Optimal Low-Rank Matrix Completion:
Semidefinite Relaxations and Eigenvector Disjunctions

Dimitris Bertsimas (MIT), **Ryan Cory-Wright (ICBS, I-X)***, Sean Lo (MIT), Jean Pauphilet (LBS)

Imperial College London Control and Optimization Seminar Series

Preprint: arXiv 2305.12292

* Imperial College Business School and Imperial-X
Email: r.cory-wright@imperial.ac.uk
Website: ryancorywright.github.io

Save the Date: London Operations Research Day 2024

We have nine excellent speakers lined up, and will have a poster session for PhD students (to be announced)

- Date: **19 April 2024** (all day)
- Place: London Business School
- Website: londonorday.github.io

Confirmed Speakers

Sonya Crowe, University College London
Feryal Erhun, University of Cambridge
Hamza Fawzi, University of Cambridge
Raphael Hauser, University of Oxford
Rouba Ibrahim, University College London
Nitish Jain, London Business School
Ruth Misener, Imperial College
James Taylor, University of Oxford
Wolfram Wiesemann, Imperial College

Organizing Committee

Ryan Cory-Wright, Imperial College Business School and Imperial-X
Agni Orfanoudaki, Saïd Business School, Oxford and Exeter College
Jean Pauphilet, London Business School

Home Registration Program

This is the first iteration of the **London Operations Research Day (LORD)**, which brings together researchers in operations and associated fields from the London area. LORD is a single-track workshop consisting of a limited number of invited presentations and ample free time for interaction and discussion among participants. The workshop will be held at London Business School on Friday, April 19, 2024.

The workshop will also feature a poster session to give an opportunity to PhD students/postdocs to present their work. Those wishing to apply to apply for the poster session will need to fill out a form (to be announced). Acceptance notifications will subsequently be sent out along with instructions.

Stay tuned for more news about LORD by signing up to the mailing list here!

What is Low-Rank Matrix Completion?

Formulation:

Explain data well on average

$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} (X_{i,j} - A_{i,j})^2$$

Regularize

Restrict complexity

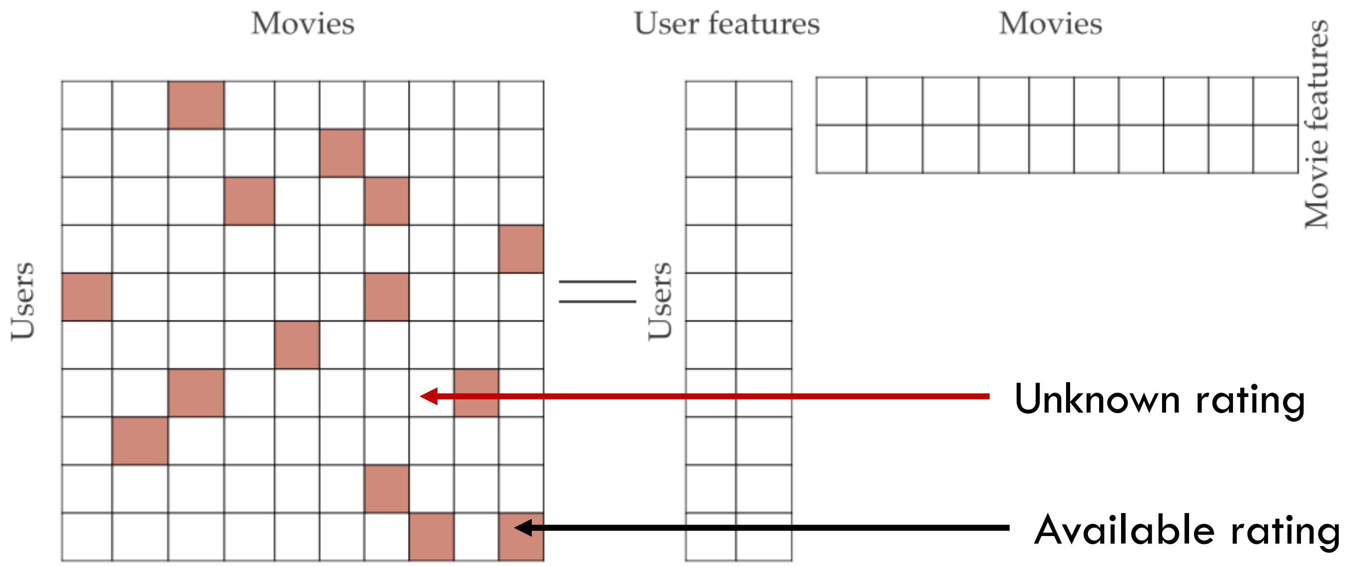
$$\text{s.t.} \quad \text{Rank}\,(\boldsymbol{X}) \leq k$$

Decision variables/Problem data

$X_{i,j}$: Predicted rating movie *j* by user *i*
$A_{i,j}$: Reported rating movie *j* by user *i*

Movie Recommendation

- Given user movie ratings, predict ratings for unseen movies.
- To make tractable, assume ratings depend on k factors (lead actor, lead actress, director, year, ..)



Why Solve Low-Rank Matrix Completion to Optimality?

Three Reasons

- Statistical Data regimes where global methods recover ground truth, polynomial time methods don't
  - Work by David Gamarnik's group (MIT Sloan) on *Overlap Gap Property*
- Reliability In high-stakes applications, important to make best imputations-And *know* best possible
- Performance out-of-sample Solving training problem to optimality improves test-set performance
  - Prior attempt that scaled to ~n=30: 0.6% MSE improvement on test set from certifiable optimality vs. AM

How do we Get There? A Tale of Two Problems

Low-Rank Matrix Completion

Explain data well on average

Restrict complexity

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} (X_{i,j} - A_{i,j})^2 \quad \text{s.t.} \quad \text{Rank}(\boldsymbol{X}) \leq k$$

Regularize

Not mixed-integer representable (Lubin et al. 2022), no methods solve it to optimality for k>1

"When you aren't sure what to do next, start with what you know and build from there" - Dimitris

Sparse Linear Regression

Explain data well on average

Regularize

Restrict complexity

$$\min_{\boldsymbol{w} \in \mathbb{R}^p} \quad \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \frac{1}{2\gamma}\|\boldsymbol{w}\|_2^2 + \mu\|\boldsymbol{w}\|_0$$

NP-hard, considered intractable 5-10 years ago

Often solved to optimality for $p = 10^6$ features (Bertsimas and Van Parys, Hazimeh/Mazumder/Saab)

Why Does Branch-and-Bound Scale for Sparse Regression?

