THE UNIVERSITY OF TEXAS AT AUSTIN

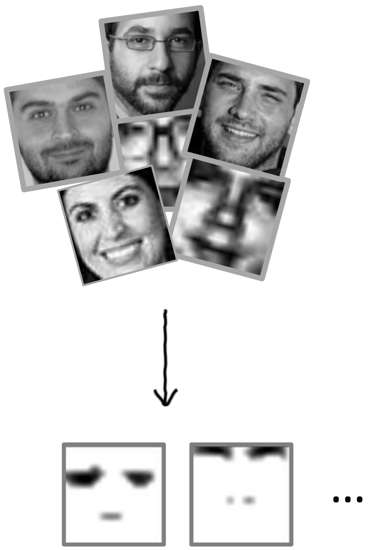
Nonnegative Sparse PCA

with Provable Guarantees

Megasthenis Asteris
Dimitris Papailiopoulos
Alex Dimakis

Nonnegative Sparse PCA

Input: set of data points (e.g. images)
multiple features (e.g. pixels)



sparse

nonnegative

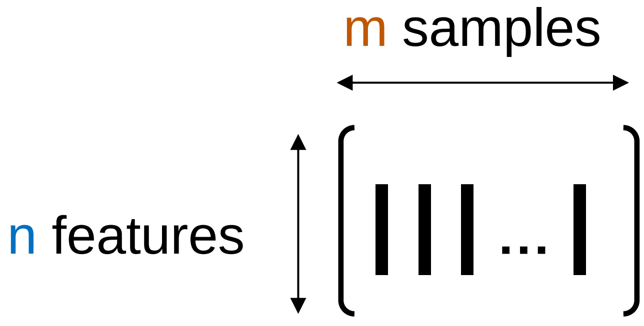
Goal: find few features
with positive influences
that capture most variance.

principal component

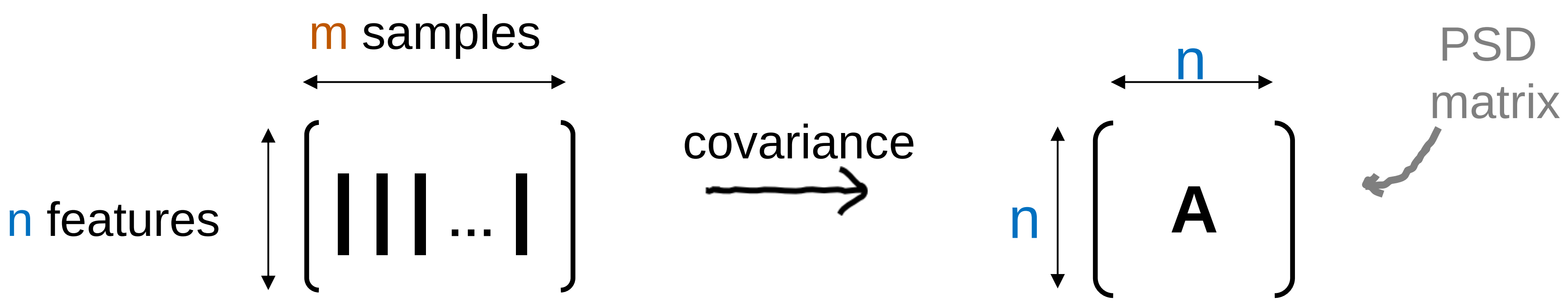
- Bioinformatics: chemical concentrations; gene expression
- Computer vision: extraction of image parts

[Lee & Seung, '99], [Kim & Park, '07] , [Badea & Tilivea, '05]

