Neural Information Processing Systems 2018

Wed Dec 5th 5 - 7 PM @ Room 210 & 230 AB

Poster #46

How Much Restricted Isometry is Needed in Nonconvex Matrix Recovery?

Richard Y. Zhang

Cédric Josz

Somayeh Sojoudi

Javad Lavaei

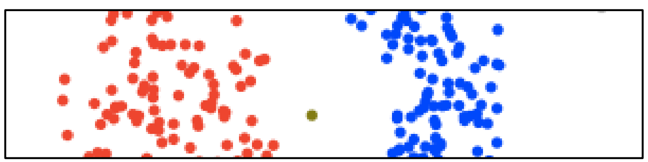
Berkeley
UNIVERSITY OF CALIFORNIA

Nonconvex matrix recovery (Burer & Monteiro 2003)

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

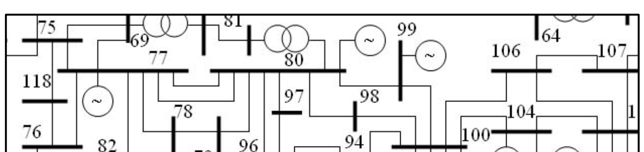
$$\mathcal{A}(X) = [\text{trace}(A_1 X) \quad \cdots \quad \text{trace}(A_m X)]^T$$

Recommendation engines



Cluster analysis

Phase retrieval



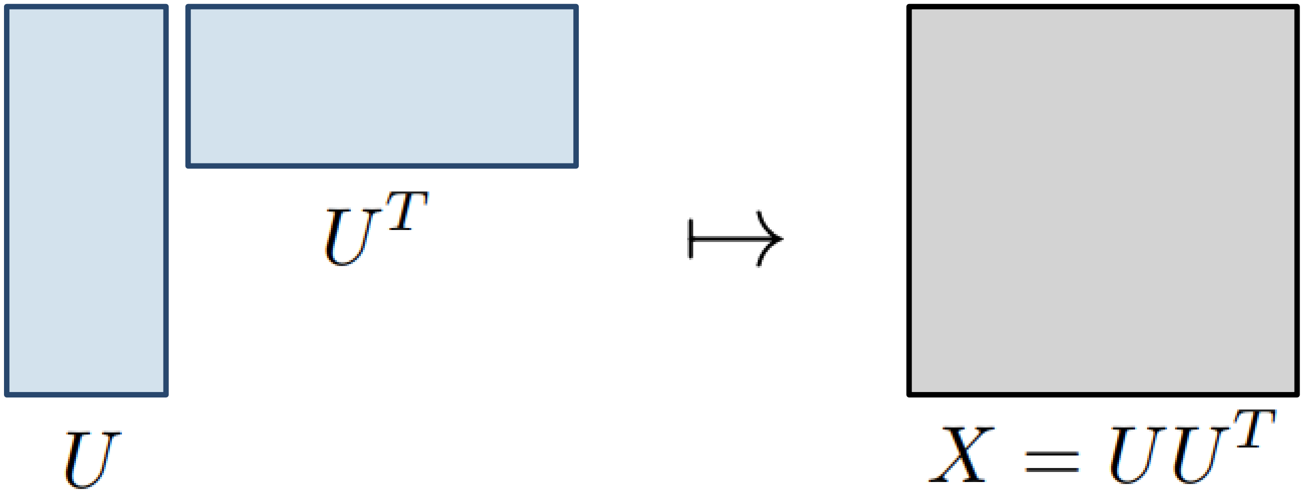
Power system state estimation

Nonconvex matrix recovery (Burer & Monteiro 2003)

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

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

1. Express low-rank matrix as product of factors



movie genres
user preferences

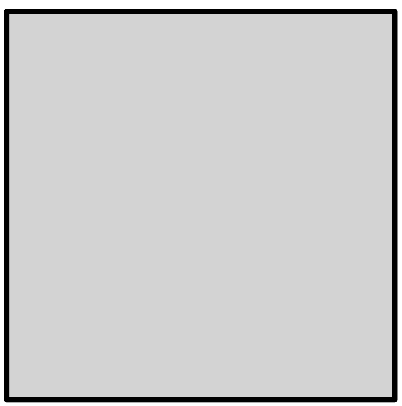
table of movie ratings

Nonconvex matrix recovery (Burer & Monteiro 2003)

