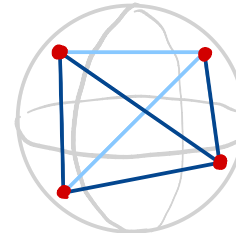
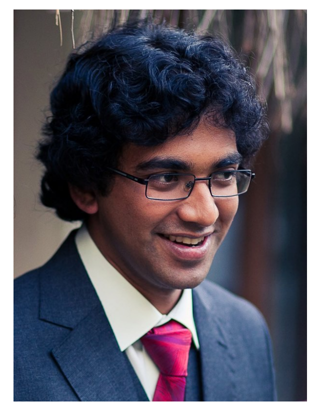
$$X = Y Y^T$$



The Burer-Monteiro SDP method can fail even above the Barvinok-Pataki bound

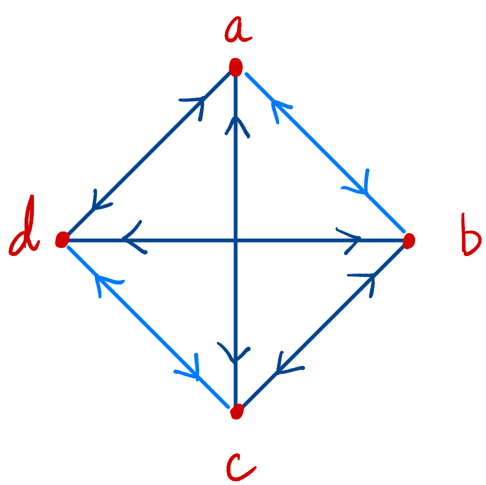
$$\sqrt{2m} \le p$$

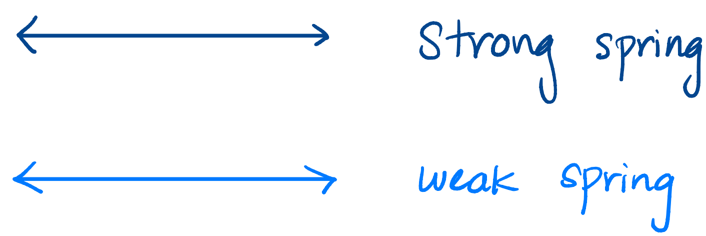


Liam O'Carroll   Vaidehi Srinivas   Aravindan Vijayaraghavan

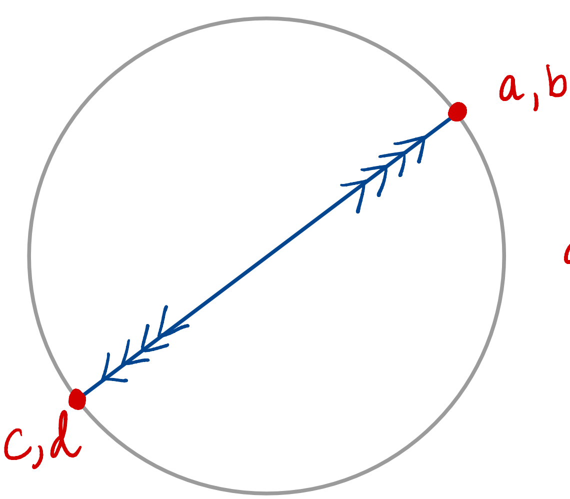
Northwestern University

TEASER TRAILER

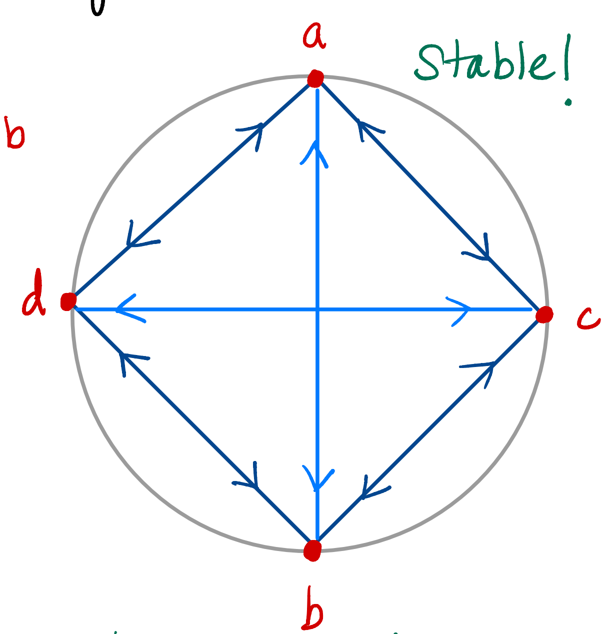




CHALLENGE: arrange vertices on circle in a "Stable" position



low potential energy



high potential energy

- Physics: "coupled oscillators"
- Algorithms: Goemans-Williamson Max Cut SDP

Semidefinite Programs

variable: $X \in \mathbb{R}^{n \times n}$

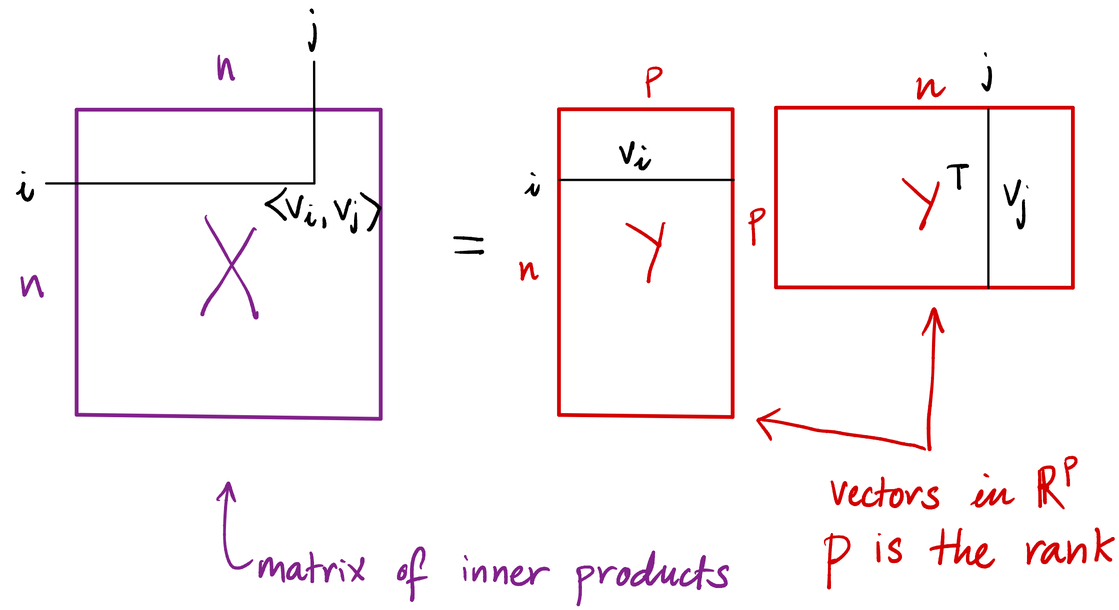
objective: minimize $\langle C, X \rangle$

subject to:
$$\langle A_1, X \rangle = b_1$$
$$\langle A_2, X \rangle = b_2$$
$$\vdots$$
$$\langle A_m, X \rangle = b_m$$

$$X \succeq 0$$

"X is positive semidefinite"

$$X = YY^T$$



SEMIDEFINITE PROGRAMS

variable: $X \in \mathbb{R}^{n \times n}$

objective: minimize $\langle C, X \rangle$

subject to: $\langle A_1, X \rangle = b_1$

$\langle A_2, X \rangle = b_2$

$\vdots$

