Minkowski Centers via Robust Optimization: Computation and Applications

Dick den Hertog
Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands
d.denhertog@uva.nl

Jean Pauphilet
Management Science and Operations, London Business School, London, United Kingdom
ORCID: 0000-0001-6352-0984
jpauphilet@london.edu

Mohamed Yahya Soali
Industrial Engineering and Operations Research, Columbia University, New York City, NY, United States
ms6352@columbia.edu

Centers of convex sets are geometric objects that have received extensive attention in the mathematical and optimization literature, both from a theoretical and practical standpoint. For instance, they serve as initialization points for many algorithms such as interior-point, hit-and-run, or cutting-planes methods. First, we observe that computing a Minkowski center of a convex set can be formulated as the solution of a robust optimization problem. As such, we can derive tractable formulations for computing Minkowski centers of polyhedra and convex hulls. Computationally, we illustrate that using Minkowski centers, instead of analytic or Chebyshev centers, improves the convergence of hit-and-run and cutting-plane algorithms. We also provide efficient numerical strategies for computing centers of the projection of polyhedra and of the intersection of two ellipsoids.

Key words: Minkowski center; Geometry; Robust optimization;
History: February 2, 2023

1. Introduction

Centers of convex sets have played a fundamental role in all areas of applied mathematics, especially in mathematical programming. Historically, the development of efficient linear optimization algorithms is deeply connected with the definition and computation of the center of a polytope. The ellipsoid algorithm solves linear optimization problems by constructing a volume-decreasing sequence of circumscribed ellipsoids (see Bland et al. 1981, for a review). This algorithm, proposed by Yudin and Nemirovskii (1976), sparked interest on computing the minimum-volume circumscribed ellipsoid of a polytope or a generic convex body (see, e.g., Todd 1982). Alternatively, Tarasov et al. (1988) proposed the inscribed ellipsoid method, where, at each step, one needs to compute numerically an approximation to the maximum-volume inscribed ellipsoid of a polytope. Since optimizing over ellipsoids is easier than over a general convex set, the minimum-volume circumscribed and the maximum-volume inscribed ellipsoids provide inner and outer approximations that can also be used to approximately solve optimization problems over a convex set. Karmarkar (1984) introduced another polynomial-time interior point algorithm for linear optimization. At each iteration, the algorithm constructs a mapping of the feasible set into a standard simplex and associates the current iterate with "the center" of the simplex, without providing a generic definition of center. Analytic centers were first defined and analyzed by Huard (1967), and further analyzed by Sonnevend (1986), Renegar (1988), Jarre (1989). Modern interior point methods rely mostly on the analytic center due to its computational benefits (see, e.g., Roos et al. 2005).

Besides extremal ellipsoids problems, the Minkowski measure of symmetry (Minkowski 1911) has been proposed as another geometric definition of the center of a convex set. Yet, compared to the other definitions, the Minkowski center has driven mostly theoretical interest and there is, to the best of our knowledge, no computational evidence on the tractability and the practical benefits of Minkowski centers. The present paper provides a first answer to these issues.

1.1. Contributions and structure

This paper shows that Minkowski centers of a convex set are solutions of a robust optimization problem. Under this robust lens, we provide computationally tractable reformulations or approximations for a series of sets including polyhedra and projections of polyhedra. We can also derive known and new analytic expressions for the symmetry measure of simple sets by analyzing the optimization formulation directly, instead of the geometry of the set. We demonstrate numerically that Minkowski centers are viable alternatives to other centers, such as Chebyshev or analytic centers, and can speed up convergence of numerical algorithms.

After presenting the notations in Section 1.2 and the existing literature on centers of convex bodies in Section 1.3, the rest of the paper is organized as follows:

- We introduce Minkowski centers in Section 2 and connect them with existing definitions of centers. In particular, we show that Minkowski centers are special cases of Helly centers, like the centroid, the John or the volumetric center. We then derive a robust optimization formulation for computing Minkowksi centers of a convex set (Proposition 3). Under this lens, we derive tractable reformulations of this optimization problem for polyhedra and the convex hull of a finite number of points and provide known and new analytical bounds in simple cases.
- Numerically, analytic centers are widely used in the initialization of many algorithms despite the fact that they are analytical and not geometric. We demonstrate empirically in Section 3 that using Minkowski centers instead can provide substantial benefit in terms of algorithmic convergence, using the hit-and-run and the cutting-plane algorithms as illustrating examples.

- In Section 4, we consider the case of a convex set defined as the projection of a polyhedron. We show that computing Minkowski centers for such a set is equivalent to solving an *adjustable* robust optimization problem. We propose an approximation based on linear decision rules and evaluate its practical relevance on numerical simulations.
- Finally, in Section 5, we propose a numerical strategy for approximating a Minkowski center of the intersection of two ellipsoids. Our algorithm relies on a second-order cone (SOC) relaxation and bisection search. We also provide a (numerically verifiable) condition under which our approximation is tight, together with a constant factor approximation bound for our approach. We also discuss the extension to intersection of $m > 2$ ellipsoids.

1.2. Notations

We use nonbold face characters (x or λ) to denote scalars, lowercase bold faced characters ($\boldsymbol{x}$) to denote vectors, uppercase bold faced characters ($\boldsymbol{X}$) to denote matrices, and calligraphic characters such as $\mathcal{X}$ to denote sets. We let $\boldsymbol{e}$ (resp. $\boldsymbol{0}$) denote the vector of all 1's (resp. 0's), with dimension implied by the context. We denote by $\boldsymbol{e}_i$ the unit vector with 1 at the ith coordinate and zero elsewhere. $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$ denote the set of real numbers, non-negative real numbers, and non-negative integers, respectively. For a positive integer $n \in \mathbb{N}$, we define $[n] := \{1, \dots, n\}$. Given two n-dimensional vectors $\boldsymbol{x}, \boldsymbol{y}$, we use the notation $\boldsymbol{x}^\top \boldsymbol{y}$ for the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x}^\top \boldsymbol{y} := \sum_{i \in [n]} x_i y_i$, and $\|\boldsymbol{x}\|$ for the Euclidean norm of $\boldsymbol{x}$, $\|\boldsymbol{x}\| := \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$. For $p \geq 1$, the p-norm of $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\|_p = \left(\sum_{i \in [n]} |x_i|^p\right)^{1/p}$, so that $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2$.

1.3. Literature review

In this section, we present the various definitions of centers that have been proposed in the applied mathematics literature.

Historically, the first definition of a center is the center of mass (or barycenter), used primarly in physics and motion geometry (see, e.g., Schwartz and Sharir 1988). The center of mass of a set is defined as the weighted arithmetic mean position of all its points, i.e.,

$$\frac{1}{\int_{\mathbb{R}^n} \mu(\boldsymbol{x})\, d\boldsymbol{x}} \int_{\mathbb{R}^n} \boldsymbol{x} \mu(\boldsymbol{x})\, d\boldsymbol{x},$$

where $\mu(\cdot)$ is a given mass density function over the set of interest $\mathcal{C}$. When μ is uniform, the center of mass is also called the centroid. In particular, the centroid of a finite number of m points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$, is $\frac{1}{m} \sum_{i \in [m]} \boldsymbol{x}_i$. Computing the centroid of a polytope is $\#P-$hard (Rademacher 2007), but it can be efficiently approximated via random sampling. In data science, the notion of centroid is the building block of the k-means clustering algorithm (see, e.g., Kanungo et al. 2002).

For convex sets, an important geometrical definition of a center is the notion of Helly center:

Definition 1. For a convex set $\mathcal{C}$, we say that $\boldsymbol{x}_H \in \mathcal{C}$ is a Helly center if for any chord $[\boldsymbol{u}, \boldsymbol{v}]$ passing through $\boldsymbol{x}_H$, we have

$$\frac{1}{n+1} \leq \frac{\|\boldsymbol{x}_H - \boldsymbol{u}\|}{\|\boldsymbol{v} - \boldsymbol{u}\|} \leq \frac{n}{n+1}.$$

Intuitively, the ratio of distances in Definition 1 measures how close $\boldsymbol{x}_H$ is to $\boldsymbol{u}$ and $\boldsymbol{v}$ (in relative terms). Helly centers are thus points that are sufficiently far away from the boundary. Klee (1963) proves that any convex compact body of $\mathbb{R}^n$ admits a Helly center, as a consequence of Helly's theorem. However, it is in general not unique. For instance, Barnes and Moretti (2005) prove that an ellipsoid admits an infinity of Helly centers (theorem 2.5).

Another class of centers encompasses centers defined via extremal ellipsoids (see Gler and Grtuna 2012, for a complete treatment). For instance, the center of the minimum-volume ellipsoid that contains a set $\mathcal{C}$ is referred to as the John (or Löwner-John) center of $\mathcal{C}$. The John center is well defined for convex bodies and unique (John 1948). Alternatively, the center of the maximum-volume ellipsoid contained in $\mathcal{C}$ is called the volumetric center of $\mathcal{C}$ (Vaidya 1996). However, even for polyhedra, known algorithms for finding the maximum-volume ellipsoid and its center require solving a semi-definite optimization problem (see, e.g., Boyd and Vandenberghe 2004, section 8.4.2). Recently, Zhen and den Hertog (2018) use Fourier-Motzkin decomposition and adjustable robust optimization techniques to approximate it in a tractable fashion for projection of polyhedra of the form $\{\boldsymbol{x} : \exists \boldsymbol{z} \text{ s.t. } \boldsymbol{A}_{\boldsymbol{x}} \boldsymbol{x} + \boldsymbol{A}_{\boldsymbol{z}} \boldsymbol{z} \leq \boldsymbol{b}\}$. If we further restrict our attention to isotropic ellipsoids, the center of a maximum-volume ball enclosed in $\mathcal{C}$ is often called a Chebyshev center of $\mathcal{C}$:

Definition 2. A Chebyshev center of a convex set $\mathcal{C}$ is a solution of the optimization problem

$$\max_{\boldsymbol{x} \in \mathcal{C}, r \in \mathbb{R}_+} r \text{ s.t. } \mathcal{B}(\boldsymbol{x}, r) \subseteq \mathcal{C},$$

where $\mathcal{B}(\boldsymbol{x}, r)$ denotes the ball centered in $\boldsymbol{x}$ and of radius r (in Euclidean norm).

Chebyschev centers of a polyhedron can be computed by solving a linear optimization problem (see, e.g., Boyd and Vandenberghe 2004, section 4.3.1 and 8.5.1). Note that the definition of Chebyshev centers is ambiguous in the literature and that some authors, e.g., Eldar et al. (2008), Xia et al. (2021), alternatively define them as the centers of the minimum-volume circumscribed balls. To avoid any ambiguity, we will consistently use Definition 2 in this paper. Finally, the main limitation of centers defined via extremal ellipsoids is that they require the convex set $\mathcal{C}$ to be fully-dimensional (or they require to restrict our attention to ellipsoids in the affine hull of $\mathcal{C}$).

Finally, the most used definition of a center in optimization is certainly the analytic center:

Definition 3. The analytic center of the convex set $\mathcal{C} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b};\ f_i(\boldsymbol{x}) \leq 0,\ \forall i \in [m]\}$ is the solution of the optimization problem

$$\max_{\boldsymbol{x}} \sum_i \log\left(-f_i(\boldsymbol{x})\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

The maximization problem above aims to find a strictly feasible point $\boldsymbol{x} \in \mathcal{C}$ with the largest sum of log-slacks. When $\mathcal{C}$ is bounded, the logarithmic barrier terms $\log(-f_i(\boldsymbol{x}))$ are bounded above, the optimization problem is well defined. For polyhedra, the analytic center, when it exists, is unique. Being defined as the solutions of a convex optimization problem, analytic centers can be computed in a tractable fashion. However, a major deficiency of this definition is that it is not a geometric concept but rather depends on the analytical description of the set $\mathcal{C}$. For example, the analytic center of the n-dimensional standard simplex defined as $\{\boldsymbol{x} \in \mathbb{R}^n_+ : \boldsymbol{e}^\top \boldsymbol{x} = 1\}$ is the vector $\frac{1}{n}\boldsymbol{e}$, but the analytic center of the geometrically equivalent set $\{\boldsymbol{x} \in \mathbb{R}^n_+ : \boldsymbol{e}^\top \boldsymbol{x} \leq 1, \boldsymbol{e}^\top \boldsymbol{x} \geq 1\}$ does not exist. Similarly, duplicating or adding redundant constraints in the description of $\mathcal{C}$ pushes the analytic center arbitrarily close to the boundary (Caron et al. 2002). Yet, the analytic center remains very popular and a cornerstone in optimization algorithms.

From a geometric perspective, analytic or Chebyshev centers are not necessarily Helly centers, as formally stated below:

Proposition 1. *Chebyshev or analytic centers are not necessarily Helly centers.*

The proof of Proposition 1 relies on counter-examples provided in Appendix A. This negative result motivates our investigation into alternative definitions of centers. In the rest of this paper, we study Minkowski centers, which are special cases of Helly center—as we will prove in the next section. For illustration purposes, Figure 1 represents the Chebyshev, analytic, and Minkowski centers of a 2-dimensional polytope. The presence of multiple constraints on the left pushes the analytic center to the right, while the pointy shape of the polyhedron settles the Chebyshev center on the left-hand side. The Minkowski center, on the other hand, appears more 'central'.

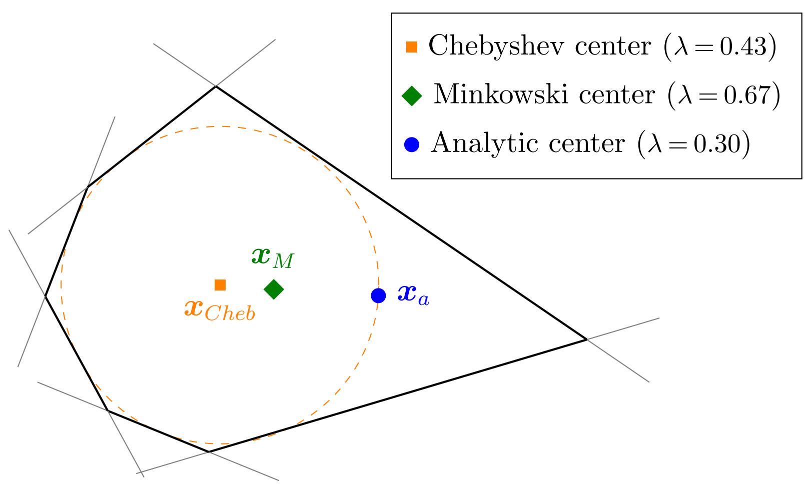


Figure 1 **Example of the Chebyshev (x_{Cheb}), Analytic (x_a), and Minkowski (x_M) center of a 2-dimensional polyhedron.**

2. Minkowski center and robust optimization formulation

In this paper, we study Minkowski centers of a closed, bounded, and convex body $\mathcal{C} \subseteq \mathbb{R}^n$. Minkowski centers are related to the notion of symmetry of the set. Let us first define the symmetry of $\mathcal{C}$ with respect to a point $\boldsymbol{x} \in \mathcal{C}$ as

$$\text{sym}(\boldsymbol{x}, \mathcal{C}) := \max_{\lambda \geq 0} \lambda \text{ s.t. } \boldsymbol{x} + \lambda(\boldsymbol{x} - \boldsymbol{y}) \in \mathcal{C}, \ \forall \boldsymbol{y} \in \mathcal{C}.$$

This measure of symmetry, initially proposed by Minkowski, intuitively states that $\text{sym}(\boldsymbol{x}, \mathcal{C})$ is the largest scalar λ such that every point $\boldsymbol{y} \in \mathcal{C}$ can be reflected through $\boldsymbol{x}$ by the factor λ and still lies in $\mathcal{C}$. Among other properties, we have $\text{sym}(\boldsymbol{x}, \mathcal{C}) \leq 1$. We refer to Belloni and Freund (2007) for an analysis of some fundamental properties of $\text{sym}(\boldsymbol{x}, \mathcal{C})$. Then, a Minkowski center is defined as a point $\boldsymbol{x} \in \mathcal{C}$ maximizing symmetry, i.e.,

Definition 4. $\boldsymbol{x}^\star$ is called a Minkowski center or symmetric point of $\mathcal{C}$ if $\boldsymbol{x}^\star$ is a solution of the optimization problem $\max_{\boldsymbol{x} \in \mathcal{C}} \text{sym}(\boldsymbol{x}, \mathcal{C})$. The optimal objective value, $\text{sym}(\mathcal{C}) := \text{sym}(\boldsymbol{x}^\star, \mathcal{C})$, is called the symmetry of $\mathcal{C}$.

In particular, Minkowski centers are not necessarily unique (for instance, Minkowski centers of $\{\boldsymbol{x} \in [0,1]^3 : x_1 + x_2 \leq 1\}$ are all points of the form $(1/3, 1/3, t)$ for $t \in [1/3, 2/3]$) and the set $\mathcal{C}$ is symmetric with respect to some $\boldsymbol{x}_0$ (i.e., $\forall \boldsymbol{x} \in \mathcal{C}, \ 2\boldsymbol{x}_0 - \boldsymbol{x} \in \mathcal{C}$) if and only if $\text{sym}(\mathcal{C}) = 1$.

2.1. Minkowski centers are Helly centers

Here, we connect the definition of Minkowski centers with the notion of Helly centers. We first provide a sufficient condition for a point $\boldsymbol{x}$ to be a Helly center.

Proposition 2. *If $\boldsymbol{x} \in \mathcal{C}$ satisfies $\frac{1}{n} \leq \text{sym}(\boldsymbol{x}, \mathcal{C})$, then $\boldsymbol{x}$ is a Helly center of $\mathcal{C}$.*

The proof of Proposition 2 is provided in Appendix B.1. From Proposition 2, we can prove that Minkowski centers, among other definitions of centers, are special cases of Helly centers:

Corollary 1. *If $\mathcal{C}$ is full dimensional, (a) the centroid, $\boldsymbol{x}_c$, (b) the John center, $\boldsymbol{x}_J$, (c) the volumetric center, $\boldsymbol{x}_v$, (d) any Minkowski center, $\boldsymbol{x}_M$, are Helly centers.*

Proof We prove that the symmetry of $\mathcal{C}$ at each center is at least $1/n$. The results then follows from Proposition 2. (a) Hammer (1951) proved that $\text{sym}(\boldsymbol{x}_c, \mathcal{C}) \geq 1/n$. (b) The John center is the center of the minimum-volume circumscribed ellipsoid $\mathcal{E}$, $\mathcal{C} \subseteq \mathcal{E}$. John (1948) showed that $(1/n)\mathcal{E} \subseteq \mathcal{C}$ (Theorem 3), which implies that $\text{sym}(\boldsymbol{x}_J, \mathcal{C}) \geq 1/n$. (c) Similarly, the maximum-volume inscribed ellipsoid (whose center is the volumetric center) satisfies $\mathcal{E}' \subseteq \mathcal{C} \subseteq n\mathcal{E}'$ so $\text{sym}(\boldsymbol{x}_v, \mathcal{C}) \geq 1/n$. (d) Since a Minkowski center maximizes symmetry, $\text{sym}(\boldsymbol{x}_M, \mathcal{C}) \geq \text{sym}(\boldsymbol{x}_c, \mathcal{C}) \geq 1/n$. □

2.2. Robust optimization formulation

As a starting point to our analysis, we would like to emphasize that Minkowski centers are the solution of a robust optimization problem. From Definition 4, we can obviously write a Minkowski center as the solution of:

$$\max_{\boldsymbol{x}\in\mathcal{C},\lambda\geq 0} \lambda \text{ s.t. } \boldsymbol{x}+\lambda(\boldsymbol{x}-\boldsymbol{y})\in\mathcal{C},\ (\forall \boldsymbol{y}\in\mathcal{C}), \tag{1}$$

which resembles a robust optimization problem where the set $\mathcal{C}$ defines both the uncertainty set and the constraints. However, the constraints involve products of decision variables, $\lambda\boldsymbol{x}$, hence might be non-convex in $(\lambda,\boldsymbol{x})$. Still, we can reformulate the above optimization problem into one that is convex in its decision variables and uncertain parameters:

Proposition 3. *Assume that $\mathcal{C}$ can be described via linear equality constraints and m convex inequality constraints, i.e., $\mathcal{C}=\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b};\ f_i(\boldsymbol{x})\leq 0, \forall i\in[m]\}$. Consider $(\boldsymbol{w}^\star,\lambda^\star)$, solutions of the following robust convex optimization problem:*

$$\begin{aligned}\max_{\boldsymbol{w},\lambda\geq 0} \lambda \text{ s.t. } \quad & \boldsymbol{A}\boldsymbol{w}=(1+\lambda)\boldsymbol{b},\\ & (1+\lambda)f_i\left(\frac{\boldsymbol{w}}{1+\lambda}\right)\leq 0, \quad \forall i\in[m],\\ & f_i(\boldsymbol{w}-\lambda\boldsymbol{y})\leq 0, \quad \forall \boldsymbol{y}\in\mathcal{C},\forall i\in[m].\end{aligned}\tag{2}$$

Then, the point $\boldsymbol{x}^\star := \boldsymbol{w}^\star/(1+\lambda^\star)$ is a Minkowski center of $\mathcal{C}$ (with symmetry measure $\lambda^\star$).

