}^n$, we denote its associated indicator function by $\mathbb{1}_S : \mathbb{R}^n \to \{0, 1\}$, where $\mathbb{1}_S(x) = 1$ when $x \in S$ and zero otherwise. Similarly when $0 \in S$, its associated Minkowski or gauge function is denoted by $\kappa_S : \mathbb{R}^n \to \mathbb{R}_+$, where $\kappa_S(x) := \inf \{\lambda > 0 \mid x \in \lambda S\}$.

## 2 Expectation Inequalities for Structured Distributions

Throughout, we will consider worst-case expectation problems of the type

$$B_{\text{wc}}\left(\max_{i \in \mathcal{I}} \ f_i(A_i x), \mathcal{P}_n, \mu, S\right) = \sup_{\mathbb{P} \in \mathcal{P}_n} \ \mathbb{E}_{\mathbb{P}}\left\{\max_{i \in \mathcal{I}} \ f_i\left(A_i \xi\right)\right\} \quad \text{s.t.} \quad \mathbb{P} \in \mathcal{P}(\mu, S) \tag{$P$}$$

where each $f_i : \mathbb{R}^d \to \mathbb{R}$ and $A_i \in \mathbb{R}^{d\times n}$, and $\max_{i\in\mathcal{I}}\ f_i(A_i x) \geq 0$. It is clear that problem $(P)$ is a special case of the more general problem $(P_{\text{wc}})$ for a loss function $L(x) = \max_i [f_i(A_i x)]$ comprised of the maximum of functions $f_i$ of the $d$ dimensional projections $A_i\xi$.

Although the problem $(P)$ is by itself quite general, we will restrict our attention initially to problems with limited moment information only, absent any additional special structure in the distributions $\mathbb{P}$. In other words, we assume for the moment that the ambiguity set $\mathcal{P}_s$ corresponds to the standard probability simplex $\mathcal{P}_n$. In subsequent sections we will show how worst-case expectation problems over more restrictive convex ambiguity sets $\mathcal{P}_s \subseteq \mathcal{P}_n$ can be recast in the form $(P)$ via a transformation of the problem data.

### 2.1 The dual problem and known results

The problem $(P)$ is an infinite dimensional linear program (LP) over a convex set of distributions, and can be dualized yielding a finite dimensional linear semi-infinite program

$$
\begin{array}{l}
B_{\text{wc}}(L, \mathcal{P}_n, \mu, S) \leq \\
B^d_{\text{wc}}(L, \mathcal{P}_n, \mu, S) := \inf\limits_{(Y,y,y_0)} \quad \text{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \right\} \qquad (D) \\
\qquad\qquad\qquad \text{s.t.} \qquad Y \in \mathbb{S}^n,\ y \in \mathbb{R}^n,\ y_0 \in \mathbb{R} \\
\qquad\qquad\qquad\qquad\quad x^\top Y x + 2x^\top y + y_0 \geq L(x), \quad \forall x \in \mathbb{R}^n.
\end{array}
$$

Under standard and quite mild regularity assumptions, e.g. $\left[S, \mu; \mu^\top, 1\right] \in \mathbb{S}^{n+1}_{++}$, strong duality $B_{\text{wc}}(L, \mathcal{P}_n, \mu, S) = B^d_{\text{wc}}(L, \mathcal{P}_n, \mu, S)$ holds [22]. We will refer to both the primal problem $(P)$ and the dual problem $(D)$ as *worst-case expectation problems with second-order moment information.*

Note that the final constraint in problem $(D)$ is convex, since it represents an infinite collection of convex constraints in the variables $(Y, y, y_0)$, parametrized by $x$. Whether the problem $(D)$ can be solved conveniently or not is a separate matter, since it is not obvious for a general function $L$ how to cleanly eliminate the universal quantifier.

However, there are two important special cases in which problem $(D)$ can be converted into a standard-form semi-definite program (SDP). The first is when $L$ is the indicator function

of the intersection of polyhedral or elliptical sets, in which case $(D)$ was shown in [27] to be transformable to an SDP via application of the S-procedure. The result is a method for easily computing a (tight) worst-case bound on the probability of a random vector with known first and second moments falling outside of a convex set, which provides a multi-dimensional analog of the Chebyshev inequality[1].

The second is the case when $L$ is a convex piecewise affine function with a finite number of pieces, in which case it is again possible to convert the problem $(D)$ to an SDP using the procedure described in [29]. As in the previous case, the method described in [29] is applicable only to situations in which the first two moments of the uncertainty are known, but the ambiguity set is otherwise unstructured.

### 2.2 Expectation inequalities over a Choquet simplex

We have thus far described expectation bounds $B_{\text{wc}}(L, \mathcal{P}_n, \mu, S)$ over the standard probability simplex $\mathcal{P}_n$. The principle aim of this work, however, is to describe worst-case expectation problems over convex ambiguity sets $\mathcal{P}_s \subseteq \mathcal{P}_n$. We will argue that the worst-case expectation problem $(P)$ is in fact rich enough to handle worst-case expectation bounds $B_{\text{wc}}(L, \mathcal{P}_s, \mu, S)$ as well, with $\mathcal{P}_s \subseteq \mathcal{P}_n$ a convex ambiguity set. The main theoretical tool necessary to handle convex classes of probability distributions $\mathcal{P}_s \subseteq \mathcal{P}_n$ is their Choquet representation [18, 24].

**Definition 2.1** (Extreme distributions)**.** *A distribution $\mathbb{P} \in \mathcal{P}_s$ is said to be an extreme point of a convex ambiguity set $\mathcal{P}_s$ if it is not representable as a strict convex combination of two distinct distributions in $\mathcal{P}_s$. The set of all extreme points of $\mathcal{P}_s$ is denoted as $\operatorname{ex} \mathcal{P}_s$.*

**Definition 2.2** (Choquet representation)**.** *We say that an ambiguity set $\mathcal{P}_s$ admits a (unique) Choquet representation if*

$$\forall \mathbb{P} \in \mathcal{P}_s, \ \exists (!) \bar{\mathfrak{m}} : \quad \mathbb{P} = \int_{\mathcal{P}_s} \mathbb{Q} \ \bar{\mathfrak{m}}(\mathrm{d}\mathbb{Q})$$

*where $\bar{\mathfrak{m}} : \mathcal{B}(\mathcal{P}_s) \to [0, 1]$ is supported on $\operatorname{ex} \mathcal{P}_s$ and is referred to as the mixture representation of $\mathbb{P}$ over $\operatorname{ex} \mathcal{P}_s$.*

[1] Note that in the case of structured distributions, one can also derive a multidimensional analog to the Gauss inequality as shown in [24]. In contrast to the present work, which operates on the dual problem $(D)$, the bounds in [24] are produced by operating directly on the primal problem $(P)$.

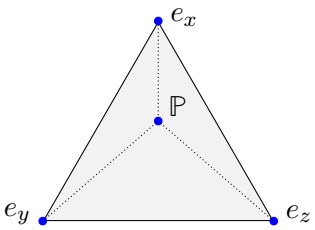



Figure 1: A simplicial ambiguity set $\mathcal{P}_s$ with extreme points $\operatorname{ex}\mathcal{P}_s = \{e_x, e_y, e_z\}$. Every distribution $\mathbb{P} \in \mathcal{P}_s$ has a unique mixture representation over $\operatorname{ex}\mathcal{P}_s$. For the depicted distribution we have the representation in term of the extreme points $\mathbb{P} = \frac{1}{2}e_x + \frac{1}{4}e_y + \frac{1}{4}e_z$.

The Choquet representation of a convex ambiguity set $\mathcal{P}_s$ will enable us to reduce the worst-case expectation problem ($P_{\text{wc}}$) over the ambiguity set $\mathcal{P}_s$ to a related worst-case expectation problem over the standard simplex $\mathcal{P}_n$. The existence or otherwise of a Choquet representation for an ambiguity set $\mathcal{P}_s$ is the topic of Choquet theory [17]. It can be shown that under the relatively mild assumption that $\operatorname{ex}\mathcal{P}_s$ is metrisable, convex and compact, such Choquet representations always exist. However, not all Choquet representable sets have necessarily a compact set of extreme points. Indeed, in Section 3 we will encounter sets of distributions with non-compact sets of extreme points which nevertheless admit a Choquet representation. It should also be remarked that when $\mathcal{P}_s$ is finite dimensional, the preceding statement is closely related to Minkowski's theorem stating that a compact convex set is the closed convex hull of its extreme points; see Figure 1.

In this paper, we will mainly encounter ambiguity sets $\mathcal{P}_s$ that enjoy the slightly stronger notion of Choquet star representability.

**Definition 2.3** (Choquet star representation)**.** *Suppose that $T$ is a distribution on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, and define a family of distributions $T_x$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that, for every $x \in \mathbb{R}^n$ and every $C \in \mathcal{B}(\mathbb{R}^n)$,*

$$T_x(C) = T(\{\lambda \geq 0 \mid \lambda x \in C\}).$$

*We say that the ambiguity set $\mathcal{P}_s$ admits a* Choquet star representation *if it admits a unique Choquet representation over*

$$\operatorname{ex}\mathcal{P}_s = \{T_x \mid \forall x \in \mathbb{R}^n\}. \tag{2}$$

*In this case we say that $\mathcal{P}_s$ is* generated by *$T$.*

Observe that in Definition 2.3 each distribution $T_x \in \operatorname{ex} \mathcal{P}_s$ is supported on the ray $\{\lambda x \mid \lambda \geq 0\}$. We will refer to distributions with support on rays emanating from the origin as *radial distributions*.

From Definition 2.3 it is evident that if $\mathcal{P}_s$ admits a Choquet star representation, then it is isomorphic to the standard probability simplex $\mathcal{P}_n$. We might therefore also refer to a Choquet star representable set as a Choquet star simplex. The extreme points of a Choquet star simplex $\mathcal{P}_s$ admit the spatial parametrization (2), which enables us to specialise Definition 2.2 to

$$\mathbb{P} \in \mathcal{P}_s \iff \exists! \bar{\mathbb{m}} : \ \mathbb{P} = \int_{\operatorname{ex} \mathcal{P}_s} \mathbb{Q} \ \bar{\mathbb{m}}(\mathrm{d}\mathbb{Q}) \iff \exists! \mathbb{m} \in \mathcal{P}_n : \ \mathbb{P} = \int_{\mathbb{R}^n} T_x \ \mathbb{m}(\mathrm{d}x). \tag{3}$$

Observe that the mixture representation $\bar{\mathbb{m}}$ in Definition 2.2 is a distribution on the set of distributions $\mathcal{P}_s$. With (3) many subtle problems arising from the need to endow $\mathcal{P}_s$ with a $\sigma$-algebra in which $\operatorname{ex} \mathcal{P}_s$ is measurable are circumvented. Indeed, the mixture representations $\mathbb{m}$ in (3) are elements of the standard probability simplex $\mathcal{P}_n$. In the context of Choquet star simplices, we will refer to both $\bar{\mathbb{m}}$ and $\mathbb{m}$ as mixture representations.