Hazimeh, Mazumder & Saab (2022) propose custom branch and bound strategy that scales to $p = 10^6$.
Four key ingredients:

1. Strong Root Node Relaxation—leverage perspective relaxation (Frangioni and Gentile 2006)

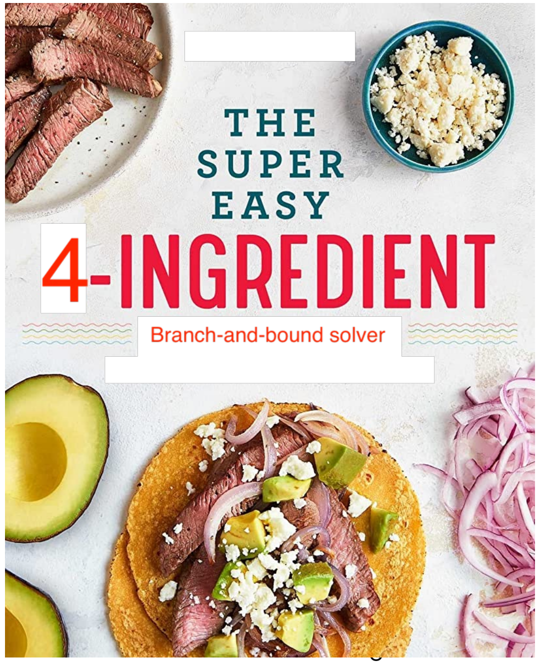
$$\min_{\boldsymbol{w},\boldsymbol{\rho}\in\mathbb{R}^p,\boldsymbol{z}\in\{0,1\}^p} \quad \frac{1}{2n}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{w}\|_2^2+\frac{1}{2\gamma}\boldsymbol{e}^\top\boldsymbol{\rho}+\mu\cdot\boldsymbol{e}^\top\boldsymbol{z} \quad \text{s.t.} \quad z_i\rho_i \geq w_i^2 \quad \forall i\in[p].$$

2. Efficient Branching Strategy—strong branching

3. High-Quality Incumbent Solutions—cyclic coordinate descent via L0Learn package

4. Efficient Nodal Subproblem Strategy
- Solve nodal relaxations via first-order method on dual, warm-started from parent node

With these ingredients, B&B *usually* scales



Agenda for Today

Use ideas from sparse regression (e.g. Hazimeh/Mazumder/Saab) as *roadmap*
Solve low-rank matrix completion to optimality for n~=150, k~=5 using ideas from MINLP

1. Strong Root Node Relaxation—leverage matrix perspective relaxation (Bertsimas et al. 2023)

2. Efficient Branching Strategy—eigenvector disjunctions (like in Saxena/Bonami/Lee 2010)

3. High-Quality Incumbent—alternating minimization with relaxation induced neighborhood search

4. Numerical Benchmarking, Comparison With Literature

Remark: Roadblock to n>150 is the scalability of semidefinite solvers

* With sincerest apologies if I missed one of your papers!
Very broad literature

Related Work*

Exact methods for related problems

MIQCP: Saxena, Bonami and Lee (2010)
Sparse Plus Low-Rank: Lee and Zou (2014)
ACOPF: Kocuk, Dey and Sun (2017)
Factor Analysis: Bertsimas, Copenhaver and Mazumder (2017)
Binary Matrices/Tensors: Kovács, Günlük and Hauser (2021), Soni, Linderoth, Luedtke and Pimentel-Alarcón (2023), Del Pia and Khajavirad (2023)
Trust Region: Anstreicher (2022)

Convex relaxations

Nuclear norm: Shapiro (1982), Fazel (2002), Candès and Recht (2009), Recht, Fazel and Parrilo (2010)
Log determinant: Fazel (2002)
Matrix perspective: Bertsimas, Cory-Wright and Pauphilet (2022, 23)
Perm-Invariant: Kim, Tawarmalani and Richard (2023)
Dantzig-Wolfe: Li and Xie (2022, 23)

Characterizing when relaxations tight

SOC/SDP relaxations: Barvinok (1995), Pataki (1998), Kim and Kojima (2003), Lavaei and Low (2012), Burer and Ye (2019), Wang and KIlInç-Karzan (2022, 23)
SOS relaxations: Goveia, Parrilo and Thomas (2010), Josz and Molzahn (2018), Barak and Moitra (2022)

Heuristics

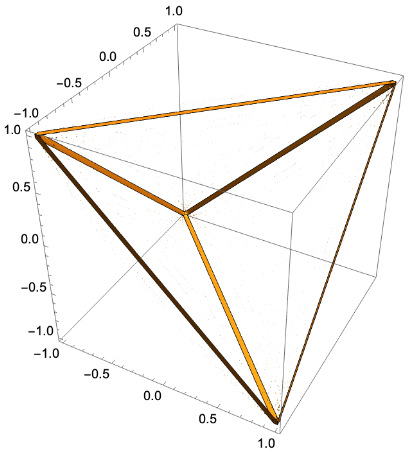
Alternating minimization: Burer and Monteiro (2003, 2005), Jain (2013), Waldspurger and Waters (2020)
Stochastic gradient descent: Recht and Ré (2013)
Frank-Wolfe: Freund, Grigas and Mazumder (2017)
Subgradient: Charisopoulos, Chen, Davis, Diaz, Ding and Drusvyatskiy (2021)
Non-convex penalties: Mazumder, Saldana and Weng (2020), Sagan and Mitchell (2021)

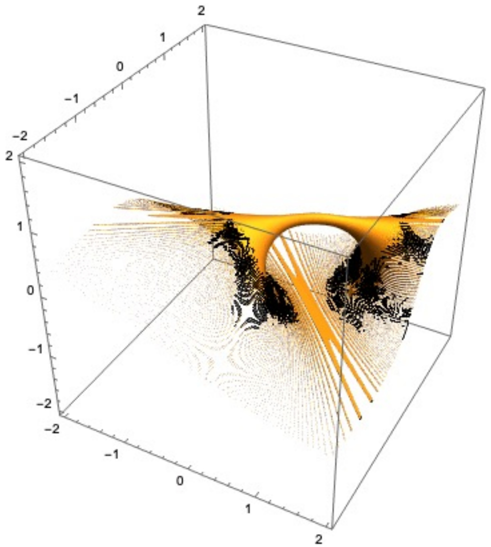
My Take on Related Work

- With heuristics, usually obtain high-quality solutions quickly
  - Burer-Monteiro alternating minimization usually performs remarkably well!
- But no guarantees on heuristic quality
  - Local methods sometimes 50% or more suboptimal; can't know if this happens without a *certificate*
- No generically applicable certifiably optimal methods that scale to $k>1$, $n>30$
  - If lots of problem structure (e.g., binaries, factor analysis), can solve to optimality by exploiting structure
  - Today: We propose method that applies to *any* low-rank problem, solve matrix completion to optimality

What do Rank Constraints Look Like?

Can be highly non-convex





$$\text{Rank}\begin{pmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{pmatrix} = 1$$

$$\text{Rank}\begin{pmatrix} x & y & z \\ y & z & 1-x \\ z & 1-x & 1-y \end{pmatrix} = 1$$

Left: 3D elliptope

Right: slice of Hankel matrix

Part I: A Strong Root Node Relaxation

A New Perspective on Low-Rank Optimization
D. Bertsimas, **R. Cory-Wright**, J. Pauphilet, Mathematical Programming, 2023.

