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

Deborah Hendrych, Mathieu Besançon, Sebastian Pokutta

hendrych@zib.de

Zuse Institute Berlin

SEA 2024 · July 25, 2024

Think Pancakes

Outline

Table of Contents

Initial Setup and Motivation

- 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.
- Problem: Running all m experiments too costly and/or too time intensive.

Information Matrix

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

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

- Variable $x_i \in \mathbb{Z}_{\geq 0}$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.

Information Matrix

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

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

- Variable $x_i \in \mathbb{Z}_{\geq 0}$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.
- $A^\top A$ is positive definite (PD).
- $X(\mathbf{x})$ is positive semi definite (PSD).

Information Matrix

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

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

- Variable $x_i \in \mathbb{Z}_{\geq 0}$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.
- $A^\top A$ is positive definite (PD).
- $X(\mathbf{x})$ is positive semi definite (PSD).
- The inverse of $X(\mathbf{x})$ (if existent) is called the *Dispersion Matrix* $D(\mathbf{x})$ and is a measure of the variance of the parameters Ahipaşaoğlu 2021.
- A design $\mathbf{x}$ is "useful" if $X(\mathbf{x})$ is regular.

Information measure

- Function $\phi : \mathbb{S}^n_{++} \rightarrow \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:

Information measure

- Function $\phi : \mathbb{S}^n_{++} \to \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:
    - Respect Loewner Ordering: Let $D, B \in \mathbb{S}^n_+$. Then $D \succcurlyeq B$ if and only $D - B \in \mathbb{S}^n_+$. If $D \succcurlyeq B$, we require $\phi(D) \geq \phi(B)$.

Information measure

- Function $\phi : \mathbb{S}^n_{++} \to \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:
    - Respect Loewner Ordering: Let $D, B \in \mathbb{S}^n_+$. Then $D \succcurlyeq B$ if and only $D - B \in \mathbb{S}^n_+$. If $D \succcurlyeq B$, we require $\phi(D) \geq \phi(B)$.
    - Concave: Prohibit linear interpolation.

Information measure

- Function $\phi : \mathbb{S}^n_{++} \to \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:
    - Respect Loewner Ordering: Let $D, B \in \mathbb{S}^n_+$. Then $D \succcurlyeq B$ if and only $D - B \in \mathbb{S}^n_+$. If $D \succcurlyeq B$, we require $\phi(D) \geq \phi(B)$.
    - Concave: Prohibit linear interpolation.
    - Positively homogeneous: $\phi(\lambda X) = \lambda \phi(X)$ for all $\lambda \geq 0$.

Information measure

- Function $\phi : \mathbb{S}^n_{++} \to \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:
  - Respect Loewner Ordering: Let $D, B \in \mathbb{S}^n_+$. Then $D \succcurlyeq B$ if and only $D - B \in \mathbb{S}^n_+$. If $D \succcurlyeq B$, we require $\phi(D) \geq \phi(B)$.
  - Concave: Prohibit linear interpolation.
  - Positively homogeneous: $\phi(\lambda X) = \lambda \phi(X)$ for all $\lambda \geq 0$.
  - Non-negative and non-constant: Convention.

Information measure

- Function $\phi : \mathbb{S}^n_{++} \to \mathbb{R}$ where $\mathbb{S}^n_{++}$ is the cone of $n \times n$ positive definite matrix.
- Necessary properties Pukelsheim 2006:
    - Respect Loewner Ordering: Let $D, B \in \mathbb{S}^n_+$. Then $D \succcurlyeq B$ if and only $D - B \in \mathbb{S}^n_+$. If $D \succcurlyeq B$, we require $\phi(D) \geq \phi(B)$.
    - Concave: Prohibit linear interpolation.
    - Positively homogeneous: $\phi(\lambda X) = \lambda \phi(X)$ for all $\lambda \geq 0$.
    - Non-negative and non-constant: Convention.
    - Upper semi-continuous: The upper level sets $\{M \in \mathbb{S}^n_{++} \mid \phi(M) \geq \lambda\}$ are closed for all $\lambda \in \mathbb{R}$.

.