$\langle A_m, X \rangle = b_m$

$X \succeq 0$

"X is positive semidefinite"

$X = YY^T$

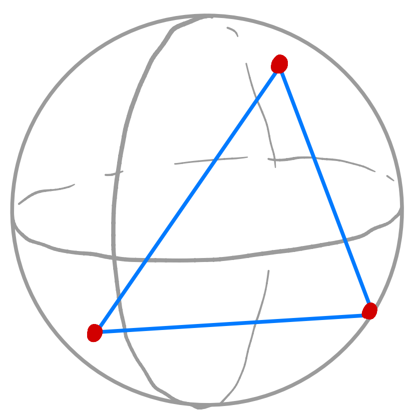
Example: Goemans-Williamson MaxCut SDP

minimize $\langle C, X \rangle$

adjacency matrix of graph

subject to: $X_{ii} = 1 \quad \forall i$

$X \succeq 0 \quad \left(\| Y_i \|_2 = 1 \right)$



n vertices in n dimensions!

- vertex → unit vector
- edge → cost

(edge weight) · (inner product of endpoints)

cos (∡ between endpoints)

"want heavy edges to be longer"

Semidefinite Programs

variable: $X \in \mathbb{R}^{n \times n}$

objective: minimize $\langle C, X \rangle$

subject to: $\langle A_1, X \rangle = b_1$

$\langle A_2, X \rangle = b_2$

$\vdots$

$\langle A_m, X \rangle = b_m$

$X \succcurlyeq 0$

"X is positive semidefinite"

$X = YY^T$

CONVEX!

Example: Goemans-Williamson MaxCut SDP

$$\text{minimize } \langle C, X \rangle$$

adjacency matrix of graph

subject to: $X_{ii} = 1 \quad \forall i$

$X \succcurlyeq 0 \qquad (\|Y_i\|_2 = 1)$

Not just for Max Cut!

- Grothendieck problem
- Quadratic programming
- Community detection

... and more!

SDP Memory Bottleneck

$$\underbrace{X}_{n \times n}$$

$O(n^2)$-size matrix in memory!

