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

Problem Setting

General low-rank problems with conic constraints:

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

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

Modeling Rank Nonlinearly

Cardinality can be modeled using **binaries**

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

Rank can be modeled using **projection matrices**

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

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

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

Summary of Contributions

- We **model** rank via projection matrices.
- Mixed-Projection Optimization (MPO) **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 when $n = 1000$s in hours.
- Future work: custom solver to branch over $\mathcal{Y}_n^k$.

A Min-Max Formulation

Rewrite (1) as projection-only minimization problem

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

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

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

- **Key result:** $\Omega^\star$ linear in $\boldsymbol{Y}$
- Strong duality removes the non-linearity $\boldsymbol{X} = \boldsymbol{YX}$.
- Solve exactly via outer-approximation.

Scalability of Exact Method: Matrix Completion

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

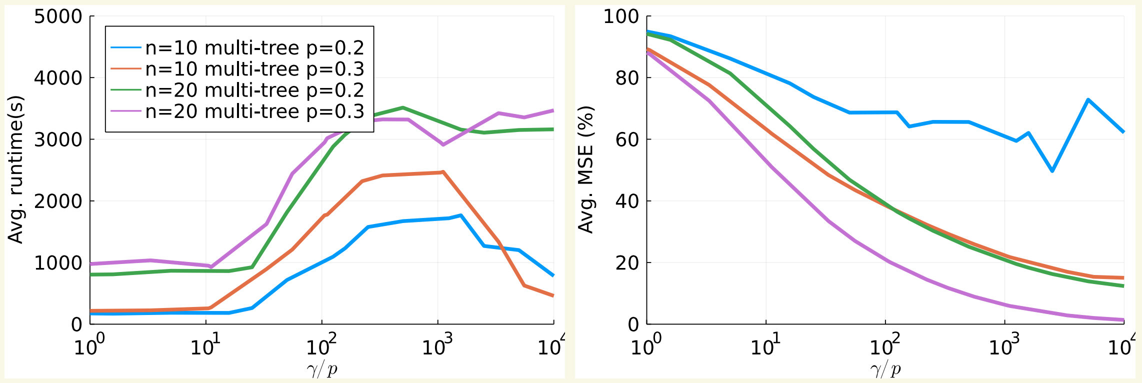

Figure: Vary γ, dimensionality $n \in \{10, 20\}$, proportion of entries observed $p\in\{0.2, 0.3\}$, fix rank $r = 1$, measure runtime (left), MSE (right).

Relaxation and Penalty Interpretation

Let $\Omega(\boldsymbol{X}) = \mathrm{tr}(f(\boldsymbol{X}))$. Valid relaxation of (1)-(3) is:

$$\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^k)} \min_{\boldsymbol{X}\in\mathcal{K}:\boldsymbol{AX}=\boldsymbol{B}} \langle \boldsymbol{C}, \boldsymbol{X}\rangle + \mathrm{tr}(g_f(\boldsymbol{X},\boldsymbol{Y})) + \lambda\cdot\mathrm{tr}(\boldsymbol{Y}), \quad (4)$$

where $g_f(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{Y} f(\boldsymbol{Y}^\dagger \boldsymbol{X})$ is matrix perspective of f.

Example: $\Omega(\boldsymbol{X}) = \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2$. Eliminate $\boldsymbol{Y}$: alternative to nuclear norm, generalizes reverse Huber penalty in sparse regression:

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

Solving the Relaxation at Scale.

- (4) decomposes into SDP-free problems in $\boldsymbol{X}$'s eigenvectors and $\boldsymbol{Y}$'s eigenvalues.
- Relaxation jointly convex in $\boldsymbol{X}, \boldsymbol{Y}$. Solve relaxation to optimality via alternating min.
- Bound gaps when $n = 1,000$s via alt. min for lower bound, greedy rounding for upper bound.

Comparison With Nuclear Norm

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

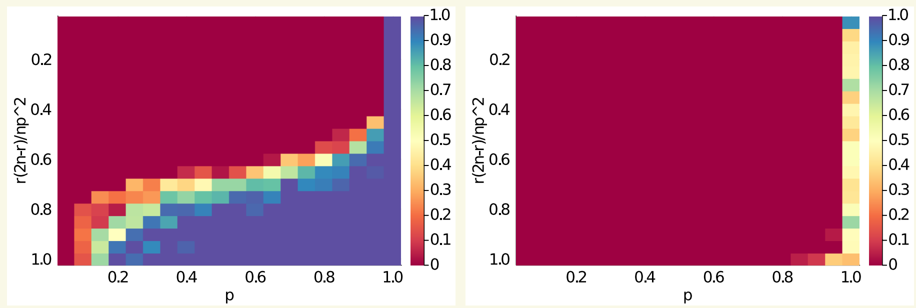

Figure: Prob. recovery relax+round (left), nuclear norm (right).

- New penalty dominates (more purple=better).

Application: Rank Regression

- Recover noisy rank-10 $50\times m$ matrix
- Vary no. observations $(\boldsymbol{X}_i, \boldsymbol{Y}_i)_{i=1}^m \in \mathbb{R}^{50}\times\mathbb{R}^{50}$
- Compare new penalties (blue/red), N.N. (green)

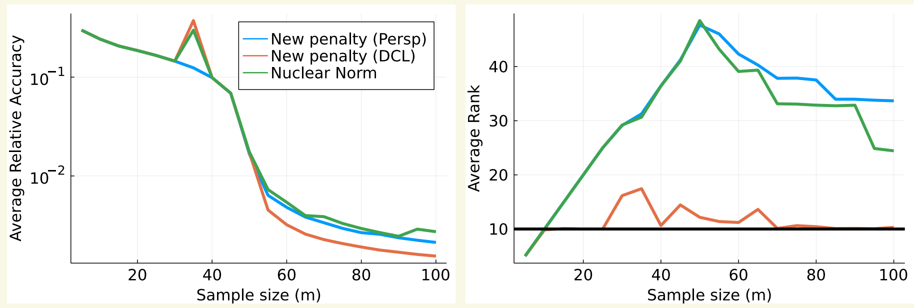

Figure: Relative accuracy (left), recovered rank (right).

- New penalties dominate nuclear norm (N.N.).

Dimitris Bertsimas, Ryan Cory-Wright, Jean Pauphilet.

Paper accepted at Operations Research, 2021; Winner of INFORMS George Nicholson Prize, 2020.

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

Operations Research Center, Massachusetts Institute of Technology

OPERATIONS RESEARCH CENTER