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

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

2. Minimize least-squares loss of linear model

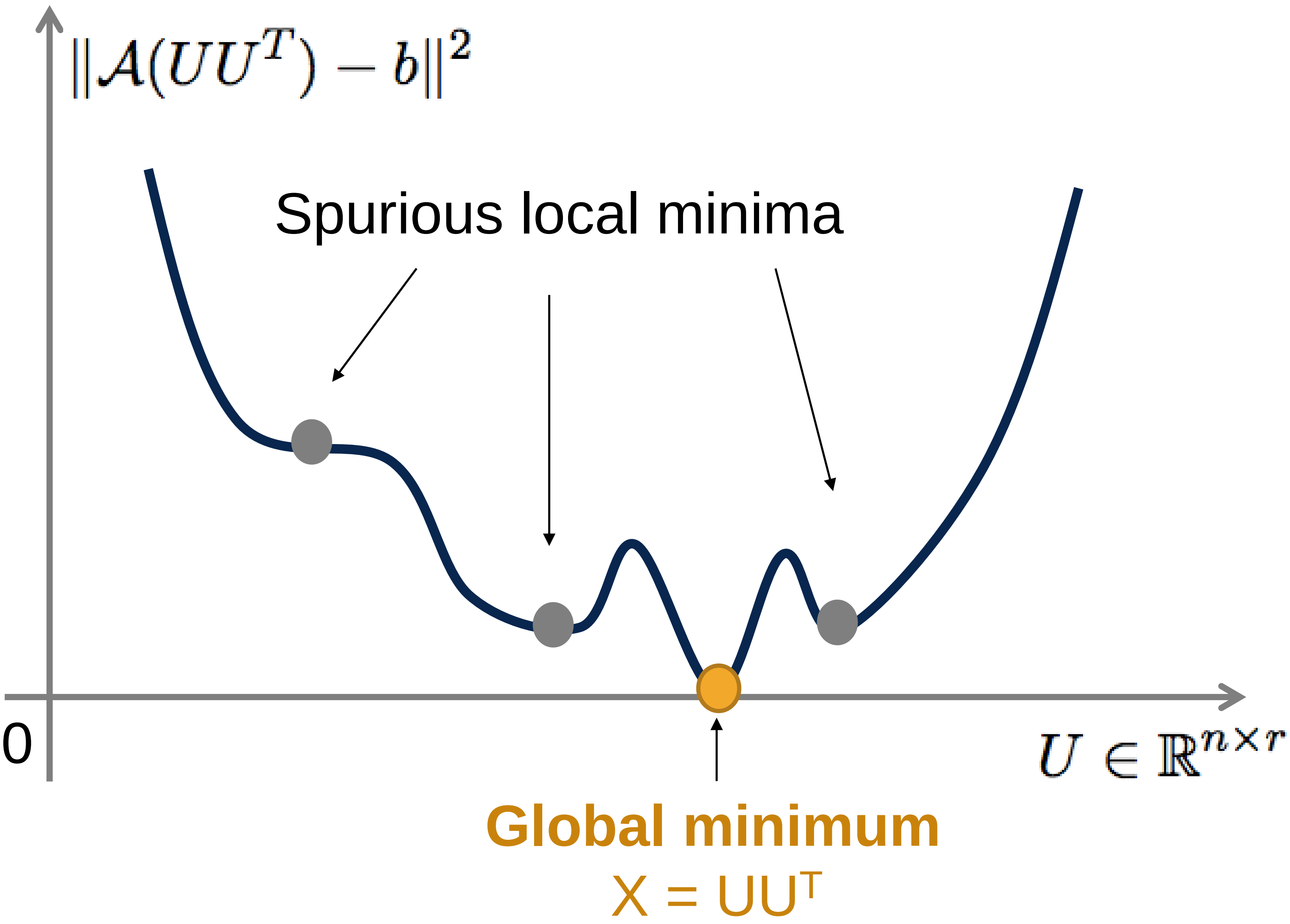


$$X = UU^T \mapsto \left\| \begin{bmatrix} \text{trace}(A_1, X) \\ \text{trace}(A_2, X) \\ \vdots \\ \text{trace}(A_m, X) \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \right\|^2$$

table of movie ratings

specific table elements

known ratings


$\|\mathcal{A}(UU^T) - b\|^2$
Spurious local minima
0
$U \in \mathbb{R}^{n\times r}$
Global minimum
$X = UU^T$

Exact recovery guarantee (Bhojanapalli et al. 2016)

$$\text{minimize } \|\mathcal{A}(UU^T) - b\|^2 \quad \text{over} \quad U \in \mathbb{R}^{n\times r}$$
$$\mathcal{A}(X) = [\text{trace}(A_1 X) \quad \cdots \quad \text{trace}(A_m X)]^T$$

δ-Restricted isometry property (δ-RIP)

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

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

See also (Ge et al. 2017; Li & Tang 2017; Zhu et al. 2017)

Exact recovery guarantee (Bhojanapalli et al. 2016)

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

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

δ-Restricted isometry property (δ-RIP)

$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq (1+\delta)\|X\|_F^2$$

$$\forall \, \text{rank}(X) \leq 2r$$

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

Local search is **guaranteed** to succeed.

See also (Ge et al. 2017; Li & Tang 2017; Zhu et al. 2017)

Exact recovery guarantee (Ge et al. 2016)

When $A_1, A_2, \ldots, A_m$ are sparse

$$\text{minimize } \|\mathcal{A}(UU^T) - b\|^2 \quad + R(U)$$

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

δ-Concentration inequality

$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq (1+\delta)\|X\|_F^2$$

$$\forall \; \text{rank}(X) \leq 2r \text{ and } X \text{ is incoherent}$$

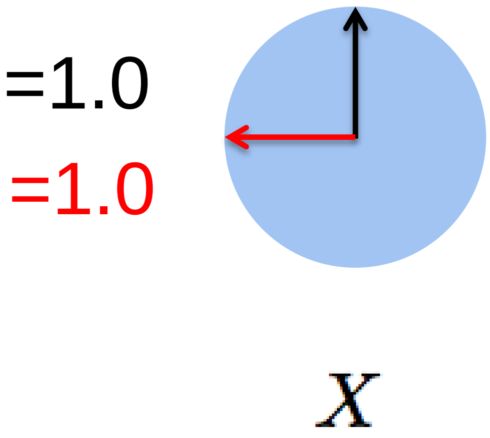
If δ very small, then no spurious local minima.

Similar idea drives many proofs

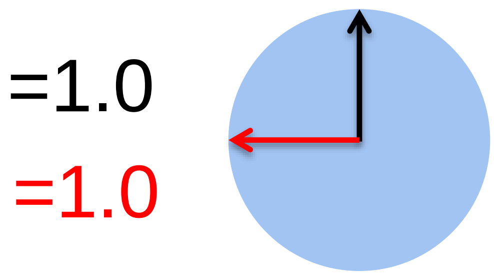
See also (Ge et al. 2015; 2017; Sun et al. 2015; 2016; Park et al. 2017; etc.)

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

If δ < 1/5, then no spurious local min.



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



X

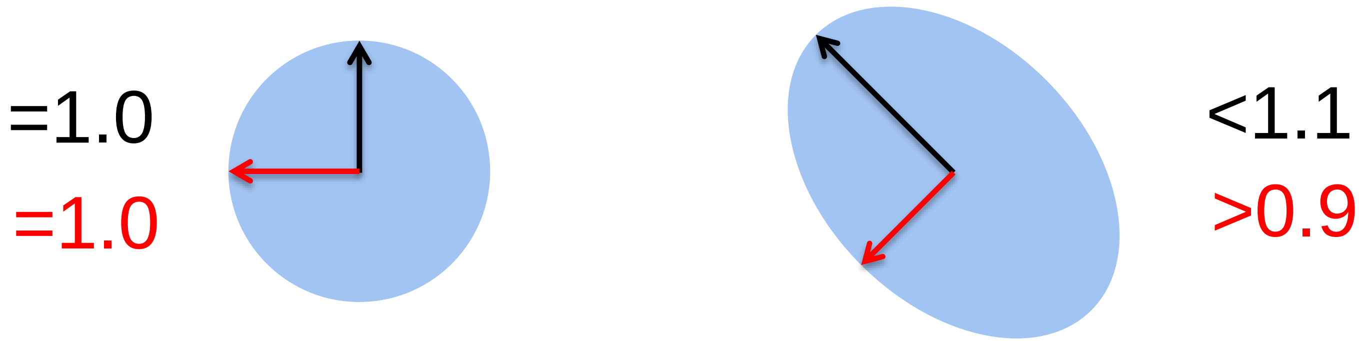
< 1/5

$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq (1+\delta)\|X\|_F^2$$

$$\forall \operatorname{rank}(X) \leq 2r$$

< 1/5

If δ < 1/5, then no spurious local min.



$$X \mapsto \mathcal{A}(X)$$

< 1/5

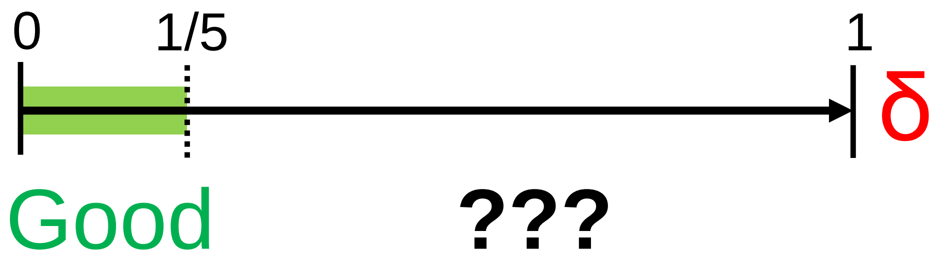
$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq (1+\delta)\|X\|_F^2$$

$$\forall \ \text{rank}(X) \leq 2r$$

< 1/5

Preserve lengths with <10% distortion

If δ < 1/5, then no spurious local min.



