On the Complexity of Optimization : Curved Spaces and Benign Landscapes

PhD Oral Exam

Candidate: Christopher Criscitiello

Advisor: Nicolas Boumal

EPFL

OPTIM, Chair of Continuous Optimization

July 15, 2025

Outline

Motivation

Part I: Geodesic convexity

Part II: Benign Landscapes

Outline

Motivation

Part I: Geodesic convexity

Part II: Benign Landscapes

Based on:

- "Negative curvature obstructs acceleration for strongly geodesically convex optimization" - C & **Boumal** – COLT'22
- "Curvature and Complexity: Better lower bounds for geodesically convex optimization" - C & **Boumal** – COLT'23
- "Synchronization on circles and spheres with nonlinear interactions" - C, **Rebjock, McRae, Boumal** - under review
- "The sensor network localization problem has benign landscape under mild rank relaxation" - C, **Rebjock, McRae, Boumal** – not yet public

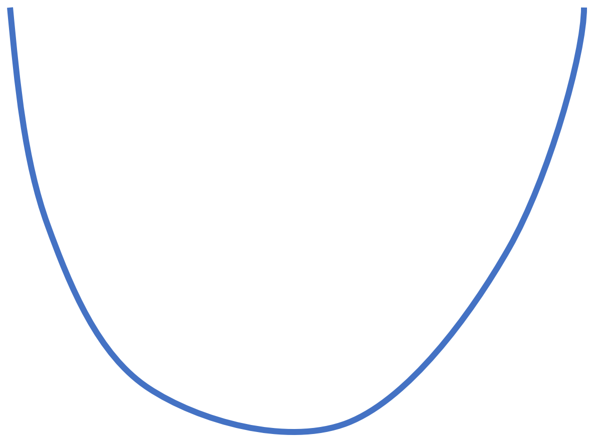
Motivation

$$\min_{x \in \mathcal{M}} f(x)$$

Optimization is fundamental in many areas

Motivation

$$\min_{x \in \mathcal{M}} f(x)$$

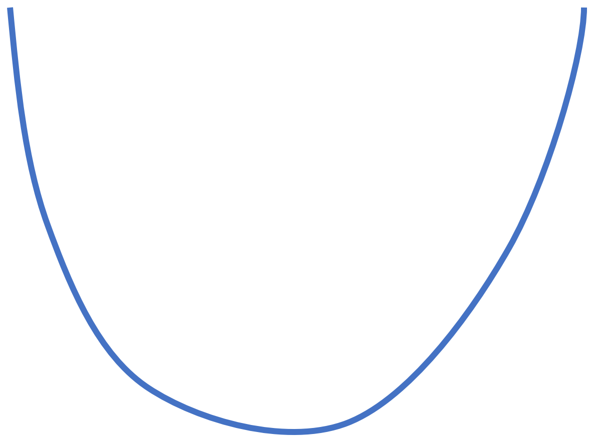


Optimization is fundamental in many areas

Success story: convex optimization in $\mathbb{R}^d$

Motivation

$$\min_{x \in \mathcal{M}} f(x)$$



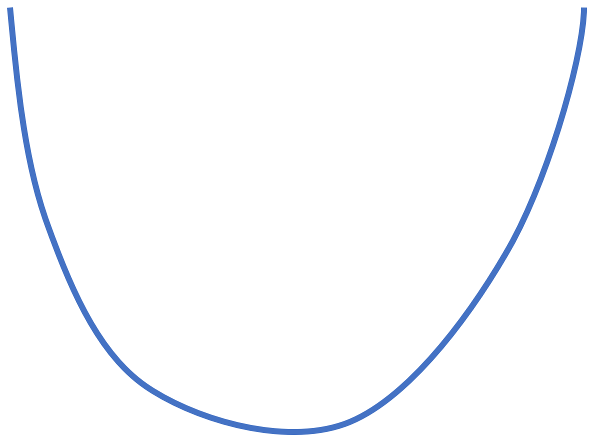
Optimization is fundamental in many areas

Success story: convex optimization in $\mathbb{R}^d$

But many problems are not convex!

Motivation

$$\min_{x \in \mathcal{M}} f(x)$$



Optimization is fundamental in many areas

Success story: convex optimization in $\mathbb{R}^d$

But many problems are not convex!

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

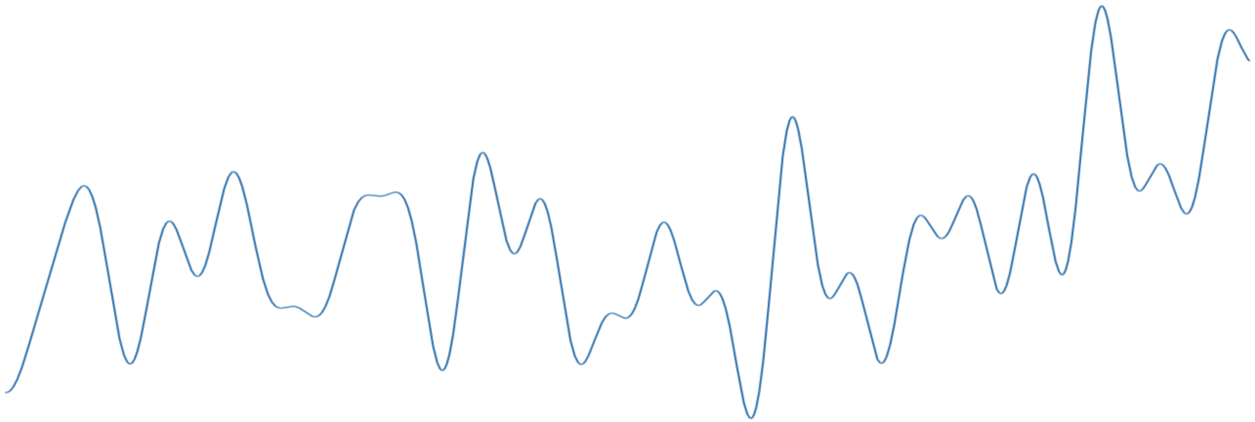
Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?



General nonconvex optimization is hard!

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Q3. Fundamental limits?

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Q3. Fundamental limits?

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Q3. Fundamental limits?

Thesis focuses on 2 special structures:

I. Geodesic convexity – Q2 and Q3

II. Benign landscapes – Q1 (and Q2)

Motivation

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Q3. Fundamental limits?

Thesis focuses on 2 special structures:

I. Geodesic convexity – Q2 and Q3

II. Benign landscapes – Q1 (and Q2)

Part I: Geodesic convexity

Example: robust covariance estimation

Part I: Geodesic convexity

Example: robust covariance estimation

iid samples $y_1, \dots, y_m \in \mathbb{R}^d$ from centered Gaussian with covariance Σ^*

Part I: Geodesic convexity

Example: robust covariance estimation

iid samples $y_1, \ldots, y_m \in \mathbb{R}^d$ from centered Gaussian with covariance Σ^*

Normalize: $x_i = \frac{y_i}{\|y_i\|}$

Target: estimate Σ^* given only $x_1, \ldots, x_m$

Part I: Geodesic convexity

Example: robust covariance estimation

iid samples $y_1, \dots, y_m \in \mathbb{R}^d$ from centered Gaussian with covariance Σ^*

Normalize: $x_i = \frac{y_i}{\|y_i\|}$

Target: estimate Σ^* given only $x_1, \dots, x_m$

MLE, aka Tyler's M-estimator:

$$\hat{\Sigma} = \underset{\Sigma \succ 0,\ \det(\Sigma)=1}{\operatorname{argmin}} \frac{d}{m} \sum_{i=1}^{m} \log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Part I: Geodesic convexity

Example: robust covariance estimation

iid samples $y_1, \dots, y_m \in \mathbb{R}^d$ from centered Gaussian with covariance Σ^*

Normalize: $x_i = \frac{y_i}{\|y_i\|}$

Target: estimate Σ^* given only $x_1, \dots, x_m$

MLE, aka Tyler's M-estimator:

$$\hat{\Sigma} = \operatorname*{argmin}_{\Sigma \succ 0,\ \det(\Sigma)=1} \frac{d}{m} \sum_{i=1}^{m} \log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Nonconvex!

Part I: Geodesic convexity

Example: robust covariance estimation

iid samples $y_1, \ldots, y_m \in \mathbb{R}^d$ from centered Gaussian with covariance Σ^*

Normalize: $x_i = \frac{y_i}{\|y_i\|}$

Target: estimate Σ^* given only $x_1, \ldots, x_m$

MLE, aka Tyler's M-estimator:

$$\hat{\Sigma} = \underset{\Sigma \succ 0,\ \det(\Sigma)=1}{\operatorname{argmin}} \frac{d}{m} \sum_{i=1}^{m} \log\left(x_i^\top \Sigma^{-1} x_i\right)$$

Nonconvex!

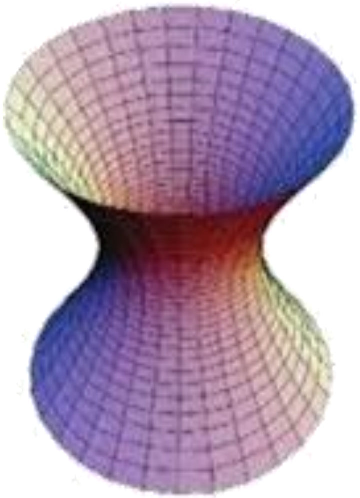
But geodesically convex on a Hadamard manifold!

Sources: Tyler, Weisel & Zhang, Franks & Moitra, etc.

