Mixed-Projection Conic Optimization: A New Paradigm for Modeling Rank Constraints

Problem Setting

General low-rank problems with conic constraints:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \langle \boldsymbol{C},\boldsymbol{X}\rangle + \Omega(\boldsymbol{X}) + \lambda\cdot \mathrm{Rank}(\boldsymbol{X}) \quad (1)$$
$$\text{s.t. } \boldsymbol{A}\boldsymbol{X}=\boldsymbol{B}, \mathrm{Rank}(\boldsymbol{X})\le k,\ \boldsymbol{X}\in\mathcal{K}.$$

- $\mathcal{K}$ a proper cone.
- $\Omega(\boldsymbol{X})$ a spectral function, e.g., $\Omega(\boldsymbol{X})=\|\boldsymbol{X}\|_F^2$.
- Modeling power: matrix completion, ACOPF.
- Complexity: **we prove** $\exists\mathbb{R}$ complete.

Modeling Rank Nonlinearly

Cardinality can be modeled using **binaries**

$\|\boldsymbol{x}\|_0\le k \iff \exists \boldsymbol{z}\in\{0,1\}^n : \boldsymbol{e}^\top\boldsymbol{z}\le k, \boldsymbol{x}=\boldsymbol{z}\circ\boldsymbol{x}.$

Rank can be modeled using **projection matrices**

$$\mathrm{Rank}(\boldsymbol{X})\le k \iff \exists \boldsymbol{Y}\in\mathcal{Y}_n^k : \boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X},$$

where $\mathcal{Y}_n^k=\{\boldsymbol{Y}\in S^n : \boldsymbol{Y}^2=\boldsymbol{Y}, \mathrm{tr}(\boldsymbol{Y})\le k\}$.

- "Right" extension of binaries which satisfy $z^2=z$.

How to Model Projection Matrices

Formulate with QCQP Constraints in Gurobi

$$\mathcal{Y}_n^k=\{\boldsymbol{Y}\in S^n : \boldsymbol{U}\in\mathbb{R}^{n\times k}, \boldsymbol{Y}=\boldsymbol{U}\boldsymbol{U}^\top, \boldsymbol{U}^\top\boldsymbol{U}=\mathbb{I}\}.$$

Strengthen with SOCP approx of convex hull

$$Y_{i,i}Y_{j,j}\ge Y_{i,j}^2\ \forall i,j\in[n], Y_{i,i}\ge\sum_{t=1}^k U_{i,t}^2\ \forall i\in[n],$$
$$\pm 2Y_{i,j}+Y_{i,i}+Y_{j,j}\ge\|\boldsymbol{U}_i\pm\boldsymbol{U}_j\|_2^2\ \forall i,j\in[n].$$

Where $\mathrm{Conv}(\mathcal{Y}_n^k)=\{\boldsymbol{Y}\in S_+^n : \boldsymbol{Y}\preceq\mathbb{I}, \mathrm{tr}(\boldsymbol{Y})\le k\}$ is not representable in Gurobi.

A Min-Max Formulation

Rewrite as projection-only minimization problem

$$\min_{\boldsymbol{Y}\in\mathcal{Y}_n^k} f(\boldsymbol{Y})+\lambda\cdot\mathrm{tr}(\boldsymbol{Y}) \quad (2)$$

$$\text{with } f(\boldsymbol{Y}) := \min_{\boldsymbol{X}\in\mathcal{K}:\boldsymbol{A}\boldsymbol{X}=\boldsymbol{B},} \langle\boldsymbol{C},\boldsymbol{X}\rangle+\Omega(\boldsymbol{X}) \text{ s.t. } \boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X}$$

$$f(\boldsymbol{Y})=\max_{\boldsymbol{\alpha}} h(\boldsymbol{\alpha})-\Omega^\star(\boldsymbol{\alpha},\boldsymbol{Y}) \leftarrow \text{strong duality} \quad (3)$$

- **Key result:** $\Omega^\star$ is linear in $\boldsymbol{Y}$
- Strong duality removes the non-linearity $\boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X}$.
- Solve exactly via outer-approximation.
- Solve approximately by relaxing, rounding $\boldsymbol{Y}$ greedily.

