# Solving the Optimal Experiment Design Problem using Convex Mixed-Integer methods

Deborah Hendrych (ZIB, TUB), Mathieu Besançon (Inria, UGA), Sebastian Pokutta (ZIB, TUB)

hendrych@zib.de

Zuse Institute Berlin

STOGO 2025 Stockholm · September 3, 2025

Berlin Mathematics Research Center

# Think cinnamon rolls

# Outline

# Table of Contents

# Set-up

- The ultimate goal is fitting a regression model.

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \|A\theta - \mathbf{y}\|_2^2$$

- $A = [-\mathbf{a}_i-]_{i=1}^m$ with $\mathbf{a}_i \in \mathbb{R}^n$ is the *Experiment Matrix*. Assumed to have full column rank.
- $\theta$ is the unknown set of parameters to be found.
- $\mathbf{y}$ is the **not-yet** measured response.
- Issue: Running all $m$ experiments too costly and/or too time intensive.

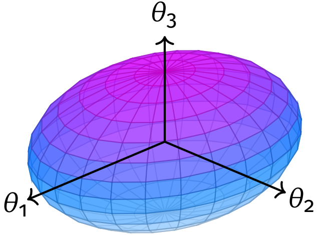
# *Some information theory later ...*

Pukelsheim 2006

Information matrix: $X(x) := A^\intercal \operatorname{diag}(\mathbf{x})A$

# D-Optimal Experiment Design

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & -\log\det\left(A^\top \operatorname{diag}(\mathbf{x}) A\right) \\
\text{s.t.} \quad & \sum_{i=1}^{m} x_i = N \\
& \mathbf{l} \le \mathbf{x} \le \mathbf{u} \\
& \mathbf{x} \in \mathbb{N}_0^m,
\end{aligned}
\tag{D}
$$

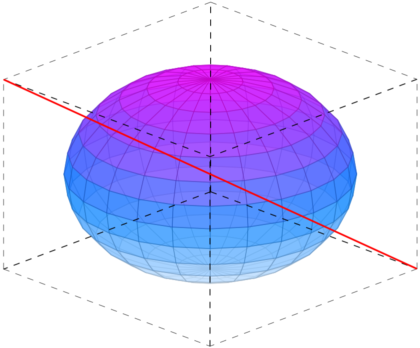



Ponte, Fampa, and Lee 2025; Ahipaşaoğlu 2021; Li et al. 2024

# A-Optimal Experiment Design

$$
\begin{aligned}
\min_{\mathbf{x}} \ & \operatorname{tr}\left( \left(A^\top \operatorname{diag}(\mathbf{x}) A\right)^{-1} \right) \\
\text{s.t.} \ & \sum_{i=1}^{m} x_i = N \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{x} \in \mathbb{N}_0^m,
\end{aligned}
$$

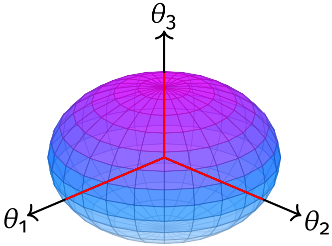
(A)



Ahipaşaoğlu 2015; Sagnol and Pauwels 2019; Nikolov, Singh, and Tantipongpipat 2022

# E-Optimal Experiment Design

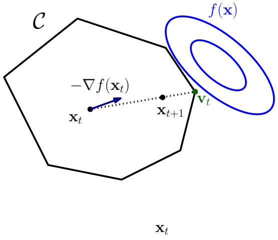
$$
\begin{aligned}
\min_{\mathbf{x}} \ & \operatorname{eigmax}\left(-A^\top \operatorname{diag}(\mathbf{x}) A\right) \\
\text{s.t.} \ & \sum_{i=1}^{m} x_i = N \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{x} \in \mathbb{N}_0^m,
\end{aligned}
\tag{D}
$$




Ponte, Fampa, and Lee 2025; Ahipaşaoğlu 2021; Li et al. 2024

# Table of Contents

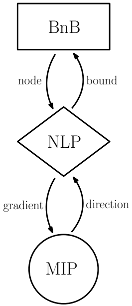
# Integer Frank-Wolfe: `Boscia.jl`




Frank and Wolfe 1956; Levitin and Polyak 1966; Braun et al. 2022

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{s.t. } \mathbf{x} \in \mathcal{X}$$
$$\mathbf{x}_j \in \mathbb{Z} \quad \forall j \in J$$

- $f$ continuously differentiable, convex, $L$-smooth.
- $\mathcal{X}$ is a compact convex set, usually polyhedral.