Geodesically convex optimization

$$\min_{x \in D \subset \mathcal{M}} f(x)$$

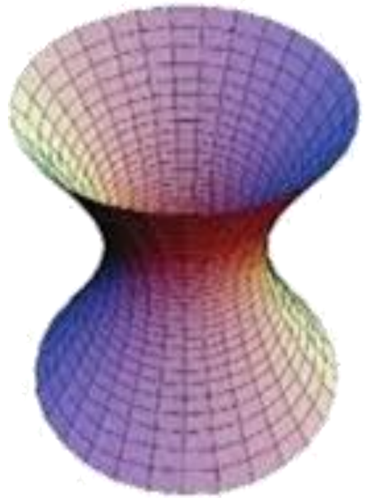
Geodesically convex optimization



$$\min_{x \in D \subset \mathcal{M}} f(x)$$

D is a g-convex subset of a Riemannian manifold $\mathcal{M}$:

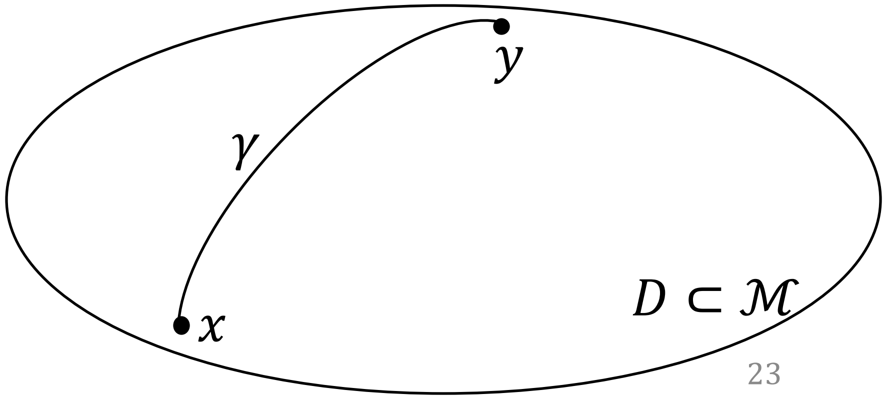
Geodesically convex optimization



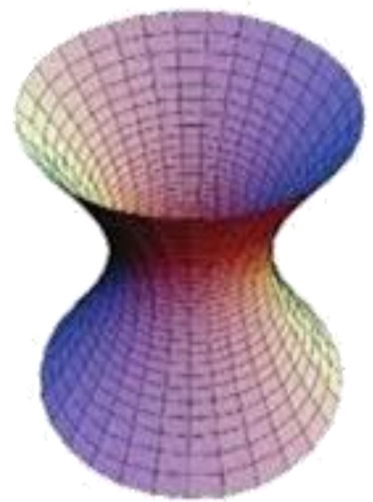
$$\min_{x \in D \subset \mathcal{M}} f(x)$$

D is a g-convex subset of a Riemannian manifold $\mathcal{M}$:

For each $x, y \in D$, there is a unique minimizing geodesic $t \mapsto \gamma(t)$ contained in D, connecting x, y.



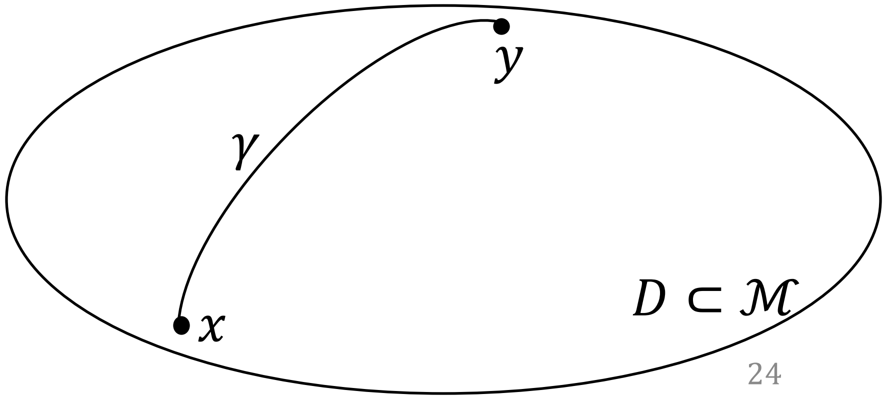
Geodesically convex optimization



$$\min_{x \in D \subset \mathcal{M}} f(x)$$

D is a g-convex subset of a Riemannian manifold $\mathcal{M}$:

For each $x, y \in D$, there is a unique minimizing geodesic $t \mapsto \gamma(t)$ contained in D, connecting x, y.



Cost f is μ-strongly g-convex:

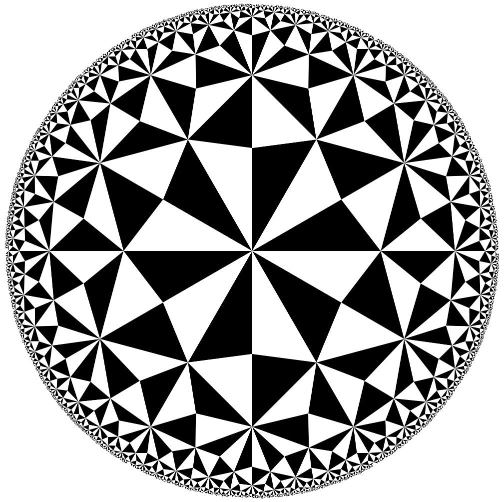
$$t \mapsto f(\gamma(t))$$

is μ-strongly convex for any geodesic γ in D.

Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Complexity of g-convex optim



Our focus: What is the query complexity of g-convex optimization?

Computational task:

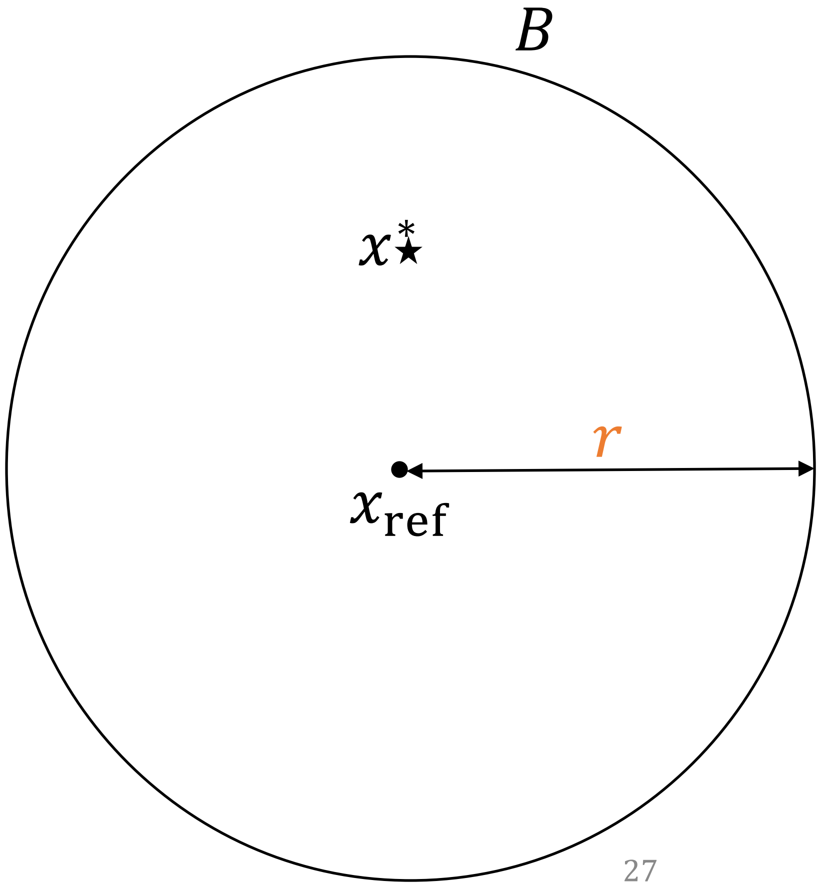
- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$

Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Computational task:

- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$
- f has a unique minimizer x^* in $B = B(x_{\mathrm{ref}}, r)$

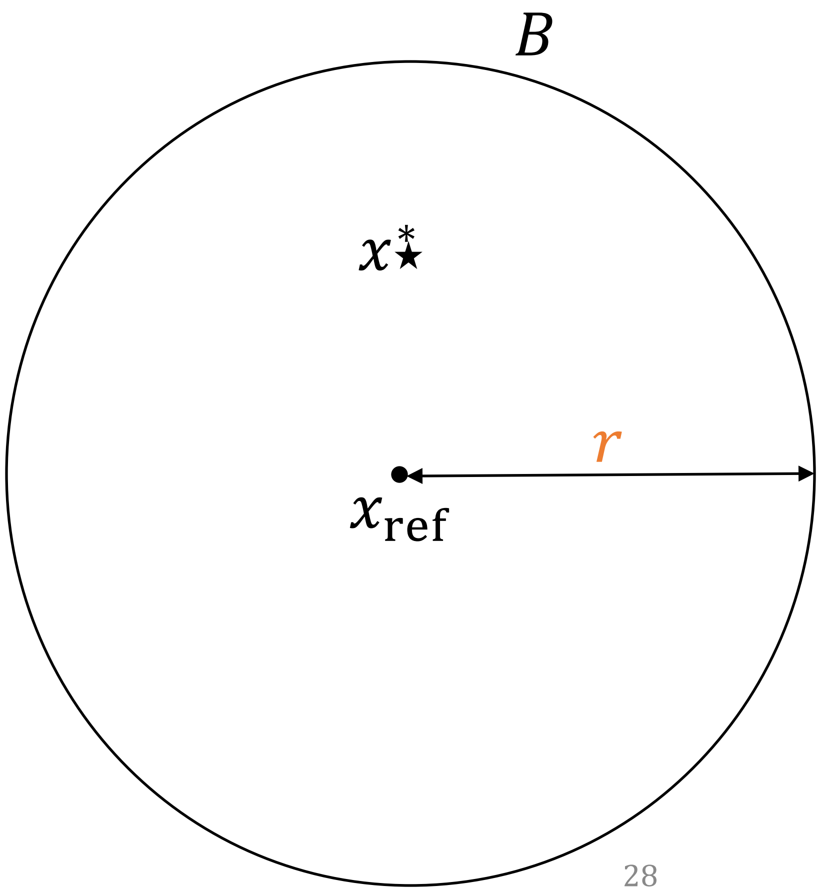


Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Computational task:

- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$
- f has a unique minimizer x^* in $B = B(x_{\text{ref}}, r)$
- f has some regularity (e.g., 1-Lipschitz/1-smooth)



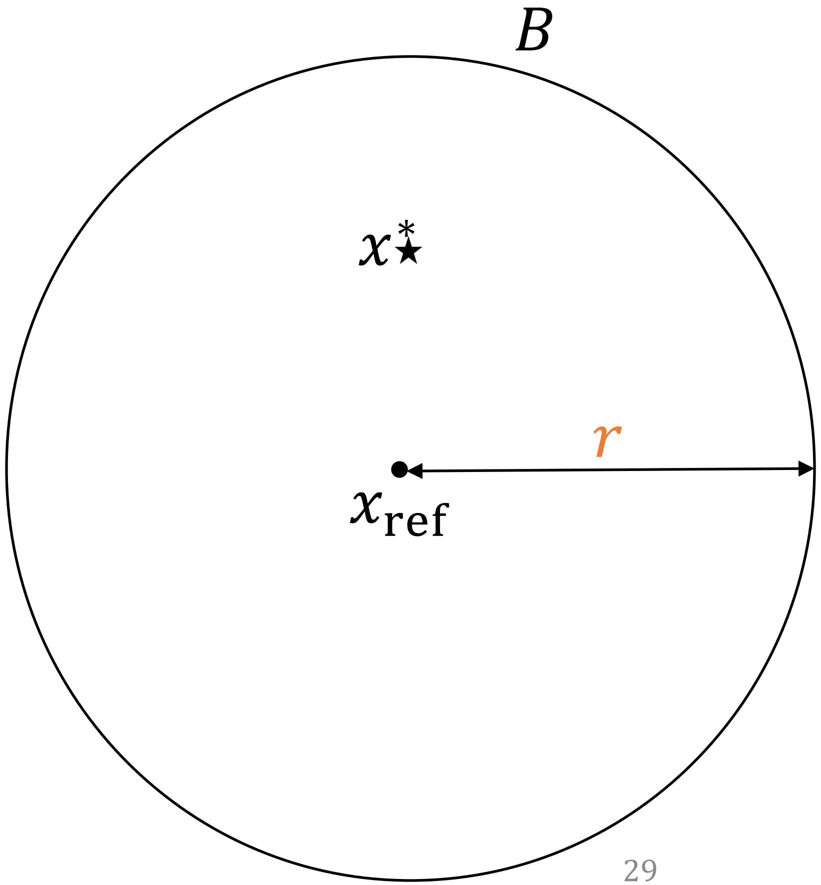
Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Computational task:

- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$
- f has a unique minimizer x^* in $B = B(x_{\mathrm{ref}}, r)$
- f has some regularity (e.g., 1-Lipschitz/1-smooth)

You can query an oracle at x to get $f(x), \nabla f(x)$



Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Computational task:

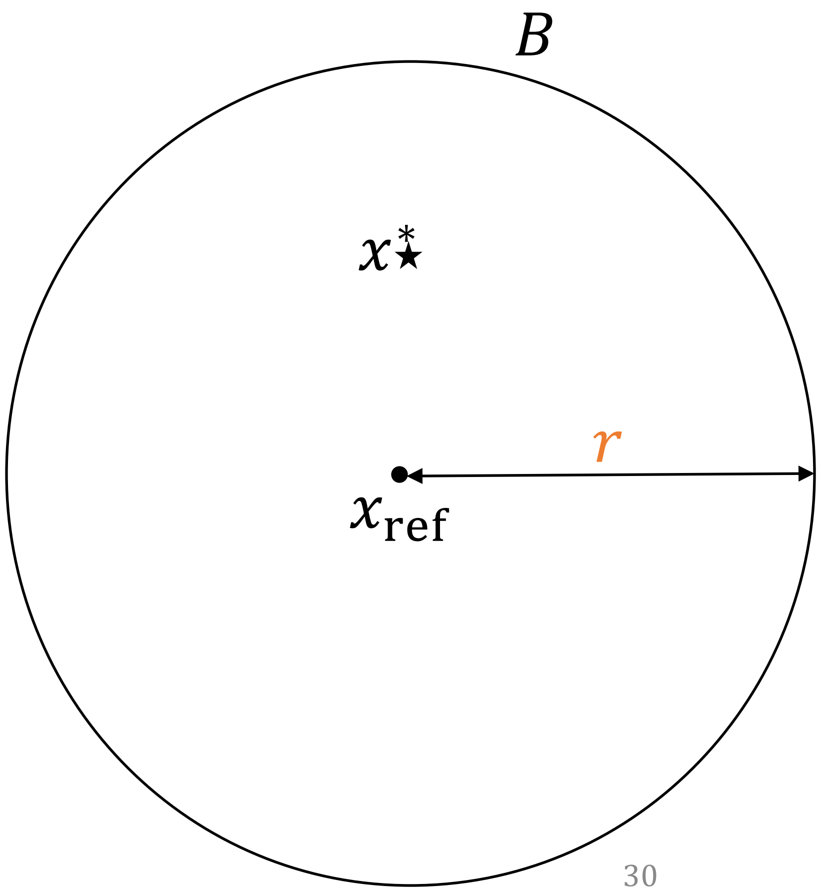
- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$
- f has a unique minimizer x^* in $B = B(x_{\mathrm{ref}}, r)$
- f has some regularity (e.g., 1-Lipschitz/1-smooth)

You can query an oracle at x to get $f(x), \nabla f(x)$

Task: find a point x with accuracy ϵ:

$$f(x) - f^* \leq \epsilon$$

Least number of oracle queries necessary?



Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Computational task:

- $\mathcal{M}$ is hyperbolic space of constant curvature $K < 0$
- f has a unique minimizer x^* in $B = B(x_{\mathrm{ref}}, r)$
- f has some regularity (e.g., 1-Lipschitz/1-smooth)

Complexity (i.e., min # of oracle queries) depends on two parameters:

- Accuracy ϵ
- Curvature term $\zeta := 1 + r\sqrt{|K|}$

Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Complexity (i.e., min # of oracle calls) depends on two parameters:

- Accuracy ϵ
- Curvature term $\zeta := 1 + r\sqrt{|K|}$

Main result: At least $\widetilde{\Omega}(\zeta)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\text{ref}}) - f^*\right)$$

Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Complexity (i.e., min # of oracle calls) depends on two parameters:

- Accuracy ϵ
- Curvature term $\zeta := 1 + r\sqrt{|K|}$

Main result: At least $\widetilde{\Omega}(\zeta)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\text{ref}}) - f^*\right)$$

Builds on Hamilton & Moitra '20

Complexity of g-convex optim

Our focus: What is the query complexity of g-convex optimization?

Complexity (i.e., min # of oracle calls) depends on two parameters:

- Accuracy ϵ
- Curvature term $\zeta := 1 + r\sqrt{|K|}$

Main result: At least $\widetilde{\Omega}(\zeta)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\mathrm{ref}}) - f^*)$$

Builds on Hamilton & Moitra '20

Why?

Why important? Consequences?

Complexity of g-convex optim

Main result: At least $\tilde{\Omega}(\zeta) = \tilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why?

Complexity of g-convex optim

Main result: At least $\tilde{\Omega}(\zeta) = \tilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\mathrm{ref}}) - f^*)$$

Why?

Geodesic ball volumes grow much faster in negatively curved spaces than flat spaces.

Complexity of g-convex optim

Main result: At least $\tilde{\Omega}(\zeta) = \tilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why?

Geodesic ball volumes grow much faster in negatively curved spaces than flat spaces.

	$K = 0$	$K < 0$
Constant curvature:	Euclidean space $\mathbb{R}^d$	Hyperbolic space $\mathbb{H}^d$ $(K = -1)$
Volume of ball of radius r:	$\mathrm{Vol} \sim r^d = e^{d\log(r)}$	$\mathrm{Vol} \sim e^{dr}$

Complexity of g-convex optim

Main result: At least $\tilde{\Omega}(\zeta) = \tilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\mathrm{ref}}) - f^*\right)$$

Why?

Geodesic ball volumes grow much faster in negatively curved spaces than flat spaces.

Curvature measures how quickly geodesics diverge

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\mathrm{ref}}) - f^*\right)$$

Why?

Geodesic ball volumes grow much faster in negatively curved spaces than flat spaces.

Curvature measures how quickly geodesics diverge

In negatively curved spaces, geodesics diverge rapidly, and so a small error can blow up quickly

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why important? Consequences?

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\mathrm{ref}}) - f^*\right)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\text{ref}}) - f^*\right)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\mathrm{ref}}) - f^*\right)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces

Riem GD has complexity $O\left(\frac{1}{\epsilon}\right)$, which seems to match convex counterpart

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces

Riem GD has complexity $O\left(\frac{1}{\epsilon}\right)$, which seems to match convex counterpart

Is there a Riem acceleration scheme with complexity $O\left(\frac{1}{\sqrt{\epsilon}}\right)$? No

Zhang & Sra; Ahn & Sra; Alimisis, Orvieto, Lucchi; Martinez-Rubio; Kim & Yang; etc.

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces
3. Scaling problems

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}\left(f(x_{\text{ref}}) - f^*\right)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces
3. Scaling problems
   - Encompass range of applications in stats, quantum, TCS

Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson '21

Complexity of g-convex optim

Main result: At least $\widetilde{\Omega}(\zeta) = \widetilde{\Omega}\left(r\sqrt{|K|}\right)$ oracle calls are needed to find x such that

$$f(x) - f^* \leq \frac{1}{100}(f(x_{\text{ref}}) - f^*)$$

Why important? Consequences?

