Multidimensional scaling has benign landscape under mild rank relaxation

DACO, March 19, 2024

Chris Criscitiello
with
Andrew McRae, Quentin Rebjock, Nicolas Boumal

OPTIM, Chair of Continuous Optimization
Institute of Mathematics, EPFL

EPFL

Landscape analysis

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

Run local algorithm (gradient descent, trust regions, …)

Landscape analysis

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

Run local algorithm (gradient descent, trust regions, …)

Will converge to 2-critical point (w/ prob 1)

[1-critical] $\nabla f(x) = 0$

[2-critical] $\nabla f(x) = 0$ and $\nabla f(x) \succcurlyeq 0$

Landscape analysis

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

Run local algorithm (gradient descent, trust regions, …)

Will converge to 2-critical point (w/ prob 1)

[1-critical] $\nabla f(x) = 0$

[2-critical] $\nabla f(x) = 0$ and $\nabla f(x) \succcurlyeq 0$

Goal: Show all 2-critical points x are global minima.

Landscape analysis

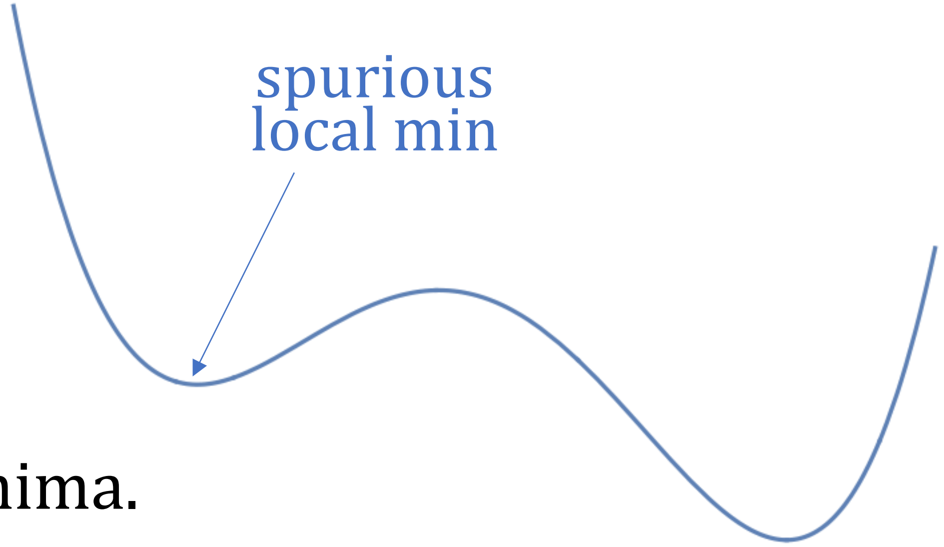
$$\min_{x \in \mathcal{M}} f(x)$$

Run local algorithm (gradient descent, trust regions, …)

Will converge to 2-critical point (w/ prob 1)

[1-critical] $\nabla f(x) = 0$

[2-critical] $\nabla f(x) = 0$ and $\nabla f(x) \succcurlyeq 0$



Goal: Show all 2-critical points x are global minima.

Landscape analysis

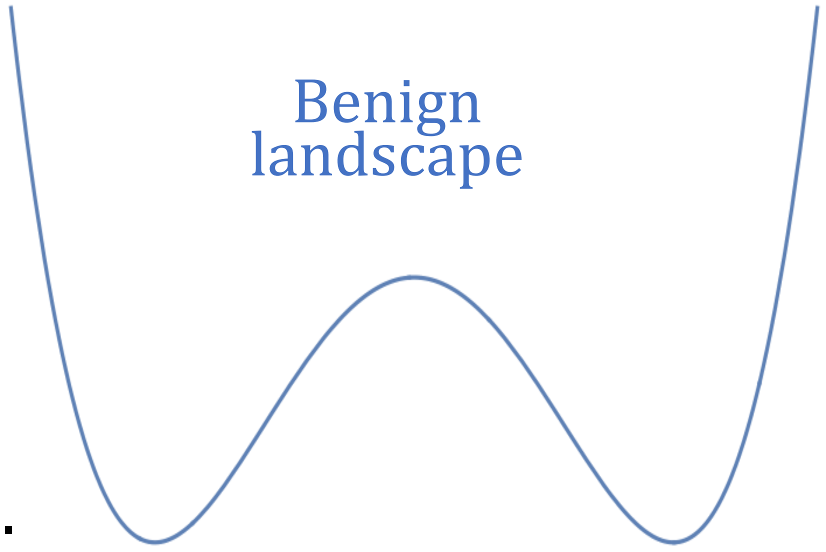
$$\min_{x \in \mathcal{M}} f(x)$$

Run local algorithm (gradient descent, trust regions, …)

Will converge to 2-critical point (w/ prob 1)

[1-critical] $\nabla f(x) = 0$

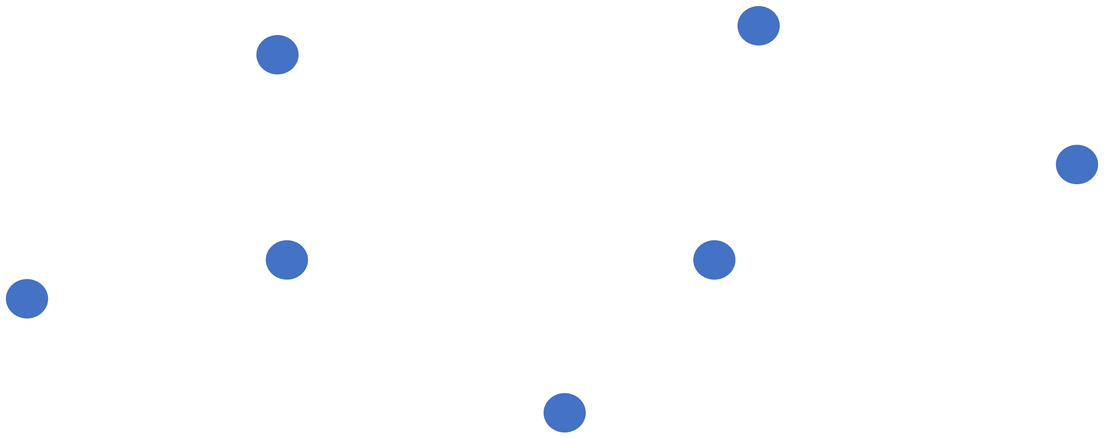
[2-critical] $\nabla f(x) = 0$ and $\nabla f(x) \succcurlyeq 0$



Goal: Show all 2-critical points x are global minima.