With this in mind, the power of Choquet star representable ambiguity sets becomes clear. We now show that a Choquet star representation of $\mathcal{P}_s$ can be utilized to remodel a structured problem in the form ($P_{\mathrm{wc}}$) as an equivalent unstructured problem (i.e. one with ambiguity set $\mathcal{P}_n$) via an appropriate transformation of the loss function and moments.

**Theorem 2.1.** *Assume that the ambiguity set $\mathcal{P}_s$ admits a Choquet star representation with generating distribution $T$, then*

$$B_{\mathrm{wc}}\left(L, \mathcal{P}_s, \mu, S\right) = B_{\mathrm{wc}}\left(L_s, \mathcal{P}_n, \mu_s, S_s\right) \tag{4}$$

*for $L_s(x) := \int_0^\infty L(\lambda x)\ T(\mathrm{d}\lambda)$, $S_s \cdot \int_0^\infty \lambda^2\ T(\mathrm{d}\lambda) = S$ and $\mu_s \cdot \int_0^\infty \lambda\ T(\mathrm{d}\lambda) = \mu$.*

*Proof.* Since the set $\mathcal{P}_s$ admits a Choquet star representation, we can optimize of the mixture representations $\mathbb{m}$ instead of $\mathbb{P}$. Indeed, using the reparametrization $\mathbb{P} = \int_{\mathbb{R}^n} T_y\, \mathbb{m}(\mathrm{d}y)$ we obtain

$$\begin{array}{rcl} \sup_{\mathbb{P}} \ \int_{\mathbb{R}^n} L(x)\, \mathbb{P}(\mathrm{d}x) & = & \sup_{\mathbb{m}} \ \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} L(\lambda)\, T_y(\mathrm{d}\lambda)\right] \mathbb{m}(\mathrm{d}y) \\ \text{s.t.} \ \ \mathbb{P} \in \mathcal{P}_s \cap \mathcal{P}(\mu, S) & & \text{s.t.} \ \ \int_{\mathbb{R}^n} T_y\, \mathbb{m}(\mathrm{d}y) \in \mathcal{P}(\mu, S). \end{array}$$

Indeed, we have that $\mathbb{P} = \int_{\mathbb{R}^n} T_y \, \mathbb{m}(\mathrm{d}y) \in \mathcal{P}_s \cap \mathcal{P}(\mu, S)$ is equivalent to $\int_{\mathbb{R}^n} T_y \, \mathbb{m}(\mathrm{d}y) \in \mathcal{P}(\mu, S)$. Furthermore, we have the identity

$$\int_{\mathbb{R}^n} [x^\top, 1]^\top \cdot [x^\top, 1] \; \mathbb{P}(\mathrm{d}x) = \int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^n} [x^\top, 1]^\top \cdot [x^\top, 1] \; T_y(\mathrm{d}x) \right] \; \mathbb{m}(\mathrm{d}y),$$

which equals using Fubini's Theorem and the Choquet star property of $T_y$

$$\int_{\mathbb{R}^n} [x^\top, 1]^\top \cdot [x^\top, 1] \; \mathbb{P}(\mathrm{d}x) = \int_{\mathbb{R}^n} \begin{pmatrix} \int_0^\infty \lambda^2 \, T(\mathrm{d}\lambda) \; y \cdot y^\top & \int_0^\infty \lambda \, T(\mathrm{d}\lambda) \; y \\ \int_0^\infty \lambda \, T(\mathrm{d}\lambda) \; y^\top & 1 \end{pmatrix} \; \mathbb{m}(\mathrm{d}y).$$

Hence $\mathbb{P} = \int_{\mathbb{R}^n} T_y \, \mathbb{m}(\mathrm{d}y) \in \mathcal{P}(\mu, S)$ is equivalent to $\mathbb{m} \in \mathcal{P}(\mu_s, S_s)$. We have lastly that the expectation $\mathbb{E}_{\mathbb{P}} \{L(\xi)\}$ for $\mathbb{P} = \int_{\mathbb{R}^n} T_y \, \mathbb{m}(\mathrm{d}y)$ equals the expectation $\mathbb{E}_{\mathbb{m}} \{L_s(\xi)\}$ where $L_s(y) := \mathbb{E}_{T_y} \{L(\xi)\} = \int_0^\infty L(\lambda y) \; T(\mathrm{d}\lambda)$ again using Fubini's Theorem concluding the proof. $\square$

Hence a worst-case expectation problem over a Choquet star simplex $\mathcal{P}_s$ can be reduced to an equivalent problem over the standard probability simplex $\mathcal{P}_n$. Both worst-case expectation problems are related in terms of their loss functions, since

$$L_s(x) = \mathbb{E}_{T_x} \{L(\xi)\} \tag{5}$$

according to the result presented in Theorem 2.1.

In the remainder of this section, we will discuss the transformation, via Theorem 2.1, of two important types of loss function $L$. These include indicator functions $L = \mathbb{1}_{\mathbb{R}^n \setminus \Xi}$ of polytopic sets $\Xi$ which arise in worst-case probability inequalities, and certain piecewise affine functions $L = \max_{i \in \mathcal{I}} \; a_i^\top x - b_i$ which come about when dealing with convex cost functions.

In section 4 we will then show how, for either type of loss function, the associated worst-case expectation bound $B_{\mathrm{wc}}(L_s, \mathcal{P}_n, \mu_s, S_s)$ amounts to a tractable SDP when the ambiguity set describes the set of all unimodal or monotone distributions.

### 2.3 Worst-case probability inequalities

We address first the problem of bounding the probability of the event $\xi \notin \Xi$ where $\Xi$ is an open convex polytope and $\mathbb{P} \in \mathcal{P}_s$ is a structured ambiguity set with known mean $\mu$ and second moment $S$. In this case we can use the standard identity between the probability of an event $\mathbb{P}(\xi \notin \Xi) = \mathbb{E}_{\mathbb{P}}\left\{\mathbb{1}_{\mathbb{R}^n \setminus \Xi}(\xi)\right\}$ and the expectation of its indicator function to state $\sup_{\mathbb{P} \in \mathcal{P}_s \cap \mathcal{P}(\mu,S)} \mathbb{P}(\xi \notin \Xi) = B_{\mathrm{wc}}(\mathbb{1}_{\mathbb{R}^n \setminus \Xi}, \mathcal{P}_s, \mu, S)$.

In what follows, we assume that the set $0 \in \Xi$ has a half-space representation in the form $\Xi := \left\{x \in \mathbb{R}^n \;\middle|\; a_i^\top x < b_i,\ \forall i \in \mathcal{I}\right\}$. It can be seen that the associated indicator function can be represented as the point-wise maximum of the indicator functions associated with the half-spaces from which the set $\Xi$ is composed, i.e.

$$L = \max_{i \in \mathcal{I}} \ \mathbb{1}_{a_i^\top x \geq b_i} = \mathbb{1}_{\mathbb{R}^n \setminus \Xi}. \tag{6}$$

The next proposition shows how to transform, via (5), such an indicator function for radial extreme distributions $T_y$ into a loss function $L_s$ for use in (4):

**Proposition 2.1.** *If the set $\mathcal{P}_s$ admits a Choquet star representation with generating distribution $T$, then*

$$L_s(y) = \mathbb{E}_{T_y}\left\{\max_{i \in \mathcal{I}} \ \mathbb{1}_{a_i^\top x \geq b_i}(\xi)\right\} = \max_{i \in \mathcal{I}} \ T\left(\left[b_i / a_i^\top y, \infty\right)\right) \cdot \mathbb{1}_{a_i^\top x \geq b_i}(y).$$

Hence, we have according to Theorem 2.1 that the worst-case probability problem over $\mathcal{P}_s$ can be reduced to an equivalent worst-case probability problem over the standard simplex $\mathcal{P}_n$

$$\begin{aligned}\sup_{\mathbb{P} \in \mathcal{P}_s \cap \mathcal{P}(\mu,S)} \mathbb{P}(\xi \notin \Xi) &= B_{\mathrm{wc}}\left(\mathbb{1}_{\mathbb{R}^n \setminus \Xi}, \mathcal{P}_s, \mu, S\right), \\ &= B_{\mathrm{wc}}\left(\max_{i \in \mathcal{I}} \ T\left(\left[b_i / a_i^\top y, \infty\right)\right) \cdot \mathbb{1}_{a_i^\top x \geq b_i}(y), \mathcal{P}_n, \mu_s, S_s\right).\end{aligned}$$

### 2.4 Worst-case expectation inequalities

As mentioned in Section 2, the worst-case expectation problem over the standard simplex with second-moment information is tractable when the loss function $L$ is in the form

$$L(x) = \max_{i \in \mathcal{I}} \; a_i^\top x - b_i \tag{7}$$

and thus convex. Because the set of all functions consisting of the point-wise maximum of affine functions coincides with the class of lower semi-continuous (l.s.c.) convex functions, the following fact is of interest.

**Fact 2.1.** *If the set $\mathcal{P}_s$ admits a Choquet star representation with generating distribution $T$ and $L$ is convex then*

$$L_s(x) = \mathbb{E}_{T_x}\left\{L(\xi)\right\} = \int_0^\infty L(\lambda x)\; T(\mathrm{d}\lambda)$$

*is convex as well.*

*Proof.* The statement can be proved almost immediately from the definition of convexity. For all $\theta \in [0,1]$

$$\begin{aligned} L_s(\theta a + (1-\theta)b) &= \int_0^\infty L(\lambda(\theta a + (1-\theta)b))\; T(\mathrm{d}\lambda) \\ &= \int_0^\infty L(\theta(\lambda a) + (1-\theta)(\lambda b))\; T(\mathrm{d}\lambda) \\ &\leq \int_0^\infty \theta L(\lambda a) + (1-\theta)L(\lambda b)\; T(\mathrm{d}\lambda) \\ &\leq \theta L_s(a) + (1-\theta)L_s(b) \end{aligned}$$

showing convexity of $L_s$. □

Despite the previous encouraging result, it is generally *not* the case that the function $L_s$ can be represented as the maximum of a *finite* number of affine functions when $L$ is in the form (7). Indeed, Fact 2.1 merely establishes that convexity is preserved, but does not otherwise address the structure of $L_s$.