The Optimal Experiment Design Problem (OEDP)

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

- Let $m \gg N \geq n$ be the *number of allowed experiments*.
- $\mathbf{l}$ and $\mathbf{u}$ are lower and upper bounds, respectively. Usually, $\mathbf{l} = 0$.
- Denote $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^m \mid \langle 1, \mathbf{x} \rangle = N, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$.

The Optimal Experiment Design Problem (OEDP)

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

- Let $m \gg N \geq n$ be the *number of allowed experiments*.
- $\mathbf{l}$ and $\mathbf{u}$ are lower and upper bounds, respectively. Usually, $\mathbf{l} = 0$.
- Denote $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^m \mid \langle 1, \mathbf{x} \rangle = N, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$.
- If $C = A^\intercal \operatorname{diag}(\mathbf{l}) A$ is positive definite, we have the *Fusion Problem*.

The A-Optimality and D-Optimality Criteria

The D-Optimal Problem ($p = 0$)

$$\begin{aligned} \min_{\mathbf{x}} \ & -\log\det\left(X(\mathbf{x})\right) \\ \text{s.t.} \ & \mathbf{x} \in \mathcal{P} \cap \mathbb{Z}^m_{\geq 0} \end{aligned} \tag{DO}$$

minimal volume of standard ellipsoidal confidence region of θ

Ponte, Fampa, and Lee 2023; Ahipaşaoğlu 2021; Li et al. 2022

The A-Optimal Problem ($p = -1$)

$$\begin{aligned} \min_{\mathbf{x}} \ & \operatorname{Tr}\left(\left(X(\mathbf{x})\right)^{-1}\right) \\ \text{s.t.} \ & \mathbf{x} \in \mathcal{P} \cap \mathbb{Z}^m_{\geq 0} \end{aligned} \tag{AO}$$

minimize the diagonal of the bounding box of the confidence ellipsoid

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

Both are $\mathcal{NP}$-hard. (Welch 1982; Nikolov, Singh, and Tantipongpipat 2022)

The D-Fusion Problem is also $\mathcal{NP}$-hard (Li et al. 2022). For the A-Fusion it is not known.

Table of Contents

Previous Approach

- Due to the hardness of the MINLP formulation, one often solves a continuous version.
- The *Limit* Problem is created by letting $N \to \infty$.
- $\mathbf{x}$ can then be found of as a probability distribution.
- Different rounding schemes and exchange heuristics can then be employed to generate integer solution.
- Problem 1: Sparsity of the probability solution is often required.
- Problem 2: No guarantee of the optimality.

Branch-and-Bound with Frank-Wolfe: `Boscia.jl`

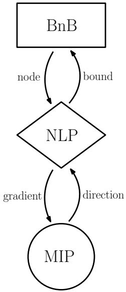
Hendrych et al. 2023

Problem Form

$$\begin{aligned}\min_{\mathbf{x}}\ & f(\mathbf{x}) \\ \text{s.t. } & \mathbf{x} \in \mathcal{X} \\ & x_j \in \mathbb{Z}\ \forall j \in J\end{aligned}$$

- f is a differentiable, non-linear, convex, L-smooth function.
- $\mathcal{X}$ is polyhedral with combinatorial and integrality constraints.

Illustration of the Solution Approach



- Frank-Wolfe variants as node solver.
- Frank-Wolfe implemented in `FrankWolfe.jl` Besançon, Carderera, and Pokutta 2022.
- Exploits Frank-Wolfe's error-adaptiveness.
- Frank-Wolfe enables warm-starting.
- Linear Minimization Oracle is a *Bounded (Mixed-Integer) Linear Minimization Oracle (BLMO)*, usually a MIP solver but can also be a combinatorial solver.
- Integer feasible solutions from the root node.

Is `Boscia.jl` applicable?

- We have L-smoothness for the Fusion Problems, not for the Optimal Problems.

Is `Boscia.jl` applicable?

- We have L-smoothness for the Fusion Problems, not for the Optimal Problems.
- By Carderera, Besançon, and Pokutta 2021, we also have convergence if the objective is generalized self-concordant.
- Known for the $-\log\det$ as the barrier of the PSD cone.

