New Perspectives on Mixed-Integer Convex Optimization with Applications in Statistical Learning

Simge Küçükyavuz

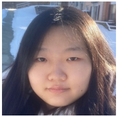
NORTHWESTERN UNIVERSITY

Joint work with...

Alper Atamtürk
Berkeley

Salar Fattahi
Michigan

Andrés Gómez
USC



Peijing Liu
USC

Linchuan Wei
Northwestern

Supported by NSF grants 2007814 and 2006762 (with A. Gómez)

Time travel

Before we talk about Mixed-Integer **Convex Quadratic** Programs, let's do an experiment to see how far we've come in Mixed-Integer **Linear** Programming (MILP)

Impact of cutting planes in Mixed-Integer **Linear** Programming (MILP) software

Without cuts

```
Explored 1933736 nodes (3989094 simplex iterations) in 38.70 seconds
Thread count was 8 (of 8 available processors)
```

With cuts

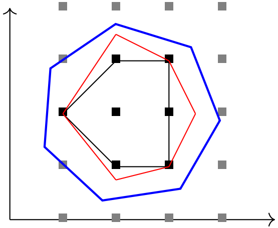
```
Cutting planes:
  Gomory: 4
  Implied bound: 22
  MIR: 19
  Flow cover: 35
  Flow path: 18
  Relax-and-lift: 2

Explored 1 nodes (426 simplex iterations) in 0.10 seconds
Thread count was 8 (of 8 available processors)
```

Extended formulation

```
Explored 0 nodes (238 simplex iterations) in 0.03 seconds
Thread count was 8 (of 8 available processors)
```

What's the Secret Sauce?

Polyhedral Theory for MILP



Original formulation
Stronger formulation
Ideal formulation (convex hull, facets)

- Structured cutting planes (Cover, flow cover, flow path, etc.)
- General-purpose cutting planes (Gomory, MIR, disjunctive, etc.)
- Presolve, heuristics, branching, ...

See, also, "Progress in Mathematical Programming Solvers from 2001 to 2020," Koch et al, 2021.

Agenda

MIQP with indicators

$$\min_{x,z} a^\top x + b^\top z + \frac{1}{2}x^\top Qx \qquad (Q \succeq 0)$$

$$\text{s.t. } x_j(1 - z_j) = 0, \quad j \in [n] := \{1, \dots, n\} \qquad (x \circ (\mathbf{1} - z) = \mathbf{0})$$

$$x \in \mathbb{R}^n,\ z \in Z \subseteq \{0,1\}^n$$

or equivalently, in its epigraph form,

$$\min_{x,z,t} a^\top x + b^\top z + \frac{1}{2}t$$

$$\text{s.t. } t \geq x^\top Qx, x \circ (\mathbf{1} - z) = \mathbf{0}, x \in \mathbb{R}^n,\ z \in Z$$

Alternative formulation of non-convex complementarity constraint

$$-Mz \leq x \leq Mz \qquad \text{(Big-M constraint)}$$

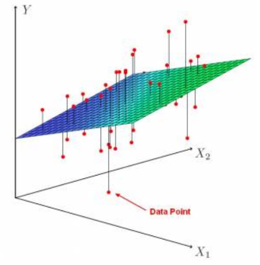
Weak continuous relaxation

Motivating Example: Best Subset Selection

Given model matrix $A_{m \times n}$ and response vector $y \in \mathbb{R}^m$

$$\min_{x : \|x\|_0 \leq k} \|y - Ax\|_2^2,$$

where $\|x\|_0 = \sum_{i=1}^n \mathbb{1}_{\{x_i \neq 0\}}$ is the "ℓ_0 norm," $k \in \mathbb{Z}$ is a given cardinality.



