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

AA3-15

Deborah Hendrych
hendrych@zib.de

Deborah Hendrych, Mathieu Besançon, Sebastian Pokutta

Zuse Institute Berlin

Math+ Spotlight · May 8, 2024

Berlin Mathematics Research Center

Outline

1. Branch-and-Bound with Frank-Wolfe Methods

2. The Optimal Experiment Design Problem (OEDP)

3. Computational Experiments
    - Other MINLP approaches
    - Results

4. Summary and Outlook

Table of Contents

Branch-and-Bound with Frank-Wolfe Methods

The Optimal Experiment Design Problem (OEDP)

Computational Experiments

Other MINLP approaches

Results

Summary and Outlook

Setting

Mixed-Integer Convex Problems

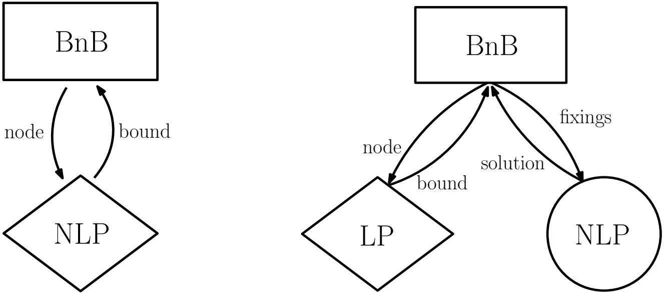
We want to solve MINLPs of the form

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

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

Previous Strategies

Kronqvist et al. 2019



The Frank-Wolfe algorithm I

Frank and Wolfe 1956; Levitin and Polyak 1966

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

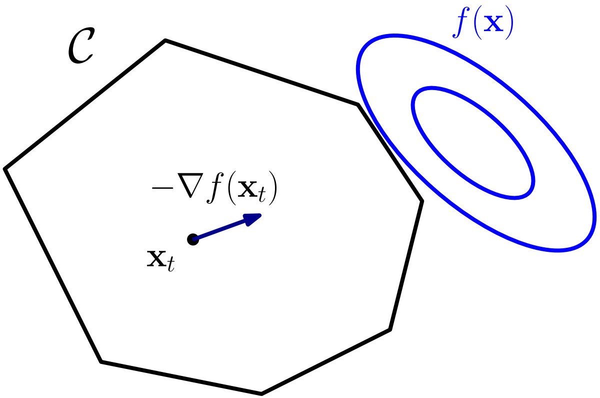
- $\mathcal{C}$ is a compact convex set.
- f continuously differentiable, convex function.
- Requires f to be L-smooth, i.e. gradient ∇f Lipschitz-continuous.
- Assumes that the Linear Minimization Oracle (LMO) $\min_{\mathbf{x}\in\mathcal{C}} \langle \mathbf{d}, \mathbf{x} \rangle$ is easy to solve.

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

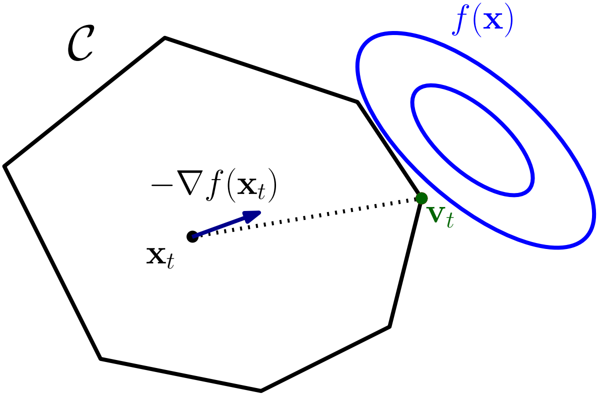
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm III

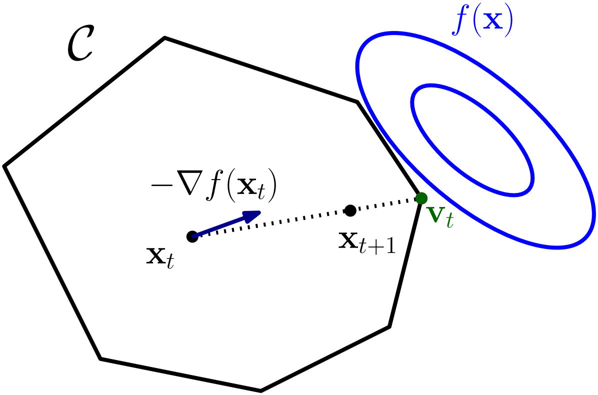
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm IV

Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

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

Hendrych et al. 2023

- 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.

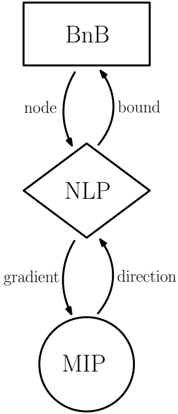


Table of Contents

Branch-and-Bound with Frank-Wolfe Methods

The Optimal Experiment Design Problem (OEDP)

Computational Experiments
- Other MINLP approaches
- Results

Summary and Outlook

Think Pancakes

Set-up

- The ultimate goal is fitting a regression model.