Note that (2) is a robust optimization problem with a linear objective and constraints that are convex in the decision variables $(\boldsymbol{w},\lambda)$ for a fixed $\boldsymbol{y}$, and convex in the uncertain parameters $\boldsymbol{y}$ for a fixed $(\boldsymbol{w},\lambda)$. This class of robust constraints can be approximated using the so-called reformulation-perspectification technique (see Bertsimas et al. 2022).

Proof Since $\lambda\geq 0$, $1+\lambda>0$ and we can consider the bijective change of variable $(\boldsymbol{x},\lambda)\mapsto(\boldsymbol{w},\lambda)$ with $\boldsymbol{w}=(1+\lambda)\boldsymbol{x}$. Problem (1) becomes

$$\begin{aligned}\max_{\boldsymbol{w},\lambda\geq 0} \lambda \text{ s.t. } \quad & \frac{\boldsymbol{w}}{1+\lambda}\in\mathcal{C},\\ & \boldsymbol{w}-\lambda\boldsymbol{y}\in\mathcal{C},\ \forall\boldsymbol{y}\in\mathcal{C}.\end{aligned}\tag{3}$$

To enforce $\boldsymbol{w}/(1+\lambda)\in\mathcal{C}$, we need to impose

$$\begin{aligned}\boldsymbol{A}\frac{\boldsymbol{w}}{1+\lambda}=\boldsymbol{b} &\iff \boldsymbol{A}\boldsymbol{w}=(1+\lambda)\boldsymbol{b},\\ f_i\left(\frac{\boldsymbol{w}}{1+\lambda}\right)\leq 0 &\iff (1+\lambda)f_i\left(\frac{\boldsymbol{w}}{1+\lambda}\right)\leq 0, \quad \forall i\in[m].\end{aligned}$$

Observe that $(\boldsymbol{x},t)\mapsto t f_i(\boldsymbol{x}/t)$ is the perspective function of f_i and is jointly convex in $(\boldsymbol{x},t)$ over $\{(\boldsymbol{x},t)\in\mathbb{R}^n\times\mathbb{R}_+ : \boldsymbol{x}/t\in\mathrm{dom}(f_i)\}$ (see, e.g., Boyd and Vandenberghe 2004, section 3.2.6). So, all constraints are convex constraints in $(\boldsymbol{w},\lambda)$.

Regarding the robust constraints, $\boldsymbol{w} - \lambda \boldsymbol{y} \in \mathcal{C},\ \forall \boldsymbol{y} \in \mathcal{C}$, we consider the equality and inequality constraints separately. First, $(\boldsymbol{w}, \lambda)$ should satisfy $\boldsymbol{A}\boldsymbol{w} - \lambda \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}, \forall \boldsymbol{y} \in \mathcal{C}$. However, since $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$ for $\boldsymbol{y} \in \mathcal{C}$, these constraints are equivalent to $\boldsymbol{A}\boldsymbol{w} = (1+\lambda)\boldsymbol{b}$, which are already enforced. Second, the inequality constraints can be written as $f_i(\boldsymbol{w} - \lambda \boldsymbol{y}) \leq 0, \forall \boldsymbol{y} \in \mathcal{C}$, which are robust constraints, convex in the decision variables and convex in the uncertain parameters $\boldsymbol{y}$. □

This observation prompts us to investigate whether tools and techniques developed for robust optimization problems could be usefully and successfully applied to compute Minkowski centers of convex sets.

2.3. Tractable reformulations for polyhedra

In robust optimization, tractable reformulations are obtained when the robust constraints are concave in the uncertain parameter (Ben-Tal et al. 2015). When they are convex in the uncertain parameter, like in (2), even computing the worst case scenario, i.e., solving $\max_{\boldsymbol{y} \in \mathcal{C}} f_i(\boldsymbol{w} - \lambda \boldsymbol{y})$ for a fixed $(\boldsymbol{w}, \lambda)$, is challenging. Accordingly, we first consider the easy case where the f_i's are linear, hence both convex and concave.

First, we consider the case where $\mathcal{C}$ is described via linear constraints.

Proposition 4. *Consider* $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b};\ \boldsymbol{C}\boldsymbol{x} \leq \boldsymbol{d}\}$*, where* $\boldsymbol{C} \in \mathbb{R}^{m \times n}$*,* $\boldsymbol{d} \in \mathbb{R}^m$*. For* $i \in [m]$*, define* $\delta_i := \min_{\boldsymbol{y} \in \mathcal{C}} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}$*. Then,* (2) *is equivalent to*

$$\max_{\boldsymbol{w}, \lambda \geq 0} \lambda \ \textit{s.t.} \quad \boldsymbol{A}\boldsymbol{w} = (1+\lambda)\boldsymbol{b}, \quad \boldsymbol{C}\boldsymbol{w} - \lambda \boldsymbol{\delta} \leq \boldsymbol{d}. \tag{4}$$

Remark 1. Since we assume throughout the paper that the set $\mathcal{C}$ is bounded, we know that $\delta_i > -\infty$ for all $i \in [m]$. For an unbounded polyhedron $\mathcal{C}$, however, the behavior of Problem (4) depends on the recession cone of $\mathcal{C}$. Let us consider a decomposition of $\mathcal{C}$ into $\mathcal{C} = \mathcal{C}_0 + \mathcal{H}$, where $\mathcal{C}_0$ is a convex bounded set and $\mathcal{H}$ is the recession cone of $\mathcal{C}$. If $\mathcal{H}$ is not a subspace, then there exists an index i such that $\delta_i = -\infty$. Consequently, the set of feasible (and a fortiori optimal) solutions to (4) is $\mathcal{C} \times \{0\}$, and all points in $\mathcal{C}$ are Minkowski centers. If $\mathcal{H}$ is a subspace, the δ_i's are all finite but the set of Minkowski centers of $\mathcal{C}$ (and the set of optimal solutions to Problem 4) is invariant by translation by any $\boldsymbol{h} \in \mathcal{H}$.

Proof From Proposition 3, we know that a Minkowski center can be obtained by rescaling the solution of the following optimization problem:

$$\begin{aligned}
\max_{\boldsymbol{w}, \lambda \geq 0} \lambda \ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{w} = (1+\lambda)\boldsymbol{b}, \\
& \boldsymbol{C}\boldsymbol{w} \leq (1+\lambda)\boldsymbol{d}, \\
& \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{w} - \lambda \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y} \leq \boldsymbol{e}_i^\top \boldsymbol{d}, \quad \forall \boldsymbol{y} \in \mathcal{C}, \forall i \in [m].
\end{aligned}$$

The ith robust constraint, $i \in [m]$, is equivalent to

$$\boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{w} + \max_{\boldsymbol{y}\in\mathcal{C}} \left\{-\lambda \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}\right\} \le \boldsymbol{e}_i^\top \boldsymbol{d} \iff \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{w} - \lambda \min_{\boldsymbol{y}\in\mathcal{C}} \left\{\boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}\right\} \le \boldsymbol{e}_i^\top \boldsymbol{d},$$

where the equivalence follows from the fact that $\lambda > 0$.

By definition of $\mathcal{C}$, note that $\delta_i := \min_{\boldsymbol{y}\in\mathcal{C}} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y} \quad \le d_i$, so that the constraints $\boldsymbol{C}\boldsymbol{w} - \lambda\boldsymbol{\delta} \le \boldsymbol{d}$ imply $\boldsymbol{C}\boldsymbol{w} \le (1+\lambda)\boldsymbol{d}$, which is then redundant with the robust constraint. □

According to Proposition 4, computing a Minkowski center of a polyhedron can be achieved by solving $m+1$ linear optimization problems, including m optimization problems over the same feasible set $\mathcal{C}$. Proposition 4 hence recovers the numerical approach presented in Belloni and Freund (2007, section 5.2), yet from an optimization perspective. Our approach is also numerically more efficient. Indeed, the number of optimization problems to be solved, m, does not depend on the number of equality constraints but only on the number of linear inequalities in the description of $\mathcal{C}$. On the contrary, the approach of Belloni and Freund (2007) applies to $\mathcal{C}$ described as $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b};\ -\boldsymbol{A}\boldsymbol{x} \le -\boldsymbol{b};\ \boldsymbol{C}\boldsymbol{x} \le \boldsymbol{d}\}$, which is more prodigal in linear inequalities.

Second, we consider the case where $\mathcal{C}$ is described as the convex hull of a finite number of points. Consider m points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ and $\mathcal{C} = \operatorname{conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$. For notation convenience, let us define $\Lambda_m := \{\boldsymbol{\lambda} \in \mathbb{R}^m_+ \mid \sum_{i\in[m]} \lambda_i = 1\}$, so that $\mathcal{C} = \{\sum_{i\in[m]} \lambda_i \boldsymbol{x}_i,\ \mid \boldsymbol{\lambda} \in \Lambda_m\}$.

PROPOSITION 5. *Consider m points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ and $\mathcal{C} = \operatorname{conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$. The optimization problem* (2) *is equivalent to*

$$\max_{\substack{\boldsymbol{w},\lambda\ge 0,\\ \boldsymbol{\nu}^1,\ldots,\boldsymbol{\nu}^m \in \Lambda_m}} \lambda \text{ s.t. } \boldsymbol{w} = \lambda \boldsymbol{x}_i + \sum_{j\in[m]} \nu_j^i \boldsymbol{x}_j,\ \forall i \in [m].$$

Proposition 5 (proved in Appendix B.2) recovers exactly the result provided in Belloni and Freund (2007, section 5.1). Unfortunately, this formulation involves in the order of m^2 decision variables and constraints, so column-and-constraint generation procedures could be investigated to improve practical tractability.

2.4. Analytic expressions for simple sets

Deriving analytic expressions or bounds for the symmetry measure of a set can be of theoretical interest. For instance, in a robust optimization context, Bertsimas et al. (2011b) derive closed-form expression for the symmetry measure of many uncertainty sets by using the following result of Belloni and Freund (2007, equation (40)):

LEMMA 1. *Consider $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b};\ \boldsymbol{C}\boldsymbol{x} \le \boldsymbol{d}\}$, where $\boldsymbol{C} \in \mathbb{R}^{m\times n}$, $\boldsymbol{d} \in \mathbb{R}^m$. For $i \in [m]$, define $\delta_i^: = \min_{\boldsymbol{y}\in\mathcal{C}} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}$. Then, for any $\boldsymbol{x} \in \mathcal{C}$,*

$$\operatorname{sym}(\boldsymbol{x}, S) = \min_{i\in[m]} \frac{d_i - \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{x}}{\boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{x} - \delta_i},$$

with the convention that $0/0 = +\infty$.

In particular, Lemma 1 remains valid if $\mathcal{C}$ is described as the intersection of an infinite number of half-spaces. Based on this observation, Bertsimas et al. (2015) are able to derive explicit formulae for the symmetry measure of some non-polyhedral sets.

In this section, we give new, direct, and simple proofs for some of these results. Our proof technique relies on analyzing the robust optimization formulation (3) directly and naturally leads to generalizations to a broader class of sets than previously studied. In particular, we will consider two special structures, namely permutation-invariant sets and packing constrained sets. Some examples of convex sets and their symmetry measures are reported in Table 1.

First, we can easily compute the Minkowski measures of sets that are permutation-invariant. Indeed, in this case, (3) simplifies into a two-dimensional problem:

Lemma 2. *Assume that $\mathcal{C}$ is permutation-invariant, i.e., for any $\boldsymbol{x} \in \mathcal{C}$ and any permutation σ, $\boldsymbol{x}_\sigma := (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \in \mathcal{C}$. Then, there exists an optimal solution to* (3) *satisfying $\boldsymbol{w} = t\boldsymbol{e}$ for some $t \in \mathbb{R}$.*

Proof Consider a feasible solution for (3), $(\lambda, \boldsymbol{w})$. For any permutation σ, $(\lambda, \boldsymbol{w}_\sigma)$ is also feasible, with same objective value. Define $\bar{\boldsymbol{w}} = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \boldsymbol{w}_\sigma$, where Σ_n is the set of all permutations of $[n]$. Then $(\lambda, \bar{\boldsymbol{w}})$ is also feasible with objective value λ. Applying this construction with an optimal solution $\boldsymbol{w}$ yields the result. □

To illustrate the implications of this observation, we consider the intersection of the p-norm unit ball and the non-negative orthant:

Proposition 6. *Consider $\mathcal{B}_p^+ = \left\{ \boldsymbol{x} \in \mathbb{R}_+^n \mid \|\boldsymbol{x}\|_p \leq 1 \right\}$. Then, $(\lambda^\star, \boldsymbol{w}^\star) = \left(\frac{1}{n^{1/p}}, \frac{1}{n^{1/p}} \boldsymbol{e} \right)$ is an optimal solution of* (3).

The proof is deferred to Appendix B.3 and relies directly on applying Lemma 2 to $\mathcal{B}_p^+$.

We now consider a broad class of polyhedra referred to as packing constrained sets, i.e., sets of the form $\mathcal{P} := \{\boldsymbol{x} \geq \boldsymbol{0} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$, where $\boldsymbol{A} \in \mathbb{R}_+^{m \times n}$ is a matrix with non-negative entries and $\boldsymbol{b} \in \mathbb{R}_+^m$. Among others, such sets appear in the studied multi-dimensional knapsack problem (Kellerer et al. 2004) or in robust optimization with budgeted uncertainty set (Bertsimas and Sim 2004), intersections of budgeted uncertainty sets, CLT sets (Bandi and Bertsimas 2012), and inclusion-constrained budgeted sets (Gounaris et al. 2016).

Proposition 7. *Consider $\mathcal{P} := \{\boldsymbol{x} \geq \boldsymbol{0} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$, with $\boldsymbol{A} \in \mathbb{R}_+^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}_+^m$. For $i \in [n]$, define*

$$y_i^\star := \max_{\boldsymbol{y} \in \mathcal{P}} \boldsymbol{e_i}^\top \boldsymbol{y} = \min_{j \in [m]} \left(\frac{b_j}{A_{ji}} \right),$$

with the convention $1/0 = +\infty$. The Minkowski measure and a scaled Minkowski center of $\mathcal{P}$ are

$$\lambda^\star = \min_{j \in [m]} \left(\frac{b_j}{\boldsymbol{e_j}^\top \boldsymbol{A} \boldsymbol{y}^\star} \right), \quad \boldsymbol{w}^\star = \lambda^\star \boldsymbol{y}^\star.$$

Among others, we can readily apply Proposition 7 to the budgeted uncertainty set. For example, for the budgeted uncertainty set with equal weights, $\Delta_\gamma^e = \left\{ \boldsymbol{x} \in [0,1]^n \;\middle|\; \sum_{i\in[n]} x_i \le \gamma \right\}$ with $\gamma \ge 0$, we have $y_i^\star = \min(1,\gamma)$. If $\gamma \ge n$, we obtain $\lambda^\star = 1$, which is intuitive because in this case the budget constraint is redundant and $\Delta_\gamma^e = [0,1]^n$ is perfectly symmetric. If $\gamma \le n$, we have $\lambda^\star = \dfrac{\gamma}{n\min(1,\gamma)}$. In particular, if $\gamma \le 1$, we have $\lambda^\star = 1/n$, which is consistent with the fact that Δ_γ^e corresponds to a scaled simplex in this case. In the less trivial case where $1 \le \gamma \le n$, we obtain $\lambda^\star = \gamma/n$. A similar discussion can be conducted for a generic budgeted uncertainty set.

Furthermore, a similar line of proof can be applied to the intersection of a class of generalized ellipsoids with the non-negative orthant, that is, sets of the form $\mathcal{E}_p^+ := \{\boldsymbol{x} \ge \boldsymbol{0} \mid \|\boldsymbol{A}\boldsymbol{x}\|_p \le 1\}$, where $\boldsymbol{A} \in \mathbb{R}_+^{m\times n}$, as reported in Table 1. These "non-negative" ellipsoids are also important uncertainty sets in the literature and have been used as baselines in many robust optimization settings (Bertsimas et al. 2011a). The corresponding proofs can be found in Appendix B.4-B.5.

Table 1 Analytical expression of the Minkowski symmetry measure and Minkowski center of simple sets. A box indicates results not already derived in the literature.

No	Convex set	Symmetry measure	Minkowski center
.5 1	$p-$norm unit ball $\mathcal{B}_p^+ = \{\boldsymbol{x} \ge \boldsymbol{0} \mid \|\boldsymbol{x}\|_p \le 1\}$	.5 $\dfrac{1}{n^{1/p}}$	.5 $\dfrac{1}{n^{1/p}+1}\boldsymbol{e}$
2	Standard simplex .5 $\Delta = \left\{\boldsymbol{x} \ge \boldsymbol{0} \mid \sum_{i\in[n]} x_i \le 1\right\}$	$\dfrac{1}{n}$	$\dfrac{1}{n+1}\boldsymbol{e}$
.5 3	Budgeted uncertainty set, equal weights .5 $\Delta_\gamma^e = \left\{\boldsymbol{x} \in [0,1]^m \mid \sum_{i\in[n]} x_i \le \gamma\right\}$, with $\gamma \le n$	.5 $\dfrac{\gamma}{n\min(1,\gamma)}$	.5 $\dfrac{\gamma\min(1,\gamma)}{\gamma + n\min(1,\gamma)}\boldsymbol{e}$
.5 $\boxed{4}$	Budgeted uncertainty set .5 $\Delta_\gamma = \left\{\boldsymbol{x} \in [0,1]^n \mid \sum_{i\in[n]} u_i x_i \le \gamma\right\}$, with $\gamma \le \sum_i u_i$, and $u_i \ge 0$	.5 $\dfrac{\gamma}{\sum_{i\in[n]} \min(u_i,\gamma)}$	.5 $\dfrac{\gamma}{\gamma + \sum_{i\in[n]} \min(u_i,\gamma)}\boldsymbol{e}$
$\boxed{5}$	p-norm ellipsoidal set .5 $\mathcal{E}_p^+ = \left\{\boldsymbol{x} \in \mathbb{R}_+^n \mid \|\boldsymbol{A}\boldsymbol{x}\|_p \le 1\right\}$, with $\boldsymbol{A} \in \mathbb{R}_+^{m\times n}$	$\lambda^\star = \frac{1}{\|\boldsymbol{A}\boldsymbol{y}^\star\|_p}$ with $y_i^\star = \frac{1}{\|\boldsymbol{A}^\top \boldsymbol{e}_i\|_p}, i \in [n]$	.5 $\dfrac{\lambda^\star}{1+\lambda^\star}\boldsymbol{y}^\star$

2.5. Choosing among Minkowski centers

As previously discussed, Minkowksi centers are not uniquely defined, not even for bounded sets. When $\mathcal{C}$ is a compact, convex set with a nonempty interior, Belloni and Freund (2007, proposition 6) prove that the set of its Minkowksi centers is a compact convex set with empty interior. Under our robust optimization lens, multiplicity of Minkowski centers relates to the multiplicity of robust optimal solutions. Indeed, it has been observed (e.g., Iancu and Trichakis 2014) that different robust optimal solutions, although leading to the same worst-case cost, can provide very different average performance. In this section, we propose two methods to choose one center among the set of all Minkowksi centers and describe them in the particular case of polyhedra.

The first method computes a Minkowski center of the set of Minkowski centers. Let $\lambda^\star$ be the objective value of (3). The set of Minkowski centers of $\mathcal{C}$ can thus be described as

$$\mathcal{M}(\mathcal{C}) = \left\{ \boldsymbol{x} \;\middle|\; \begin{array}{l} \boldsymbol{x} \in \mathcal{C} \\ (1+\lambda^\star)\boldsymbol{x} - \lambda^\star \boldsymbol{y} \in \mathcal{C},\, \forall \boldsymbol{y} \in \mathcal{C} \end{array} \right\}.$$

Hence, by applying Proposition 3 to $\mathcal{M}(\mathcal{C})$, we can obtain a Minkowski center of $\mathcal{M}(\mathcal{C})$ by solving the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{v},\mu\geq 0} \mu \text{ s.t.} \quad & \boldsymbol{v}/(1+\mu) \in \mathcal{M}(\mathcal{C}), \\ & \boldsymbol{v} - \mu \boldsymbol{z} \in \mathcal{M}(\mathcal{C}),\ \forall \boldsymbol{z} \in \mathcal{M}(\mathcal{C}). \end{aligned}$$

The difficulty in the above formulation is that the description of $\mathcal{M}(\mathcal{C})$ contains robust constraints and appears at three different places in the optimization problem: as constraints on $\boldsymbol{v}/(1+\mu)$, as constraints in the uncertainty set ($\boldsymbol{z} \in \mathcal{M}(\mathcal{C})$), and as constraints that need to be "robustified" ($\boldsymbol{v} - \mu\boldsymbol{z} \in \mathcal{M}(\mathcal{C})$). Fortunately, we can obtain a tractable formulation in the case of polyhedra:

PROPOSITION 8. *Assume* $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b};\ \boldsymbol{C}\boldsymbol{x} \leq \boldsymbol{d}\}$ *and denote* $\lambda^\star$ *the objective value of* (4). *Define* $\delta_i := \min_{\boldsymbol{y}\in\mathcal{C}} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}$ *and* $\tilde{\delta}_i := \min_{\boldsymbol{y}\in\mathcal{M}(\mathcal{C})} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}$ *for* $i \in [m]$. *The point* $\boldsymbol{v}^\star/(1+\mu^\star)$, *with* $(\boldsymbol{v}^\star, \mu^\star)$ *solutions of*

$$\begin{aligned} \max_{\boldsymbol{v},\mu\geq 0} \mu \textit{ s.t.} \quad & \boldsymbol{A}\boldsymbol{v} = (1+\mu)\boldsymbol{b}, \\ & (1+\lambda^\star)\left(\boldsymbol{C}\boldsymbol{v} - \mu\tilde{\boldsymbol{\delta}}\right) \leq \boldsymbol{d} + \lambda^\star \boldsymbol{\delta}, \end{aligned}$$

is a Minkowski center of $\mathcal{M}(\mathcal{C})$.