Here, $Z = \{z \in \{0,1\}^n : \sum_{i=1}^n z_i \leq k\}$, $Q = A^\top A$, $a = -y^\top A$

NP-hard. (Chen et al, 2017)

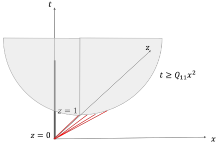
Other Applications

- Structured regression (e.g., Bertsimas et al, 2021; Hazimeh and Mazumder, 2020)
- Probabilistic graphical models (e.g., Küçükyavuz et al., 2020)
- Portfolio optimization (e.g., Bienstock, 1996)
- Power systems (e.g., Bacci et al., 2019)
- Machine scheduling (e.g., Aktürk et al., 2009)

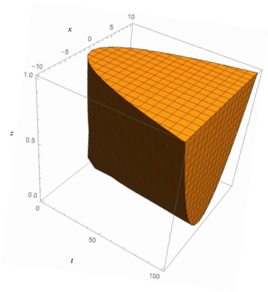
Agenda

Special Cases: $n = 1$ (or Q is diagonal)

$$R \equiv \{(z,x,t) \mid t \geq Q_{11}x^2,\ x \circ (\mathbf{1} - z) = \mathbf{0},\ z \in \{0,1\}\}$$



$$\mathrm{cl\ conv}(R) \equiv \{(z,x,t) \mid t \geq Q_{11}\frac{x^2}{z},\ z \in [0,1]\}$$ (Big-M free, SOCP)



Convention: $\frac{0}{0} = 0$.

Perspective reformulation: Ceria and Soares (1999), Frangioni and Gentile (2006), Aktürk et al. (2008), Günlük and Linderoth (2010)...

Why perspective? For convex function $f : \mathbb{R} \to \mathbb{R}$ with $f(0) = 0$ its perspective function $\phi(x,z) = zf\left(\frac{x}{z}\right) : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is also convex.

Solution approaches leveraging perspective formulation

1. Find "good" diagonal matrix $D, D_{ii} \geq 0$ such that $Q - D \succeq 0$
   - Using minimum eigenvalue of Q (Frangioni, 2006)
   - Using SDP heuristics (Frangioni, 2007)
   - Using ridge regularization (Bertsimas and Van Parys, 2020)
   - Maximizing relaxation quality (Zheng et al., 2014; Dong et al., 2015)

2. Use branch-and-bound based on the perspective reformulation

$$\begin{aligned}
\min_{x,z}\ & a^\top x + b^\top z + \frac{1}{2}x^\top (Q - D)x + \frac{1}{2}\sum_{i=1}^{n} \frac{D_{ii}x_i^2}{z_i} \\
\text{s.t.}\ & -Mz \leq x \leq Mz \qquad \text{(Big-M constraint)} \\
& x \in \mathbb{R}^n,\ z \in Z \subseteq \{0,1\}^n
\end{aligned}$$

Research Questions

- Can we exploit matrix and constraint structure to obtain stronger relaxations? (Part 1)

- What does strong mean for MIQP?
  Can we leverage polyhedral theory for MIQP? (Part 2)

Special Cases: Rank-one convex function f

$X = \left\{(z,x,t) \in \{0,1\}^n \times \mathbb{R}^{n+1} \mid t \geq f(q^\top x), x \circ (\mathbf{1} - z) = \mathbf{0}, z \in Z\right\}$
Quadratic: $f(q^\top x) = (q^\top x)^2$ for a given vector $q \in \mathbb{R}^n$, i.e., $Q = qq^\top \succeq 0$.

Theorem (Wei, Gómez, Küçükyavuz, 2022)

If f is convex, $f(0) = 0$, and Z is 'connected', then

$$\text{cl conv}(X) = \left\{(z,x,t) | z \in \text{conv}(Z), t \geq f(q^\top x), t \geq (\pi^\top z) f\left(\frac{q^\top x}{\pi^\top z}\right), \forall \pi \in \mathcal{F}\right\},$$

where $\mathcal{F}$ is a family of strong separating inequalities for

$$\text{conv}(Z \setminus \{\mathbf{0}\}) \;=\; \text{conv}(Z) \bigcap \{z \in \mathbb{R}^n \;:\; \pi^\top z \geq 1,\; \forall \pi \in \mathcal{F}\}$$

- New perspectives
  Subsumes all related convexifications to date; first convexification for a logistic loss function.

How to characterize $\mathcal{F}$? Can use ideas in Angulo et al, "Forbidden Vertices," 2015.