Matrix Completion as a Mixed-Projection Problem

Original formulation:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} (X_{i,j} - A_{i,j})^2 \quad \text{s.t.} \quad \text{Rank}\,(\boldsymbol{X}) \leq k$$

This like trying to solve a sparse regression problem without using binary variables

Rank constraints can be modeled using projection matrices

$$\text{Rank}(\mathbf{X}) \leq k \iff \exists \mathbf{Y} \in \mathcal{Y}_n : \text{tr}(\mathbf{Y}) \leq k, \; \mathbf{X} = \mathbf{Y}\mathbf{X}$$

where $\mathcal{Y}_n := \{\mathbf{P} \in S^n : \mathbf{P}^2 = \mathbf{P}\}$

Mixed-Projection reformulation:

Matrix Analog of Logical Constraint

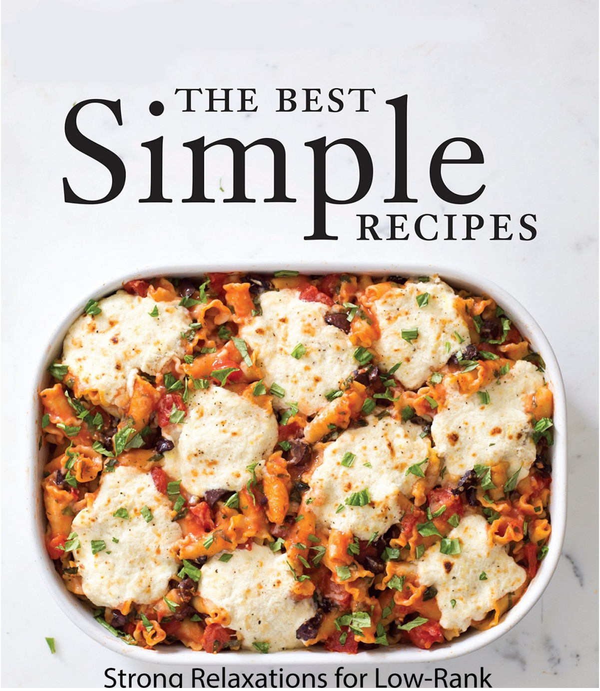
$$\min_{\mathbf{Y} \in \mathcal{Y}_n^k} \; \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} (X_{i,j} - A_{i,j})^2 \quad \text{s.t.} \quad \text{tr}(\mathbf{Y}) \leq k, \; \mathbf{X} = \mathbf{Y}\mathbf{X}$$

Matrix Analog of Binary Variables

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

D. Bertsimas, **R. Cory-Wright**, J. Pauphilet, Operations Research, 2022.

A Matrix Perspective Reformulation



Theorem: Can rewrite low-rank matrix completion w.l.o.g. as:

$$\min_{\boldsymbol{Y} \in \mathcal{Y}_n^k} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{\Theta} \in \mathcal{S}^m} \frac{1}{2\gamma} \mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j) \in \mathcal{I}} (X_{i,j} - A_{i,j})^2$$

$$\text{s.t.} \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \mathbf{0}.$$

Proof : $||X||_F^2 = tr(X^\top X)\ s.t. X = YX$ trace of matrix convex $f(X) = X^\top X$ under projection constraint

Replace f with matrix perspective g_f w.l.o.g.

$$g_{f_\omega}(\boldsymbol{\beta}, \boldsymbol{P}) = \begin{cases} \boldsymbol{P}^{\frac{1}{2}} f_\omega \left(\boldsymbol{P}^{-\frac{1}{2}} \boldsymbol{\beta} \boldsymbol{P}^{-\frac{1}{2}} \right) \boldsymbol{P}^{\frac{1}{2}} & \text{if } \boxed{\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\boldsymbol{P})} \\ \infty & \text{otherwise} \end{cases}$$

Captures bilinear constraint β=Pβ

g_f jointly convex in (X,Y) by construction

A Strong Root Node Relaxation

Apply matrix perspective reformulation technique to matrix completion, obtain:

$$\min_{\boldsymbol{Y} \in \mathcal{Y}_n^k} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{\Theta} \in \mathcal{S}^m} \frac{1}{2\gamma} \mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j) \in \mathcal{I}} (X_{i,j} - A_{i,j})^2 \text{ s.t. } \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \mathbf{0}.$$

Non-convexity: Just relax!

Impose rank constraint implicitly
via domain of Schur complement

A Strong Root Node Relaxation

Our Matrix Completion Formulation:

$$\min_{\boldsymbol{Y} \in \mathrm{Conv}\left(\mathcal{Y}_n^k\right)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m},\ \boldsymbol{\Theta} \in \mathcal{S}^m} \quad \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j) \in \mathcal{I}} (X_{i,j} - A_{i,j})^2 \text{ s.t. } \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \mathbf{0}.$$

where $\mathrm{Conv}\,(\mathcal{Y}_n^k) = \{\, \boldsymbol{P} \in S^n : \mathbf{0} \preceq \boldsymbol{P} \preceq \mathbb{I},\ \mathrm{tr}\,(\boldsymbol{P}) \leq k \,\}$ is semidefinite representable.

Generalizes the perspective relaxation, and often very tight—just like the perspective relaxation!

Part II: An Efficient Branching Strategy

Improving the Root Node Relaxation: Eigenvector Branching

Suppose we solve relaxation, get (X*, Y*). If Y* has binary eigenvalues, done

Otherwise, want to separate Y* from $\boldsymbol{Y} \in \mathcal{Y}_n^k$. Hard to do in original space, so lift!

Introduce new $n \times k$ matrix U, ideally, $Y = UU^\top, U^\top U = I$. New (equivalent) relaxation:

$$\min_{\substack{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n^k) \\ \boldsymbol{U} \in \mathbb{R}^{n \times k}}} \min_{\substack{\boldsymbol{X} \in \mathbb{R}^{n \times m} \\ \boldsymbol{\Theta} \in \mathcal{S}^m}} \frac{1}{2\gamma} \mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j) \in \mathcal{I}} (X_{i,j} - A_{i,j})^2 \quad \text{s.t.} \quad \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \mathbf{0}, \ \boldsymbol{Y} \succeq \boldsymbol{U}\boldsymbol{U}^\top$$

Improving the Root Node Relaxation: Eigenvector Branching

Given relaxation $\min_{\substack{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^k)\\ \boldsymbol{U}\in\mathbb{R}^{n\times k}}} \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{n\times m}\\ \boldsymbol{\Theta}\in\mathcal{S}^m}} \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j)\in\mathcal{I}} (X_{i,j} - A_{i,j})^2 \text{ s.t. } \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \boldsymbol{0},\ \boldsymbol{Y} \succeq \boldsymbol{U}\boldsymbol{U}^\top$

Want solution where $\hat{Y} \preccurlyeq \hat{U}\hat{U}^T$, then $\hat{Y} = \hat{U}\hat{U}^T$ and we are done. Suppose not.