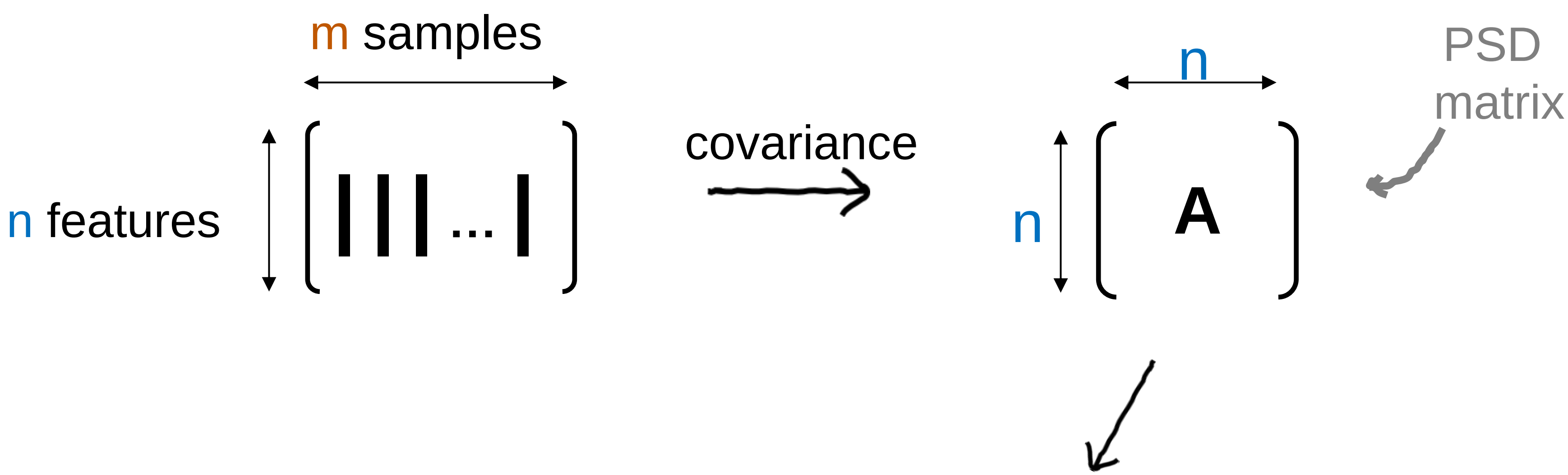
Nonnegative Sparse PCA



Nonnegative Sparse PCA



Nonnegative Sparse PCA



(NNSPCA) $$\max_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A}\, \mathbf{x}$$

Sparse: $\|\mathbf{x}\|_0 \leq k$

Nonnegative: $\mathbf{x} \geq 0$

Nonnegative Sparse PCA

- **NP hard**:

  Sparse: best support?

  Nonnegative: even if support is known…

  [Murty & Kabadi, '87] [Parrilo, '00]

- Few algorithms: **heuristics**, **no guarantees**

  [Lee & Seung, '99] [Zass & Shashua, '07] [Sigg & Buhmann, '08]

Today:

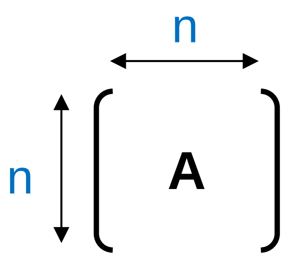
New algorithm,

…with guarantees!

Our NNSPCA Algorithm

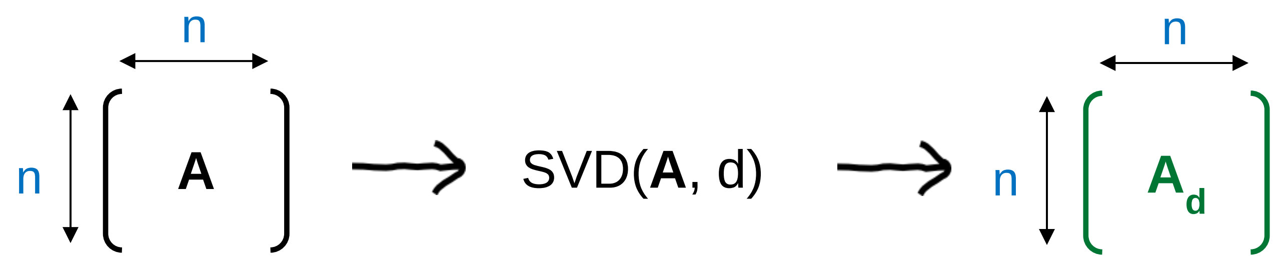
(NNSPCA) $$\max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Our NNSPCA Algorithm



$$\text{(NNSPCA)} \quad \max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} \mathbf{x}^T \mathbf{A}\, \mathbf{x}$$