Instead of considering convex piecewise linear loss functions $L$ as done in (7), we focus our

attention in what follows on loss functions in the form

$$L(x) = (\ell \circ \kappa_\Xi)(x) \tag{8}$$

where $\ell : \mathbb{R}_+ \to \mathbb{R}$ is a monotonically increasing function and $0 \in \Xi$ a convex set. Loss functions in the form (8) arise in distributionally robust optimisation [29, 30] and control [25, 26] when bounding the expected violation of a constraint $\xi \in \Xi$ using

$$\mathbb{E}_{\mathbb{P}}\{(\ell \circ \kappa_\Xi)(\xi)\} \leq \alpha, \quad \forall \mathbb{P} \in \mathcal{P}$$

as $L$ increases with decreasing proximity to the set $\Xi$. Moreover, the loss function (8) generalises the loss function (6) for $\ell = \mathbb{1}_{t\geq 1}$. The next proposition establishes that the structure of a loss function in the form (8) is preserved under the transformation (5) when $T_x$ are radial distributions.

**Proposition 2.2.** *If the set $\mathcal{P}_s$ admits a Choquet star representation with generating distribution $T$ and $L$ is in the form (8) with $0 \in \Xi = \left\{x \in \mathbb{R}^n \;\middle|\; a_i^\top x < b_i,\ \forall i \in \mathcal{I}\right\}$ then*

$$L_s(x) = \mathbb{E}_{T_x}\{L(\xi)\} = (\ell_s \circ \kappa_\Xi)(x) = \max_{i\in\mathcal{I}}\ \ell_s(a_i^\top x/b_i),$$

*with $\ell_s(t) := \int_0^\infty \ell(\lambda t)\ T(\mathrm{d}\lambda)$.*

*Proof.* We have the following chain of equalities proving the claim

$$\begin{aligned} L_s(x) = \mathbb{E}_{T_x}\{L(\xi)\} &= \int_0^\infty L(\lambda x)\ T(\mathrm{d}\lambda) = \int_0^\infty \ell\left(\kappa_\Xi(\lambda x)\right)\ T(\mathrm{d}\lambda) \\ &= \int_0^\infty \ell\left(\lambda \cdot \kappa_\Xi(x)\right)\ T(\mathrm{d}\lambda) \end{aligned}$$

where the last equality follows from the positive homogeneity of $\kappa_\Xi$. □

Hence, we have according to Theorem 2.1 that the worst-case expectation problem over $\mathcal{P}_s$ can be reduced to an equivalent worst-case probability problem over the standard simplex $\mathcal{P}_n$, i.e.

$$B_{\mathrm{wc}}(\ell \circ \kappa_\Xi, \mathcal{P}_s, \mu, S) = B_{\mathrm{wc}}(\max_{i\in\mathcal{I}}\ \ell_s(a_i^\top x/b_i), \mathcal{P}_n, \mu_s, S_s).$$

In the next section we discuss specific ambiguity sets $\mathcal{P}_s$ that admit Choquet star representations, with a focus on unimodal and monotone distributions. Both structural properties are shown to be closely related and their corresponding ambiguity sets admit Choquet star representations. We will then be able to exploit the particular structure of these ambiguity sets in combination with Propositions 2.1 and 2.2 to produce tractable optimization problems in Section 4 for the computation of the worst case bound (4) via the solution of its dual.

## 3 Unimodal and monotone distributions

We next identify two important classes of distributions that are amenable to Choquet representation. These include the family of *unimodal* (and more generally $\alpha$-*unimodal*) distributions – which have previously been employed to produce multi-dimensional generalizations to the Gauss and Chebyshev inequalities in [24] – and the related (and more restrictive) class of monotone (and more generally $\gamma$-*monotone*) distributions.

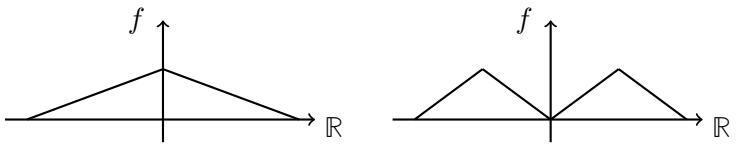
### 3.1 Unimodal distributions and their Choquet representations

A minimal structural property commonly encountered in practical situations is unimodality. Informally, a continuous probability distribution is unimodal if it has a centre $m$, 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 the remainder we adopt the following standard definition of unimodality; see e.g. Dharmadhikari and Joag-Dev [5].

**Definition 3.1** ($\alpha$-Unimodal distributions [5,15])**.** *For any fixed $\alpha \in \mathbb{R}_+$, a distribution $\mathbb{P} \in \mathcal{P}_n$ is called $\alpha$-unimodal with mode 0 if $t^\alpha \mathbb{P}(B/t)$ is non-decreasing in $t \in (0, \infty)$ for every Borel set $B \in \mathcal{B}(\mathbb{R}^n)$. The set of all $\alpha$-unimodal distributions with mode 0 is denoted as $\mathcal{U}_\alpha$.*

To develop an intuitive understanding of Definition 3.1, it is instructive to study the special case of continuous distributions. The density function of a continuous $\alpha$-unimodal distribution




(a) 1-unimodal distribution (b) 2-unimodal distribution

Figure 2: Univariate $\alpha$-unimodal probability distributions and their density functions.

may increase along rays, but the rate of increase is controlled by the parameter $\alpha$. We have that a distribution $\mathbb{P} \in \mathcal{P}_n$ with a continuous density function $f(x)$ is $\alpha$-unimodal about 0 if and only if $t^{n-\alpha} f(tx)$ is non-increasing in $t \in (0, \infty)$ for every fixed $x \neq 0$. This implies that if an $\alpha$-unimodal distribution on $\mathbb{R}^n$ has a continuous density function $f(x)$, then $f(x)$ does not grow faster than $\|x\|^{\alpha-n}$. In particular, for $\alpha = n$ the density is non-increasing along rays emanating from the origin. In this case, the notion of $\alpha$-unimodality coincides with star unimodality [5]. Hence $\alpha$ can be seen as a characterization of the degree of unimodality of a distribution; see Figure 2.

**Definition 3.2** (Radial $\alpha$-unimodal Distributions)**.** *For any $\alpha > 0$ and $x \in \mathbb{R}^n$ we denote by $u_x^\alpha$ the radial distribution supported on the line segment $[0, x] \subset \mathbb{R}^n$ with the property that*

$$u_x^\alpha([0, tx]) = \alpha \int_0^t \lambda^{\alpha-1} \, \mathrm{d}\lambda \quad \forall t \in [0, 1].$$

The importance of the radial distributions $u_x^\alpha$ is highlighted in the following theorem, stating that the set of radial unimodal distributions are the extreme points of the ambiguity set $\mathcal{U}_\alpha$:

**Theorem 3.1** ([5])**.** *The set $\mathcal{U}_\alpha$ admits a Choquet star representation of the form*

$$\forall \mathbb{P} \in \mathcal{U}_\alpha, \ \exists! \mathbb{m} \in \mathcal{P}_n : \quad \mathbb{P}(\cdot) = \int_{\mathbb{R}^n} u_x^\alpha(\cdot) \ \mathbb{m}(\mathrm{d}x).$$

*for the generating distribution $T([0, t]) = \alpha \int_0^t \lambda^{\alpha-1} \, \mathrm{d}\lambda$, $\forall t \in [0, 1]$.*

Theorem 3.1 asserts that every $\alpha$-unimodal distribution admits a unique Choquet star representation in terms of the extreme radial distributions $u_x^\alpha$. Thus, $\mathcal{U}_\alpha$ is a Choquet simplex over the set of radial $\alpha$-unimodal distributions.

The ambiguity sets $\mathcal{U}_\alpha$ enjoy the nesting property $\mathcal{U}_\alpha \subseteq U_\beta$ whenever $\alpha \leq \beta$. It easy to

verify that the radial distribution $u_x^\alpha$ converges weakly to the Dirac distribution $\delta_x$ as $\alpha$ tends to infinity. This allows us to conclude that the weak closure of $\cup_{\alpha \geq 0} \mathcal{U}_\alpha$ coincides with the standard simplex $\mathcal{P}_n$. Hence, the standard simplex $\mathcal{P}_n$ is included as the limit of the hierarchy of $\alpha$-unimodal ambiguity sets $\mathcal{U}_\alpha$ for $\alpha$ tending to infinity.

### 3.2 Monotone distributions and their Choquet representations

A structural property which is closely related to unimodality is monotonicity. Where unimodality requires intuitively that the density function of a continuous distribution should be decreasing with increasing distance from the mode, monotonicity additionally requires that this decrease is *smooth*. Indeed, monotonicity is often used in mathematics to model the notion of *smoothness* of a distribution [16].

In the remainder we adopt the following standard definitions of monotonicity [16] of distributions, which are inspired on the notion of monotone functions.

**Definition 3.3** ($\gamma$-monotone functions)**.** *A univariate function $f : \mathbb{R}_+ \to \mathbb{R}$ is denoted as $\gamma$-monotone if it is $\gamma$ times differentiable and*

$$(-1)^k f^{(k)}(t) \geq 0, \quad \forall t > 0,\ k \in \{0, \ldots, \gamma\}.$$

**Definition 3.4** ($\gamma$-Monotone distributions[2])**.** *For any $1 \leq \gamma \in \mathbb{N}$, a distribution $\mathbb{P}$ is called $\gamma$-monotone with mode 0 if $t^{\gamma+n-1}\mathbb{P}(B/t)$ is $\gamma$-monotone in $t \in (0, \infty)$ for every Borel set $B \in \mathcal{B}(\mathbb{R}^n)$. The set of all $\gamma$-monotone distributions with mode 0 is denoted as $\mathcal{M}_\gamma$.*

Again, it is instructive to consider the case of continuous distributions. We have that a continuous distribution $\mathbb{P}$ is $\gamma$-monotone if and only if its density function $f(tx)$ is $\gamma$-monotone in $t \in (0, \infty)$ for every fixed $x$ [2]. This means that if a $\gamma$-monotone distribution $\mathbb{P}$ admits a continuous density $f$, then $f$ is $\gamma$-monotone along rays emanating from the mode. Hence $\gamma$ can be seen as a characterization of how smooth the distribution is, see also Figure 3.

**Definition 3.5** (Radial $\gamma$-monotone distributions)**.** *For any $\gamma \in \mathbb{N}_0$ and $x \in \mathbb{R}^n$ we denote by*

[2]The class of $\gamma$-monotone distributions defined here can be identified with the class of $(n, \gamma)$-unimodal distributions discussed in [2, Theorem 3.1.14].


(a) 1-monotone distribution (b) 2-monotone distribution

Figure 3: Univariate $\gamma$-monotone probability distributions and their density functions.