The problem

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

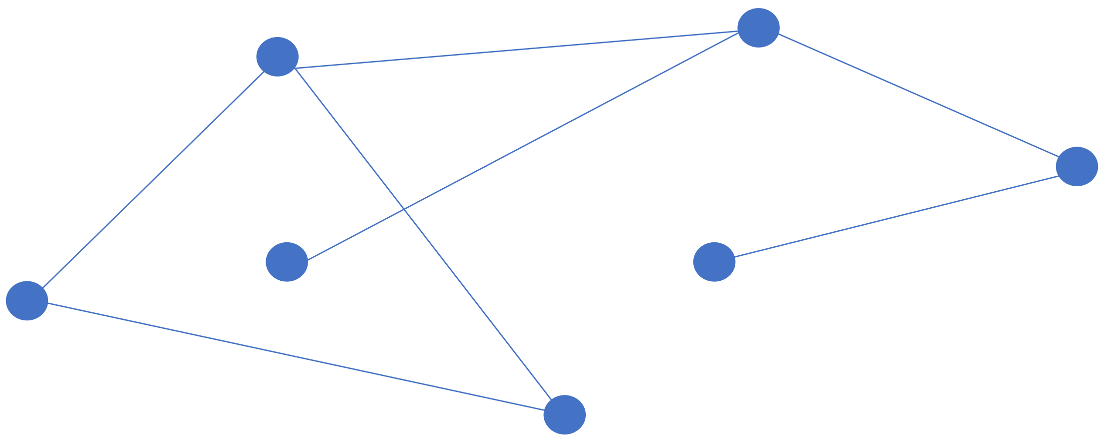


The problem

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

Know a subset of the pairwise distances (measurements)

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



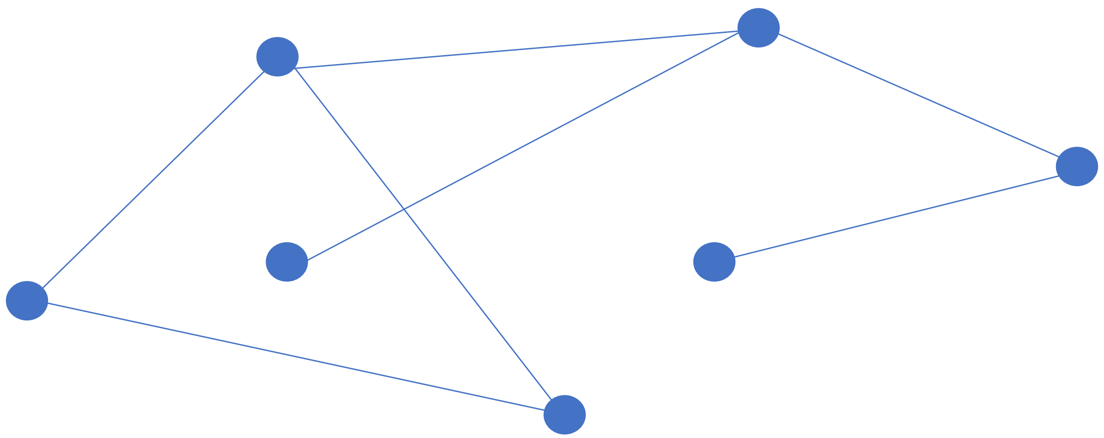
The problem

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)



Applications

Robotics (**sensor network localization**), $\ell = 2,3$

Data analysis (metric **multidimensional scaling**)

Graph theory (rigidity)

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

"s-stress"

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

"s-stress"

Solved via local algorithms. Guarantees?

Nonconvex! How bad?

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

"s-stress"

Solved via local algorithms. Guarantees?

Nonconvex! How bad?

Possible variations: Noisy measurements, landmarks, …

Our focus: (nearly) complete graphs, no noise

Synthetic experiments

Recipe (all distances known):

(1) Choose ground truths $z_1^*, z_2^*, \ldots, z_n^*$ at random (normal iid)

(2) Run gradient descent/trust regions/etc.

(3) Find global min?

(4) Repeat

Synthetic experiments

Recipe (all distances known):

(1) Choose ground truths $z_1^*, z_2^*, \dots, z_n^*$ at random (normal iid)

(2) Run gradient descent/trust regions/etc.

(3) Find global min?

(4) Repeat

Always finds global min!

Synthetic experiments

Recipe (all distances known):

(1) Choose ground truths $z_1^*, z_2^*, \dots, z_n^*$ at random (normal iid)

(2) Run gradient descent/trust regions/etc.

(3) Find global min?

(4) Repeat

Always finds global min!