Proof From Proposition 4, we have that the set of Minkowski centers of $\mathcal{C}$ is a polyhedron $\mathcal{M}(\mathcal{C}) = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}; \boldsymbol{C}\boldsymbol{x} \leq \tilde{\boldsymbol{d}}\}$, with

$$\tilde{\boldsymbol{d}} := \frac{1}{1+\lambda^\star}(\boldsymbol{d} + \lambda^\star \boldsymbol{\delta}).$$

In other words, $\mathcal{M}(\mathcal{C})$ is also a polyhedron defined with the same equality constraints as $\mathcal{C}$ and the same inequality constraints except with a different right-hand side vector $\tilde{\boldsymbol{d}}$.

Applying Proposition 4 to $\mathcal{M}(\mathcal{C})$, we can obtain a Minkowski center of $\mathcal{M}(\mathcal{C})$ by rescaling the solution of:

$$\max_{\boldsymbol{v},\mu\geq 0} \mu \text{ s.t.} \quad \boldsymbol{A}\boldsymbol{v} = (1+\mu)\boldsymbol{b}, \quad \boldsymbol{C}\boldsymbol{v} - \mu\tilde{\boldsymbol{\delta}} \leq \tilde{\boldsymbol{d}}.$$

$\square$

REMARK 2. For an unbounded set $\mathcal{C}$ (which admits an infinity of Minkowski centers), the behavior of our approach would again depend on $\mathcal{H}$, the recession cone of $\mathcal{C}$. If it is not a subspace, then $\mathcal{M}(\mathcal{C}) = \mathcal{C}$ and computing a Minkowski center of $\mathcal{M}(\mathcal{C})$ is not useful. If it is a subspace, then $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathcal{C}_0) + \mathcal{H}$ and $\mathcal{H}$ is also the recession cone of $\mathcal{M}(\mathcal{C})$, so Proposition 8 can be used to identify better Minkowski centers, but could never lead to a unique one.

Since Minkowski centers can be viewed as robust optimal solutions of a given optimization problem, the second method we propose to select one center is to consider Pareto robust optimal solutions, as defined in Iancu and Trichakis (2014).

DEFINITION 5. Consider a polyhedron $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b};\ \boldsymbol{C}\boldsymbol{x} \le \boldsymbol{d}\}$ with m inequality constraints. Denote $\delta_i := \min_{\boldsymbol{y}\in\mathcal{C}} \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{y}$ for $i \in [m]$ and $\lambda^\star$ the objective value of (4). Then, we call a solution of the optimization problem

$$\max_{\boldsymbol{x}} \boldsymbol{v}^\top(\boldsymbol{d} - (1+\lambda^\star)\boldsymbol{C}\boldsymbol{x} + \lambda^\star \boldsymbol{C}\bar{\boldsymbol{y}}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ (1+\lambda^\star)\boldsymbol{C}\boldsymbol{x} \le \boldsymbol{d} + \lambda^\star \boldsymbol{\delta},$$

for some $\bar{\boldsymbol{y}}$ in the relative interior of $\mathcal{C}$ and some valuation of the constraints $\boldsymbol{v} \in \mathbb{R}^m_+$, a Pareto-optimal Minkowski center of $\mathcal{C}$.

In other words, a Pareto-optimal Minkowski center is a center that maximizes the penalized sum of the slacks in the constraints $(1+\lambda^\star)\boldsymbol{C}\boldsymbol{x} - \lambda^\star \boldsymbol{C}\bar{\boldsymbol{y}} \le \boldsymbol{d}$, at some predefined point $\bar{\boldsymbol{y}}$.

REMARK 3. If $\mathcal{C}$ is unbounded and its recession cone is not a subspace, then we can find a non-trivial Pareto-optimal Minkowski center of $\mathcal{C}$ whenever the valuation vector $\boldsymbol{v}$ is chosen so that $v_i = 0$ if $\delta_i = -\infty$. If $\mathcal{C}$ is unbounded and its recession cone is a subspace, however, the set of Pareto-optimal Minkowski centers is invariant by translation by $\mathcal{H}$, hence it is unbounded.

3. Practical benefits of Minkowski centers

While Minkowski centers have mostly been regarded as theoretical objects, the previous section shows that it can be expressed as the solution of a tractable linear optimization problem for polyhedra. In this section, we illustrate numerically the tractability and potential practical benefits from using Minkowski centers (instead of available alternatives) in two popular algorithms.

3.1. Computational tractability

We first evaluate the numerical scalability of computing Minkowski centers of polyhedra and how it compares to other known and used centers, namely analytic and Chebyshev centers, on 78 polyhedra from the NETLIB library (Gay 1985) and 37 from the MIPLIB library. As reported in Table 2, these computational times are one order of magnitude higher than those needed to compute analytic and Chebyshev centers. To the best of our knowledge, our paper is the first to investigate the numerical tractability of Minkowski centers, although it has been extensively used for theoretical purposes. We also report measures of centrality: the measure of symmetry sym($\boldsymbol{x}$,$\mathcal{C}$), the depth, and the average sum of log-slacks $\frac{1}{m}\sum_{i\in[m]}\log(d_i - \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{x})$. These three metrics are maximized (by definition) by Minkowski, Chebyshev, and analytic centers, respectively. By reporting these measures, we want to emphasize how complex and ambiguous it is to properly define a center of a set and how varied the current definitions are. In the next two sections, we adopt a more pragmatic approach and evaluate the benefit from using Minkowski centers as initialization points of two numerical algorithms.

Table 2 Median (and inter-quartile range) for runtime and 3 performance metrics for analytic, Chebyshev and Minkowski centers. We report runtime (in seconds), the symmetry measure sym($\boldsymbol{x}$,$\mathcal{C}$) and the average sum of log-slacks $\frac{1}{m}\sum\log(d_i - \boldsymbol{e}_i^\top \boldsymbol{C}\boldsymbol{x})$.

Method	Runtime	Symmetry measure	Depth	Average sum of log slacks
Chebyshev	0.014 (0.036)	0.0 (0.004)	0.025 (0.598)	-1.584 (20.371)
Analytic	0.231 (0.349)	0.009 (0.075)	0.0 (0.148)	0.066 (7.172)
Minkowski	5.308 (23.278)	0.056 (0.13)	0.0 (0.109)	-0.756 (5.322)

3.2. Hit-And-Run algorithm

Hit-and-run is a standard algorithm for sampling random points from an arbitrary density on a high dimensional Euclidian space (see Chen and Schmeiser 1993, for a comparison of sampling schemes), initially proposed by Boneh and Golan (1979) and Smith (1984). In particular here, we apply it to sample points uniformly over a polyhedron $\mathcal{P}$.

The hit-and-run (HAR) algorithm starts at an initial point $\boldsymbol{x}_0 \in \mathcal{P}$ and generates a sequence $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ in $\mathcal{P}$ with random increments $\boldsymbol{x}_{m+1} - \boldsymbol{x}_m$. Precisely, at step m, we generate a random direction $\boldsymbol{d}_m$. The line passing through $\boldsymbol{x}_m$ with direction $\boldsymbol{d}_m$ *hits* the boundary of $\mathcal{P}$ at some points $\boldsymbol{y}_m^-$ and $\boldsymbol{y}_m^+$. We sample $\boldsymbol{x}_{m+1}$ uniformly over the segment $[\boldsymbol{y}_m^-,\boldsymbol{y}_m^+]$. Algorithm 2 in Appendix C.1 describes the algorithm in pseudo-code for a polyhedron $\mathcal{P}$ described as the intersection of halfspaces (see Bélisle et al. 1993, for its extension to generic compact convex sets). It was later shown to have polynomial mixing time for sampling from convex sets (Lovász 1999), and seems to be much faster in practice. We refer to Bélisle et al. (1998) for a careful review of the literature.

The sequence of points generated, $\boldsymbol{x}_0,\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$, is an ergodic Markov chain that geometrically converges to the uniform distribution over $\mathcal{P}$ (Chen and Schmeiser 1993, section 2-3). To estimate the expected value of some functional of $\tilde{\boldsymbol{x}}$, $\mathbb{E}[h(\tilde{\boldsymbol{x}})]$, using N uniformly sampled points from $\mathcal{P}$, two options are possible: (a) Run Algorithm 2 with m steps, N times and consider $\{\boldsymbol{x}_m^{(i)}, i\in[N]\}$; (b) Run Algorithm 2 with $m\times N$ steps and consider $\{\boldsymbol{x}_{im}, i\in[N]\}$. Generally speaking, for a fixed value of $m\times N$, option (b) will provide better point estimates but worse standard errors, due to auto-correlations between the samples (see Chen and Schmeiser 1993, section 5.1). In any case, it is crucial that the distribution of the sequence generated by the algorithm converges as fast as possible (in terms of number of steps m) towards the uniform distribution. Intuitively, starting from a "central" point $\boldsymbol{x}_0$ should speed up convergence.

Formally, we want to test the null hypothesis:

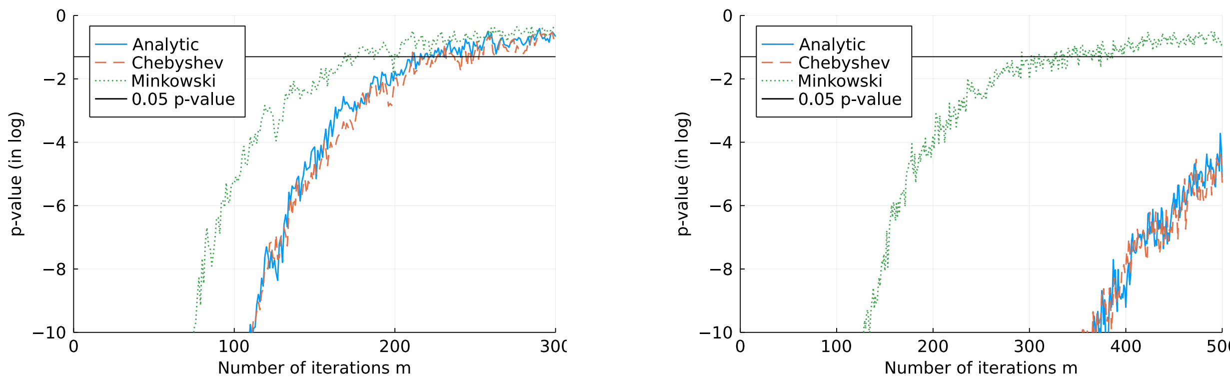
$$(H_0^m): \tilde{\boldsymbol{x}}_m \text{ is uniformly distributed on } \mathcal{P}$$

using an i.i.d. random sample of size $N = 5,000$. Daz et al. (2006) develop a method for testing this hypothesis called the distance to boundary (DB) test.

For a compact subset of $\mathbb{R}^n$, $\mathcal{C}$, define the distance of any point $\boldsymbol{x}$ to the boundary as $D(\boldsymbol{x}, \partial\mathcal{C}) = \min\{\|\boldsymbol{x} - \boldsymbol{y}\| \ : \ \boldsymbol{y} \in \partial\mathcal{C}\}$ and denote by R the maximum distance to the boundary that can be attained on $\mathcal{C}$, i.e., $R = \max\{D(\boldsymbol{x}, \partial\mathcal{C}) \ : \ \boldsymbol{x} \in \mathcal{C}\}$. The quantity R is sometimes called the depth of $\mathcal{C}$ and $D(\boldsymbol{x}, \delta\mathcal{C})/R$ the relative depth at $\boldsymbol{x}$. For a wide class of sets, namely sets that are "invariant by erosion", Daz et al. (2006) show that, under (H_0^m), the relative depth $\tilde{y}_m = D(\tilde{\boldsymbol{x}}_{\boldsymbol{m}}, \delta\mathcal{S})/R$ follows a beta distribution with parameters $(1, d)$, i.e., its cumulative distribution function is $y \mapsto 1-(1-y)^d$, for $y \in [0,1]$. Accordingly, we can test (H_0^m) by testing whether $\tilde{y}_m$ follows the right distribution via a Kolmogorov-Smirnov test. In particular, this result holds for a convex polyhedron circumscribed to a ball, i.e., defined as the intersection of halfspaces that are all tangent to a ball. We will use this type of polyhedra in our experiments.

For our experiment, we generate random convex polyhedra circumscribed to a ball in dimension $n \in \{10, 20, 50, 100\}$ (see Algorithm 3 in Appendix C). We run the HAR algorithm with different initial points $\boldsymbol{x}_0$. In particular, we compare the Minkowski, Chebyshev, and analytic centers. Figure 2 represents the p-value of the DB-Test for (H_0^m) as a function of the number of steps m. Recall that one can reject the null hypothesis (H_0^m) (i.e., conclude that the sample is not uniformly distributed) when the p-value is low. We also display a 0.05 cut-off. We observe that the hit-and-run algorithm initialized with a Minkowski center converges faster to a uniform distribution than when initialized with either analytic or Chebyshev centers. In particular, the benefit from a Minkowski center increases as the dimension of the space n increases.

We also compute the number of iterations m required for Algorithm 2 to achieve a p-value of 0.05, for each initialization point. Table 3 reports the average number of additional iterations required with the analytic center vs. the Minkowski center. Note that we had to limit the total number of iterations in the HAR algorithm (we used 500 in our experiments) and that on some instances, some methods failed to reach the 0.05 p-value target within the limit. In these cases, to allow for a fair comparison, we compare the number of iterations required to achieve a lower p-value target, equal to the best p-value achieved by the worst performing method. Table 3 confirms the benefit from the Minkowski center, especially in high dimensions. In low dimension, we observe that the



(a) $n = 50$ (b) $n = 100$

Figure 2 p**-value of a DB-Test for the hit-and-run algorithm, as the number of interactions** m **increases. Results are averaged over 20 random polyhedra defined as the intersection of 10 halfspaces.**

analytic center seems to perform better when more constraints define the polyhedron. We confirm these findings by doing a regression analysis of the number of additional iterations (in log terms) as a function of the dimension n and the number of halfspaces defining the polyhedron (see Table 6 in Appendix D.1). In Appendix D.1 Table 5, we conduct a similar analysis for the Chebyshev center and observe similar (though marginally stronger) benefits.

Table 3 **Number of additional iterations required by Algorithm 2 when initialized with the analytic center vs. the Minkowski center. We report the average number over 20 random polyhedra (and standard errors).**

	# halfspaces (p)				
Dimension (n)	10	20	30	40	50
10	0.3 (0.3)	-1.0 (0.5)	-3.1 (0.5)	-2.8 (0.7)	-3.2 (0.6)
20	4.1 (1.3)	3.8 (1.1)	1.6 (1.3)	-4.0 (1.2)	-5.6 (1.3)
50	47.9 (5.3)	69.5 (4.3)	61.8 (5.5)	54.9 (4.0)	44.7 (4.6)
100	283.6 (8.9)	362.1 (4.9)	362.0 (7.4)	375.4 (7.4)	376.1 (6.7)

3.3. Cutting-plane algorithm

Cutting-plane methods (CPMs) are a broad family of algorithms for solving convex or quasiconvex nondifferentiable optimization problems (see Elhedhli et al. 2009, for a comprehensive overview). In order to motivate the use of Minkowski centers within CPM schemes, we consider in this section the basic implementation of a CPM algorithm to minimize a piece-wise linear convex function over an intersection of ellipsoids and evaluate its performance on random instances by following the methodology of Boyd et al. (2008), depending on the definition of center used.

We consider a generic problem of the form

$$\min_{\boldsymbol{x},t} t \text{ s.t. } (\boldsymbol{x},t) \in \mathcal{C}, \tag{5}$$

where $\mathcal{C}$ is a convex set. Typically, (5) arises as the epigraph formulation of a constrained minimization problem. In our implementation, we will consider the minimization of a piecewise linear function over an intersection of ellipsoids, i.e.,

$$\min_{\boldsymbol{x},t} t \text{ s.t. } \boldsymbol{a}_i^\top \boldsymbol{x} + b_i \leq t, \forall i \in [m],$$
$$\|\boldsymbol{F}_i \boldsymbol{x} + \boldsymbol{g}_i\|_2 \leq 1, \forall i \in [k].$$

In order to apply the CPM described in Algorithm 1, three ingredients are needed: First, the ability to test whether the current solution is feasible, $(\boldsymbol{x}_k, t_k) \in \mathcal{C}$. Second, an oracle that, given an infeasible solution $(\boldsymbol{x}_k, t_k)$, provides a hyperplane that separates the current solution from the feasible set $\mathcal{C}$. In our case, we will simply consider the most violated constraint and obtain a separating hyperplane by linearizing it around the current solution. Finally, and most relevant to our experiments, we need a `query` function that returns a point from a given polyhedron. From a convergence perspective, it is understood that the query point should be "central" so that the volume of $\mathcal{P}_k$ decreases fast. In our experiments, we will numerically compare the convergence of this algorithm when a Minkowski, analytic, or Chebyshev center is used as a query point. We shall denote the variants as MC-, AC-, and CC-CPM algorithms. Regarding the termination criterion, we impose a limit on the total number of iterations (600 for MC-CPM and 60,000 for AC- and CC-CPM in our experiments) and the bound gap (10^{-4}). We consider instances in $n \in \{10, 20, 50\}$ dimensions, with $m \in \{100, 200, 500\}$ linear pieces and $k \in \{0, 1, 5, 10\}$ ellipsoids. As in Boyd et al. (2008), instances are generated randomly by sampling the entries of $\boldsymbol{a}_i$ and b_i independently from a standard normal distribution. We sample $\boldsymbol{F}_i \sim \mathcal{U}([0,1])^{n/2 \times n}$ and $\boldsymbol{g}_i$ uniformly in the unit ball, so that $\boldsymbol{x} = \boldsymbol{0}$ is always feasible. We use $\mathcal{P}_0 = \{(\boldsymbol{x}, t) : \boldsymbol{a}_i^\top \boldsymbol{x} + b_i \leq t, \forall i \in [m]\}$ as our initial polyhedron.

Figure 3 displays the convergence profile (in terms of the number of iterations) of the suboptimality gap, averaged over 20 instances in dimension $n = 50$ with $m = 500$ linear pieces. The left and right panels report results for instances with $k = 0$ and $k = 10$ second-order cone constraints respectively. First, we should emphasize that no method is a clear winner and that the quality of a query point depends heavily on the instance. For example, in the linear case ($k = 0$), the CPM with Chebyshev center demonstrates the fastest convergence, while it requires two orders of magnitude more iterations than MC- and AC-CPM with $k = 10$ SOC constraints. Second, we observe that Minkowski centers provide a competitive alternative to analytic or Chebyshev centers. In particular, when $k = 10$, it nearly halves the number of iterations compared with using analytic centers. However, in Figure 3, convergence is measured in terms of number of iterations. Since computing Minkowski centers requires 1 to 2 orders of magnitude more time and since the query function is invoked at each iteration, the reduction in number of iterations does not translate into a reduction in computational time (see Figure 7 in Appendix D.2).

Algorithm 1: Cutting-Plane Method (CPM) for solving (5)

Input: Initial polytope $\mathcal{P}_0$ enclosing $\mathcal{C}$.

Output: A solution $\boldsymbol{x}$ to (5).

