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

EUROPT 2025 Southampton · July 2, 2025

Berlin Mathematics Research Center

MATH+

ZUSE INSTITUTE BERLIN

Think scones

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.
- θ 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.

ZUSE INSTITUTE BERLIN

Information Matrix I

Information Matrix

The *Information Matrix* is a linear map $X : \mathbb{R}^m_{\geq 0} \to \mathbb{R}^{n\times n}$.

$$X(\mathbf{x}) = \sum_{i=1}^{m} x_i \mathbf{a}_i \mathbf{a}_i^\top = A^\top \operatorname{diag}(\mathbf{x}) A$$

- Variable $x_i \in \mathbb{N}_0$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.
- A design $\mathbf{x}$ is "useful" if $X(\mathbf{x})$ is regular.

Information Matrix II

The Fusion Information Matrix

Let $C \in \mathbb{S}^n_{++}$ denote the already performed experiments.

$$X_C(\mathbf{x}) = C + A^\intercal \operatorname{diag}(\mathbf{x}) A = C + X(\mathbf{x})$$

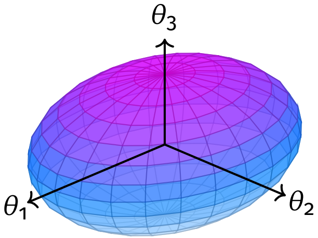
- Even for $\mathbf{x} = 0$, $X_C(\mathbf{x})$ is PD.
- You can think of $C = A^\intercal \operatorname{diag}(\mathbf{l}) A$ where $\mathbf{l}$ are some non-trivial lower bounds on the experiments.

Some information theory later ...

Pukelsheim 2006

D-Optimal Experiment Design

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

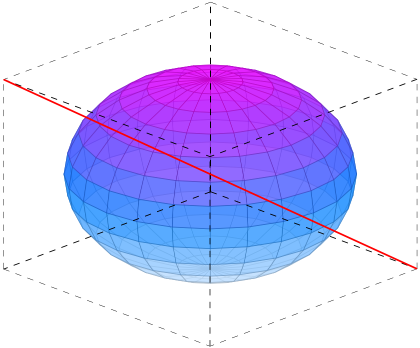


Ponte, Fampa, and Lee 2025; Ahipaşaoğlu 2021; Li et al. 2024

A-Optimal Experiment Design

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

(A)



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

Generalized-Trace-Inverse (GTI) Optimal Design

We generalize the A-optimal formulation for any real $p > 0$:

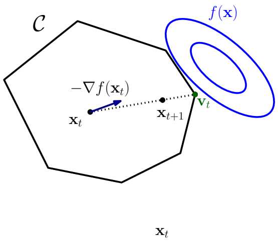
$$\begin{aligned} \min_{\mathbf{x}} \ & \log\left(\operatorname{tr}\left((X(\mathbf{x}))^{-p}\right)\right) \\ \text{s.t.} \ & \sum_{i=1}^{m} x_i = N \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\ & \mathbf{x} \in \mathbb{N}_0^m \end{aligned} \qquad \text{(logGTI-Opt)}$$

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

- $p = 1$ corresponds to the A-optimal problem.
- Also known as Kiefer's criteria.
- We conjecture that OEDP under the GTI-criterion is $\mathcal{NP}$-hard for any $p > 0$.

Table of Contents

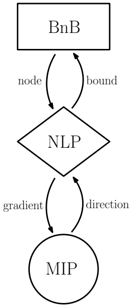
Integer Frank-Wolfe: `Boscia.jl`



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 et al. 2025

Table of Contents

Convergence results

Objective	Convergence	Linear Convergence
$f_F(\mathbf{x}) = \log\det(X_C(\mathbf{x}))$	✓	✓
$g_F(\mathbf{x}) = \operatorname{tr}(X_C(\mathbf{x})^{-p})$	✓	✓
$k_F(\mathbf{x}) = \log(\operatorname{tr}(X_C(\mathbf{x})^{-p}))$	✓	✓
$f(\mathbf{x}) = -\log\det(X(\mathbf{x}))$	✓	✓*
$g(\mathbf{x}) = \operatorname{tr}(X(\mathbf{x})^{-p})$	✓	Not guaranteed
$k(\mathbf{x}) = \log(\operatorname{tr}(X(\mathbf{x})^{-p}))$	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