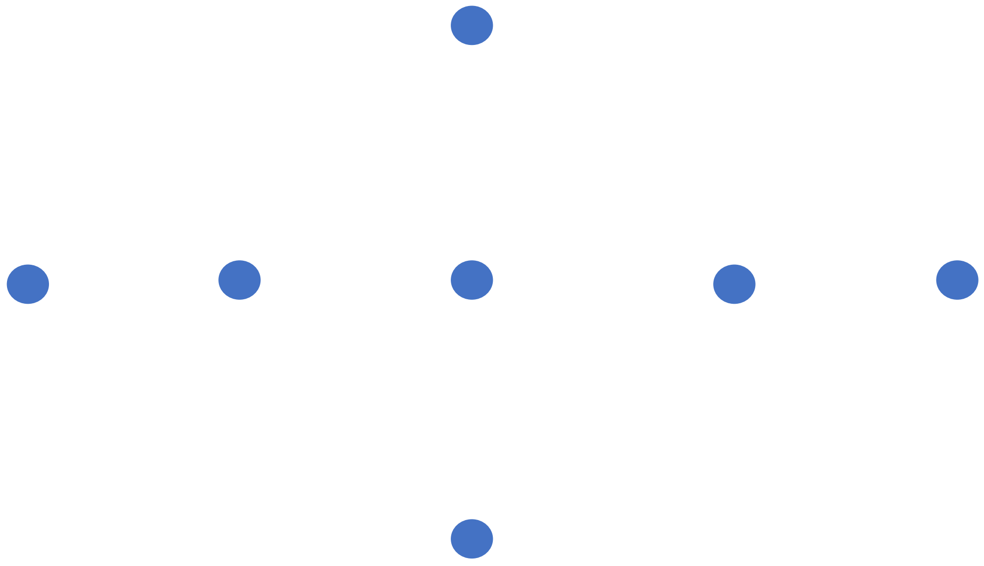
Open Question: Does s-stress have spurious local minima? Are all 2-critical points global minima?

* Malone & Trosset 2000, Parhizkar 2013, etc.

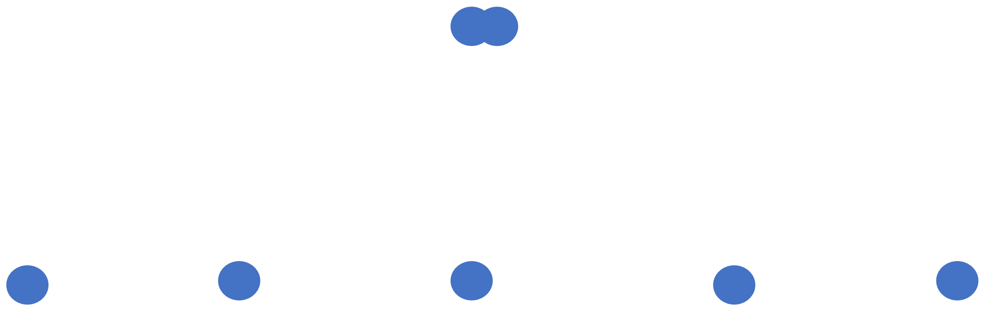
Counterexamples

s-stress can have spurious strict local minima!

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



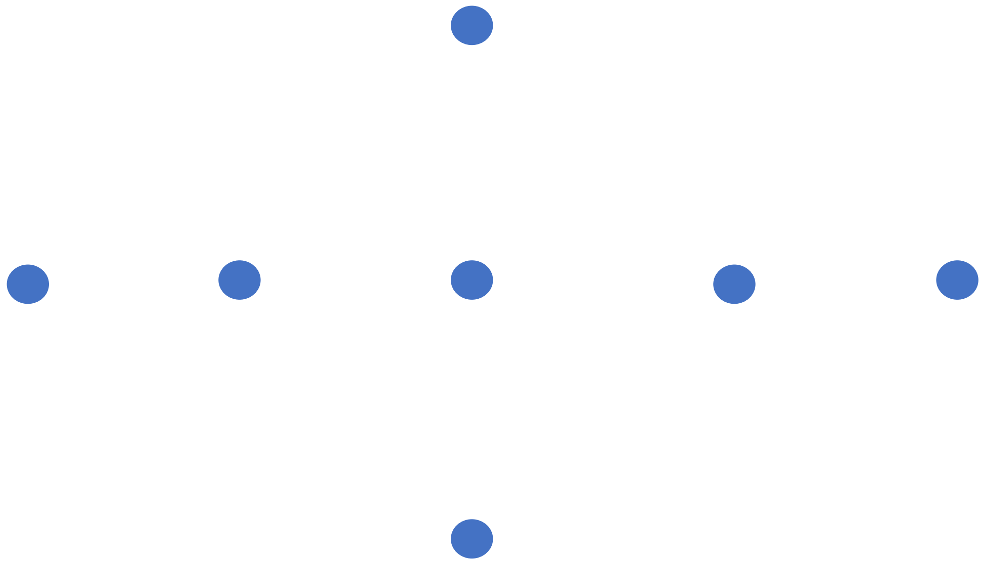
Spurious configuration $z_1, z_1, \ldots$



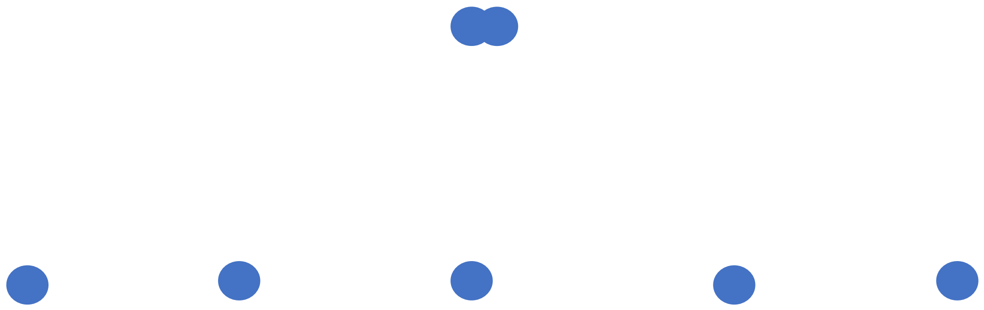
Counterexamples

s-stress can have spurious strict local minima!