<u>Theorem</u> [Barvinok, Pataki '94]

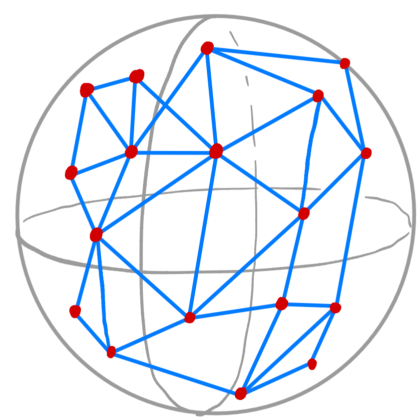
An SDP has an optimal solution of rank at most $\sqrt{2m}$.

m: # of constraints

e.g. $\sqrt{2n}$ for MaxCut

$$\underbrace{X}_{n \times n} = \underbrace{Y}_{n \times \sqrt{2n}} \; \underbrace{Y^T}_{\sqrt{2n} \times n}$$

Many fewer variables!



many more vertices on low-dimensional sphere!

Burer-Monteiro Method

Replace X with YY^T for skinny Y. ← low memory!

$$\begin{aligned} \text{minimize} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \\ & X \succeq 0 \end{aligned} \longrightarrow \begin{aligned} \text{minimize} \quad & \langle C, YY^T \rangle \\ \text{s.t.} \quad & \langle A_i, YY^T \rangle = b_i \\ & Y \in \mathbb{R}^{n \times \sqrt{2n}} \end{aligned}$$

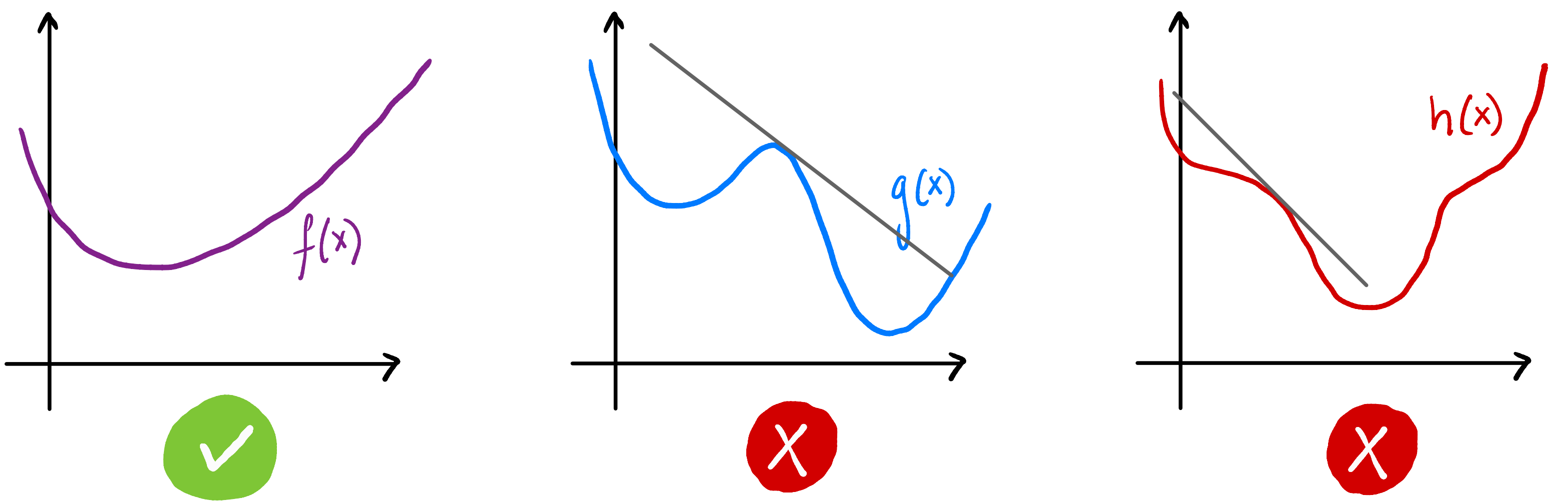
Solve via gradient descent or other local methods

[Burer Monteiro '03]

How SDPs are solved in practice!

NON-CONVEXITY: FRIEND OR FOE?

Select all images that contain convex* functions:



But this one is friendly!
(i.e. no spurious local minima)

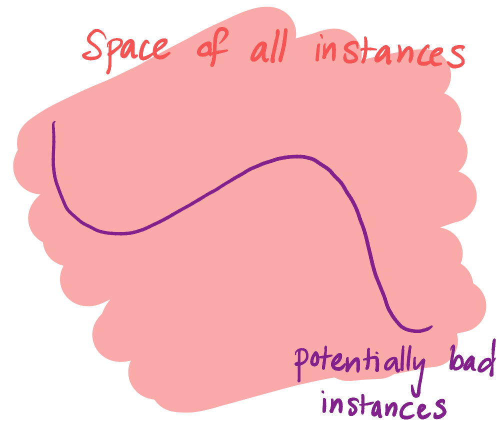
* a convex function always lies above the tangent

So BM is Non-convex, but is it Friendly?