*$m_x^\gamma$ the radial distribution supported on the line segment $[0, x] \subset \mathbb{R}^n$ with the property that*

$$m_x^\gamma\left([0, tx]\right) = \frac{1}{B(n,\gamma)} \cdot \int_0^t \lambda^{n-1} \cdot (1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda \quad \forall t \in [0,1].$$

The importance of the radial distributions $m_x^\gamma$ is highlighted in the following theorem, stating that the set of radial monotone distributions are the extreme points of $\mathcal{M}_\gamma$.

**Theorem 3.2** ([2])**.** *The set $\mathcal{M}_\gamma$ admits a Choquet star representation of the form*

$$\forall \mathbb{P} \in \mathcal{M}_\gamma, \ \exists! \mathbb{m} \in \mathcal{P}_n : \quad \mathbb{P}(\cdot) = \int_{\mathbb{R}^n} m_x^\gamma(\cdot) \ \mathbb{m}(\mathrm{d}x).$$

*for the generating distribution $T([0,t]) = B^{-1}(n,\gamma) \cdot \int_0^t \lambda^{n-1} \cdot (1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda$, $\forall t \in [0,1]$.*

Theorem 3.2 asserts that every $\gamma$-monotone distribution admits a unique Choquet star representation in term of the extreme radial monotone distributions $m_x^\gamma$. Thus, $\mathcal{M}_\gamma$ is a Choquet simplex over the set of radial $\gamma$-monotone distributions.

The ambiguity sets $\mathcal{M}_\gamma$ enjoy the nesting property $\mathcal{M}_\delta \subseteq \mathcal{M}_\gamma$ whenever $\gamma \leq \delta$. Historically, a distribution in $\cap_{\gamma \in \mathbb{N}_0} \mathcal{M}_\gamma$ has been denoted as a completely monotone distribution [1]. It easy to verify that the sequence $m_{x\gamma}^\gamma$ of radial monotone distributions converges weakly when $\gamma$ tends to infinity to a radial distribution $m_x$ supported on the ray $\{\lambda x \mid \lambda \in \mathbb{R}_+\}$ with the property

$$m_x\left([0, tx]\right) = \frac{1}{\Gamma(n)} \int_0^t \lambda^{n-1} \mathrm{e}^{-\lambda} \, \mathrm{d}\lambda \quad \forall t \in [0, \infty).$$

In the light of this observation, Theorem 3.2 reduces in the case of univariate completely monotone distributions to Bernstein's representation theorem [1]. Hence, the set of completely monotone distributions is included as the limit of the hierarchy of $\gamma$-monotone ambiguity sets $\mathcal{M}_\gamma$ for $\gamma$ tending to infinity.

## 4 Structured expectation inequalities as semidefinite programs

Having identified two classes of structured distributions amenable to Choquet representation, we are now free in principle to apply our approach to solving the dual problem $(D)$. Specifically, given the generating distributions $T$ identified in either Theorem 3.1 (for unimodal distributions) or Theorem 3.2 (for monotone distributions) and either an indicator or convex piecewise affine (PWA) function $L$, we can apply the appropriate transformations from Section 2.2 to transform $(L, S, \mu) \mapsto (L_s, S_s, \mu_s)$ and then solve the dual problem $(D)$ with this new data.

However, as a practical matter this remains problematic, since our transformed function $L_s$ will be neither an indicator function nor PWA. In order to circumvent this difficulty we require the following result, which transforms the semi-infinite constraints over $\mathbb{R}^n$ in the dual problem $(D)$ to an equivalent semi-infinite constraint over $\mathbb{R}^d$.

**Theorem 4.1.** *The worst-case expectation problem with second-order moment information* $(D)$ *can be reformulated as* $B^d_{\mathrm{wc}}\left(\max_{i\in\mathcal{I}}\ f_i(A_i x), \mathcal{P}_n, \mu, S\right) =$

$$
\begin{array}{ll}
\inf & \operatorname{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix} \right\} \\
\text{s.t.} & \left.\begin{array}{l} \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \in \mathbb{S}_+^{n+1}, \quad \begin{pmatrix} T_{1,i} & T_{2,i} \\ T_{2,i}^\top & T_{3,i} \end{pmatrix} \in \mathbb{S}_+^{d+1}, \quad \Lambda_{1,i} \in \mathbb{R}^{d\times d}, \quad \Lambda_{2,i} \in \mathbb{R}^d \\ \begin{pmatrix} \Lambda_{1,i} + \Lambda_{1,i}^\top - T_{1,i} & \Lambda_{2,i} - T_{2,i} & -\Lambda_{1,i}^\top A_i \\ \Lambda_{2,i}^\top - T_{2,i}^\top & y_0 - T_{3,i} & y^\top - \Lambda_{2,i}^\top A_i \\ -A_i^\top \Lambda_{1,i} & y - A_i^\top \Lambda_{2,i} & Y \end{pmatrix} \succeq 0, \qquad \forall i \in \mathcal{I} \end{array} \right\} \qquad (\mathcal{C}_1) \\
& T_{3,i} + 2q^\top T_{2,i} + q^\top T_{1,i} q \geq f_i(q), \quad \forall q \in \mathbb{R}^d, \quad \forall i \in \mathcal{I} \qquad (\mathcal{C}_2)
\end{array}
$$

*Proof.* The constraint in the dual problem $(D)$ can be reformulated as

$$
\forall i \in \mathcal{I}, \quad \forall q \in \mathbb{R}^d : \quad \inf_{A_i x = q}\ x^\top Y x + 2x^\top y + y_0 \ \geq\ f_i(q).
$$

As we assume throughout that $\max_{i\in\mathcal{I}}\ f_i(A_i x) \geq 0$, it must hence follow that $Y$ is positive semidefinite and $x^\top Y x + 2x^\top y$ is bounded from below. The claim now follows immediately from

Theorem A.1 applied to the parametric optimization problem $\inf_{A_i x = q} \; x^\top Y x + 2x^\top y + y_0$. $\square$

Note that this reformulation of the standard dual problem $(D)$ into the more unconventional form in Theorem 4.1 is motivated by a desire to replace the semi-infinite constraint over $\mathbb{R}^n$ with one over $\mathbb{R}^d$. Hence when $d \ll n$, the reformulation offered by Theorem 4.1 is preferable to the standard dual $(D)$. It is well known that the semi-infinite constraint in $\mathbb{R}^d$ of Theorem 4.1 for piece-wise polynomial $f_i$ admits a tractable reformulation in the univariate case when $d = 1$, or when the functions $f_i$ are quadratically representable. In particular, when the functions $f_i$ are univariate piecewise polynomial, the semi-infinite constraints in Theorem 4.1 are known to admit tractable linear matrix inequality (LMI) reformulations based on exact sum-of-squares (SOS) representations [12].

### 4.1 Unstructured distributions

Before considering the structured classes of distributions identified in Section 3, it is instructive to apply Theorem 4.1 to the unstructured case $\mathcal{P}_s = \mathcal{P}_n$, and restate two well-known results for problems with univariate $f_i$ in terms of Theorem 4.1.

When the functions $f_i = \mathbb{1}_{a_i^\top x \geq b_i}$ are indicator functions, problem $(P)$ describes the worst-case probability of $\xi \notin \Xi = \left\{ x \in \mathbb{R}^n \;\middle|\; a_i^\top x < b_i, \; \forall i \in \mathcal{I} \right\}$ as found in [27]:

**Example 4.1** (Vandenberghe et al. [27])**.** *Suppose that* $\Xi = \left\{ x \in \mathbb{R}^n \;\middle|\; a_i^\top x < b_i, \; \forall i \in \mathcal{I} \right\}$ *and define a loss function* $L = \mathbb{1}_{\mathbb{R}^n \setminus \Xi}$*. Then the worst-case probability problem for the event* $\xi \notin \Xi$ *can be modelled as in Theorem 4.1. The constraint* $(\mathcal{C}_2)$ *becomes*

$$T_{3,i} - 1 + 2q^\top T_{2,i} + q^\top T_{1,i} q \geq 0, \quad \forall q \geq b_i, \quad \forall i \in \mathcal{I},$$

*which can be rewritten using a LMI representation. The worst-case probability problem is there-*

*fore equivalent to an SDP, i.e.*

$$\begin{aligned} B_{\mathrm{wc}}(L, \mathcal{P}_n, \mu, S) = \inf \quad & \mathrm{Tr}\left\{YS\right\} + 2y^\top \mu + y_0 \\ \text{s.t.} \quad & (\mathcal{C}_1), \\ & \exists \tau_i \in \mathbb{R}_+, \ \begin{pmatrix} T_{1,i} & T_{2,i} \\ T_{2,i} & T_{3,i} - 1 \end{pmatrix} \succeq \tau_i \begin{pmatrix} 0 & 1 \\ 1 & -2b_i \end{pmatrix}, \ \forall i \in \mathcal{I}. \end{aligned}$$

Similarly, for affine functions $f_i$, problem $(P)$ describes the worst-case expectation problem for convex piecewise affine loss functions, as can be found in [29]:

**Example 4.2** (Zymler et al. [29])**.** *For a piecewise affine loss function $L(x) = \max_{i\in\mathcal{I}} \ a_i^\top x - b_i$, the constraint $(\mathcal{C}_2)$ in Theorem 4.1 becomes*

$$T_{3,i} + 2q^\top T_{2,i} + q^\top T_{3,i} q \geq q - b_i, \quad \forall q \in \mathbb{R}, \quad \forall i \in \mathcal{I},$$

*which can be rewritten using an LMI representation. The worst-case expectation problem for the piece-wise affine loss function $L$ is therefore equivalent to an SDP, i.e.*

$$\begin{aligned} B_{\mathrm{wc}}(L, \mathcal{P}_n, \mu, S) = \inf \quad & \mathrm{Tr}\left\{YS\right\} + 2y^\top \mu + y_0 \\ \text{s.t.} \quad & (\mathcal{C}_1), \\ & \begin{pmatrix} T_{1,i} & T_{2,i} - \frac{1}{2} \\ T_{2,i} - \frac{1}{2} & T_{3,i} + b_i \end{pmatrix} \succeq 0, \quad \forall i \in \mathcal{I}. \end{aligned}$$

### 4.2 Unimodal distributions

We can now make the abstract result of Theorem 2.1 concrete and explicitly state SDP reformulations of a worst-case probability problem for a polytopic set $\Xi$ and a worst-case expectation problem for a loss function $L(x) = (\kappa_\Xi(x) - 1)^+$ for the case of unimodal ambiguity sets.

Specifically, our method is as follows: Theorem 3.1 provides us with the appropriate generating distribution $T$ for $\alpha$-unimodal distributions. We then use this generating distribution to transform $(L, \mu, S) \mapsto (L_s, \mu_\alpha, \mu_\alpha)$ via Theorem 2.1, where the mapping $L \mapsto L_s$ in particu-

lar is supplied by either Proposition 2.1 or Proposition 2.2. This produces a transformed loss function in the form $L_s(x) = \max_i f_i(A_i x)$ for some univariate functions $f_i$. Finally, we apply Theorem 4.1 and identify the appropriate expression for the constraint $\mathcal{C}_2$ for our particular functions $f_i$.