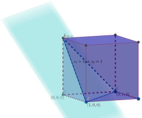
Special Case: $f(q^\top x) = (q^\top x)^2, Z = \{0,1\}^n$

$$R \equiv \{(z, x, t) \in \{0,1\}^n \times \mathbb{R}^{n+1} \ : \ t \geq (\mathbf{q}^\top x)^2, \ x \circ (\mathbf{1} - z) = \mathbf{0}\}.$$

Theorem (Atamtürk and Gómez, 2019)

$$\text{cl conv}(R) = \left\{(z, x, t) \in [0,1]^n \times \mathbb{R}^{n+1} \mid t \geq (\mathbf{q}^\top x)^2, \ t \geq \frac{(\mathbf{q}^\top x)^2}{\sum_{i\in[n]} z_i}\right\}$$

- $\sum_{i=1}^n z_i \geq 1$ is a strong inequality separating $\mathbf{0}$ from the set Z



$$\text{conv}(Z\backslash\{\mathbf{0}\}) = \{z \in [0,1]^n \ : \ \sum_{i=1}^n z_i \geq 1\}$$

Special Case: Diagonal Q, General $Z \subset \{0,1\}^n$

$$R \equiv \{(z,x,t) \mid t_i \geq Q_{ii}x_i^2, i \in [n],\ x \circ (\mathbf{1}-z) = \mathbf{0},\ z \in Z\}$$

Corollary (Wei, Gómez, Küçükyavuz, 2022)

$$\text{cl conv}(R) = \left\{(z,x,t) \middle| t_i \geq \frac{Q_{ii}x_i^2}{z_i}, i \in [n], z \in \text{conv}(Z)\right\}.$$

Xie and Deng (2020) show this for $Z = \{z \in \{0,1\}^n : \sum_{i=1}^n z_i \leq k\}$.

Numerical Results

Least squares regression with **strong hierarchy** constraints on pairwise interactions.

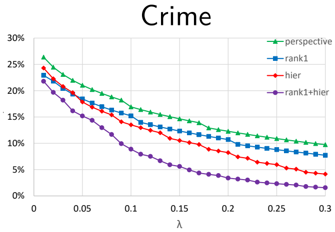
$$
\begin{aligned}
\min_{z,x} \ & \sum_{\ell=1}^{p}\left(y_\ell - \sum_{i=1}^{n} A_{\ell i}x_i - \sum_{i=1}^{n}\sum_{j=i}^{n} A_{\ell i}A_{\ell j}x_{ij}\right)^2 + \lambda\|x\|_2^2 + \mu\|z\|_1 \\
\text{s.t. } & x_i(1-z_i) = 0 \qquad \forall i \\
& x_{ij}(1-z_{ij}) = 0 \qquad i \le j \\
& z_{ii} \le z_i \qquad \forall i \\
& z_{ij} \le z_i,\ z_{ij} \le z_j \qquad i \le j \\
& z \in \{0,1\}^{\frac{n(n+3)}{2}}
\end{aligned}
$$

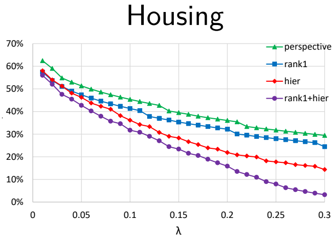
- $z_i + z_j - z_{ij} \ge 1$ is a strong inequality separating $\mathbf{0}$ from the set Z

Relaxation comparisons

- **Perspective**: Optimal perspective relaxation (Dong et al., 2015)
- **Rank1**: Rank-one relaxation (Atamtürk and Gómez, 2019)
- **Hier**: Hierarchical strengthening (the formulation we proposed)
- **Rank1 + Hier**: Combine these two methods

Optimality Gaps: Varying λ

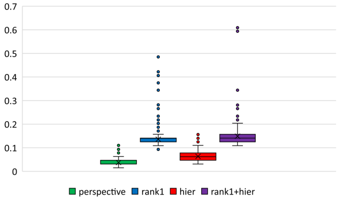




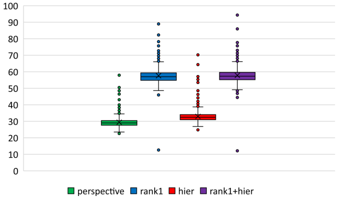
- **Hier** (vs. **Persp**): Significant improvement in lower bound
- **Rank1+Hier** (vs. **Rank1**): Gives the best optimality gap