Empirical evidence: the Burer-Monteiro method works surprisingly well!

Theory:

[Boumal Voroninski Bandeira '18] Almost all ("generic") instances of Burer-Monteiro above Barvinok-Pataki are friendly.



[Cifuentes Moitra '22] Algorithmic guarantee:

"if an instance is randomly perturbed by magnitude σ, it will be well-conditioned enough to run local methods in poly$(1/\sigma)$ time."

Smoothed analysis

For Max Cut:

[WW20] Not always friendly

[BVB18] almost always friendly...

[BVB18] always friendly!

rank of X = width of Y

$\sqrt{2n}$

Barvinok-Pataki bound

$\frac{n}{2}$

SDP optimum value = BM optimum value

So BM is Non-convex, but is it Friendly?

For Max Cut:

[WW20] Not always friendly

[BVB18] almost always friendly ...

[OSV 22] ... but not always!

[BVB 18] always friendly!

rank of X = width of Y

$\sqrt{2n}$

Barvinok - Pataki bound

$\frac{n}{2}$

SDP optimum value = BM optimum value

NEW!

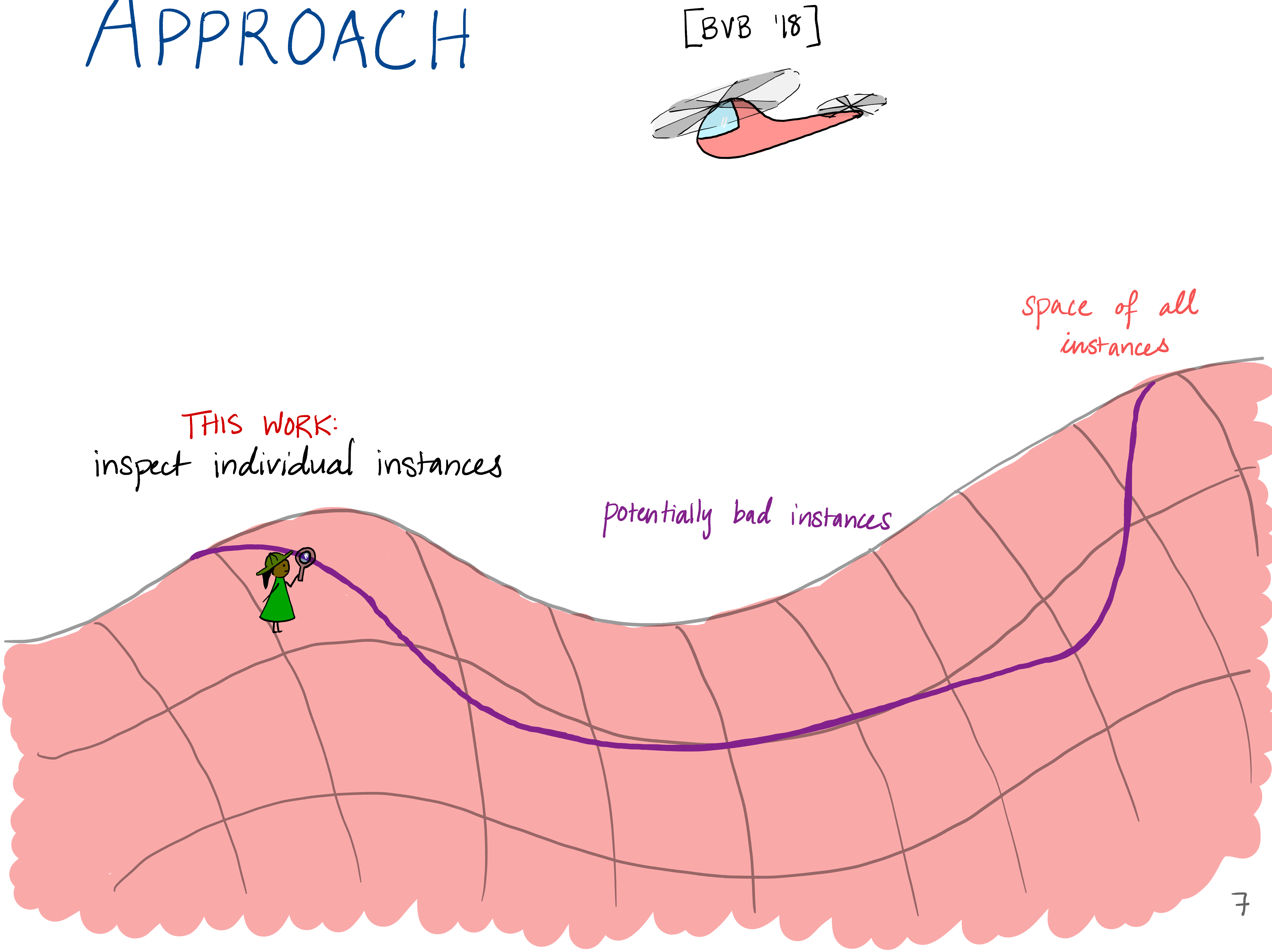
Theorem [O'Carroll, **S.**, Vijayaraghavan '22]

For every n, there exists a Burer-Monteiro instance with spurious local minima for all settings of rank $p \leq \frac{n}{2}$.