1. Explains why all known upper bounds worsen with curvature (e.g., Zhang & Sra '16)
2. Rules out *full* Nesterov acceleration in negatively curved spaces
3. Scaling problems
   - Encompass range of applications in stats, quantum, TCS
   - For tensor scaling (and others), radius r is exponentially large in input size (Franks & Reichenbach'21) – so lower bound points to difficulty of finding poly-time algos

Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson '21

Part I: Geodesic convexity

Much more in thesis!

Part I: Geodesic convexity

Much more in thesis!

- Bounds depending on both ϵ and ζ.
- Other problem classes.

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d)$	Center of gravity (Criscitiello, et al. '23)
(P2) Lipschitz	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta^2 \epsilon^2}\right)$	$O\left(\frac{\zeta}{\epsilon^2}\right)$	Subgradient descent (Zhang & Sra'16)
(P3) Smooth	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta\sqrt{\epsilon}}\right)$	$\tilde{O}\left(\sqrt{\zeta/\epsilon}\right)$	RNAG-C (Kim & Yang'22 / Martinez-Rubio et al.'22)
(P4) Smooth, strongly g-convex	$\tilde{\Omega}(\zeta + \sqrt{\kappa})$	$\tilde{O}(\sqrt{\zeta\kappa})$	RNAG-SC (Kim & Yang'22 / Martinez-Rubio et al.'22)

Cutting-planes game	$\tilde{\Omega}(\zeta d)$	$O(\zeta d)$	Center of gravity (Criscitiello, et al. '23)

$\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size)

Part I: Geodesic convexity

Much more in thesis!

- Bounds depending on both ϵ and ζ.
- Other problem classes.

Open question:

Matching upper and lower bounds?

g-convex setting	Lower bound	Upper bound	Algorithm
(P1) Lipschitz, lo-dim	$\tilde{\Omega}(\zeta + d)$	$O(\zeta d)$	Center of gravity (Criscitiello, et al. '23)
(P2) Lipschitz	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta^2 \epsilon^2}\right)$	$O\left(\frac{\zeta}{\epsilon^2}\right)$	Subgradient descent (Zhang & Sra'16)
(P3) Smooth	$\tilde{\Omega}\left(\zeta + \frac{1}{\zeta \sqrt{\epsilon}}\right)$	$\tilde{O}\left(\sqrt{\zeta / \epsilon}\right)$	RNAG-C (Kim & Yang'22 / Martinez-Rubio et al.'22)
(P4) Smooth, strongly g-convex	$\tilde{\Omega}(\zeta + \sqrt{\kappa})$	$\tilde{O}(\sqrt{\zeta \kappa})$	RNAG-SC (Kim & Yang'22 / Martinez-Rubio et al.'22)
Cutting-planes game	$\tilde{\Omega}(\zeta d)$	$O(\zeta d)$	Center of gravity (Criscitiello, et al. '23)

$\Omega\left(\frac{\zeta}{\epsilon^2}\right)$ lower bound for subgradient descent (with Polyak step size)

Part II: Benign landscapes

Goal of thesis: beyond convex optimization in $\mathbb{R}^d$?

Q1. Special structure in nonconvex problems?

Q2. Algorithms and how much computational resources?

Q3. Fundamental limits?

Thesis focuses on 2 special structures:

I. Geodesic convexity – Q2 and Q3

II. Benign landscapes – Q1 (and Q2)

Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Some problems are g-convex, but many are not!

Need a more general **structure!**

Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

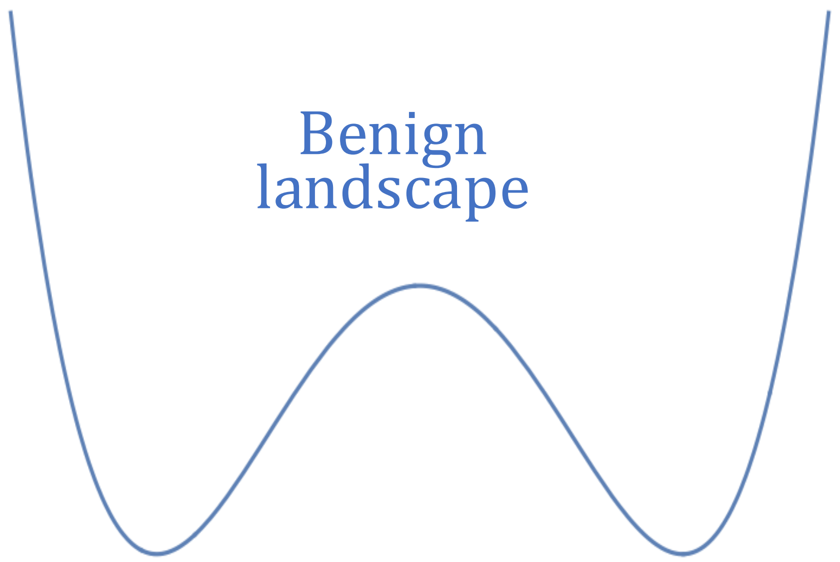
$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

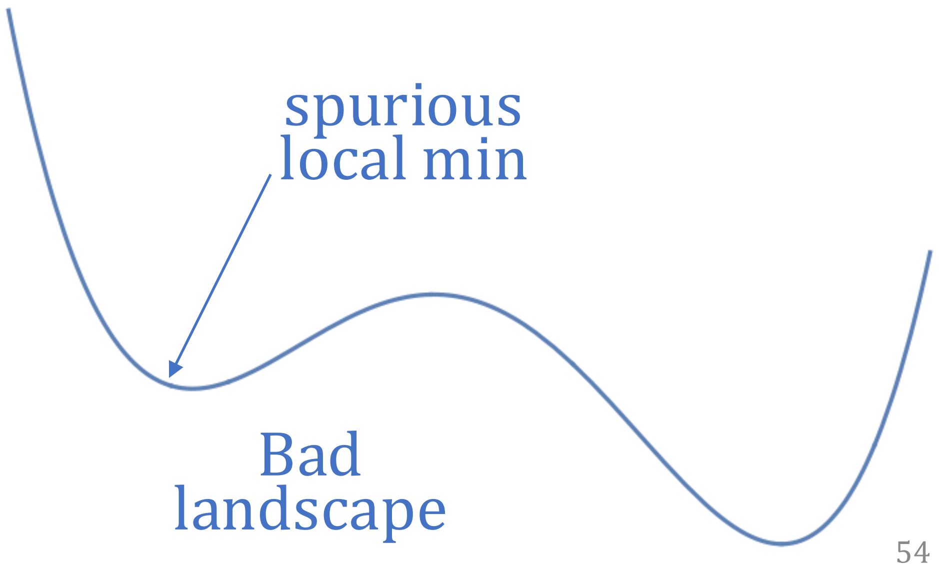
Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min





Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

Why useful?

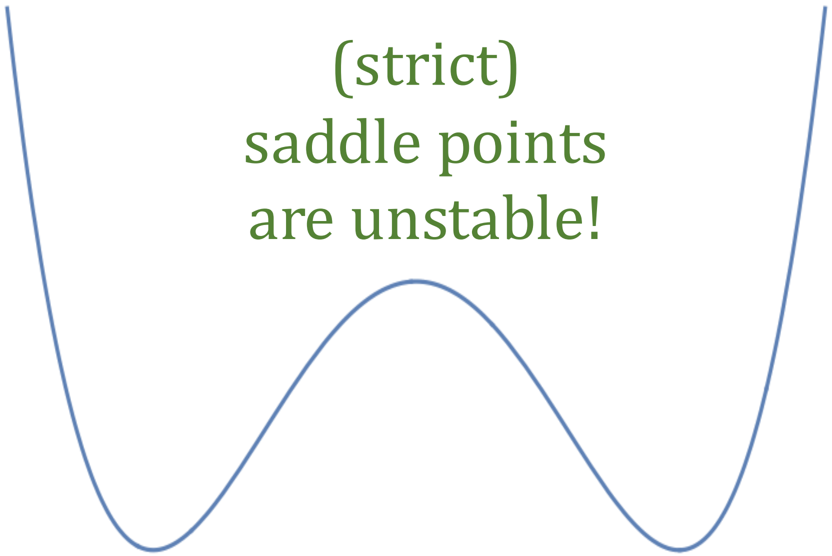
Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

Why useful?



(strict)
saddle points
are unstable!

Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

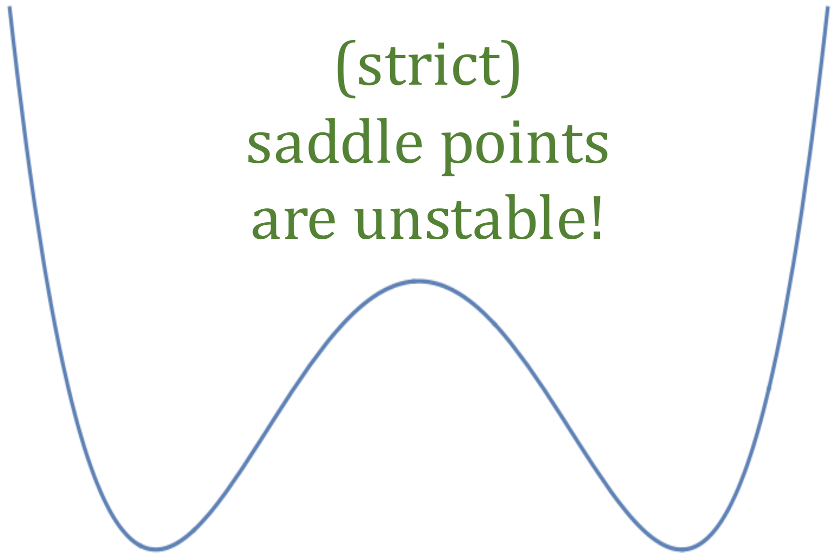
$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

Why useful?