Solution Time

Crime



Housing

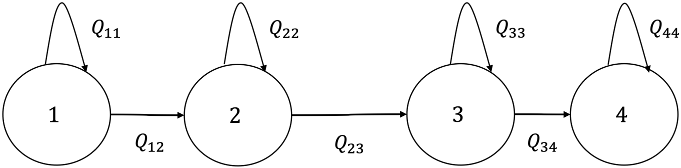


- **Hier** (vs. **Persp**): Only a slight increase in solution time
- **Rank1+Hier** (vs. **Rank1**): No increase in solution time

Special Case: Tridiagonal Q

$$Q = \begin{pmatrix} * & * & 0 & 0 & \dots & 0 \\ * & * & * & 0 & \dots & 0 \\ 0 & * & * & * & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & * & * \end{pmatrix}$$

Support graph of Q: Arc (i,j) for $i \le j$ with $Q_{ij} \neq 0$



Special Case: Tridiagonal Q

$$\begin{aligned} \min_{x,z}\ & a^\top x + b^\top z + \frac{1}{2} x^\top Q x \qquad\qquad (Q \succ 0,\ \text{tridiagonal}) \\ \text{s.t.}\ & x \circ (\mathbf{1} - z) = \mathbf{0} \\ & x \in \mathbb{R}^n,\ z \in \{0,1\}^n \end{aligned}$$

Suppose $z = \mathbf{1}$

- Optimality condition: Solve $Qx = -a$

- Thomas Algorithm for tridiagonal Q takes $O(n)$ time

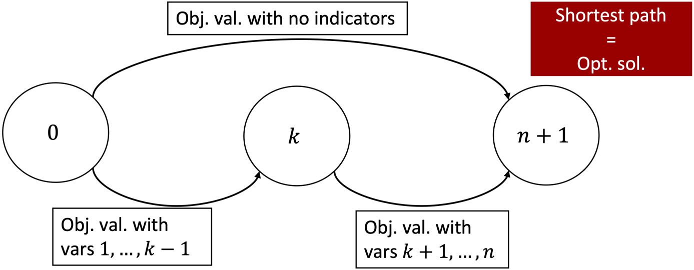
Single indicator: $z_k \in \{0, 1\}$

Now suppose $z_j = 1$ for $j \in [n] \setminus \{k\}$:

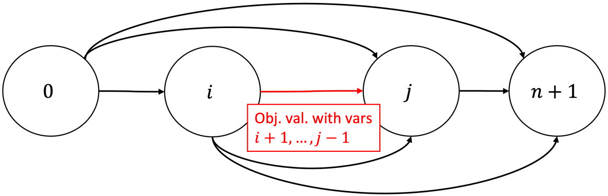
$$\min_{x, z_k} a^\top x + b_k z_k + \frac{1}{2} \sum_{i \in [n]} Q_{ii} x_i^2 + \sum_{i \in [n-1]} Q_{i,i+1} x_i x_{i+1}$$

$$\text{s.t. } x_k(1 - z_k) = 0$$

$$x \in \mathbb{R}^n, \ z_k \in \{0, 1\}$$



All indicators: $z \in \{0,1\}^n$



Proposition (Liu, Fattahi, Gómez, Küçükyavuz, 2022)

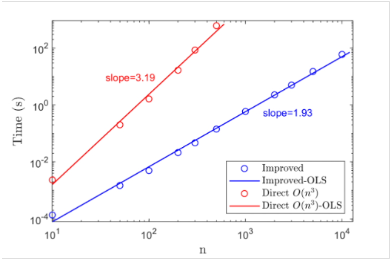
MIQP with tridiagonal matrices can be solved by solving a shortest path problem.

Complexity:

- Direct: $O(n^2)$ arcs $\times O(n)$ arc cost calculation $= O(n^3)$
- Improved: $O(n^2)$

Leads to a shortest path-based compact tight extended formulation.

