How Much Restricted Isometry is Needed in Nonconvex Matrix Recovery?

Richard Y. Zhang[1], Cédric Josz[2], Somayeh Sojoudi[2], Javad Lavaei[1] [1]IEOR, UC Berkeley, [2]EECS, UC Berkeley

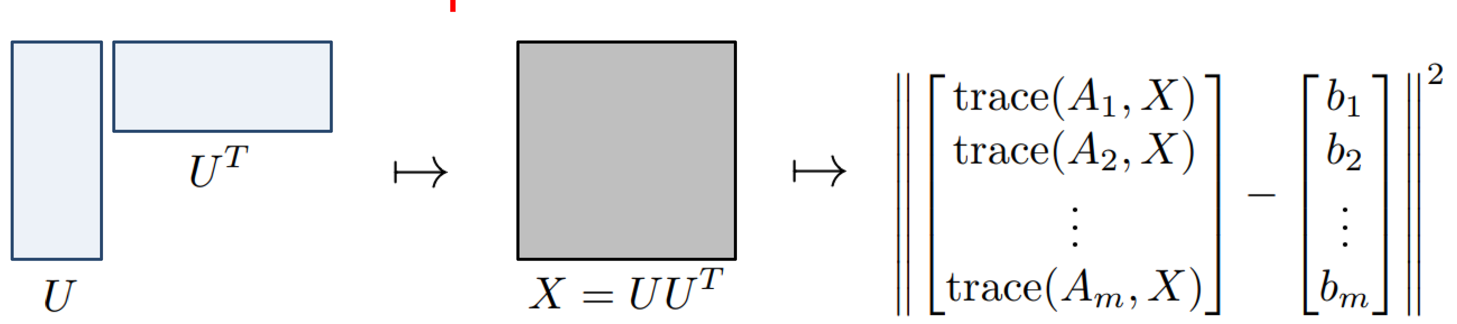
- Existing guarantees for nonconvex matrix recovery based on restricted isometry (or similar ideas).
- How much restricted isometry? Strong RIP.
- Modest RIP not enough. If δ-RIP with δ ≥ 1/2, then many counterexamples.
- SGD may succeed, may fail. No guarantees.
- RIP-like ideas most useful for nearly isotropic measurements or models.

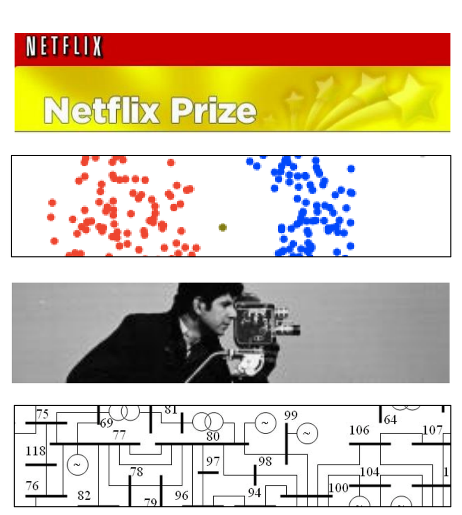
Nonconvex Matrix Recovery

$$\text{minimize } \|\mathcal{A}(UU^T)-b\|^2 \quad \text{over} \quad U\in\mathbb{R}^{n\times r}$$

$$\mathcal{A}(X) = [\mathrm{trace}(A_1 X) \quad \cdots \quad \mathrm{trace}(A_m X)]^T$$