Can this be significantly improved?

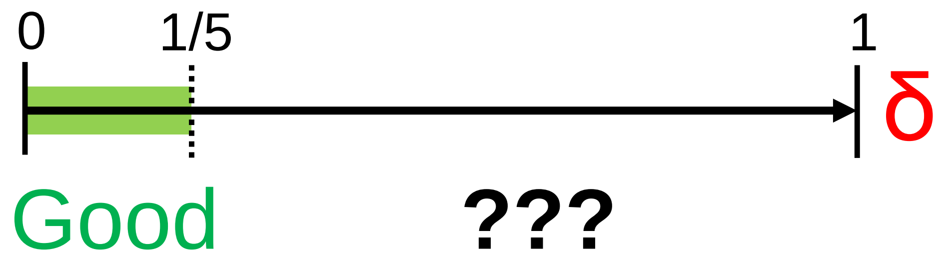
Yes

- Problem **easy**.
- **Agnostic** to algorithm.
- Proof idea is **powerful**.

NO

- Problem **hard**.
- **Specific** to algorithm.
- Proof idea is **limited**.

If δ < 1/5, then no spurious local min.



Can this be significantly improved?

Yes

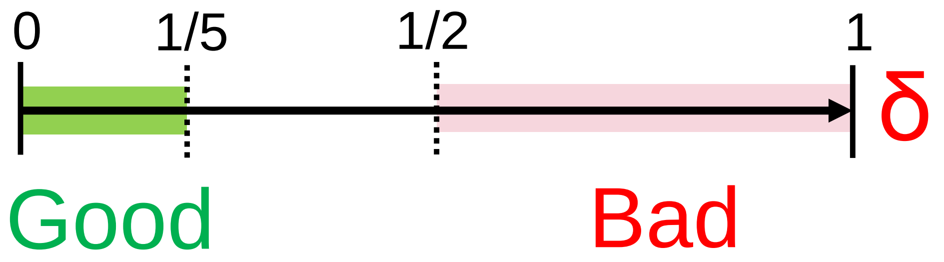
- Problem **easy**.
- **Agnostic** to algorithm.
- Proof idea is **powerful**.

NO

- Problem **hard**.
- **Specific** to algorithm.
- Proof idea is **limited**.

Previous attempts all stuck at 1/5
(Bhojanapalli et al. 2016) (Ge et al. 2017) (Li & Tang 2017) (Zhu et al. 2017) etc.

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



Can this be significantly improved?

NO

- Problem **hard**.
- **Specific** to algorithm.
- Proof idea is **limited**.

Contribution 1.

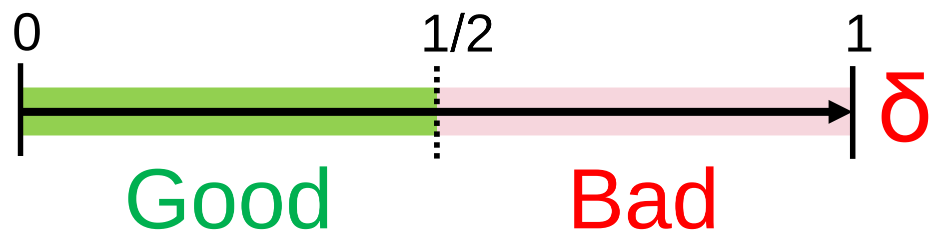
If $\delta \geq 1/2$, many counterexamples.

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Contribution 2.

Let rank $r = 1$.

If $\delta < 1/2$, then no spurious local min.



If $\delta \geq 1/2$, many counterexamples.

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

Counterexample with δ = 1/2

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

Satisfies ½-RIP.

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)
Zhang, Sojoudi, Lavaei, Submitted to *JMLR* (2018)

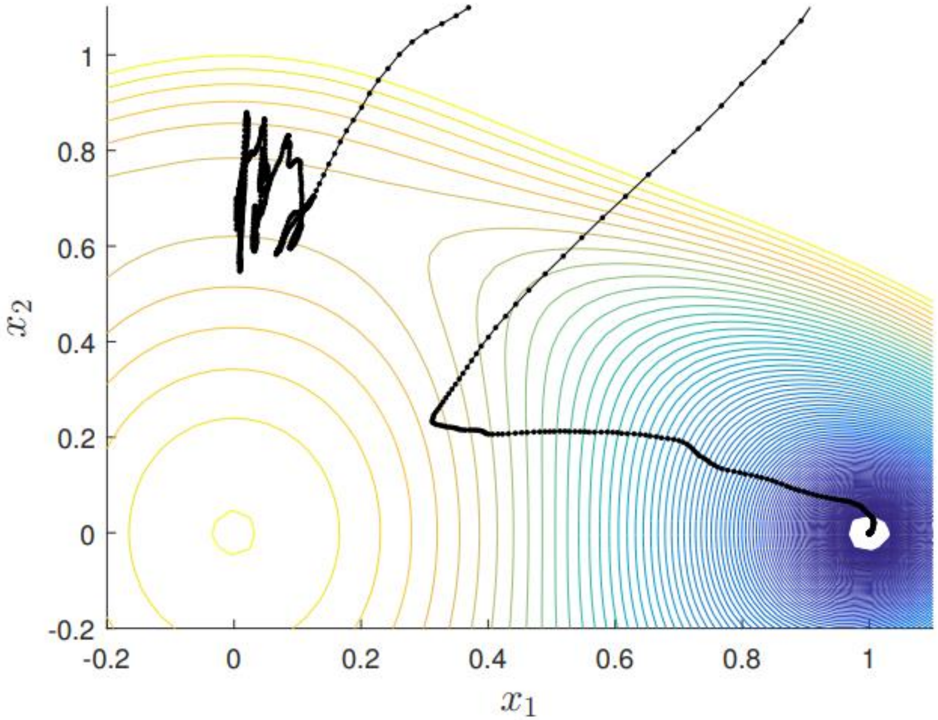
Counterexample with δ = 1/2

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

Satisfies ½-RIP.