Experiments



n	50	100	200	500	1000	10000
Improved	**<0.01s**	**<0.01s**	**<0.01s**	**0.14s**	**0.59s**	**59.6s**
Direct	0.16s	1.63s	18.9s	608.9s	>3600s	>3600s
Big-M	0.20s	214.1s	>3600s	>3600s	>3600s	>3600s

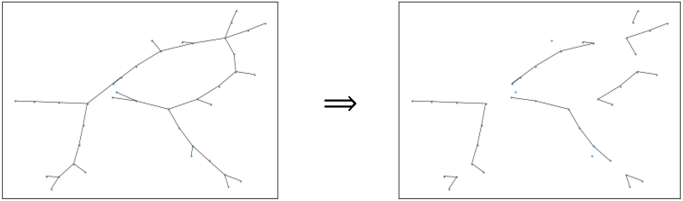
Can we leverage this efficient algorithm to solve the problem for non-tridiagonal $Q \succ 0$?

Sparse, Strictly Diagonally Dominant Matrix Q

Idea: Split Q into tridiagonal submatrices: $T_1, \ldots, T_\ell$ and a remainder R of off-tridiagonals

- Use convexification and Fenchel duality for off-tridiagonals

- Decomposes to path subproblems ($O(n^2)$ algorithm)



Convexification and Fenchel duality

Rewriting the problem

$$\min_{x,z,t} a^\top z + b^\top z + \frac{1}{2}\sum_{k=1}^{\ell} t_k + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=i+2}^{n} |Q_{i,j}|(x_i \pm x_j)^2$$

$$\text{s.t. } t_k \geq x^\top T_k x, k = 1, \dots, \ell, x \circ (\mathbf{1} - z) = \mathbf{0}, x \in \mathbb{R}^n, z \in \{0,1\}^n, t \in \mathbb{R}^\ell.$$

Convexify the rank-one terms to obtain relaxation objective

$$\min_{x,z,t} a^\top x + b^\top z + \frac{1}{2}\sum_{k=1}^{\ell} t_k + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=i+2}^{n} |Q_{i,j}|\frac{(x_i \pm x_j)^2}{\min\{1, z_i + z_j\}}$$

Relax complicating terms via Fenchel dual to obtain relaxation

$$\zeta_p = \min_{x,z,t}\ \max_{\alpha,\beta}\ a^\top x + b^\top z + \frac{1}{2}\sum_{k=1}^{\ell} t_k$$
$$+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=i+2}^{n} |Q_{ij}|\Big(\alpha_{ij}(x_i \pm x_j) - \beta_{ij,i} z_i - \beta_{ij,j} z_j - f^*(\alpha_{ij}, \beta_{ij,i}, \beta_{ij,j})\Big)$$

Fenchel Decomposition

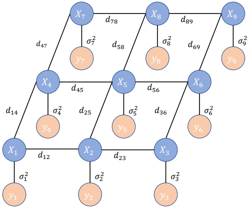
Strong duality holds, so

$$\zeta_p = \max_{\alpha,\beta} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=i+2}^{n} |Q_{ij}| f^*(\alpha_{ij}, \beta_{ij,i}, \beta_{ij,j}) + \min_{x,z,t} \left\{\psi(x, z, t, \alpha, \beta)\right\}$$

- Inner min Independent tridiagonal problems
- Outer max Subgradient ascent

Computational Study

Inference with graphical models



- Given: noisy observations (orange)
- Goal: find true values (blue)
- Arc (i, j): connection between variables i, j with $Q_{ij} \neq 0$

Computational Results

$n = 100$

Noise	Low		High	
Method	Time(s)	Gap	Time(s)	Gap
Decomposition	**0.1**	**<1%**	**0.5**	**<1%**
Big-M	0.3	0.0%	3600	4.7%

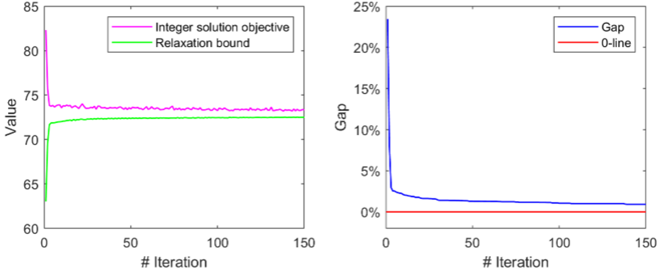
$n = 1600$

Noise	Low		High	
Method	Time(s)	Gap	Time(s)	Gap
Decomposition	**43.6**	**<1%**	**190.5**	**1.0%**
Big-M	3600	3.9%	3600	30.9%