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



Spurious configuration $z_1, z_2, \ldots$



Set of ground truths with spurious local minima has positive measure

Nonconvex relaxation

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

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

Nonconvex relaxation

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

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

Relax to dimension $k > \ell$

Nonconvex relaxation

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

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

Relax to dimension $k > \ell$

Minimizer of relaxed problem same as original?

Yes if graph is complete (or more generally if it is *universally rigid*)

Nonconvex relaxation

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

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

Relax to dimension $k > \ell$

Minimizer of relaxed problem same as original?
Yes if graph is complete (or more generally if it is *universally rigid*)

Want k small; new problem has kn variables

Nonconvex relaxation

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

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

Relax to dimension $k > \ell$

Minimizer of relaxed problem same as original?
Yes if graph is complete (or more generally if it is *universally rigid*)

Want k small; new problem has kn variables
If $k = n - 1$, landscape is benign (later)
Can we do better?

Results

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

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.

Results

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.

Conjecture [**arbitrary GT**]: Relaxing to $k = \ell + 1$ is enough.

Conjecture [**isotropic GT**]: Relaxing is not necessary.

Results

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.

Conjecture [arbitrary GT]: Relaxing to $k = \ell + 1$ is enough.

Conjecture [isotropic GT]: Relaxing is not necessary.

- True if $n \leq \ell + 3$
- Numerical optim to explicitly search for counterexamples.

Results

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.

Conjecture [arbitrary GT]: Relaxing to $k = \ell + 1$ is enough.

Conjecture [isotropic GT]: Relaxing is not necessary.

- True if $n \leq \ell + 3$
- Numerical optim to explicitly search for counterexamples.

