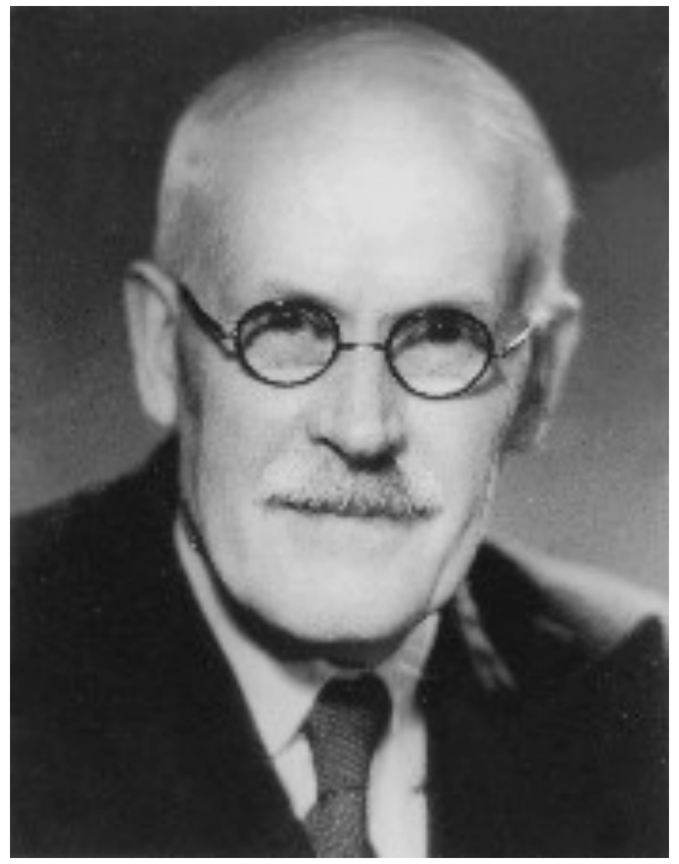
Constrained Optimization of Rank-One Functions with Indicator Variables

Soroosh Shafiee, Cornell University
Fatma Kılınç-Karzan, Carnegie Mellon University

Occam's razor

Principle of Parsimony



"The existence of simple laws is, then, apparently, to be regarded as a quality of nature". [1)]

1) Wrinch & Jeffreys, Philosophical Magazine Series, 1921

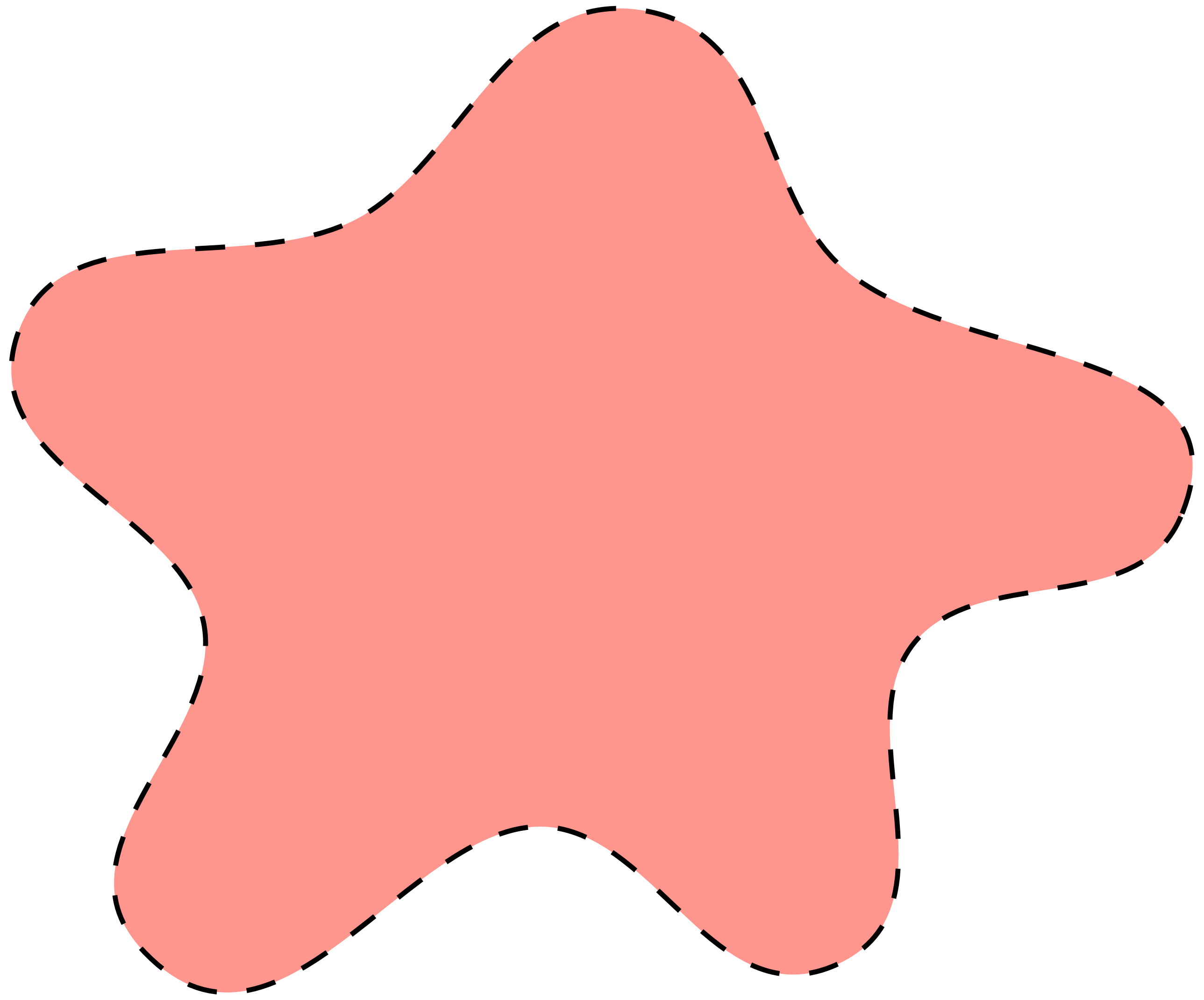
Sparse Learning

$$\min_{\boldsymbol{x},\boldsymbol{z}} \{H(\boldsymbol{x}) : \ (\boldsymbol{x},\boldsymbol{z}) \in \mathcal{X} \times \mathcal{Z}, \ x_i(1 - z_i) = 0, \ \forall i \in [d]\}$$

- $\mathcal{X} \times \mathcal{Z} \subseteq \mathbb{R}^d \times \{0,1\}^d$
- $z_i = 0 \implies x_i = 0$
- $H(\boldsymbol{x}) = \sum_{k\in[n]} h_k(\boldsymbol{a}_k^\top \boldsymbol{x}) + h_0(\boldsymbol{x})$

Epigraph Set

$$\min_{\boldsymbol{x},\boldsymbol{z}} \left\{ H(\boldsymbol{x}) : \ (\boldsymbol{x},\boldsymbol{z}) \in \mathcal{X} \times \mathcal{Z}, \ x_i(1 - z_i) = 0, \ \forall i \in [d] \right\}$$

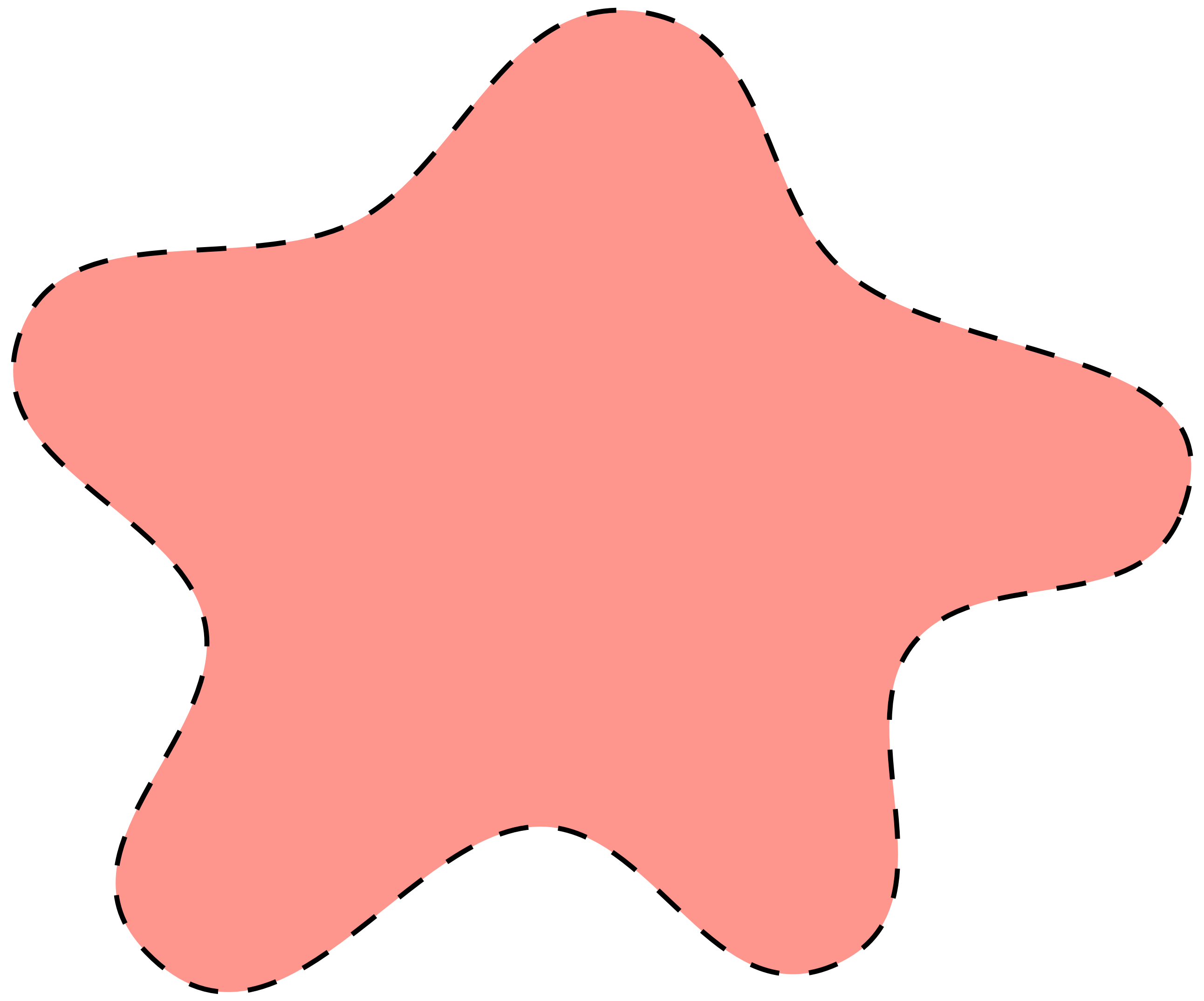


Epigraph Set

$$\min_{\boldsymbol{x},\boldsymbol{z}} \{H(\boldsymbol{x}) : \; (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X} \times \mathcal{Z}, \; x_i(1 - z_i) = 0, \; \forall i \in [d]\}$$