Gap=(Upper Bound- Lower Bound)/Upper Bound

Decomposition Method

$n = 1600$, high noise



Research Questions

- Can we exploit matrix and constraint structure to obtain stronger relaxations? (Part 1) ✓

- What does strong mean for MIQP?
  Can we leverage polyhedral theory for MIQP? (Part 2)

Agenda

General $Q \succ 0$

A Combinatorial View

For a subset $S \in Z$, $z_i = 1$ if $i \in S$ and Q_S is the submatrix of Q indexed by S (similarly a_S, b_S, x_S)

$$\begin{aligned} \min_x \ & a^\top x + b_S + \frac{1}{2} x^\top Q x \quad = \quad \min_{x_S} a_S^\top x_S + b_S + \frac{1}{2} x_S^\top Q_S x_S \\ \text{s.t. } & x_i = 0, \ \forall i \notin S. \end{aligned}$$

- $x_S^* = -Q_S^{-1} a_S$.
- A combinatorial problem of selecting subset S

$$\min_{S \subseteq [n]} \ b_S - \frac{1}{2} a_S^\top Q_S^{-1} a_S$$

Notation

- Given $S \subseteq \{1, \dots, n\}$:
  - e_S = n-dimensional indicator vector of S
  - Q_S = $|S| \times |S|$ submatrix of Q induced by S
  - $\hat{Q}_S^{-1}$ = $n \times n$ matrix corresponding to Q_S^{-1} in the rows/columns of S, and 0 elsewhere

Example: $Q = \begin{pmatrix} d_1 & 1 \\ 1 & d_2 \end{pmatrix}$

$d_1 d_2 > 1$

S	e_S	Q_S	$\hat{Q}_S^{-1}$
$\emptyset$	$(0\ 0)$	$\emptyset$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
$\{1\}$	$(1\ 0)$	(d_1)	$\begin{pmatrix} 1/d_1 & 0 \\ 0 & 0 \end{pmatrix}$
$\{2\}$	$(0\ 1)$	(d_2)	$\begin{pmatrix} 0 & 0 \\ 0 & 1/d_2 \end{pmatrix}$
$\{1,2\}$	$(1\ 1)$	$\begin{pmatrix} d_1 & 1 \\ 1 & d_2 \end{pmatrix}$	$\frac{1}{d_1 d_2 - 1}\begin{pmatrix} d_2 & -1 \\ -1 & d_1 \end{pmatrix}$

Structure of the convex hull (extended formulation)

$$X \equiv \left\{(z,x,t) \in Z \times \mathbb{R}^{n+1} \mid t \geq x^\top Q x,\ x \circ (\mathbf{1} - z) = \mathbf{0}\right\}.$$

$$P \equiv \operatorname{conv}\left(\left\{(e_S, \hat{Q}_S^{-1})_{S \in Z}\right\}\right).$$

Theorem (Wei, Atamtürk, Gómez and Küçükyavuz, 2022)

If Q is positive definite, then

$$\operatorname{cl}\operatorname{conv}(X) = \left\{(z,x,t) \in [0,1]^n \times \mathbb{R}^{n+1} \mid \exists W \in \mathbb{R}^{n\times n},\ \begin{pmatrix} W & x \\ x^\top & t \end{pmatrix} \succeq \mathbf{0}, (z, W) \in P\right\}.$$

Can be extended to the psd/low rank case (a more compact extended formulation)

Preliminaries

Definition

Given a matrix $W \in \mathbb{R}^{p \times q}$, its pseudoinverse $W^\dagger \in \mathbb{R}^{q \times p}$ is the unique matrix satisfying: $WW^\dagger W = W$, $W^\dagger W W^\dagger = W^\dagger$, $(WW^\dagger)^\top = WW^\dagger$, $(W^\dagger W)^\top = W^\dagger W$.

