Benefits and limits of rank overparametrization in low-rank matrix optimization

Andrew D. McRae

CERMICS, École nationale des ponts et chausées

Journées SMAI-MODE
March 20, 2026

Project status

- Collaboration with Richard Zhang (UIUC, Illinois, USA)
- Preprint coming "soon"

Low-rank matrix least squares

Consider the problem

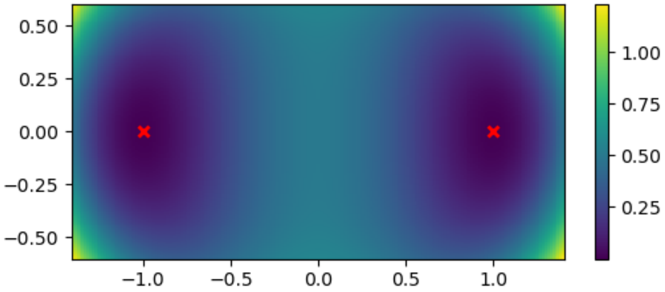
$$\min_{M \in \mathbf{R}^{d_1 \times d_2}} \frac{1}{2}\|\mathcal{A}(M) - b\|^2 \quad \text{s.t.} \quad \text{rank}(M) \leq r$$

- $\mathcal{A} \colon \mathbf{R}^{d_1 \times d_2} \to \mathbf{R}^n$ is linear
- We typically assume $b \approx \mathcal{A}(M_*)$ for some low-rank M_*
- Many applications in statistics/learning/signal processing
- Rank r is a hyperparameter we can choose
- Rank constraint makes the problem **nonconvex** and nonsmooth

Nonconvex landscapes

Simple example (symmetric, $r = r_* = 1$, Burer-Monteiro factored formulation):

$$\min_{u \in \mathbf{R}^2} \frac{1}{2}\|uu^T - zz^T\|_F^2, \qquad z = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



The nonconvex **landscape** is "benign"

- The only local optima are $u = \pm z$ (exact recovery of zz^T)
- Global property: a step toward global algorithmic guarantees

Key assumption: matrix restricted isometry property (RIP)

We are studying problems of the form

$$\min_{M \in \mathbf{R}^{d_1 \times d_2}} \frac{1}{2}\|\mathcal{A}(M) - b\|^2 \quad \text{s.t.} \quad \text{rank}(M) \le r$$

Example with benign landscape: $\mathcal{A}$ was the identity

- Loosen this assumption to "almost identity on low-rank matrices"

$\mathcal{A}$ has $(\boldsymbol{p}, \boldsymbol{\delta})$**-RIP** if

$$\text{rank}(H) \le p \quad \Longrightarrow \quad (1-\delta)\|H\|_F^2 \le \|\mathcal{A}(H)\|^2 \le (1+\delta)\|H\|_F^2$$

Much nonconvex landscape work assuming RIP

- Bhojanapalli et al., 2016; Ge et al., 2017; Park et al., 2017; Li et al., 2019; Zhu et al., 2018; Zhu et al., 2021...

Admittedly, **RIP is a strong assumption**

- Does not hold for many important practical problems
- Nevertheless a useful starting point for theory

Generalizations

This theory extends to non-quadratic functions

- $\frac{1}{2}\|\mathcal{A}(M) - b\|^2 \longrightarrow$ any convex and twice-differentiable $\phi(M)$
- RIP of $\mathcal{A} \longrightarrow$ restricted smoothness/strong convexity of ϕ:

$$\operatorname{rank}(H) \le p, \operatorname{rank}(M) \le q \quad \Longrightarrow \quad \mu\|H\|_F^2 \le \nabla^2\phi(M)[H,H] \le L\|H\|_F^2$$

In this talk, we only consider quadratic ϕ for clarity.

