Boscia.jl and the Optimal Experiment Design Problem

Deborah Hendrych, Mathieu Besançon, Sebastian Pokutta

hendrych@zib.de

Zuse Institute Berlin

IOL Seminar · June 11, 2025

Berlin Mathematics Research Center

Introduction

- 2019: B.Sc. in Mathematics at BHT
- 2019/20: Study Abroad at UOW, Australia
- 2021: Student Assistant at IOL Lab
- 2023: M.Sc. in Mathematics at FU
- 2023: Started PhD at IOL Lab

Topics of interest

- Discrete geometry
- Discrete mathematics
- First-order optimization algorithms
- Mixed Integer Non-Linear Problems

Think Pancakes

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.

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*.

Information Matrix

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

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

- Variable $x_i \in \mathbb{N}_0$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.
- $A^\intercal A$ is positive definite (PD).
- $X(\mathbf{x})$ is positive semi definite (PSD).

Information Matrix

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

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

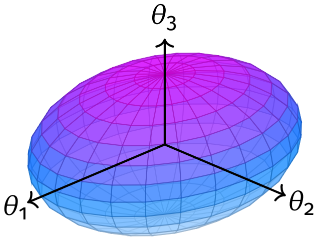
- Variable $x_i \in \mathbb{N}_0$ denotes how often experiment i is to be run. $\mathbf{x}$ is a *design*.
- $A^\intercal A$ is positive definite (PD).
- $X(\mathbf{x})$ is positive semi definite (PSD).
- The inverse of $X(\mathbf{x})$ (if existent) is called the *Dispersion Matrix* $D(\mathbf{x})$ and is a measure of the variance of the parameter vector θ Ahipaşaoğlu 2021.
- A design $\mathbf{x}$ is ”useful” if $X(\mathbf{x})$ is regular.

Some information theory later ...

Pukelsheim 2006

D-Optimal Experiment Design

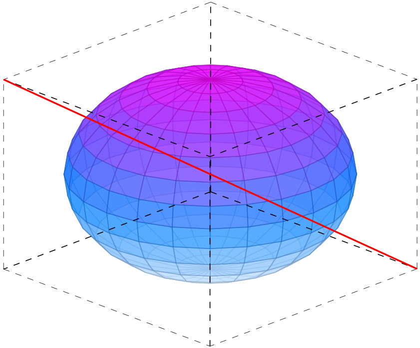
$$
\begin{aligned}
\min_{\mathbf{x}} \quad & -\log\det\left(X(\mathbf{x})\right) \\
\text{s.t.} \quad & \sum_{i=1}^{m} x_i = N \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{x} \in \mathbb{N}_0^m,
\end{aligned}
\qquad \text{(DO)}
$$



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

A-Optimal Experiment Design

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



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

Table of Contents

Boscia - Problem 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 (L-smooth) function.
- $\mathcal{X}$ is polyhedral potentially with combinatorial constraints.

$\longrightarrow$ Branch-and-Bound with Frank-Wolfe

The Frank-Wolfe algorithm I

Frank and Wolfe 1956; Levitin and Polyak 1966

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

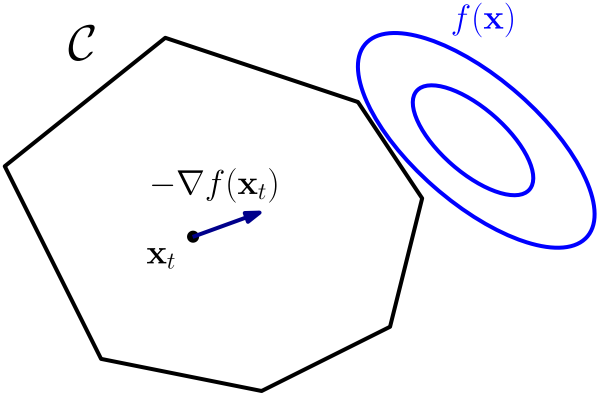
- $\mathcal{C}$ is a compact convex set.
- f continuously differentiable, convex function.
- Requires f to be L-smooth, i.e. gradient ∇f Lipschitz-continuous.
- Assumes that the **Linear Minimization Oracle (LMO)**

$$\min_{\mathbf{v} \in \mathcal{C}} \langle \mathbf{d}, \mathbf{v} \rangle$$

is comparatively easy to solve.