`query` a point $(\boldsymbol{x}_0, t_0) \in \mathcal{P}_0$.

while *termination criterion not met* **do**

  if $(\boldsymbol{x}_k, t_k) \in \mathcal{C}$ **then**

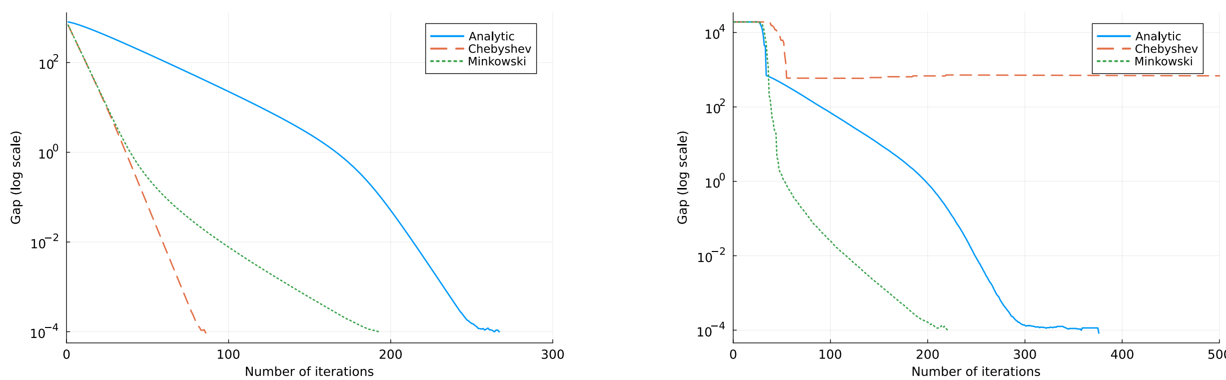
    set $\mathcal{P}_{k+1} = \mathcal{P}_k \cap \{(\boldsymbol{x}, t) \mid t \leq t_k\}$.

  else

    an oracle finds a separating hyperplane, i.e., $(\boldsymbol{a}, a, b)$ s.t. $\boldsymbol{a}^\top \boldsymbol{x}_k + a t_k > b$ but $\mathcal{C} \subseteq \{(\boldsymbol{x}, t) \mid \boldsymbol{a}^\top \boldsymbol{x} + at \leq b\}$.

    set $\mathcal{P}_{k+1} = \mathcal{P}_k \cap \{(\boldsymbol{x}, t) \mid \boldsymbol{a}^\top \boldsymbol{x} + at \leq b\}$.

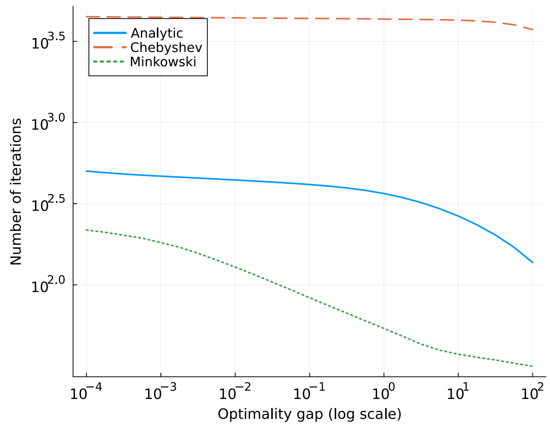
  `query` $(\boldsymbol{x}_{k+1}, t_{k+1}) \in \mathcal{P}_{k+1}$.



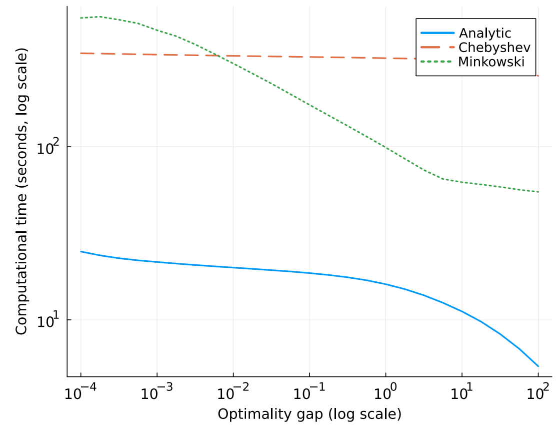
(a) $k = 0$ (b) $k = 10$

Figure 3 **Convergence profile (in terms of iterations) of the CPM for different query points. Results are averaged over 20 random instances in dimension** $n = 20$ **with** $m = 100$ **linear pieces.**

To verify this finding across various problem sizes, we compare, for different targets of optimality gap, the total number of iterations and total computational time needed by each algorithm to achieve this gap, averaged over all instances generated. As displayed in Figure 4, we observe that MC-CPM requires fewer iterations than other query strategies on average, across all tolerance level. However, given the additional computational burden, it is usually slower than AC-CPM, but competes with CC-CPM for moderate optimality gaps. While these metrics are averaged across all instances, we have observed in Figure 3 that the performance (and especially that of CC-CPM) depends greatly on the presence of SOC constraints, among other problem dimensions. To capture the effect of n, m, and k on the performance of each method, we report the average number of iterations and computational time to achieve a 10^{-2} optimality gap for each method, for all values of n and m, and some values of k, in Appendix D.2.



(a) Number of iterations



(b) Computational time

Figure 4 Average convergence speed of the CPM for different query points and different target optimality gaps.

Contrasting our findings on instances of the HAR algorithm (Section 3.2) and the CPM (Section 3.3), we conclude that there is no absolute best definition of centers but that the convergence of numerical algorithms can be significantly (and positively) impacted by choosing the 'right' one– for the algorithm and the particular instance at hand. On this matter, we argue that Minkowski centers, which have been largely overlooked for computational purposes, should be considered as a potential candidate. Given the additional burden of computing Minkowski centers, they could be most impactful for algorithms that query a strictly feasible point once (e.g., at initialization). Given the drastic reduction in the number of iterations we observe for the CPM, further work could also investigate the design of tailored solvers for computing Minkowski centers more efficiently.

4. Tractable approximations for projections of polyhedra

In this section, we consider the important case where the convex set $\mathcal{C}$ is the projection of a polyhedron. Precisely, we consider a polyhedron

$$\mathcal{P} = \left\{(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{n_x + n_z} \,|\, \boldsymbol{A}_x \boldsymbol{x} + \boldsymbol{A}_z \boldsymbol{z} = \boldsymbol{b},\ \boldsymbol{C}_x \boldsymbol{x} + \boldsymbol{C}_z \boldsymbol{z} \leq \boldsymbol{d}\right\},$$

and its projection onto the $\boldsymbol{x}-$space, i.e, $\mathcal{P}_{\boldsymbol{x}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \,|\, \exists \boldsymbol{z} \in \mathbb{R}^{n_z} \text{ s.t. } (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{P}\}$. In optimization, and combinatorial optimization in particular, such definition of sets as polyhedral projections are commonly referred to as extended or lifted formulations (see, e.g., Conforti et al. 2010).

The general approach for computing a (Minkowski) center for $\mathcal{P}_{\boldsymbol{x}}$ would be to first derive an explicit algebraic description of $\mathcal{P}_{\boldsymbol{x}}$ which does not rely on any additional variables $\boldsymbol{z}$, for instance by using Fourier-Motzkin elimination (FME, Motzkin 1936). However, the number of constraints resulting from this procedure grows exponentially in n_z. Moreover, FME introduces many redundant constraints which would need to be identified and removed or might negatively impact the quality of the analytic center. Hence, an algorithm that could compute a center of $\mathcal{P}_{\boldsymbol{x}}$ by working directly on its lifted description would be extremely tractable and valuable.

Also, the projection of a Minkowski center of $\mathcal{P}$ seems like a natural candidate for a Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$. However, we show that this approach fails.

LEMMA 3. *The projection onto the $\boldsymbol{x}$-space of a Minkowski center of $\mathcal{P}$ is not necessarily a Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$.*

Proof Our proof is based on the following counter example. In dimension n, consider the set $\mathcal{P}_n = \{\boldsymbol{x} \in \mathbb{R}^n_+ \,|\, \boldsymbol{e}^\top \boldsymbol{x} \leq 1\}$. The Minkowski center of $\mathcal{P}_n$ is the vector $\frac{1}{n+1}\boldsymbol{e}$ and its measure of symmetry is $\frac{1}{n}$. In particular, $\mathcal{P}_1 = [0,1]$ and its center is $1/2$. If we consider the projection of $\mathcal{P}_n$ onto the first coordinate, we recover $\mathcal{P}_1$. However, the projection of the Minkowski center is $1/(n+1) \neq 1/2$ for $n \geq 2$. □

Actually, the proof of Lemma 3 shows that the projection of the Minkowski center of $\mathcal{P}_n$ onto the first coordinate is not even a Helly center of the set $\mathcal{P}_1$.

Furthermore, one can show that projection can only improve symmetry:

LEMMA 4. *$\text{sym}(\mathcal{P}) \leq \text{sym}(\mathcal{P}_{\boldsymbol{x}})$.*

Proof Consider a center of Minkowski of $\mathcal{P}$, $(\boldsymbol{x}, \boldsymbol{z})$. Then, $\text{sym}(\boldsymbol{x}, \mathcal{P}_{\boldsymbol{x}}) \geq \text{sym}(\mathcal{P})$. □

4.1. Adjustable robust optimization reformulation

We now derive an analogous of Proposition 3 which applies to the case where the set is described as the projection of a polyhedron directly. In this case, computing a Minkowski center for $\mathcal{P}_{\boldsymbol{x}}$ is equivalent to an adjustable robust optimization (ARO) problem.

PROPOSITION 9. *Consider the set*

$$\mathcal{P}_{\boldsymbol{x}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{z} \in \mathbb{R}^{n_z} \,:\, \boldsymbol{A}_x \boldsymbol{x} + \boldsymbol{A}_z \boldsymbol{z} = \boldsymbol{b},\ \boldsymbol{C}_x \boldsymbol{x} + \boldsymbol{C}_z \boldsymbol{z} \leq \boldsymbol{d}\}.$$

Let $(\boldsymbol{w}^\star, \boldsymbol{z}^\star_{\boldsymbol{w}}, \lambda^\star)$ be solution of the adjustable robust optimization problem

$$\begin{aligned} \max_{\boldsymbol{w}, \boldsymbol{z}_w, \lambda \geq 0} \lambda \ \text{s.t.} \quad & \boldsymbol{A}_{\boldsymbol{x}} \boldsymbol{w} + \boldsymbol{A}_{\boldsymbol{z}} \boldsymbol{z}_w = (1+\lambda)\boldsymbol{b}, \\ & \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{w} + \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{z}_w \leq (1+\lambda)\boldsymbol{d}, \\ & \forall (\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P},\ \exists \boldsymbol{z} \,:\, (\boldsymbol{w} - \lambda \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{P}. \end{aligned} \tag{6}$$

Then $\boldsymbol{x}^\star = \boldsymbol{w}^\star/(1+\lambda^\star)$ is a Minkowski center for $\mathcal{P}_{\boldsymbol{x}}$.

Proof From the proof of Proposition 3, we know that the result holds with $(\boldsymbol{w}^\star, \lambda^\star)$ solution of

$$\begin{aligned} \max_{\boldsymbol{w}, \lambda \geq 0} \lambda \ \text{s.t.} \quad & \frac{\boldsymbol{w}}{1+\lambda} \in \mathcal{P}_{\boldsymbol{x}}, \\ & \boldsymbol{w} - \lambda \boldsymbol{y} \in \mathcal{P}_{\boldsymbol{x}},\ \forall \boldsymbol{y} \in \mathcal{P}_{\boldsymbol{x}}. \end{aligned}$$

By an appropriate rescaling of the additional variables,

$$\frac{\boldsymbol{w}}{1+\lambda} \in \mathcal{P}_{\boldsymbol{x}} \iff \exists \boldsymbol{z}_{\boldsymbol{w}} : \boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{z}_w = (1+\lambda)\boldsymbol{b}, \quad \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{z}_w \le (1+\lambda)\boldsymbol{d}.$$

Finally, the robust constraints can be rewritten as: $\forall(\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P},\ \exists \boldsymbol{z} : (\boldsymbol{w} - \lambda\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{P}$. □

The term "adjustable" comes from the fact that in the robust constraints, the additional variable $\boldsymbol{z}$, needed to certify that $\boldsymbol{w} - \lambda\boldsymbol{y} \in \mathcal{P}_{\boldsymbol{x}}$, can be adjusted to the uncertain parameter $\boldsymbol{y}$. Effectively, $\boldsymbol{z}$ is a function of $\boldsymbol{y}$ (and potentially of $\boldsymbol{z}_{\boldsymbol{y}}$ as well). Linear ARO optimization problems are NP-hard in general, as stated in Ben-Tal et al. (2004) and proved in Guslitser (2002, section 3.3). Instead of solving (6) exactly, we can obtain tractable approximations by restricting our attention to parametrized functional forms for $\boldsymbol{z}$ (as a function of $\boldsymbol{y}$ and $\boldsymbol{z}_{\boldsymbol{y}}$).

For instance, restricting our attention to $\boldsymbol{z}$ of the form $\boldsymbol{z} = \boldsymbol{z}_{\boldsymbol{w}} - \lambda\boldsymbol{z}_{\boldsymbol{y}}$, is equivalent to computing a Minkowski center for $\mathcal{P}$ and taking its projection onto the $\boldsymbol{x}$-space. Exploring a larger class of policies might lead to stronger formulations and better approximations. In the following section, we propose restricting the scope to general affine decision rules to derive tractable approximations of Minkowski centers, and a lower bound on the symmetry measure of $\mathcal{P}_{\boldsymbol{x}}$.

Remark 4. The optimization formulations and algorithms we present in this section are described for evaluating (or approximating) the symmetry measure of the set $\mathcal{P}_{\boldsymbol{x}}$. By adding the linear constraints $\boldsymbol{w} = (1+\lambda)\boldsymbol{x}$, they can be used to evaluate $\mathrm{sym}(\boldsymbol{x}, \mathcal{P}_{\boldsymbol{x}})$ for a given $\boldsymbol{x}$ as well.

4.2. Converging exact algorithm

The adjustable robust optimization problem (6) can be approximated by replacing the uncertainty set $\mathcal{P}$ by a finite number of scenarios and solving the fully adjustable problem (6) on this discrete uncertainty set. Precisely, given $\boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(k)} \in \mathcal{P}_{\boldsymbol{x}}$, we solve

$$\begin{aligned} \max_{\boldsymbol{w}, \boldsymbol{z}_w, \boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(k)}, \lambda \ge 0} \lambda \text{ s.t.} \quad & \boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{z}_w = (1+\lambda)\boldsymbol{b}, \\ & \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{z}_w \le (1+\lambda)\boldsymbol{d}, \\ & \forall i \in [k], (\boldsymbol{w} - \lambda\boldsymbol{y}^{(i)}, \boldsymbol{z}^{(i)}) \in \mathcal{P}. \end{aligned} \tag{7}$$

Since (7) is less constrained than (6), its value provides an upper bound on $\mathrm{sym}(\mathcal{P}_{\boldsymbol{x}})$. This generic approach in adjustable robust optimization was first presented by Hadjiyiannis et al. (2011) and is known to guarantee tight upper bounds for a large class of ARO problems and applications. We will later refer to this method via the acronym HGK, the initials of the authors.

Zeng and Zhao (2013) proposed a column-and-constraint generation scheme that converges to the exact robust solution of (6) by solving a sequence of problems of the form (7). Their method relies on an oracle that, for any solution $(\boldsymbol{w}, \lambda)$, can either certify that it satisfies the robust constraint "$\forall(\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P},\ \exists \boldsymbol{z} : (\boldsymbol{w} - \lambda\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{P}$", or returns a scenario $\bar{\boldsymbol{y}} \in \mathcal{P}_{\boldsymbol{x}}$ for which no such $\boldsymbol{z}$ exist. In

C.3, we describe how this oracle amounts to solving a non-convex maximization problem. At each iteration of their algorithm, a robust optimization problem with finite uncertainty of the form (7) is solved. If the resulting solution is robust feasible, then it is an optimal solution to the original problem (6). Otherwise, the oracle provides a new scenario $\bar{\boldsymbol{y}}$ to be added to (7). This algorithm is guaranteed to converge in a finite number of iterations, yet that can be linear in the number of extreme points of $\mathcal{P}_{\boldsymbol{x}}$ (Zeng and Zhao 2013, proposition 2). In practice, we typically impose a limit on the number of iterations or total computational time. In our setting, if the algorithm does not converge within the allocated budget, then the resulting solution $(\boldsymbol{w}, \lambda)$ is not robust feasible. In other words, the point $\boldsymbol{x} := \boldsymbol{w}/(1+\lambda)$ belongs to $\mathcal{P}_{\boldsymbol{x}}$ and is a candidate for being a Minkowski center, but the value of λ provided is an upper bound on both $\text{sym}(\boldsymbol{x}, \mathcal{P}_{\boldsymbol{x}})$ and $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$.

Alternatively, Zhen and den Hertog (2018) proposed a general method based on FME to solve adjustable linear robust optimization problems like (6). In our case, their approach is equivalent to using FME to derive an explicit description of $\mathcal{P}_{\boldsymbol{x}}$ and then compute the Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$ using the formulations from Section 2.

4.3. Approximations with computationable sub-optimality gaps

We restrict our attention to adjustable variables of the form

$$\boldsymbol{z} = \boldsymbol{Y}\boldsymbol{y} + \boldsymbol{Z}\boldsymbol{z}_{\boldsymbol{y}} + \boldsymbol{z}_0,$$

where $\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{z}_0$ are here-and-now decision variables. For instance, taking $\boldsymbol{Y} = \boldsymbol{0}$, $\boldsymbol{Z} = -\lambda \boldsymbol{I}$, and $\boldsymbol{z}_0 = \boldsymbol{z}_{\boldsymbol{w}}$ recovers the projection of a Minkowski center of $\mathcal{P}$. Among others, such linear decision rules (LDR in short, first proposed by Ben-Tal et al. 2004) are simple, tractable, and often enjoy strong empirical and theoretical performance for adjustable robust optimization problems (Bertsimas et al. 2010, Bertsimas and Goyal 2012, Housni and Goyal 2021). All in all, we solve

$$\begin{aligned} \max_{\boldsymbol{w}, \boldsymbol{z}_w, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{z}_0, \lambda \geq 0} \lambda \text{ s.t.} \quad & \boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{z}_w = (1+\lambda)\boldsymbol{b}, \\ & \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{w} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{z}_w \leq (1+\lambda)\boldsymbol{d}, \\ & \forall (\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P}, \ (\boldsymbol{w} - \lambda \boldsymbol{y}, \boldsymbol{Y}\boldsymbol{y} + \boldsymbol{Z}\boldsymbol{z}_{\boldsymbol{y}} + \boldsymbol{z}_0) \in \mathcal{P}. \end{aligned} \tag{8}$$

The objective value of the above optimization problem $\lambda^\star_{LDR}$ provides a lower bound on the actual symmetry of $\mathcal{P}_{\boldsymbol{x}}$, i.e., $\lambda^\star_{LDR} \leq \text{sym}(\mathcal{P}_{\boldsymbol{x}})$. Among others, Bertsimas et al. (2010), Ben-Ameur et al. (2018) show that linear decision rules are optimal (hence, the inequality is tight) when the uncertainty set (here, $\mathcal{P}$) is a standard simplex.

The robust constraints in (8) can be written explicitly

$$\begin{aligned} \forall (\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P}, \ & \boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{w} - \lambda \boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{y} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{Y}\boldsymbol{y} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{Z}\boldsymbol{z}_{\boldsymbol{y}} + \boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{z}_0 = \boldsymbol{b}, \\ & \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{w} - \lambda \boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{y} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{Y}\boldsymbol{y} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{Z}\boldsymbol{z}_{\boldsymbol{y}} + \boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{z}_0 \leq \boldsymbol{d}. \end{aligned}$$

They can then be enforced numerically either by adopting a cutting-plane approach or by computing the robust counterpart of each constraint separately via strong duality (Bertsimas et al. 2016), thus leading to a linear optimization problem. We implement the later approach for our numerical experiments.

To measure the quality of the approximation provided by using linear decision rules, we also derive an upper bound on $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$ using the generic approach described in the previous section. To identify the scenarios $\boldsymbol{y^j}, j \in [k]$, Hadjiyiannis et al. (2011) suggest considering each robust constraint and compute the binding scenarios for each of them, the decision variables being fixed. In our implementation, we follow their recommendation to obtain a finite number of scenarios and solve one instance of (7) to obtain a valid upper bound. We denote the objective value of (7) λ^k_{HGK} –or $\lambda^k_{HGK}(\boldsymbol{x})$ if we upper bound $\text{sym}(\boldsymbol{x}, \mathcal{P}_{\boldsymbol{x}})$ by adding the constraints $\boldsymbol{w} = (1+\lambda)\boldsymbol{x}$.

Remark 5. In Appendix C.4, we describe a potentially stronger approximation based on quadratic decision rules. Unfortunately, the robust counterpart of these problems can only be approximated by large semidefinite optimization problems, which we were unable to solve to reasonable accuracy, even for the smallest polyhedron we considered (dimension $n = 10$).

4.4. Numerical experiments

In this section, we evaluate the performance of our method for computing approximate values of the Minkowski measure (lower bounds via (8) and upper bounds via (7)) for polytopic projections.

