Electronic Companion for the paper "Minkowski Centers via Robust Optimization: Computation and Applications"

EC.1. 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 $m > n$. □

EC.2. Robust perspective on Minkowski centers: Omitted proofs

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

EC.2.1. 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)

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

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 $\mathcal{C}$. □

EC.2.2. 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

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

□

EC.2.3. 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

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

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

EC.2.4. 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):

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

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 $\lambda^\star$. □

EC.2.5. Minkowski measure for a class of generalized ellipsoids

Proposition EC.1. *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*

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

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

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

□

EC.3. Pseudo-codes and algorithmic details

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

EC.3.1. Hit-and-Run

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

EC.3.2. 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 2: Hit-and-run (HAR) algorithm

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

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

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

for $i=0,1,\dots,m-1$ **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 $\lambda^+=\min\{\lambda_k \mid \lambda_k\ge 0\}$, $\lambda^-=\max\{\lambda_k \mid \lambda_k\le 0\}$.
- Define $\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\lambda\boldsymbol{d}_i$, with $\lambda\sim\mathcal{U}([\lambda^-,\lambda^+])$.

Algorithm 3: Generation of a polyhedron circumscribed to a sphere

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

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

for $i=1,\dots,p$ **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 $d_i=R$.

EC.3.3. 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

$$\max_{(\boldsymbol{y},\boldsymbol{z}_y)\in\mathcal{P}}\ \min_{\substack{\boldsymbol{z}:\boldsymbol{A}_{\boldsymbol{x}}(\boldsymbol{w}-\lambda\boldsymbol{y})+\boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{z}=\boldsymbol{b}\\ \boldsymbol{C}_{\boldsymbol{x}}(\boldsymbol{w}-\lambda\boldsymbol{y})+\boldsymbol{C}_{\boldsymbol{z}}\boldsymbol{z}\le\boldsymbol{d}}} 0 \quad \le 0,$$

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. $\boldsymbol{q}\ge 0$) 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:

$$\max_{(\boldsymbol{y},\boldsymbol{z}_y)\in\mathcal{P}}\ \max_{\boldsymbol{q}\ge 0,\boldsymbol{p}:\boldsymbol{A}_{\boldsymbol{z}}^\top\boldsymbol{p}-\boldsymbol{C}_{\boldsymbol{z}}^\top\boldsymbol{q}=\boldsymbol{0}} \boldsymbol{p}^\top(\boldsymbol{b}-\boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{w}+\lambda\boldsymbol{A}_{\boldsymbol{x}}\boldsymbol{y})-\boldsymbol{q}^\top(\boldsymbol{d}+\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{w}-\lambda\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{y}). \tag{EC.1}$$

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).

EC.3.4. 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., $\mathcal{P}_{\boldsymbol{x}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{z} \in \mathbb{R}^{n_z} : \boldsymbol{C}_x \boldsymbol{x} + \boldsymbol{C}_z \boldsymbol{z} \le \boldsymbol{d}\}$.

We restrict our attention to adjustable variables $\boldsymbol{z}$ in (6) that are quadratic functions of $(\boldsymbol{y}, \boldsymbol{z}_{\boldsymbol{y}})$. 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 $\boldsymbol{z}_{\boldsymbol{y}}$, 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

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

which we concisely write

$$\boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{w} + \boldsymbol{e}_j^\top \boldsymbol{C}_{\boldsymbol{z}} \boldsymbol{z}_0 + \sup_{\boldsymbol{\omega} \in \mathcal{P}, \boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\omega}^\top} \left\{ \boldsymbol{a}^\top \boldsymbol{\omega} + \langle \boldsymbol{A}, \boldsymbol{\Omega} \rangle \right\} \le \boldsymbol{e}_j^\top \boldsymbol{d}, \tag{EC.2}$$

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 (EC.2) 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

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

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

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

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

EC.4. Additional numerical results

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

EC.4.1. 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 EC.1 reports the average number of additional iterations required when using a Chebyshev center vs. a Minkowski center. Table EC.2 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 EC.1 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)

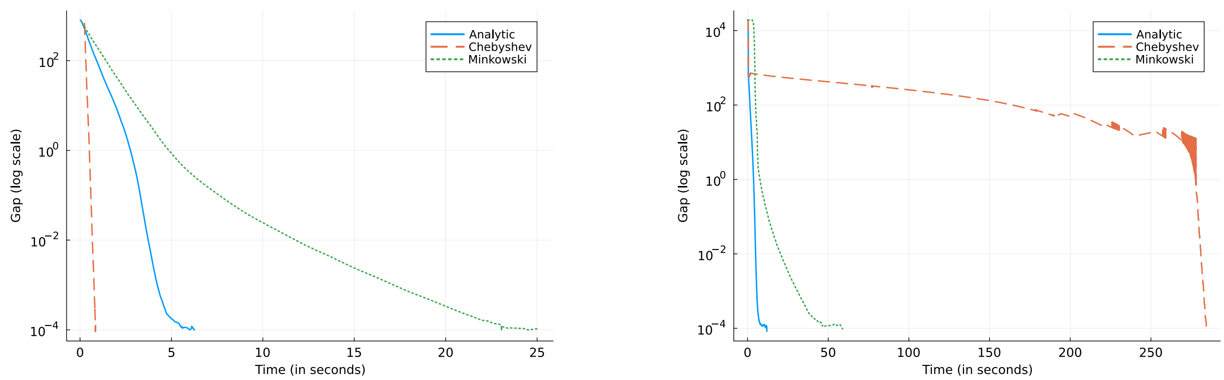
Table EC.2 Regression analysis of the benefit from using the Minkowski center to initialize Algorithm 2. The outcome variable is the number of iterations saved (in log terms).