Schematic of the algorithm.

Hendrych, Troppens, et al. 2025

# Table of Contents

# Feasible region vs Domain

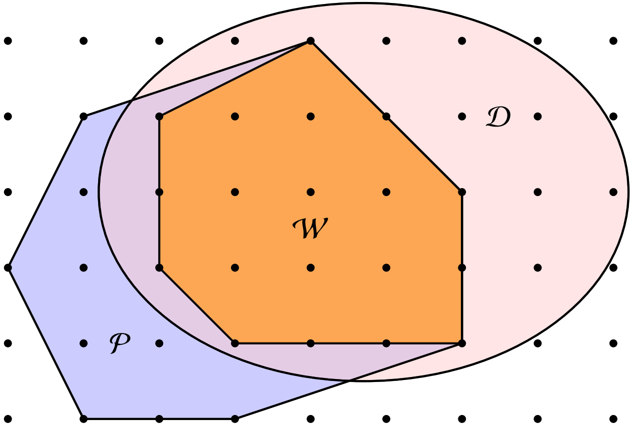



Figure: A schematic representation of the feasible region $\mathcal{P}$, the domain of the objective $\mathcal{D}$ and the convex hull of vertices that are both feasible and in the domain denoted as $\mathcal{W}$.

# Generalized self-concordance

By Carderera, Besançon, and Pokutta 2021, we also have convergence if the objective is generalized self-concordant. Self-concordance was already proved for the $-\log\det(X)$ on $\mathbb{S}^n_{++}$.

# Generalized self-concordance

By Carderera, Besançon, and Pokutta 2021, we also have convergence if the objective is generalized self-concordant. Self-concordance was already proved for the $-\log\det(X)$ on $\mathbb{S}^n_{++}$.

**Theorem (GSC A-criterion** Hendrych, Besançon, and Pokutta 2024**)**

*The function $g(X) = \operatorname{tr}(X^{-p})$, with $p > 0$, is $\left(3, \frac{(p+2)\sqrt[4]{a^{2p}n}}{\sqrt{p(p+1)}}\right)$-generalized self-concordant on $\operatorname{dom}(g) = \left\{X \in \mathbb{S}^n_{++} \mid 0 \prec X \preccurlyeq aI\right\}$ where $a \in \mathbb{R}_{>0}$ bounds the maximum eigenvalue of $X$.*

# Other approaches

## Co-BnB Ahipaşaoğlu 2021

- Branch-and-Bound with customized coordinate-descent-like algorithm for the node problems, exploits connection to Minimum Volume Enclosing Ellipsoid Problem.
- Developed for $m >> n$.

## Outer Approximation approaches

- Using the mixed-integer convex solver with conic certificates `Pajarito.jl` Coey, Kapelevich, and Vielma 2022a.
  1. **Direct Conic:** Natural formulation via `Hypatia.jl` Coey, Kapelevich, and Vielma 2022b.
  2. **SOCP:** as introduced in Sagnol 2011; Sagnol and Harman 2015.