1. Express low-rank matrix as product of factors
2. Minimize least-squares loss of linear model



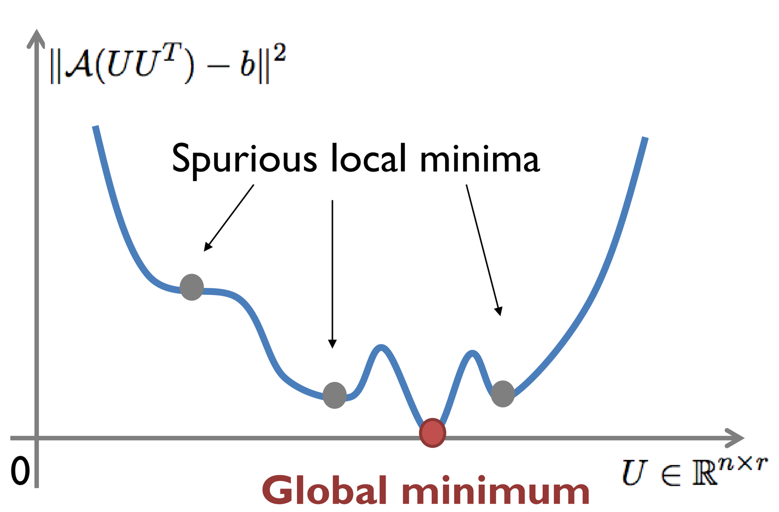


Recommendation engines

Cluster analysis

Phase retrieval

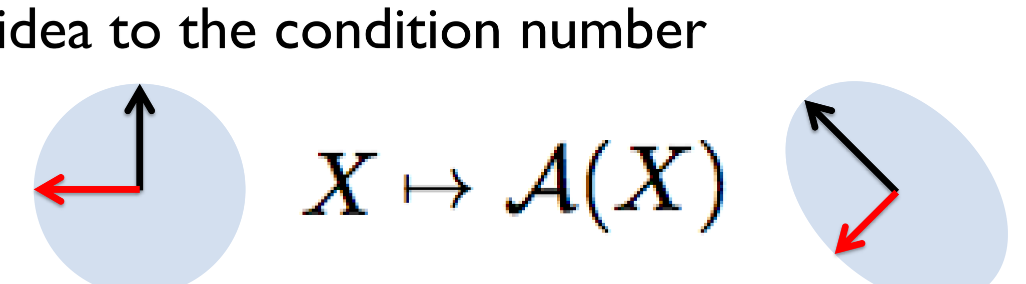
Power system state estimation



δ-Restricted Isometry Property (δ-RIP)

$$(1-\delta)\|X\|_F^2 \le \|\mathcal{A}(X)\|^2 \le (1+\delta)\|X\|_F^2 \qquad \forall\, \mathrm{rank}(X)\le 2r$$

Similar idea to the condition number



How much restricted isometry?

If $\delta < 1/5$, then no spurious local minima.

(Bhojanapalli et al. 2016; Ge et al. 2017; Li & Tang 2017; Zhu et al. 2017)

Suppose $Z\in\mathbb{R}^{n\times r}$ such that $\mathcal{A}(ZZ^T)=b$

$$\|\mathcal{A}(UU^T)-b\|^2 = \|\mathcal{A}(UU^T-ZZ^T)\|^2 \approx \|UU^T-ZZ^T\|_F^2$$

objective is δ-approximation of function with no spurious local minima.

Similar proof idea for sparse measurements.

(Ge et al. 2015; 2016; 2017; Sun et al. 2015; 2016; Park et al. 2017; etc.)

Can $\delta < 1/5$ be significantly improved?

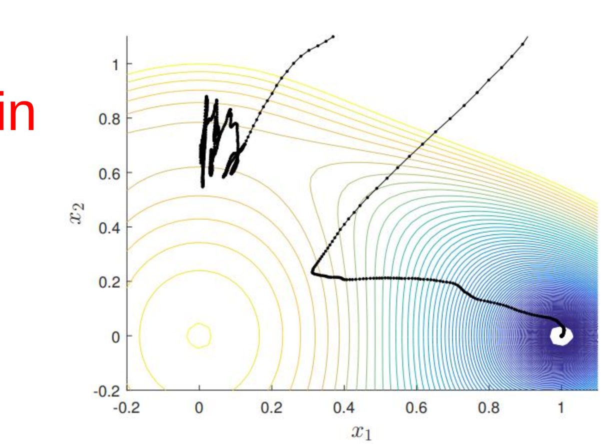
- Conservative. Real-world data $\delta \approx 0.99$.
- Generality and usefulness of RIP and similar assumptions.
- Spurious local min vs property of SGD?
- Success as m → infinity?