Our NNSPCA Algorithm



$$\text{(NNSPCA)} \qquad \max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \leq k \\ \mathbf{x} \geq 0}} \ \mathbf{x}^T \mathbf{A}\, \mathbf{x}$$

Our NNSPCA Algorithm

n × n $\mathbf{A}$ → SVD($\mathbf{A}$, d) → n × n $\mathbf{A}_d$

(NNSPCA)

low rank

$$\max \ \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (\mathbf{A} \to \mathbf{A}_d)$$

$$\|\mathbf{x}\|_2 = 1$$

$$\|\mathbf{x}\|_0 \leq k$$

$$\mathbf{x} \geq 0$$

Approximation Results

Thm 1: For any parameter d, our algorithm outputs a nonnegative k-sparse vector $\mathbf{x}_d$ such that

$$\mathbf{x}_d^T \mathbf{A}\, \mathbf{x}_d \geq \rho\ \mathbf{x}_*^T \mathbf{A}\, \mathbf{x}_*$$

OPT

where

$$\rho = \left(1 + 2 * \frac{n}{k} * \frac{\lambda_{d+1}}{\lambda_1} \right)^{-1}$$

in $O(n^{2d})$.

Approximation Results

Thm 2: For any parameter d, our algorithm outputs a nonnegative k-sparse vector x_d such that

$$\mathbf{x_d}^T \mathbf{A}\, \mathbf{x_d} \geq (1-\delta)\, \rho\, \mathbf{x_*}^T \mathbf{A}\, \mathbf{x_*}$$

OPT

where

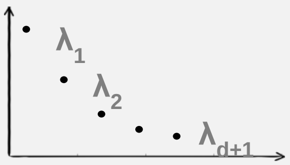
$$\rho = \left(1 + 2 * \frac{n}{k} * \frac{\lambda_{d+1}}{\lambda_1}\right)^{-1}$$