# Experimental Results I

## Termination over time for the A-Optimal Problem

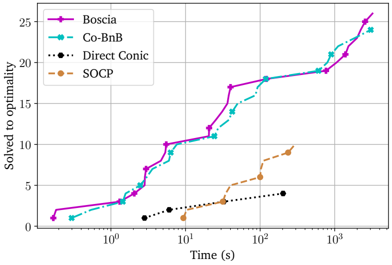



Figure: Independent data

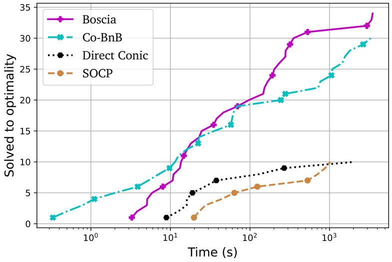



Figure: Correlated data

# Experimental Results II

## Termination over time for the D-Optimal Problem

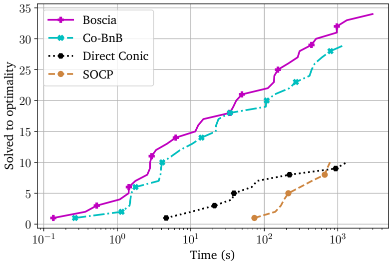



Figure: Independent data

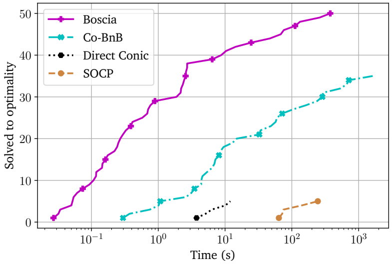



Figure: Correlated data

# Table of Contents

# E-Criterion and smoothing

## Issue

The subgradients of $\lambda_i(X)$ are the corresponding eigenvectors.
$\Rightarrow$ the gradient $\nabla\lambda_i$ only exists if $\mathrm{mult}(\lambda_i) = 1$.

# E-Criterion and smoothing

## Issue

The subgradients of $\lambda_i(X)$ are the corresponding eigenvectors.
$\Rightarrow$ the gradient $\nabla \lambda_i$ only exists if $\mathrm{mult}(\lambda_i) = 1$.

## Solution - Smoothing

We use the smoothing introduced in Nesterov 2007.

$$\max \lambda_{\min}(X) \Leftrightarrow \min \lambda_{\max}(-X) := f(X)$$

# E-Criterion and smoothing

## Issue

The subgradients of $\lambda_i(X)$ are the corresponding eigenvectors.
$\Rightarrow$ the gradient $\nabla\lambda_i$ only exists if $\mathrm{mult}(\lambda_i) = 1$.

## Solution - Smoothing

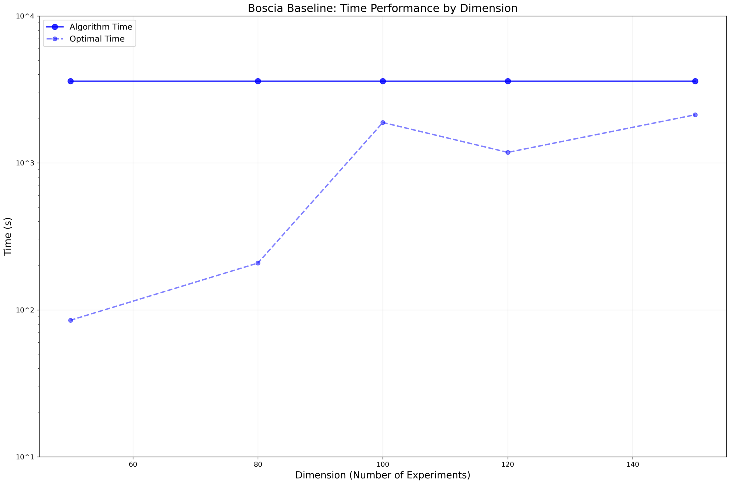
We use the smoothing introduced in Nesterov 2007.

$$\max \lambda_{\min}(X) \Leftrightarrow \min \lambda_{\max}(-X) := f(X)$$

$$f_\mu(X) = \mu \log\left(\sum_{i=1}^{n} e^{-\lambda_i(X)/\mu}\right) - \mu \log n \tag{1}$$

$$f_\mu(X) \leq f(X) \leq f_\mu(X) + \mu \log n \quad \forall X \in \mathbb{S}^n \tag{2}$$