Ground truth
z = (1,0)

Spurious local min
x = (0,1/√2)



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)
Zhang, Sojoudi, Lavaei, Submitted to *JMLR* (2018)

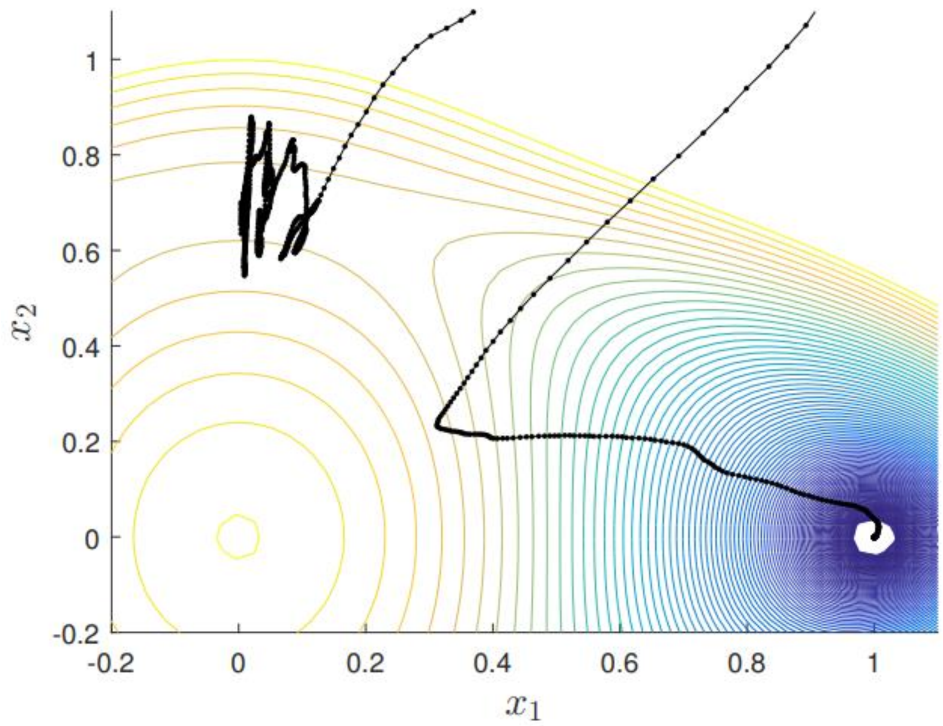
Counterexample with δ = 1/2

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

Satisfies ½-RIP.

Ground truth
$z = (1,0)$

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



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

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)
Zhang, Sojoudi, Lavaei, Submitted to *JMLR* (2018)

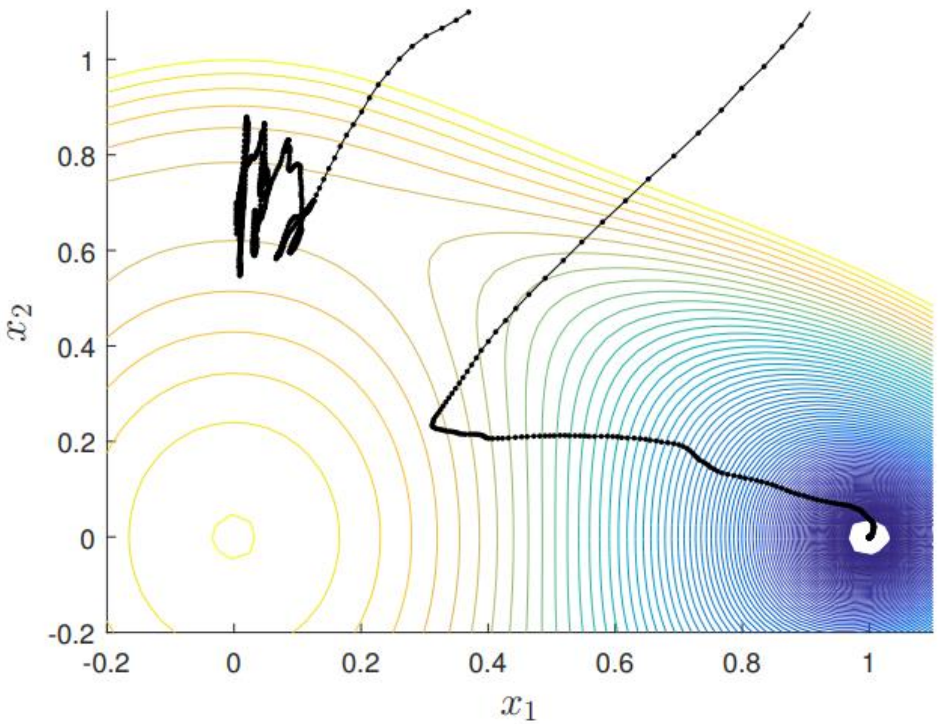
Counterexample with δ = 1/2

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

Satisfies ½-RIP.

Ground truth
z = (1,0)

Spurious local min
x = (0,1/√2)



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

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

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)
Zhang, Sojoudi, Lavaei, Submitted to *JMLR* (2018)

Proof idea. Counterexamples via convex optimization

$$\begin{aligned} \text{find} \quad & f(U) = \|\mathcal{A}(UU^T - ZZ^T)\|^2 \\ \text{such that} \quad & \nabla f(X) = 0, \\ & \nabla^2 f(X) \succeq 0, \\ & \mathcal{A} \text{ satisfies } \delta\text{-RIP}. \end{aligned}$$

Key insight. Relax into a semidefinite program

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Main Result 1. Counterexamples are almost everywhere

Theorem 1 (Zhang, Josz, Sojoudi, Lavaei 2018).
Given x, z not colinear and nonzero,
there exists a counterexample that

- satisfy δ-RIP and $1/2 \le \delta < 1$
- has z as ground truth
- has x as spurious local min.

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Main Result 1. Counterexamples are almost everywhere

Theorem 1 (Zhang, Josz, Sojoudi, Lavaei 2018).
Given x, z not colinear and nonzero,
there exists a counterexample that

- satisfy δ-RIP and $1/2 \le \delta < 1$
- has z as ground truth
- has x as spurious local min.

Take-away.
If δ-RIP with $\delta \ge 1/2$, then expect spurious local minima.

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Conjecture (Zhang, Josz, Sojoudi, Lavaei 2018).
If δ-RIP with δ < 1/2, then no spurious local min.

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Main Result 2. Sharp RIP-based guarantee

Theorem 2 (Zhang, Sojoudi, Lavaei 2018).
If δ-RIP with $\delta < 1/2$ and $r = 1$, then no spurious local min.