The Frank-Wolfe algorithm II

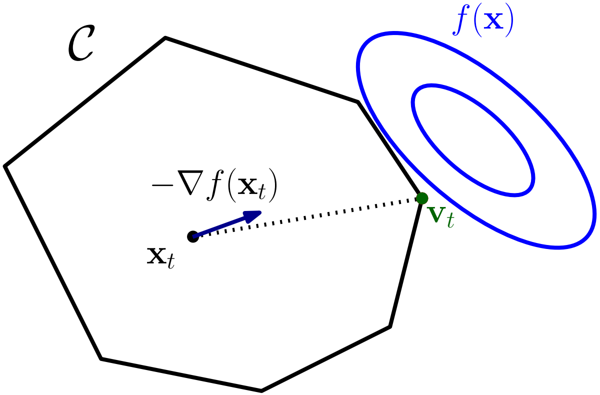
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm III

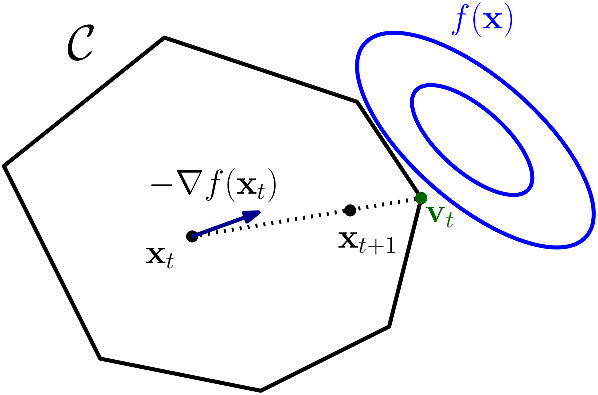
Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm IV

Frank and Wolfe 1956; Levitin and Polyak 1966



$\mathbf{x}_t$

The Frank-Wolfe algorithm V

Frank and Wolfe 1956; Levitin and Polyak 1966

The Frank-Wolfe Gap

By convexity, we have

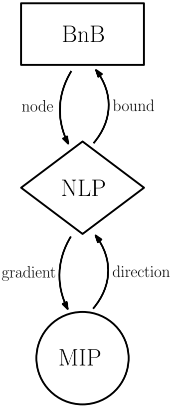
$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \leq \min_{\mathbf{v} \in \mathcal{C}} \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{v} \rangle .$$

Also, this provides us with a lower bound on the optimum.

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

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

Hendrych et al. 2023



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

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{s.t. } \mathbf{x} \in \operatorname{conv}\{\mathcal{X} \cap \mathbb{Z}_j\}$$

$$\min_{\mathbf{v}} \langle \nabla f(x_t), \mathbf{v} \rangle$$
$$\text{s.t. } \mathbf{v} \in \mathcal{X} \cap \mathbb{Z}_j$$

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

Hendrych et al. 2023

- Frank-Wolfe implemented in `FrankWolfe.jl` Besançon, Carderera, and Pokutta 2022.
- Linear Minimization Oracle is a *Bounded (Mixed-Integer) Linear Minimization Oracle (BLMO)*, usually this is a MIP solver but it can also be a combinatorial solver.
- Integer feasible solutions from the root node → an inbuild heuristic.
- Exploits Frank-Wolfe's precision-adaptiveness → the solution tolerance is tightening with the node depth.
- Frank-Wolfe enables warm-starting via the active set being split while branching.

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

Hendrych et al. 2023

- Frank-Wolfe's lazification techniques to avoid calling the BLMO too often.
- Collecting dropped vertices into the *shadow set* as a further lazification.
- Dynamically stop node evaluation if
  - the node's lower bound is larger than the incumbent.
  - there are k open nodes with a better lower bounds.

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.

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.

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

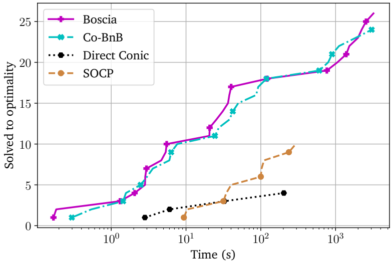


Figure: Independent data

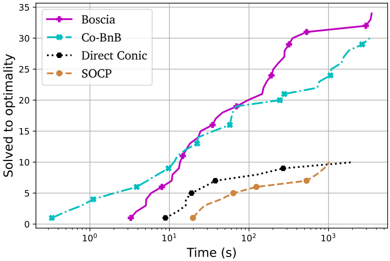


Figure: Correlated data

Experimental Results III

Termination over time for the D-Optimal Problem

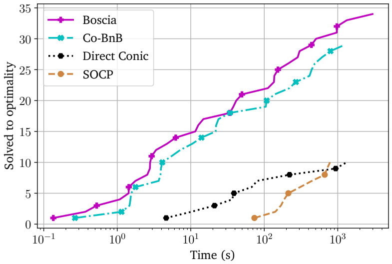


Figure: Independent data

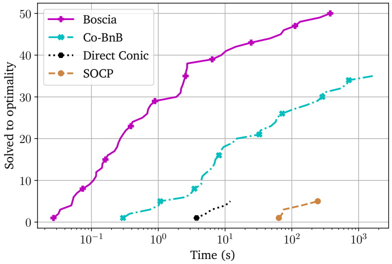


Figure: Correlated data

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?
- Preprocessing.
- Gradient feasible region.