Separation oracle x: $\boldsymbol{x}^\top(\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^\top - \hat{\boldsymbol{Y}})\boldsymbol{x} < 0,\ \|\boldsymbol{x}\|_2 = 1.$

Impose 2^k-term disjunction

$$\bigvee_{L\subseteq[k]} \left\{ (\boldsymbol{U},\boldsymbol{Y}) \;\middle|\; \begin{array}{ll} \boldsymbol{U}_j^\top \boldsymbol{x} \in [-1, \hat{\boldsymbol{U}}_j^\top \boldsymbol{x}] & \forall\, j \in L, \\ \boldsymbol{U}_j^\top \boldsymbol{x} \in (\hat{\boldsymbol{U}}_j^\top \boldsymbol{x}, 1] & \forall\, j \in [k]\setminus L, \\ \boldsymbol{x}^\top \boldsymbol{Y}\boldsymbol{x} \leq \sum_{j\in L} \left(\boldsymbol{x}^\top \boldsymbol{U}_j \hat{\boldsymbol{U}}_j^\top \boldsymbol{x} + (\hat{\boldsymbol{U}}_j - \boldsymbol{U}_j)^\top \boldsymbol{x} \right) & \\ \quad + \sum_{j\in[k]\setminus L} \left(\boldsymbol{x}^\top \boldsymbol{U}_j \hat{\boldsymbol{U}}_j^\top \boldsymbol{x} + (\boldsymbol{U}_j - \hat{\boldsymbol{U}}_j)^\top \boldsymbol{x} \right) & \end{array} \right\}$$

Theorem: Disjunction Separates $\hat{Y}$ from $\boldsymbol{Y}\in\mathcal{Y}_n^k$ ➡️ regions for branch-and-bound

Eigenvector Branching, Visualized

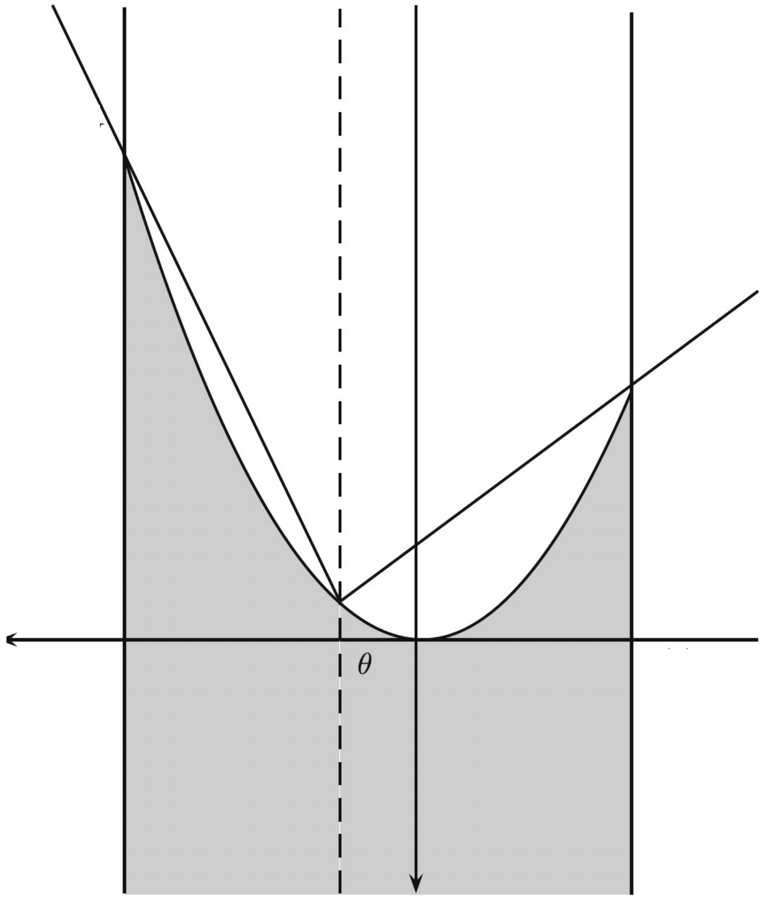
Would like to model expression

$$x^\top Y\, x \leq \left|\left|U^\top x\right|\right|_2^2.$$

Requires piecewise linear overestimator of $\left(U_i^\top x\right)^2$ on [-1, 1]

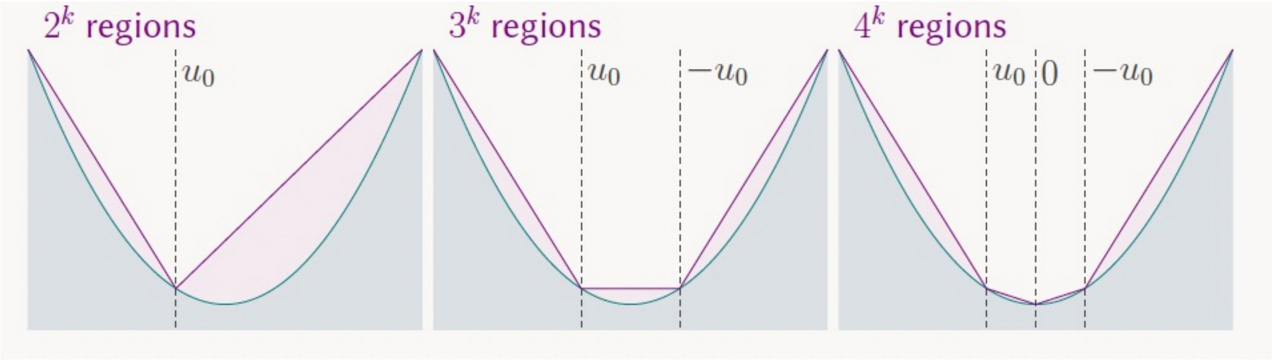
Therefore use $\theta = \widehat{U}_i^\top x$ as breakpoint, refine PWL upper approx.

Aim: develop good approximation of $\left(U_i^\top x\right)^2$ near optimal solution, without too many breakpoints



We Can Also Use Multiple Breakpoints

Branching factor becomes (no. pieces)^k



- Trade-off between strength of disjunction and no. nodes that need expanding
- 4 pieces better than 2 pieces for small n; breaks symmetry
- 2 pieces about as good as 4 pieces as n increases

Why Not McCormick Regions?