Local algorithm (GD, TR, …)

converges to 2-critical point (w/ prob 1)

(strict)
saddle points
are unstable!



Benign landscape

$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

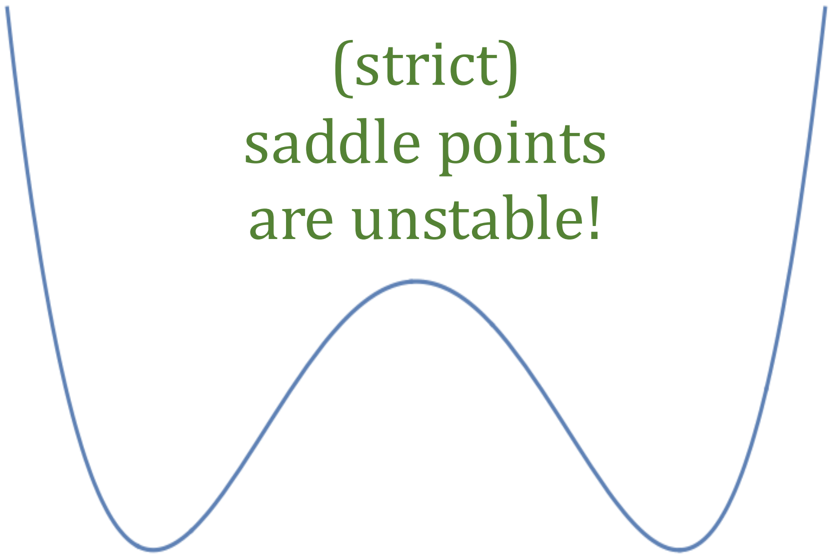
$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

Why useful?

Local algorithm (GD, TR, …)

converges to 2-critical point (w/ prob 1)

Stable manifold theorems
+
Łojasiewicz theorem



(strict)
saddle points
are unstable!

Benign landscape

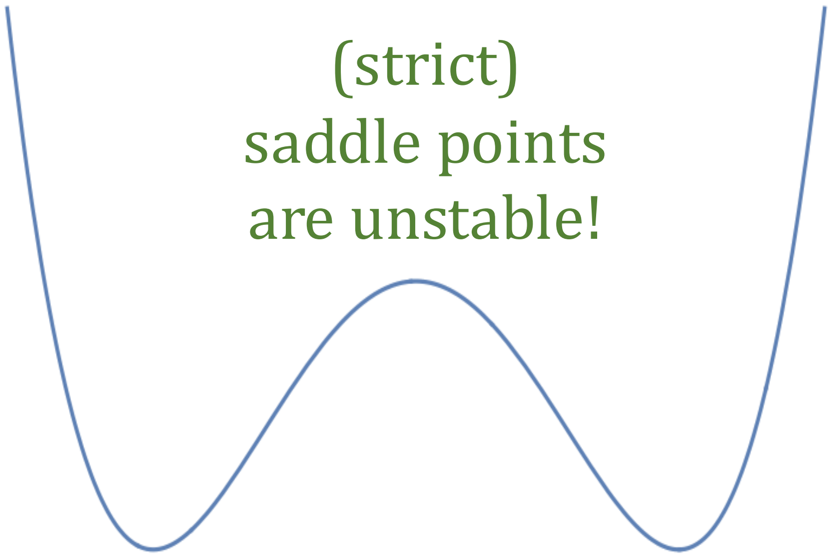
$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

Part II of thesis:
Understand when several problems have benign landscape. With an eye towards new tools.

(strict)
saddle points
are unstable!



Benign landscape

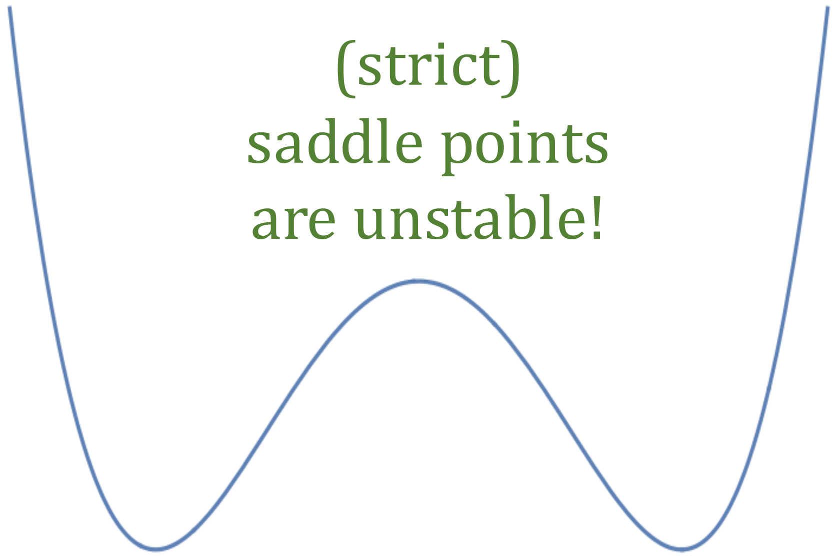
$$\min_{x \in \mathbb{R}^d} f(x)$$

Definition: f has a **benign landscape** if all 2-critical points are optimal:

$\nabla f(x) = 0$ and $\nabla^2 f(x) \succcurlyeq 0$ implies x is a global min

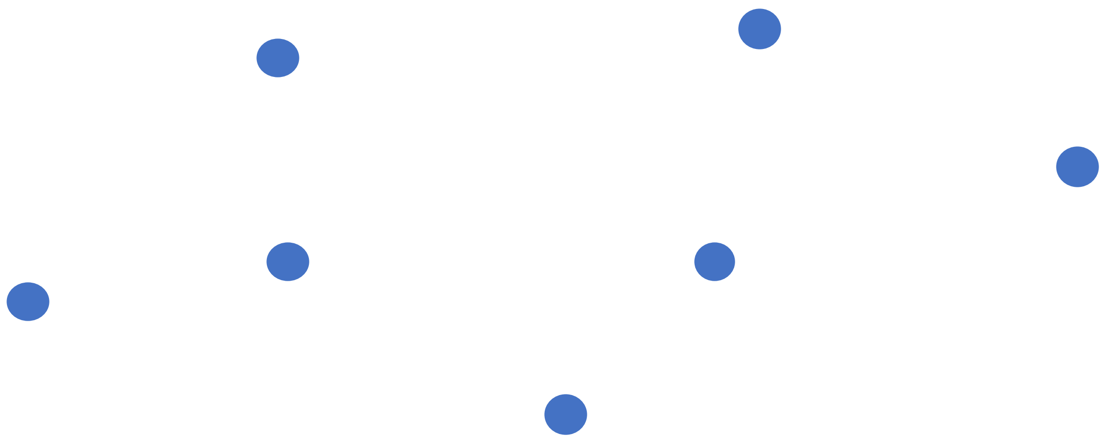
Part II of thesis:
Understand when several problems have benign landscape. With an eye towards new tools.

Our focus next: sensor network localization



The problem: SNL

n unknown points $z_1^*, z_2^*, \dots, z_n^*$ in $\mathbb{R}^\ell$.

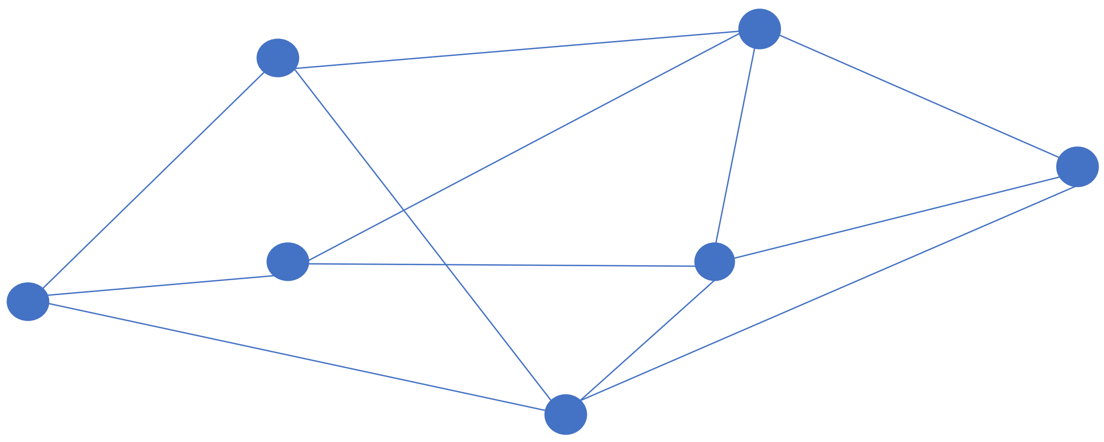


The problem: SNL

n unknown points $z_1^*, z_2^*, \ldots, z_n^*$ in $\mathbb{R}^\ell$.

Know a subset of the pairwise distances (measurements)

$$d_{ij} = \|z_i^* - z_j^*\| \text{ for } ij \in E.$$



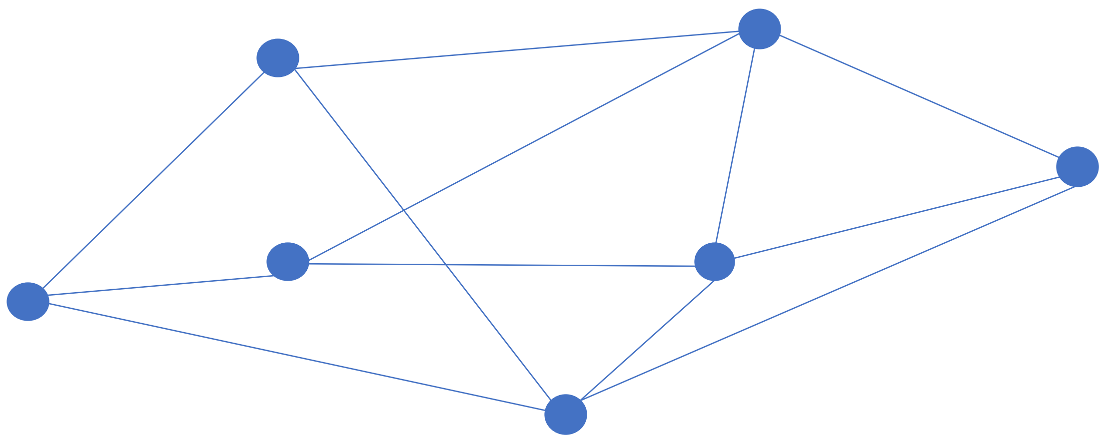
The problem: SNL

n unknown points $z_1^*, z_2^*, \ldots, z_n^*$ in $\mathbb{R}^\ell$.