Feasible region vs Domain

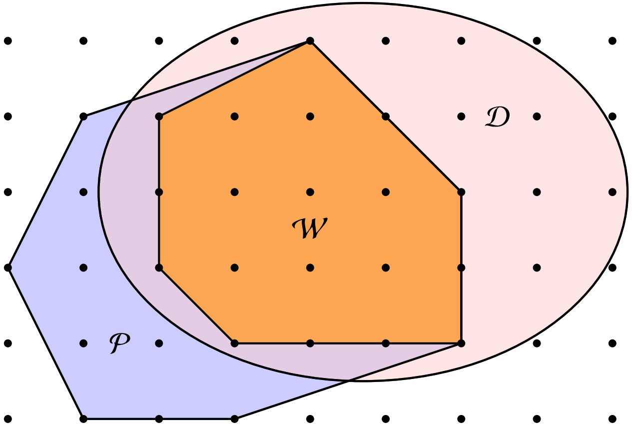


Figure: A schematic representation of the feasible region $\mathcal{P}$, the domain of the objective $\mathcal{D}$ and the convex hull of vertices that are both feasible and in the domain denoted as $\mathcal{W}$.

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.*

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 (Generalized self concordance A-criterion)

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.

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 (Generalized self concordance A-criterion)

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.

Conjecture: The function $k(X) = \log\operatorname{tr}(X^{-p})$, with $p > 0$, is generalized self-concordant on some bounded set of the PD cone.

Sharpness

Theorem (Strong convexity)

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

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.

Table of Contents

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.

Epigraph-based Outer Approximation: SCIP+OA

Bestuzheva et al. 2021; Kronqvist et al. 2019

- Requires Epigraph Formulation

$$
\begin{aligned}
\min_{t,\mathbf{x}}\ & t \\
\text{s.t.}\ & t \geq \phi(\mathbf{x}) \\
& \sum_{i=1}^{m} x_i = N \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{x} \in \mathbb{Z}^m_{\geq 0}.
\end{aligned}
$$

- Would require additional domain cuts, i.e. adding cutting planes to separate domain infeasible points. Hence, it's only used on the Fusion Problems.

Direct Conic Formulation

coey2022conic; Coey, Kapelevich, and Vielma 2022; 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:

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

- A-Criterion: Dual of

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

- $\mathcal{Q}$ is the PSD cone and φ 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.

Experimental Results I

Set up

- m between 50 and 120
- $n = \lfloor m/4 \rfloor$ and $n = \lfloor m/10 \rfloor$
- $\#$allowed experiments $N = \lfloor 1.5n \rfloor$.
- Independent and correlated data for the experiment matrix A.
- Five random seeds, leading to 50 instances per problem and data.
- Carried out in `Julia` with a time limit of 1 hour.