McCormick doesn't improve on root node without multiple partitions!

$$
\begin{aligned}
\min_{\substack{\boldsymbol{Y}\in \mathrm{Conv}(\mathcal{Y}_n^k),\\ \boldsymbol{U}\in\mathbb{R}^{n\times k}, \boldsymbol{V}\in\mathbb{R}^{n\times k\times k}}} \quad \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \boldsymbol{\Theta}\in\mathcal{S}^m} \quad & \frac{1}{2\gamma}\mathrm{tr}\left(\boldsymbol{\Theta}\right) + g(\boldsymbol{X}) \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \boldsymbol{0}, \ \boldsymbol{Y} \succeq \boldsymbol{U}\boldsymbol{U}^\top, \\
& \sum_{i=1}^n V_{i,j,j} = 1 \ \forall\ j \in [k], \ \sum_{i=1}^n V_{i,j_1,j_2} = 0 \ \forall\ j_1, j_2 \in [k]:\ j_1 \neq j_2, \\
& (V_{i,j_1,j_2}, U_{i,j_1}, U_{i,j_2}) \in \mathcal{M}(\underline{U}_{i,j_1}, \overline{U}_{i,j_2}, \underline{U}_{i,j_2}, \overline{U}_{i,j_2}), \quad \begin{matrix} \forall\ i \in [n], \\ \forall\ j_1, j_2 \in [k] \end{matrix}
\end{aligned}
$$

- Theoretically disjuncting on only one $U_{i,j}$ in each column j cannot improve root relaxation, no matter how many regions we partition [-1, 1] into!
- Practically McCormick routinely fails to improve root relaxation after expanding millions of nodes

Part III: A Branch-and-bound Scheme

Incumbent Generation

- Warm-start via Burer-Monteiro (BM) method at root node. $X = U V^{\top}, U \in R^{n\times k}, V \in R^{n\times k}$

Iteratively solve

$$\hat{V}^{t+1} = \arg\min_{V \in \mathbb{R}^{k\times m}} \frac{1}{2} \sum_{(i,j)\in\mathcal{I}} \left((\hat{U}^t V)_{i,j} - A_{i,j} \right)^2 + \frac{1}{2\gamma} \|\hat{U}^t V\|_F^2$$

$$\hat{U}^{t+1} = \arg\min_{U \in \mathbb{R}^{n\times k}} \frac{1}{2} \sum_{(i,j)\in\mathcal{I}} \left((U\hat{V}^{t+1})_{i,j} - A_{i,j} \right)^2 + \frac{1}{2\gamma} \|U\hat{V}^{t+1}\|_F^2$$

- Incumbent generation by Relaxation-Induced Neighborhood Search-type BM at "promising" leaf nodes

Math. Program., Ser. A 102: 71–90 (2005)

Digital Object Identifier (DOI) 10.1007/s10107-004-0518-7

Emilie Danna · Edward Rothberg · Claude Le Pape

Exploring relaxation induced neighborhoods to improve MIP solutions

Overall Branch-and-Bound Scheme

- Root node: Matrix perspective relaxation (Bertsimas et al. 2023)

$$\min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n^k)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m},\ \boldsymbol{\Theta} \in \mathcal{S}^m} \frac{1}{2\gamma} \mathrm{tr}(\boldsymbol{\Theta}) + \sum_{(i,j) \in \mathcal{I}} (X_{i,j} - A_{i,j})^2 \text{ s.t. } \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \mathbf{0}.$$

- Branching: Eigenvector disjunctions (Saxena et al. 2010)
- Incumbent: Burer-Monteiro (BM) at root node, RINS-BM at “promising” leaf nodes
- Node expansion: Solve SDPs using Mosek

Algorithm implemented in Julia. Code available: github.com/sean-lo/OptimalMatrixCompletion.jl

A New Semidefinite Relaxation

Improving Our Relaxation

- Useful Fact If a matrix is rank-k, all (k+1) x (k+1) minors have determinant zero
- In particular, if matrix rank-1, all 2x2 minors have determinant zero
- Therefore, take Shor relaxation of (vectorized) 2x2 minor, and obtain:

$$\min_{\substack{\boldsymbol{X},\boldsymbol{W}\in\mathbb{R}^{n\times m},\\ \boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^1),\\ \boldsymbol{\Theta}\in\mathcal{S}_+^m,\boldsymbol{V}}} \quad \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\Theta}) + \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} \left(A_{i,j}^2 - 2X_{i,j}A_{i,j} + W_{i,j}\right) \tag{5a}$$

$$\text{s.t.} \quad \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta} \end{pmatrix} \succeq \boldsymbol{0}, \tag{5b}$$

$$\begin{pmatrix} 1 & X_{i_1,j_1} & X_{i_1,j_2} & X_{i_2,j_1} & X_{i_2,j_2} \\ X_{i_1,j_1} & W_{i_1,j_1} & V^1_{i_1,(j_1,j_2)} & V^2_{(i_1,i_2),j_1} & V^3_{(i_1,i_2),(j_1,j_2)} \\ X_{i_1,j_2} & V^1_{i_1,(j_1,j_2)} & W_{i_1,j_2} & V^3_{(i_1,i_2),(j_1,j_2)} & V^2_{(i_1,i_2),j_2} \\ X_{i_2,j_1} & V^2_{(i_1,i_2),j_1} & V^3_{(i_1,i_2),(j_1,j_2)} & W_{i_2,j_1} & V^1_{i_2,(j_1,j_2)} \\ X_{i_2,j_2} & V^3_{(i_1,i_2),(j_1,j_2)} & V^2_{(i_1,i_2),j_2} & V^1_{i_2,(j_1,j_2)} & W_{i_2,j_2} \end{pmatrix} \succeq \boldsymbol{0}, \quad \begin{array}{l} \forall\ i_1 < i_2 \in [n], \\ \forall\ j_1 < j_2 \in [m], \end{array} \tag{5c}$$

$$\Theta_{j_1,j_2} = \sum_{i\in[n]} V^1_{i,(j_1,j_2)}\ \forall\ j_1 < j_2 \in [m], \quad \Theta_{j,j} = \sum_{i\in[n]} W_{i,j}, \ \forall\ j \in [m]. \tag{5d}$$

Closes most of gap between matrix perspective relaxation and optimal solution!