First, we evaluate the quality of our ARO-based approximation on small instances where exact methods apply. We generate random polyhedra, following the same generation methodology as Section 3.2, in $n = 10$ dimensions and using $p = 10$ linear inequalities. For each polyhedron, we consider its projection onto the first n_x coordinates, $n_x \in [n]$. Hence, $n - n_x$ corresponds to the number of dimensions eliminated. We compute the approximate Minkowski center obtained by solving (8), and obtain λ^*_{LDR}, and λ^*_{HGK}. As a comparison, we apply two exact approaches: We perform a FME procedure to obtain an explicit description of the projected polyhedron and then compute its Minkowski center by solving (4). Alternatively, we implement the column-and-constraint generation (C&CG) algorithm of Zeng and Zhao (2013). We impose a 1000-iteration (resp. 2000-iteration) limit and use an exact spatial-branch-and-bound solver (resp. the heuristic linearization technique of Bertsimas et al. 2012) as the oracle.

Figure 5 compares the lower and upper bounds, λ^*_{LDR} and λ^*_{HGK}, with the exact value of $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$ for different values of n_x. Notice that $n_x = n = 10$ corresponds to the case $\mathcal{P} = \mathcal{P}_{\boldsymbol{x}}$ so we naturally expect $\lambda^*_{LDR} = \lambda^*_{HGK} = \text{sym}(\mathcal{P}_{\boldsymbol{x}})$. At the other extreme, when $n_x = 1$, $\mathcal{P}_{\boldsymbol{x}}$ is a segment, which is perfectly symmetric so one should conclude that $\text{sym}(\mathcal{P}_{\boldsymbol{x}}) = 1$. First of all, we observe that the lower bound, the upper bound and the exact value of symmetry $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$ are non-increasing with

n_x. In other words, projecting increases symmetry, which validates experimentally Lemma 4. We also observe on Figure 5 that our adjustable robust optimization approach provides valid and small (within 5%) intervals on $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$. In particular, the upper bound derived from (7) is almost tight. Further improvement should thus mainly come from improving the lower bound. Note, however, that the width of the interval $[\lambda^*_{LDR}, \lambda^*_{HGK}]$ does not necessarily imply a bound on the distance between the returned solution and the set of Minkowski centers for $\mathcal{P}_{\boldsymbol{x}}$, although intuition suggests the tighter the interval the closer the solution is to an actual Minkowski center (see Figure 9 in Appendix D.3).

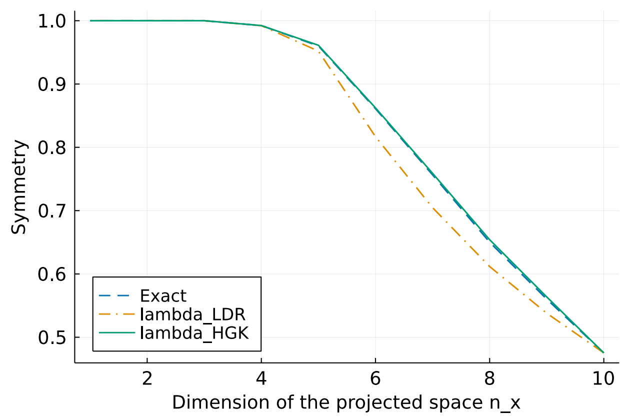


Figure 5 **Comparison of λ^*_{LDR}, λ^*_{HGK}, $\text{sym}(\mathcal{P}_{\boldsymbol{x}})$ for different values of n_x. Results are averaged over 20 polyhedra in dimension $n = 10$.**

In addition to providing high-quality solutions, the adjustable robust optimization approach is also significantly more computationally tractable than the exact approaches, as reported in Table 4. After accounting for the time required by the FME procedure, the ARO approach is 10^4 times faster than the FME-based approach. We should also mention that FME requires substantial memory and we could not perform simulations on larger instances with 16GB of RAM. The C&CG algorithm terminates in a short number of iterations for $n_x = 1, 2, 3$, yet requires 1 to 2 orders of magnitude more time than ARO. For larger values of n_x, however, the method often fails to converge within 1,000 (resp. 2,000) iterations, which represent 3-9 hours of computation. We should emphasize here that, when terminated early, the C&CG algorithm cannot provide any sub-optimality gap on the incumbent solution it returns.

We conduct further experiments in higher dimensions, $n \in \{10, 20, 50\}$, and for polyhedra defined with $p \in \{10, 20, 30, 40, 50\}$ inequalities. Figure 6 represents the distribution of the gap $(\lambda^*_{HGK} - \lambda^*_{LDR})/\lambda^*_{HGK}$ for varying values of n_x/n and varying values of n. We observe a similar qualitative behavior as in Figure 5: the width of the interval $[\lambda^*_{LDR}, \lambda^*_{HGK}]$ increases and then decreases with

Table 4 **Average runtimes (in seconds) for the adjustable robust optimization approach (both lower and upper bounds) compared with two exact approaches: FME followed by solving (4) and the column-and-constraint generation (C&CG) approach of Zeng and Zhao (2013) with an exact or heuristic oracle. Results are averaged over 20 iterations.**

	ARO		FME-based Method		C&CG Method	
n_x	(8)	(7)	FME	(4)	Exact	Heuristic
10	0.02	0.05	0.0	0.01	29,107.41	20,840.75
9	0.03	0.06	7.50	0.01	25,946.54	37,127.03
8	0.03	0.06	22.72	0.02	21,612.46	27,550.29
7	0.03	0.06	242.16	0.02	14,063.92	28,630.70
6	0.03	0.06	337.29	0.01	15,031.63	31,139.72
5	0.03	0.05	346.70	0.00	14,027.92	21,724.48
4	0.03	0.06	347.67	0.00	19,228.26	11,924.57
3	0.04	0.06	347.69	0.00	1.03	1.93
2	0.04	0.06	347.69	0.00	0.61	0.95
1	0.04	0.06	347.69	0.00	0.46	0.30

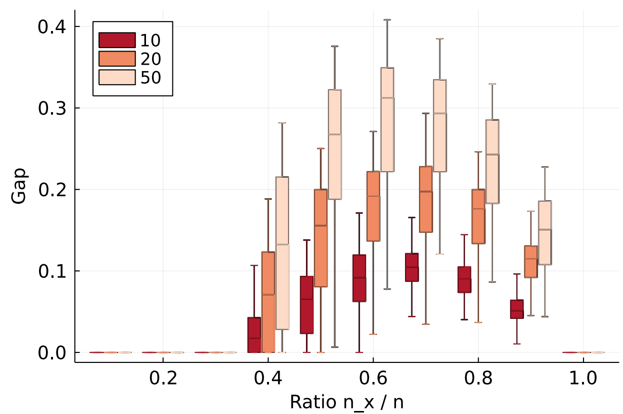


Figure 6 **Distribution (box plot) of the gap** $(\lambda^*_{HGK} - \lambda^*_{LDR})/\lambda^*_{HGK}$ **for varying values of** n_x/n **and varying values of** n.

n_x. A more detailed regression analysis (Table 10 in Appendix D.3) suggests that the gap scales as $0.9(n_x/n) - 0.7(n_x/n)^2$, hence maximized for $n_x/n \approx 0.64$, which is consistent with our observations. We also observe that the gap increases with the total dimension n and with the number of inequality constraints defining $\mathcal{P}$.

Regarding computational time, we observe that the effort required for solving (8) is fairly independent of the number of linear inequalities p but depends primarily on the dimension of the projected and of the full space, n_x and n respectively. On the contrary, solving (7) primarily depends on p and not on n_x/n, which is intuitive since the number of constraints p directly impacts the number of binding scenarios involved in (7). Tables 11 and 12 in Appendix D.3 summarize the average computational time required for both problems for varying input sizes.

5. Intersection of two ellipsoids

For $i=1,2$, we define the ellipsoid $\mathcal{E}_i=\{\boldsymbol{x}\in\mathbb{R}^n \mid \|\boldsymbol{A}_i(\boldsymbol{x}-\boldsymbol{x}_i)\|\leq 1\}$, where $\boldsymbol{x}_i\in\mathbb{R}^n$ and $\boldsymbol{A}_i\in\mathbb{R}^{n\times n}$. We are interested in computing a Minkowski center of the intersection of these two ellipsoids, $\mathcal{E}_1\cap\mathcal{E}_2$. In this case, we make an additional assumption on the matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$.

Assumption 1. *There exists an invertible matrix* $\boldsymbol{P}$ *such that, for* $i=1,2$, $\boldsymbol{A}_i^\top\boldsymbol{A}_i=\boldsymbol{P}^\top\boldsymbol{D}_i\boldsymbol{P}$ *for some diagonal matrix* $\boldsymbol{D}_i=\operatorname{diag}(\boldsymbol{d}_i)$.

When matrices $\boldsymbol{A}_1^\top\boldsymbol{A}_1$ and $\boldsymbol{A}_2^\top\boldsymbol{A}_2$ satisfy Assumption 1, we say that they are "diagonalized simultaneously by a congruence relationship" (Uhlig 1973). For instance, Assumption 1 is satisfied whenever one of the matrices $\boldsymbol{A}_i^\top\boldsymbol{A}_i$ is non-singular (Uhlig 1973, theorem 0.2). After a proper change of variable, $\boldsymbol{w}\leftarrow\boldsymbol{P}\boldsymbol{w}$ and $\boldsymbol{y}\leftarrow\boldsymbol{P}\boldsymbol{y}$, we can assume, without further loss of generality, that the matrices $\boldsymbol{A}_i$ are diagonal, i.e., that we have

$$\mathcal{E}_i=\{\boldsymbol{x}\in\mathbb{R}^n \mid \|\boldsymbol{D}_i^{1/2}(\boldsymbol{x}-\boldsymbol{x}_i)\|\leq 1\},$$

where $\boldsymbol{D}_i^{1/2}=\operatorname{diag}(\sqrt{d_1},\dots\sqrt{d_n})$. Let us denote $\boldsymbol{b}_i:=\boldsymbol{D}_i\boldsymbol{x}_i$ and $c_i:=\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i$. The objective of this section is to propose an efficient approach, based on second-order cone relaxation and bisection search, to obtain a lower bound on $\operatorname{sym}(\mathcal{E}_1\cap\mathcal{E}_2)$ together with an approximate Minkowski center. We also provide conditions (that can be numerically verified) under which our proposed approximation is tight.

Remark 6. Assumption 1 is a much weaker assumption than simultaneous diagonalizability, i.e., $\boldsymbol{A}_i=\boldsymbol{P}^{-1}\boldsymbol{D}_i\boldsymbol{P}$, $i=1,2$. If $\boldsymbol{A}_i^\top\boldsymbol{A}_i$, $i=1,2$, are simultaneously diagonalizable, then Assumption 1 is satisfied. The reverse implication is not true.

5.1. Second-order cone approximation

We start by reformulating the optimization problem defining Minkowski centers of $\mathcal{E}_1\cap\mathcal{E}_2$.

Lemma 5. *For* $\mathcal{C}=\mathcal{E}_1\cap\mathcal{E}_2$, *Problem* (2) *is equivalent to*

$$\begin{aligned}\max_{\substack{\boldsymbol{w},\boldsymbol{\xi},\lambda\geq 0\\ \boldsymbol{\eta}^\star}} \lambda \;\; s.t. \quad & \boldsymbol{d}_i^\top\boldsymbol{\xi}-2\boldsymbol{b}_i^\top\boldsymbol{w}+(1+\lambda)c_i\leq(1+\lambda), \quad \forall i\in\{1,2\},\\ & w_j^2\leq(1+\lambda)\xi_j, \quad \forall j\in[n],\\ & \|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\boldsymbol{x}_i)\|_2^2+\eta_i^\star(\boldsymbol{w},\boldsymbol{\lambda})\leq 1, \quad \forall i\in\{1,2\},\end{aligned}$$

where each $\eta_i^\star(\boldsymbol{w},\lambda)$, $i=1,2$, *is the objective value of a non-convex quadratic optimization problem:*

$$\begin{aligned}\eta_i^\star(\boldsymbol{w},\lambda)=\max_{\boldsymbol{y},\boldsymbol{z}} \lambda^2\boldsymbol{d}_i^\top\boldsymbol{z}-2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\boldsymbol{y} \;\; s.t. \;\; & \boldsymbol{d}_k^\top\boldsymbol{z}-2\boldsymbol{b}_k^\top\boldsymbol{y}\leq 1-c_k, \; \forall k\in\{1,2\},\\ & y_j^2=z_j, \; \forall j\in[n].\end{aligned} \tag{9}$$

The proof of Lemma 5 relies on simple algebraic manipulations on Problem (2) and is hence deferred to Appendix E.

The maximization problem defining $\eta_i^\star$ is not convex due to the quadratic equality constraints $z_j = y_j^2$. Instead, we now propose a valid convex upper bound on $\eta_i^\star$, under constraint qualification conditions.

Assumption 2. *There exists $\boldsymbol{x} \in \mathbb{R}^n$ such that, for all $i \in \{1,2\}$, $\|\boldsymbol{D}_i^{1/2}(\boldsymbol{x}-\boldsymbol{x}_i)\| < 1$.*

In other words, we assume that $\mathcal{E}_1 \cap \mathcal{E}_2$ has a non-empty relative interior.

Lemma 6. *Under Assumption 2, for each $i \in \{1,2\}$, we have $\eta_i^\star(\boldsymbol{w},\lambda) \le \eta_i(\boldsymbol{w},\lambda)$ with*

$$\begin{aligned}\eta_i(\boldsymbol{w},\lambda) = \min_{\boldsymbol{u}\in\mathbb{R}^n_+, \boldsymbol{v}\in\mathbb{R}^2_+, \boldsymbol{\theta}\in\mathbb{R}^n_+} \quad & v_1(1-c_1) + v_2(1-c_2) + \boldsymbol{e}^\top \boldsymbol{\theta} \\ \text{s.t. } & \lambda^2 \boldsymbol{d}_i - v_1 \boldsymbol{d}_1 + v_2 \boldsymbol{d}_2 + \boldsymbol{u} \le \boldsymbol{0}, \\ & (v_1 b_{1,j} + v_2 b_{2,j} - \lambda d_{i,j}(w_j - x_{i,j}))^2 \le u_j \theta_j, \ \forall j \in [n].\end{aligned} \tag{10}$$

Proof Fix $i \in \{1,2\}$. Relaxing the constraint $y_j^2 = z_j$ into the second-order cone constraints $y_j^2 \le z_j$ leads to $\eta_i^\star(\boldsymbol{w},\lambda) \le \eta_i(\boldsymbol{w},\lambda)$ with

$$\begin{aligned}\eta_i(\boldsymbol{w},\lambda) = \max_{\boldsymbol{y},\boldsymbol{z}} \ \lambda^2 \boldsymbol{d}_i^\top \boldsymbol{z} - 2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y} \text{ s.t. } & \boldsymbol{d}_k^\top \boldsymbol{z} - 2\boldsymbol{b}_k^\top \boldsymbol{y} \le 1 - c_k, \ \forall k \in \{1,2\} \quad [\boldsymbol{v}] \\ & y_j^2 \le z_j, \ \forall j \in [n]. \quad [\boldsymbol{u}]\end{aligned} \tag{11}$$

By introducing dual variables $(\boldsymbol{v},\boldsymbol{u})$ for the constraints in (11), we have that

$$\eta_i(\boldsymbol{w},\lambda) = \max_{\boldsymbol{y},\boldsymbol{z}\ge\boldsymbol{0}} \ \min_{\boldsymbol{u}\in\mathbb{R}^n_+, \boldsymbol{v}\in\mathbb{R}^2_+} \mathcal{L}(\boldsymbol{y},\boldsymbol{z};\boldsymbol{u},\boldsymbol{v}),$$

where $\mathcal{L}$ is the Lagrangian of the problem and is defined as

$$\mathcal{L}(\boldsymbol{y},\boldsymbol{z};\boldsymbol{u},\boldsymbol{v}) = \lambda^2 \boldsymbol{d}_i^\top \boldsymbol{z} - 2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y} + \sum_{k=1}^2 v_k (1 - c_k - \boldsymbol{d}_k^\top \boldsymbol{z} + 2\boldsymbol{b}_k^\top \boldsymbol{y}) + \sum_{j\in[L]} u_j (z_j - y_j^2).$$

Assumption 2 implies that there exists a strictly feasible solution to (11). Hence, strong duality holds and we can invert the order of the maximization and minimization. For a fixed $(\boldsymbol{u},\boldsymbol{v})$, by partially maximizing with respect to $\boldsymbol{z}$, we obtain

$$\max_{\boldsymbol{z}\ge\boldsymbol{0}} \left(\lambda^2 \boldsymbol{d}_i - \sum_{k=1}^2 v_k \boldsymbol{d}_k + \boldsymbol{u}\right)^\top \boldsymbol{z} = \begin{cases} 0 & \text{if } \lambda^2 \boldsymbol{d}_i - v_1 \boldsymbol{d}_1 - v_2 \boldsymbol{d}_2 + \boldsymbol{u} \le \boldsymbol{0}, \\ +\infty & \text{otherwise.} \end{cases}$$

Observe that for any $a, u \in \mathbb{R}$,

$$\max_y \ \left\{ay - uy^2\right\} = \begin{cases} \frac{a^2}{4u} & \text{if } u > 0 \\ +\infty & \text{otherwise} \end{cases} = \min_\theta \left\{ \ \theta \text{ s.t. } a^2 \le 4\theta u \right\}.$$

So, maximizing with respect to y_j,

$$\max_{y_j} \ 2\boldsymbol{e}_j^\top \left(-\lambda \boldsymbol{D}_i(\boldsymbol{w}-\boldsymbol{x}_i) + \sum_{k=1}^2 v_k \boldsymbol{b}_k\right) y_j - u_j y_j^2,$$

is equivalent to minimizing θ_j subject to the constraint detailed in the final formulation. ☐

Remark 7. Our approach could be generalized to matrices not satisfying Assumption 1. In this case, however, (9) would involve the additional variables $\boldsymbol{Z} : Z_{i,j} = y_i y_j$ and its convex relaxation (11) would be a semidefinite optimization problem (instead of second-order cone) over $(\boldsymbol{y}, \boldsymbol{Z}) : \boldsymbol{Z} \succeq \boldsymbol{y}\boldsymbol{y}^\top$. Hence, Assumption 1 substantially improves computational tractability without great loss of generality in the case of two matrices.

5.2. Final formulation and numerical algorithm

Overall, an approximate Minkowski center for $\mathcal{E}_1 \cap \mathcal{E}_2$ can be obtained by solving

$$\begin{aligned}
\max_{\substack{\boldsymbol{w},\boldsymbol{\xi},\lambda \\ (\eta_i, v_{1,i}, v_{2,i}, \boldsymbol{u}_i, \boldsymbol{\theta}_i)_{i=1,2}}} \lambda \text{ s.t. } \quad & \boldsymbol{d}_i^\top \boldsymbol{\xi} - 2\boldsymbol{b}_i^\top \boldsymbol{w} + (1+\lambda)c_i \leq (1+\lambda), \ \forall i \in \{1,2\}, \\
& w_j^2 \leq (1+\lambda)\xi_j, \ \forall j \in [n], \\
& \|\boldsymbol{D}_i(\boldsymbol{w} - \boldsymbol{x}_i)\|_2^2 + \eta_i \leq 1, \ \forall i \in \{1,2\}, \\
& v_{1,i}(1-c_1) + v_{2,i}(1-c_2) + \boldsymbol{e}^\top \boldsymbol{\theta}_i \leq \eta_i, \ \forall i \in \{1,2\}, \\
& \lambda^2 \boldsymbol{d}_i - v_{1,i}\boldsymbol{d}_1 - v_{2,i}\boldsymbol{d}_2 + \boldsymbol{u}_i \leq \boldsymbol{0}, \ \forall i \in \{1,2\}, \\
& (v_{1,i}b_{1,j} + v_{2,i}b_{2,j} - \lambda d_{i,j}(w_j - x_{i,j}))^2 \leq u_j \theta_j, \ \forall i \in \{1,2\}, j \in [n], \\
& \boldsymbol{\xi}, \boldsymbol{u}_i, \boldsymbol{\theta}_i \geq \boldsymbol{0}, \\
& \lambda, v_{1,i}, v_{2,i} \geq 0.
\end{aligned} \tag{12}$$

In this formulation, the variables η_i satisfy $\eta_i \geq \eta_i^\star(\boldsymbol{w}, \lambda)$ so any solution $(\boldsymbol{w}, \lambda)$ feasible for (12) is feasible for the original problem and solving (12) provides a lower bound on $\mathrm{sym}(\mathcal{E}_1 \cap \mathcal{E}_2)$. Solving (12) is challenging, however, due to the bilinear product of decision variables $\lambda d_{i,j}(w_j - x_{i,j})$ in the constraints. To do so efficiently, we propose to conduct a bisection search over λ. Indeed, $\lambda \in [0,1]$ and, for a fixed λ, (12) is a second-order cone optimization problem. Consequently, we can obtain an ϵ-approximation of the objective value of (12) after solving $\log_2(\varepsilon)$ second-order cone optimization problems.