~~in $O(n^{2d})$.~~

in $O(\delta^{-d}\, n\log n) + T_{SVD}$

Approximation Results

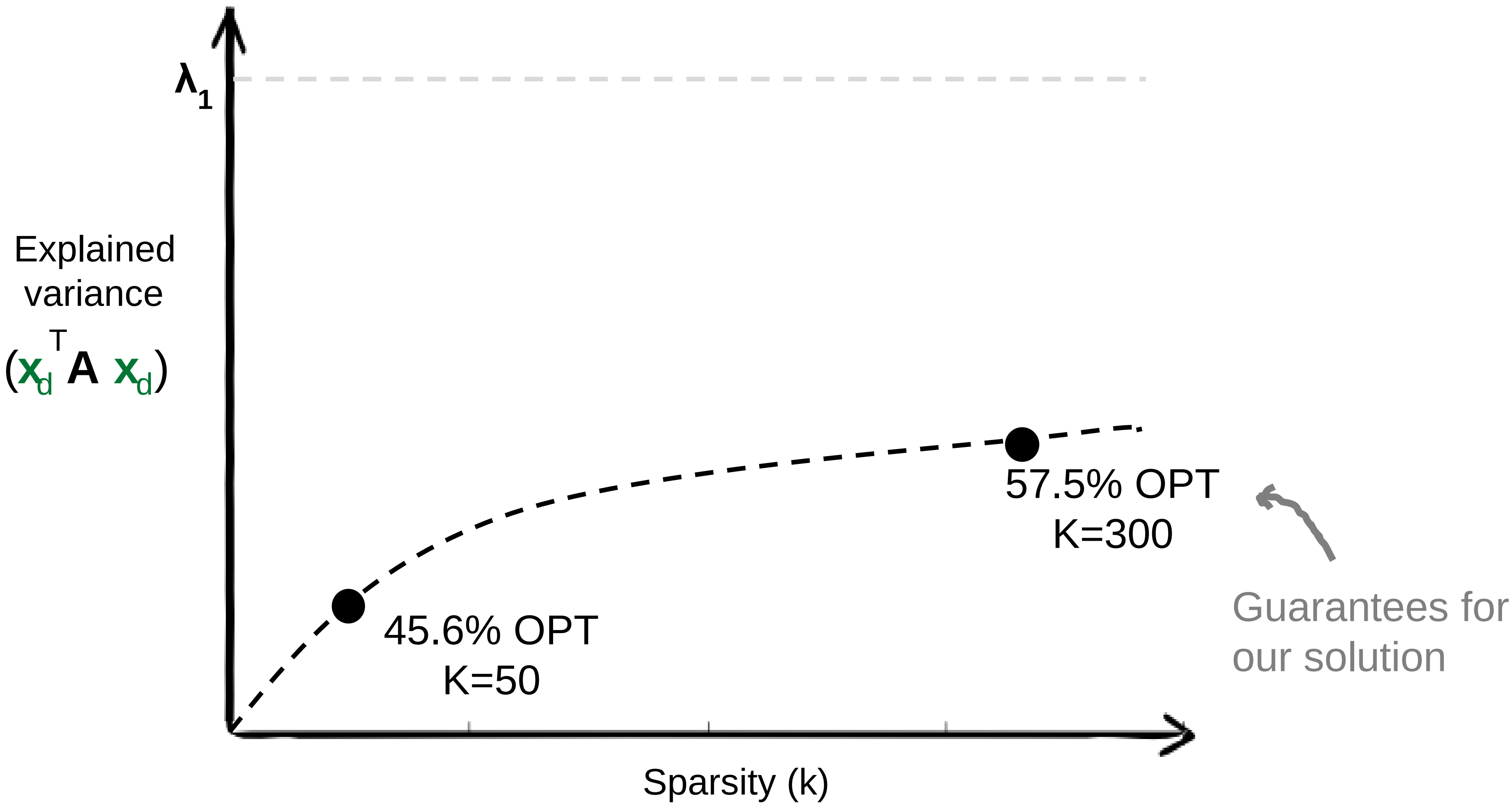
Cor. 1 : (1-ε) * OPT approximation
for <u>any</u> vanishing eigenvalue decay function
in poly(n), but not 1/ ε