Penalty Interpretation of Relaxation

$\Omega(\boldsymbol{X})=\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2$. Dual of (3) generalizes the perspective relax.

$$\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)}\ \min_{\boldsymbol{X},\boldsymbol{\Theta}} \langle\boldsymbol{C},\boldsymbol{X}\rangle+\frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\Theta})+\lambda\cdot\mathrm{tr}(\boldsymbol{Y}) \text{ s.t. } \begin{pmatrix}\boldsymbol{\Theta} & \boldsymbol{X}\\ \boldsymbol{X}^\top & \boldsymbol{Y}\end{pmatrix}\succeq\boldsymbol{0}.$$

Eliminate $\boldsymbol{Y},\boldsymbol{\Theta}$ for alternative to nuclear norm which generalizes the reverse Huber penalty from sparse linear regression:

$$\min_{\boldsymbol{X}} \langle\boldsymbol{C},\boldsymbol{X}\rangle+\sum_{i=1}^n \min\left(\frac{2\lambda}{\gamma}\sigma_i(\boldsymbol{X}), \lambda+\frac{\sigma_i(\boldsymbol{X})^2}{2\gamma}\right).$$

Scalability of Exact Method: Matrix Completion

Multi-tree branch+cut: optimal solutions after 20 cuts in 3000s.

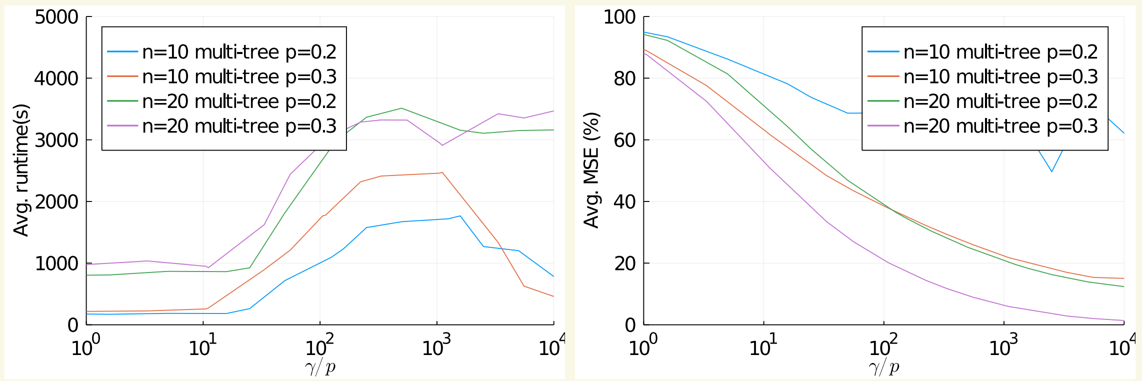

Figure: Vary γ, dimensionality $n\in\{10,20\}$, proportion of entries observed $p\in\{0.2,0.3\}$, fix rank $r=1$, measure runtime (left), MSE (right).

Comparison With Nuclear Norm

Noiseless 100×100 matrix completion problem.
Vary proportion of entries observed (p) and rank (r)

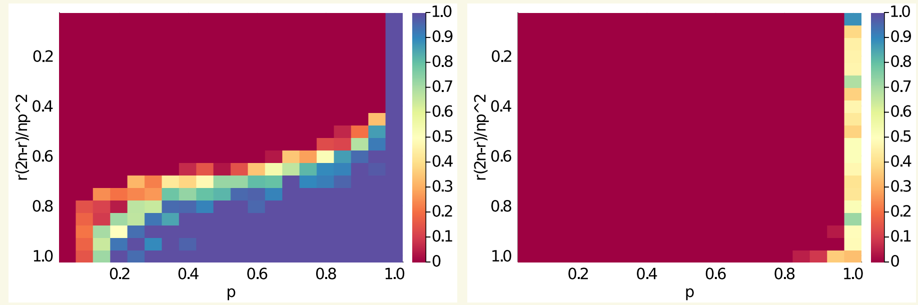

Figure: Prob. recovery relax+round (left), nuclear norm (right).

- New penalty dominates (more purple=better).

Solving the Relaxation at Scale.

- (2)'s relaxation decomposes into SDP-free problems in $\boldsymbol{X}$ and $\boldsymbol{Y}$'s eigenvalues.
- $\boldsymbol{Y}^\star=\sum_{i=1}^n\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^\top$ where $\boldsymbol{X}^\star=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ SVD.
- Relaxation amenable to alternating min.
- Solve relaxations when $n=1,000$s by iteratively solving QPs and doing top-k SVD.

Summary

- We **model** rank via projection matrices.
- Mixed-Projection Optimization **strictly generalizes** Mixed-Integer Optimization.
- We **extend** tools from MIO, including branch-and-cut and relax-and-round, to MPO.
- Branch-and-cut finds certifiably optimal solutions when $n=30$s in hours.
- Relax-and-round finds solutions with bound gap in hours when $n=1000$s.
- Further improvement: use custom solver.

Dimitris Bertsimas, Ryan Cory-Wright, Jean Pauphilet

email: ryancw@mit.edu, website: ryancorywright.github.io, code: github.com/ryancorywright/MixedProjectionSoftware

Operations Research Center, Massachusetts Institute of Technology

OPERATIONS RESEARCH CENTER