# Initial Computational Experiments I

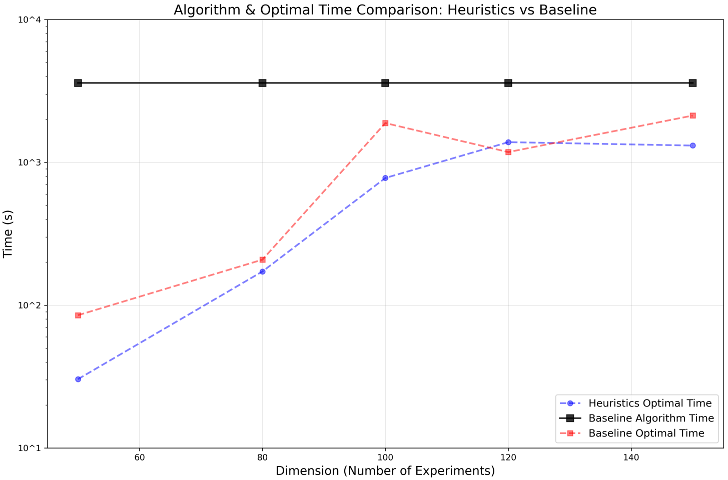


# Initial Computational Experiments II

Table: Heuristic Performance Comparison

| Heuristic | Optimal Time (s) | Dual Gap |
|---|---|---|
| Baseline | 6.092e+02 $*$/ 5.825 | 3.204e-01 $*$/ 1.794 |
| Fedorov<br>Fedorov 2013 | 1.979e+03 $*$/ 2.236 | 6.231e-01 $*$/ 2.840 |
| Follow<br>Subgradient | 1.738e+03 $*$/ 2.486 | 6.302e-01 $*$/ 2.810 |
| Pipage<br>Rounding<br>Brown, Laddha, and Singh 2024 | 2.114e+03 $*$/ 1.615 | 9.723e-01 $*$/ 2.765 |
| Simple<br>Randomized<br>Rounding<br>Lamperski, Yang, and Prokopyev n.d. | **4.962e+02** $*$/ **6.893** | **1.510e-01** $*$/ **1.752** |

# Initial Computational Experiments III




ZUSE INSTITUTE BERLIN

# Table of Contents

# Outlook

## E-Optimal Design

- Adapt lower bounds from Li 2025.
- Lower bound on the distance in the objective.
- More extensive experiments.

# Outlook

## E-Optimal Design

- Adapt lower bounds from Li 2025.
- Lower bound on the distance in the objective.
- More extensive experiments.

## On the Optimal Design side

- Sequential designs.
- Knowledge of experiment matrix $A \to$ Robustness.
- Non-Linear regression and nearly convex functions.

# Outlook

## E-Optimal Design

- Adapt lower bounds from Li 2025.
- Lower bound on the distance in the objective.
- More extensive experiments.

## On the Optimal Design side

- Sequential designs.
- Knowledge of experiment matrix $A \longrightarrow$ Robustness.
- Non-Linear regression and nearly convex functions.

## On the general `Boscia.jl` side

- Extended smoothness conditions.
- Handling of non-convex objectives.

ZUSE INSTITUTE BERLIN

# *Thank you for your attention!*

Convex mixed-integer optimization with Frank–Wolfe methods

Solving the Optimal Experiment Design Problem with Mixed-Integer Convex Methods

ZUSE INSTITUTE BERLIN

# References I


Ahipaşaoğlu, Selin Damla (2015). “A first-order algorithm for the A-optimal experimental design problem: a mathematical programming approach”. In: *Statistics and Computing* 25.6, pp. 1113–1127.

— (2021). “A branch-and-bound algorithm for the exact optimal experimental design problem”. In: *Statistics and Computing* 31.5, p. 65.

Braun, Gábor et al. (2022). “Conditional gradient methods”. In: *arXiv preprint arXiv:2211.14103*.

Brown, Adam, Aditi Laddha, and Mohit Singh (2024). “Maximizing the minimum eigenvalue in constant dimension”. In: *arXiv preprint arXiv:2401.14317*.