5.3. Tightness

In this section, we fix $i \in \{1,2\}$ and analyze the tightness of the relaxation $\eta_i(\boldsymbol{w}, \lambda)$. First, we provide a (numerically verifiable) condition for our relaxation to be tight:

Proposition 10. *Fix $i \in \{1,2\}$. Let $(\boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{u}^\star, \boldsymbol{v}^\star)$ be a primal-dual optimal pair of* (11)-(10). *If $v_1^\star v_2^\star = 0$, then $\eta_i^\star(\boldsymbol{w}, \lambda) = \eta_i(\boldsymbol{w}, \lambda)$.*

Proof The result is a special case of Ben-Tal and den Hertog (2014, Theorem 7) after noting that assumption 5 in Ben-Tal and den Hertog (2014) is automatically satisfied in our case and that their assumption 6 is equivalent to the condition $v_1^\star v_2^\star = 0$. □

Second, we show that $\eta_i(\boldsymbol{w}, \lambda)$ provides a constant factor approximation on $\eta_i^\star(\boldsymbol{w}, \lambda)$ under the additional assumption that $\boldsymbol{0}$ lies in the relative interior of $\mathcal{E}_1 \cap \mathcal{E}_2$.

PROPOSITION 11. *Fix $i \in \{1,2\}$. Further assume that Assumption 2 is satisfied for $\boldsymbol{x} = \boldsymbol{0}$. Then,*

$$\eta_i^\star(\boldsymbol{w}, \lambda) \geq \left(\frac{1-\gamma}{\sqrt{2}+\gamma} \right)^2 \eta_i(\boldsymbol{w}, \lambda),$$

where $\gamma = \max_k \|\boldsymbol{D}_k^{1/2} \boldsymbol{x}_k\| = \max_k \sqrt{c_k} < 1$.

The proof of Proposition 11 relies on a similar construction as in Xia et al. (2021, Theorem 8). However, Xia et al. (2021) consider the special case of spheres, i.e., $d_{k,j} = 1$ for all $k \in \{1,2\}$, $j \in [p]$. We extend their proof technique to the non-isotropic case (see details in Appendix E) after making the following observation:

LEMMA 7. *There exists an optimal solution of* (11), $(\boldsymbol{y}^\star, \boldsymbol{z}^\star)$, *such that, for any $j \in [p]$,*

$$(y_j^\star)^2 < z_j^\star \implies d_{i,j} > 0.$$

Proof Let $(\boldsymbol{y}^\star, \boldsymbol{z}^\star)$ be an optimal solution of (11). Define $\mathcal{J} := \{ j \in [p] \mid (y_j^\star)^2 < z_j^\star \}$. We assume there exists $j \in \mathcal{J}$ such that $d_{i,j} = 0$. Let us define $\bar{\boldsymbol{y}} = \boldsymbol{y}^\star$ and

$$\bar{z}_{j'} = \begin{cases} z_{j'}^\star & \text{if } j' \neq j, \\ (y_j^\star)^2 & \text{if } j' = j. \end{cases}$$

Then, $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{z}})$ satisfies

$$\begin{aligned} \lambda^2 \boldsymbol{d}_i^\top \bar{\boldsymbol{z}} - 2(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \bar{\boldsymbol{y}} + &= -2(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}^\star + \lambda^2 \boldsymbol{d}_i^\top \boldsymbol{z}^\star, \\ \boldsymbol{d}_k^\top \bar{\boldsymbol{z}} - 2\boldsymbol{b}_k^\top \bar{\boldsymbol{y}} &\leq \boldsymbol{d}_k^\top \boldsymbol{z}^\star - 2\boldsymbol{b}_k^\top \boldsymbol{y}^\star \leq 1 - c_k, \ \forall k \in \{1,2\}. \end{aligned}$$

In other words, $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{z}})$ is feasible and optimal for (11) and $\{ j \in [p] \mid (\bar{y}_j)^2 < \bar{z}_j \} = \mathcal{J} \setminus \{j\}$. □

5.4. Discussion: Extension to the intersection of $m \geq 2$ ellipsoids

The approach we outlined in this section could be extended to the intersection of $m \geq 2$ ellipsoids. However, the conditions for Assumption 1 are more stringent in this case and overly restrictive (Grimus and Ecker 1986). Consequently, as mentioned in Remark 7, our approach in the case of m ellipsoids would entail relaxing each non-convex problem (9) into a semidefinite optimization problem, similar to the approach of Eldar et al. (2008) for the Chebyshev center of an intersection of ellipsoids. Eventually, the resulting formulation would be a semidefinite optimization problem with m $n \times n$ semidefinite matrices to optimize over, analogous to the one described in Ben-Tal et al. (2009, Chapter 7.2.1). Alternatively, one could follow the approach developed in Bertsimas et al. (2022) to derive safe approximation in the case of m ellipsoids of the form $\mathcal{E}_i = \{ \boldsymbol{y} \ : \ \|\boldsymbol{A}_i(\boldsymbol{y} - \boldsymbol{x}_i)\| \leq 1 \}$. The resulting safe approximation would involve an additional uncertain parameter, $\boldsymbol{V} \in \mathbb{R}^{n \times n}$, with bounded singular values. Again, the resulting robust counterpart is a semidefinite optimization problem that can be approximated by a second-order cone problem by bounding the matrix 2-norm by the Frobenius norm. Their approach could be also applied to derive approximate Minkowski center of arbitrary convex sets.

6. Conclusion

This paper provides a robust optimization formulation for the Minkowski centers of convex sets. Building up on this formulation, we propose tractable reformulations and efficient approximation techniques to numerically compute the Minkowski centers of a variety of sets (polyhedra, convex hulls, projections of polyhedra, intersections of ellipsoids). Theoretical benefits of Minkowski centers are numerous and well documented: They are geometrically defined and do not depend on the analytic description of the set (unlike the analytic center). Moreover, they naturally adapt to the dimension of the convex set and do not require the set to be fully dimensional (unlike centers of extremal ellipsoids such as Chebyshev centers). In addition, we illustrate their computational appeal by analyzing the algorithmic convergence of hit-and-run and cutting-plane method examples. While the actual gains ultimately depend on the particular algorithm and instance at hand, we believe our work sheds new and practical light on Minkowski centers and exposes their potential benefits as a computational tool.

References

Chaithanya Bandi and Dimitris Bertsimas. Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134(1):23–70, 2012.

Earl R. Barnes and Antonio Carlos Moretti. Some results on centers of polytopes. *Optimization Methods and Software*, 20(1):924, 2005.

Claude Bélisle, Arnon Boneh, and Richard J Caron. Convergence properties of hit–and–run samplers. *Stochastic Models*, 14(4):767–800, 1998.

Claude J. P. Bélisle, H. Edwin Romeijn, and Robert L. Smith. Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266, 1993.

Alexandre Belloni and Robert M. Freund. On the symmetry function of a convex set. *Mathematical Programming*, 111(12):5793, 2007.

Walid Ben-Ameur, Adam Ouorou, Guanglei Wang, and Mateusz Żotkiewicz. Multipolar robust optimization. *EURO Journal on Computational Optimization*, 6(4):395–434, 2018.

Aharon Ben-Tal and Dick den Hertog. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Mathematical Programming*, 143(1):1–29, 2014.

Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization.* Princeton University Press, 2009.

Aharon Ben-Tal, Dick den Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1):265–299, 2015.

Dimitris Bertsimas and Vineet Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 134(2):491–531, 2012.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

Dimitris Bertsimas, Dan A. Iancu, and Pablo A. Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

Dimitris Bertsimas, David B. Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011a.

Dimitris Bertsimas, Vineet Goyal, and Xu Andy Sun. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, 36(1):2454, 2011b.

Dimitris Bertsimas, Eugene Litvinov, Xu Andy Sun, Jinye Zhao, and Tongxin Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2012.

Dimitris Bertsimas, Vineet Goyal, and Brian Y. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, 150(2):281319, 2015.

Dimitris Bertsimas, Iain Dunning, and Miles Lubin. Reformulation versus cutting-planes for robust optimization. *Computational Management Science*, 13(2):195–217, 2016.

Dimitris Bertsimas, Dick den Hertog, Jean Pauphilet, and Jianzhe Zhen. Robust convex optimization: A new perspective that unifies and extends. *Mathematical Programming*, pages 1–42, 2022.

Robert G. Bland, Donald Goldfarb, and Michael J. Todd. The ellipsoid method: A survey. *Operations Research*, 29(6):1039–1091, 1981.

Arnon Boneh and A. Golan. Constraints redundancy and feasible region boundedness by random feasible point generator (RFPG). *Third European Congress on Operations Research - EURO III*, 1979.

Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Stephen P. Boyd, Lieven Vandenberghe, and Joëlle Skaf. Analytic center cutting-plane method. *Lecture Notes from Stanford University*, 2008.

Richard J. Caron, Harvey J. Greenberg, and Allen G. Holder. Analytic centers and repelling inequalities. *European Journal of Operational Research*, 143(2):268290, 2002.

Ming-Hui Chen and Bruce Schmeiser. Performance of the Gibbs, hit-and-run, and metropolis samplers. *Journal of Computational and Graphical Statistics*, 2:251–272, 1993.

Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8(1):1–48, 2010.

José Daz, Antonio Cuevas, and Francisco Grande. Testing multivariate uniformity: The distance-to-boundary method. *Canadian Journal of Statistics*, 34, 2006.

Yonina C Eldar, Amir Beck, and Marc Teboulle. A minimax Chebyshev estimator for bounded error estimation. *IEEE Transactions on Signal Processing*, 56(4):1388–1397, 2008.

Samir Elhedhli, Jean-Louis Goffin, and Jean-Philippe Vial. *Nondifferentiable optimization: cutting plane methods*, pages 2590–2595. Springer US, Boston, MA, 2009. ISBN 978-0-387-74759-0.

David M. Gay. Electronic mail distribution of linear programming test problems. *Mathematical Programming Society COAL Newsletter*, 13:10–12, 1985.

Chrysanthos E. Gounaris, Panagiotis P. Repoussis, Christos D. Tarantilis, Wolfram Wiesemann, and Christodoulos A. Floudas. An adaptive memory programming framework for the robust capacitated vehicle routing problem. *Transportation Science*, 50(4):1239–1260, 2016.

Walter Grimus and G. Ecker. On the simultaneous diagonalizability of matrices. Technical report, Vienna University (Austria). Institute für Theoretische Physik, 1986.

Elana Guslitser. Uncertainty-immunized solutions in linear programming. Master's thesis, Technion, Israeli Institute of Technology, IE&M faculty, 2002.

Osman Gler and Filiz Grtuna. Symmetry of convex sets and its applications to the extremal ellipsoids of convex bodies. *Optimization Methods and Software*, 27(45):735759, 2012.

Michael J. Hadjiyiannis, Paul J. Goulart, and Daniel Kuhn. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 7386–7391. IEEE, 2011.

Preston C. Hammer. The centroid of a convex body. *Proceedings of the American Mathematical Society*, 2 (4):522–525, 1951.

Omar El Housni and Vineet Goyal. On the optimality of affine policies for budgeted uncertainty sets. *Mathematics of Operations Research*, 46(2):674–711, 2021.

Pierre Huard. Resolution of mathematical programming with nonlinear constraints by the method of centers. *Nonlinear Programming*, pages 207–219, 1967.

Dan A Iancu and Nikolaos Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1): 130–147, 2014.

Florian Jarre. On the method of analytic centers for solving smooth convex programs. In *Optimization*, pages 69–85. Springer, 1989.

Fritz John. Extreme problems with inequalities as subsidiary conditions, studies and essays. In *Traces and Emergence of Nonlinear Programming*, pages 187–204. Springer Basel, 1948.

Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881–892, 2002.

Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, pages 302–311, 1984.

Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, Berlin, Germany, 2004.

Victor Klee. Convexity. In *Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society*, volume 7. American Mathematical Society, 1963.

László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

Hermann Minkowski. Allegemeine lehzätze über konvexe polyeder. *Gesellschaft Abhandlungen, Leipzog-Berlin*, 1:103–121, 1911.

Theodore S. Motzkin. *Beiträge zur Theorie der Linearen Ungleichungen*. Azriel Press, 1936.

Luis A Rademacher. Approximating the centroid is hard. In *Proceedings of the Twenty-Third Annual Symposium on Computational Geometry*, pages 302–305, 2007.

James Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40(1):59–93, 1988.

Cornelis Roos, Tamás Terlaky, and Jean-Philippe Vial. *Interior Point Methods for Linear Optimization*. Springer Science & Business Media, 2005.

Jacob T. Schwartz and Micha Sharir. A survey of motion planning and related geometric algorithms. *Artificial Intelligence*, 37(1-3):157–169, 1988.

Robert L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):12961308, 1984.

Gy Sonnevend. An "analytical center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In *System Modelling and Optimization*, pages 866–875. Springer, 1986.

Sergei P. Tarasov, Leonid G. Khachiyan, and Ivan I. Erlikh. The method of inscribed ellipsoids. *Soviet Mathematics-Doklady*, 37(1):226–230, 1988.

Michael J. Todd. On minimum volume ellipsoids containing part of a given ellipsoid. *Mathematics of Operations Research*, 7(2):253–261, 1982.

Frank Uhlig. Definite and semidefinite matrices in a real symmetric matrix pencil. *Pacific Journal of Mathematics*, 49(2):561–568, 1973.

Pravin M. Vaidya. A new algorithm for minimizing convex functions over convex sets. *Mathematical Programming*, 73(3):291–341, 1996.

Yong Xia, Meijia Yang, and Shu Wang. Chebyshev center of the intersection of balls: complexity, relaxation and approximation. *Mathematical Programming*, 187(12):287315, 2021.

David B Yudin and Arkadi S Nemirovskii. Informational complexity and efficient methods for the solution of convex extremal problems. *Matekon*, 13(2):22–45, 1976.

Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.

Jianzhe Zhen and Dick den Hertog. Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31–42, 2018.

Jianzhe Zhen, Dick den Hertog, and Melvyn Sim. Adjustable robust optimization via Fourier-Motzkin elimination. *Operations Research*, 66(4):1086–1100, 2018.

Appendix A: Analytic or Chebyshev centers are not Helly centers

Proof For Chebyshev centers, in dimension n, we consider the polytope $\mathcal{P}$ defined as the convex hull of the ∞-norm ball $\mathcal{B}_\infty = \{\boldsymbol{x} : \|\boldsymbol{x}\|_\infty \leq 1\}$ and the point $\boldsymbol{u} = t\boldsymbol{e}_1$ where $t > 0$ is a positive scalar and $\boldsymbol{e}_1$ is the first vector of the canonical basis. A Chebyshev center of $\mathcal{P}$ is $\boldsymbol{x}_{Cheb} = \boldsymbol{0}$. Consider the cord $[\boldsymbol{u}, \boldsymbol{v}]$ with $\boldsymbol{v} = -\boldsymbol{e}_1$. Then, $\frac{\|\boldsymbol{x}_{Cheb} - \boldsymbol{v}\|}{\|\boldsymbol{u} - \boldsymbol{v}\|} = \frac{1}{t+1} < \frac{1}{n+1}$ if we take $t > n$.

For the analytic center, in dimension n, we describe $\mathcal{B}_\infty$ as $\mathcal{B}_\infty = \{\boldsymbol{x} \ : -\boldsymbol{e} \leq \boldsymbol{x} \leq \boldsymbol{e}, \ \forall j \in [m], x_1 \leq 1\}$, where the constraint $x_1 \leq 1$ is redundantly added m times. The analytic center of $\mathcal{B}_\infty$ hence-described is $\boldsymbol{x}_a = -\frac{m}{m+2}\boldsymbol{e}_1$. If we consider the chord $[\boldsymbol{u}, \boldsymbol{v}] = [\boldsymbol{e}_1, -\boldsymbol{e}_1]$, we get $\frac{\|\boldsymbol{x}_a - \boldsymbol{v}\|}{\|\boldsymbol{u} - \boldsymbol{v}\|} = \frac{1}{m+2} < \frac{1}{n+1}$ if $m > n$. □

Appendix B: Robust perspective on Minkowski centers: Omitted proofs

This section details the proof of some of the results presented in Section 2.

B.1. Proof of Proposition 2

Proof Consider a chord $[\boldsymbol{u}, \boldsymbol{v}]$ passing through $\boldsymbol{x}$. Then, by definition of the symmetry measure (see Bertsimas et al. 2011b, for a formal proof)

$$\operatorname{sym}(\boldsymbol{x}, \mathcal{C}) \leq \min\left(\frac{\|\boldsymbol{x} - \boldsymbol{u}\|}{\|\boldsymbol{x} - \boldsymbol{v}\|}, \frac{\|\boldsymbol{x} - \boldsymbol{v}\|}{\|\boldsymbol{x} - \boldsymbol{u}\|}\right) \leq 1.$$

Assume without loss of generality that $r := \dfrac{\|\boldsymbol{x} - \boldsymbol{u}\|}{\|\boldsymbol{x} - \boldsymbol{v}\|} \leq 1$, then $\dfrac{\|\boldsymbol{x} - \boldsymbol{u}\|}{\|\boldsymbol{v} - \boldsymbol{u}\|} = \dfrac{r}{1+r}$. Since $r \in [1/n, 1]$ and $r \mapsto r/(1+r)$ is increasing in r,

$$\frac{1}{1+n} \leq \frac{\|\boldsymbol{x} - \boldsymbol{u}\|}{\|\boldsymbol{v} - \boldsymbol{u}\|} \leq \frac{1}{2} \leq \frac{n}{n+1}.$$

In other words, $\boldsymbol{x}$ is a Helly center of $\mathcal{C}$. □

B.2. Proof of Proposition 5

Proof We reformulate each constraint in (3) separately. By convexity

$$\boldsymbol{w} - \lambda \boldsymbol{y} \in \mathcal{C}, \forall \boldsymbol{y} \in \mathcal{C} \iff \boldsymbol{w} - \lambda \boldsymbol{x}_i \in \mathcal{C}, \forall i \in [m].$$

We can enforce the ith constraint by introducing additional variables $\boldsymbol{\nu}^i$ satisfying $\boldsymbol{w} - \lambda \boldsymbol{x}_i = \sum_{j \in [m]} \nu_j^i \boldsymbol{x}_j$. In particular, such a constraint ensures that

$$\frac{\boldsymbol{w}}{1+\lambda} = \frac{\lambda}{1+\lambda}\boldsymbol{x}_i + \sum_{j \in [m]} \frac{\nu_j^i}{1+\lambda}\boldsymbol{x}_j \in \operatorname{conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} = \mathcal{C}.$$

□

B.3. Proof of Proposition 6

Proof First, remark that $\mathcal{B}_p^+$ is permutation-invariant. According to Lemma (2), we can search for solutions of the form $\boldsymbol{w} = t\boldsymbol{e}$ without loss of optimality. Hence, we solve

$$\max_{\lambda \geq 0, t \geq 0} \lambda \text{ s.t. } n\left(\frac{t}{1+\lambda}\right)^p \leq 1, \quad t\boldsymbol{e} - \lambda \boldsymbol{y} \in \mathcal{B}_p^+, \forall \boldsymbol{y} \in \mathcal{B}_p^+.$$

Evaluating the robust constraint at $\boldsymbol{y} = (1, 0, ..., 0)$ and $\boldsymbol{y} = \boldsymbol{0}$, we get $t \geq \lambda$ and $nt^p \leq 1$ respectively, which leads to $\lambda \leq (1/n)^{1/p}$. Hence, we must have $\lambda^\star \leq (1/n)^{1/p}$. Finally, we verify that $(\lambda, t) = \left(\frac{1}{n^{1/p}}, \frac{1}{n^{1/p}}\right)$ is feasible. Indeed,

$$\frac{t}{1+\lambda} = \frac{1}{n^{1/p}+1} \leq \frac{1}{n^{1/p}},$$

and for every $\boldsymbol{y} \in \mathcal{B}_p^+$,

$$t - \lambda y_i \geq t - \lambda = 0, \text{ and } \sum_{i \in [n]} (t - \lambda y_i)^p = \lambda^p \sum_{i \in [n]} (1 - y_i)^p \leq \lambda^p n = 1,$$

so $t\boldsymbol{e} - \lambda \boldsymbol{y} \in \mathcal{B}_p^+$. □

B.4. Proof of Proposition 7

Proof Let $(\lambda, \boldsymbol{w})$ be an optimal solution for (3) for $\mathcal{C} = \mathcal{P}$. The robust constraints can be reformulated as

$$\begin{cases} \boldsymbol{w} \geq \lambda \boldsymbol{y}, & \forall \boldsymbol{y} \in \mathcal{P} \\ \boldsymbol{A}\boldsymbol{w} \leq \lambda \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}, & \forall \boldsymbol{y} \in \mathcal{P} \end{cases} \iff \begin{cases} \boldsymbol{w} \geq \lambda \boldsymbol{y}^\star \\ \boldsymbol{A}\boldsymbol{w} \leq \boldsymbol{b} \end{cases},$$

where the equivalence follows by evaluating each constraint at the worst-case scenario ($\boldsymbol{y}^\star$ and $\boldsymbol{0}$ respectively). By the non-negativity of $\boldsymbol{A}$, $\lambda \boldsymbol{A}\boldsymbol{y}^\star \leq \boldsymbol{A}\boldsymbol{w}$ so $\lambda \boldsymbol{A}\boldsymbol{y}^\star \leq \boldsymbol{b}$ and $\lambda \leq \lambda^\star$ (as defined in the statement of Proposition 7). Hence, $\lambda^\star$ constitutes an upper bound on the Minkowski measure of $\mathcal{P}$. It remains to prove that this bound is achievable.