Part IV: Numerical Results

Experiment I: Justifying Algorithmic Design Decisions

Problem Setting

Recover low-rank **nxn rank-1** matrix

- Generate synthetic nxn rank-1 matrices
- Inject small amount of i.i.d. noise
- Sample $p = 2n \log n$ entries at random
- Vary n, branch-and-bound strategy
- Measure average relative optimality gap after one hour, over 20 instances
- Terminate early if gap of 10^-4

n	Alternating minimization	With McCormick disjunctions: Best-first	With McCormick disjunctions: Breadth-first	With McCormick disjunctions: Depth-first	With eigenvector disjunctions: Best-first	With eigenvector disjunctions: Breadth-first	With eigenvector disjunctions: Depth-first
10	✗	2.37×10^{-2}	3.06×10^{-2}	5.02×10^{-2}	5.28×10^{-3}	1.10×10^{-2}	2.60×10^{-2}
10	✓	3.29×10^{-4}	4.90×10^{-4}	7.92×10^{-3}	2.93×10^{-4}	4.91×10^{-4}	5.22×10^{-3}
20	✗	4.78×10^{-3}	4.78×10^{-3}	4.78×10^{-3}	2.61×10^{-4}	4.03×10^{-4}	4.03×10^{-3}
20	✓	5.51×10^{-4}	8.01×10^{-4}	8.01×10^{-4}	1.32×10^{-4}	1.92×10^{-4}	6.37×10^{-4}
30	✗	1.77×10^{-2}	1.77×10^{-2}	1.77×10^{-2}	2.00×10^{-3}	4.16×10^{-3}	1.35×10^{-2}
30	✓	2.01×10^{-3}	3.13×10^{-3}	3.13×10^{-3}	2.82×10^{-4}	4.53×10^{-4}	1.98×10^{-3}
40	✗	1.32×10^{-3}	1.32×10^{-3}	1.32×10^{-3}	3.28×10^{-4}	7.12×10^{-4}	6.11×10^{-4}
40	✓	1.12×10^{-4}	1.12×10^{-4}	1.12×10^{-4}	1.57×10^{-5}	1.94×10^{-5}	8.25×10^{-5}
50	✗	6.18×10^{-4}	6.18×10^{-4}	6.18×10^{-4}	8.11×10^{-5}	3.99×10^{-4}	8.11×10^{-4}
50	✓	6.37×10^{-5}	6.37×10^{-5}	6.40×10^{-5}	9.99×10^{-6}	1.13×10^{-5}	7.57×10^{-5}

Eigenvector disjunctions improve relative gap by order of magnitude
Alternating minimization exhibits similar improvement
Best-first search better than breadth-first or depth-first search

Experiment I: Justifying Algorithmic Design Decisions

Aside:

- Commercial non-convex solvers typically use McCormick relaxations (spatial branching), not eigenvector disjunctions
- Our results+related results in Anstreicher (2022) suggest commercial solvers may benefit from eigenvector disjunctions when solving non-convex (MI)QCPs
- Please implement this 😀

		With McCormick disjunctions			With eigenvector disjunctions		
n	Alternating minimization	Best-first	Breadth-first	Depth-first	Best-first	Breadth-first	Depth-first
10	✗	2.37×10^{-2}	3.06×10^{-2}	5.02×10^{-2}	5.28×10^{-3}	1.10×10^{-2}	2.60×10^{-2}
10	✓	3.29×10^{-4}	4.90×10^{-4}	7.92×10^{-3}	2.93×10^{-4}	4.91×10^{-4}	5.22×10^{-3}
20	✗	4.78×10^{-3}	4.78×10^{-3}	4.78×10^{-3}	2.61×10^{-4}	4.03×10^{-4}	4.03×10^{-3}
20	✓	5.51×10^{-4}	8.01×10^{-4}	8.01×10^{-4}	1.32×10^{-4}	1.92×10^{-4}	6.37×10^{-4}
30	✗	1.77×10^{-2}	1.77×10^{-2}	1.77×10^{-2}	2.00×10^{-3}	4.16×10^{-3}	1.35×10^{-2}
30	✓	2.01×10^{-3}	3.13×10^{-3}	3.13×10^{-3}	2.82×10^{-4}	4.53×10^{-4}	1.98×10^{-3}
40	✗	1.32×10^{-3}	1.32×10^{-3}	1.32×10^{-3}	3.28×10^{-4}	7.12×10^{-4}	6.11×10^{-4}
40	✓	1.12×10^{-4}	1.12×10^{-4}	1.12×10^{-4}	1.57×10^{-5}	1.94×10^{-5}	8.25×10^{-5}
50	✗	6.18×10^{-4}	6.18×10^{-4}	6.18×10^{-4}	8.11×10^{-5}	3.99×10^{-4}	8.11×10^{-4}
50	✓	6.37×10^{-5}	6.37×10^{-5}	6.40×10^{-5}	9.99×10^{-6}	1.13×10^{-5}	7.57×10^{-5}

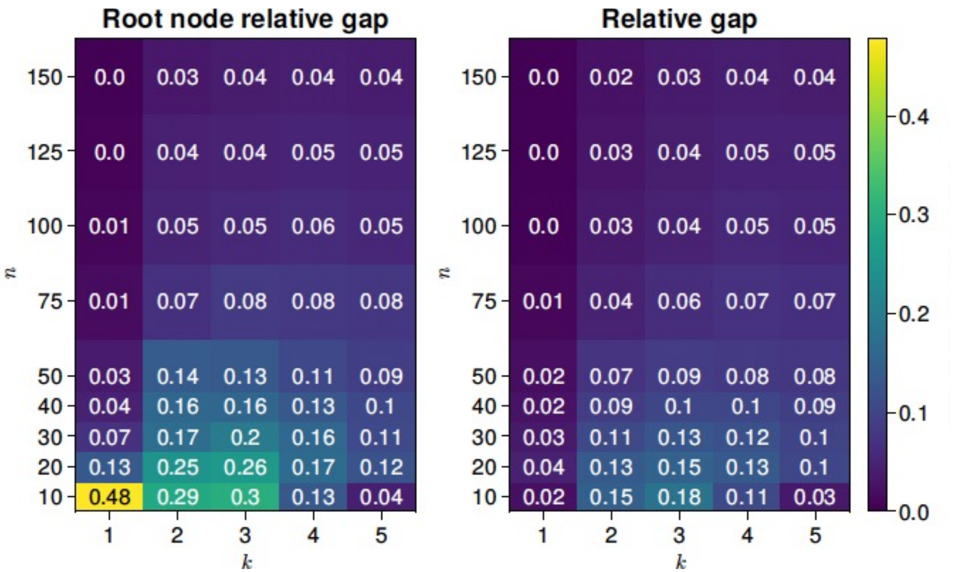
Eigenvector disjunctions improve relative gap by order of magnitude
Alternating minimization exhibits similar improvement
Best-first search better than breadth-first or depth-first search