Carderera, Alejandro, Mathieu Besançon, and Sebastian Pokutta (2021). “Simple steps are all you need: Frank–Wolfe and generalized self-concordant functions”. In: *Advances in Neural Information Processing Systems* 34, pp. 5390–5401.

Coey, Chris, Lea Kapelevich, and Juan Pablo Vielma (2022a). “Performance enhancements for a generic conic interior point algorithm”. In: *Mathematical Programming Computation*. DOI: https://doi.org/10.1007/s12532-022-00226-0.

— (2022b). “Solving natural conic formulations with Hypatia.jl”. In: *INFORMS Journal on Computing* 34.5, pp. 2686–2699. DOI: https://doi.org/10.1287/ijoc.2022.1202.

Coey, Chris, Miles Lubin, and Juan Pablo Vielma (2020). “Outer approximation with conic certificates for mixed-integer convex problems”. In: *Mathematical Programming Computation* 12.2, pp. 249–293.

Fedorov, Valerii Vadimovich (2013). *Theory of optimal experiments*. Elsevier.

Frank, Marguerite and Philip Wolfe (1956). “An algorithm for quadratic programming”. In: *Naval research logistics quarterly* 3.1-2, pp. 95–110.

# References II

Hendrych, Deborah, Mathieu Besançon, and Sebastian Pokutta (2024). "Solving the Optimal Experiment Design Problem with Mixed-Integer Convex Methods". en. In: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. DOI: 10.4230/LIPICS.SEA.2024.16. arXiv: 2312.11200. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.SEA.2024.16.

Hendrych, Deborah, Hannah Troppens, et al. (June 28, 2025). "Convex Integer Optimization with Frank-Wolfe Methods". In: *Mathematical Programming Computation*. DOI: 10.1007/s12532-025-00288-w. arXiv: 2208.11010 [math.OC]. URL: https://link.springer.com/article/10.1007/s12532-025-00288-w.

Lamperski, Jourdain, Haeseong Yang, and Oleg Prokopyev (n.d.). "Simple Randomized Rounding for Max-Min Eigenvalue Augmentation". In: *Forty-second International Conference on Machine Learning*.

Levitin, Evgeny S and Boris T Polyak (1966). "Constrained minimization methods". In: *USSR Computational mathematics and mathematical physics* 6.5, pp. 1–50.

Li, Yongchun (2025). "Strong Formulations and Algorithms for Regularized A-optimal Design". In: *arXiv preprint arXiv:2505.14957*.

Li, Yongchun et al. (2024). "D-Optimal Data Fusion:: Exact and Approximation Algorithms". In: *INFORMS Journal on Computing* 36.1, pp. 97–120.

Nesterov, Yurii (2007). "Smoothing technique and its applications in semidefinite optimization". In: *Mathematical Programming* 110.2, pp. 245–259.

# References III


Nikolov, Aleksandar, Mohit Singh, and Uthaipon Tantipongpipat (2022). "Proportional volume sampling and approximation algorithms for A-optimal design". In: *Mathematics of Operations Research* 47.2, pp. 847–877.

Ponte, Gabriel, Marcia Fampa, and Jon Lee (2025). "Branch-and-bound for D-Optimality with fast local search and variable-bound tightening". In: *Mathematical Programming*, pp. 1–38.

Pukelsheim, Friedrich (2006). *Optimal design of experiments*. SIAM.