Know a subset of the pairwise distances (measurements)

$$d_{ij} = \|z_i^* - z_j^*\| \text{ for } ij \in E.$$

Goal: recover the n points (up to translation & rotation)



The problem: SNL

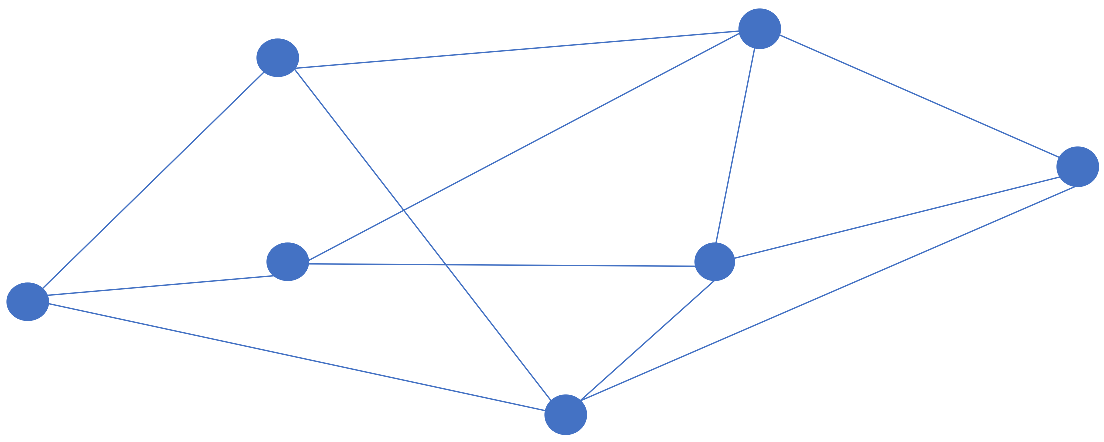
n unknown points $z_1^*, z_2^*, \dots, z_n^*$ in $\mathbb{R}^\ell$.

Know a subset of the pairwise distances (measurements)

$$d_{ij} = \|z_i^* - z_j^*\| \text{ for } ij \in E.$$

Goal: recover the n points (up to translation & rotation)

Sensor network localization (SNL) – Torgerson ’58, Shepard ‘62



The problem: SNL

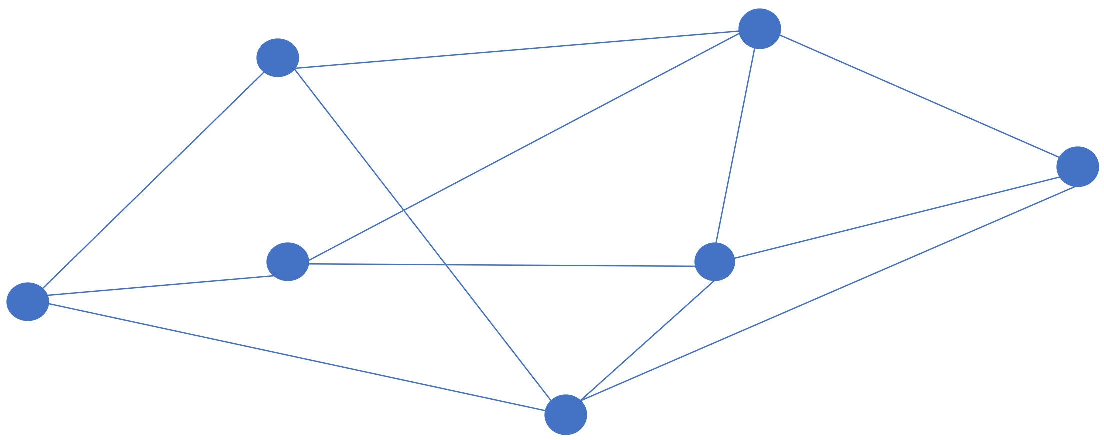
n unknown points $z_1^*, z_2^*, \ldots, z_n^*$ in $\mathbb{R}^\ell$.

Know a subset of the pairwise distances (measurements)

$$d_{ij} = \|z_i^* - z_j^*\| \text{ for } ij \in E.$$

Goal: recover the n points (up to translation & rotation)

Sensor network localization (SNL) – Torgerson '58, Shepard '62



Applications in robotics, determining molecular conformations, data analysis

Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \dots, z_n \in \mathbb{R}^\ell$

Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^\ell$

Nonconvex!

Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^\ell$

Relax to dimension $k > \ell$

Optimization problem

$$\min \sum_{ij \in E} \left(\left\| z_i - z_j \right\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \| z_i^* - z_j^* \|$$

over $z_1, z_2, \dots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$

Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

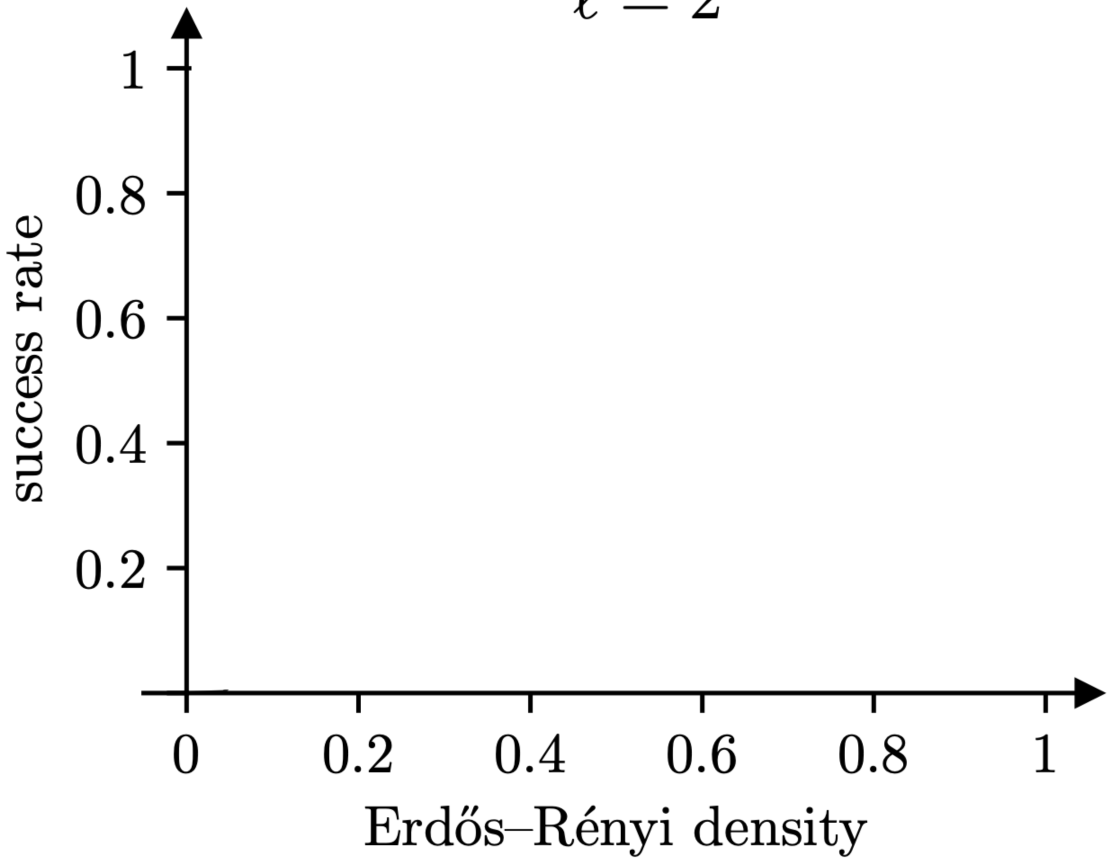
over $z_1, z_2, \dots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$

Experiment:

- $n = 50, \ell = \text{dimension} = 2$
- Ground truth = iid Gaussian points
- Graph = ER
- Run TR from random initialization