Cor. 2 : Data dependent upper bound on OPT
Real datasets: 40% – 90% OPT

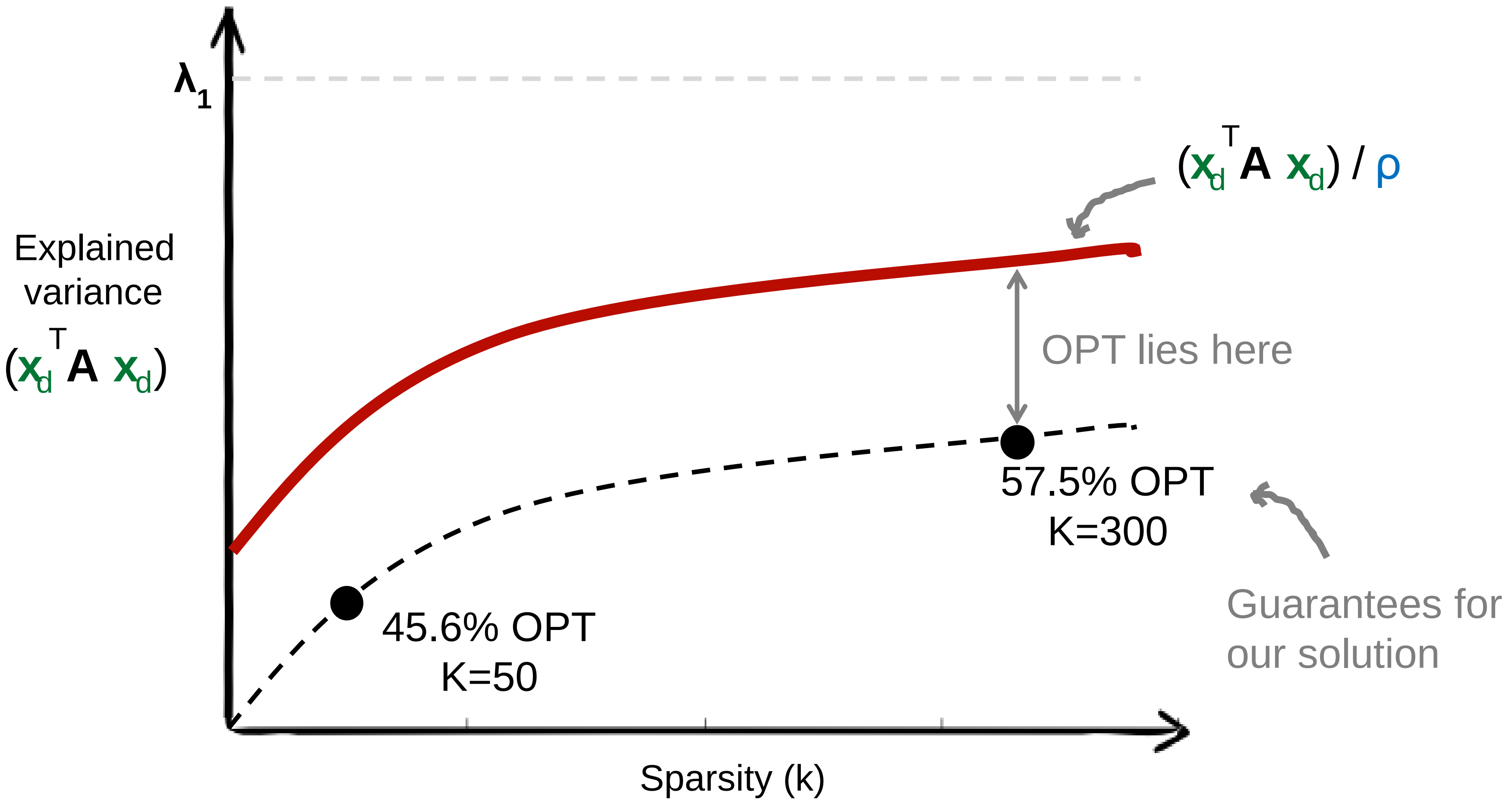
Approximation Results

Example: Leukemia dataset ~12,000 features [UCI]



Approximation Results

Example: Leukemia dataset ~12,000 features [UCI]



The low rank solver

Rank($\mathbf{A_d}$) = 1

$$\max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} \mathbf{x}^T \mathbf{A_d} \mathbf{x}$$

The low rank solver

Rank($\mathbf{A_d}$) = 1

$$\max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} \mathbf{x}^T \mathbf{A_d}\, \mathbf{x} \quad \xrightarrow{\mathbf{A_d} = \mathbf{v}\mathbf{v}^T} \quad \max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} (\mathbf{x}^T \mathbf{v})^2$$

The low rank solver

Rank($\mathbf{A}_d$) = 1

$$\max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} \mathbf{x}^T \mathbf{A}_d \mathbf{x} \quad \xrightarrow{\mathbf{A}_d = \mathbf{v}\mathbf{v}^T} \quad \max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le k \\ \mathbf{x} \ge 0}} (\mathbf{x}^T \mathbf{v})^2$$

Example:

$$\max_{\substack{\|\mathbf{x}\|_2 = 1 \\ \|\mathbf{x}\|_0 \le 2 \\ \mathbf{x} \ge 0}} \left(\mathbf{x}^T \overset{\mathbf{v}}{\begin{bmatrix} 5 \\ 3 \\ -2 \\ -4 \end{bmatrix}} \right)^2$$

The low rank solver

$\text{Rank}(\mathbf{A}_d) = 1$

$$\max \; \mathbf{x}^T \mathbf{A}_d \mathbf{x} \quad \text{s.t.} \; \|\mathbf{x}\|_2 = 1,\; \|\mathbf{x}\|_0 \le k,\; \mathbf{x} \ge 0$$

$\mathbf{A}_d = \mathbf{v}\mathbf{v}^T$ $\longrightarrow$

$$\max \; (\mathbf{x}^T \mathbf{v})^2 \quad \text{s.t.} \; \|\mathbf{x}\|_2 = 1,\; \|\mathbf{x}\|_0 \le k,\; \mathbf{x} \ge 0$$

