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

Dimitris Bertsimas, Ryan Cory-Wright, Sean Lo, Jean Pauphilet.

website: https://sean-lo.github.io, **paper**: https://tinyurl.com/olrmc2023, **code**: https://github.com/sean-lo/OptimalMatrixCompletion.jl

Operations Research Center, Massachusetts Institute of Technology

Problem Setting

Find low-rank $\boldsymbol{X}$ that approx. fits observed entries $\mathcal{I}$ in noisy data $\boldsymbol{A}$:

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

Noiseless: $\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \|\boldsymbol{X}\|_F^2$ s.t. $X_{i,j} = A_{i,j}\ \forall (i,j)\in\mathcal{I},\ \mathrm{Rank}(\boldsymbol{X}) \le k$

Summary of Contributions

1. Branch-and-bound by **eigenvector disjunctions**
2. Tighter **semidefinite inequalities** via determinant minors
3. Restricted **alternating minimization** at each node

⇒ Solves (nonconvex) low-rank problem to **global optimality**

Matrix Perspective Reformulation Technique

Reformulate (1) exactly as [c.f. 2]:

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

- Isolates nonconvexity of $\mathrm{Rank}(\boldsymbol{X}) \le k$ in set of rank-k projections $\mathcal{Y}_n^k$
- Relaxing to $\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n^k) = \{\boldsymbol{0} \preceq \boldsymbol{Y} \preceq \mathbb{I}, \mathrm{tr}(\boldsymbol{Y}) \le k\}$ gives a SDP!

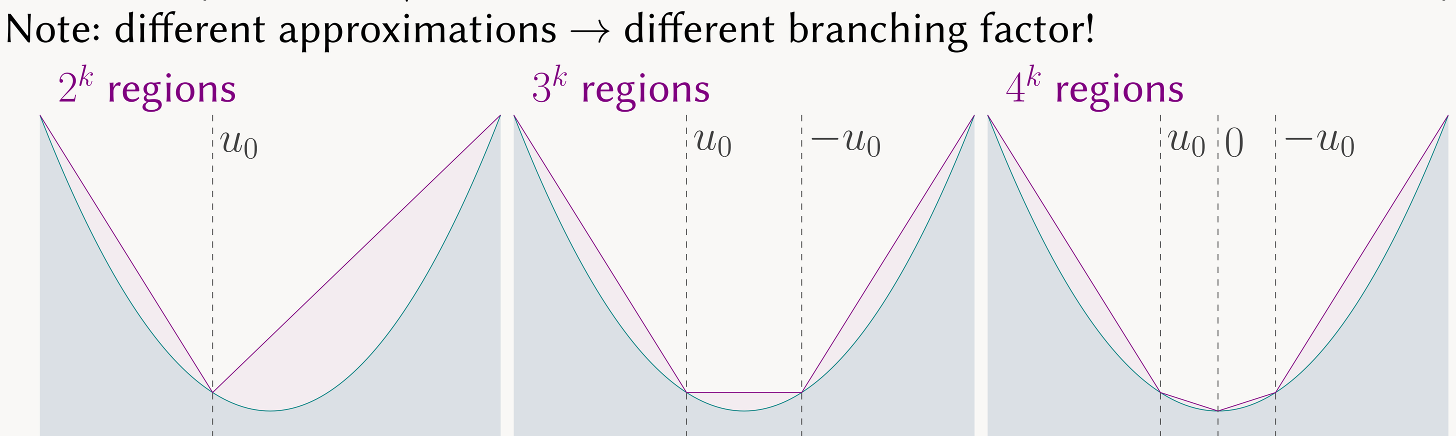
1. Eigenvector Disjunctions

Introduce $\boldsymbol{U}\in\mathbb{R}^{n\times k}$, and relax $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{U}^\top$ to $\boldsymbol{Y} \succeq \boldsymbol{U}\boldsymbol{U}^\top$:

$$\min_{\substack{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^k)\\ \boldsymbol{U}\in\mathbb{R}^{n\times k}}} \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{n\times m}\\ \boldsymbol{\Theta}\in\mathcal{S}^m}} \frac{1}{2}\sum_{(i,j)\in\mathcal{I}} (X_{i,j}-A_{i,j})^2 + \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\Theta}) \quad \text{s.t. } \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{\Theta}\end{pmatrix} \succeq \boldsymbol{0},\ \boldsymbol{Y} \succeq \boldsymbol{U}\boldsymbol{U}^\top \tag{3}$$