Proof for rank-1 case

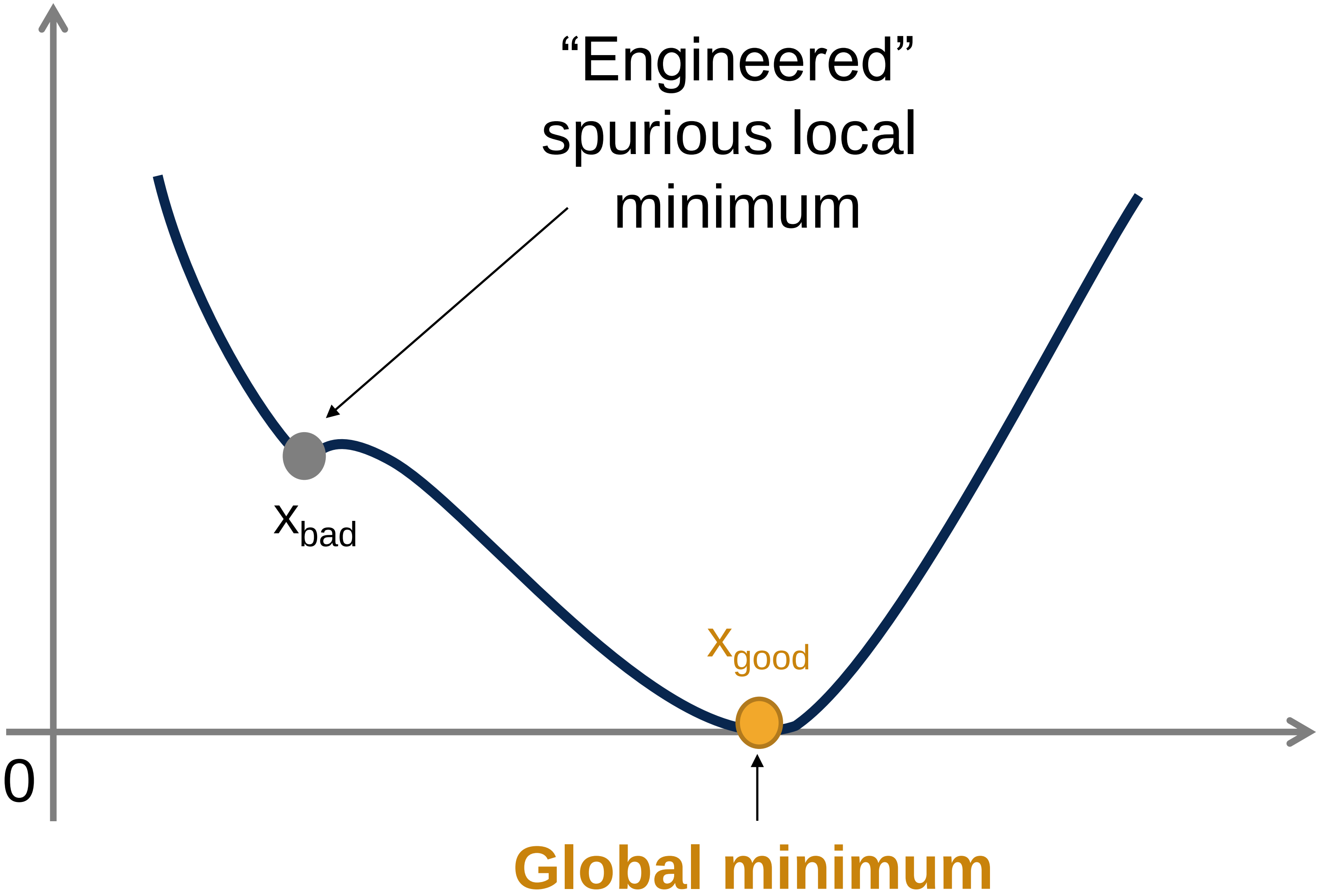
Zhang, Sojoudi, Lavaei. Submitted to *JMLR* (2018)

Theorem 2 (Zhang, Sojoudi, Lavaei 2018).
If δ-RIP with $\delta < 1/2$ and $r = 1$, then no spurious local min.

Ongoing work.
Generalization to rank-r

Practical implications?

δ-RIP with $1/2 \leq \delta < 1$



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

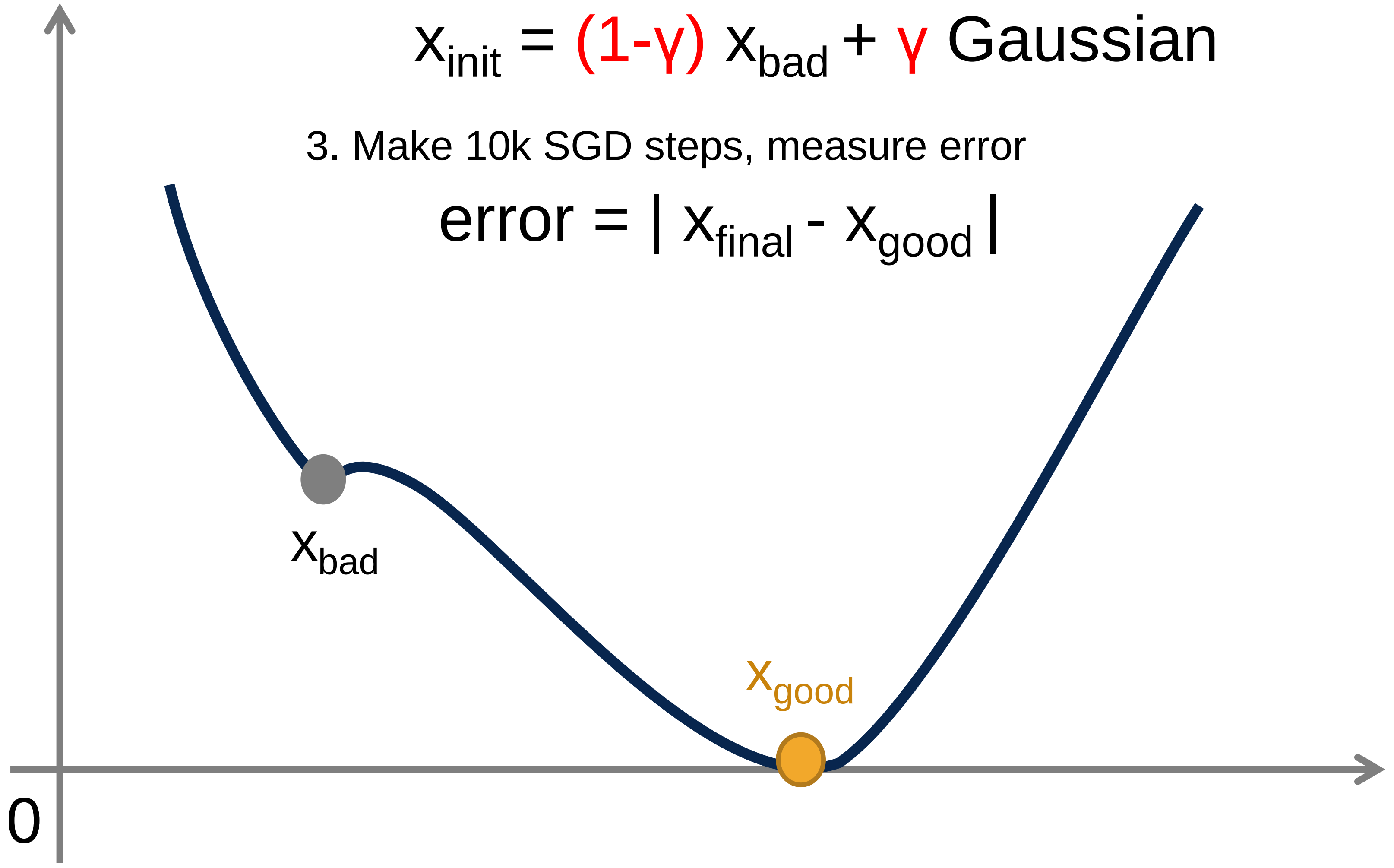
1. Select γ in [0,1]