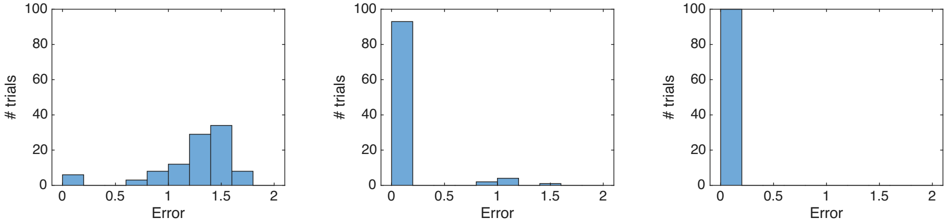
Rank overparametrization

Low-rank least-squares problem:

$$\min_{M \in \mathbf{R}^{d_1 \times d_2}} \frac{1}{2}\|\mathcal{A}(M) - b\|^2 \quad \text{s.t.} \quad \operatorname{rank}(M) \leq r, \quad \text{where} \quad b \approx \mathcal{A}(M_*)$$

Setting the rank hyperparameter r **larger** than the ground truth rank $r_* = \operatorname{rank}(M_*)$ often helps:

Empirical landscape results for phase retrieval, $d = 50$, $r_* = 1$ (trust-region, random initialization)



$r = 1$ $\qquad$ $r = 2$ $\qquad$ $r = 3$

Recent analysis of rank overparametrization under RIP assumption (Ma et al., 2023; Zhang, 2024; Zhang, 2025)

Current state-of-the-art for noisy recovery

Consider the symmetric problem (for $M_* \succeq 0$) with Burer-Monteiro factorization

$$\min_{U \in \mathbf{R}^{d \times r}} \frac{1}{2}\|\mathcal{A}(UU^T) - b\|^2, \qquad b = \mathcal{A}(M_*) + \xi. \tag{BM}$$

Theorem (Zhang, 2025)

Suppose $\operatorname{rank}(M_*) = r_* \leq r$, *and suppose*

$$\mathcal{A} \text{ has } (r + r_*, \delta)\text{-RIP} \qquad \text{with} \qquad \delta < \delta_c := \frac{1}{1 + \sqrt{r_*/\boldsymbol{r}}}.$$

Then every ***second-order critical point*** U *of (BM) satisfies*

$$\|UU^T - M_*\|_F \leq (\delta_c - \delta)^{-1}\sqrt{r + r_*}\|\mathcal{A}^*(\xi)\|_{op}.$$

- **SOCP** $\iff$ zero gradient and positive demidefinite Hessian in U
- Larger $\boldsymbol{r}$ allows RIP constant δ to be larger
- Condition on δ, r, r_* is optimal (counterexamples for any $\delta \geq \delta_c$)

Caveats

Theoretical guarantees in the asymmetric case lacking

- Impose a smaller value of critical RIP constant δ_c
- State-of-the-art results use an unnatural "balancing" regularizer:

$$\min_{\substack{U \in \mathbf{R}^{d_1 \times r} \\ V \in \mathbf{R}^{d_2 \times r}}} \frac{1}{2}\|\mathcal{A}(UV^T) - b\|^2 + \gamma\|U^T U - V^T V\|_F^2$$

  for some carefully chosen $\gamma > 0$
- Imbalance ($U^T U \neq V^T V$) significantly complicates analysis (cf. Valérie Castin's talk)

Error depends on rank hyperparameter

- Error bound is

$$\|M - M_*\|_F \lesssim \sqrt{r + r_*}\|\mathcal{A}^*(\xi)\|_{op}$$

- This dependence on r is inevitable
- Overparametrization may help landscape but **increases sensitivity to noise**

Adding regularization

Classic remedy to noise in low-rank matrix recovery: add nuclear norm regularizer

$$\min_{\substack{U \in \mathbf{R}^{d_1 \times r} \\ V \in \mathbf{R}^{d_2 \times r}}} \frac{1}{2}\|\mathcal{A}(UV^T) - b\|^2 \quad + \quad \lambda \frac{\|U\|_F^2 + \|V\|_F^2}{2}$$

- Follows from identity

$$\|M\|_{\text{nuc}} = \min_{U,V} \frac{\|U\|_F^2 + \|V\|_F^2}{2} \quad \text{s.t.} \quad UV^T = M$$

- Limited study in nonconvex landscapes literature (Li et al., 2019; McRae, 2026)

Main positive landscape result

$$\min_{\substack{U\in\mathbf{R}^{d_1\times r}\\ V\in\mathbf{R}^{d_2\times r}}} \frac{1}{2}\|\mathcal{A}(UV^T)-b\|^2 \quad + \quad \lambda\frac{\|U\|_F^2+\|V\|_F^2}{2}, \qquad b=\mathcal{A}(M_*)+\xi \qquad \text{(BM)}$$