Experimental Results II

Termination over time for the A-Optimal Problem

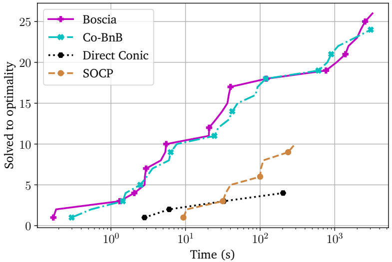


Figure: Independent data

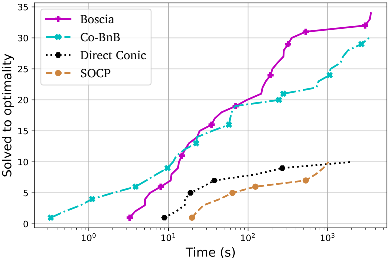


Figure: Correlated data

Experimental Results III

Termination over time for the D-Optimal Problem

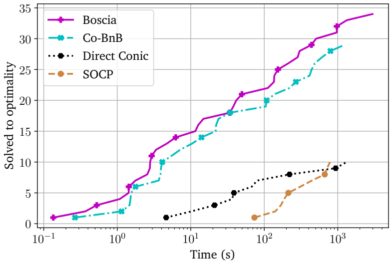


Figure: Independent data

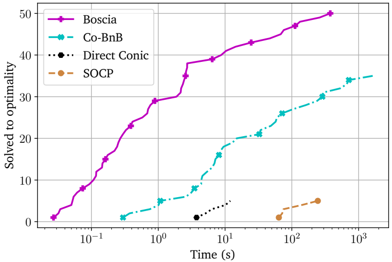


Figure: Correlated data

Experimental Results IV

Termination over time for the A-Fusion Problem

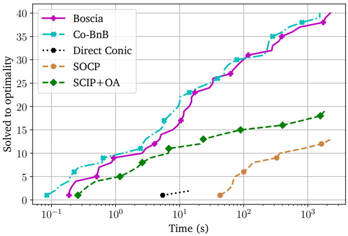


Figure: Independent data

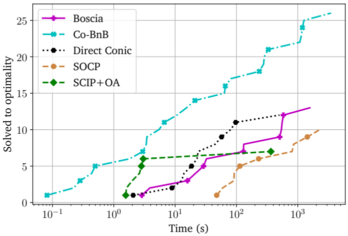


Figure: Correlated data

Experimental Results V

Termination over time for the D-Fusion Problem

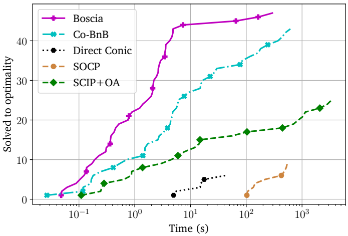


Figure: Independent data

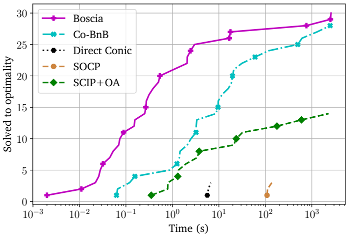


Figure: Correlated data

Table of Contents

Summary

- The Outer Approximation approaches are fast for small instances.
- Co-BnB is fast for instances where n is very small compared to m, i.e. $n = m/10$.
- `Boscia.jl` outperforms the other solvers, especially for medium to large scale instances.
- It also keeps the problem structure intact.
- Can easily handle additional constraints.

Outlook

On the Optimal Design side

- The E-Optimal Criterion and smoothing techniques.
- Sequential designs.
- Knowledge of experiment matrix $A \longrightarrow$ Robustness.
- Non-Linear regression and nearly convex functions.

On the general `Boscia.jl` side

- Handling of non-convex objectives.
- Extended smoothness conditions.
- Preprocessing.
- Gradient feasible region.

Thank you for your attention!

We’re Hiring!

- **International Student Research Visits (2-6 months):**
  - For undergraduates or Master’s students from other universities.
  - Requirements: Background in mathematics, CS, or related fields.
  - Apply 2-4 months before preferred start date, including CV, transcript, and motivational letter.
- **Bachelor and Master Theses (Berlin universities):**
  - Supervision for students at Berlin universities.
  - Requirements: Alignment with our research and engagement in respective courses.
  - Limited slots available; apply early with a CV, transcript, and motivational letter.
- **PhD Positions (Berlin universities):**
  - Research as a PhD student with potential teaching opportunities.
  - Requirements: Strong background in mathematics, CS, or related fields; specific research interests.
  - Apply 3-6 months before preferred start date; include detailed CV, transcripts, research statement, and recommendations.
- **Postdoctoral Research Opportunities:**
  - Collaborate on advanced research projects.
- **Interested? Contact Us!**
  - Email: `hendrych@zib.de`
  - Website: `https://iol.zib.de`
  - Include all relevant documents and information in your email.