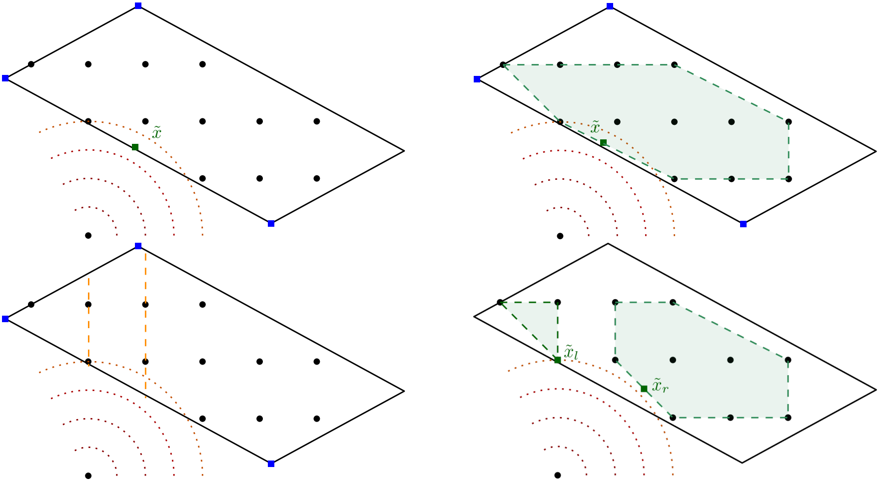
Sagnol, Guillaume (2011). "Computing optimal designs of multiresponse experiments reduces to second-order cone programming". In: *Journal of Statistical Planning and Inference* 141.5, pp. 1684–1708.

Sagnol, Guillaume and Radoslav Harman (2015). "Computing exact D-optimal designs by mixed integer second-order cone programming". In: *The Annals of Statistics* 43.5, pp. 2198–2224.

Sagnol, Guillaume and Edouard Pauwels (2019). "An unexpected connection between Bayes A-optimal designs and the group lasso". In: *Statistical Papers* 60.2, pp. 565–584.

Zhao, Renbo (2025). "New Analysis of an Away-Step Frank-Wolfe Method for Minimizing Log-Homogeneous Barriers". In: *Mathematics of Operations Research*.

# Convex relaxation vs integer hull



# Convergence results

| Objective | Convergence | Linear Convergence |
|---|---|---|
| $f(\mathbf{x}) = -\log\det(X(\mathbf{x}))$ | ✓ | ✓* |
| $g(\mathbf{x}) = \mathrm{tr}(X(\mathbf{x})^{-p})$ | ✓ | Not guaranteed |
| $k(\mathbf{x}) = \log\left(\mathrm{tr}(X(\mathbf{x})^{-p})\right)$ | Conjecture | Not guaranteed |

Table: Convergence of Frank-Wolfe on different objective functions.

*We adapted the proof from Zhao 2025 to the Blended Pairwise Conditional Gradient (BPCG).

Hendrych, D. and Besançon, M. and Pokutta, S. (2025). "Solving the Optimal Experiment Design Problem with Mixed-Integer Convex Methods." https://arxiv.org/abs/2312.11200

# What do we need for convergence?

In short: Frank-Wolfe has to converge.

$$f(X) = -\log\det(X) \quad \text{and} \quad g(X) = \operatorname{tr}(X^{-p}) \quad \text{and} \quad k(X) = \log\operatorname{tr}(X^{-p})$$

We need:

- Convexity of the objective functions. $\checkmark$
- $L$-smoothness of the objective functions. $\rightarrow$ Issue with the domain.
- Generalized self-concordance of the objective functions.
- Strong convexity/Sharpness of the objective functions.

Hendrych, D. and Besançon, M. and Pokutta, S. (2025). "Solving the Optimal Experiment Design Problem with Mixed-integer Convex Methods." `https://arxiv.org/abs/2312.11200`

# $L$-smoothness

**Theorem ($L$-smoothness for the Fusion Problems)**

*The functions* $f_F(\mathbf{x}) = \log\det(X_C(\mathbf{x}))$, $g_F(\mathbf{x}) = \operatorname{tr}(X_C(\mathbf{x})^{-p})$ *and* $k_F(\mathbf{x}) = \log\left(\operatorname{tr}(X_C(\mathbf{x})^{-p})\right)$ *are* $L$*-smooth on* $x \in \mathbb{R}_{\geq 0}$.