Theorem

Suppose $\operatorname{rank}(M_*) = r_* \leq r$, *and suppose*

$$\mathcal{A} \text{ has } (r+r_*, \delta)\text{-RIP} \qquad \text{with} \qquad \delta < \delta_c = \frac{1}{1+\sqrt{r_*/r}}.$$

For any $\lambda \geq 0$, *every second-order critical point* (U,V) *of (BM) satisfies*

$$\|UV^T - M_*\|_F \leq (\delta_c - \delta)^{-1}[6\sqrt{r_*}\lambda + \sqrt{r+r_*}(\|\mathcal{A}^*(\xi)\|_{op} - \lambda)_+].$$

- Identical δ_c as previous (optimal) state-of-the-art
- Allows **asymmetry/imbalance** (new analysis tools needed)
- Under similar conditions, we have exact recovery of low-rank global optima
- Error bound...

Interpretation of the error bound

We showed

$$\|M - M_*\|_F \lesssim \sqrt{r_*}\lambda \quad + \quad \sqrt{r + r_*}(\|\mathcal{A}^*(\xi)\|_{op} - \lambda)_+$$

This interpolates two previous bounds:

1. Convex relaxation from matrix sensing literature:

$$\|M_{cvx} - M_*\|_F \lesssim \sqrt{r_*}\lambda \quad \text{for} \quad \lambda \gtrsim \|\mathcal{A}^*(\xi)\|_{op}$$

2. Unregularized nonconvex result we saw before:

$$\|M_{unreg} - M_*\|_F \lesssim \sqrt{r + r_*}\|\mathcal{A}^*(\xi)\|_{op} \quad \text{for} \quad \lambda = 0$$

If $\lambda \approx \|\mathcal{A}^*(\xi)\|_{op}$, we obtain $\|M - M_*\|_F \lesssim \sqrt{r_*}\|\mathcal{A}^*(\xi)\|_{op}$

- Statistically optimal (matrix sensing literature)
- **No dependence on rank hyperparameter *r***
  - (Except through the RIP constants δ, δ_c, whose effect is often minor)

Another caveat: RIP constant depends on rank parameter

Our main assumption is

$$\mathcal{A} \text{ has } (r + r_*, \delta)\text{-RIP} \qquad \text{with} \qquad \delta < \delta_c = \frac{1}{1 + \sqrt{r_*/\boldsymbol{r}}}.$$

This means that

$$(1-\delta)\|H\|_F^2 \leq \|\mathcal{A}(H)\|^2 \leq (1+\delta)\|H\|_F^2 \quad \text{for all} \quad \operatorname{rank}(H) \leq r_* + \boldsymbol{r}.$$

- Larger r **improves** δ_c...
- But δ potentially **also increases** with r!

From this, it is **unclear** whether overparametrization improves the landscape or not!

Second contribution: counterexamples

Consider

$$\min_{\substack{U \in \mathbf{R}^{d_1 \times r} \\ V \in \mathbf{R}^{d_2 \times r}}} \quad \frac{1}{2}\|\mathcal{A}(UV^T) - b\|^2 + \lambda \frac{\|U\|_F^2 + \|V\|_F^2}{2}, \qquad b = \mathcal{A}(M_*) + \xi \qquad \text{(BM)}$$

Theorem (Informal)

For any $\lambda \geq 0$, $r_1 \geq r_* \geq 1$, and $\delta < \frac{1}{1+\sqrt{r_*/r_1}}$, there exist $\mathcal{A}$, M_*, ξ with $\|\mathcal{A}^*(\xi)\|_{op} \approx \lambda$ for which

1. *$\mathcal{A}$ has $(r_1 + r_*, \delta)$-RIP, so (BM) has a "benign landscape" with $r = r_1$, but,*
2. *For certain $r_2 > r_1$, (BM) with $r = r_2$ has a spurious local optimum (U, V) with $\|UV^T - M_*\|_F \geq \|M_*\|_F$.*

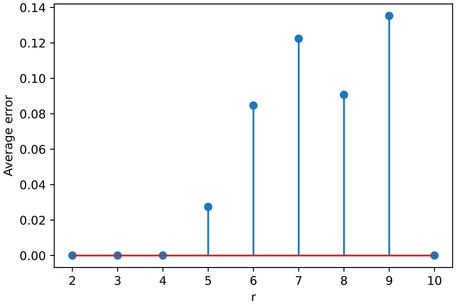
Takeaway: **too much overparametrization can introduce spurious local optima!**