$\ell = 2$

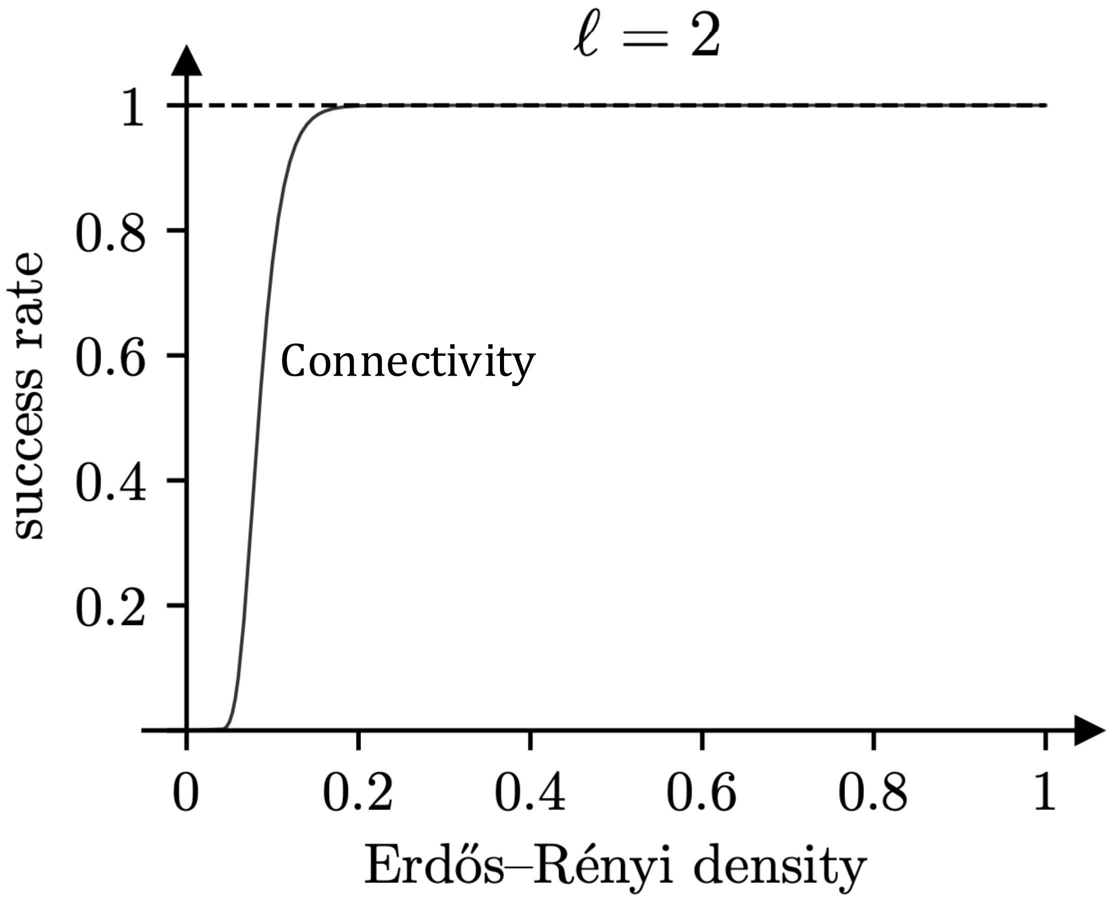


Optimization problem

$$\min \sum_{ij\in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$



Experiment:

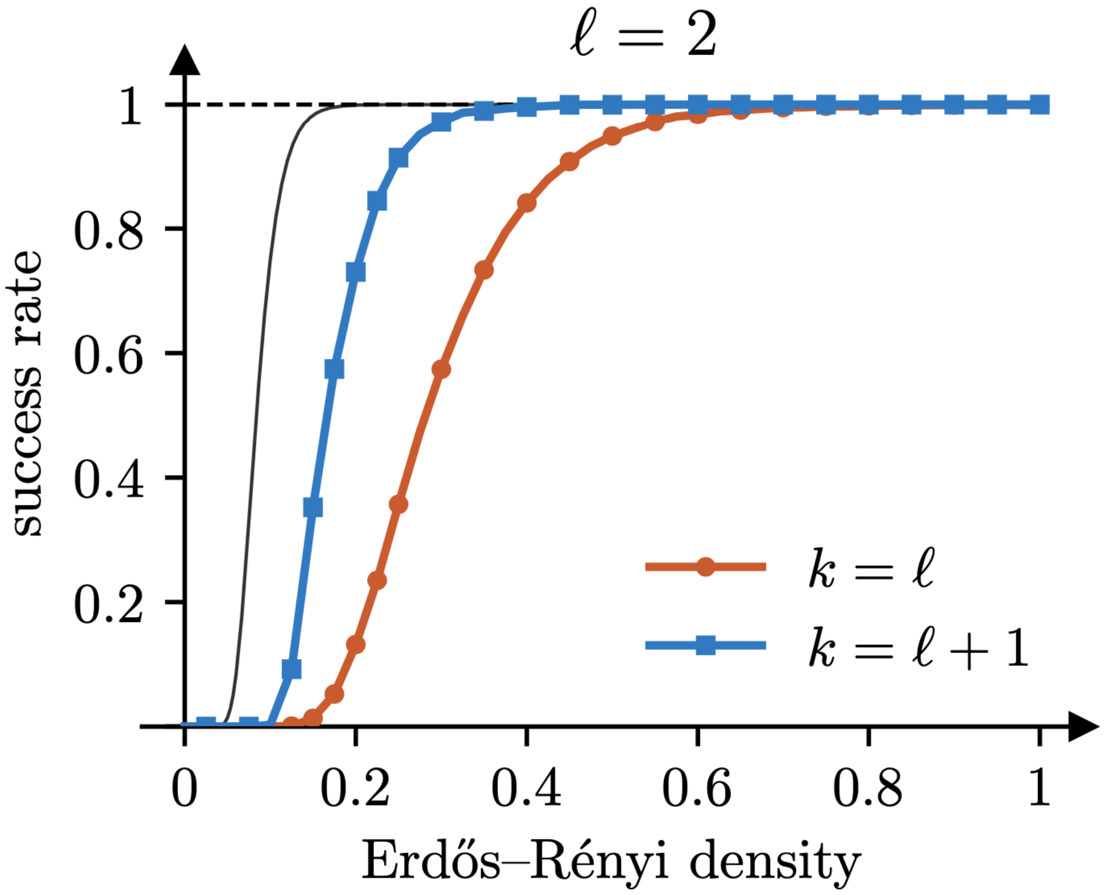
- $n = 50, \ell = \text{dimension} = 2$
- Ground truth = iid Gaussian points
- Graph = ER
- Run TR from random initialization

Optimization problem

$$\min \sum_{ij\in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$



Experiment:

- $n = 50, \ell =$ dimension $= 2$
- Ground truth = iid Gaussian points
- Graph = ER
- Run TR from random initialization

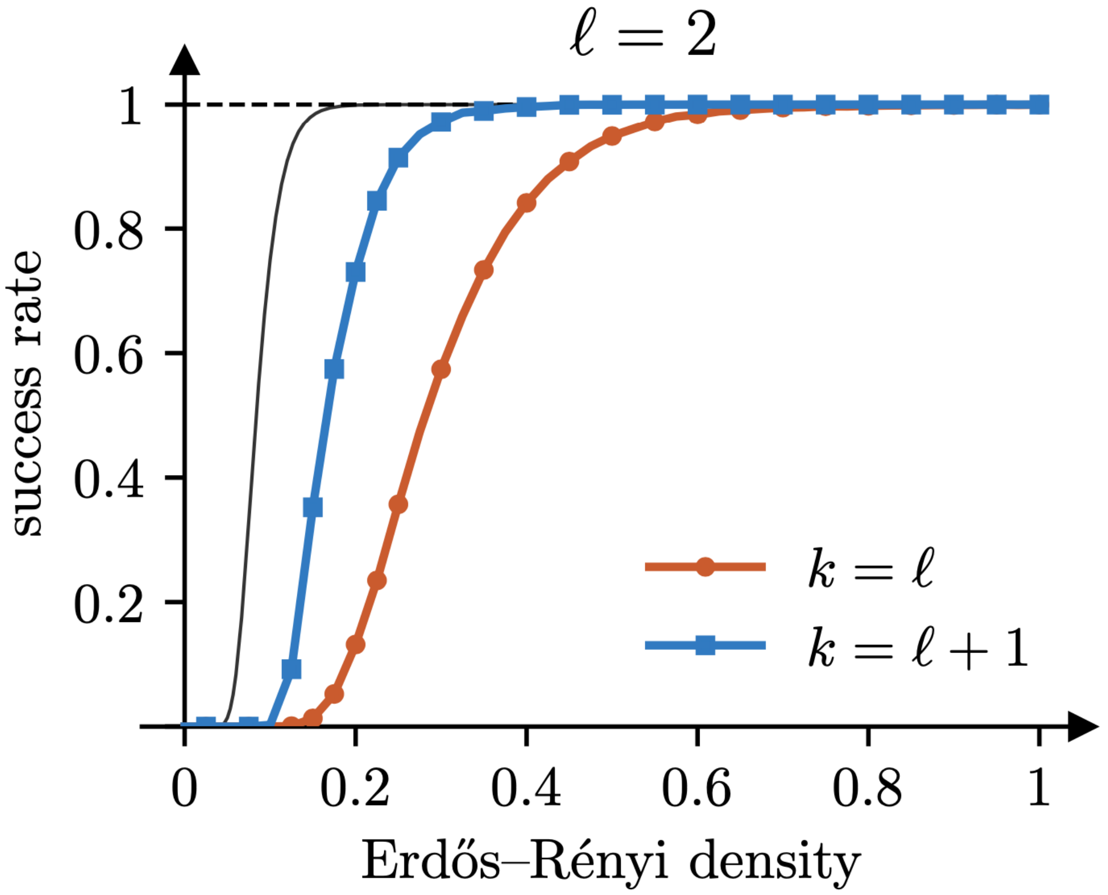
Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$

Theoretical explanation?



Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \dots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$

Theoretical explanation?

Want k small as possible! New problem has kn variables

If $k = n - 1$, landscape is benign (Song, Goncalves, Jung, Lavor, Mucherino, Wolkowicz, 2024)

Can we do better?

Optimization problem

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

Relax to dimension $k > \ell$

Theoretical explanation?

Want k small as possible! New problem has kn variables

If $k = n - 1$, landscape is benign (Song, Goncalves, Jung, Lavor, Mucherino, Wolkowicz, 2024)

Can we do better?

Our focus: (nearly) all distances known.
Landscape for s-stress and its relaxations unknown even in this simplest of cases!

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

"s-stress"

Results

Theorem [relaxation is necessary]: If $k = \ell$ (no relaxation), landscape may be non-benign even if graph is complete.

Also see:
Song, Goncalves, Jung, Lavor, Mucherino, Wolkowicz, 2024

Results

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$ "s-stress"

Theorem [relaxation is necessary]: If $k = \ell$ (no relaxation), landscape may be non-benign even if graph is complete.