Is `Boscia.jl` applicable?

- We have L-smoothness for the Fusion Problems, not for the Optimal Problems.
- By Carderera, Besançon, and Pokutta 2021, we also have convergence if the objective is generalized self-concordant.
- Known for the $-\log\det$ as the barrier of the PSD cone.

Theorem

The function $g(X) = \operatorname{Tr}\left(X^{-p}\right)$, 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.

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

Ahipaşaoğlu 2021

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

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

- **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

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

- Approximate feasible region by linear cuts using the gradient of the objective.
- 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

Coey, Kapelevich, and Vielma 2022a; Coey, Kapelevich, and Vielma 2022b; 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 φ is the negative square root.
- The convex conjugate of the negative square root is the trace inverse.

Table of Contents

Set-up

- m between 50 and 120
- $n = \lfloor m/4 \rfloor$ and $n = \lfloor m/10 \rfloor$
- #allowed experiments $N = \lfloor 1.5n \rfloor$ in case of the Optimal Problems.
- #allowed experiments $N \in [m/20, m/3]$ in case of the Fusion Problem.
- 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

		Boscia		Co-BnB		Direct Conic		SCIP+OA	
Type	Corr.	% solved	Time (s)	% solved	Time (s)	% solved	Time (s)	% solved	Time (s)
A	no	**58 %**	**208.53**	42 %	640.68	14 %	1901.7		
A	yes	**82 %**	**98.5**	50 %	541.22	20 %	1591.74		
AF	no	**80 %**	**54.82**	78 %	82.96	12 %	2006.81	38 %	464.82
AF	yes	26 %	1359.35	**50 %**	**370.57**	20 %	1132.66	14 %	1471.59
D	no	**74 %**	**81.07**	58 %	442.28	24 %	732.57		
D	yes	**100 %**	**1.26**	68 %	223.34	10 %	755.88		
DF	no	**94 %**	**3.32**	86 %	38.28	14 %	1576.5	50 %	333.25
DF	yes	**60 %**	**50.68**	54 %	185.07	14 %	1761.18	28 %	753.56

Table: Relative number of instances solved and the geometric mean of the time shifted by 1 second for all solvers. Note that the average time includes the timeouts!

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.
- It can easily handle additional constraints.

Outlook

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

Thank you for your attention!

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.

Besançon, Mathieu, Alejandro Carderera, and Sebastian Pokutta (2022). "FrankWolfe.jl: A High-Performance and Flexible Toolbox for Frank–Wolfe Algorithms and Conditional Gradients". In: *INFORMS Journal on Computing.*

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

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). "Conic optimization with spectral functions on Euclidean Jordan algebras". In: *Mathematics of Operations Research.*

— (2022b). "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.

References II

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. (2023). *Convex mixed-integer optimization with Frank–Wolfe methods*. arXiv: 2208.11010 [math.OC].

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. (2022). "D-optimal Data Fusion: Exact and Approximation Algorithms". In: *arXiv preprint arXiv:2208.03589*.

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 (2023). "Branch-and-bound for D-Optimality with fast local search and variable-bound tightening". In: *arXiv preprint arXiv:2302.07386*.

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.

References III

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.

Welch, William J (1982). “Branch-and-bound search for experimental designs based on D optimality and other criteria”. In: *Technometrics* 24.1, pp. 41–48.

Matrix Means

Definition (Matrix Mean)

Let $B \in \mathbb{S}_+^n$ and let $\lambda(B)$ denote its eigenvalues. The *matrix mean* ϕ_p of B is defined as

$$\phi_p(B) = \begin{cases} \lambda_{\max}(B), & \text{for } p = \infty, \\ \left(\frac{1}{n}\operatorname{Tr}(B^p)\right)^{\frac{1}{p}}, & \text{for } p \neq 0, \pm\infty, \\ \det(B)^{\frac{1}{n}}, & \text{for } p = 0, \\ \lambda_{\min}(B), & \text{for } p = -\infty, \\ 0 & \text{for } p = [-\infty, 0] \text{ and } B \text{ singular.} \end{cases} \tag{1}$$

For $p \leq 1$, ϕ_p is a valid information measure (Pukelsheim 2006).