- If solution $(\hat{\boldsymbol{Y}}, \hat{\boldsymbol{U}})$ has $\hat{\boldsymbol{Y}} \preceq \hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^\top$, done!
- Else, there is eigenvector $\boldsymbol{x}$ such that $\boldsymbol{x}^\top(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^\top)\boldsymbol{x} > 0$. Cut:
$$\boldsymbol{x}^\top\boldsymbol{Y}\boldsymbol{x} \le \boldsymbol{x}^\top\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{x} = \|\boldsymbol{U}^\top\boldsymbol{x}\|_2^2 \tag{4}$$
- $k=1$: Following [6], linearize $u \mapsto u^2$ on $[-1,1]$ as below:
$$u^2 \le \begin{cases} f_L(u;u_0) := -u + uu_0 + u_0 & \text{if } u\in[-1,u_0] \\ f_R(u;u_0) := u + uu_0 - u_0 & \text{if } u \in (u_0, 1]\end{cases}$$
and set $u_0 = \hat{\boldsymbol{U}}^\top\boldsymbol{x}$, $u = \boldsymbol{U}^\top\boldsymbol{x}$.
- $k>1$: Yields 2^k disjunctive regions:
$$\bigvee_{L\subseteq[k]} \left\{ (\boldsymbol{U},\boldsymbol{Y}) \;\middle|\; \begin{array}{ll} \boldsymbol{U}_j^\top\boldsymbol{x} \in [-1, \hat{\boldsymbol{U}}_j^\top\boldsymbol{x}] & \forall j\in L, \\ \boldsymbol{U}_j^\top\boldsymbol{x} \in (\hat{\boldsymbol{U}}_j^\top\boldsymbol{x}, 1] & \forall j \in [k]\setminus L, \\ \boldsymbol{x}^\top\boldsymbol{Y}\boldsymbol{x} \le \sum_{j\in L} f_L(\boldsymbol{U}_j^\top\boldsymbol{x}; \hat{\boldsymbol{U}}_j^\top\boldsymbol{x}) + \sum_{j\in[k]\setminus L} f_R(\boldsymbol{U}_j^\top\boldsymbol{x}; \hat{\boldsymbol{U}}_j^\top\boldsymbol{x}) & \end{array}\right\} \tag{5}$$

Note: different approximations → different branching factor!



2. Tight SDP Inequalities via Determinant Minors

In rank-1 matrices, each 2-by-2 minor has determinant 0. Taking the moment matrix of $(1\ X_{i_1,j_1}\ X_{i_1,j_2}\ X_{i_2,j_1}\ X_{i_2,j_2})$ gives:

$$\det\begin{pmatrix} X_{i_1,j_1} & X_{i_1,j_2} \\ X_{i_2,j_1} & X_{i_2,j_2}\end{pmatrix} = 0 \implies \begin{pmatrix} 1 & X_{i_1,j_1} & X_{i_1,j_2} & X_{i_2,j_1} & X_{i_2,j_2} \\ X_{i_1,j_1} & W_{i_1,j_1} & V^1_{i_1,(j_1,j_2)} & V^2_{(i_1,i_2),j_1} & V^3_{(i_1,i_2),(j_1,j_2)} \\ X_{i_1,j_2} & V^1_{i_1,(j_1,j_2)} & W_{i_1,j_2} & V^3_{(i_1,i_2),(j_1,j_2)} & V^2_{(i_1,i_2),j_2} \\ X_{i_2,j_1} & V^2_{(i_1,i_2),j_1} & V^3_{(i_1,i_2),(j_1,j_2)} & W_{i_2,j_1} & V^1_{i_2,(j_1,j_2)} \\ X_{i_2,j_2} & V^3_{(i_1,i_2),(j_1,j_2)} & V^2_{(i_1,i_2),j_2} & V^1_{i_2,(j_1,j_2)} & W_{i_2,j_2}\end{pmatrix} \succeq \boldsymbol{0}$$