2. Start SGD at

$$x_{init} = (1-\gamma)\, x_{bad} + \gamma \text{ Gaussian}$$

3. Make 10k SGD steps, measure error

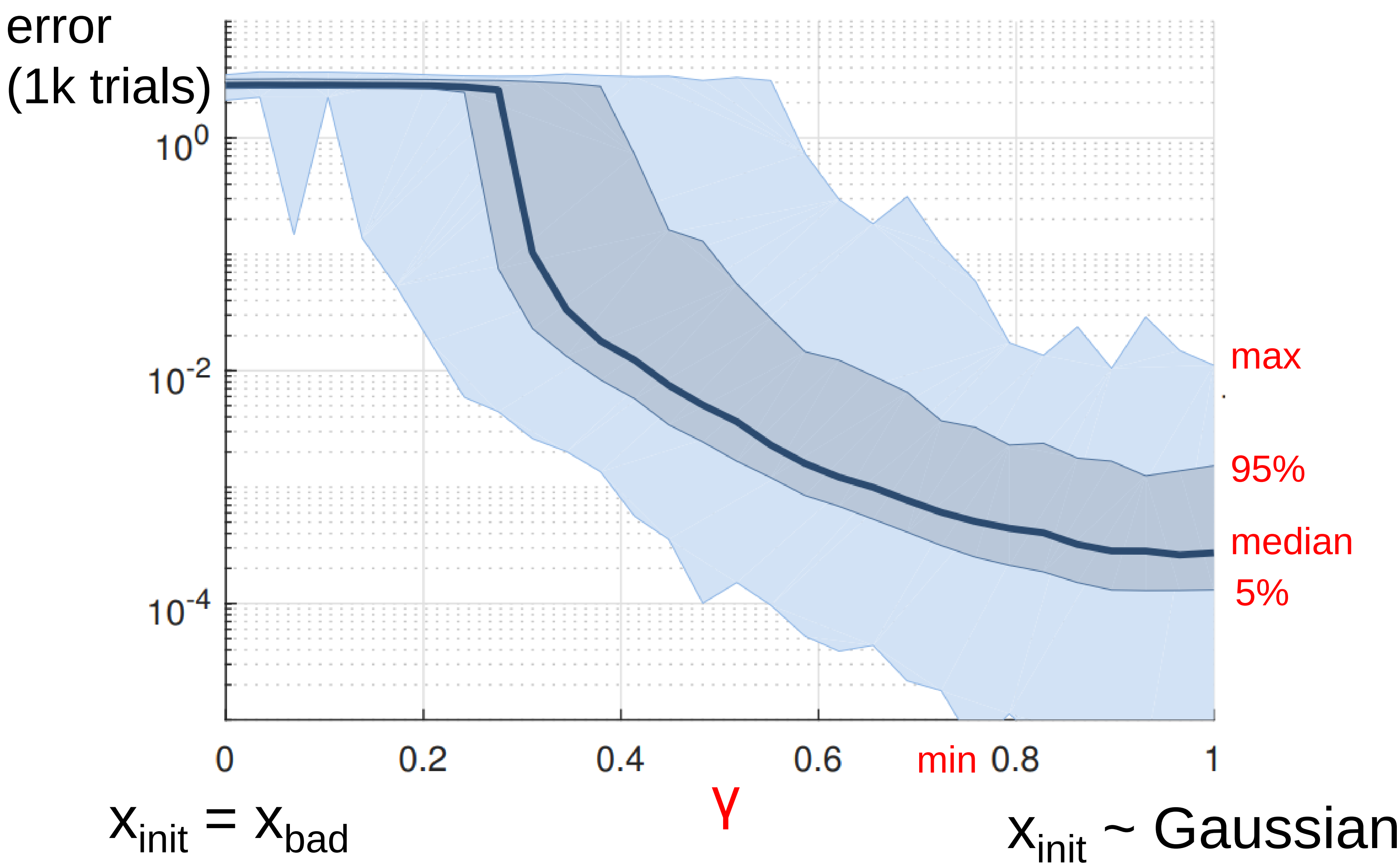
$$\text{error} = |\, x_{final} - x_{good} \,|$$