Surprisingly, can construct Max Cut instance!

(Recall title: "The Burer-Monteiro SDP method can fail, even above the Barvinok-Pataki bound")

APPROACH

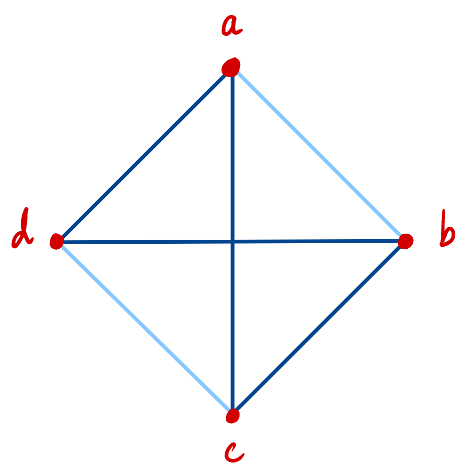


[BVB '18]
space of all instances
THIS WORK:
inspect individual instances
potentially bad instances

WHAT DOES AN INSTANCE LOOK LIKE?

OBJECTIVE

Cost matrix C (a graph)

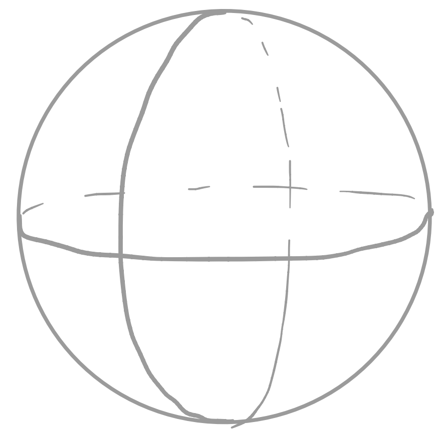


minimize $\langle C, YY^T \rangle$

minimize $\sum_{\text{edges}}$ (edge weight) (cos ∡ between endpoints)

CONSTRAINTS

Place vertices on unit sphere $\in \mathbb{R}^p$



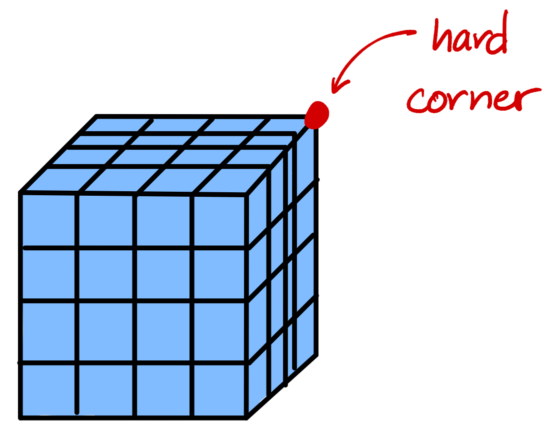
Each row of Y has $\|\cdot\|_2 = 1$

WANT TO CONSTRUCT: a C and a Y, so Y is a spurious local minimum for C

how to optimize with constraints?

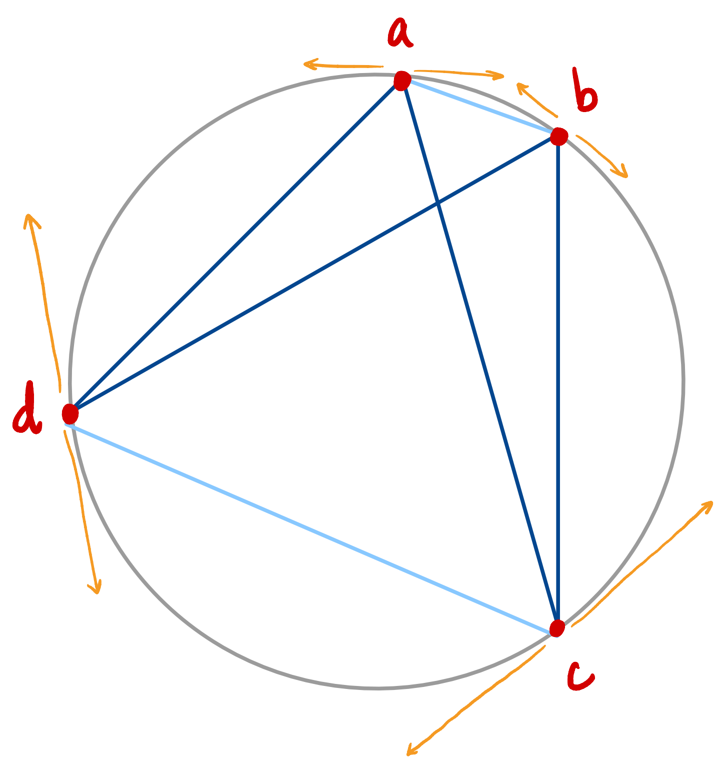
HOW TO OPTIMIZE?