$$\{(\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \; H(\boldsymbol{x}) \leq \tau, \; x_i(1 - z_i) = 0, \; \forall i \in [d]\}$$

- Valid inequalities?
- Big-M free formulations?
- Closed convex hull?

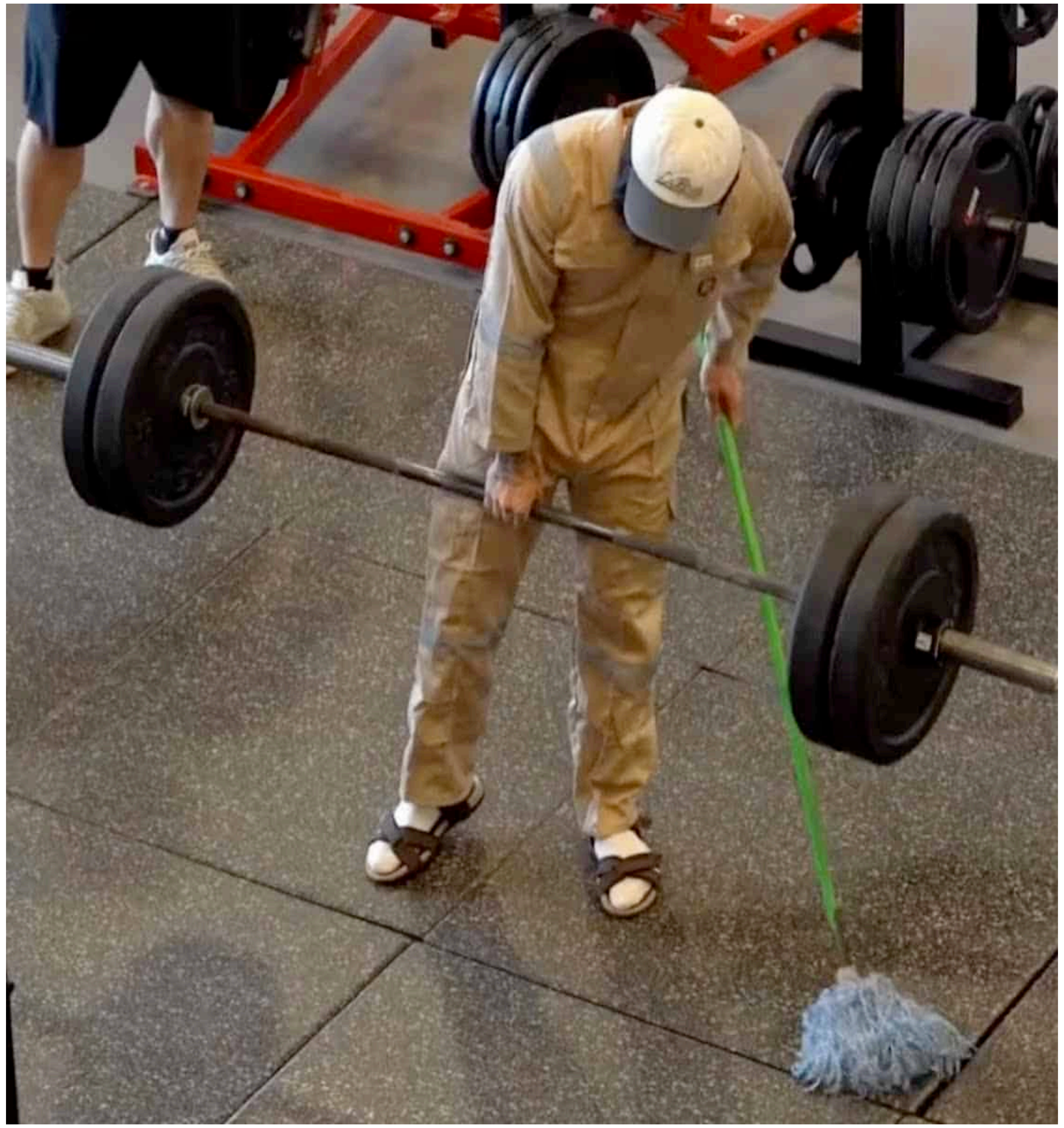


Separable and Non-Separable Structure

$$\mathcal{H} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \begin{array}{l} \sum_{i \in [d]} h_i(x_i) \leq \tau, \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$$

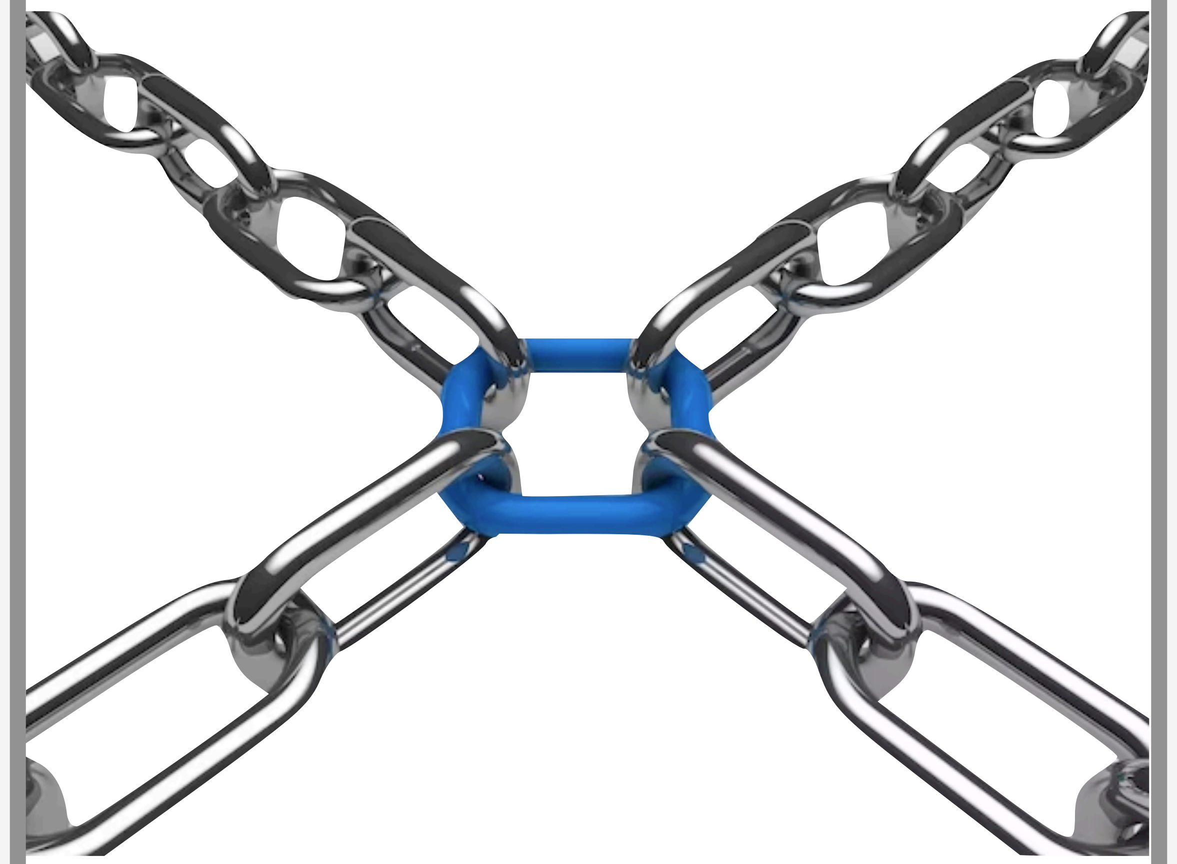
$$\mathcal{T} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \begin{array}{l} h(\boldsymbol{a}^\top \boldsymbol{x}) \leq \tau, \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$$