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.*

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.

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

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

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.

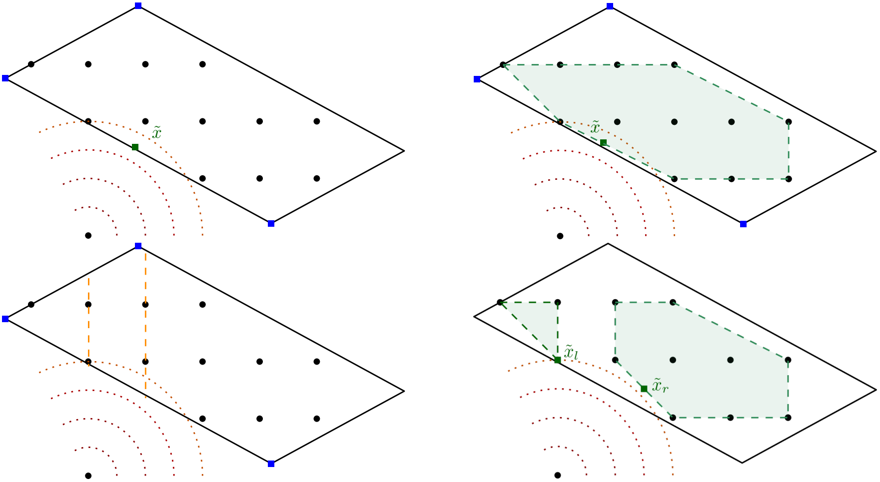
Pukelsheim, Friedrich (2006). *Optimal design of experiments.* SIAM.

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.

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.

Convex relaxation vs integer hull



Using `Boscia.jl` to solve OEDP I

Theorem (L-smoothness for the Fusion Problems)

The functions

$$f(X_C(\mathbf{x})) = -\log\det\left(X_C(\mathbf{x})\right)$$
$$g(X_C(\mathbf{x})) = \operatorname{tr}\left(X_C(\mathbf{x})^{-p}\right) \qquad p > 0$$
$$k(X_C(\mathbf{x})) = \log\operatorname{tr}\left(X_C(\mathbf{x})^{-p}\right) \quad p > 0$$

are L-smooth on $x \in \mathbb{R}_{\geq 0}$ where $X_C(\mathbf{x})$ denotes the information matrix for the Fusion Problems.

Using `Boscia.jl` to solve OEDP II

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

$$\mathcal{L}_0 = \left\{ \mathbf{x} \in D \cap \sum_{i=1}^{m} x_i = N \,\middle|\, (*)(\mathbf{x}) \leq (*)(\mathbf{x}_0) \right\}$$

where $(*)$ is a placeholder for each function.

Using `Boscia.jl` to solve OEDP III

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)

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.

Using `Boscia.jl` to solve OEDP IV

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 $D := \{X \in \mathbb{S}^n_{++} \mid \lambda_{\max}(X) \leq \alpha, \kappa(X) \leq \kappa\}$.

Theorem

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

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

Using `Boscia.jl` to solve OEDP V

Objective	Convergence	Linear Convergence
$f_F(\mathbf{x}) = \log\det(X_C(\mathbf{x}))$	✓	✓
$g_F(\mathbf{x}) = \mathrm{tr}(X_C(\mathbf{x})^{-p})$	✓	✓
$k_F(\mathbf{x}) = \log\left(\mathrm{tr}(X_C(\mathbf{x})^{-p})\right)$	✓	✓
$f(\mathbf{x}) = -\log\det(X(\mathbf{x}))$	✓	✓
$g(\mathbf{x}) = \mathrm{tr}(X(\mathbf{x})^{-p})$	✓	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.

Feasible region vs Domain

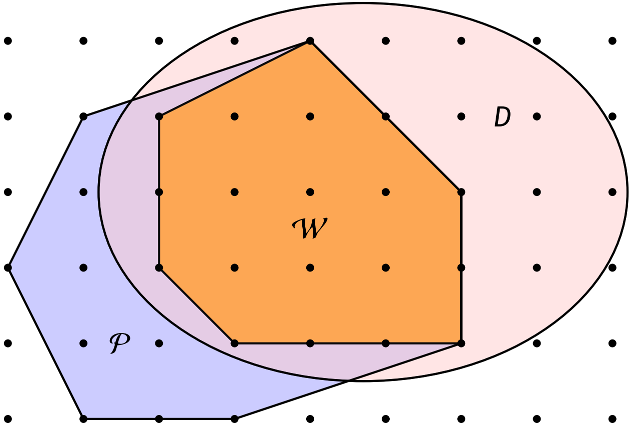


Figure: A schematic representation of the feasible region $\mathcal{P}$, the domain of the objective $\mathcal{D}$ and the convex hull of vertices that are both feasible and in the domain denoted as $\mathcal{W}$.

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

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