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

- $A = [-\mathbf{a}_i-]_{i=1}^m$ with $\mathbf{a}_i \in \mathbb{R}^n$ is the *Experiment Matrix*.
- Problem: Running all m experiments too costly and/or too time intensive.

Set-up

- The ultimate goal is fitting a regression model.

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

- $A = [-\mathbf{a}_i-]_{i=1}^m$ with $\mathbf{a}_i \in \mathbb{R}^n$ is the *Experiment Matrix*.
- Problem: Running all m experiments too costly and/or too time intensive.
- 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.

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)

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

- $\mathbf{l}$ and $\mathbf{u}$ are lower and upper bounds, respectively.
- Let $m \gg N \geq n$ be the *number of allowed experiments*.
- 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 ($p = 0$)

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

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

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

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

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

- They stem from the matrix means ϕ_p (Pukelsheim 2006; Ahipaşaoğlu 2021).
- For $p \leq 1$, ϕ_p is a valid information measure (Pukelsheim 2006).
- Both are $\mathcal{NP}$-hard. (Welch 1982; Nikolov, Singh, and Tantipongpipat 2022)

Table of Contents

Branch-and-Bound with Frank-Wolfe Methods

The Optimal Experiment Design Problem (OEDP)

Computational Experiments

- Other MINLP approaches
- Results

Summary and Outlook

Solution Approaches I

Solution via `Boscia.jl`

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

Solution Approaches II

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.

Solution Approaches III

Outer Approximation with SCIP (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.

Solution Approaches IV

Outer Approximation via Direct Conic Formulation

Coey, Kapelevich, and Vielma 2022a; Coey, Kapelevich, and Vielma 2022b; Coey, Lubin, and Vielma 2020

- Using Julia packages `Pajarito.jl` and `Hypatia.jl`.
- $\mathcal{K}_{\log\det} := \operatorname{cl}\left\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{S}^n_{++} \mid u \leq v \log\det(W/v)\right\}$
- Dual of $\mathcal{K}_{\text{sepspec}} := \operatorname{cl}\left\{(u, v, w) \in \mathbb{R} \times \mathbb{R}_{>0} \times \operatorname{int}(\mathcal{Q}) \mid u \geq v\varphi(w/v)\right\}$
- $\mathcal{Q}$ is the PSD cone and φ is the negative square root.
- The convex conjugate of the negative square root is the trace inverse.

Outer Approximation via Second Order Cones

Sagnol 2011; Sagnol and Harman 2015

- The models need far more variables and constraints.
- For D-optimal, $m + nm + nm + \frac{n(n+1)}{2}$ variables

Experimental Results I

Set up

- m between 50 and 120
- $n = \lfloor m/4 \rfloor$ and $n = \lfloor m/10 \rfloor$
- #allowed experiments $N = \lfloor 1.5n \rfloor$ 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 II

		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

Experimental Results III

Termination over time for the A-Criterion

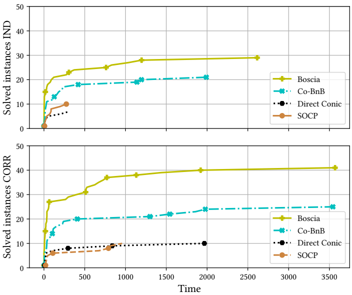


Figure: A-Optimal Problem

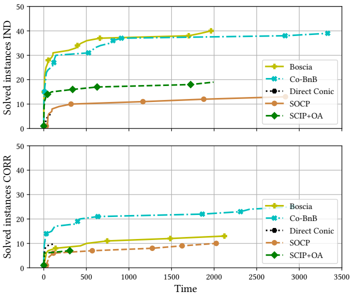


Figure: A-Fusion Problem

Experimental Results IV

Termination over time for the D-Criterion

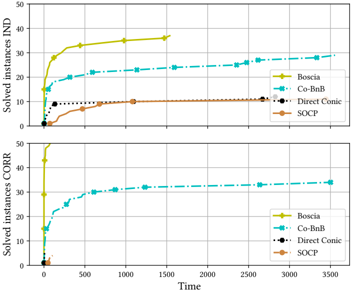


Figure: D-Optimal Problem

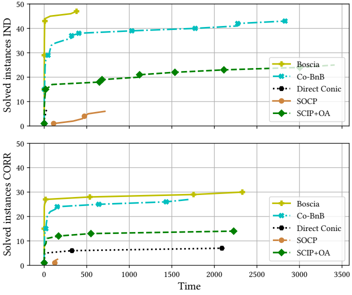


Figure: D-Fusion Problem

Table of Contents

Branch-and-Bound with Frank-Wolfe Methods

The Optimal Experiment Design Problem (OEDP)

Computational Experiments

- Other MINLP approaches
- Results

Summary and Outlook

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.
- Does not require a reformulation and is general, i.e. non-zero lower bounds are hand-able and additional constraints can be added.

Outlook

On the Optimal Design side

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

On the general `Boscia.jl` side

- Exploiting Sharpness.
- Improvement of the Branch-and-Bound strategies.
- Preprocessing.
- Heuristics.
- Handling of non-linear constraints?

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.