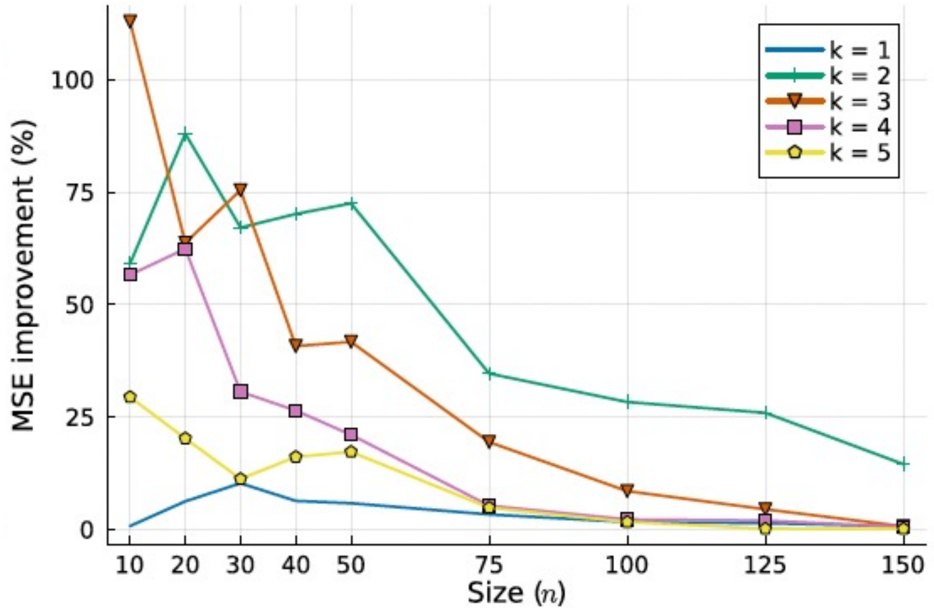
Experiment II: Scalability

Problem Setting

Recover low-rank n×n rank-k matrix

- Using best method in Experiment I
- Sample $p = 2nk \log n$ entries at random
- Vary n, k
- Measure average optimality gap at root node, after one hour over 50 instances
- Measure average MSE improvement compared to Burer-Monteiro





Surprisingly large MSE improvement from branch-and-bound!

Although edge decreases as n,k increases.

- many local minima and branch-and-bound needed when n,k small
- but few local minima, existing methods work well when n,k large

Do Local Methods Solve Matrix Completion to Optimality?

Many Positive Results in Literature

Deterministic guarantees for Burer–Monteiro factorizations of smooth semidefinite programs

NICOLAS BOUMAL
Mathematics Department and Program in Applied and Computational Mathematics, Princeton University
VLADISLAV VORONINSKI
Helm.ai
AND
AFONSO S. BANDEIRA
Department of Mathematics and Center for Data Science, Courant Institute of Mathematical Sciences, New York University

Rank optimality for the Burer-Monteiro factorization

Irène Waldspurger* Alden Waters†

ON THE BURER-MONTEIRO METHOD FOR GENERAL SEMIDEFINITE PROGRAMS

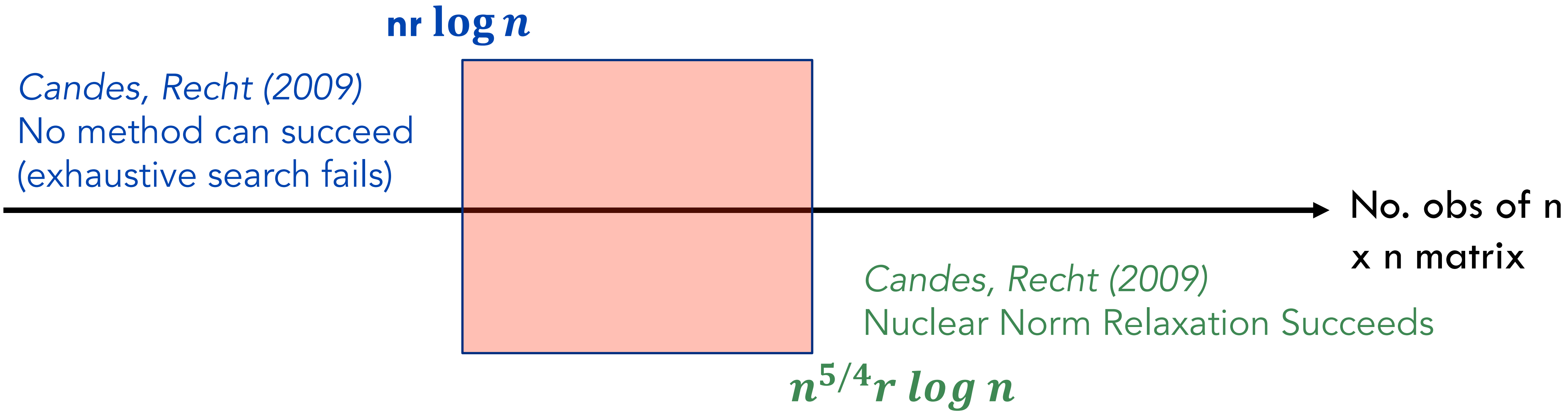
DIEGO CIFUENTES

Literature says: Given *enough* data, Burer-Monteiro solves low-rank matrix completion to optimality! 🎉
- BM is very fast! So, if assumptions on *enough* data hold, you should use it

But… Assumptions on *enough* data may not hold! ⚠️
- Our results add: When they don't, there are often many local optima, and global methods are needed

Why Do Exact Methods Perform Better?

The Literature Does Not Rule Out The Possibility of an “Overlap Gap”

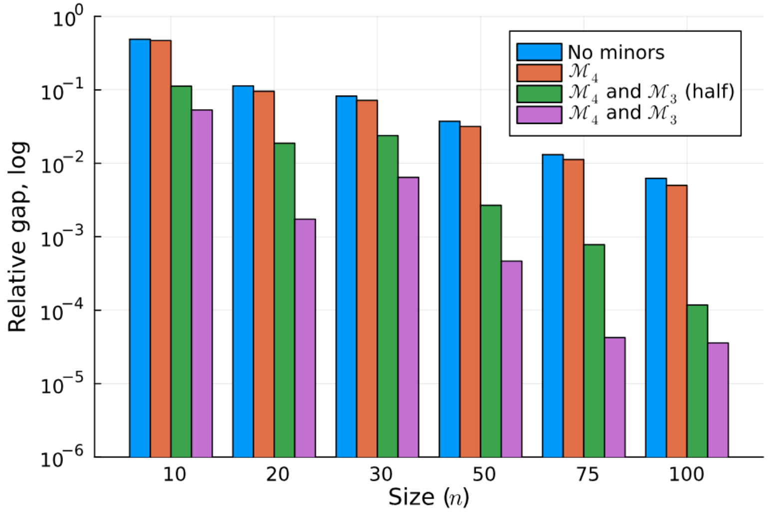


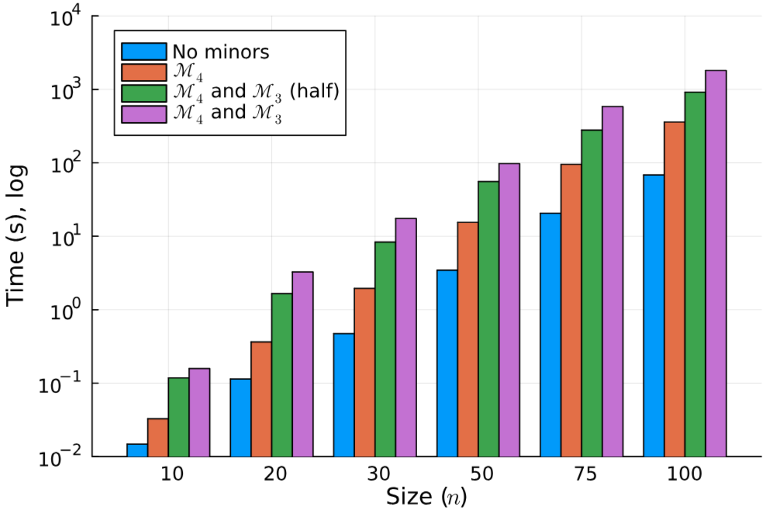
Previous slide confirms empirically:
branch-and-bound better in practice