The Frank-Wolfe algorithm I

Frank and Wolfe 1956; Levitin and Polyak 1966

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{s.t. } \mathbf{x} \in \mathcal{C}$$

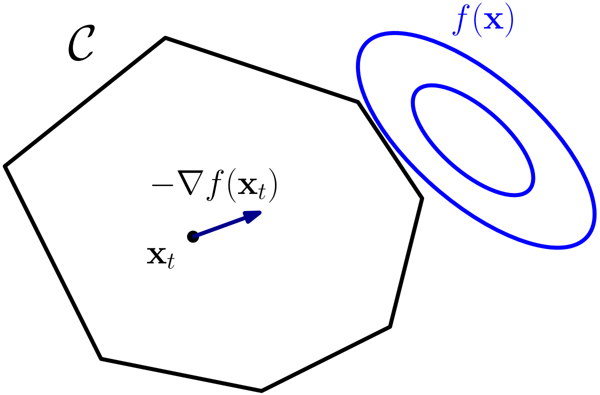
- $\mathcal{C}$ is a compact convex set.
- f continuously differentiable, convex function.
- Requires f to be L-smooth, i.e. gradient ∇f Lipschitz-continuous.
- Assumes that the Linear Minimization Oracle (LMO) $\min_{\mathbf{x} \in \mathcal{C}} \langle \mathbf{d}, \mathbf{x} \rangle$ is easy to solve.

The Frank-Wolfe Gap or Dual Gap

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \leq \min_{\mathbf{v} \in \mathcal{C}} \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{v} \rangle$$

The Frank-Wolfe algorithm II

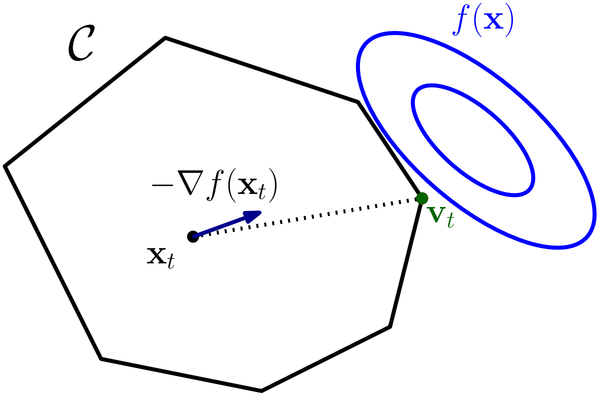
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm III

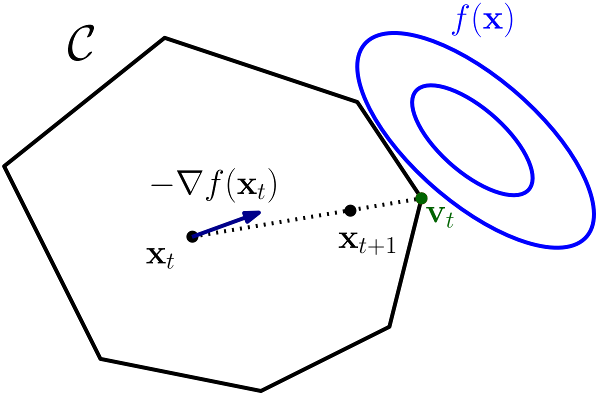
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm IV

Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

Plots I

Termination over time for the A-Criterion

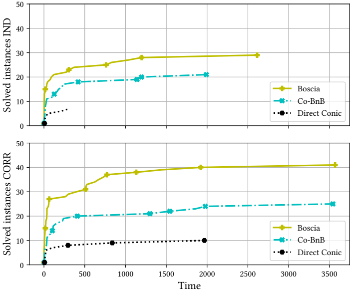


Figure: A-Optimal Problem

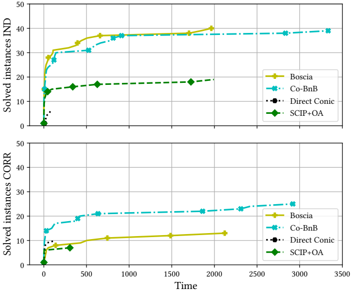


Figure: A-Fusion Problem

Plots II

Termination over time for the D-Criterion

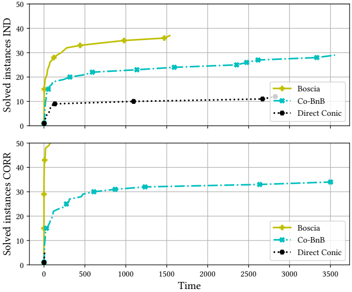


Figure: D-Optimal Problem

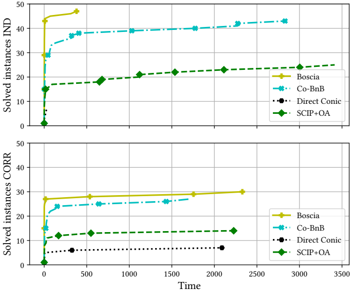


Figure: D-Fusion Problem

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

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

Experimental Results including SOCP

		Boscia		Co-BnB		Direct Conic		SOCP		SCIP+OA	
Type	Corr.	% solved	Time (s)	% solved	Time (s)	% solved	Time (s)	% solved	Time (s)	% solved	Time (s)
A	no	**58 %**	**208.53**	42 %	640.68	14 %	1901.7	20 %	1499.12		
A	yes	**82 %**	**98.5**	50 %	541.22	20 %	1591.74	20 %	1844.74		
AF	no	**80 %**	**54.82**	78 %	82.96	12 %	2006.81	26 %	1591.48	38 %	464.82
AF	yes	26 %	1359.35	**50 %**	**370.57**	20 %	1132.66	20 %	2002.94	14 %	1471.59
D	no	**74 %**	**81.07**	58 %	442.28	24 %	732.57	22 %	2192.52		
D	yes	**100 %**	**1.26**	68 %	223.34	10 %	755.88	8 %	2623.3		
DF	no	**94 %**	**3.32**	86 %	38.28	14 %	1576.5	12 %	2748.23	50 %	333.25
DF	yes	**60 %**	**50.68**	54 %	185.07	14 %	1761.18	6 %	2970.18	28 %	753.56

Table: Relative number of instances solved and the geometric mean of the time shifted by 1 second for all solvers. Note that the average time includes the timeouts!

Figures including SHOT I

Termination over time for the A-Criterion

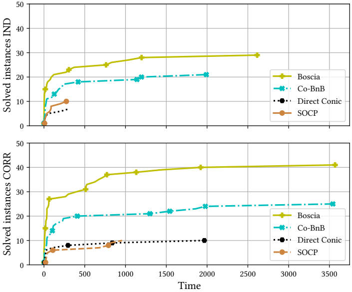


Figure: A-Optimal Problem

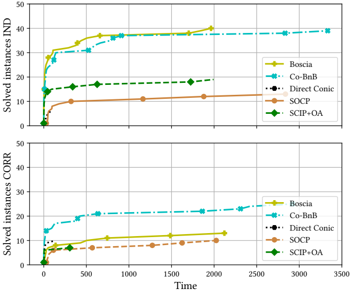


Figure: A-Fusion Problem

Figures including SHOT II

Termination over time for the D-Criterion

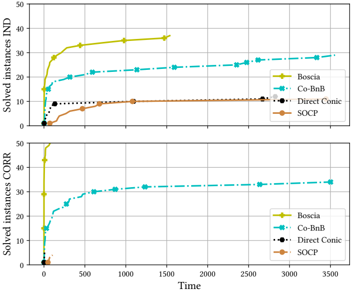


Figure: D-Optimal Problem

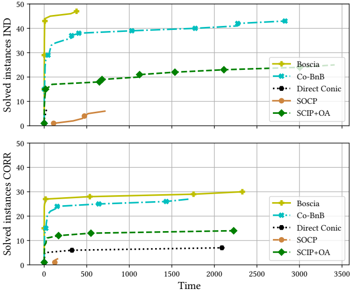


Figure: D-Fusion Problem