- Let $W_{i,j}$ model $X_{i,j}^2$ in objective and constraints;
- Link $\Theta_{i,j}$ (which models $\sum_l X_{i,l}X_{j,l}$) to appropriate W and V terms:
$$\Theta_{j_1,j_2} = \sum_{i\in[n]} V^1_{i,(j_1,j_2)}, \quad \forall j_1 < j_2 \in [m], \qquad \Theta_{j,j} = \sum_{i\in[n]} W_{i,j}, \quad \forall j\in[m].$$
- Extend to rank $k>1$ by modelling $\boldsymbol{X}$ as sum of rank-1 slices $\sum_{t\in[k]} \boldsymbol{X}^t$:
  - Requires Shor relaxation on the moment matrix of $(1\ X_{i,j}^1\ X_{i,j}^2\ \ldots\ X_{i,j}^k)$
  - New terms $H_{i,j}^{t_1,t_2}$ modelling $X_{i,j}^{t_1}X_{i,j}^{t_2}$

Key insight: modelling *some* minors → cutting planes for (3)

3. Alternating Minimization

Following [3], express $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}$, iterate until local minima reached

- **Improvement:** compute high-quality solutions at child nodes [e.g., in 4]
- Restrict $\boldsymbol{U}$ in feasible region of child node: linear constraints, and SOC approximation of $\boldsymbol{U}^\top\boldsymbol{U} \preceq \mathbb{I}$ [c.f. 1]
- *Randomly* initiate altmin, with probability decaying with tree depth.

Minimize $\boldsymbol{V}$ given $\boldsymbol{U}^*$

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

→ $\boldsymbol{V}^*$ →

Minimize $\boldsymbol{U}$ given $\boldsymbol{V}^*$

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

s.t. linear constraints on $\boldsymbol{U}$ from disjunctions

$$\left.\begin{array}{ll} \|\boldsymbol{U}_i\|_2^2 \le 1, & \forall i\in[k] \\ \|\boldsymbol{U}_i \pm \boldsymbol{U}_j\|_2^2 \le 2, & \forall i,j\in[k]\end{array}\right\} \text{SOC approx. of } \boldsymbol{U}^\top\boldsymbol{U} \preceq \mathbb{I}$$

→ $\boldsymbol{U}^*$ →

Computational Experiments: Root Node Relaxations

- Imposing some Shor SDP constraints provide tradeoff between relaxation strength and solve time!