Proof Strategy



Proof Strategy

Conic Binary Optimization

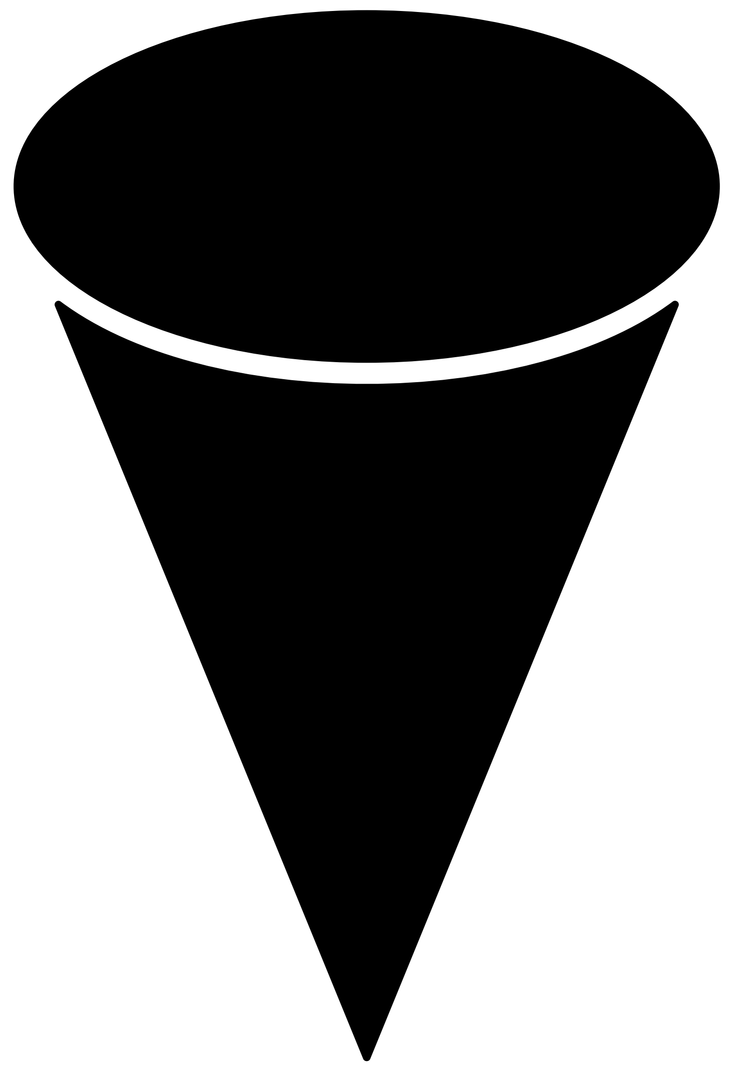


Perspective Functions

Conic Binary Sets

$$\mathcal{S}(\Delta, \mathbb{K}) = \{(\boldsymbol{\alpha}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \Delta : \boldsymbol{A}_i \boldsymbol{\alpha}_i + \boldsymbol{B}_i \delta_i \in \mathbb{K}_i, \ \forall i \in [p]\}$$

- $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)$
- $\Delta \subseteq \{0, 1\}^n$
- $\mathbb{K}_i$ is a convex cone & $\mathbf{0} \in \mathbb{K}_i$
- $\mathbb{K} = \times_{i \in [d]} \mathbb{K}_i$



Main Result

Theorem 1. The following holds

- ◇ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) = \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$

Main Result

Theorem 1. The following holds

- $\operatorname{conv}(\mathcal{S}(\Delta, \mathbb{K})) = \mathcal{S}(\operatorname{conv}(\Delta), \mathbb{K})$
- $\operatorname{cl}\operatorname{conv}(\mathcal{S}(\Delta, \mathbb{K})) = \mathcal{S}(\operatorname{conv}(\Delta), \operatorname{cl}(\mathbb{K}))$

Assumption 1. $\mathbb{K}_i$ is *nice*:

- $\delta_i > 0$ & $\boldsymbol{A}_i\boldsymbol{\alpha}_i + \boldsymbol{B}_i\delta_i \in \operatorname{cl}(\mathbb{K}_i) \implies \boldsymbol{A}_i\boldsymbol{\alpha}_i + \boldsymbol{B}_i\delta_i \in \mathbb{K}_i$
- $\boldsymbol{B}_i \in \mathbb{K}_i$

Proof of Theorem 1

- ◇ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \subseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 😉
- ◇ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \supseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 🤔

  □ $(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\delta}}) \in \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K}) \implies \bar{\boldsymbol{\delta}} \in \mathrm{conv}(\Delta) \implies \bar{\boldsymbol{\delta}} = \sum_{k \in [q]} \lambda_k \boldsymbol{\delta}^k$

Proof of Theorem 1

- ◇ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \subseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 😉
- ◇ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \supseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 🤔

□ $(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\delta}}) \in \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K}) \implies \bar{\boldsymbol{\delta}} \in \mathrm{conv}(\Delta) \implies \bar{\boldsymbol{\delta}} = \sum_{k \in [q]} \lambda_k \boldsymbol{\delta}^k$

□ $$\boldsymbol{\alpha}_i^k := \begin{cases} \bar{\boldsymbol{\alpha}}_i, & \text{if } \bar{\delta}_i = 0 \\ \bar{\boldsymbol{\alpha}}_i / \bar{\delta}_i, & \text{if } \bar{\delta}_i \neq 0 \text{ and } \delta_i^k = 1 \\ \mathbf{0}, & \text{if } \bar{\delta}_i \neq 0 \text{ and } \delta_i^k = 0 \end{cases}$$

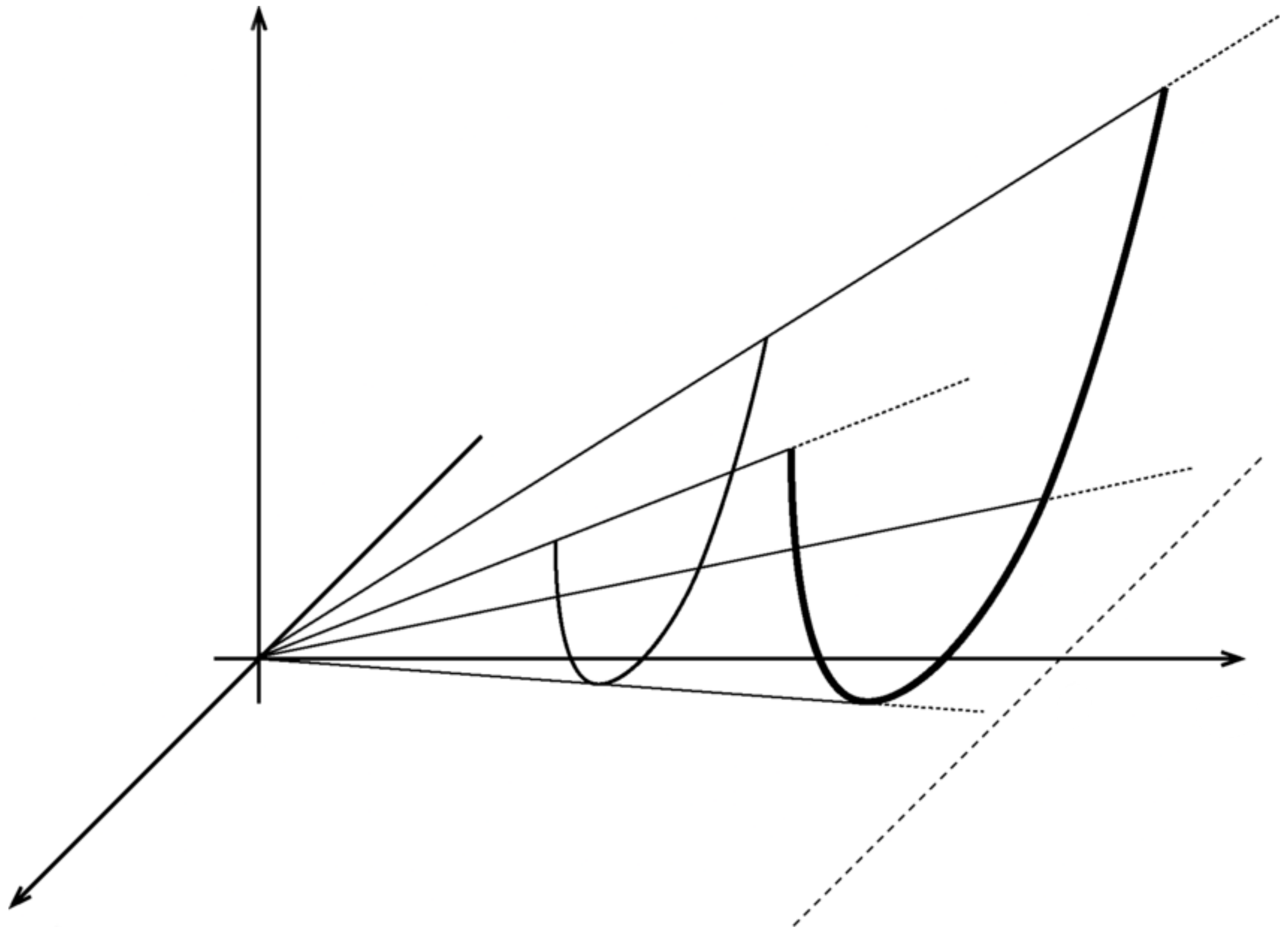
Proof of Theorem 1

- $\diamond$ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \subseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 😉
- $\diamond$ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) \supseteq \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$ 🤔

  - $\square$ $(\boldsymbol{\alpha}^k, \boldsymbol{\delta}^k) \in \mathcal{S}(\Delta, \mathbb{K})$
  - $\square$ $(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\delta}}) = \sum_{k \in [q]} \lambda_k (\boldsymbol{\alpha}^k, \boldsymbol{\delta}^k)$

  $\Longrightarrow$ $(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\delta}}) \in \mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K}))$