To do so, it suffices to show that $(\lambda^\star, \boldsymbol{w}^\star)$ is feasible for (3):

$$\frac{\boldsymbol{w}^\star}{1+\lambda^\star} \geq \boldsymbol{0},$$
$$\boldsymbol{A}\frac{\boldsymbol{w}^\star}{1+\lambda^\star} = \frac{\lambda^\star}{1+\lambda^\star}\boldsymbol{A}\boldsymbol{y}^\star \leq \frac{1}{1+\lambda^\star}\boldsymbol{b} \leq \boldsymbol{b}.$$

Also, for every $\boldsymbol{y} \in \mathcal{P}$, $\boldsymbol{w}^\star - \lambda^\star \boldsymbol{y} \geq \boldsymbol{w}^\star - \lambda^\star \boldsymbol{y}^\star = \boldsymbol{0}$ and $\boldsymbol{A}(\boldsymbol{w}^\star - \lambda^\star \boldsymbol{y}) \leq \lambda^\star \boldsymbol{A}\boldsymbol{y}^\star - \lambda^\star \boldsymbol{A}\boldsymbol{0} \leq \lambda^\star \boldsymbol{A}\boldsymbol{y}^\star \leq \boldsymbol{b}$ by definition of $\lambda^\star$. □

B.5. Minkowski measure for a class of generalized ellipsoids

PROPOSITION 12. *Consider $\mathcal{E}_p^+ := \{\boldsymbol{x} \geq \boldsymbol{0} \mid \|\boldsymbol{A}\boldsymbol{x}\|_p \leq 1\}$ with $\boldsymbol{A} \in \mathbb{R}_+^{m \times n}$. For $i \in [n]$, define*

$$y_i^\star := \max_{\boldsymbol{y} \in \mathcal{E}_p^+} \boldsymbol{e}_i^\top \boldsymbol{y} = \frac{1}{\|\boldsymbol{A}^\top \boldsymbol{e}_i\|_p}.$$

Let $\lambda^\star = \dfrac{1}{\|\boldsymbol{A}\boldsymbol{y}^\star\|_p}$ and $\boldsymbol{w}^\star = \lambda^\star \boldsymbol{y}^\star$. Then, $(\lambda^\star, \boldsymbol{w}^\star)$ are the Minkowski measure and a scaled Minkowski center of $\mathcal{E}_p^+$.

Proof The proof structure is analogous to the proof of Proposition 7. Let $(\lambda, \boldsymbol{w})$ be an optimal solution of (3). We first provide an upper bound on the value of λ. By evaluating the robust (non-negativity) constraint in (3) at $\boldsymbol{y} = \boldsymbol{y}^\star$, we obtain $w_i \geq \lambda y_i^\star$ for every $i \in [n]$. Since the entries of $\boldsymbol{A}$ are non-negative, we get $\boldsymbol{A}\boldsymbol{w} \geq \lambda \boldsymbol{A}\boldsymbol{y}^\star$ and $\lambda \|\boldsymbol{A}\boldsymbol{y}^\star\|_p \leq \|\boldsymbol{A}\boldsymbol{w}\|_p$. Evaluating the robust (p-norm) constraint in (3) at $\boldsymbol{y} = \boldsymbol{0}$ yields $\|\boldsymbol{A}\boldsymbol{w}\|_p \leq 1$ so $\lambda \leq \lambda^\star$.

Finally, we verify that the proposed solution $(\lambda^\star, \boldsymbol{w}^\star)$ is feasible. Obviously, $\boldsymbol{w}^\star/(1+\lambda^\star) \geq \boldsymbol{0}$.

$$\left\| \boldsymbol{A} \frac{\boldsymbol{w}^\star}{1+\lambda^\star} \right\|_p = \frac{\lambda^\star}{1+\lambda^\star} \|\boldsymbol{A}\boldsymbol{y}^\star\|_p = \frac{1}{1+\lambda^\star} \leq 1.$$

Finally, for any $\boldsymbol{y} \in \mathcal{E}_p^+$, $\boldsymbol{w}^\star - \lambda^\star \boldsymbol{y} \geq \boldsymbol{w}^\star - \lambda^\star \boldsymbol{y}^\star = \boldsymbol{0}$ and

$$\|\boldsymbol{A}\,(\boldsymbol{w}^\star - \lambda^\star \boldsymbol{y})\|_p \leq \|\boldsymbol{A}\boldsymbol{w}^\star\|_p = 1.$$

□

Appendix C: Pseudo-codes and algorithmic details

We report here the detailed pseudocode of the hit-and-run algorithm and the random polyhedron generation methodology.

C.1. Hit-and-Run

Algorithm 2 describes the hit-and-run algorithm for a polyhedron defined as the intersection of halfspaces, $\mathcal{P} = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$.

Algorithm 2: Hit-and-run (HAR) algorithm

Input: A polytope $\mathcal{P} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$, a starting point $\boldsymbol{x}_0 \in \mathcal{P}$, number of iterations $m \in \mathbb{N}$

Output: Sample path $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathcal{P}$

Initialize $\boldsymbol{x}_0 \in \mathcal{P}$

for $i = 0, 1, ..., m-1$ **do**

- Generate a random direction on the hypersphere $\boldsymbol{d}_i = \frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|_2}$ where $\boldsymbol{u}_i \sim \mathcal{N}(\boldsymbol{0}_n, \boldsymbol{I}_n)$.
- Let $\lambda_k = \frac{b_k - \boldsymbol{A}_k^\top \boldsymbol{x}_i}{\boldsymbol{A}_k^\top \boldsymbol{d}_i}$ for each constraint k.
- Set $\lambda^+ = \min\{\lambda_k \mid \lambda_k \geq 0\}$, $\lambda^- = \max\{\lambda_k \mid \lambda_k \leq 0\}$.
- Define $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \lambda \boldsymbol{d}_i$, with $\lambda \sim \mathcal{U}([\lambda^-, \lambda^+])$.

C.2. Random polyhedron generation

Algorithm 3 presents the methodology we use to generate a random polyhedron circumscribed to a sphere of radius R. To avoid generating unbounded polyhedra, we add the constraints $-R \leq \boldsymbol{x} \leq R$. In our experiments, we typically take $R = 1000$, $n \in \{10, 20, 50, 100\}$, and $p \in \{10, 20, 30, 40, 50\}$.

Algorithm 3: Generation of a polyhedron circumscribed to a sphere

Input: Dimension n, number of tangents p, radius R

Output: Polyhedron $\mathcal{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid -R \leq \boldsymbol{x} \leq R;\ \boldsymbol{c}_i^\top \boldsymbol{x} \leq d_i, \forall i \in [p]\}$

for $i = 1, ..., p$ **do**
    Generate a random direction on the hypersphere $\boldsymbol{c}_i = R\frac{\tilde{\boldsymbol{u}}_i}{\|\tilde{\boldsymbol{u}}_i\|_2}$ where $\tilde{\boldsymbol{u}}_i \sim \mathcal{N}(\boldsymbol{0}_n, \boldsymbol{I}_n)$.
    Set $d_i = R$.

C.3. Projection of polyhedra: Separation oracle for the algorithm of Zeng and Zhao (2013)

For a given $(\boldsymbol{w}, \lambda)$, the adjustable robust constraint $\forall (\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P},\ \exists \boldsymbol{z}\ : (\boldsymbol{w} - \lambda \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{P}$ in (6) can be expressed as

$$\max_{(\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P}} \quad \min_{\substack{\boldsymbol{z}: \boldsymbol{A_x}(\boldsymbol{w} - \lambda \boldsymbol{y}) + \boldsymbol{A_z} \boldsymbol{z} = \boldsymbol{b} \\ \boldsymbol{C_x}(\boldsymbol{w} - \lambda \boldsymbol{y}) + \boldsymbol{C_z} \boldsymbol{z} \leq \boldsymbol{d}}} 0 \quad \leq 0,$$

where the value of saddle-point max-min problem is either 0 if $(\boldsymbol{w}, \lambda)$ and $+\infty$ otherwise. Alternatively, we can introduce dual variables $\boldsymbol{p}$ (resp. $\boldsymbol{q} \geq 0$) associated with the equality (resp. inequality) constraints in $\mathcal{P}$ and, by strong duality, reformulate the saddle-point problem as a non-convex maximization problem:

$$\max_{(\boldsymbol{y}, \boldsymbol{z}_y) \in \mathcal{P}} \quad \max_{\boldsymbol{q} \geq 0, \boldsymbol{p}: \boldsymbol{A_z}^\top \boldsymbol{p} - \boldsymbol{C_z}^\top \boldsymbol{q} = \boldsymbol{0}} \boldsymbol{p}^\top (\boldsymbol{b} - \boldsymbol{A_x} \boldsymbol{w} + \lambda \boldsymbol{A_x} \boldsymbol{y}) - \boldsymbol{q}^\top (\boldsymbol{d} + \boldsymbol{C_x} \boldsymbol{w} - \lambda \boldsymbol{C_x} \boldsymbol{y}). \tag{13}$$

The maximization problem above is non-convex due to bilinear products of decision variables in the objective and can be solved by spatial branch-and-bound (in our implementation, we will simply used the the commercial solver `Gurobi`), as a mixed-integer optimization problem, or via an alternating minimization heuristic (Bertsimas et al. 2012).

C.4. Projection of polyhedra: Semidefinite approximation for quadratic decision rules

In this section, we consider the problem of finding a Minkowski center of the projection of a polyhedron, as in Section 4, and use quadratic decision rules for approximating the adjustable robust optimization problem. For ease of notations, we consider a polyhedron described by linear *in*equalities only, i.e., $\mathcal{P}_{\boldsymbol{x}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{z} \in \mathbb{R}^{n_z}\ : \boldsymbol{C}_x \boldsymbol{x} + \boldsymbol{C}_z \boldsymbol{z} \leq \boldsymbol{d}\}$.

We restrict our attention to adjustable variables $\boldsymbol{z}$ in (6) that are quadratic functions of $(\boldsymbol{y}, \boldsymbol{z_y})$. Hence, without loss of generality, we assume that $\boldsymbol{z}$ can be expressed as an affine function of $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$, where $\boldsymbol{\omega} \in \mathcal{P}$ is simply the concatenation of $\boldsymbol{y}$ and $\boldsymbol{z_y}$, and $\boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\omega}^\top$. Formally, we restrict our attention to adjustable variables of the form

$$\boldsymbol{z} = \boldsymbol{z}_0 + \boldsymbol{Y}\boldsymbol{\omega} + \sum_{i=1}^{n_{\boldsymbol{z}}} \langle \boldsymbol{F}_i, \boldsymbol{\Omega} \rangle \boldsymbol{e}_i,$$

where $\boldsymbol{z}_0$, $\boldsymbol{Y}$, and the $\boldsymbol{F}_i$'s are decision variables that parametrize the quadratic policy.

Let us consider one particular constraint, $j \in [m]$, defining $\mathcal{P}$. We want that

$$\forall \boldsymbol{\omega} \in \mathcal{P}, \boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\omega}^\top,\ \boldsymbol{e}_j^\top \left(\boldsymbol{C_x} \boldsymbol{w} - \lambda \boldsymbol{C_x} \boldsymbol{\omega_x} + \boldsymbol{C_z} \boldsymbol{z}_0 + \boldsymbol{C_z} \boldsymbol{Y} \boldsymbol{\omega} + \sum_{i=1}^{n_{\boldsymbol{z}}} \langle \boldsymbol{F}_i, \boldsymbol{\Omega} \rangle \boldsymbol{C_z} \boldsymbol{e}_i \right) \leq \boldsymbol{e}_j^\top \boldsymbol{d},$$

which we concisely write

$$\boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{w} + \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{z}_0 + \sup_{\boldsymbol{\omega} \in \mathcal{P}, \boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\omega}^\top} \left\{ \boldsymbol{a}^\top \boldsymbol{\omega} + \langle \boldsymbol{A}, \boldsymbol{\Omega} \rangle \right\} \le \boldsymbol{e}_j^\top \boldsymbol{d}, \tag{14}$$

with $\boldsymbol{a}^\top := -\lambda \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{x}} + \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{Y}$, and $\boldsymbol{A} := \sum_{i=1}^{n_{\boldsymbol{z}}} \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{e}_i$.

The inner-maximization problem is challenging due to the non-convex constraints $\boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\omega}^\top$. Instead, we relax this constraint by imposing the semidefinite constraint $\boldsymbol{\Omega} \succeq \boldsymbol{\omega}\boldsymbol{\omega}^\top$. In addition, the vector $\boldsymbol{\omega}$ must satisfy $\boldsymbol{C}\boldsymbol{\omega} \le \boldsymbol{d}$, which in turns yield a linear constraint on $(\boldsymbol{\omega}, \boldsymbol{\Omega})$:

$$(\boldsymbol{d} - \boldsymbol{C}\boldsymbol{\omega})(\boldsymbol{d} - \boldsymbol{C}\boldsymbol{\omega})^\top = \boldsymbol{d}\boldsymbol{d}^\top - (\boldsymbol{C}\boldsymbol{\omega}\boldsymbol{d}^\top + \boldsymbol{d}\boldsymbol{\omega}^\top \boldsymbol{C}^\top) + \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}^\top \ge \boldsymbol{0}.$$

Formally, we replace the robust constraint (14) by the following safe approximation

$$\boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{w} + \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{z}_0 + \sup_{(\boldsymbol{\omega}, \boldsymbol{\Omega}) \in \mathcal{O}} \left\{ \boldsymbol{a}^\top \boldsymbol{\omega} + \langle \boldsymbol{A}, \boldsymbol{\Omega} \rangle \right\} \le \boldsymbol{e}_j^\top \boldsymbol{d},$$

with

$$\mathcal{O} := \left\{ (\boldsymbol{\omega}, \boldsymbol{\Omega}) \,\middle|\, \begin{array}{ll} \boldsymbol{d} & \ge \boldsymbol{C}\boldsymbol{\omega} \\ \boldsymbol{\Omega} & \succeq \boldsymbol{\omega}\boldsymbol{\omega}^\top \\ \boldsymbol{d}\boldsymbol{d}^\top & \ge (\boldsymbol{C}\boldsymbol{\omega}\boldsymbol{d}^\top + \boldsymbol{d}\boldsymbol{\omega}^\top \boldsymbol{C}^\top) - \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}^\top \end{array} \right\}.$$

Finally, by weak duality, the inner maximization problem can be upper bounded by its dual:

$$\min_{\begin{pmatrix} \boldsymbol{R} & \boldsymbol{r} \\ \boldsymbol{r}^\top & r_0 \end{pmatrix} \succeq \boldsymbol{0}} \min_{\substack{\boldsymbol{p} \ge 0 \\ \boldsymbol{Q} \ge 0}} \quad r_0 + \boldsymbol{d}^\top \boldsymbol{p} + \langle \boldsymbol{d}\boldsymbol{d}^\top, \boldsymbol{Q} \rangle \text{ s.t. } \boldsymbol{C}^\top \boldsymbol{q} + (\boldsymbol{Q} + \boldsymbol{Q}^\top) \boldsymbol{C}^\top \boldsymbol{d} = \boldsymbol{a} + 2\boldsymbol{r}$$

$$\boldsymbol{C}^\top \boldsymbol{Q} \boldsymbol{C} = \boldsymbol{A} + \boldsymbol{R}.$$

We then solve the overall robust optimization by adding the dual variables as additional decision variables in our original problem in $(\boldsymbol{w}, \lambda)$.

Appendix D: Additional numerical results

In this section, we provide additional supporting evidence to our numerical experiments.

D.1. Convergence of the Hit-And-Run algorithm

In Section 3.2, we quantify the benefit from using a Minkowski center on the convergence of the HAR algorithm. In particular, we compute the number of iterations m required for the DB-test to achieve a p-value of 0.05.

Table 5 reports the average number of additional iterations required when using a Chebyshev center vs. a Minkowski center. Table 6 reports the results from a regression analysis predicting the additional number of iterations required (in log terms) when using the analytic and Chebyshev center as a function of the problem size, i.e., the dimension n and the number of halfspaces defining the polyhedron p.

Table 5 Number of additional iterations required by Algorithm 2 when initialized with a Chebyshev center vs. a Minkowski center. We report the average number over 20 random polyhedra (and standard errors).

	# halfspaces (p)				
Dimension (n)	10	20	30	40	50
10	1.5 (0.5)	1.4 (0.5)	0.0 (0.4)	0.4 (0.6)	0.2 (0.7)
20	6.1 (1.3)	7.5 (1.3)	6.4 (1.1)	3.9 (1.2)	2.6 (0.9)
50	58.4 (3.8)	78.0 (4.8)	69.9 (6.6)	70.8 (5.6)	61.2 (5.2)
100	284.1 (9.4)	381.7 (5.4)	389.8 (6.6)	395.8 (5.6)	397.0 (4.7)

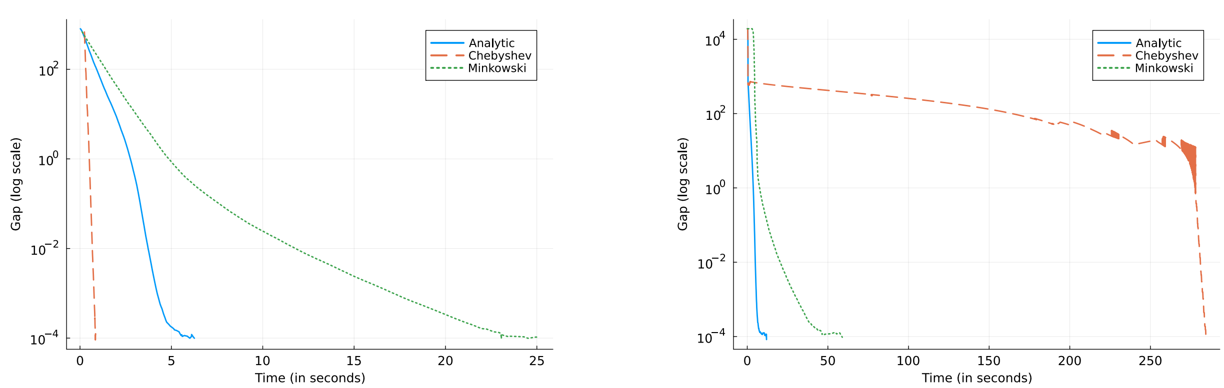
Table 6 Regression analysis of the benefit from using the Minkowski center to initialize Algorithm 2. The outcome variable is the number of iterations saved (in log terms).

	Analytic		Chebyshev	
	Coefficient (SE)	p-value	Coefficient (SE)	p-value
(Intercept)	2.542 (0.043)	$< 10^{-16}$	2.651 (0.029)	$< 10^{-16}$
Dimension n	0.035 (0.001)	$< 10^{-16}$	0.033 (0.003)	$< 10^{-16}$
# halfspaces p	-0.004 (0.001)	$2 \cdot 10^{-4}$		
Adjusted R^2	0.9374		0.9668	

Number of observations: 400

D.2. Convergence of the Cutting-Plane Method

Figure 3 in Section 3.3 displays a convergence profile in terms of the number of iterations of the CPM with different query point methods. However, since computing Minkowski centers requires 1 to 2 orders of magnitude more time and since the query function is invoked at each iteration of the algorithm, a reduction in number of iterations might not translate into a reduction in computational time. Figure 7 displays the same convergence plot as in Figure 3, except that convergence is measured in terms of computational time. We observe that AC-CPM is systematically faster than MC-CPM, although MC-CPM requires less iterations. MC-CPM, on the other hand, improves over CC-CPM whenver the number of SOC constraints $k > 0$.



(a) $k = 0$ (b) $k = 10$

Figure 7 Convergence profile (in terms of computational time) of the CPM for different query points. Results are averaged over 20 random instances in dimension $n = 20$ with $m = 100$ linear pieces.

In Section 3.3, we observed that initializing the CPM with a Minkowski center provides generally faster convergence in terms of number of iterations than with an analytic or Chebyshev center, but that using an analytic center is typically the fastest in terms of overall compuational time. To better understand how the dimension of the space n, the number of linear pieces in the objective m, and the number of SOC constraints in the feasible space k impact the convergence and scalability of each approach, we report the number of iterations and computational time required to achieve a 10^{-2} optimality gap, for all values of n and m and for all three query methods, for $k=0$ in Table 7 and $k=10$ in Table 8. Performance metrics are averaged over 20 random replications of the same instance.