# $L$-smoothness

## Theorem ($L$-smoothness for the Fusion Problems)

*The functions* $f_F(\mathbf{x}) = \log\det(X_C(\mathbf{x}))$, $g_F(\mathbf{x}) = \mathrm{tr}(X_C(\mathbf{x})^{-p})$ *and* $k_F(\mathbf{x}) = \log\left(\mathrm{tr}(X_C(\mathbf{x})^{-p})\right)$ *are* $L$*-smooth on* $x \in \mathbb{R}_{\geq 0}$.

## Theorem (Local $L$-smoothness for the Optimal Problems)

*The functions* $f(X(\mathbf{x}))$, $g(X(\mathbf{x}))$ *and* $k(X(\mathbf{x}))$ *are locally* $L$*-smooth on*

$$\mathcal{L}_0 = \{\mathbf{x} \in \mathcal{D} \cap \mathcal{P} \mid (*)(\mathbf{x}) \leq (*)(\mathbf{x}_0)\}$$

*where* $(*)$ *is a placeholder for each function and* $x_0 \in \mathcal{D}$ *an initial point.*

# Sharpness

## Theorem (Strong convexity)

*The functions* $f(X) = -\log\det(X)$ *and* $g(X) = \mathrm{tr}(X^{-p})$, $p > 0$, *are strongly convex on* $D := \{X \in \mathbb{S}^n_{++} \mid \lambda_{\max}(X) \leq \alpha\}$.
*The function* $k(X) = \log\mathrm{tr}(X^{-p})$, $p > 0$, *is strongly convex on* $D := \{X \in \mathbb{S}^n_{++} \mid \lambda_{\max}(X) \leq \alpha, \kappa(X) \leq \kappa\}$.

ZUSE INSTITUTE BERLIN

# Sharpness

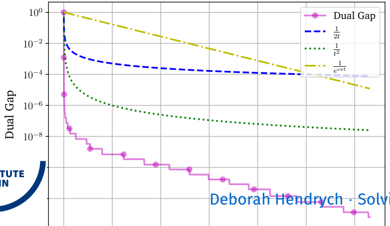
## Theorem (Strong convexity)

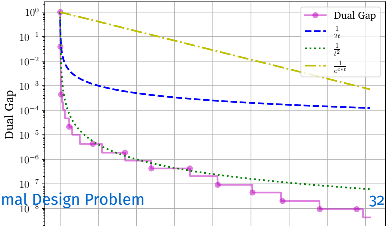
*The functions $f(X) = -\log\det(X)$ and $g(X) = \mathrm{tr}(X^{-p})$, $p > 0$, are strongly convex on $D := \{X \in \mathbb{S}^n_{++} \mid \lambda_{\max}(X) \leq \alpha\}$.*
*The function $k(X) = \log\mathrm{tr}(X^{-p})$, $p > 0$, is strongly convex on $D := \{X \in \mathbb{S}^n_{++} \mid \lambda_{\max}(X) \leq \alpha, \kappa(X) \leq \kappa\}$.*

## Theorem

*The compositions of $f$, $g$ and $k$ with the information matrix $X(\mathbf{x})$, respectivily, are sharp on $\mathcal{W}$.*

In general, the compositions do not have a unique minimizer.



# A Custom Branch-and-Bound for OEDP under Matrix Means (Co-BnB)

Ahipaşaoğlu 2021

- Coordinate-Descent-like algorithm for the nodes.
- Developed for $m \gg n$.

$$\begin{aligned} \max_{\mathbf{w}} \ & \log(\phi(X(\mathbf{w}))) \\ \text{s.t.} \ & \sum_{i=1}^{m} w_i = 1 \\ & \mathbf{w} \in [0,1]^m \\ & N\mathbf{w} \in \mathbb{Z}^m_{\geq 0} \end{aligned} \tag{M-OEDP}$$

- $\mathbf{w}$ can be interpreted as a probability distribution.
- Exploit connection to the Minimum Volume Enclosing Ellipsoid Problem (MVEP) for the termination criteria.