- Based on a construction from (Zhang, 2024; Zhang, 2025)

Numerical illustrations

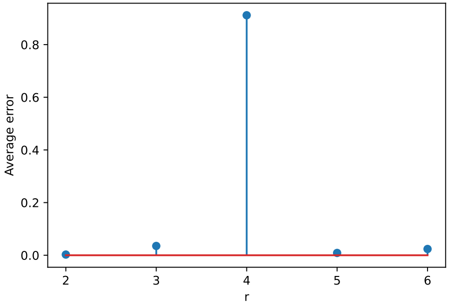
Explicit construction from theorem

- $r_* = 2$, $d = 10$
- Trust-regions, random init.
- "Pathological"?



Gaussian measurements (matrix sensing)

- $r_* = 2$, $d = 100$
- $n = 420 = 2.1 r_* d$ random measurements



Conclusion

We considered the nonconvex landscape of

$$\min_{\substack{U \in \mathbf{R}^{d_1 \times r} \\ V \in \mathbf{R}^{d_2 \times r}}} \frac{1}{2}\|\mathcal{A}(UV^T) - b\|^2 + \lambda \frac{\|U\|_F^2 + \|V\|_F^2}{2}, \quad b \approx \mathcal{A}(M_*)$$

with possible rank overparametrization $r \geq r_* = \mathrm{rank}(M_*)$

Main contributions:

1. Refined landscape analysis under restricted isometry property (RIP)
   - Recovers optimal dependence on r_*, r, and RIP constant δ
   - Improvements with asymmetry ($U \neq V$)
   - Optimal error rate independent of rank hyperparameter r
2. Counterexamples showing that rank overparametrization can hurt landscape

Future/ongoing work

1. Understand better the (complicated) effect of nuclear norm regularization
2. Algorithmic considerations (cf. Irène Waldspurger's talk)
   - Likely requires better understanding of balancing (cf. Valérie Castin's talk)

Thanks!

References I

Bhojanapalli, Srinadh, Behnam Neyshabur, and Nati Srebro (Dec. 2016). "Global Optimality of Local Search for Low Rank Matrix Recovery". In: *Proc. Conf. Neural Inf. Process. Syst. (NeurIPS)*. Barcelona, Spain, pp. 3873–3881.

Ge, Rong, Chi Jin, and Yi Zheng (Aug. 2017). "No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis". In: *Proc. Int. Conf. Mach. Learn. (ICML)*. Sydney, Australia, pp. 1233–1242.

Li, Qiuwei, Zhihui Zhu, and Gongguo Tang (2019). "The non-convex geometry of low-rank matrix optimization". In: *Inform. Inference.* 8.1, pp. 51–96.

Ma, Ziye, Yingjie Bi, Javad Lavaei, and Somayeh Sojoudi (2023). "Geometric Analysis of Noisy Low-Rank Matrix Recovery in the Exact Parametrized and the Overparametrized Regimes". In: *INFORMS J. Opt.* 5.4, pp. 356–375.

McRae, Andrew D. (2026). "Low solution rank of the matrix LASSO under RIP with consequences for rank-constrained algorithms". In: *Math. Program.* 215, pp. 717–741. Pre-published.

References II

Park, Dohyung, Anastasios Kyrillidis, Constantine Carmanis, and Sujay Sanghavi (2017). "Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach". In: *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*. Fort Lauderdale, FL, USA, pp. 65–74.

Zhang, Richard Y. (2024). "Improved global guarantees for the nonconvex Burer-Monteiro factorization via rank overparameterization". In: *Math. Program.*

— (2025). "Sharp Global Guarantees for Nonconvex Low-Rank Recovery in the Noisy Overparameterized Regime". In: *SIAM J. Optim.* 35.3, pp. 2128–2154.

Zhu, Zhihui, Qiuwei Li, Gongguo Tang, and Michael B. Wakin (2018). "Global Optimality in Low-Rank Matrix Optimization". In: *IEEE Trans. Signal Process.* 66.13, pp. 3614–3628.

— (2021). "The Global Optimization Geometry of Low-Rank Matrix Optimization". In: *IEEE Trans. Inf. Theory* 67.2, pp. 1308–1331.