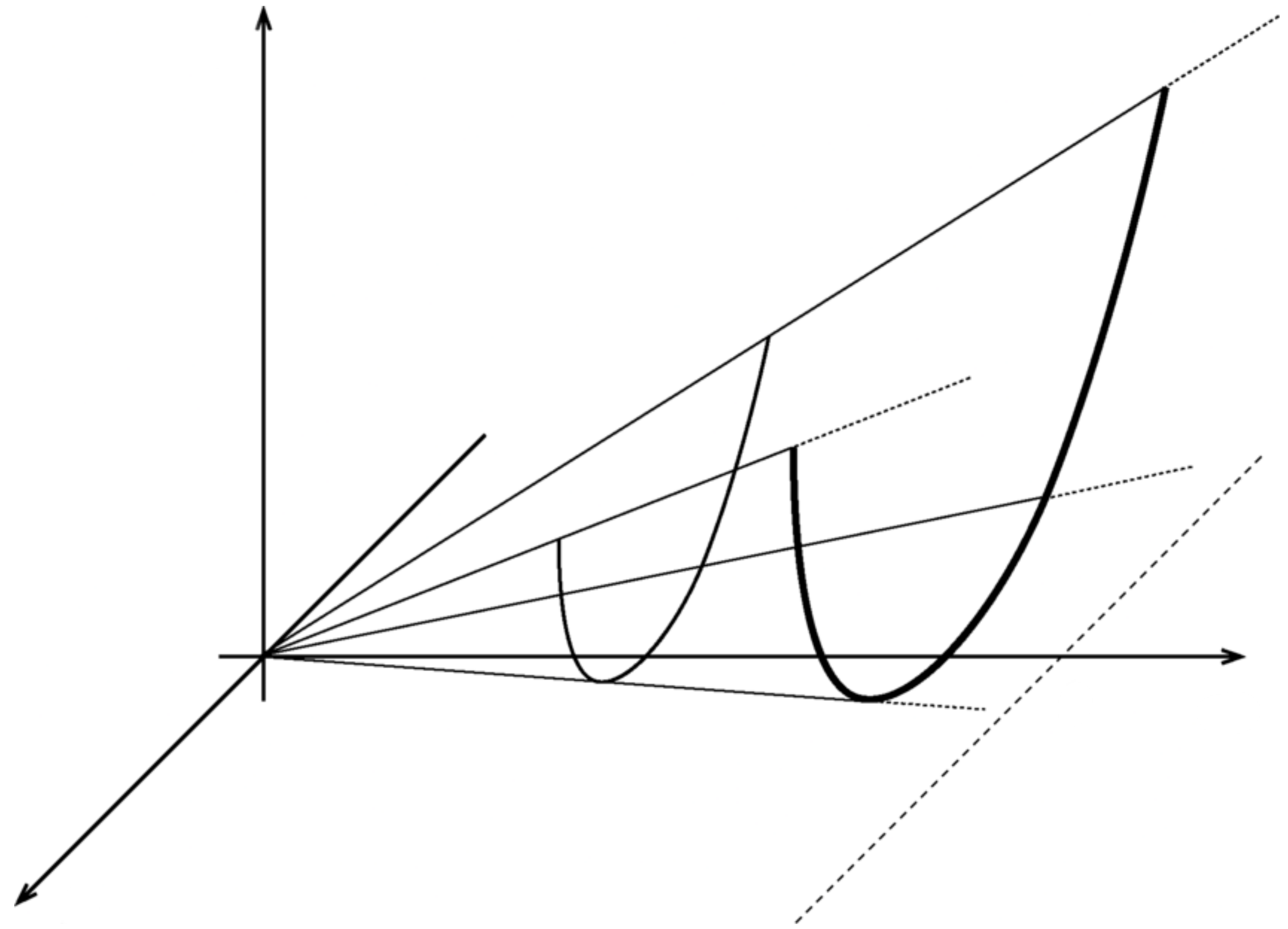
Perspective Function



- $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ is proper l.s.c. and convex
- $h(\mathbf{0}) = 0$

$$h^+(\boldsymbol{x}, w) := \begin{cases} w\, h(\boldsymbol{x}/w), & \text{if } w > 0, \\ 0, & \text{if } w = 0 \text{ and } \boldsymbol{x} = \mathbf{0}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Closed Perspective Function



- $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ is proper l.s.c. and convex
- $h(\mathbf{0}) = 0$

$$h^\pi(\boldsymbol{x}, w) := \begin{cases} w\, h(\boldsymbol{x}/w), & \text{if } w > 0, \\ \lim_{s \downarrow 0} s\, h(\boldsymbol{x}/s), & \text{if } w = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

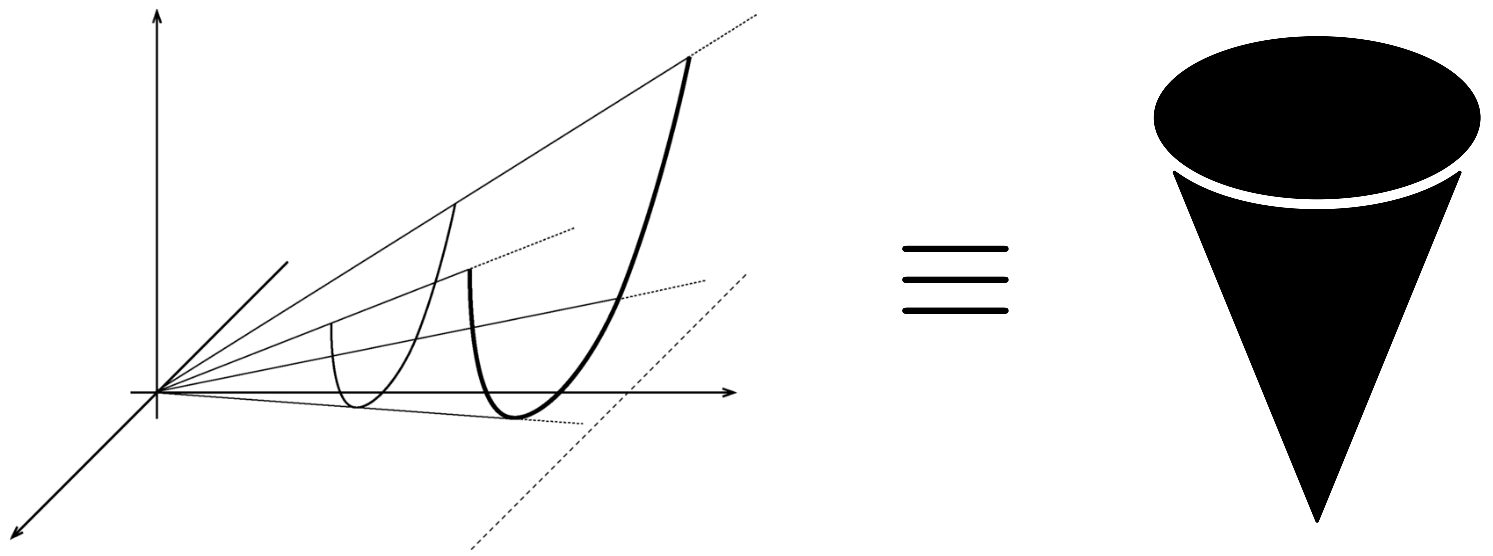
Perspective Cone

Assumption 1. $\mathbb{K}_i$ is *nice*:

- $\diamond$ $\delta_i > 0$ & $\boldsymbol{A}_i\boldsymbol{\alpha}_i + \boldsymbol{B}_i\delta_i \in \mathrm{cl}(\mathbb{K}_i) \implies \boldsymbol{A}_i\boldsymbol{\alpha}_i + \boldsymbol{B}_i\delta_i \in \mathbb{K}_i$
- $\diamond$ $\boldsymbol{B}_i \in \mathbb{K}_i$

Lemma 1. The following holds.

- $\diamond$ $\mathrm{epi}(h^+)$ is a convex cone
- $\diamond$ $\mathbf{0} \in \mathrm{epi}(h^+)$
- $\diamond$ $\mathrm{epi}(h^+)$ is nice
- $\diamond$ $\mathrm{cl}(\mathrm{epi}(h^+)) = \mathrm{epi}(h^\pi)$



- $\diamond$ $\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) = \mathcal{S}(\mathrm{conv}(\Delta), \mathbb{K})$
- $\diamond$ $\mathrm{cl}\,\mathrm{conv}(\mathcal{S}(\Delta, \mathbb{K})) = \mathcal{S}(\mathrm{conv}(\Delta), \mathrm{cl}(\mathbb{K}))$

Perspective Reformulation Technique

$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \le \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1 - \delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Perspective Reformulation Technique

$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \leq \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1 - \delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Theorem 2. $\text{conv}(\mathcal{W})$ is given by

$$\left\{ (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \text{conv}(\Delta) : \begin{array}{l} h_i^+(\beta_{i,1}, \delta_i) \leq \gamma_i, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Perspective Reformulation Technique

$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \leq \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1 - \delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Theorem 2. $\operatorname{cl}\operatorname{conv}(\mathcal{W})$ is given by

$$\left\{ (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \operatorname{conv}(\Delta) : \begin{array}{l} h_i^{\pi}(\beta_{i,1}, \delta_i) \leq \gamma_i, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Proof of Theorem 2

$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \leq \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1 - \delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

$$h^+(\boldsymbol{x}, w) := \begin{cases} w\, h(\boldsymbol{x}/w), & \text{if } w > 0, \\ 0, & \text{if } w = 0 \text{ and } \boldsymbol{x} = \boldsymbol{0}, \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\mathcal{W} = \left\{ (\boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i^+(\beta_{i,1}, \delta_i) \leq \gamma_i, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

Proof of Theorem 2

$$\mathcal{S}(\Delta, \mathbb{K}) = \{(\boldsymbol{\alpha}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \Delta : \boldsymbol{A}_i \boldsymbol{\alpha}_i + \boldsymbol{B}_i \delta_i \in \mathbb{K}_i, \ \forall i \in [p]\}$$

$$\mathcal{W} = \left\{ (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i^+(\beta_{i,1}, \delta_i) \leq \gamma_i, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

$$h_i^+(\beta_{i,1}, \delta_i) \leq \gamma_i \ \& \ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i \iff \boldsymbol{A}_i \boldsymbol{\alpha}_i + \boldsymbol{B}_i \delta_i \in \mathbb{K}_i$$

$$\boldsymbol{\alpha}_i = (\boldsymbol{\beta}_i, \gamma_i), \ \boldsymbol{A}_i = \begin{bmatrix} \mathbf{0}^\top & 1 \\ \boldsymbol{e}_1^\top & 0 \\ \mathbf{0}^\top & 0 \\ \boldsymbol{C}_i & 0 \end{bmatrix}, \ \boldsymbol{B}_i = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \mathbf{0} \end{bmatrix}, \ \mathbb{K}_i = \operatorname{epi}(h_i^+) \times \mathbb{C}_i$$