Counterexample for $\delta = 1/2$

$$A_1 = \begin{bmatrix}\sqrt{2} & 0\\ 0 & 1/\sqrt{2}\end{bmatrix},\quad A_2 = \begin{bmatrix}0 & \sqrt{3/2}\\ \sqrt{3/2} & 0\end{bmatrix},\quad A_3 = \begin{bmatrix}0 & 0\\ 0 & \sqrt{3/2}\end{bmatrix}$$

Spurious local min $x = (0, 1/\sqrt{2})$

Ground truth $z = (1,0)$

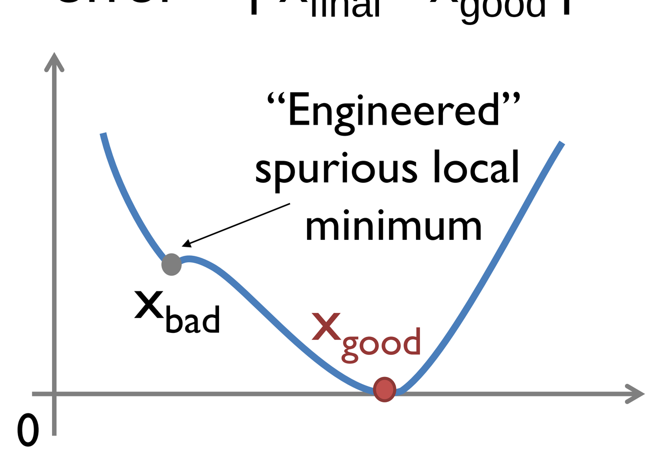


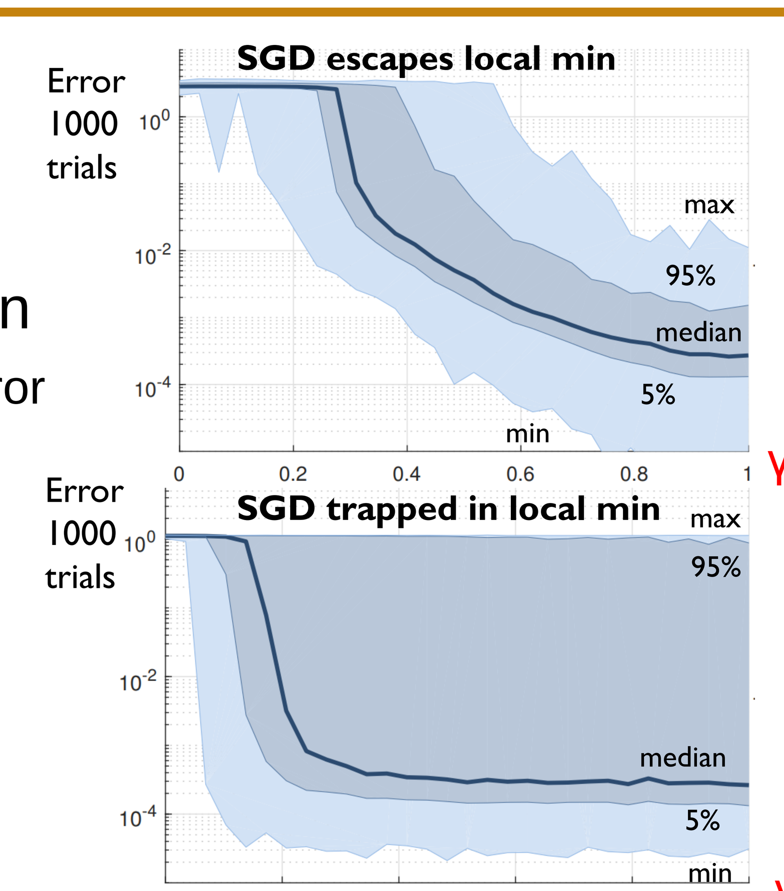
- 100,000 trials w/ SGD
- 87,947 successful
- 12% failure rate

Generalization to **arbitrary rank-1** ground truth

SGD and local min?

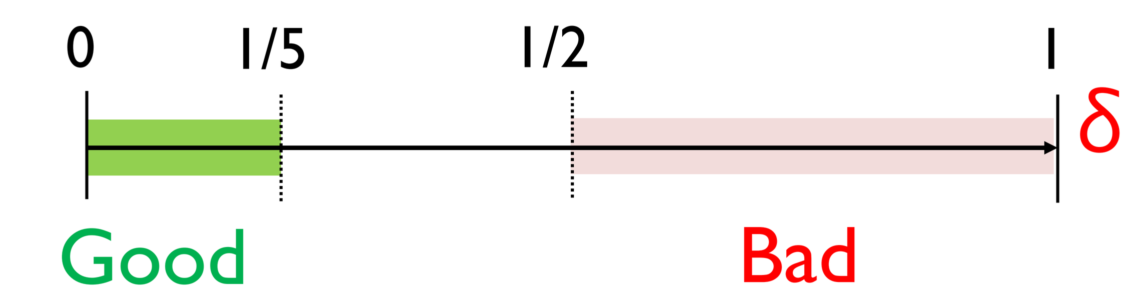
1. Select γ in [0,1]
2. Start SGD at $x_{init} = (1-\gamma)\, x_{bad} + \gamma$ Gaussian
3. Make 10k SGD steps, measure error $\text{error} = |x_{final} - x_{good}|$