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Example 1

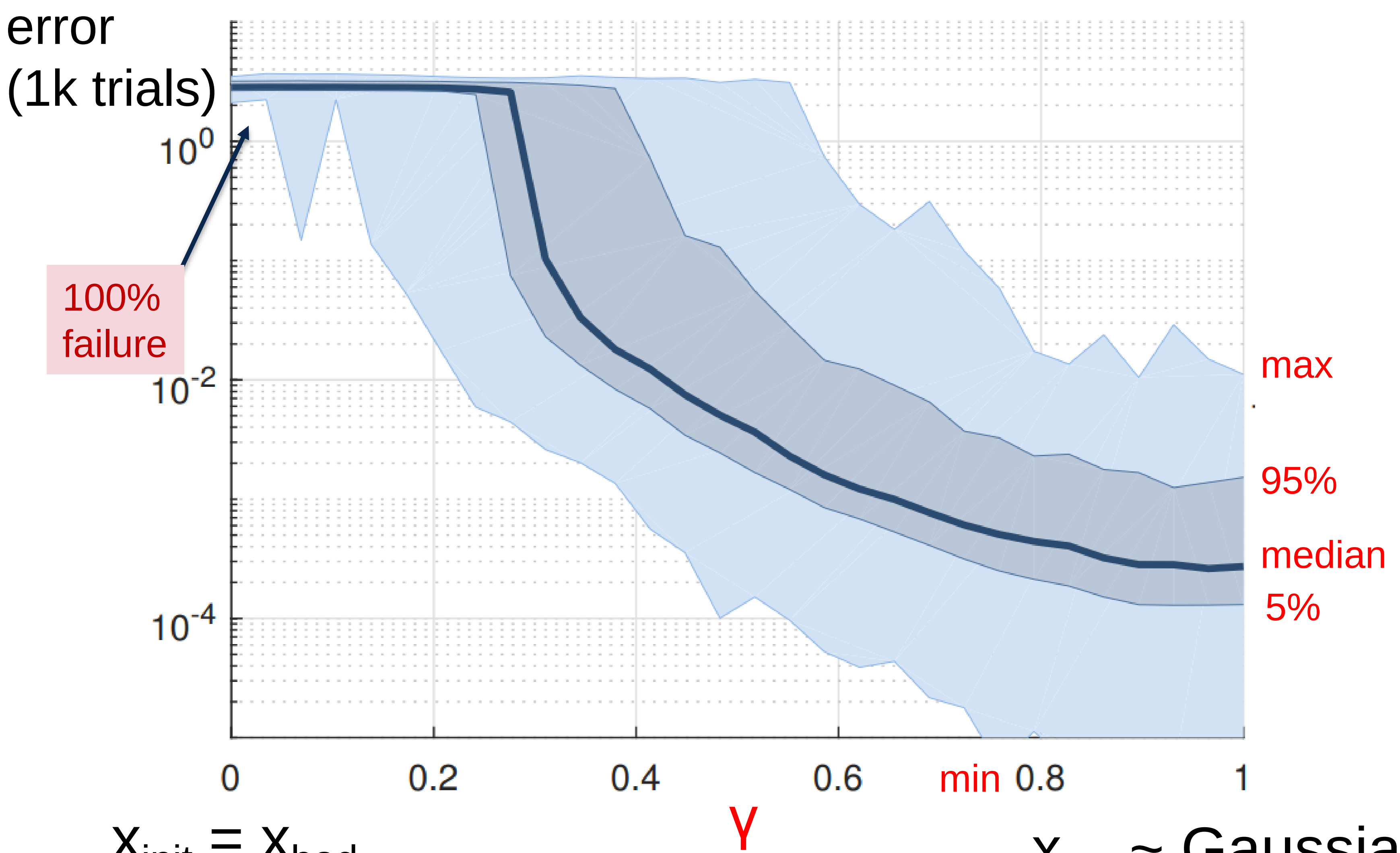
100% success rate



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Example 1

100% success rate

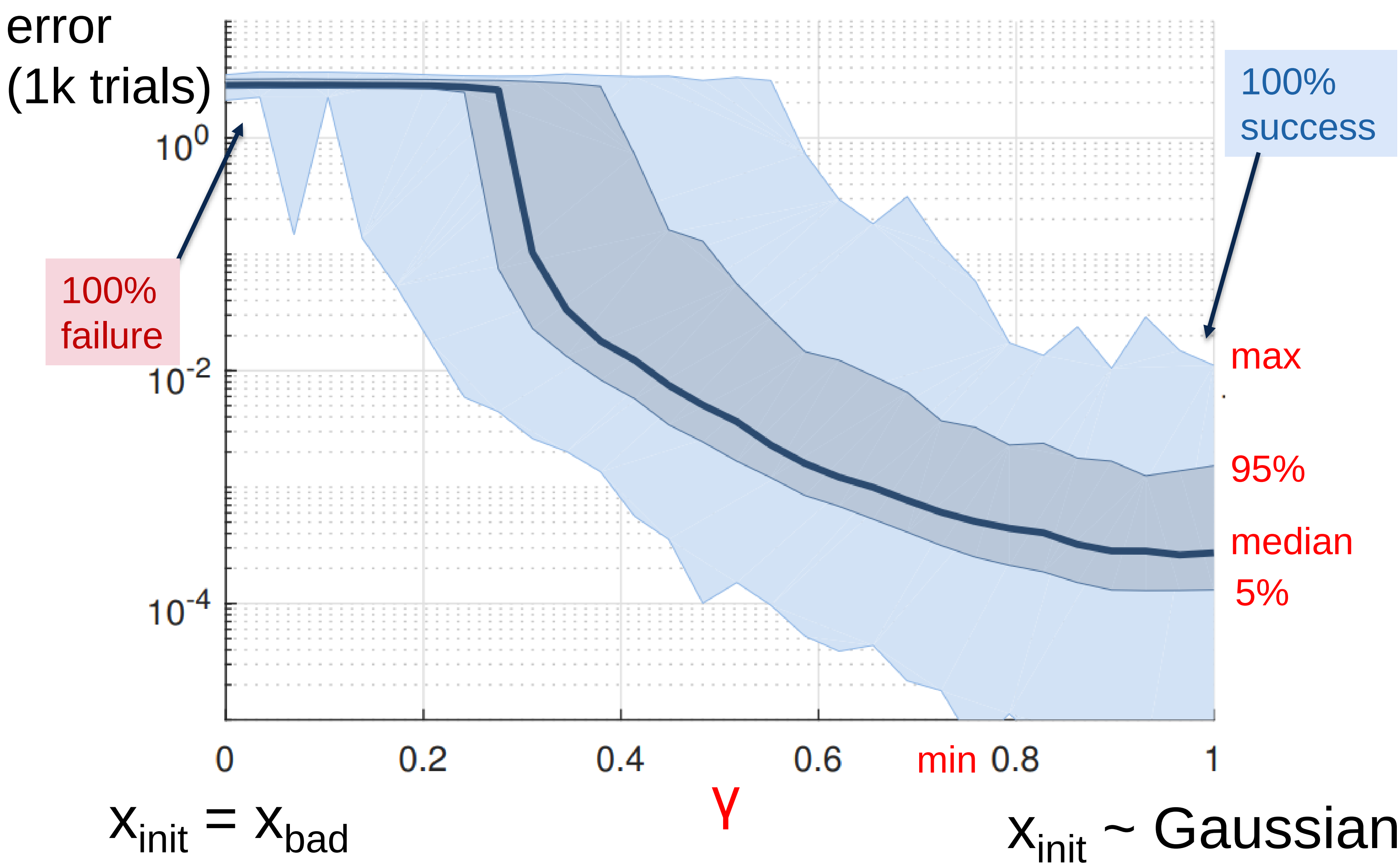


$x_{init} = x_{bad}$ $x_{init} \sim$ Gaussian

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Example 1

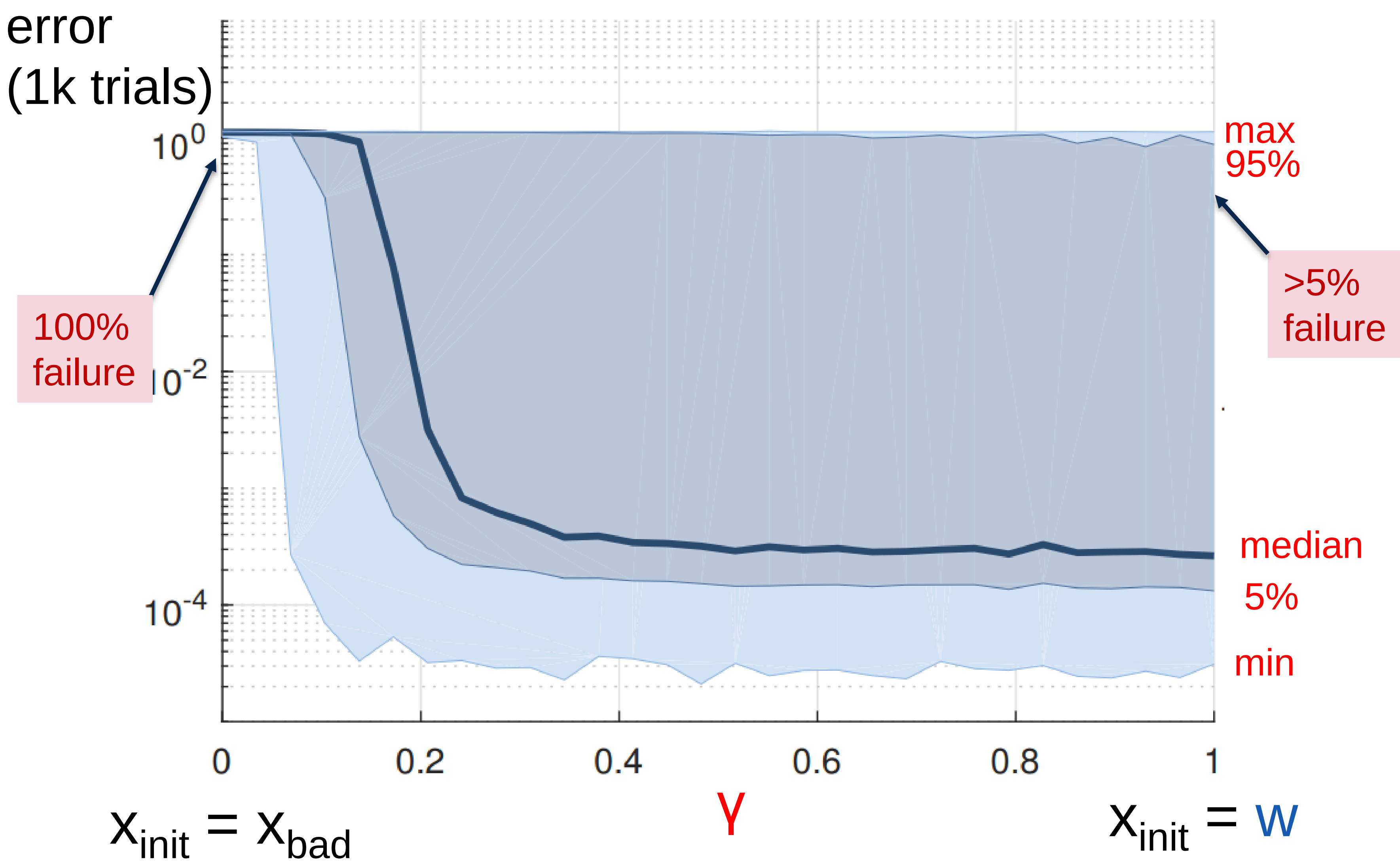
100% success rate



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Example 2

<95% success rate



Zhang, Josz, Sojoudi, Lavaei, *NeurIPS* (2018)

Practical implications?

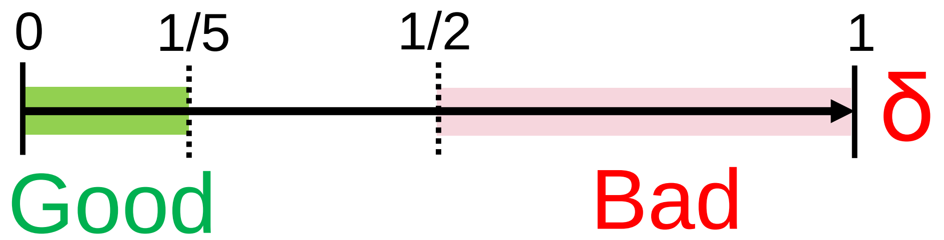
δ-RIP with 1/2 ≤ δ < 1

spurious local min	$\not\Rightarrow$	> 0% failure
100% success	$\not\Rightarrow$	no spurious local min

Limitations of “no spurious local min” guarantees

How Much Restricted Isometry is Needed in Nonconvex Matrix Recovery?