Separable Structure

$$\mathcal{H} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \begin{array}{l} \sum_{i\in[d]} h_i(x_i) \le \tau, \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$$

Theorem 3. If $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^d : x_i \ge 0, \ \forall i \in \mathcal{I}\}$, then

$$\operatorname{cl}\operatorname{conv}(\mathcal{H}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \operatorname{conv}(\mathcal{Z}) : \ \sum_{i\in[d]} h_i^{\pi}(x_i, z_i) \le \tau \right\}$$

[1] Günlük & Linderoth, *Mathematical Programming*, 2010

[2] Xie & Deng, *SIAM Journal on Optimization*, 2020

[3] Wei, Gómez & Küçükyavuz., *Mathematical Programming*, 2022

[4] Bacci, Frangioni, Gentile & Tavlaridis-Gyparakis, *Operations Research*, 2023

Separable Structure

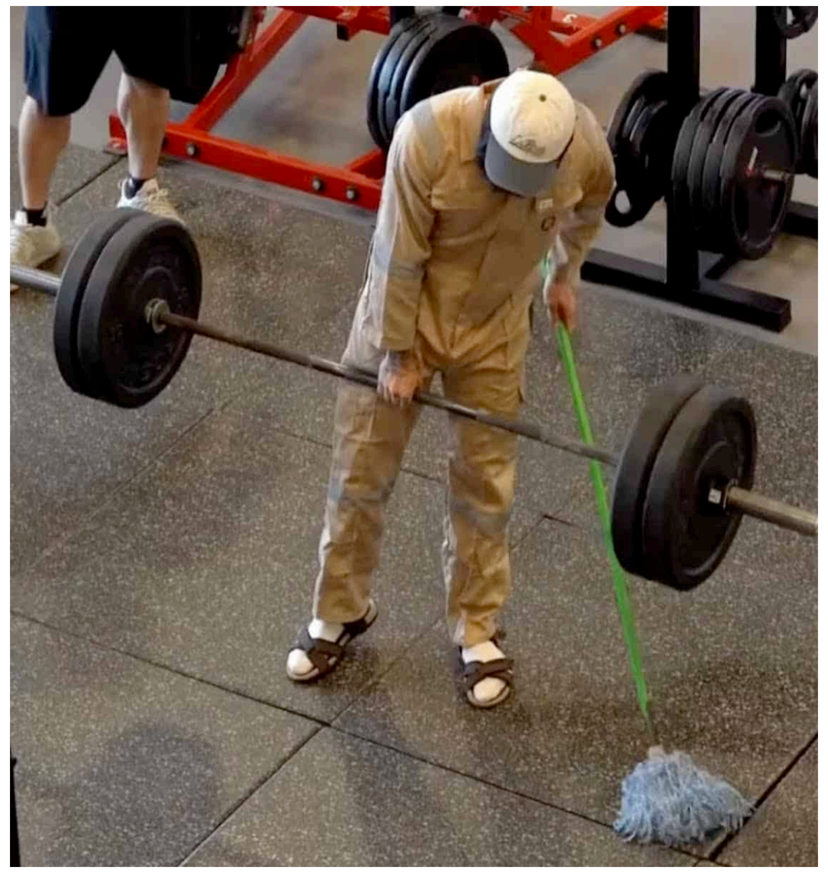
	$\mathcal{X}$	$\mathcal{Z}$	h	Proof
[1]	$\mathbb{R}^d$	$\{0,1\}^d$	super-linear	disjunctive programming
[2]	$\mathbb{R}^d$	cardinality	quadratic	constructive
[3]	$\mathbb{R}^d$	$\subseteq \{0,1\}^d$	real-valued	support function
[4]	$\mathbb{R}^d$	$\subseteq \{0,1\}^d$	slater	support function

[1] Günlük & Linderoth, *Mathematical Programming*, 2010
[2] Xie & Deng, *SIAM Journal on Optimization*, 2020
[3] Wei, Gómez & Küçükyavuz., *Mathematical Programming*, 2022
[4] Bacci, Frangioni, Gentile & Tavlaridis-Gyparakis, *Operations Research*, 2023

Proof of Theorem 3

$\diamond$ $\overline{\mathcal{H}} := \left\{ (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathcal{X} \times \mathcal{Z} \times \mathbb{R}^d : \begin{array}{l} h_i(x_i) \le t_i, \ \forall i \in [d], \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$

$\diamond$ $\mathcal{H} = \{(\tau, \boldsymbol{x}, \boldsymbol{z}) : \exists \boldsymbol{t} \text{ s.t. } (\tau, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathbb{R} \times \overline{\mathcal{H}}, \ \mathbf{1}^\top \boldsymbol{t} = \tau\}$



Proof of Theorem 3

$\diamond\ \overline{\mathcal{H}} := \left\{ (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathcal{X} \times \mathcal{Z} \times \mathbb{R}^d : \begin{array}{l} h_i(x_i) \le t_i,\ \forall i \in [d], \\ x_i(1 - z_i) = 0,\ \forall i \in [d] \end{array} \right\}$

$\diamond\ \operatorname{cl}\operatorname{conv}(\mathcal{H}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) : \exists \boldsymbol{t} \text{ s.t. } \begin{array}{l} (\tau, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathbb{R} \times \operatorname{cl}\operatorname{conv}(\overline{\mathcal{H}}), \\ \mathbf{1}^\top \boldsymbol{t} = \tau \end{array} \right\}$

Proof of Theorem 3

$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \leq \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1 - \delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

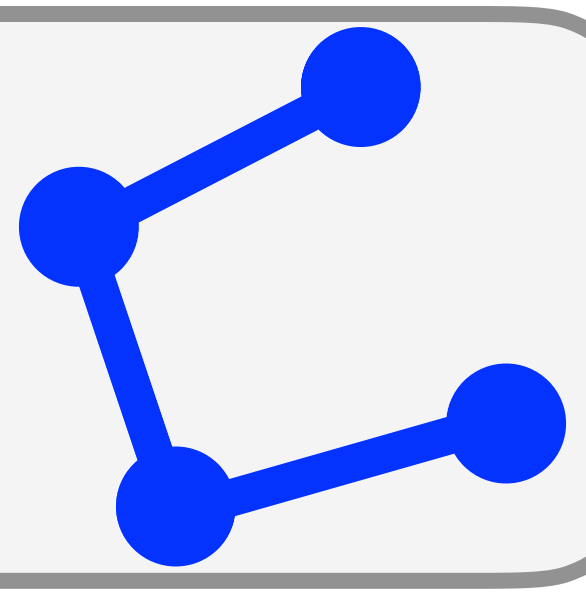
- $\overline{\mathcal{H}} := \left\{ (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathcal{X} \times \mathcal{Z} \times \mathbb{R}^d : \begin{array}{l} h_i(x_i) \leq t_i, \ \forall i \in [d], \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$

- $\operatorname{cl}\operatorname{conv}(\mathcal{H}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) : \exists \boldsymbol{t} \text{ s.t. } \begin{array}{l} (\tau, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathbb{R} \times \operatorname{cl}\operatorname{conv}(\overline{\mathcal{H}}), \\ \mathbf{1}^\top \boldsymbol{t} = \tau \end{array} \right\}$

- $\operatorname{cl}\operatorname{conv}(\overline{\mathcal{H}}) = \left\{ (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{t}) \in \mathcal{X} \times \operatorname{conv}(\mathcal{Z}) \times \mathbb{R}^d : h_i^\pi(x_i, z_i) \leq t_i, \ \forall i \in [d] \right\}$

Rank-One Structure I

$$\mathcal{T} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \begin{array}{l} h(\boldsymbol{a}^\top \boldsymbol{x}) \le \tau, \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$$



Theorem 4. If $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Z}$ is *connected*, then

$$\operatorname{cl}\operatorname{conv}(\mathcal{T}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d : \exists w \in \mathbb{R} \text{ s.t. } \begin{array}{l} h^\pi(\boldsymbol{a}^\top \boldsymbol{x}, w) \le \tau, \\ (w, \boldsymbol{z}) \in \operatorname{conv}(\Delta_1) \end{array} \right\}$$

$$\Delta_1 := \{(w, \boldsymbol{z}) \in \{0, 1\} \times \mathcal{Z} : w \le \mathbf{1}^\top \boldsymbol{z}\}$$

[1] Atamtürk & Gómez, *Mathematical Programming*, 2022

[2] Wei, Gómez & Küçükyavuz, IPCO, 2020.