Select all images that contain Riemannian manifolds*:



* at every point: "if you zoom-in close enough it looks like Euclidean space."

GRADIENTS AND HIGHER MOMENTS



Tangent space : Set of valid directions $\vec{u}$ on the manifold

Riemannian gradient : first-derivative in directions $\vec{u}$ in tangent space

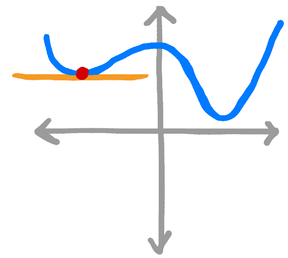
Riemannian Hessian : second-derivative in directions $\vec{u}$ in tangent space

CALCULUS REVIEW

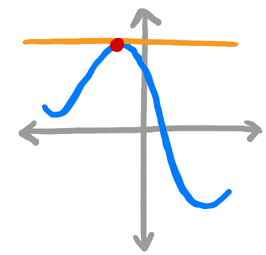
WANT TO SHOW: a point is a local minimum

NECESSARY CONDITIONS:

Objective function



(not sufficient!



)

① gradient $= 0$

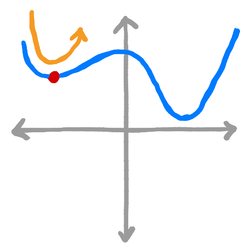
i.e. first-derivative in all directions $= 0$

"first-order critical"

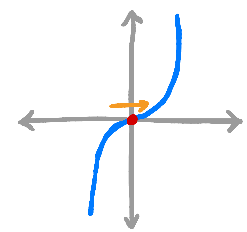
② Hessian $\succcurlyeq 0$

i.e. second-derivative in all directions ≥ 0

"second-order critical"



(not sufficient!

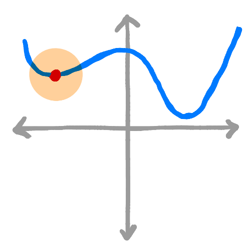


)

would need strict > 0

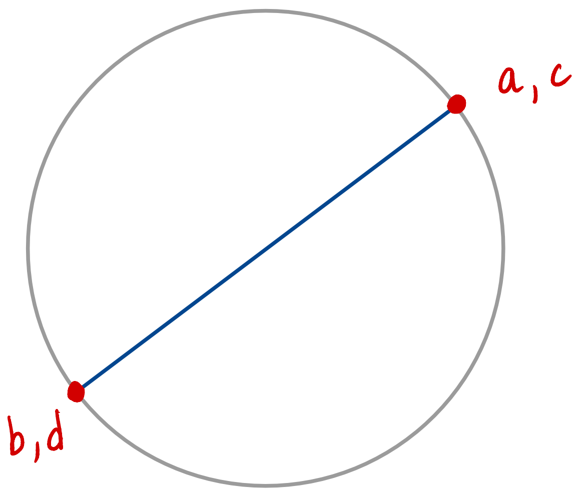
NEED:

point is at least as good as all other points in its neighborhood.