Example:

$$\max \; \left(\mathbf{x}^T \underbrace{\begin{bmatrix} 5 \\ 3 \\ -2 \\ -4 \end{bmatrix}}_{\mathbf{v}}\right)^2 \quad \text{s.t.} \; \|\mathbf{x}\|_2 = 1,\; \|\mathbf{x}\|_0 \le 2,\; \mathbf{x} \ge 0$$

k largest pos. entries of **v** →

$$\mathbf{X}^+ = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 0 \end{bmatrix} / \sqrt{34}$$

k smallest neg. entries of **v** →

$$\mathbf{X}^- = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 4 \end{bmatrix} / \sqrt{20}$$

Output the best…

The low rank solver

Rank($\mathbf{A}_d$) = 2, 3, … d

The low rank solver

Rank($\mathbf{A}_d$) = 2, 3, … d

1. Find possible supports for **x**

2. For each support, find **x**

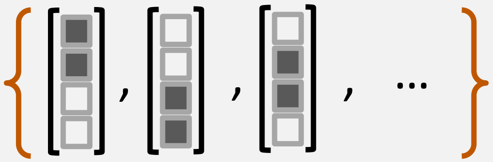
3. Output best candidate, $\mathbf{x}_d$.

The low rank solver

Rank($\mathbf{A}_d$) = 2, 3, … d

1. Find possible supports for $\mathbf{x}$ [Asteris, Papailiopoulos, Karystinos, '11]

$\mathbf{A}_d$ "Spannogram" ⟶ supp($\mathbf{x}$) in



$O(n^d) \ll O(n^k)$ candidate supports

2. For each support, find $\mathbf{x}$

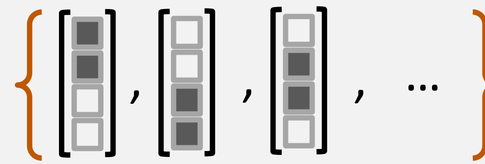
3. Output best candidate, $\mathbf{x}_d$.

The low rank solver

Rank($\mathbf{A}_d$) = 2, 3, … d

1. Find possible supports for $\mathbf{x}$ [Asteris, Papailiopoulos, Karystinos, '11]

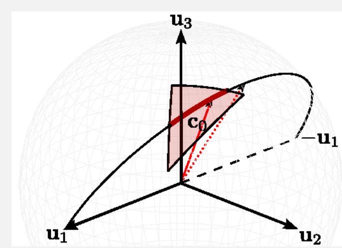
$\mathbf{A}_d$ "Spannogram" ⟶ supp($\mathbf{x}$) in



$O(n^d) \ll O(n^k)$ candidate supports

2. For each support, find $\mathbf{x}$

$$\max_{\substack{\|\mathbf{c}\|_2 = 1 \\ \mathbf{R}\mathbf{c} \geq \mathbf{0}}} \mathbf{c}^T \mathbf{Q} \mathbf{c}$$



NP-Hard, but d-dimensional

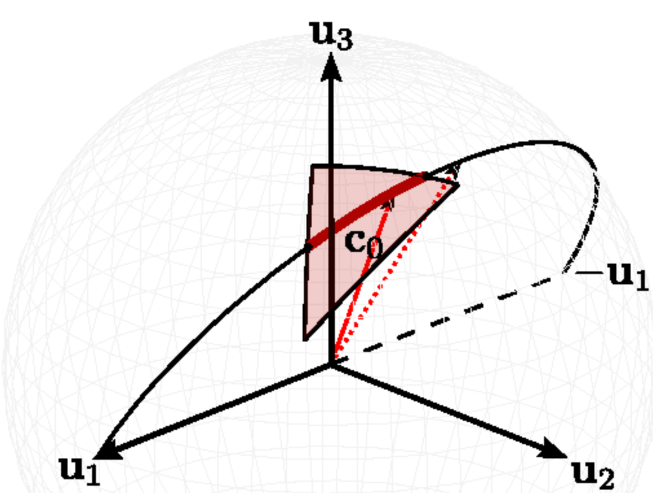
3. Output best candidate, $\mathbf{x}_d$.