# Direct Conic Formulation

**coey2022conic**; Coey, Kapelevich, and Vielma 2022a; Coey, Lubin, and Vielma 2020

- `Pajarito.jl` is a mixed-integer convex solver with conic certificates.
- `Hypatia.jl` is an interior point solver for conic optimization problems.
- D-Criterion:

$$\mathcal{K}_{\log\det} := \mathrm{cl}\left\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{S}^n_{++} \mid u \leq v \log\det(W/v)\right\}$$

- A-Criterion: Dual of

$$\mathcal{K}_{\mathrm{sepspec}} := \mathrm{cl}\left\{(u, v, w) \in \mathbb{R} \times \mathbb{R}_{>0} \times \mathrm{int}(\mathcal{Q}) \mid u \geq v\varphi(w/v)\right\}$$

- $\mathcal{Q}$ is the PSD cone and $\varphi$ is the negative square root.
- The convex conjugate of the negative square root is the trace inverse.

# Second-Order Conic Formulation (SOCP)

Sagnol 2011; Sagnol and Harman 2015

- In Sagnol 2011 the SOCP formulation for the continuous problem was introduced.
- The SOCP formulation of the exact formulation of OEDP was shown in Sagnol and Harman 2015.
- In theory, a nice result, but in practice, the problem size becomes much larger.
- For the A-Optimal Problem, we have $2m(n+1)$ variables and $2(n+1)+m$ constraints.
- For the D-Optimal Problem, we have $2m(1+n)+n^2+1$ variables and $n(m+1)+3m+4$ constraints.

# SOCP A-Criterion

$$
\begin{aligned}
\min_{\substack{\boldsymbol{\mu}\in\mathbb{R}^m \\ \mathbf{x}\in\mathbb{Z}^m \\ \forall i\in[m]\ Y_i\in\mathbb{R}^{1xn}}} \quad & \sum_{i=1}^{m} t_i \\
\text{s.t.} \quad & \sum_{i=1}^{m} A_i Z_i = I \\
& \forall i \in [m],\ \|Z_i\|_F^2 \leq t_i x_i \\
& \sum_{i=1}^{m} x_i = N
\end{aligned}
\tag{A-SOCP}
$$

# SOCP D-Criterion

$$
\begin{aligned}
\min_{\substack{\mathbf{w}\in\mathbb{R}^m_{\geq 0},\mathbf{x}\in\mathbb{Z}^m \\ J\in\mathbb{R}^{nxn} \\ \forall i\in[m]\ Z_i\in\mathbb{R}^{1xn} \\ \forall i\in[m], j\in[n]\ t_{ij}\in\mathbb{R}_{\geq 0}}} \quad & \prod_{i=1}^{m} J_{ii}^{1/m} \\
\text{s.t.} \quad & \sum_{i=1}^{m} A_i Z_i = J \\
& J \text{ lower triangle matrix} \\
& \forall i \in [m], j \in [n]\ \|Z_i e_j\|_F^2 \leq t_{ij} w_i \\
& \forall j \in [n]\ \sum_{i=1}^{m} t_{ij} \leq J_{jj} \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \text{ and } \sum_{i=1}^{m} w_i = 1
\end{aligned}
\tag{D-SOCP}
$$

# Effect of the conditioning of the problem

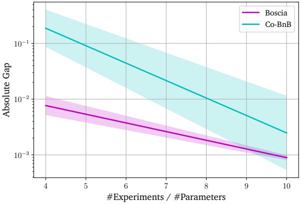



Figure: A IND

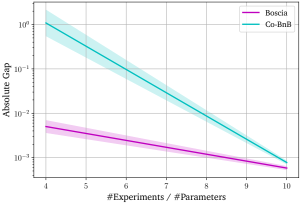



Figure: D IND

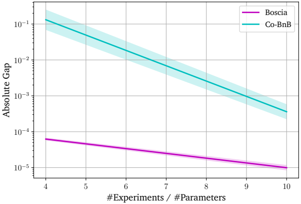



Figure: A CORR

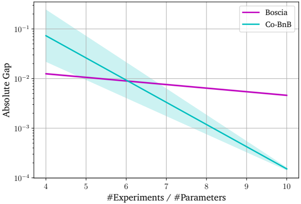



Figure: D CORR