[3] Wei, Gómez & Küçükyavuz., *Mathematical Programming*, 2022

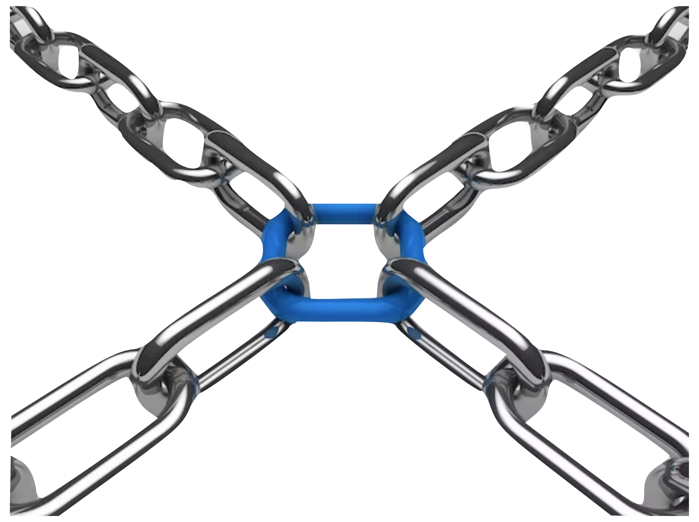
[4] Han & Gómez, *arXiv*, 2021

Rank-One Structure I

	description	$\mathcal{Z}$	h	Proof
[1]	ideal	$\{0,1\}^d$	quadratic	support function
[2]	ideal	$\subseteq \{0,1\}^d$	quadratic	constructive
[3]	ideal	$\subseteq \{0,1\}^d$	real-valued	support function
[4]	extended	$\{0,1\}^d$	real-valued	support function

[1] Atamtürk & Gómez, *Mathematical Programming*, 2022
[2] Wei, Gómez & Küçükyavuz, IPCO, 2020.
[3] Wei, Gómez & Küçükyavuz., *Mathematical Programming*, 2022
[4] Han & Gómez, *arXiv*, 2021

Proof of Theorem 4

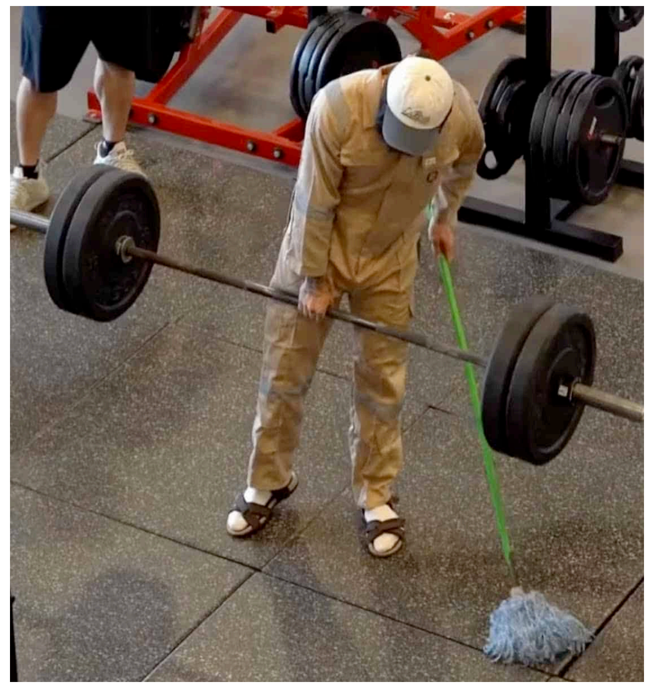


- $(\tau, \boldsymbol{x}, \boldsymbol{z}) \in \operatorname{cl}\operatorname{conv}(\mathcal{T})$ & $\boldsymbol{a}^\top \bar{\boldsymbol{x}} = 0 \implies (\tau, \boldsymbol{x} + \bar{\boldsymbol{x}}, \boldsymbol{z}) \in \operatorname{cl}\operatorname{conv}(\mathcal{T})$

- $\mathcal{R} := \left\{(\bar{\tau}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d : \ \bar{\tau} = 0, \ \bar{\boldsymbol{z}} = \mathbf{0}, \ \boldsymbol{a}^\top \bar{\boldsymbol{x}} = 0\right\}$

$$\implies \operatorname{cl}\operatorname{conv}(\mathcal{T}) = \operatorname{cl}\operatorname{conv}(\mathcal{T} + \mathcal{R})$$

Proof of Theorem 4



- $\operatorname{cl}\operatorname{conv}(\mathcal{T}) = \operatorname{cl}\operatorname{conv}(\mathcal{T} + \mathcal{R})$

- $$\overline{\mathcal{T}} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}, s, w) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} : \begin{array}{l} h(s) \leq \tau, \\ \boldsymbol{a}^\top \boldsymbol{x} = s, \\ s(1 - w) = 0, \\ (w, \boldsymbol{z}) \in \Delta_1 \end{array} \right\}$$

- $\mathcal{T} + \mathcal{R} = \operatorname{Proj}_{\tau, \boldsymbol{x}, \boldsymbol{z}}(\overline{\mathcal{T}})$

Proof of Theorem 4

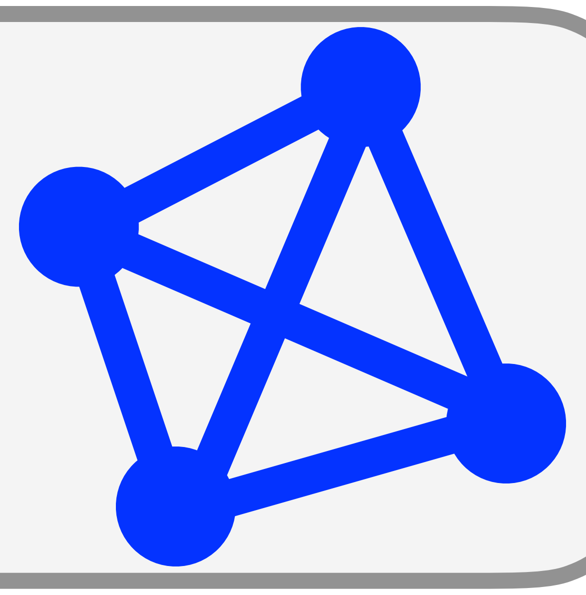
$$\mathcal{W} := \left\{ (\boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) \in \mathbb{R}^m \times \mathbb{R}^p \times \Delta : \begin{array}{l} h_i(\beta_{i,1}) \le \gamma_i, \ \forall i \in [p], \\ \beta_{i,1}(1-\delta_i) = 0, \ \forall i \in [p], \\ \boldsymbol{C}_i \boldsymbol{\beta}_i \in \mathbb{C}_i, \ \forall i \in [p] \end{array} \right\}$$

- $\operatorname{cl}\operatorname{conv}(\mathcal{T}) = \{(\tau, \boldsymbol{x}, \boldsymbol{z}) : \exists s, w \text{ s.t. } (\tau, \boldsymbol{x}, \boldsymbol{z}, s, w) \in \operatorname{cl}\operatorname{conv}(\overline{\mathcal{T}})\}$

- $\operatorname{cl}\operatorname{conv}(\overline{\mathcal{T}}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}, s, w) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} : \begin{array}{l} h^\pi(s, w) \le \tau, \\ \boldsymbol{a}^\top \boldsymbol{x} = s, \\ (w, \boldsymbol{z}) \in \operatorname{conv}(\Delta_1) \end{array} \right\}$

Rank-One Structure II

$$\mathcal{T} := \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z} : \begin{array}{l} h(\boldsymbol{a}^\top \boldsymbol{x}) \le \tau, \\ x_i(1 - z_i) = 0, \ \forall i \in [d] \end{array} \right\}$$



Theorem 5. If $\mathcal{X} = \{\mathbb{R}^d : x_i \ge 0, \ \forall i \in \mathcal{I}\}$ and $\mathcal{Z}$ is *complete*, then

$$\operatorname{cl} \operatorname{conv}(\mathcal{T}) = \left\{ (\tau, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R}^d : \exists \boldsymbol{s}, \boldsymbol{w} \text{ s.t. } \begin{array}{l} \sum_{i \in [d]} h^\pi(a_i s_i, w_i) \le \tau, \\ 0 \le s_i \le x_i, \ \forall i \in \mathcal{I}, \\ \boldsymbol{a}^\top \boldsymbol{s} = \boldsymbol{a}^\top \boldsymbol{x}, \\ (\boldsymbol{w}, \boldsymbol{z}) \in \operatorname{conv}(\Delta) \end{array} \right\}$$

$$\Delta := \{(\boldsymbol{w}, \boldsymbol{z}) \in \{0,1\}^d \times \mathcal{Z} : \ \mathbf{1}^\top \boldsymbol{w} \le 1, \ \boldsymbol{w} \le \boldsymbol{z}\}$$

[1] Han & Gómez, *arXiv*, 2021

Numerical Results: Sparse Logistic Regression

$$\min \quad \frac{1}{N} \sum_{j \in [N]} \log(1 + \exp(-b_j \boldsymbol{a}_j^\top \boldsymbol{x})) + \lambda \sum_{i \in [d]} z_i + \mu \|\boldsymbol{x}\|_2^2$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathbb{R}_+^d, \ \boldsymbol{z} \in \mathcal{Z}, \ x_i(1 - z_i) = 0, \ \forall i \in [d],$$

- $\mathcal{X}$ nonnegativity constraints
- $\mathcal{Z}$ strong hierarchy constraints
- $\boldsymbol{a}$ is sparse

Natural Relaxation

$$\begin{aligned} \min \quad & \frac{1}{N} \sum_{j \in [N]} \log(1 + \exp(-b_j \boldsymbol{a}_j^\top \boldsymbol{x})) + \lambda \sum_{i \in [d]} z_i + \mu \|\boldsymbol{x}\|_2^2 \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathbb{R}_+^d, \ \boldsymbol{z} \in \mathcal{Z}, \ \require{cancel}\xcancel{x_i(1 - z_i) = 0}, \ \forall i \in [d], \end{aligned}$$