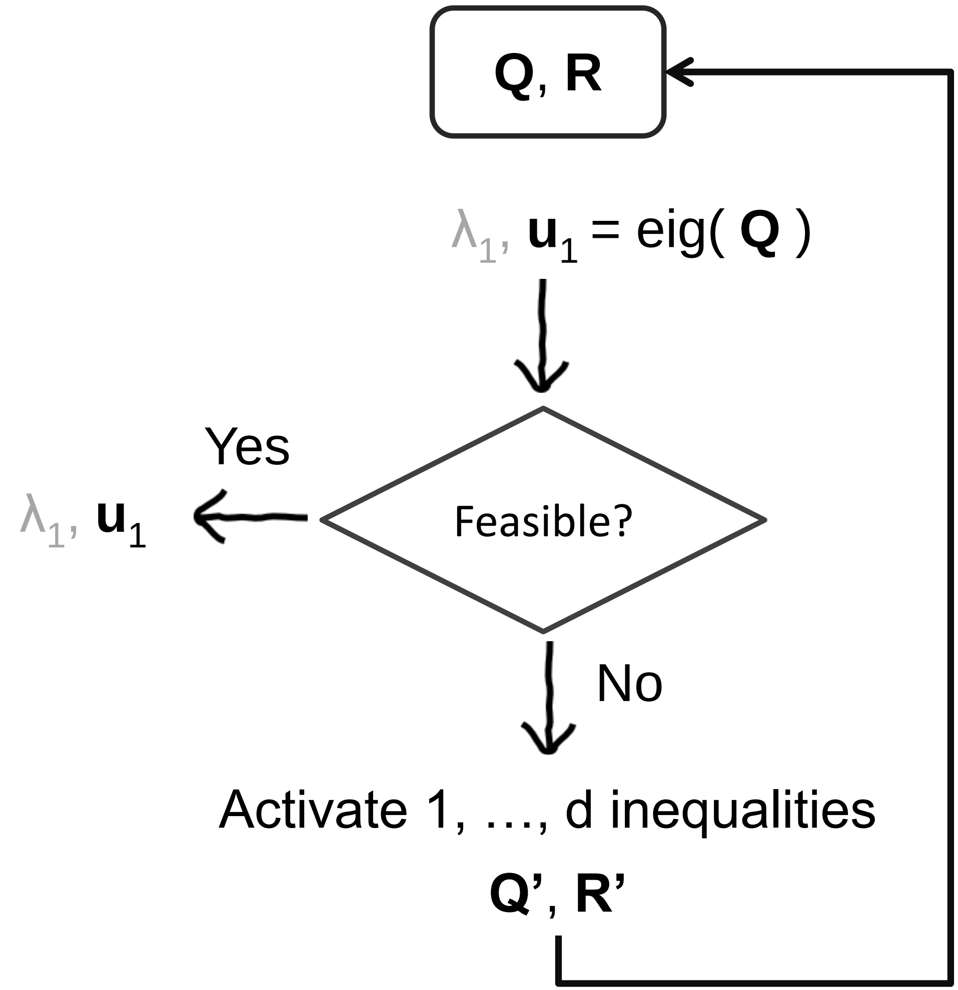
The low rank solver

$$\max \ \mathbf{c}^T \mathbf{Q} \, \mathbf{c}$$

$$\|\mathbf{c}\|_2 = 1$$

$$\mathbf{R} \, \mathbf{c} \geq \mathbf{0}$$





Summary

- **New** combinatorial approximation **algorithm** for NNSPCA
- Data dependent performance **guarantees**
- Relies on solving NNSPCA on a **low rank** matrix in **almost linear time**
- Arbitrary approximation factor for any vanishing eigenvalue decay

$$\text{(NNSPCA)} \quad \max \ \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$$\|\mathbf{x}\|_2 = 1$$

Sparse: $\|\mathbf{x}\|_0 \leq k$

Nonnegative: $\mathbf{x} \geq 0$