**Corollary 4.1** ($\alpha$-Unimodal probability inequalities)**.** *For any rational* $0 \leq \alpha = \frac{v}{w}$*, with* $(v, w) \in \mathbb{N}$ *and* $0 \in \Xi$ *we have the equality* $B_{\text{wc}}(\mathbb{1}_{\mathbb{R}^n \setminus \Xi}, \mathcal{U}_\alpha, \mu, S) =$

$$\begin{aligned} \inf \quad & \text{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S_\alpha & \mu_\alpha \\ \mu_\alpha^\top & 1 \end{pmatrix} \right\} \\ \text{s.t.} \quad & (\mathcal{C}_1), \\ & q^{2w+v} b_i^2 T_{1,i} + 2q^{w+v} b_i T_{2,i} + q^v (T_{3,i} - 1) + 1 \geq 0, \quad \forall q \geq 0 \end{aligned}$$

*where* $S_\alpha \succ \mu_\alpha \mu_\alpha^\top$ *for* $S_\alpha = \frac{\alpha+2}{\alpha} S$ *and* $\mu_\alpha = \frac{\alpha+1}{\alpha} \mu$*.*

We remark here that Corollary 4.1 generalizes the results presented in [24] as it no longer matters that $\alpha \geq 1$. However where the result in [24] follows from a direct reformulation of the primal problem $(P_{\text{wc}})$, the result in corollary 4.1 is hinges on the dual problem $(D)$ to be strong. Strong duality calls for the additional Slater type condition $S_\alpha \succ \mu_\alpha \mu_\alpha^\top$ to hold, which explains the strict inequality in the corollaries stated hereafter. The proofs of the corollaries presented in this section are deferred to Appendix B.

**Corollary 4.2** ($\alpha$-Unimodal expectation inequalities)**.** *For any rational* $0 \leq \alpha = \frac{v}{w} \in \mathbb{Q}$*, with* $v$*,* $w \in \mathbb{N}$*, we have the equality* $B_{\text{wc}}(\max\{0, \kappa_\Xi(x) - 1\}, \mathcal{U}_\alpha, \mu, S) =$

$$\begin{aligned} \inf \quad & \text{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S_\alpha & \mu_\alpha \\ \mu_\alpha^\top & 1 \end{pmatrix} \right\} \\ \text{s.t.} \quad & q^{2w+v}\, b_i^2 T_{1,i} + q^{w+v} \left( 2 b_i T_{2,i} - \frac{\alpha}{\alpha+1} \right) + q^v \left(1 + T_{3,i}\right) - \frac{1}{\alpha+1} \geq 0, \quad \forall q \geq 1 \end{aligned}$$

*where* $S_\alpha \succ \mu_\alpha \mu_\alpha^\top$ *for* $S_\alpha = \frac{\alpha+2}{\alpha} S$ *and* $\mu_\alpha = \frac{\alpha+1}{\alpha} \mu$*.*

### 4.3 Monotone distributions

We now can also make the abstract results of Theorem 2.1 concrete and explicitly state the SDP reformulations of a worst-case probability problem for a polytopic set $\Xi$ and a worst-case expectation problem for a loss function $L(x) = (\kappa_\Xi(x) - 1)^+$ for the case of monotone ambiguity sets.

Our approach is identical to that in Section 4.2, except that we now look to Theorem 3.2 to provides us with the appropriate generating distribution $T$ for $\gamma$-monotone distributions.

**Corollary 4.3** ($\gamma$-Monotone probability inequalities)**.** *For any $\gamma \in \mathbb{N}_0$ we have the equality* $B_{\mathrm{wc}}(\mathbb{1}_{\mathbb{R}^n \setminus \Xi}, \mathcal{M}_\gamma, \mu, S) =$

$$
\begin{array}{ll}
\inf & \mathrm{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S_\gamma & \mu_\gamma \\ \mu_\gamma^\top & 1 \end{pmatrix} \right\} \\
\text{s.t.} & (\mathcal{C}_1), \\
& T_{1,i} b_i^2 q^{n+\gamma+1} + 2 b_i T_{2,i} q^{n+\gamma} + (T_{3,i} - 1)\, q^{n+\gamma-1} + \\
& \qquad \frac{1}{B(n,\gamma)} \sum_{k=0}^{\gamma-1} \frac{(-1)^k}{n+k} \binom{\gamma-1}{k} q^{\gamma-k-1} \geq 0, \quad \forall q \geq 1
\end{array}
$$

*where $S_\gamma \succ \mu_\gamma \mu_\gamma^\top$ for $S_\gamma = \frac{n+\gamma}{n} \frac{n+\gamma+1}{n+1} S$ and $\mu_\gamma = \frac{n+\gamma}{n} \mu$.*

**Corollary 4.4** ($\gamma$-Monotone expectation inequalities)**.** *For any $\gamma \in \mathbb{N}_0$ we have the equality* $B_{\mathrm{wc}}(\max\{0, \kappa_\Xi(x) - 1\}, \mathcal{M}_\gamma, \mu, S) =$

$$
\begin{array}{ll}
\inf & \mathrm{Tr}\left\{ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} \cdot \begin{pmatrix} S_\gamma & \mu_\gamma \\ \mu_\gamma^\top & 1 \end{pmatrix} \right\} \\
\text{s.t.} & (\mathcal{C}_1), \\
& T_{1,i} b_i^2 q^{n+\gamma+1} + \left( 2 b_i T_{2,i} - \frac{n}{n+\gamma} \right) q^{n+\gamma} + (T_{3,i} + 1)\, q^{n+\gamma-1} - \\
& \qquad \frac{1}{B(n,\gamma)} \sum_{k=0}^{\gamma-1} \frac{(-1)^k}{(n+k)(n+k+1)} \binom{\gamma-1}{k} q^{\gamma-k-1} \geq 0, \quad \forall q \geq 1
\end{array}
$$

*where $S_\gamma \succ \mu_\gamma \mu_\gamma^\top$ for $S_\gamma = \frac{n+\gamma}{n} \frac{n+\gamma+1}{n+1} S$ and $\mu_\gamma = \frac{n+\gamma}{n} \mu$.*

As mentioned in the beginning of this section, the polynomial inequalities appearing in Corollaries 4.1 to 4.4 admit exact SDP representations [12]. Standard software tools, such as `YALMIP` [10], are available which implement this transformation automatically. We do not state the resulting SDP constraints explicitly as they offer no further insight and would only clutter the statement of previous corollaries further.

## 5 Numerical examples

We illustrate the optimal inequalities presented in this paper by bounding the value of European stock portfolios [3] and by computing worst-case bounds when aggregating random variables with known marginal information [6]. The resulting SDP problems are implemented in `Matlab` using the interface `YALMIP`, and solved numerically using `SDPT3`.

### 5.1 Optimal pricing of stock portfolios

In this example we are interested in finding an upper bound on the price of a European stock option [3] with random pay-off

$$\Phi(\xi) := \max\{0,\ a^\top \xi - k\} = k\left(\kappa_\Xi(\xi) - 1\right)^+,$$

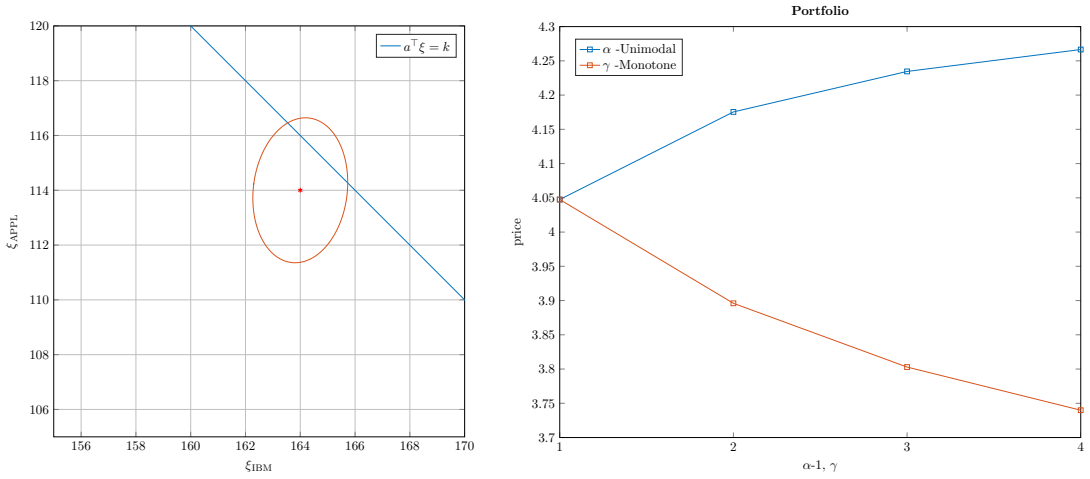
for $\Xi = \left\{x \in \mathbb{R}^n \ \middle|\ a^\top x \leq k\right\}$. This option allows its holder to buy a portfolio $a \in \mathbb{R}^n$ of stocks at a price $k \in \mathbb{R}_+$ at maturity. The payoff $\Phi$ is hence positive if the uncertain value $\xi \in \mathbb{R}^n$ of the stocks at maturity in the portfolio $a \in \mathbb{R}^n$ exceeds the negotiated price $k \in \mathbb{R}_+$. If the price of portfolio of stocks $a^\top \xi$ in the market at maturity is less then $k$, then the holder will not exercise his right to buy the stock portfolio at price $k$.

When we denote with $\mathbb{P}^\star$ the distribution of $\xi$, then for the issuer of the option it is of interest to know

$$p := \sup_{\mathbb{P} \in \mathcal{P}} \ \mathbb{E}_{\mathbb{P}}\left\{\Phi(\xi)\right\}$$

for $\mathcal{P}$ a set of distributions for which the option issuer is convinced that $\mathbb{P}^\star \in \mathcal{P}$. Indeed, the issuer would like to demand a price of the stock option buyer which exceeds $p$, as in this case he or she is convinced that on average a profit is made.

In the remainder of this section, we assume that our portfolio $\xi = (\xi_{\texttt{IBM}}, \xi_{\texttt{APPLE}})$ consists of




(a) (b)

Figure 4: Optimal pricing of a portfolio containing an equal amount of $\texttt{IBM}^{\text{TM}}$ and $\texttt{APPLE}^{\text{TM}}$ stocks. Figure 4(a) indicates the distribution of $(\xi_{\texttt{IBM}}, \xi_{\texttt{APPLE}})$ visually. The red half line indicates realizations beyond which a profit is made.