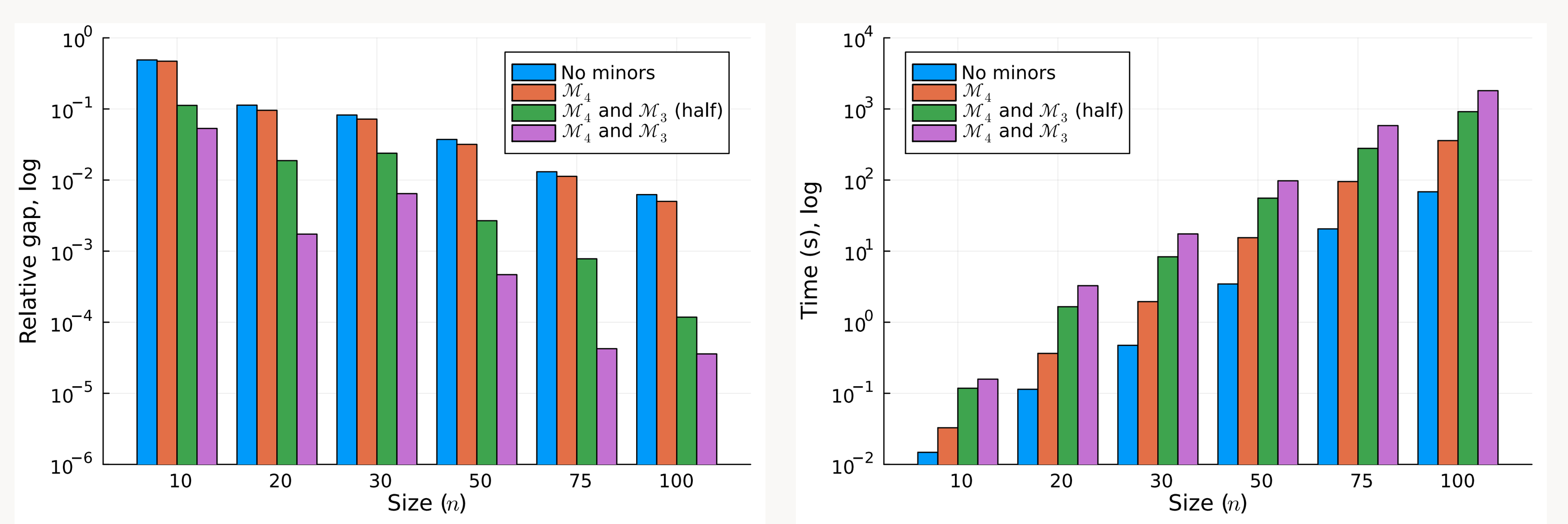


Figure: Root node relative gap (left) and time taken (right) at the root node for rank-1 matrix completion with $2kn\log_{10}(n)$ filled entries, with low regularization ($\gamma = 80.0$).

Computational Experiments: Scalability and Accuracy

- As n increases, SDP relaxations are tighter but take longer to solve!

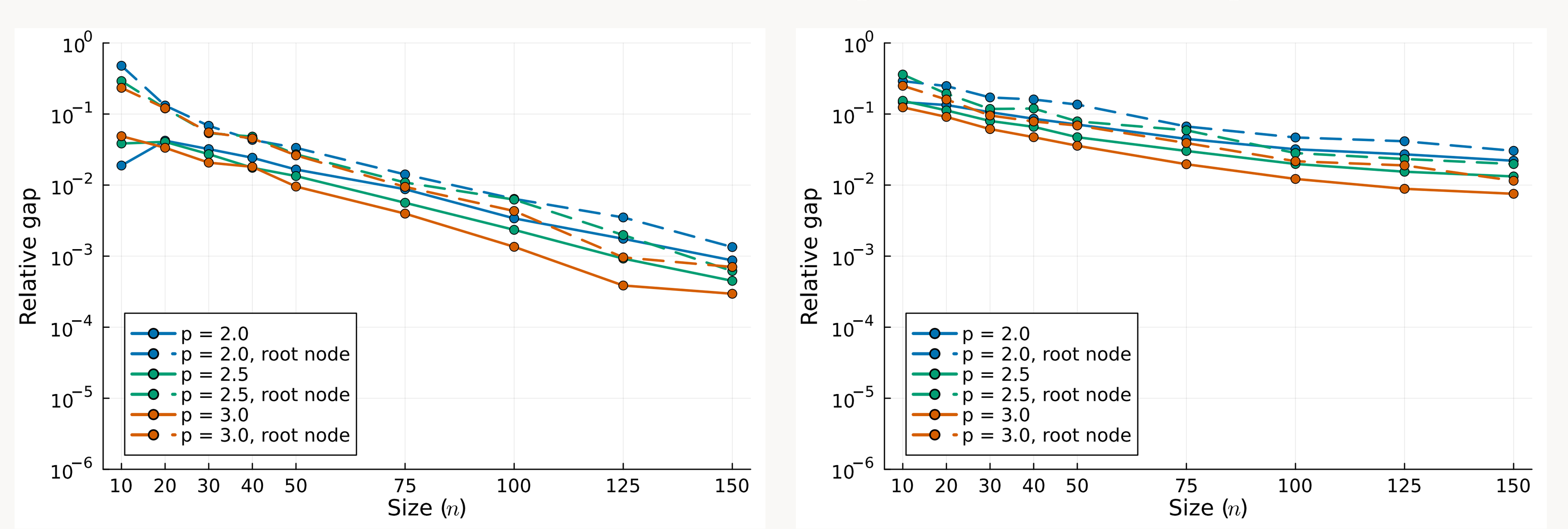


Figure: Relative optimality gap for rank-1 (left) and rank-2 (right) matrix completion with $pkn\log_{10}(n)$ filled entries, with low regularization ($\gamma = 80.0$).