Also see:
Song, Goncalves, Jung, Lavor, Mucherino, Wolkowicz, 2024

Theorem [arbitrary GT]: If graph is complete and relax to

$$k \approx \ell + \sqrt{n\ell},$$

then every 2-critical point is the ground truth.

Results

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$ "s-stress"

Theorem [relaxation is necessary]: If $k = \ell$ (no relaxation), landscape may be non-benign even if graph is complete.

Also see:
Song, Goncalves, Jung, Lavor, Mucherino, Wolkowicz, 2024

Theorem [arbitrary GT]: If graph is complete and relax to

$$k \approx \ell + \sqrt{n\ell},$$

then every 2-critical point is the ground truth.

Theorem [isotropic GT]: If graph is nearly complete*, ground truth points are isotropic* and iid, and relax to

$$k \approx \ell + \log(n),$$

then every 2-critical point is the ground truth, w.h.p.

$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^k$

"s-stress"

Results

Theorem [relaxation is necessary]: If $k = \ell$ (no relaxation), landscape may be non-benign even if graph is complete.

Also see:
Song, Goncalves, Jung,
Lavor, Mucherino,
Wolkowicz, 2024

Theorem [arbitrary GT]: If graph is complete and relax to

$$k \approx \ell + \sqrt{n\ell},$$

then every 2-critical point is the ground truth.

Theorem [isotropic GT]: If graph is nearly complete*, ground truth points are isotropic* and iid, and relax to

$$k \approx \ell + \log(n),$$

then every 2-critical point is the ground truth, w.h.p.

Why?
Why important?

Example

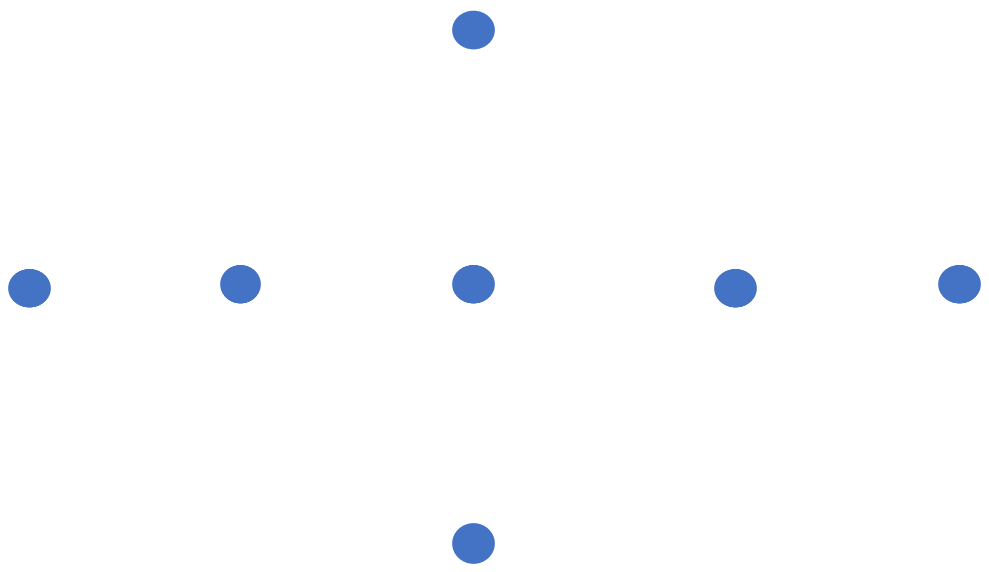
$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^\ell$

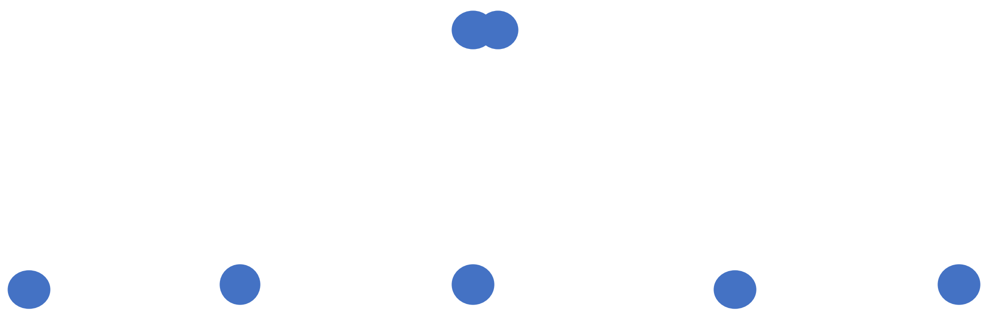
"s-stress"

s-stress can have spurious strict local minima, with complete graph

Ground truth $z_1^*, z_2^*, \ldots$



Spurious configuration $z_1, z_2, \ldots$



Example

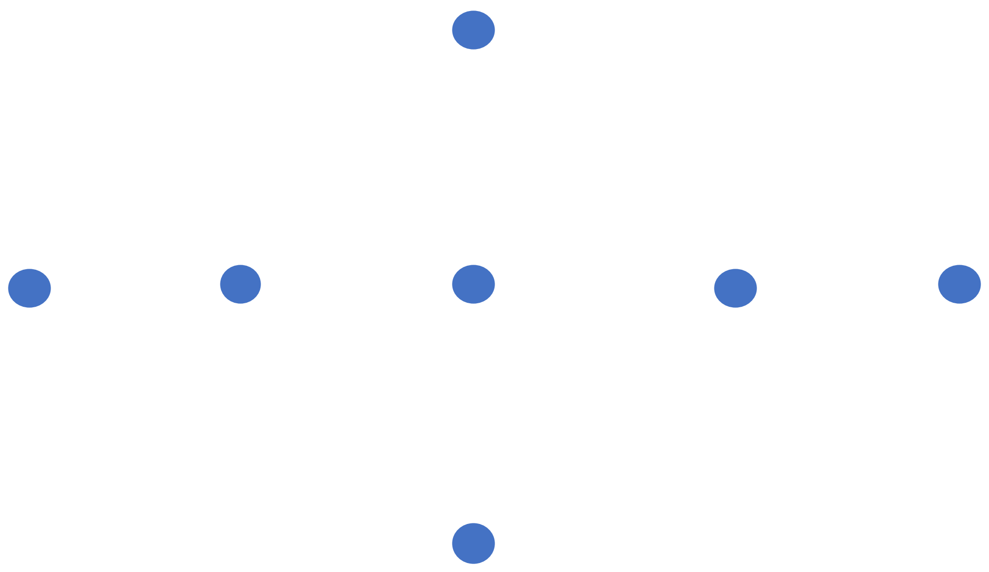
$$\min \sum_{ij \in E} \left(\|z_i - z_j\|^2 - d_{ij}^2 \right)^2, \qquad d_{ij} = \|z_i^* - z_j^*\|$$

over $z_1, z_2, \ldots, z_n \in \mathbb{R}^\ell$

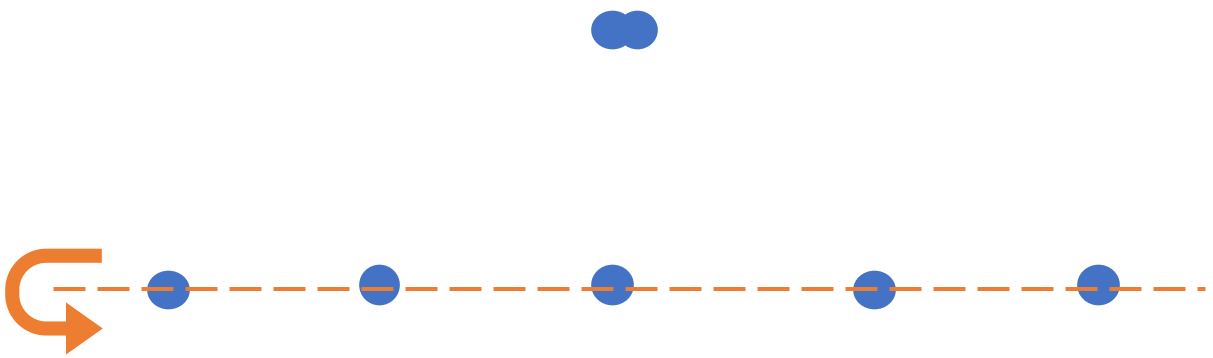
"s-stress"

s-stress can have spurious strict local minima, with complete graph

Ground truth $z_1^*, z_2^*, \ldots$



Spurious configuration $z_1, z_2, \ldots$



Optimize over points in $\mathbb{R}^3$

Results

Why important?

Results

Why important?

First nontrivial landscape results for SNL

Results

Why important?

First nontrivial landscape results for SNL

New mathematical techniques needed to prove benign landscape.

➔Beyond the Restricted Isometry Property

Results

Why important?

First nontrivial landscape results for SNL

New mathematical techniques needed to prove benign landscape.

➔Beyond the Restricted Isometry Property

Another instance where low-dimensional nonconvex relaxations are successful, empirically and theoretically (Z2/orthogonal synchron: McRae/Boumal/Bandeira/… ’16 ’23 ’24).

Results

Why important?

First nontrivial landscape results for SNL

New mathematical techniques needed to prove benign landscape.
➔Beyond the Restricted Isometry Property

Another instance where low-dimensional nonconvex relaxations are successful, empirically and theoretically (Z2/orthogonal synchron: McRae/Boumal/Bandeira/… ’16 ’23 ’24).

Provides a theoretical foundation for more sophisticated localization problems.

Results

Why important?

First nontrivial landscape results for SNL

New mathematical techniques needed to prove benign landscape.
➔Beyond the Restricted Isometry Property

Another instance where low-dimensional nonconvex relaxations are successful, empirically and theoretically (Z2/orthogonal synchron: McRae/Boumal/Bandeira/… ’16 ’23 ’24).

Provides a theoretical foundation for more sophisticated localization problems.

Going forward: tools for proving benign landscapes?