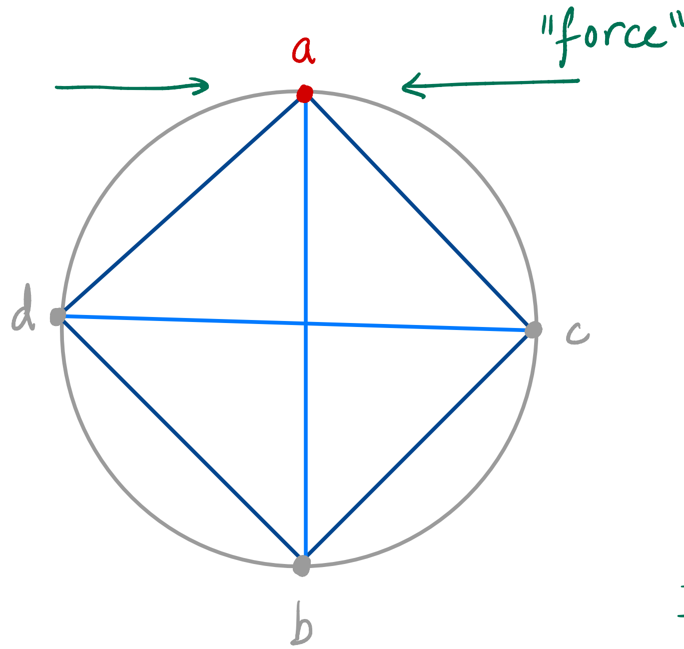


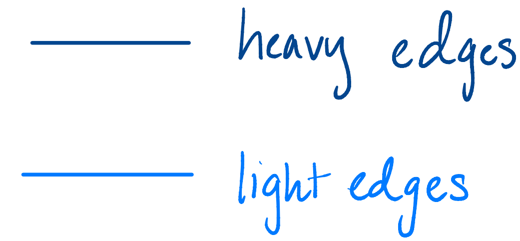
FIRST-ORDER CRITICALITY

GLOBAL MIN



SPURIOUS FIRST-ORDER CRITICAL





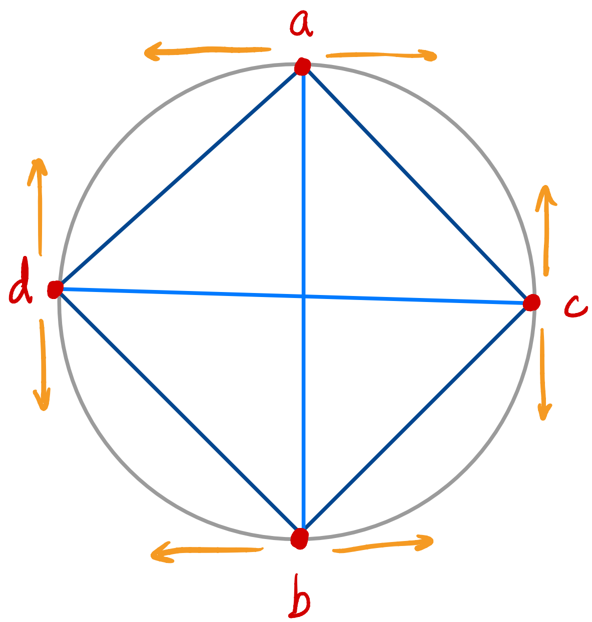
gradient is 0!

OBJECTIVE: minimize $\langle C, YY^T \rangle$ (graph: C)

$$\sum_{\text{edges}} (\text{edge weight})(\cos \measuredangle \text{ between endpoints})$$

"want heavy edges to be longer"

SECOND-ORDER CRITICALITY



heavy edges

light edges

Condition:

$$\text{second derivative} = \left\langle \begin{bmatrix} C & C \\ C & C \end{bmatrix}, u \, u^T \right\rangle \geq 0$$

for valid directions u (tangent space)

"looks" positive semidefinite to Y

Fact:

positive semidefinite

$\left\langle \begin{bmatrix} C & C \\ C & C \end{bmatrix}, uu^T \right\rangle \geq 0$ for all directions