Separable Relaxation

$$\sum_{i\in[d]} \frac{x_i^2}{z_i}$$

$$\min \quad \frac{1}{N}\sum_{j\in[N]} \log(1+\exp(-b_j \boldsymbol{a}_j^\top \boldsymbol{x})) + \lambda \sum_{i\in[d]} z_i + \mu \|\boldsymbol{x}\|_2^2$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathbb{R}_+^d, \ \boldsymbol{z} \in \mathcal{Z}, \ x_i(1-z_i) = 0, \ \forall i \in [d],$$

$$\boldsymbol{z} \in \text{conv}(\mathcal{Z})$$

Rank One Relaxations

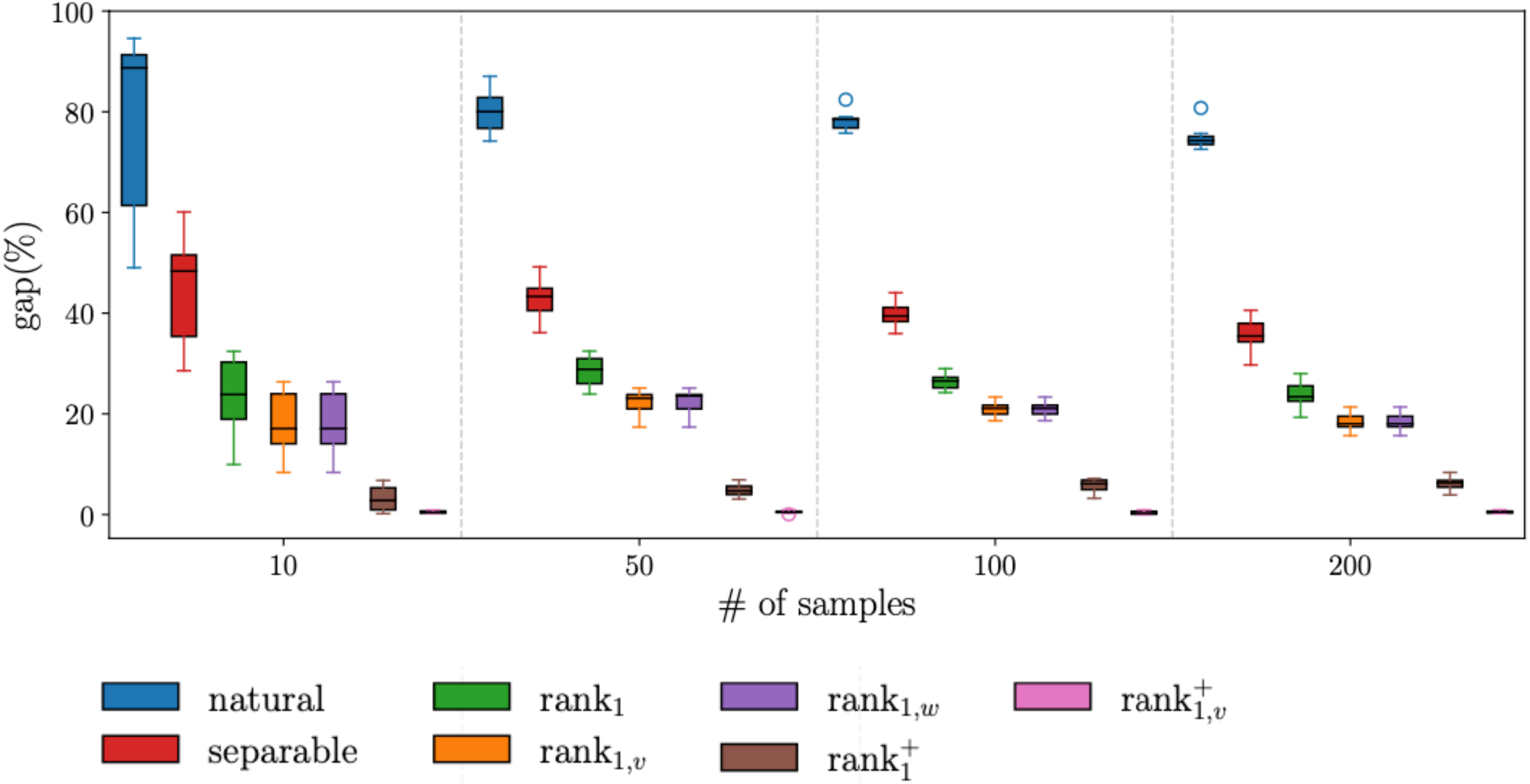
$$\sum_{i\in[d]} \frac{x_i^2}{z_i}$$

$$\min \quad \frac{1}{N}\sum_{j\in[N]} \boxed{\log(1+\exp(-b_j \boldsymbol{a}_j^\top \boldsymbol{x}))} + \lambda \sum_{i\in[d]} z_i + \mu\|\boldsymbol{x}\|_2^2$$

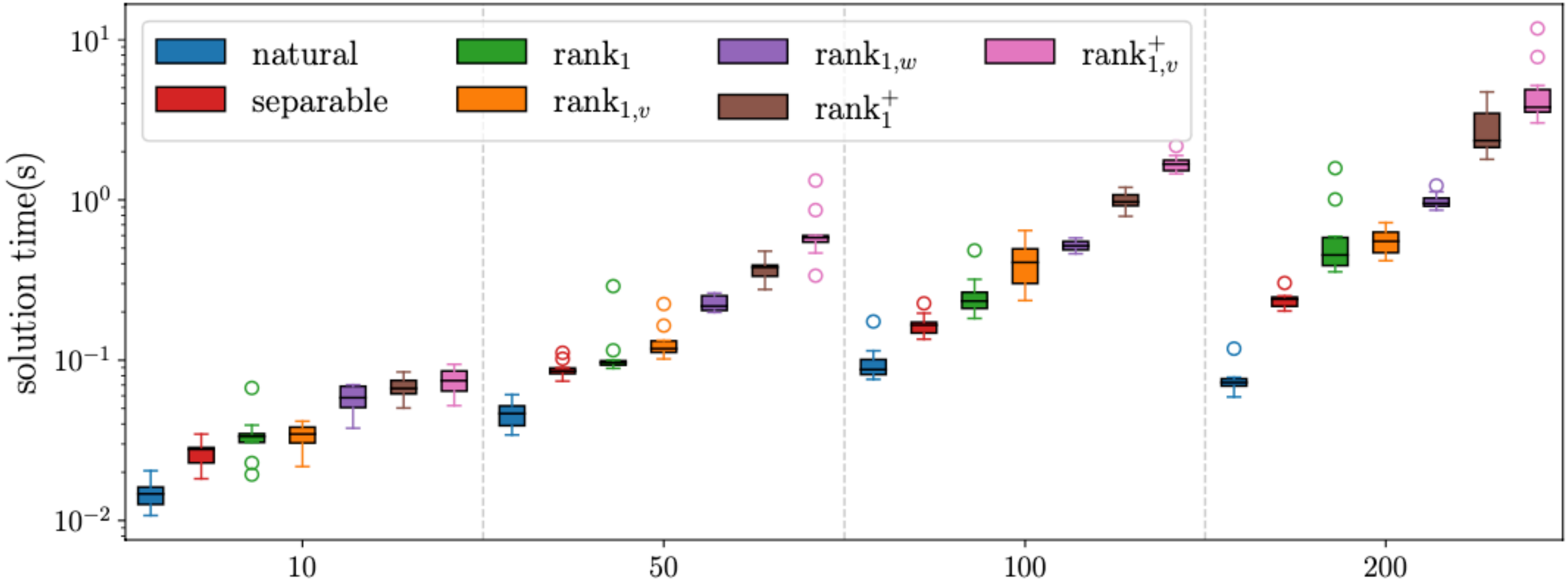
$$\text{s.t.} \quad \boldsymbol{x} \in \mathbb{R}^d_+, \ \boldsymbol{z} \in \mathcal{Z}, \ \cancel{x_i(1-z_i)=0}, \ \forall i \in [d],$$