Examples

- if W is invertible then $W^\dagger = W^{-1}$
- $W = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$, $W^\dagger = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & 0 \end{pmatrix}$

Lemma (Generalized Schur Complement)

$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{12}^\top & U_{22} \end{pmatrix}$ with $U_{11} \in \mathcal{S}^{m \times m}$ and $U_{22} \in \mathcal{S}^{n \times n}$, and $U_{12} \in \mathbb{R}^{m \times n}$. Then $U \succeq \mathbf{0}$ if and only if $U_{11} \succeq \mathbf{0}$, $U_{11} U_{11}^\dagger U_{12} = U_{12}$ and $U_{22} - U_{12}^\top U_{11}^\dagger U_{12} \succeq \mathbf{0}$.

Recall

Theorem (Wei, Atamtürk, Gómez and Küçükyavuz, 2022)

If Q is positive definite, then

$$\text{cl conv}(X) = \left\{(z,x,t) \in [0,1]^n \times \mathbb{R}^{n+1} \mid \exists W \in \mathbb{R}^{n\times n} \ \begin{pmatrix} W & x \\ x^\top & t \end{pmatrix} \succeq \mathbf{0},\ (z,W) \in P\right\}.$$

Proof idea: Optimizing over cl conv(X) is equivalent to optimizing the original problem.

Proof of the Theorem

Optimizing over cl conv(X)

$$\begin{aligned} \min_{x,z,W} \quad & a^\top x + b^\top z + t \\ \text{s.t.} \quad & \begin{pmatrix} W & x \\ x^\top & t \end{pmatrix} \succeq \mathbf{0} \\ & (z, W) \in P \equiv \text{conv}\left(\left\{(e_S, \hat{Q}_S^{-1})_{S \in Z}\right\}\right) \end{aligned}$$

- $z = e_S$ for some $S \in Z$
- $W = \hat{Q}_S^{-1} = \begin{pmatrix} Q_S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \succeq \mathbf{0}$ and $W^\dagger = \begin{pmatrix} Q_S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$
- $\color{blue}{WW^\dagger x = x} \Leftrightarrow \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_S \\ x_{[n]\setminus S} \end{pmatrix} = \begin{pmatrix} x_S \\ x_{[n]\setminus S} \end{pmatrix} \Leftrightarrow x_{[n]\setminus S} = \mathbf{0}$
- $\color{red}{t \geq x^\top W^\dagger x} \Leftrightarrow t \geq x_S^\top Q_S x_S$

Example: Quadratic with "Choose-one" constraint

$$X_{C_1} = \{(z,x,t) \in \{0,1\}^n \times \mathbb{R}^{n+1} \mid t \geq x^\top Q x,\ x \circ (\mathbf{1} - z) = \mathbf{0},\ \textstyle\sum_{i=1}^n z_i \leq 1\}$$

Corollary

$$\operatorname{cl}\operatorname{conv}(X_{C_1}) = \left\{(z,x,t) \in \mathbb{R}_+^n \times \mathbb{R}^n \times \mathbb{R} \mid t \geq \textstyle\sum_{i=1}^n Q_{ii} \frac{x_i^2}{z_i},\ \sum_{i=1}^n z_i \leq 1\right\}.$$

- $P = \operatorname{conv}\left(\left\{(0,\mathbf{0}), (e_{\{i\}}, \hat{Q}_{\{i\}}^{-1})_{i=1}^n\right\}\right)$
- $P = \{(z,W) \mid W_{ij} = 0, i \neq j, W_{ii} = \frac{z_i}{Q_{ii}}, i = 1, \ldots, n\}$

$$\begin{pmatrix} W & x \\ x^\top & t \end{pmatrix} \succeq \mathbf{0}, (z,W) \in P \Leftrightarrow \begin{pmatrix} \frac{z_1}{Q_{11}} & \ldots & 0 & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \frac{z_n}{Q_{nn}} & x_n \\ x_1 & \ldots & x_n & t \end{pmatrix} \succeq \mathbf{0} \Leftrightarrow t \geq \sum_{i=1}^n Q_{ii} \frac{x_i^2}{z_i}$$