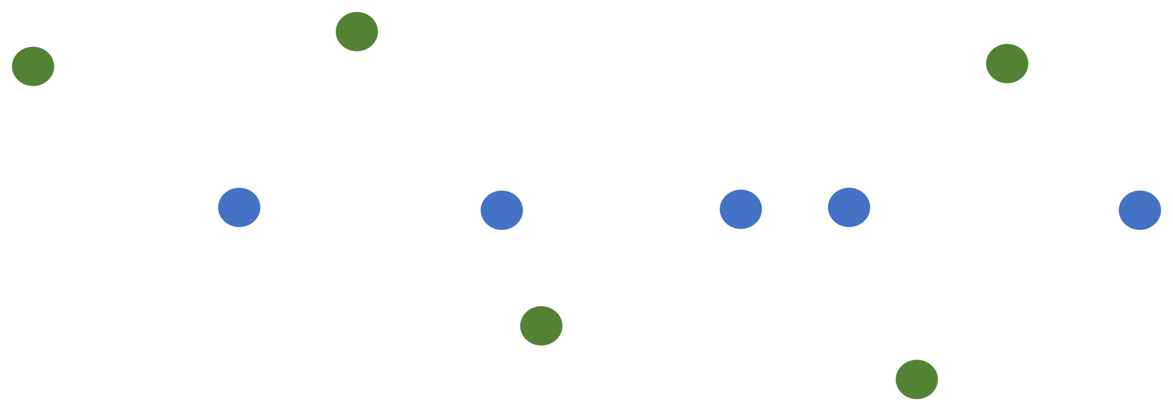
Some ideas from the proof

Ground truth $z_1^*, z_2^*, \ldots$ in dimension ℓ

Some ideas from the proof

Ground truth $z_1^*, z_2^*, \ldots$ in dimension ℓ

1-critical configuration in dimension $k > \ell$

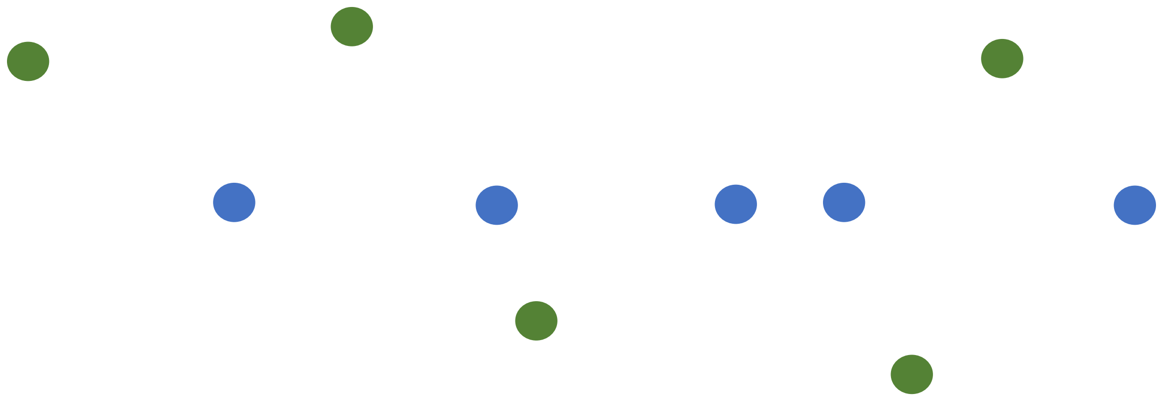


Some ideas from the proof

Ground truth $z_1^*, z_2^*, \ldots$ in dimension ℓ

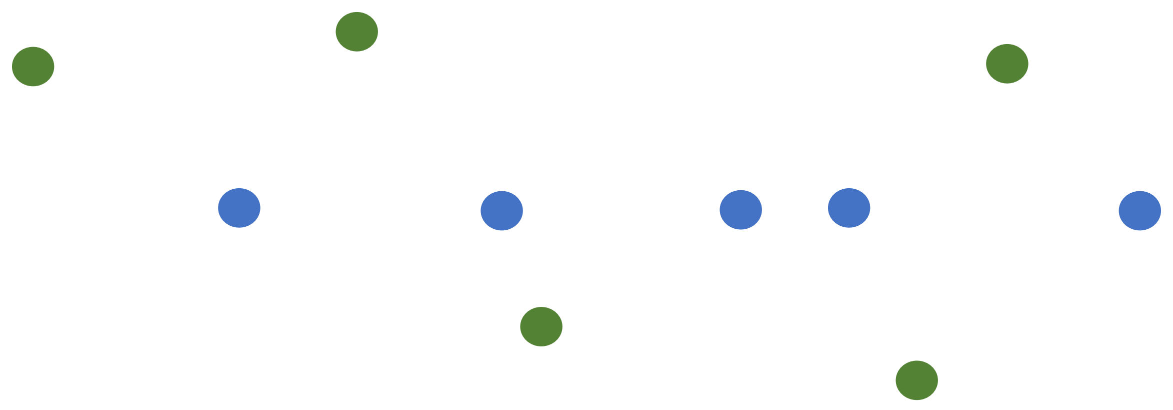
1-critical configuration in dimension $k > \ell$

Goal: perturb 1-critical configuration to decrease cost



Some ideas from the proof

Goal: perturb 1-critical configuration to decrease cost

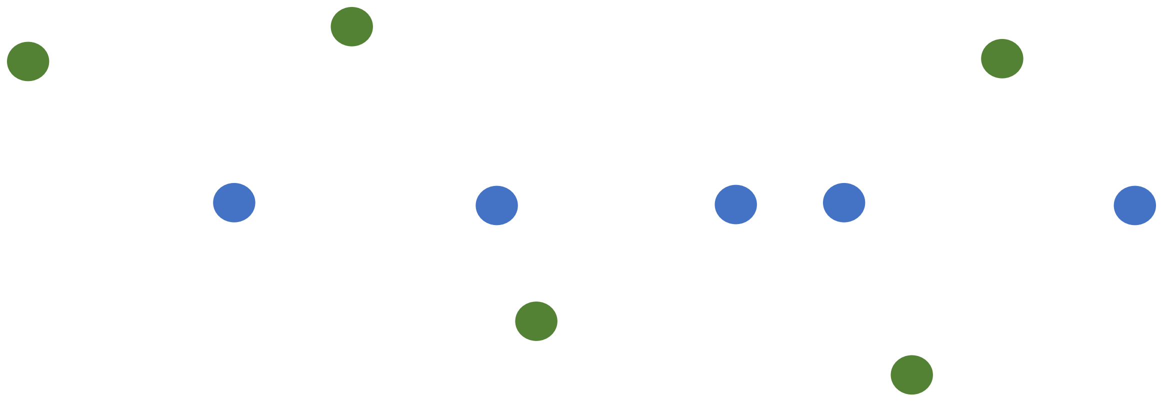


Some ideas from the proof

Goal: perturb 1-critical configuration to decrease cost

Best linear transformation $z_i \mapsto Rz_i$ mapping 1-critical config to ground truth

$$R = \text{argmin}_R \|Z_* - ZR^\top\|^2, \qquad Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n \times k}$$

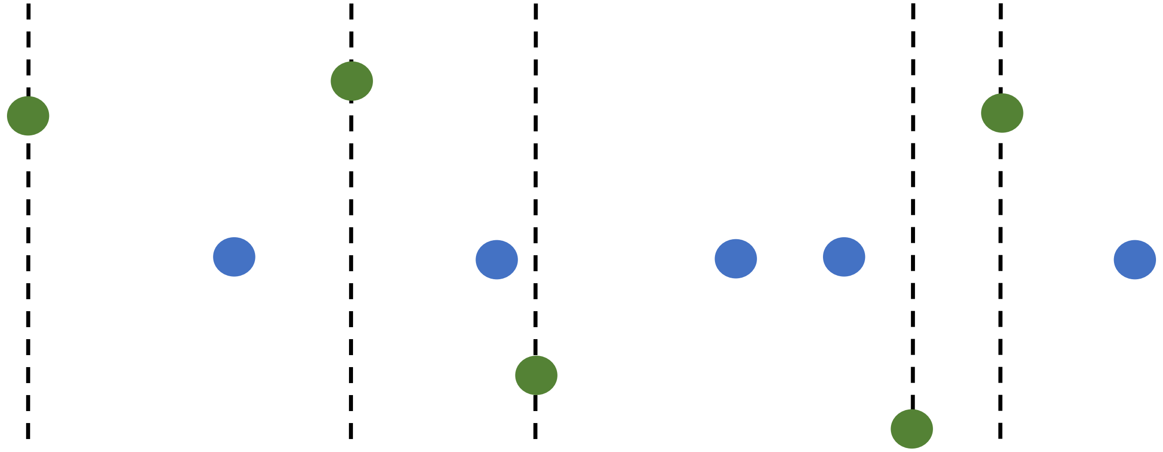


Some ideas from the proof

Goal: perturb 1-critical configuration to decrease cost

Best linear transformation $z_i \mapsto Rz_i$ mapping 1-critical config to ground truth

$$R = \operatorname{argmin}_R \|Z_* - ZR^\top\|^2, \qquad Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n \times k}$$