$a = (1,1)^\top$ an equal part of $\texttt{IBM}^{\text{TM}}$ and $\texttt{APPLE}^{\text{TM}}$ stocks. The stock holder is convinced that the distribution of $\xi$ satisfies

$$\mathbb{P}^\star \in \mathcal{P}\left(\begin{pmatrix}164\\114\end{pmatrix}, \begin{pmatrix}20 & 5\\5 & 60\end{pmatrix} + \begin{pmatrix}164\\114\end{pmatrix}\begin{pmatrix}164\\114\end{pmatrix}^\top\right)$$

for a strike price at maturity $k = 280$. This situation is sketched in Figure 4(a). The stock holder is also convinced that the distribution of $\xi$ should be well-behaved and has a mode which coincides with its mean. In Figure 4(b), the optimal price $p$ is given when the stock holder believes that either $\mathbb{P}^\star \in \mathcal{M}_\gamma$ or $\mathbb{P}^\star \in \mathcal{U}_\alpha$ in function of $\gamma \in \{1,\ldots,5\}$ and $\alpha \in \{2,\ldots,6\}$. As remarked before the bounds converge to either the bounds for arbitrary distributions when $\alpha \to \infty$ or completely monotone distributions in case $\gamma \to \infty$.

### 5.2 Factor models in insurance

Insurance companies most commonly model the size of claims $\xi_i$ incurred as a result of different types of insurance policies separately from another [6]. The claims $\xi_i$ factor the total claim

$$S_d := \textstyle\sum_{i=1}^d \xi_i$$

| CHF | Average $\mu_i$ | Standard deviation $\sigma_i$ |
|---|---|---|
| Car insurance | 15.000 | 2.000 |
| Life insurance | 7.000 | 1.000 |
| Fire insurance | 3.000 | 5.000 |
| Medical insurance | 20.000 | 2.000 |

Table 2: Marginal means and standard deviations of the size of the claims incurred by the four types of insurance policies in the portfolio.

as a sum of $d$ separate claims $\xi_i$ without a specified dependence structure. The problem of quantifying a certain statistic of $L(S_d)$ for a given loss function $L$ based on (partial) marginal information of the distributions of the factors $\xi_i$ is known as a Fréchet problem [20].

We consider a portfolio containing four types of insurance policies, i.e. car, life, fire and medical insurances. We will assume that only information on the means $\mu_i := \mathbb{E}_{\mathbb{P}}\{\xi_i\}$ and second moments $s_i^2 := \mathbb{E}_{\mathbb{P}}\left\{\xi_i^2\right\}$ of the size of the corresponding insurance claims are given. Suppose we are interested in large aggregate claims $S_d$ occurring with probability at most $\alpha = 5\%$, where that part of the claim $S_d$ exceeding the threshold $k = 150.000$ CHF is covered by a reinsurer. In what follows we therefore consider the problem of quantifying the least upper bound on the conditional value at risk $\mathrm{CVaR}_\alpha\left(L(S_d)\right)$, where

$$L(S_d) = \min\left(\max\left(S_d, 0\right), k\right)$$

using only the marginal means $\mu_i$ and second moments $s_i^2$ as given in Table 2. Additionally, it is assumed that the joint probability distribution $\mathbb{P}$ of $(\xi_1, \ldots, \xi_4)$ is star unimodal. The worst-case CVaR problem can be reduced to a worst-case expectation problem as indicated in Section 1.1 using the golden search method for the outer minimization problem over $\beta \in [0, k]$. The corresponding transformed loss function $L_s$ according to Theorem 2.1 is given in Appendix C. The worst-case excepted aggregate claim above the 5th percentile, i.e. $\mathrm{CVaR}_\alpha\left(L(S_d)\right)$, was numerically determined to be 123.325 CHF in approximately 15 seconds using `Matlab` on a PC[3] operated by `Debian GNU/Linux 7 (wheezy)`.

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

## 6 Conclusion

This paper provides a new perspective on the computational solution of worst-case expectation problems for convex classes of distributions with second-order moment information. We show that worst-case expectation inequalities over a Choquet simplex can in several interesting cases be reduced to worst-case expectation inequalities over the standard probability simplex. We focus in particular on the set of all unimodal distributions and the set of all monotone distributions. We illustrate the power of this perspective by pointing out that all known previous results concerning worst-case probability bounds with second-order moment information can be reduced to special cases of the results stated in this work. Moreover, we present how our results can be used to compute optimal bounds on the worst-case expectation and CVaR of certain loss functions when the true but unknown distribution is known to be either unimodal or monotone. We illustrate our methods by considering an option pricing problem for European stock options and an insurance risk aggregation problem with marginal information.

## A Equality constrained quadratic programs (QPs).

We will state here a relevant result concerning equality constrained QPs used throughout the rest of this paper. Assume we define a function $I : \mathbb{R}^d \to \mathbb{R}$ as follows

$$\begin{aligned} I(b) := \min_{x \in \mathbb{R}^n} \quad & x^\top G x + 2x^\top c + y \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

with $A \in \mathbb{R}^{d \times n}$ having full row rank and $G$ positive semidefinite. It is assumed that the function $x^\top G x + 2x^\top c$ in bounded from below such that $I(b) > \infty$. We can now represent the quadratic function $I$ using a dual representation as indicated in the following theorem.

**Theorem A.1** (Parametric representation of $I$)**.** *The function $I$ is lower bounded by*

$$I(b) \geq b^\top T_1 b + 2b^\top T_2 + T_3 \tag{9}$$

*for all* $T_1 \in \mathbb{S}^d$, $T_2 \in \mathbb{R}^d$ *and* $T_3 \in \mathbb{R}$ *such that there exist* $\Lambda_1 \in \mathbb{R}^{d\times d}$, $\Lambda_2 \in \mathbb{R}^d$ *with*

$$\begin{pmatrix} \Lambda_1 + \Lambda_1^\top - T_1 & \Lambda_2 - T_2 & -\Lambda_1^\top A \\ \Lambda_2^\top - T_2^\top & y - T_3 & c^\top - \Lambda_2^\top A \\ -A^\top \Lambda_1 & c - A^\top \Lambda_2 & G \end{pmatrix} \succeq 0. \tag{10}$$

*Moreover, inequality* (9) *is tight uniformly in* $b \in \mathbb{R}^d$ *for some* $T_1$, $T_2$ *and* $T_3$ *satisfying condition* (10).

*Proof.* The Lagrangian of the optimization problem defining $I(b)$ is given as

$$\mathcal{L}(x, \lambda) := x^\top G x + 2x^\top \left(c + A^\top \lambda\right) - 2\lambda^\top b + y.$$

As $x^\top G x + 2x^\top c$ is bounded from below on $\mathbb{R}^n$, we have that for all $b \in \mathbb{R}^d$ there exists a minimizer $x^\star$ such that $I(b) = (x^\star)^\top G x^\star + 2(x^\star)^\top c + y$ and $Ax^\star = b$. From the first order optimality conditions for convex QPs [14, Lemma 16.1], we have that $\min_x \max_\lambda \ \mathcal{L}(x, \lambda) = \mathcal{L}(x^\star, \lambda^\star) = \max_\lambda \min_x \ \mathcal{L}(x, \lambda)$ where the saddle point $(x^\star, \lambda^\star)$ is any solution of the linear system

$$\begin{pmatrix} G & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^\star \\ \lambda^\star \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}. \tag{11}$$

The quadratic optimization problem $\max_x \mathcal{L}(x, \lambda^\star)$ admits a maximizer if and only if $(c + A^\top \lambda^\star)$ is in the range of $G$. It must thus hold that

$$\left(\mathbb{I}_d - GG^\dagger\right)\left(c + A^\top \lambda^\star\right) = 0. \tag{12}$$

Hence when dualizing the problem defining $I(b)$, we get its dual representation $I(b) = \max_\lambda \ -\left(c + A^\top \lambda\right)^\top G^\dagger \left(c + A^\top \lambda\right) - 2\lambda^\top b + y$. From equation (11) it follows that $\lambda^\star$ is any solution of the linear equation $b + AG^\dagger A^\top \lambda^\star + AG^\dagger c = 0$. Therefore there exists an affine $\lambda^\star(b) = -\Lambda_1^\star b - \Lambda_2^\star$ with $\Lambda_1^\star \in \mathbb{R}^{d\times d}$ and $\Lambda_2^\star \in \mathbb{R}^d$ such that

$$I(b) = -\left(c - A^\top \Lambda_1^\star b - A^\top \Lambda_2^\star\right)^\top G^\dagger \left(c - A^\top \Lambda_1^\star b - A^\top \Lambda_2^\star\right) + 2b^\top \Lambda_1^{\star\top} b + 2\Lambda_2^{\star\top} b + y. \tag{13}$$

From equation (12) it follows that for all $b$ in $\mathbb{R}^d$ it holds that $\left(\mathbb{I}_d - GG^\dagger\right)\left(c - A^\top \Lambda_1^\star b - A^\top \Lambda_2^\star\right) = 0$.

We must hence also have that

$$\left(\mathbb{I}_d - GG^\dagger\right)\left(-A^\top\Lambda_1^\star,\, c - A^\top\Lambda_2^\star\right) = 0 \tag{14}$$

The dual reprenstation of $I(b)$ guarantees that for all $\lambda(b) = -\Lambda_1 b - \Lambda_2$ with $\Lambda_1 \in \mathbb{R}^{d\times d}$ and $\Lambda_2 \in \mathbb{R}^d$

$$I(b) \geq -\left(c - A^\top\Lambda_1 b - A^\top\Lambda_2\right)^\top G^\dagger\left(c - A^\top\Lambda_1 b - A^\top\Lambda_2\right) + 2b^\top\Lambda_1^\top b + 2\Lambda_2^\top b$$

Lower bounding the right hand side of the previous inequality with $b^\top T_1 b + 2T_2^\top b + T_3$ yields $I(b) \geq b^\top T_1 b + 2T_2^\top b + T_3$ if for all $b$ in $\mathbb{R}^d$ it holds that

$$\begin{pmatrix} b \\ 1 \end{pmatrix}^\top \left[\begin{pmatrix} \Lambda_1 + \Lambda_1^\top - T_1 & \Lambda_2 - T_2 \\ \Lambda_2^\top - T_2^\top & y - T_3 \end{pmatrix} - \begin{pmatrix} -\Lambda_1^\top A \\ c^\top - \Lambda_2^\top A \end{pmatrix} G^\dagger \begin{pmatrix} -A^\top\Lambda_1 & c - A^\top\Lambda_2 \end{pmatrix}\right] \begin{pmatrix} b \\ 1 \end{pmatrix} \geq 0$$

and

$$\left(\mathbb{I}_d - GG^\dagger\right)\left(-A^\top\Lambda_1,\, c - A^\top\Lambda_2\right) = 0.$$