Experiment III: Comparison of Convex Relaxations

Recover low-rank **nxn rank-1** matrix

- Vary n in 10, 20, 30, …, 75, 100
- Impose described Shor constraints on a subset of determinant minors, depending on no. observed entries
- Imposing M_4 and M_3 minors reduces optimality gap by 1-2 orders of magnitude, depending on problem setting





Summary: A Tale of Two Problems

Low-Rank Matrix Completion	Sparse Linear Regression
Parsimony rank	Parsimony sparsity
Modeling constraint **X=YX**	Modeling constraint $x = zx$ ($x = 0$ if $z = 0$)
Non-convex set $\mathbf{Y}^2 = \mathbf{Y}$ (Y projection matrix)	Non-convex set $z^2 = z$ (z binary)
Root node matrix perspective relaxation	Root node perspective relaxation
Branching eigenvectors	Branching 0-1 (strong)
Incumbent alternating minimization	Incumbent coordinate descent

Main contribution of talk: Build bridge from MIO to rank constraints, leverage MIO marketplace of ideas to solve low-rank matrix completion via branch-and-bound

Selected References I

Solving Two-Trust-Region Subproblems Using Semidefinite Optimization With Eigenvector Branching
K. Anstreicher, Journal of Optimization Theory and Applications, 2022.

Mixed-Projection Conic Optimization: A New Paradigm for Modeling Rank Constraints
D. Bertsimas, **R. Cory-Wright**, J. Pauphilet, Operations Research, 2022.

A New Perspective on Low-Rank Optimization
D. Bertsimas, **R. Cory-Wright**, J. Pauphilet, Mathematical Programming, 2023.

A Nonlinear Programming Algorithm for Solving Semidefinite Programs via Low-Rank Factorization
S. Burer, R. Monteiro, Mathematical Programming **95** 329-357, 2003.

Trace Inequalities and Quantum Entropy: An Introductory Course
E. Carlen, Entropy and the Quantum, 529:73-140, 2010.

Exploring Relaxation Induced Neighborhoods to Improve MIP Solutions
E. Danna, E. Rothberg, C. Le Pape, Mathematical Programming **102** 71-90, 2005.

Regularization vs. Relaxation: A Convexification Perspective of Statistical Variable Selection
H. Dong, K. Chen, J. Linderoth, submitted to Mathematical Programming, 2015.

Selected References II

Perspectives of Matrix Convex Functions
A. Ebadian, I. Nikoufar, M. E. Gordji, Proc. Natl. Acad. Sci **108**(18), 7313-7314, 2011.

A Matrix Convexity Approach to Some Celebrated Quantum Inequalities
E. G. Effros, Proc. Natl. Acad. Sci **106**(4), 1006-1008, 2009.

Semidefinite Approximations of the Matrix Logarithm
H. Fawzi, J. Saunderson, P. Parrilo. Foundations of Computational Mathematics **19**(2): 259-296, 2019

Perspective Cuts for a Class of Convex 0-1 Mixed Integer Programs
A. Frangioni, C. Gentile. Mathematical Programming **106**:225-236 (2006)

Perspective Reformulations of Mixed Integer Nonlinear Programs With Indicator Variables
O. Günlük, J. Linderoth. Mathematical Programming **124**:183-205 (2010)

Sparse Regression at Scale: Branch-and-bound Rooted in First-Order Optimization
H. Hazimeh, R. Mazumder, A. Saab. Mathematical Programming **196**:347-388 (2022)

Convex Relaxations of Non-Convex Mixed Integer Quadratically Constrained Programs: Extended Formulations A. Saxena, P. Bonami, J. Lee. Mathematical Programming **124**:383-411 (2010)

Experiment I: Backup-Average Runtime

Problem Setting

Recover low-rank **nxn rank-1** matrix:

- Generate synthetic nxn rank-1 matrices
- Inject small amount of i.i.d. noise
- Sample $p = 2n \log n$ entries at random
- Vary n, branch-and-bound strategy
- Measure average relative optimality gap after one hour
- Terminate early if gap of 10^-4

		With McCormick disjunctions			With eigenvector disjunctions		
n	Alternating minimization	Best-first	Breadth-first	Depth-first	Best-first	Breadth-first	Depth-first
10	✗	6.43×10^2	6.76×10^2	6.94×10^2	3.10×10^2	4.13×10^2	8.60×10^2
10	✓	1.40×10^2	1.36×10^2	5.70×10^2	6.37×10^1	1.04×10^2	3.98×10^2
20	✗	6.93×10^2	6.92×10^2	6.87×10^2	2.07×10^2	3.46×10^2	6.18×10^2
20	✓	2.06×10^2	2.28×10^2	2.37×10^2	5.88×10^1	9.17×10^1	2.63×10^2
30	✗	3.49×10^3	3.49×10^3	3.46×10^3	1.99×10^3	2.24×10^3	3.38×10^3
30	✓	9.21×10^2	9.04×10^2	9.28×10^2	3.07×10^2	3.35×10^2	8.86×10^2
40	✗	7.62×10^2	7.62×10^2	7.66×10^2	1.83×10^2	2.10×10^2	7.25×10^2
40	✓	5.14×10^2	5.08×10^2	5.19×10^2	8.19×10^1	9.53×10^1	4.99×10^2
50	✗	6.51×10^2	6.47×10^2	6.45×10^2	3.18×10^2	4.56×10^2	6.31×10^2
50	✓	3.22×10^2	3.26×10^2	3.26×10^2	1.08×10^2	1.47×10^2	4.35×10^2

Eigenvector disjunctions improve relative gap by order of magnitude
Alternating minimization exhibits similar improvement
Best-first search better than breadth-first or depth-first search

What does MPCO (not) generalize from MIO?

MIO captures notions of

- Finiteness: $z \in \{0, 1\}$
- Algebraicity: $z^2 - z = 0$

While MPCO captures notions of algebraicity ($Y^2 = Y$) but NOT finiteness-uncountably infinitely many Y

Therefore [what follows is conjecture]

- Results from MIO which depend on algebraic arguments (perspective reformulation, taking convex hulls)
- Or where enumeration argument can be replaced with coverage argument (branch-and-bound/cut)

Generalize from MIO. But..

- Results in MIO which depend on discreteness (e.g., MIR cuts) probably do not

Therefore, QCQP cuts (split cuts, PSD cuts) can be used by MPCO, but MIO cuts (Knapsack/flow cover) cannot

Remark: determining whether MIO result due to finiteness is non-trivial