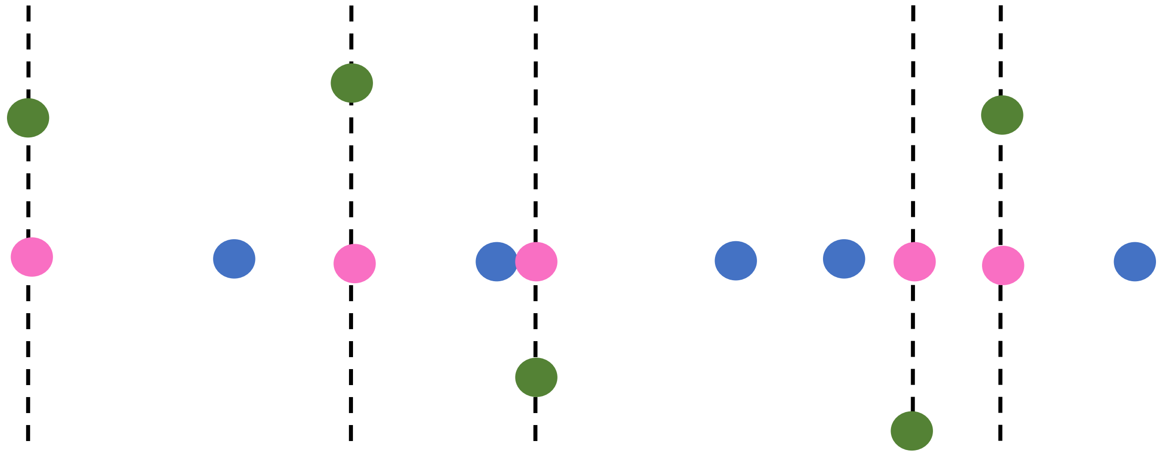


Some ideas from the proof

Goal: perturb 1-critical configuration to decrease cost

Best linear transformation $z_i \mapsto Rz_i$ mapping 1-critical config to ground truth

$$R = \text{argmin}_R \|Z_* - ZR^\top\|^2, \qquad Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n \times k}$$

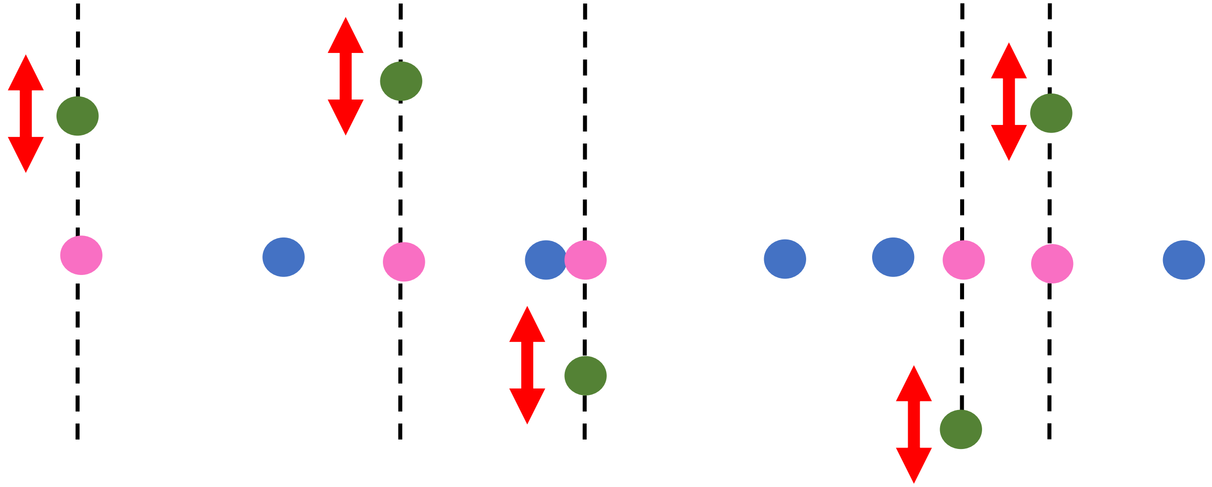


Some ideas from the proof

Goal: perturb 1-critical configuration to decrease cost

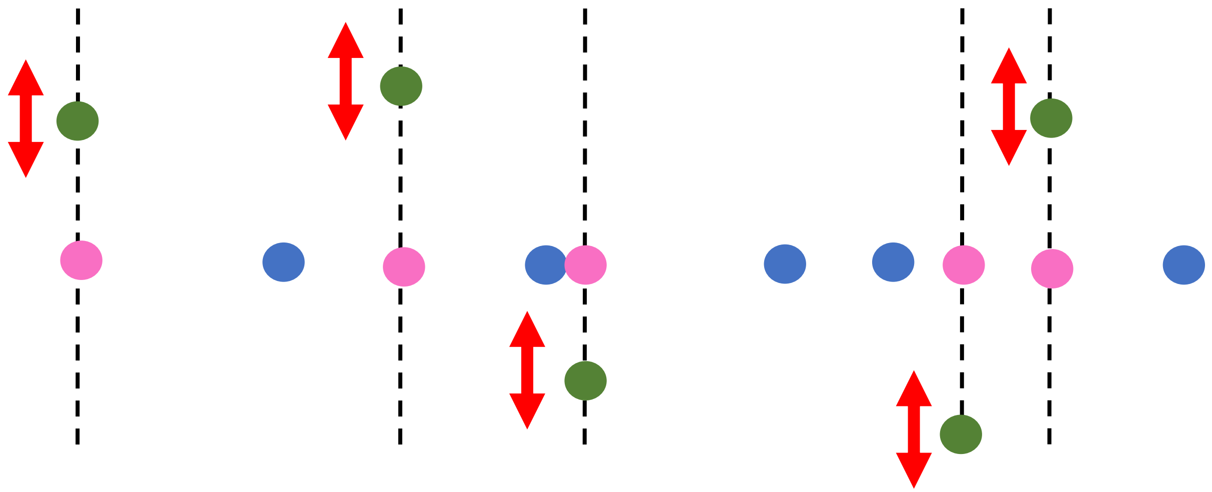
Best linear transformation $z_i \mapsto Rz_i$ mapping 1-critical config to ground truth

$$R = \mathrm{argmin}_R \|Z_* - ZR^\top\|^2, \qquad Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n\times k}$$