After a Schur complement [7, Thm 4.3], we obtain the first part of the theorem

$$\exists\Lambda_1,\Lambda_2: \quad \begin{pmatrix} \Lambda_1 + \Lambda_1^\top - T_1 & \Lambda_2 - T_2 & -\Lambda_1^\top A \\ \Lambda_2^\top - T_2^\top & y - T_3 & c^\top - \Lambda_2^\top A \\ -A^\top\Lambda_1 & c - A^\top\Lambda_2 & G \end{pmatrix} \succeq 0 \implies I(b) \geq b^\top T_1 b + 2T_2^\top b + T_3.$$

As $I(b)$ is a quadratic function there exist $T_1^\star$, $T_2^\star$ and $T_3^\star$ such that $I(b) = b^\top T_1^\star b + 2T_2^{\star\top} b + T_3^\star$. The equations (13) and (14) guarantee [7, Thm 4.3] that

$$\begin{pmatrix} \Lambda_1^\star + \Lambda_1^{\star\top} - T_1^\star & \Lambda_2^\star - T_2^\star & -\Lambda_1^{\star\top} A \\ \Lambda_2^{\star\top} - T_2^{\star\top} & y - T_3^\star & c^\top - \Lambda_2^{\star\top} A \\ -A^\top\Lambda_1^\star & c - A^\top\Lambda_2^\star & G \end{pmatrix} \succeq 0$$

completing the proof. □

## B Proofs

*Corollary 4.1:*

From Theorem 3.1, we have that the generating distribution $T$ for $\alpha$-unimodal ambiguity sets satisfies

$$T([0,t]) = \alpha \int_0^t \lambda^{\alpha-1}\,\mathrm{d}\lambda, \quad \forall t \in [0,1].$$

The moment transformations from Theorem 2.1 become

$$\mu_\alpha := \left[\int_0^\infty \lambda\, T(\mathrm{d}\lambda)\right]^{\text{-}1}\mu \;= \left[\alpha \int_0^1 \lambda^\alpha(\mathrm{d}\lambda)\right]^{\text{-}1}\mu \quad = \frac{\alpha+1}{\alpha}\mu$$

$$S_\alpha := \left[\int_0^\infty \lambda^2\, T(\mathrm{d}\lambda)\right]^{\text{-}1} S = \left[\alpha \int_0^1 \lambda^{\alpha+1}(\mathrm{d}\lambda)\right]^{\text{-}1} S = \frac{\alpha+2}{\alpha} S.$$

From Proposition 2.1, the transformed loss function $L_s$ required in Theorem 2.1 can be found as

$$\begin{aligned} L_s(y) &= \max_{i\in\mathcal{I}}\; T\left(\left[b_i/a_i^\top y, \infty\right)\right) \cdot \mathbb{1}_{a_i^\top x \geq b_i}(y) \\ &=: \max_{i\in\mathcal{I}} f_i(a_i^\top y), \end{aligned}$$

where

$$f_i(q) = \begin{cases} \alpha \displaystyle\int_{b_i/q}^1 \lambda^{\alpha-1}\,\mathrm{d}\lambda, & q \geq b_i, \\ 0 & \text{otherwise.} \end{cases}$$

In order to apply Theorem 4.1, we now need only reformulate the semi-infinite constraint $(\mathcal{C}_2)$, i.e. the constraint

$$T_{3,i} + 2qT_{2,i} + q^2 T_{1,i} \geq f_i(q) \quad \forall q \in \mathbb{R}, \quad \forall i \in \mathcal{I}.$$

Because $0 \in \Xi$ and hence $b_i > 0$, we have equivalently, for each $i \in \mathcal{I}$, and for all $q \in \mathbb{R}^+$

$$T_{3,i} + 2qT_{2,i} + q^2 T_{1,i} \geq \begin{cases} 1 - (b_i/q)^\alpha & q \geq b_i, \\ 0 & \text{otherwise.} \end{cases}$$

which can be seen to reduce to

$$T_{3,i} + 2qT_{2,i} + q^2T_{1,i} \geq 1 - \frac{b_i^\alpha}{q^\alpha}, \quad \forall q \geq 0.$$

Defining a new scalar variable $\tilde{q}$ and applying the variable substitution $\tilde{q}^{\,w} = q$, this can be rewritten as

$$\tilde{q}^{2\,w+v}T_{1,i} + 2\tilde{q}^{w+v}T_{2,i} + \tilde{q}^v(T_{3,i} - 1) + b_i^\alpha \geq 0, \quad \forall \tilde{q} \geq 0$$

after multiplying both sides with $\tilde{q}^v > 0$. The final result is obtained after the substitution $b_i^{1/w}\bar{q} = \tilde{q}$.

*Corollary 4.2:* The method of proof follows that of Corollary 4.1, except that we now apply Proposition 2.2 to generate the transformed loss function $L_s$.

In this case the loss function $L$ is equivalent to $L = \ell \circ \kappa_\Xi$ with $\ell(t) = \max\{0, t-1\}$. Recalling from Theorem 3.1 the generating distribution $T$ for $\alpha$-unimodal distributions, we set

$$\begin{aligned}\ell_s(t) &= \int_0^\infty \ell(\lambda t)T(\mathrm{d}\lambda) \\ &= \alpha \int_0^1 \max\{0, (\lambda t - 1)\}\lambda^{\alpha-1}\mathrm{d}\lambda,\end{aligned}$$

which is zero for any $t \leq 1$. For $t \geq 1$, we can evaluate the integral to get

$$\begin{aligned}\forall t \geq 1 \ : \ \ell_s(t) &= \alpha \int_{1/t}^1 (t\lambda^\alpha - \lambda^{\alpha-1})\mathrm{d}\lambda \\ &= \frac{\alpha}{\alpha+1}t - 1 + \frac{1}{\alpha+1}\left(\frac{1}{t}\right)^\alpha\end{aligned}$$

and then set $L_s(x) = \max_{i\in\mathcal{I}} f_i(a_i^\top x)$ where each $f_i(q) := \ell_s(q/b_i)$.

We can now apply Theorem 4.1 by reformulating the constraint $(\mathcal{C}_2)$ for this choice of $f_i$ for each $i \in \mathcal{I}$, resulting in the constraint

$$T_{3,i} + 2qT_{2,i} + q^2T_{1,i} \geq \frac{\alpha}{\alpha+1}\frac{q}{b_i} - 1 + \frac{1}{\alpha+1}\frac{b_i^\alpha}{q^\alpha} \quad \forall q \geq b_i$$

because $0 \in \Xi$ and hence $b_i > 0$. We define a new scalar variable $\tilde{q}$ and apply the variable

substitution $\tilde{q}^{\,w} = q$, resulting in the constraint

$$\tilde{q}^{2w+v} T_{1,i} + \tilde{q}^{w+v} \left( 2T_{2,i} - \frac{\alpha}{(\alpha+1)b_i} \right) + \tilde{q}^{v} \left( 1 + T_{3,i} \right) - \frac{b_i^{\alpha}}{\alpha+1} \geq 0, \quad \forall \tilde{q} \geq b_i^{1/w}$$

after multiplying both sides by $\tilde{q}^v > 0$. The final result is obtained after the substitution $b_i^{1/w} \bar{q} = \tilde{q}$.

*Corollary 4.3:*

We follow the same approach as the proof of Corollary 4.1, but this time use the generating distribution $T$ for $\gamma$-monotone distributions from Theorem 3.2, i.e.

$$T([0,t]) = \frac{1}{B(n,\gamma)} \cdot \int_0^t \lambda^{n-1} \cdot (1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda, \quad \forall t \in [0,1].$$

In this case the moment transformations from Theorem 2.1 become

$$\mu_\gamma := \left[ \int_0^\infty \lambda \, T(\mathrm{d}\lambda) \right]^{\text{-}1} \mu \;\; = \left[ \frac{1}{B(n,\gamma)} \int_0^1 \lambda^n (1-\lambda)^{\gamma-1} (\mathrm{d}\lambda) \right]^{\text{-}1} \mu \quad = \frac{n+\gamma}{n} \mu$$

$$S_\gamma := \left[ \int_0^\infty \lambda^2 \, T(\mathrm{d}\lambda) \right]^{\text{-}1} S = \left[ \frac{1}{B(n,\gamma)} \int_0^1 \lambda^{n+1} (1-\lambda)^{\gamma-1} (\mathrm{d}\lambda) \right]^{\text{-}1} S = \frac{n+\gamma}{n} \frac{n+\gamma+1}{n+1} S.$$

From Proposition 2.1, the transformed loss function $L_s$ required in Theorem 2.1 become

$$\begin{aligned} L_s(y) &= \max_{i \in \mathcal{I}} \; T\left( \left[ b_i / a_i^\top y, \infty \right) \right) \cdot \mathbb{1}_{a_i^\top x \geq b_i}(y) \\ &=: \max_{i \in \mathcal{I}} f_i(a_i^\top y). \end{aligned}$$

where

$$f_i(q) = \begin{cases} \dfrac{1}{B(n,\gamma)} \displaystyle\int_{b_i/q}^1 \lambda^{n-1} (1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda, & q \geq b_i, \\ 0 & \text{otherwise.} \end{cases}$$

For $q \geq b_i$, we can use a binomial expansion to evaluate this integral[4], obtaining

$$\begin{aligned} B(n,\gamma) f_i(q) &= B(n,\gamma) - \int_0^{b_i/q} \lambda^{n-1} \cdot (1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda \\ &= B(n,\gamma) - \sum_{k=0}^{\gamma-1} \int_0^{b_i/q} (-1)^k \binom{\gamma-1}{k} \lambda^{n+k-1} \, \mathrm{d}\lambda \\ &= B(n,\gamma) - b_i^n \sum_{k=0}^{\gamma-1} \frac{(-b_i)^k}{n+k} \binom{\gamma-1}{k} \frac{1}{q^{n+k}} \end{aligned}$$

In order to apply Theorem 4.1, we now need only reformulate the semi-infinite constraint $(\mathcal{C}_2)$. We obtain, for each $i \in \mathcal{I}$, the constraint

$$T_{3,i} + 2qT_{2,i} + q^2 T_{1,i} \geq 1 - \frac{b_i^n}{B(n,\gamma)} \sum_{k=0}^{\gamma-1} \frac{(-b_i)^k}{n+k} \binom{\gamma-1}{k} \frac{1}{q^{n+k}}, \ \forall q \geq b_i.$$

recalling that $0 \in \Xi$ and hence $b_i > 0$. We multiply both sides by $q^{n+\gamma-1} > 0$ to produce, for each $i \in \mathcal{I}$ the constraint

$$T_{1,i} q^{n+\gamma+1} + 2T_{2,i} q^{n+\gamma} + (T_{3,i} - 1)\, q^{n+\gamma-1} + \frac{b_i^n}{B(n,\gamma)} \sum_{k=0}^{\gamma-1} \frac{(-b_i)^k}{n+k} \binom{\gamma-1}{k} q^{\gamma-k-1} \geq 0, \quad \forall q \geq b_i.$$

The final result is obtained after the substitution $b_i \bar{q} = q$.