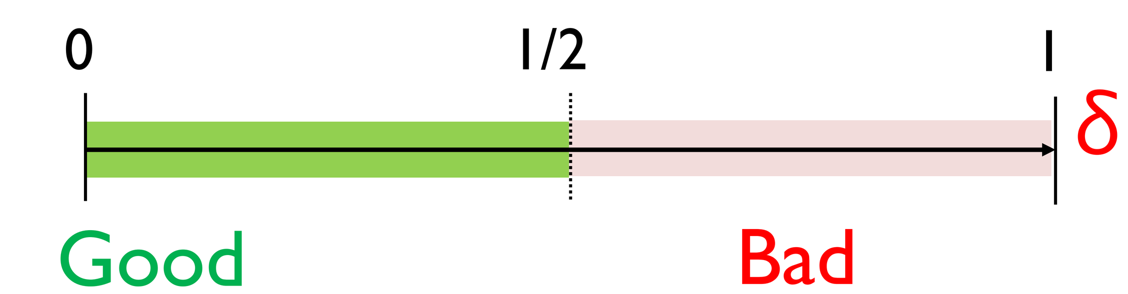
Theorem 1. Counterexamples are ubiquitous

Given vectors x, z not collinear, there exists a counterexample that satisfies δ-RIP with $1/2 \le \delta < 1$, and has z as global min and x as spurious local min.



Theorem 2. Sharp RIP-based guarantee

If $\delta < 1/2$ and **r = 1**, then no spurious local minima.



(Zhang, Sojoudi, Lavaei. Submitted to JMLR 2018)

"If λ-RIP, then x is not a spurious local min for the recovery of z"

Optimization reformulation

$$\begin{aligned} &\text{minimize } \delta \text{ over } \mathcal{A}\\ &\text{subject to } f(u)=\tfrac{1}{2}\|\mathcal{A}(uu^T-zz^T)\|^2,\\ &\nabla f(x)=0,\quad \nabla^2 f(x)\succeq 0,\\ &\mathcal{A} \text{ satisfies } \delta\text{-RIP}.\end{aligned}$$

optimal $\delta^* \le \lambda$ → refutation
optimal $\delta^* > \lambda$ → proof

Refutation via counterexample

$$\begin{aligned} &\text{find } f(u)=\tfrac{1}{2}\|\mathcal{A}(uu^T-zz^T)\|^2,\\ &\text{such that } \nabla f(x)=0,\quad \nabla^2 f(x)\succeq 0,\\ &\mathcal{A} \text{ satisfies } \delta\text{-RIP}.\end{aligned}$$

feasible for $\delta = \lambda$ → refutation
infeasible for $\delta = \lambda$ → proof

Convex upper-bound

$$\begin{aligned} &\text{minimize } \delta \text{ over } \mathcal{A}^T\mathcal{A}\\ &\text{subject to } f(u)=\tfrac{1}{2}\|\mathcal{A}(uu^T-zz^T)\|^2,\\ &\nabla f(x)=0,\quad \nabla^2 f(x)\succeq 0,\\ &\mathrm{cond}(\mathcal{A}^T\mathcal{A}) \le \frac{1-\delta}{1+\delta}.\end{aligned}$$

optimal $\delta_{ub}^* \le \lambda$ → refutation
optimal $\delta_{ub}^* > \lambda$ → nothing

Key Lemmas. For all x, z in $\mathbf{R}^n$:
1) $\delta_{ub}^* < 1$ iff x, z not collinear;
2) $\delta_{ub}^* \ge 1/2$; 3) $\delta^* = \delta_{ub}^*$.

Emails. (R.Y. Zhang) ryz@alum.mit.edu (C. Josz) cedric.josz@gmail.com (S. Sojoudi) sojoudi@berkeley.edu (J. Lavaei) lavaei@berkeley.edu

This work was supported by the ONR Awards N00014-17-1-2933 and ONR N00014-18-1-2526, NSF Award 1808859, DARPA Award D16AP00002, and AFOSR Award FA9550- 17-1-0163