Some ideas from the proof

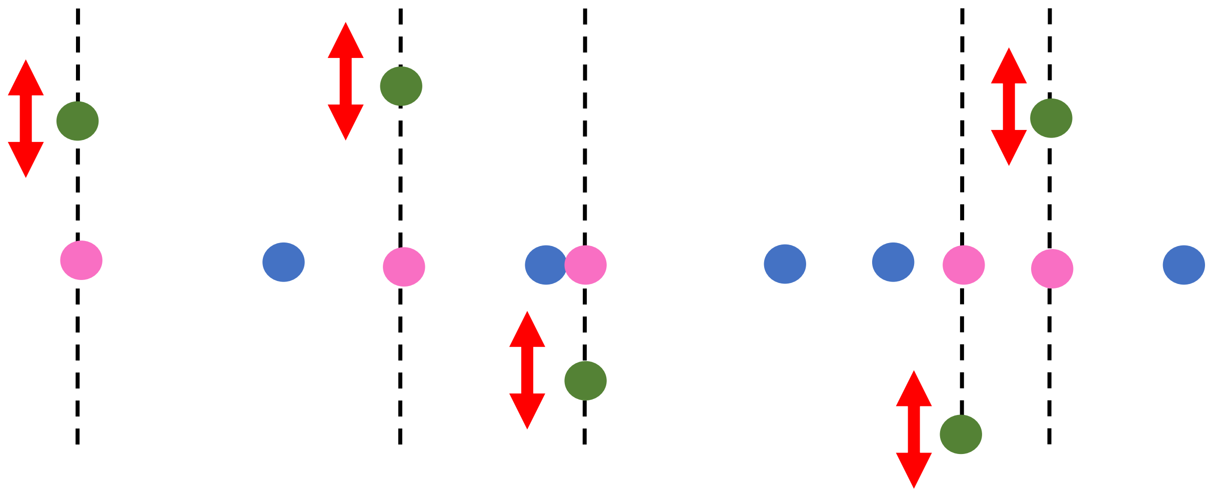
If relax enough, many ways to perturb this way



Some ideas from the proof

If relax enough, many ways to perturb this way

Use **eigenvalue interlacing** to argue that a good one exists, if relax enough

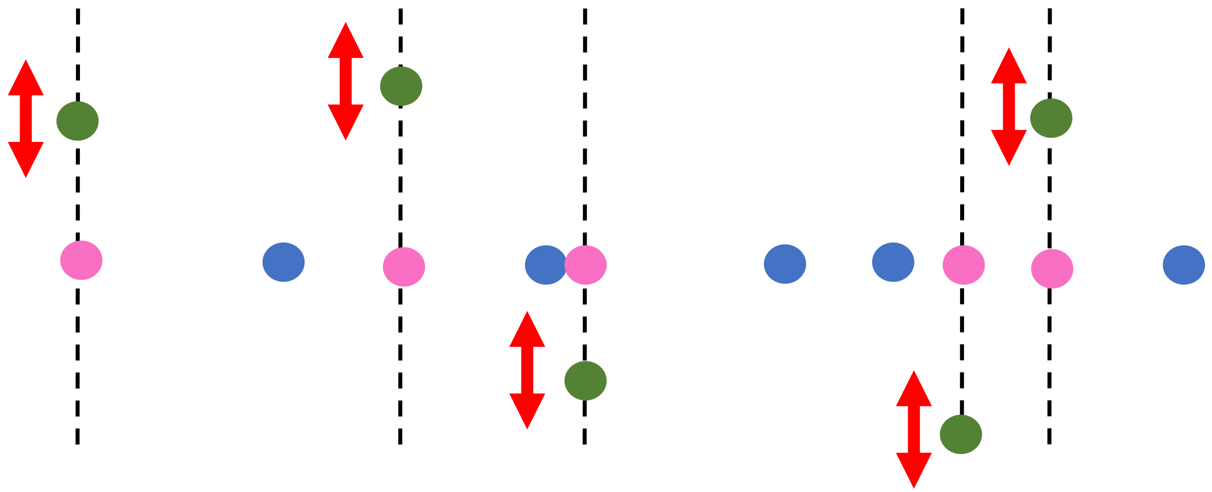


Some ideas from the proof

If relax enough, many ways to perturb this way

Use **eigenvalue interlacing** to argue that a good one exists, if relax enough

Somewhat nonconstructive



Some ideas from the proof

If relax enough, many ways to perturb this way

Use **eigenvalue interlacing** to argue that a good one exists, if relax enough

Somewhat nonconstructive

For **isotropic GT,** $k \approx \ell \log(n)$, similar descent direction

Randomize over descent directions (instead of eigenvalue interlacing)

Connection to low-rank optimization

Notation and reformulation

$$Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n\times\ell}, \qquad Z_* = \begin{pmatrix} {z_1^*}^\top \\ \vdots \\ {z_n^*}^\top \end{pmatrix} \in \mathbb{R}^{n\times\ell}$$

Notation and reformulation

$$Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n\times \ell}, \qquad Z_* = \begin{pmatrix} {z_1^*}^\top \\ \vdots \\ {z_n^*}^\top \end{pmatrix} \in \mathbb{R}^{n\times \ell}$$

Gram matrices $Y = ZZ^\top, Y_* = Z_* Z_*^\top$

Notation and reformulation

$$Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n\times\ell}, \qquad Z_* = \begin{pmatrix} {z_1^*}^\top \\ \vdots \\ {z_n^*}^\top \end{pmatrix} \in \mathbb{R}^{n\times\ell}$$

Gram matrices $Y = ZZ^\top, Y_* = Z_* Z_*^\top$

MDS map $\Delta : \mathrm{Sym}(n) \to \mathrm{Hollow}(n)$

Gram $\to$ EDM (euclidean distance matrix)

Notation and reformulation

$$Z = \begin{pmatrix} z_1^\top \\ \vdots \\ z_n^\top \end{pmatrix} \in \mathbb{R}^{n\times\ell}, \qquad Z_* = \begin{pmatrix} z_1^{*\top} \\ \vdots \\ z_n^{*\top} \end{pmatrix} \in \mathbb{R}^{n\times\ell}$$