*Corollary 4.4:*

The method of proof parallels that of Corollary 4.2, but this time using the generating distribution $T$ for $\gamma$-monotone distributions from Theorem 3.2. In this case we set

$$\ell_s(t) = \frac{1}{B(n,\gamma)} \int_0^1 \max\{0, (\lambda t - 1)\} \lambda^{n-1} (1-\lambda)^{\gamma-1} \mathrm{d}\lambda,$$

which is zero for any $t \leq 1$. For any $t \geq 1$, using a binomial expansion we can evaluate the

[4]Note that the integral amounts to $1 - \frac{1}{B(n,\gamma)} \int_0^{b_i/q} \lambda^{n-1}(1-\lambda)^{\gamma-1} \, \mathrm{d}\lambda =: 1 - I_{b_i/q}(n,\gamma)$, where $I_{b_i/q}(n,\gamma)$ is the so-called *regularized incomplete beta function*, i.e. the cumulative distribution function for the beta distribution with shape parameters $(n,\gamma)$.

integral to get

$$\begin{aligned}\forall t \geq 1 \ : \ B(n,\gamma)\ell_s(t) &= t\int_{1/t}^{1} \lambda^n(1-\lambda)^{\gamma-1}\,\mathrm{d}\lambda - \int_{1/t}^{1}\lambda^{n-1}(1-\lambda)^{\gamma-1}\,\mathrm{d}\lambda \\ &= tB(n+1,\gamma) - B(n,\gamma) + \int_0^{1/t}\lambda^{n-1}(1-\lambda)^{\gamma-1}\mathrm{d}\lambda - t\int_0^{1/t}\lambda^n(1-\lambda)^{\gamma-1}\mathrm{d}\lambda \\ &= tB(n+1,\gamma) - B(n,\gamma) + \sum_{k=0}^{\gamma-1}\left[(-1)^k\binom{\gamma-1}{k}\int_0^{1/t}\left(\lambda^{n-1}-t\lambda^n\right)\lambda^k\,\mathrm{d}\lambda\right] \\ &= tB(n+1,\gamma) - B(n,\gamma) + \sum_{k=0}^{\gamma-1}\frac{(-1)^k}{(n+k)(n+k+1)}\binom{\gamma-1}{k}\left(\frac{1}{t}\right)^{n+k}\end{aligned}$$

and then set $L_s(x) = \max_{i\in\mathcal{I}} f_i(a_i^\top x)$ where each $f_i(q) := \ell_s(q/b_i)$. In order to apply Theorem 4.1, we now need only reformulate the semi-infinite constraint $(\mathcal{C}_2)$. We obtain, for each $i \in \mathcal{I}$, the constraint

$$\begin{aligned}T_{3,i} + 2qT_{2,i} + q^2T_{1,i} \geq \frac{B(n+1,\gamma)}{b_iB(n,\gamma)}q - 1 + \\ \frac{b_i^n}{B(n,\gamma)}\sum_{k=0}^{\gamma-1}\frac{(-b_i)^k}{(n+k)(n+k+1)}\binom{\gamma-1}{k}\frac{1}{q^{n+k}} \qquad \forall q \geq b_i\end{aligned}$$

because $0 \in \Xi$ and hence $b_i > 0$. We multiply both sides by $q^{n+\gamma-1} > 0$ to produce the constraint

$$\begin{aligned}T_{1,i}q^{n+\gamma+1} + \left(2T_{2,i} - \frac{B(n+1,\gamma)}{b_iB(n,\gamma)}\right)q^{n+\gamma} + (T_{3,i}+1)\,q^{n+\gamma-1} - \\ \frac{b_i^n}{B(n,\gamma)}\sum_{k=0}^{\gamma-1}\frac{(-b_i)^k}{(n+k)(n+k+1)}\binom{\gamma-1}{k}q^{\gamma-k-1} \geq 0, \quad \forall q \geq b_i.\end{aligned}$$

The final result is obtained after the substitution $b_i\bar{q} = \tilde{q}$.

## C Factor models in insurance

As mentioned in Section 1.1, any worst-case CVaR problem can be reduced to a related worst-case expectation problem. We are therefore interested in loss functions of the form $L(S_d) = \min\left(\max\left(S_d, 0\right), k\right) - \beta$ for $0 \leq \beta \leq k$. We have that the loss function $L(S_d)$ can be written as

the gauge function $L(S_d) = \ell \circ \kappa_\Xi(S_d)$ for $\Xi = \left\{ x \in \mathbb{R}^d \;\middle|\; \sum_{i=1}^d x_i \geq 1 \right\}$ and

$$\ell = \begin{cases} 0 & \text{if } t \leq \beta, \\ t - \beta & \text{if } \beta \leq t < k, \\ k - \beta & \text{if } t \geq k. \end{cases}$$

Recalling from Theorem 3.1 the generating distribution $T$ for $\alpha$-unimodal distributions, we set $\ell_s(t) = \int_0^\infty \ell(\lambda t) T(\mathrm{d}\lambda)$ which is zero for any $t \leq \beta$. For $\beta \leq t < k$, we can evaluate the integral to get

$$\begin{aligned} \beta \leq \forall t < k \;:\; \ell_s(t) &= \alpha \int_{\beta/t}^1 (\lambda t - \beta) \lambda^{\alpha-1} \mathrm{d}\lambda \\ &= \frac{\alpha}{\alpha+1} t - \beta + \frac{\beta^{\alpha+1}}{\alpha+1} \frac{1}{t^\alpha}. \end{aligned}$$

Similarly for $t \geq k$, we get

$$\begin{aligned} \forall t \geq k \;:\; \ell_s(t) &= \alpha \int_{\beta/t}^{k/t} (\lambda t - \beta) \lambda^{\alpha-1} \mathrm{d}\lambda + \alpha \int_{k/t}^1 (k - \beta) \lambda^{\alpha-1} \mathrm{d}\lambda \\ &= k - \beta - \frac{k^{\alpha+1} - \beta^{\alpha+1}}{\alpha+1} \frac{1}{t^\alpha} \end{aligned}$$

and then set $L_s(x) = \ell_s(\sum_{i=1}^d x_i)$. In order to apply Theorem 4.1, we now need only reformulate the semi-infinite constraint $(\mathcal{C}_2)$. This can be done using methods analogous to the method described in the proof of Corollary 4.2, but is omitted here for the sake of brevity. We get finally

$$\left\{ \begin{array}{ll} T_{1,i} q^2 + 2q T_{2,i} + T_{3,i} \geq 0, & \forall q \in \mathbb{R} \\ T_{1,i} \beta^2 q^{2w+v} + q^{w+v} \beta \left( 2T_{2,i} - \dfrac{\alpha}{\alpha+1} \right) + q^v \left( T_{3,i} + \beta \right) - \dfrac{\beta}{\alpha+1} \geq 0, & 1 \leq \forall q < \left( \dfrac{k}{\beta} \right)^{1/w} \\ T_{1,i} k^2 q^{2w+v} + 2k q^{w+v} T_{2,i} + q^v \left( T_{3,i} + \beta - k \right) + k \dfrac{1 - (\beta/k)^{\alpha+1}}{\alpha+1} \geq 0, & \forall q \geq 1 \end{array} \right\} (\mathcal{C}_2)$$

## References


[1] S.N. Bernstein. Sur les fonctions absolument monotones. *Acta Mathematica*, 52(1):1–66, 1929.

[2] E.M.J. Bertin, R. Theodorescu, and I. Cuculescu. *Unimodality of Probability Measures*. Mathematics and its applications. Springer, 1997.

[3] D. Bertsimas and I. Popescu. On the relation between option and stock prices: A convex optimization approach. *Operations Research*, 50(2):358–374, 2002.

[4] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[5] S.W. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications*, volume 27 of *Probability and Mathematical Statistics*. Academic Press, 1988.

[6] P. Embrechts, G. Puccetti, and L. Rüschendorf. Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, 37(8):2750–2764, 2013.

[7] J. Gallier. The Schur complement and symmetric positive semidefinite (and definite) matrices. Technical report, Penn Engineering, 2010.

[8] J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3):502–506, 1953.

[9] A.W. Lo. Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19(2):373–387, 1987.

[10] J. Löfberg. Pre- and post-processing sum-of-squares programs in practice. *IEEE Transactions on Automatic Control*, 54(5):1007–1011, 2009.

[11] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[12] Y. Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

[13] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *Studies in Applied and Numerical Mathematics*. SIAM, 1994.

[14] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Series in Operational Research and Financial Engineering. Springer New York, 2006.

[15] R.A. Olshen and L.J. Savage. A Generalized Unimodality. Technical Report 143, Stanford University, 1969.

[16] D.D. Pestana and S. Mendonça. Higher-order monotone functions and probability theory. In *Generalized Convexity and Generalized Monotonicity*, pages 317–331. Springer, 2001.

[17] R.R. Phelps. *Lectures on Choquet's Theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer, 2001.

[18] I. Popescu. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.

[19] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[20] L. Rüschendorf. *Fréchet-bounds and their applications*. Springer, 1991.

[21] I.R. Savage. Probability inequalities of the Tchebycheff type. *Journal of Research of the National Bureau of Standards–B. Mathematics and Mathematical Physics*, 65B(3):211–226, 1961.

[22] A. Shapiro. On duality theory of conic linear problems. *Nonconvex Optimization and its Applications*, 57:135–155, 2001.

[23] A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

[24] B.P.G. Van Parys, P.J. Goulart, and D. Kuhn. Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming*, pages 1–32, 2015.

[25] B.P.G. Van Parys, D. Kuhn, P.J. Goulart, and M. Morari. Distributionally robust control of constrained stochastic systems. Technical report, ETH Zürich, 2013.

[26] B.P.G. Van Parys, B.F. Ng, P.J. Goulart, and R. Palacios. *Optimal Control for Load Alleviation in Wind Turbines*. American Institute of Aeronautics and Astronautics, 2014.

[27] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.

[28] Y. Yamada and J.A. Primbs. Value-at-risk estimation for dynamic hedging. *International Journal of Theoretical and Applied Finance*, 5(4):333–354, 2002.

[29] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

[30] S. Zymler, D. Kuhn, and B. Rustem. Worst-case value-at-risk of non-linear portfolios. *Management Science*, 59(1):172–188, 2013.