	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

EC.4.2. 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 EC.1 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 EC.1 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 EC.3 and $k = 10$ in Table EC.4. Performance metrics are averaged over 20 random replications of the same instance.

m	n	# Iterations Analytic	Chebyshev	Minkowski	Time (in s) 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 EC.3 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

m	n	# Iterations Analytic	Chebyshev	Minkowski	Time (in s) 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 EC.4 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

EC.4.3. 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 $n_x + 1$th variable from all the constraints, followed by a screening step that removes redundant constraints. Table EC.5 reports the computational time and the number of constraints created by the FME procedure. As displayed in Table EC.5, the redundant constraint screening step is computationally expensive but drastically reduces the number of constraints in our formulation, which would otherwise exponentially grow with $n - n_x$.

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)

Table EC.5 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

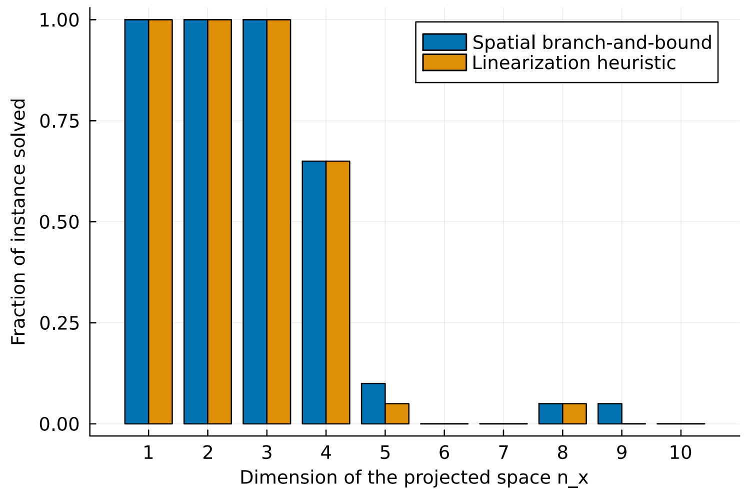


Figure EC.2 Fraction of instances solved by optimality by the column-and-constraint generation approach of Zeng and Zhao (2013).

as the oracle. Figure EC.2 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 EC.3 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 EC.3 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

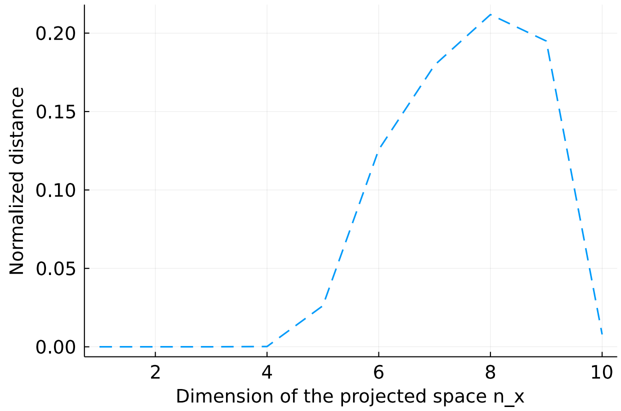


Figure EC.3 Average distance between the solution of (8) and a Minkowski center of $\mathcal{P}_{\boldsymbol{x}}$. The distance is normalized by the depth of the original polyhedron $\mathcal{P}$.

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 EC.6. 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.

Table EC.6 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		

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

EC.5. Intersection of ellipsoids: Omitted proofs

We detail the proofs of Section 5 in this section.

Table EC.7 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 EC.8 Average computational time (in seconds) for solving (7) as a funtion of n and p. Results are averaged over $20 \times 10 = 200$ polyhedra.

n	m 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

EC.5.1. 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

$$\max_{\boldsymbol{y}\in\mathcal{E}_1\cap\mathcal{E}_2} \|\boldsymbol{D}_i^{1/2}(\boldsymbol{w}-\lambda\boldsymbol{y}-\boldsymbol{x}_i)\|^2 \leq 1. \tag{EC.3}$$

We expand the norm-square term in the objective of the maximization problem in (EC.3):

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

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, (EC.3) 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. □

EC.5.2. 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

$$\lambda^2 \|\boldsymbol{D}_i^{1/2}\boldsymbol{y}(\beta)\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \boldsymbol{y}(\beta) = \eta_i(\boldsymbol{w}, \lambda). \tag{EC.4}$$

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

$$\lambda^2 \left\| \boldsymbol{D}_i^{1/2} \frac{\boldsymbol{s}_k}{u_k} \right\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \frac{\boldsymbol{s}_k}{u_k} = \eta_i(\boldsymbol{w}, \lambda), \tag{EC.5}$$

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

$$\text{(EC.4)} \iff \lambda^2 \left\| \boldsymbol{D}_i^{1/2} \frac{\boldsymbol{s}_1}{u_1} \right\|^2 - 2\lambda(\boldsymbol{w} - \boldsymbol{x}_i)^\top \boldsymbol{D}_i \frac{\boldsymbol{s}_1}{u_1} = \eta_i(\boldsymbol{w}, \lambda),$$

and for $k=2$,

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

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

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

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 $\gamma \in [0,1)$. □