Gram matrices $Y = ZZ^\top, Y_* = Z_*Z_*^\top$

MDS map $\Delta : \mathrm{Sym}(n) \to \mathrm{Hollow}(n)$

Gram $\to$ EDM (euclidean distance matrix)

$$[\Delta(Y)]_{ij} := Y_{ii} + Y_{jj} - 2Y_{ij} = \left\| z_i - z_j \right\|^2$$

Notation and reformulation

$$\min \|\Delta(ZZ^\top - Z_* Z_*^\top)\|^2 \text{ over } Z \in \mathbb{R}^{n\times\ell}$$

Notation and reformulation

$$\min \|\Delta(ZZ^\top - Z_* Z_*^\top)\|^2 \text{ over } Z \in \mathbb{R}^{n\times\ell}$$

$$\Updownarrow$$

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \le \ell$$

Notation and reformulation

$$\min \|\Delta(ZZ^\top - Z_* Z_*^\top)\|^2 \text{ over } Z \in \mathbb{R}^{n\times\ell}$$

$\Updownarrow$

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \le \ell$$

$\Downarrow$ (relax)

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \le k$$

Notation and reformulation

$$\min \|\Delta(ZZ^\top - Z_* Z_*^\top)\|^2 \text{ over } Z \in \mathbb{R}^{n\times\ell}$$

$$\Updownarrow$$

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \le \ell$$

$\Downarrow$ (relax)

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \le k$$

If $k = n$, problem is convex

Notation and reformulation

$$\min \|\Delta(ZZ^\top - Z_* Z_*^\top)\|^2 \text{ over } Z \in \mathbb{R}^{n\times\ell}$$

$$\Updownarrow$$

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with rank}(Y) \le \ell$$

$\Downarrow$ (relax)

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with rank}(Y) \le k$$

If $k = n$, problem is convex

Restricted Isometry Property? No! $\Delta^* \circ \Delta$ has condition number n

Special properties of MDS map?

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \leq k$$

Special properties of MDS map?

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \leq k$$

Special "perturbation" of the identity

$$(\Delta^* \circ \Delta)(Y) = Y + \Theta(Y)$$

$$(\Delta^* \circ \Delta)^{-1}(Y) = Y - \Gamma(Y)$$

Special properties of MDS map?

$$\min \|\Delta(Y - Y_*)\|^2 \text{ over } Y \succcurlyeq 0 \text{ with } \operatorname{rank}(Y) \leq k$$

Special "perturbation" of the identity

$$(\Delta^* \circ \Delta)(Y) = Y + \Theta(Y)$$

$$(\Delta^* \circ \Delta)^{-1}(Y) = Y - \Gamma(Y)$$

General theorem: If Γ is completely positive, contractive, "behaves well w.r.t. rank-1 matrices", and

$$\langle Y, \Theta(Y) \rangle \leq c \langle Y, \Gamma(Y) \rangle \quad \forall Y$$

then landscape is benign when relax to $k \approx \ell + \sqrt{c\ell}$.

Open questions

Conjecture [**arbitrary GT**]: Relaxing to $k = \ell + 1$ is enough.

Conjecture [**isotropic GT**]: Relaxing is not necessary.

Incomplete graphs (random, expanders, …)

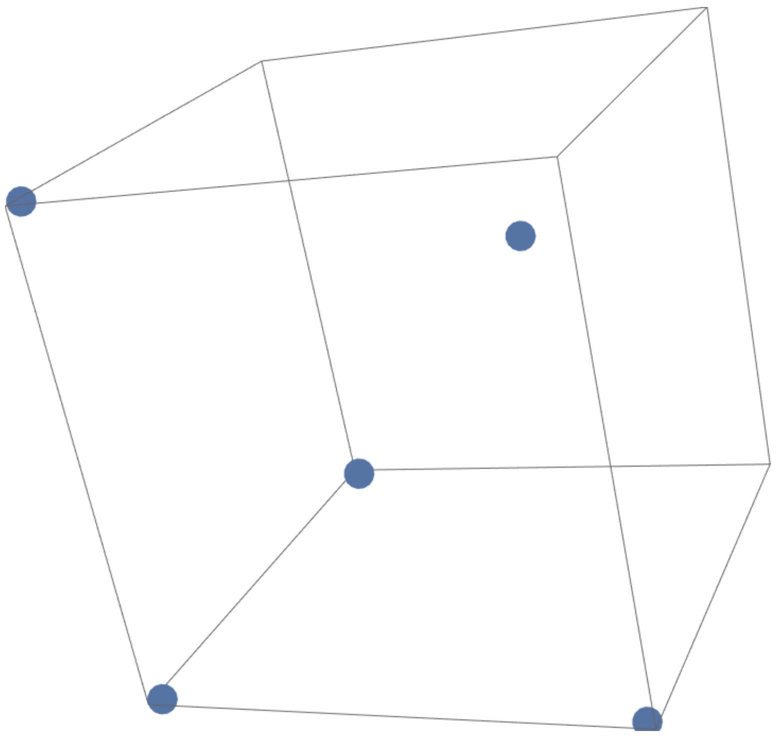
Many other SNL problems (noise models, trajectory localization, …)

Appendix

Counterexamples

Minima number of points to have spurious local minima?

$$\ell + 2 \text{ (for } \ell \geq 5\text{)}$$



Counterexamples

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.

Counterexamples

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.

The **hub** of a graph is the set of vertices which are connected to all other vertices.

$$H = \text{size of hub}$$

Counterexamples

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.

The **hub** of a graph is the set of vertices which are connected to all other vertices.

$$H = \text{size of hub}$$

Theorem [isotropic GT]: If ground truth points are isotropic and iid, and relax to

$$k \approx \text{poly}(n - H)\ell \log(n),$$

then every 2-critical point is the ground truth.