Structure of the convex hull (original space)

Let

$$X = \left\{(x,z,t) \in \mathbb{R}^n \times Z \times \mathbb{R} : t \geq x^\top Q x,\ x \circ (\mathbf{1} - z) = \mathbf{0}\right\}$$

Suppose a minimal description of P is given by

$$\begin{aligned} \langle \Gamma_i, W \rangle - \gamma_i^\top z &\leq \beta_i, \quad i = 1, \ldots, m_1 \\ \langle \Gamma_i, W \rangle - \gamma_i^\top z &= \beta_i, \quad i = m_1 + 1, \ldots, m. \end{aligned}$$

Theorem (Wei, Atamtürk, Gómez, Küçükyavuz, 2022)

$(x,z,t) \in cl\ conv(X)$ iff $z \in \mathrm{conv}(Z)$, $t \geq 0$ and

$$t \geq \frac{x^\top \left(\sum_{i=1}^m \Gamma_i s_i\right) x}{\beta^\top s + \left(\sum_{i=1}^m \gamma_i s_i\right)^\top z}$$

for all $s \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m-m_1}$ such that $\sum_{i=1}^m \Gamma_i s_i \succeq 0, \sum_{i=1}^m Tr(\Gamma_i) s_i \leq 1$.

Observations

- Semi-infinite conic quadratic program, but "finitely" generated by $(\Gamma_i, \gamma_i, \beta_i), i \in [m]$.
- The strongest conic quadratic inequality is given by

$$t \geq \max_{s \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m-m_1}} \frac{x^\top \left(\sum_{i=1}^m \Gamma_i s_i\right) x}{\beta^\top s + \left(\sum_{i=1}^m \gamma_i s_i\right)^\top z}$$

$$\text{s.t.} \sum_{i=1}^m \Gamma_i s_i \succeq 0, \ \sum_{i=1}^m \mathrm{Tr}(\Gamma_i) s_i \leq 1.$$

- How to work with $P = \mathrm{conv}\left(\left\{(e_S, \hat{Q}_S^{-1})_{S \in Z}\right\}\right)$ in practice?
  We give an MILP formulation for $\left\{(e_S, \hat{Q}_S^{-1})_{S \in Z}\right\}$.
  Preliminary tests show that this MILP is faster than perspective for some instances.

Research Questions

- Can we exploit matrix and constraint structure to obtain stronger relaxations? (Part 1) ✓

- What does strong mean for MIQP?
  Can we leverage polyhedral theory for MIQP? (Part 2) ✓

Agenda

Conclusions

- We characterize the convex hulls of MIQPs with indicators
- Convexification reduces to finding a facial description of a polytope
- We can use any tools from MILP to do so
- We can use polyhedral theory to understand strength of convexifications
- Offers insights into design of algorithms

References

- L. Wei, A. Gómez and S. Küçükyavuz, Ideal formulations for constrained convex optimization problems with indicator variables. *Mathematical Programming*, 192, 57-88, 2022.
- P. Liu, S. Fattahi, A. Gómez and S. Küçükyavuz, A Graph-based Decomposition Method for Convex Quadratic Optimization with Indicators. *Mathematical Programming*, 2022.
- L. Wei, A. Atamtürk, A. Gómez and S. Küçükyavuz, On the convex hull of convex quadratic optimization problems with indicators. *ArXiv* 2201.00387, 2021.
- H. Manzour, S. Küçükyavuz, H-H. Wu and A. Shojaie, Integer Programming for Learning Directed Acyclic Graphs from Continuous Data, *INFORMS Journal on Optimization*, 3(1), 46-73, 2021.
- L. Wei, A. Gómez and S. Küçükyavuz, On the Convexification of Constrained Quadratic Optimization Problems with Indicator Variables, In: D. Bienstock, G. Zambelli (eds), IPCO 2020, Lecture Notes in Computer Science, vol 12125, 433-447, Springer, Cham, 2020.
- S. Küçükyavuz, A. Shojaie, H. Manzour, L. Wei and H-H. Wu, Consistent Second-Order Conic Integer Programming for Learning Bayesian Networks. *ArXiv* 2005.14346, 2020.