R.Y. Zhang, C. Josz, S. Sojoudi, J. Lavaei, *NeurIPS* (2018)



If $\delta \geq 1/2$, many counterexamples.

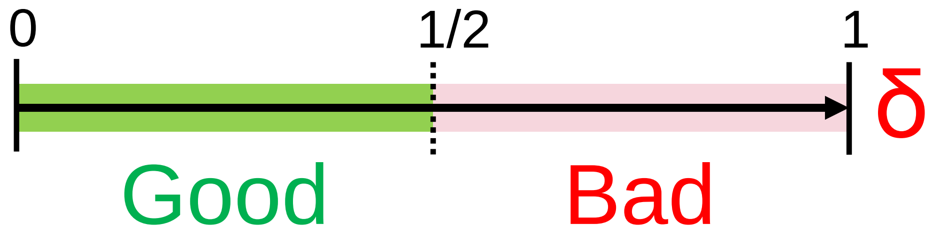
Wed Dec 5th 5 - 7 PM @ Room 210 & 230 AB

Poster #46

How Much Restricted Isometry is Needed in Nonconvex Matrix Recovery?

R.Y. Zhang, C. Josz, S. Sojoudi, J. Lavaei, *NeurIPS* (2018)

If $\delta < 1/2$, then no spurious local min (?)



If $\delta \geq 1/2$, many counterexamples.

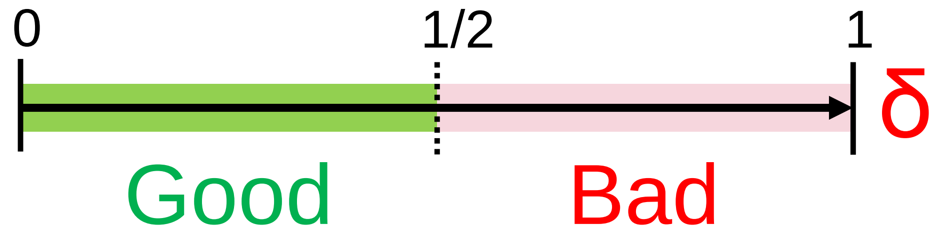
Wed Dec 5th 5 - 7 PM @ Room 210 & 230 AB

Poster #46

How Much Restricted Isometry is Needed in Nonconvex Matrix Recovery?

R.Y. Zhang, C. Josz, S. Sojoudi, J. Lavaei, *NeurIPS* (2018)

If $\delta < 1/2$, then no spurious local min (?)



If $\delta \geq 1/2$, many counterexamples.

Limitations of “no spurious local min” guarantees

Wed Dec 5th 5 - 7 PM @ Room 210 & 230 AB

Poster #46