$$\boldsymbol{z} \in \mathrm{conv}(\mathcal{Z})$$

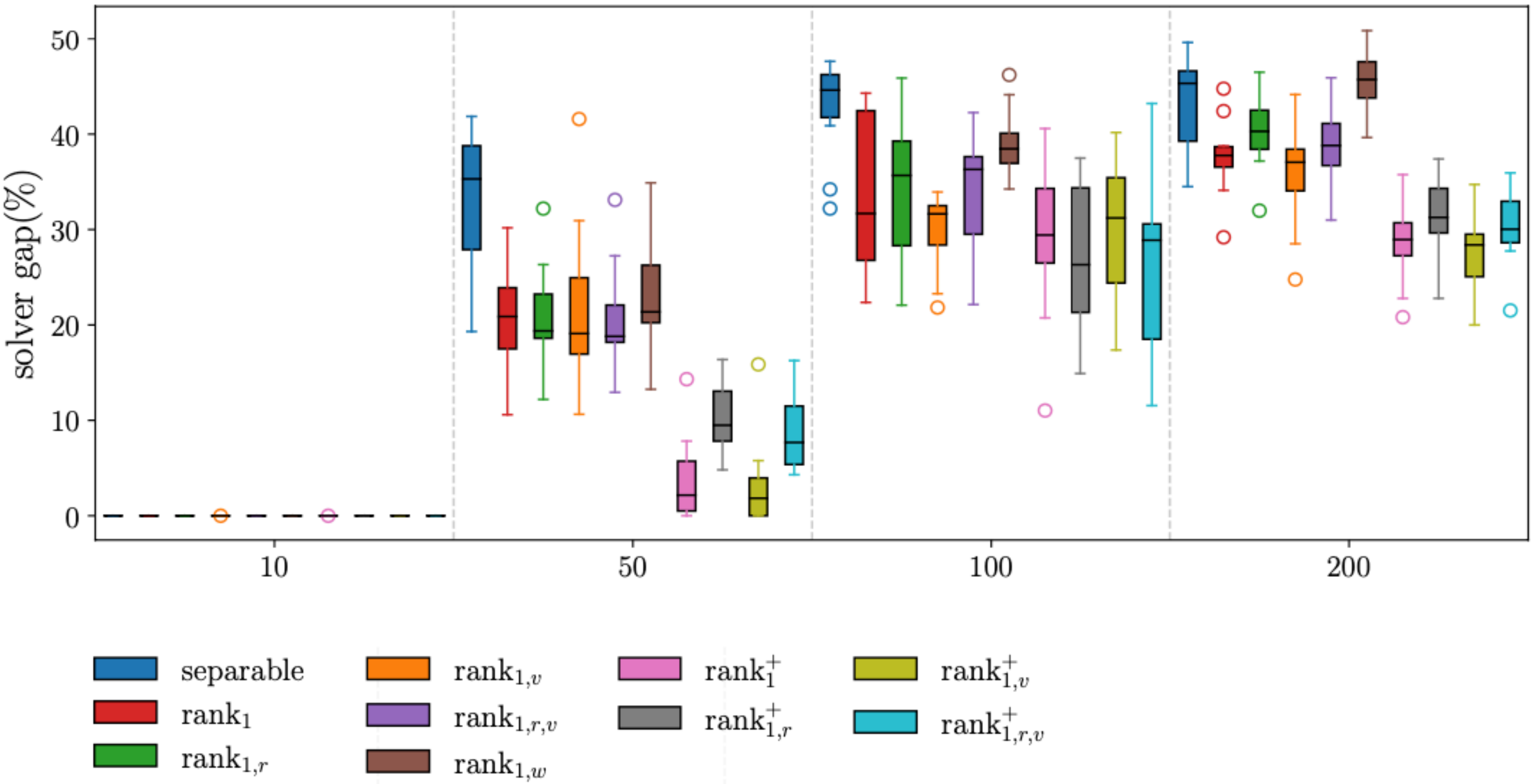
Numerical Results: Relaxations



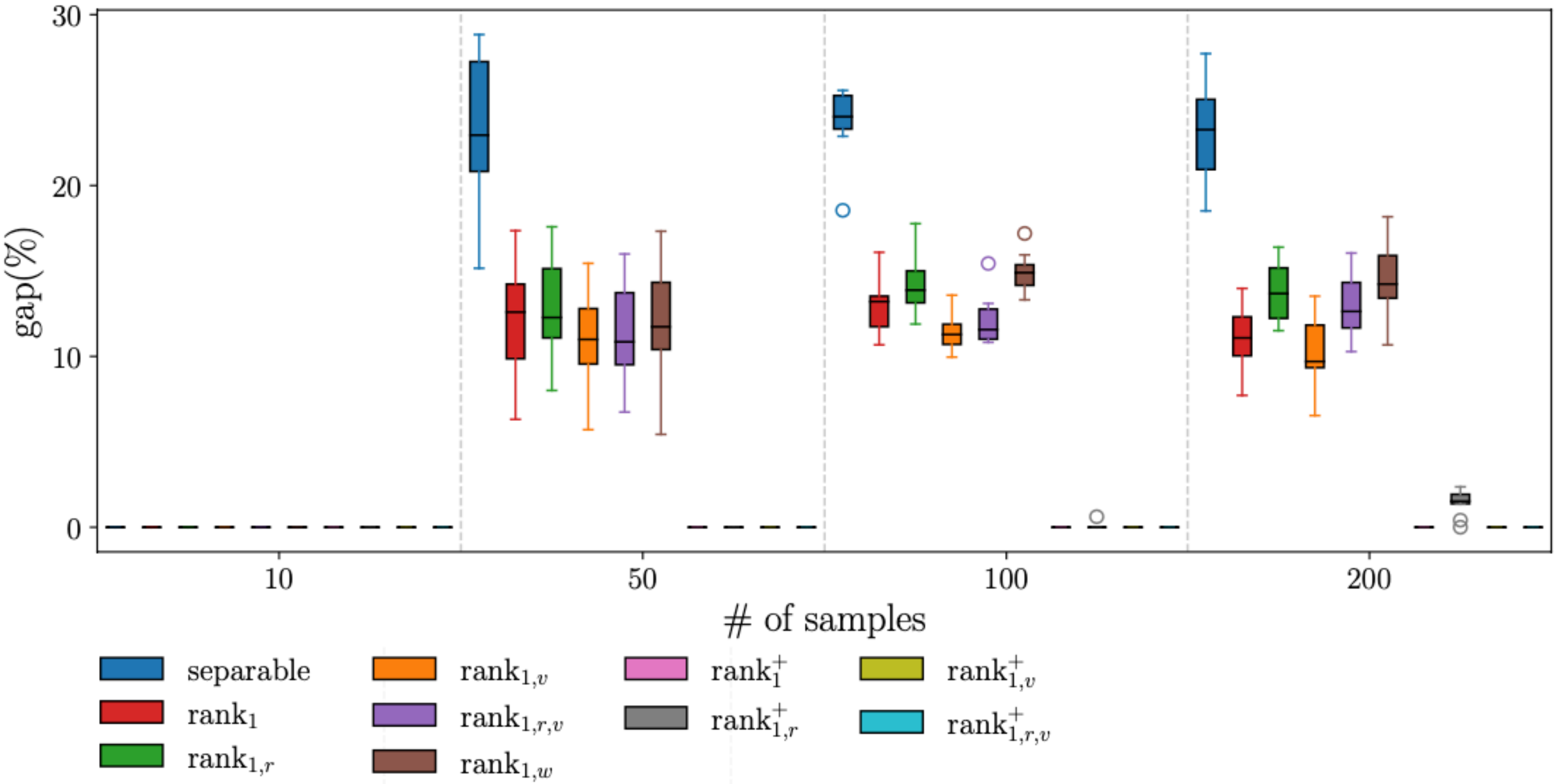
Numerical Results: Relaxations



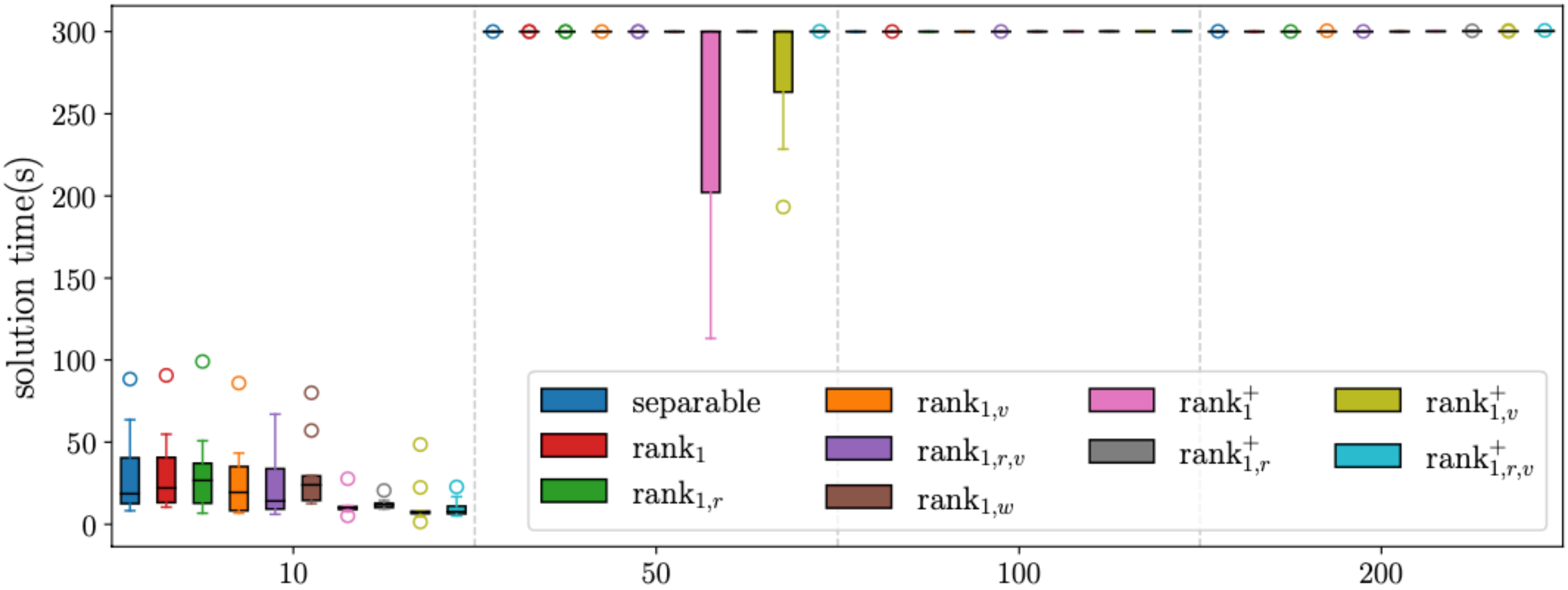
Numerical Results: Branch & Bound



Numerical Results: Branch & Bound



Numerical Results: Branch & Bound



References

[1] A. Atamtürk and A. Gómez. Supermodularity and valid inequalities for quadratic optimization with indicators. *Mathematical Programming (Forthcoming)*, 2022.

[2] T. Bacci, A. Frangioni, C. Gentile, and K. Tavlaridis-Gyparakis. New Mixed-Integer Nonlinear Programming Formulations for the Unit Commitment Problems with Ramping Constraints. *Operations Research (Forthcoming)*, 2019.

[3] O. Günlük and J. Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124(1):183–205, 2010.

[4] S. Han and A. Gómez. Compact extended formulations for low-rank functions with indicator variables. *arXiv:2110.14884*, 2021.

[5] L. Wei, A. Gómez, and S. Küçükyavuz. On the convexification of constrained quadratic optimization problems with indicator variables. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 433–447, 2020.

[6] L. Wei, A. Gómez, and S. Küçükyavuz. Ideal formulations for constrained convex optimization problems with indicator variables. *Mathematical Programming*, 192(1):57–88, 2022.

[7] W. Xie and X. Deng. Scalable algorithms for the sparse ridge regression. *SIAM Journal on Optimization*, 30(4):3359–3386, 2020.