$\Longrightarrow$

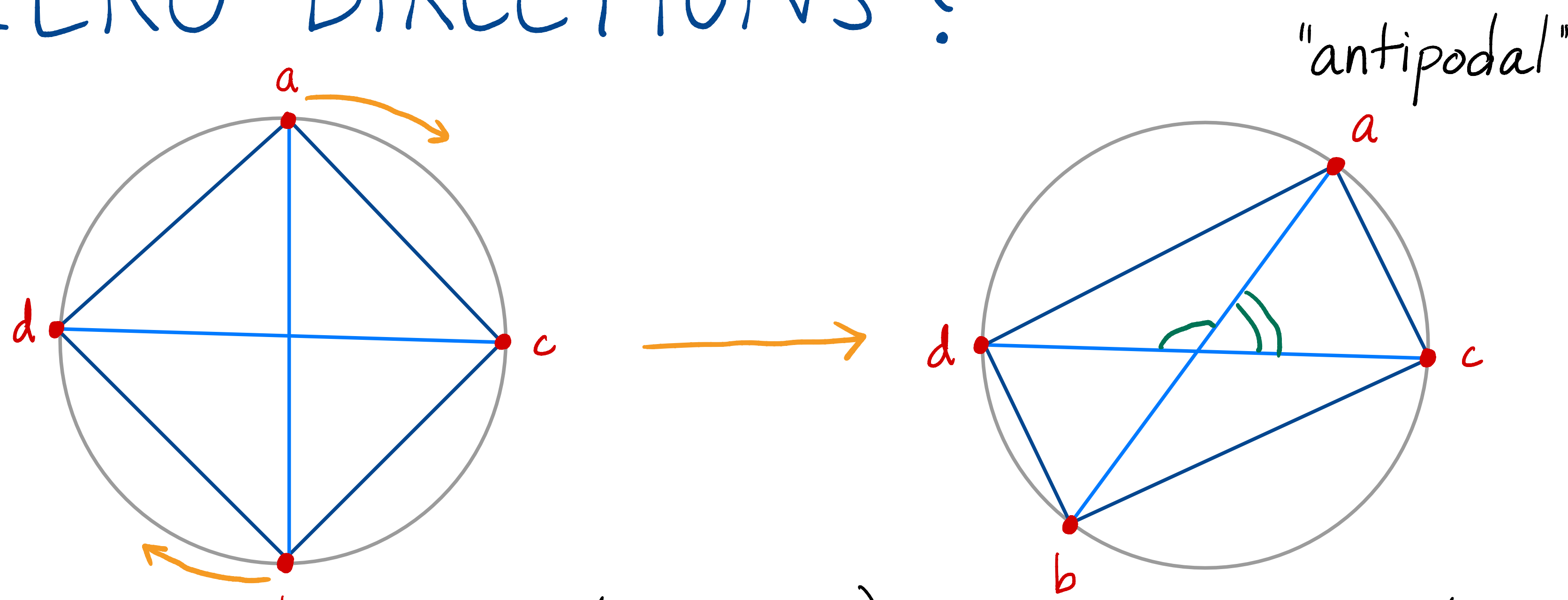
this point is globally optimal

✓ choose instance C to hide negative directions in directions Y "can't see"

C is "pseudo-PSD"

– there are still 0-directions for all higher-order moments!

ZERO-DIRECTIONS ?



OBJ = 2 (light edge)

OBJ = −2 (light edge)

Same objective value!

OBJECTIVE: minimize $\langle \overset{\text{graph}}{C}, YY^T \rangle$

$\sum_{\text{edges}}$ (edge weight) (cos $\measuredangle$ between endpoints)

"want heavy edges to be longer"

must explicitly account for flat directions of objective value

all higher-order moments are degenerate

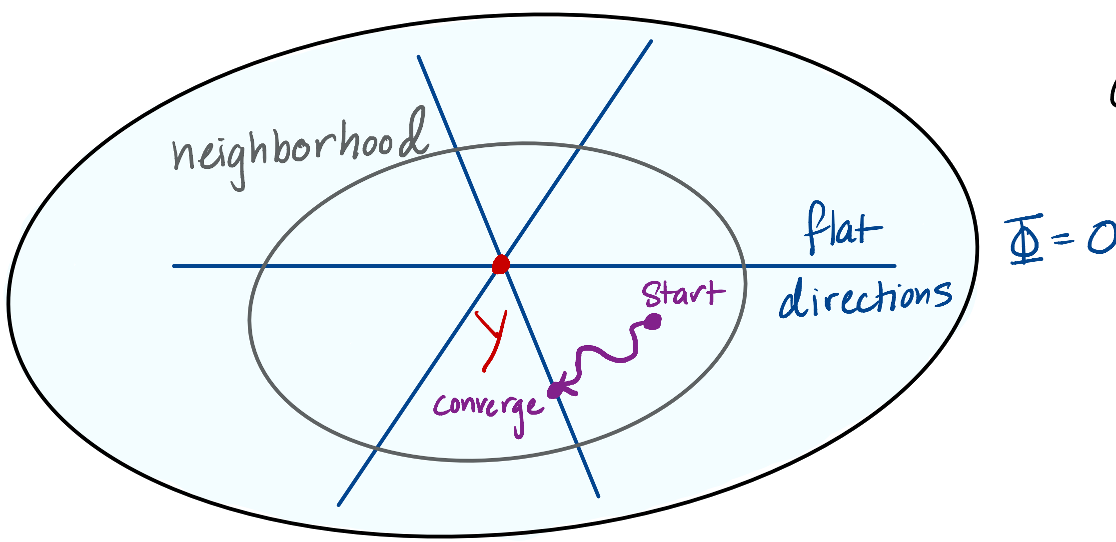
PROOF SKETCH

argue about objective value via Riemannian gradient descent

~~WANT: initialize close to Y ⟿ converge to Y~~

WANT: initialize close to Y ⟿ converge to transformation of Y

define $\Phi(\cdot)$ = how close to antipodal (same objective value as Y)



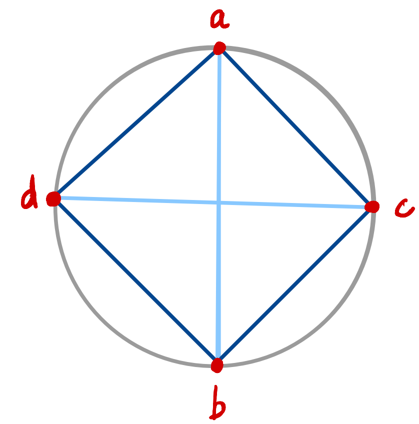
manifold for this instance

objective value only decreases with grad. descent

Φ decreases to 0 with grad. descent

OBJ(start) $\geq$ OBJ(Y)

WRAP UP



PREV. KNOWN:

- Burer-Monteiro works great in practice!
- Burer-Monteiro works on smoothed instances

THIS WORK:

- Burer-Monteiro can fail!
- NEW! way to argue about local minima with flat directions

FUTURE DIRECTIONS:

- Full characterization of spurious local minima for BM
- Spurious local minima (or lack of) for other optimization problems

THANKS!

Check out visualizer: vaidehi8913.github.io/burer-monteiro