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.

Bestuzheva, Ksenia et al. (2021). *The SCIP Optimization Suite 8.0*. arXiv: `2112.08872 [math.OC]`.

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

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 (2022). “Performance enhancements for a generic conic interior point algorithm”. In: *Mathematical Programming Computation*. DOI: `https://doi.org/10.1007/s12532-022-00226-0`.

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.

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

Hendrych, Deborah et al. (2025). “Convex mixed-integer optimization with Frank–Wolfe methods”. In: *Mathematical Programming Computation*.

References II

Kronqvist, Jan et al. (2019). “A review and comparison of solvers for convex MINLP”. In: *Optimization and Engineering* 20, pp. 397–455.

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

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.

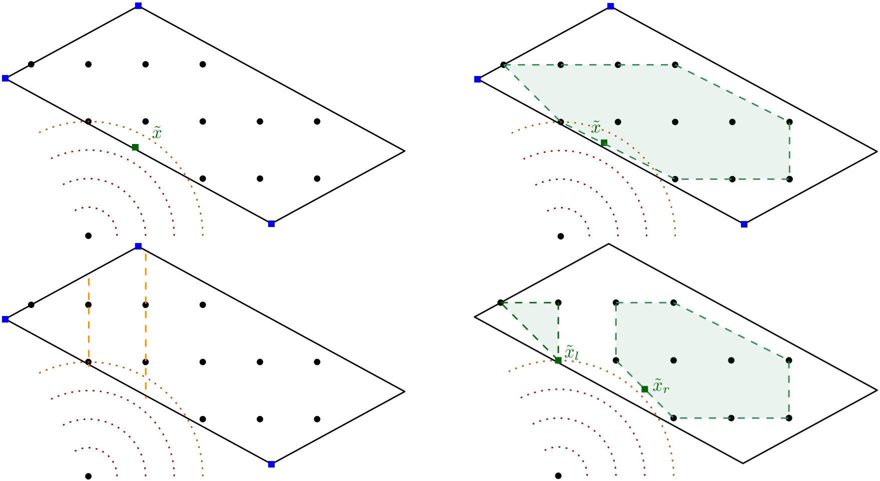
Sagnol, Guillaume (2011). “Computing optimal designs of multiresponse experiments reduces to second-order cone programming”. In: *Journal of Statistical Planning and Inference* 141.5, pp. 1684–1708.

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



SOCP A-Criterion

$$\min_{\substack{\boldsymbol{\mu}\in\mathbb{R}^m \\ \mathbf{x}\in\mathbb{Z}^m \\ \forall i\in[m]\ Y_i\in\mathbb{R}^{1xn}}} \quad \sum_{i=1}^{m} t_i$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^{m} A_i Z_i = I \\ & \forall i \in [m],\ \|Z_i\|_F^2 \le t_i x_i \\ & \sum_{i=1}^{m} x_i = N \end{aligned} \tag{A-SOCP}$$

ZUSE INSTITUTE BERLIN

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

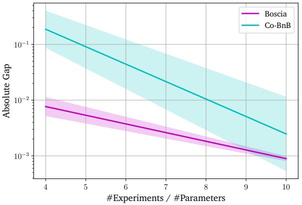


Figure: A IND

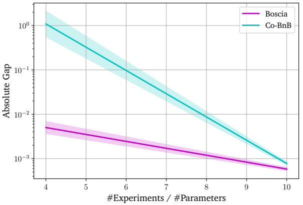


Figure: D IND

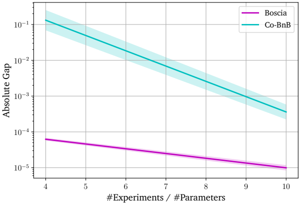


Figure: A CORR

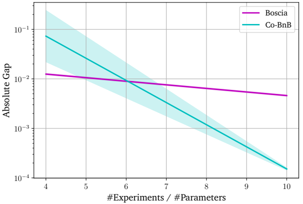


Figure: D CORR