- Outperforms B-M at root node and MFSGD [5] in terms of MSE

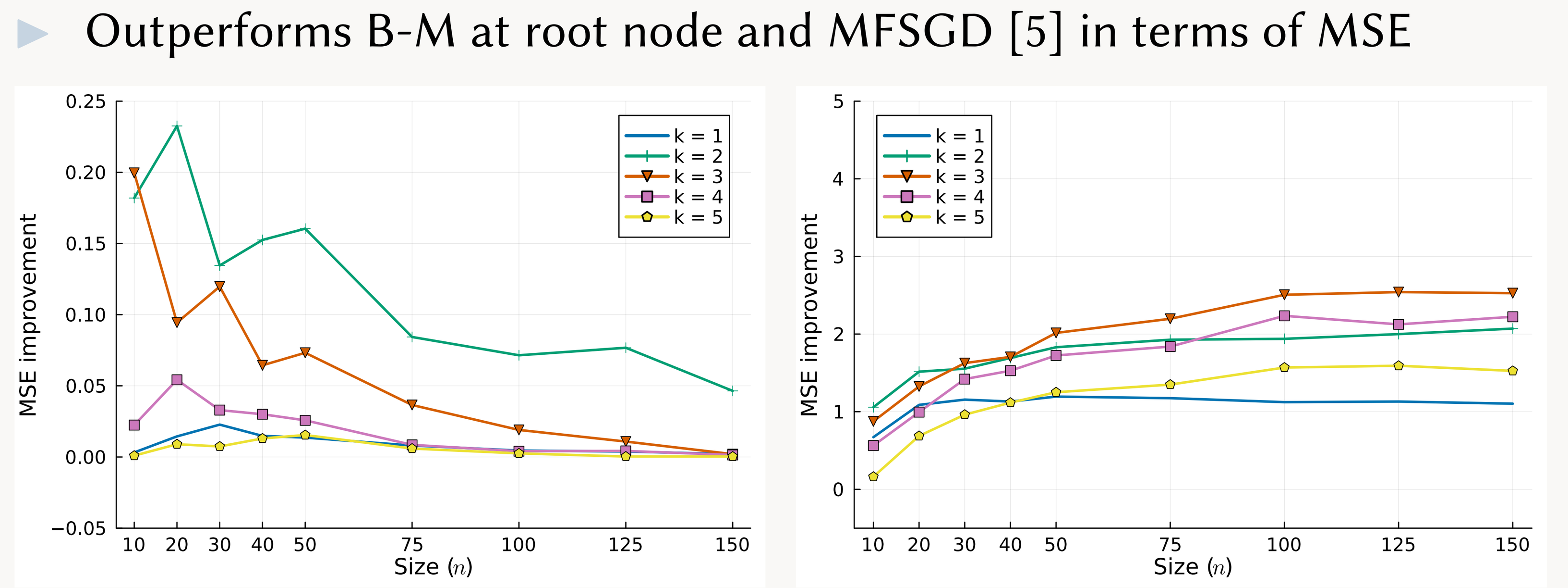


Figure: (Absolute) MSE improvement against alternating minimization at the the root node (left) and MatrixFactorization() (right) for rank-k matrix completion with $2kn\log_{10}(n)$ filled entries, with low regularization ($\gamma = 80.0$).

Bibliography

[1] Alper Atamtürk and Andrés Gómez. "Rank-one convexification for sparse regression". In: *arXiv preprint arXiv:1901.10334* (2019).

[2] Dimitris Bertsimas, Ryan Cory-Wright, and Jean Pauphilet. "A new perspective on low-rank optimization". In: *Mathematical Programming* (2023), pp. 1–46.

[3] Samuel Burer and Renato DC Monteiro. "Local minima and convergence in low-rank semidefinite programming". In: *Mathematical Programming* 103.3 (2005), pp. 427–444.

[4] Emilie Danna, Edward Rothberg, and Claude Le Pape. "Exploring relaxation induced neighborhoods to improve MIP solutions". In: *Mathematical Programming* 102.1 (2005), pp. 71–90.

[5] Chi Jin, Sham M Kakade, and Praneeth Netrapalli. "Provable efficient online matrix completion via non-convex stochastic gradient descent". In: *Advances in Neural Information Processing Systems* 29 (2016).

[6] Anureet Saxena, Pierre Bonami, and Jon Lee. "Convex relaxations of non-convex mixed integer quadratically constrained programs: projected formulations". In: *Mathematical Programming* 130.2 (2011), pp. 359–413.