		# Iterations			Time (in s)		
m	n	Analytic	Chebyshev	Minkowski	Analytic	Chebyshev	Minkowski
10	100	197.25	46.2	59.6	3.11	1.78	4.03
	200	321.45	47.25	53.95	8.25	0.43	8.69
	500	643.25	47.8	49.85	32.01	0.8	37.87
20	100	213.65	59.65	95.65	3.72	0.68	11.81
	200	340.05	60.35	81.3	10.17	0.76	27.93
	500	661.2	61.1	70.35	38.11	1.65	123.4
50	100	281.44	97.22	354.44	6.4	2.93	131.46
	200	387.85	90.8	172.55	14.14	2.24	197.14
	500	718.8	89.0	125.95	48.65	4.17	770.18

Table 7 **Number of iterations and computational time for CPM to achieve 10^{-2} optimality gap, on instances with $k=0$, for different types of query points**

		# Iterations			Time (in s)		
m	n	Analytic	Chebyshev	Minkowski	Analytic	Chebyshev	Minkowski
10	100	214.25	3113.05	62.95	3.59	74.64	4.67
	200	336.65	4190.7	59.7	8.78	125.88	10.07
	500	656.45	2688.8	58.25	33.36	92.15	43.54
20	100	247.95	4565.65	114.3	4.97	169.32	20.23
	200	371.55	4351.9	100.6	11.01	184.0	41.48
	500	696.05	3853.2	93.15	40.32	211.74	167.45
50	100	349.65	13001.3	419.0	10.05	1255.26	996.22
	200	460.55	13431.25	332.2	17.95	1241.44	1010.27
	500	788.35	12033.85	251.1	58.53	1335.14	2101.95

Table 8 **Number of iterations and computational time for CPM to achieve 10^{-2} optimality gap, on instances with $k=10$, for different types of query points**

D.3. Approximation for projections of polyhedra

In Section 4.4, we evaluate numerically the relevance of our approximation to the center of a polytopic projection. Our method provides both a lower and an upper bound on the true symmetry of the projection, $\mathrm{sym}(\mathcal{P}_{\boldsymbol{x}})$.

On small instances ($n=10$,$m=10$), we were able to compute exactly a Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$ by first obtaining an explicit description of this polyhedron via FME and then solving (4). Following the approach in

Zhen et al. (2018), we implement an iterative FME procedure with two steps: a variable elimination step that eliminates the n_x+1th variable from all the constraints, followed by a screening step that removes redundant constraints. Table 9 reports the computational time and the number of constraints created by the FME procedure. As displayed in Table 9, the redundant constraint screening step is computationally expensive but drastically reduces the number of constraints in our formulation, which would otherwise exponentially grow with $n-n_x$.

Table 9 Average number of constraints and runtime for after each step of the FME procedure. Results are averaged over 20 iterations.

	Variable Elimination		Redundant Constraint Screening	
n_x	# New Constraints	Runtime	# New Constraints	Runtime
9	34.0	2.9	32.1	4.6
8	264.9	0.2	71.0	15.0
7	1219.6	0.0	66.6	219.4
6	980.0	0.0	26.9	95.1
5	213.4	0.0	5.4	9.4
4	34.0	0.0	1.0	1.0
3	2.0	0.0	0.0	0.0
2	1.0	0.0	0.0	0.0
1	1.0	0.0	0.0	0.0
0	1.0	0.0	0.0	0.2

We also implemented the column-and-constraint generation of Zeng and Zhao (2013) with 1,000 (resp. 2,000) iterations and a spatial branch-and-bound solver (resp. a linearization heuristic) as the oracle. Figure 8 represents the proportion of instances solved for different projected dimension n_x. We observe that, even in these 10-dimensional examples, the C&CG algorithm fails to systematically converge within the allocated iteration budget for $n_x>3$. For $n_x\geq 5$, for example, termination rate for both implementations does not exceed 10%.

Figure 9 displays the distance between the approximate Minkowski center obtained by solving (8) to *one* Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$, for different values of n_x and $n=10$. The distance is normalized by the depth of the original polyhedron $\mathcal{P}$, i.e., the radius of the inscribed sphere in this case. Comparing Figure 9 with Figure 5 partially corroborates the intuition that the quality of our approximation in terms of symmetry measure (i.e., the width of the interval $[\lambda^*_{LDR},\lambda^*_{HGK}]$) is related with the quality of the approximation in terms of Minkowski center.

To further quantify the dependency of our adaptivity gap $(\lambda^*_{HGK}-\lambda^*_{LDR})/\lambda^*_{HGK}$ on characteristics of the polyhedron $\mathcal{P}$ and its projection $\mathcal{P}_{\boldsymbol{x}}$, we conduct further experiments in higher dimensions, $n\in\{10,20,50\}$, and for polyhedra defined with $p\in\{10,20,30,40,50\}$ inequalities. We perform a regression analysis, regressing $(\lambda^*_{HGK}-\lambda^*_{LDR})/\lambda^*_{HGK}$ over the dimensions of the problem, and report its results in Table 10. We observe that the gap generally increases with the dimension n and the number of inequalities defining the polyhedron m. Yet, the fraction of projected dimensions n_x/n seems to have a non-monotonous impact on the gap, first increasing then decreasing, thus confirming the behavior depicted in Figure 6.

Finally, Tables 11 and 12 summarize the average computational time required for solving (8) (the lower bound) and (7) (the upper bound) respectively, for varying input sizes.

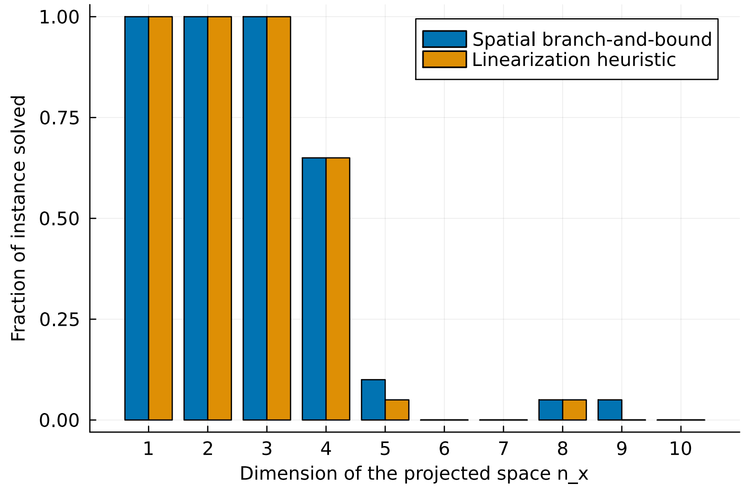


Figure 8 **Fraction of instances solved by optimality by the column-and-constraint generation approach of Zeng and Zhao (2013).**

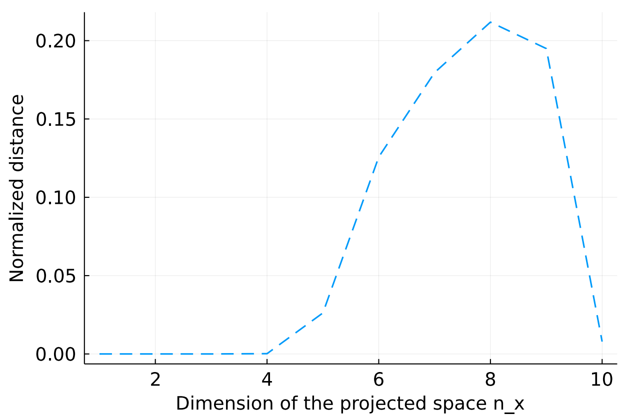


Figure 9 **Average distance between the solution of** (8) **and a Minkowski center of** $\mathcal{P}_{\boldsymbol{x}}$**. The distance is normalized by the depth of the original polyhedron** $\mathcal{P}$**.**

Table 10 **Regression analysis of the adaptivity gap** $(\lambda^*_{HGK} - \lambda^*_{LDR})/\lambda^*_{HGK}$ **depending on characteristics of the polyhedron.**

	Coefficient	p-value
(Intercept)	-0.249	$< 10^{-16}$
Dimension n	0.002	$< 10^{-16}$
# halfspaces p	0.002	$< 10^{-16}$
n_x/n	0.891	$< 10^{-16}$
$(n_x/n)^2$	-0.722	$< 10^{-16}$
Adjusted R^2	0.478	

Number of observations: 3,000

Table 11 Average computational time (in seconds) for solving (8) as a funtion of n and n_x/n. Results are averaged over $20\times 5 = 100$ polyhedra.

n	n_x/n								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.098	0.156	0.199	0.163	0.175	0.149	0.134	0.088	0.078
20	0.797	0.523	0.548	0.439	0.381	0.416	0.309	0.337	0.375
50	26.315	18.842	14.932	15.044	12.382	10.916	10.178	8.689	5.3

Table 12 Average computational time (in seconds) for solving (7) as a funtion of n and p. Results are averaged over $20\times 10 = 200$ polyhedra.

	m				
n	10	20	30	40	50
10	0.065	0.161	0.306	0.522	0.807
20	0.195	0.516	1.049	1.807	2.721
50	1.56	4.413	8.334	13.47	20.12

Appendix E: Intersection of ellipsoids: Omitted proofs

We detail the proofs of Section 5 in this section.

E.1. Proof of Lemma 5

Proof Problem (2) is equivalent to

$$\max_{\boldsymbol{w},\lambda\geq 0} \lambda \text{ s.t. } \quad \frac{\boldsymbol{w}}{1+\lambda}\in\mathcal{E}_i, \quad \forall i\in\{1,2\},$$
$$\max_{\boldsymbol{y}\in\mathcal{E}_1\cap\mathcal{E}_2} \|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\lambda\boldsymbol{y}-\boldsymbol{x}_i)\|^2\leq 1, \forall i\in\{1,2\}.$$

First, let us reformulate the membership constraints. Fix $i\in\{1,2\}$.

$$\begin{aligned}\frac{\boldsymbol{w}}{1+\lambda}\in\mathcal{E}_i &\iff \left\|\frac{1}{1+\lambda}\boldsymbol{D}_i^{1/2}\boldsymbol{w}-\boldsymbol{D}_i^{1/2}\boldsymbol{x}_i\right\|^2\leq 1\\ &\iff \frac{1}{(1+\lambda)^2}\sum_{j\in[n]} d_{i,j}w_j^2-\frac{2}{1+\lambda}\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{w}+\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i\leq 1\\ &\iff \frac{1}{(1+\lambda)}\sum_{j\in[n]} d_{i,j}w_j^2-2\underbrace{\boldsymbol{x}_i^\top\boldsymbol{D}_i}_{\boldsymbol{b}_i^\top}\boldsymbol{w}+(1+\lambda)\underbrace{\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i}_{c_i}\leq(1+\lambda).\end{aligned}$$

To obtain the final formulation, we encode the quantity $\frac{1}{1+\lambda}w_j^2$ by the additional variable ξ_j satisfying $w_j^2\leq(1+\lambda)\xi_j$. Note that the latter constraint is second-order cone representable as

$$\left\|\begin{matrix} w_j \\ \frac{\xi_j-(1+\lambda)}{2}\end{matrix}\right\|\leq\frac{\xi_j+(1+\lambda)}{2}.$$

Second, let us reformulate the robust constraints. Fix $i\in\{1,2\}$ and consider the constraint

$$\max_{\boldsymbol{y}\in\mathcal{E}_1\cap\mathcal{E}_2}\|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\lambda\boldsymbol{y}-\boldsymbol{x}_i)\|^2\leq 1. \tag{15}$$

We expand the norm-square term in the objective of the maximization problem in (15):

$$\begin{aligned}\|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\lambda\boldsymbol{y}-\boldsymbol{x}_i)\|^2 &= \|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\boldsymbol{x}_i)\|^2-2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\boldsymbol{y}+\lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{y}\|^2\\ &= \|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\boldsymbol{x}_i)\|^2-2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\boldsymbol{y}+\lambda^2\sum_{j\in[n]} d_{i,j}y_j^2.\end{aligned}$$

Similarly, the constraint $\boldsymbol{y} \in \mathcal{E}_k$, for $k = 1, 2$, write as follows

$$\|\boldsymbol{D}_k^{1/2}(\boldsymbol{y} - \boldsymbol{x}_k)\|^2 \leq 1 \iff \|\boldsymbol{D}_k^{1/2}\boldsymbol{y}\|^2 - 2\boldsymbol{x}_k^\top \boldsymbol{D}_k \boldsymbol{y} + \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2 \leq 1$$
$$\iff \sum_{j\in[n]} d_{k,j} y_j^2 - 2\underbrace{\boldsymbol{x}_k^\top \boldsymbol{D}_k}_{\boldsymbol{b}_k^\top} \boldsymbol{y} \leq 1 - \underbrace{\boldsymbol{x}_k^\top \boldsymbol{D}_k \boldsymbol{x}_k}_{c_k}.$$

Hence, (15) is equivalent to $\|\boldsymbol{D}_i^{1/2}(\boldsymbol{w} - \boldsymbol{x}_i)\|^2 + \eta_i^\star(\boldsymbol{w}, \lambda) \leq 1$, with

$$\eta_i^\star(\boldsymbol{w}, \lambda) = \max_{\boldsymbol{y}} \lambda^2 \sum_{j\in[n]} d_{i,j} y_j^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y} \text{ s.t. } \sum_{j\in[n]} d_{k,j} y_j^2 - 2\boldsymbol{b}_k^\top \boldsymbol{y} \leq 1 - c_k, \ \forall k \in \{1, 2\}.$$

Introducing additional variables z_j's such that $z_j = y_j^2$, $\forall j \in [n]$ yields the desired formulation. □

E.2. Proof of Proposition 11

Proof Let us consider an optimal solution of (11), $(\boldsymbol{y}^\star, \boldsymbol{z}^\star)$. For any $j \in [n]$, let us consider $t_j \in \mathbb{R}$ such that $z_j^\star = (y_j^\star)^2 + t_j^2$. According to Lemma 7, we can assume without loss of generality that $\|\boldsymbol{D}_i^{1/2}\boldsymbol{t}\|^2 > 0$. For any β, consider the vector $\boldsymbol{y}(\beta) = \boldsymbol{y}^\star + \beta\boldsymbol{t}$. For $\beta = 0$,

$$\lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{y}(0)\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}(0) \leq \lambda^2 \boldsymbol{d}_i^\top \boldsymbol{z}^\star - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}^\star = \eta_i(\boldsymbol{w}, \lambda),$$

while for $\beta \to \infty$,

$$\lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{y}(\beta)\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}(\beta) \sim \|\boldsymbol{D}_i^{1/2}\boldsymbol{t}\|^2\beta^2 \to +\infty.$$

So there must exist a value of β such that

$$\lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{y}(\beta)\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}(\beta) = \eta_i(\boldsymbol{w}, \lambda). \tag{16}$$

We fix β to this value in the remainder of the proof. We can now follow a similar construction as Xia et al. (2021, Theorem 8). Define $u_1 = 1/\sqrt{1+\beta^2}$, $u_2 = \beta/\sqrt{1+\beta^2}$, $\boldsymbol{s}_1 = u_1\boldsymbol{y}^\star + u_2\boldsymbol{t}$, and $\boldsymbol{s}_2 = u_2\boldsymbol{y}^\star - u_1\boldsymbol{t}$. In particular, for any $j \in [p]$,

$$s_{1,j}^2 + s_{2,j}^2 = (y_j^\star)^2 + t_j^2 = z_j^\star \text{ and } u_1 s_{1,j} + u_2 s_{2,j} = (y_j^\star).$$

Note that Xia et al. (2021) consider the case of balls, i.e., isotropic quadratic form. As a result, they can use the weaker relationships: $\boldsymbol{s}_1^\top \boldsymbol{s}_1 + \boldsymbol{s}_2^\top \boldsymbol{s}_2 = \boldsymbol{z}$ and $u_1\boldsymbol{s}_1 + u_2\boldsymbol{s}_2 = \boldsymbol{y}^\star$.

With these notations, we get

$$\lambda^2 \left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_k}{u_k}\right\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \frac{\boldsymbol{s}_k}{u_k} = \eta_i(\boldsymbol{w}, \lambda), \tag{17}$$

for any $k \in \{1, 2\}$. Indeed, for $k = 1$ we have

$$(16) \iff \lambda^2 \left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_1}{u_1}\right\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \frac{\boldsymbol{s}_1}{u_1} = \eta_i(\boldsymbol{w}, \lambda),$$

and for $k = 2$,

$$\lambda^2 \sum_j d_{i,j} z_j^{\star 2} - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}^\star = \eta_i(\boldsymbol{w}, \lambda)$$
$$\iff \lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{s}_1\|^2 - 2u_1\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{s}_1 + \lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{s}_2\|^2 - 2u_2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{s}_2 = \eta_i(\boldsymbol{w}, \lambda)$$
$$\iff u_1^2\eta_i(\boldsymbol{w}, \lambda) + \lambda^2\|\boldsymbol{D}_i^{1/2}\boldsymbol{s}_2\|^2 - 2u_2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{s}_2 = \eta_i(\boldsymbol{w}, \lambda)$$
$$\iff \lambda^2 \left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_2}{u_2}\right\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \frac{\boldsymbol{s}_2}{u_2} = \eta_i(\boldsymbol{w}, \lambda).$$

Then, from the feasibility of $(\boldsymbol{y}^\star, \boldsymbol{z}^\star)$, we have, for any $k \in \{1,2\}$,

$$\|\boldsymbol{D}_k^{1/2}\boldsymbol{s}_1 - u_1\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2 + \|\boldsymbol{D}_k^{1/2}\boldsymbol{s}_2 - u_2\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2 \le 1 - c_k + \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2 = 1.$$

Consequently,

$$\min\left\{\max_k \frac{1}{u_1^2}\|\boldsymbol{D}_k^{1/2}\boldsymbol{s}_1 - u_1\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2,\ \max_k \frac{1}{u_2^2}\|\boldsymbol{D}_k^{1/2}\boldsymbol{s}_2 - u_2\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|^2\right\} \le \min\left\{\frac{1}{u_1^2},\ \frac{1}{u_2^2}\right\} \le 2.$$

So there exists $\ell \in \{1,2\}$ such that

$$\|\boldsymbol{D}_k^{1/2}(\boldsymbol{s}_\ell/u_\ell) - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| \le \sqrt{2},\ \forall k \in \{1,2\}.$$

Finally, we define

$$\bar{\boldsymbol{y}} = \begin{cases} \boldsymbol{s}_\ell/u_l & \text{if } -2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top \boldsymbol{D}_i\boldsymbol{s}_\ell \ge 0, \\ -\boldsymbol{s}_\ell/u_l & \text{otherwise.} \end{cases}$$

For $k \in \{1,2\}$,

$$\begin{aligned} \|\boldsymbol{D}_k^{1/2}\bar{\boldsymbol{y}} - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| &\le \max\left\{\|\boldsymbol{D}_k^{1/2}(\boldsymbol{s}_\ell/u_l) - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|, \|\boldsymbol{D}_k^{1/2}(-\boldsymbol{s}_\ell/u_l) - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|\right\} \\ &\le \sqrt{2} + 2\|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|. \end{aligned}$$

So for any $\tau \in [0,1]$,

$$\begin{aligned} \|\boldsymbol{D}_k^{1/2}\tau\bar{\boldsymbol{y}} - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| &= \left\|\tau\left(\boldsymbol{D}_k^{1/2}\bar{\boldsymbol{y}} - \boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\right) + (1-\tau)\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\right\| \\ &\le \tau\left(\sqrt{2} + 2\|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|\right) + (1-\tau)\|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| \\ &= \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| + \tau\left(\sqrt{2} + \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|\right). \end{aligned}$$

Hence, $\tau\bar{\boldsymbol{y}}$ is feasible if

$$\tau \le \min_{k\in\{1,2\}} \frac{1 - \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|}{\sqrt{2} + \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|} = \frac{1 - \max_k \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|}{\sqrt{2} + \max_k \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\|}.$$

In addition, $\tau \in [0,1]$, we have

$$\begin{aligned} \lambda^2\|\boldsymbol{D}_i^{1/2}\tau\bar{\boldsymbol{y}}\|^2 - 2\tau\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\bar{\boldsymbol{y}} &= \tau^2\left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_\ell}{u_\ell}\right\|^2 + 2\tau\lambda\left|(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\frac{\boldsymbol{s}_\ell}{u_\ell}\right| \\ &\ge \tau^2\left(\left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_\ell}{u_\ell}\right\|^2 + 2\lambda\left|(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\frac{\boldsymbol{s}_\ell}{u_\ell}\right|\right) \\ &\ge \tau^2\left(\left\|\boldsymbol{D}_i^{1/2}\frac{\boldsymbol{s}_\ell}{u_\ell}\right\|^2 - 2\lambda(\boldsymbol{w}-\boldsymbol{x}_i)^\top\boldsymbol{D}_i\frac{\boldsymbol{s}_\ell}{u_\ell}\right) \\ &= \tau^2\eta_i(\boldsymbol{w},\lambda). \end{aligned}$$

Denoting $\gamma = \max_k \|\boldsymbol{D}_k^{1/2}\boldsymbol{x}_k\| = \max_k \sqrt{c_k}$ and fixing $\tau = \dfrac{1-\gamma}{\sqrt{2}+\gamma}$ yields the result. Note that since we assume that $\mathbf{0}$ is in the relative interior of $\mathcal{E}_1 \cap \mathcal{E}_2$, we have $\gamma \in [0,1)$. □