# Rank/sparsity minimization and its convex algebraic geometry

**Pablo A. Parrilo**

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

BIRS Workshop on Convex Algebraic Geometry
February 2010 – Banff

# What I will not talk about (directly)

- Theta bodies and SDP relaxations (w/Gouveia, Laurent, Thomas,...)
- Convexity and SOS-convexity (Ahmadi, Helton/Nie, Blekherman,...)

Please see their talks/ask them!

# First, my coauthors...

**Ben Recht**
(UW-Madison)

**Maryam Fazel**
(U. Washington)

**Venkat Chandrasekaran**
(MIT)

**Sujay Sanghavi**
(UT Austin)

**Alan Willsky**
(MIT)

# Rank minimization

**PROBLEM:** Find the *lowest rank* matrix in a given convex set.

- E.g., given an affine subspace of matrices, find one of lowest possible rank
- In general, NP-hard (e.g., reduction from max-cut, sparsest vector, etc.).

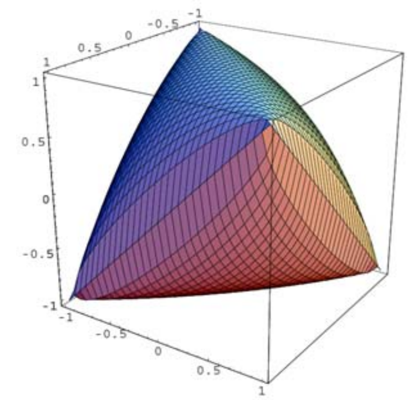
*Many* applications...

# Application 1: Quadratic optimization

- Boolean quadratic minimization (e.g., max-cut)

$$\min_{x_i \in \{-1,1\}^n} x^T Q x$$

- Relax using the substitution *X* := *xx*$^T$:



$$\min_{X \succeq 0,\, X_{ii}=1} \mathrm{Tr}\, Q X$$

- If solution *X* has *rank 1*, then we solved the original problem!

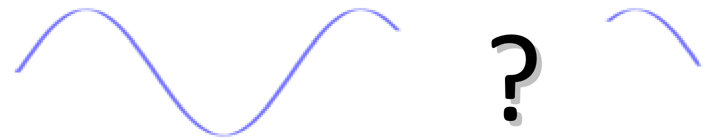
# Application 2: Sum of squares

$$P(x) = \sum_{i=1}^{r} q_i^2(x)$$

- Number of squares equal to the rank of the Gram matrix
- How to compute a SOS representation with the minimum number of squares?
- Rank minimization with SDP constraints

# Application 3: Video inpainting

*(Ding-Sznaier-Camps, ICCV 07)*

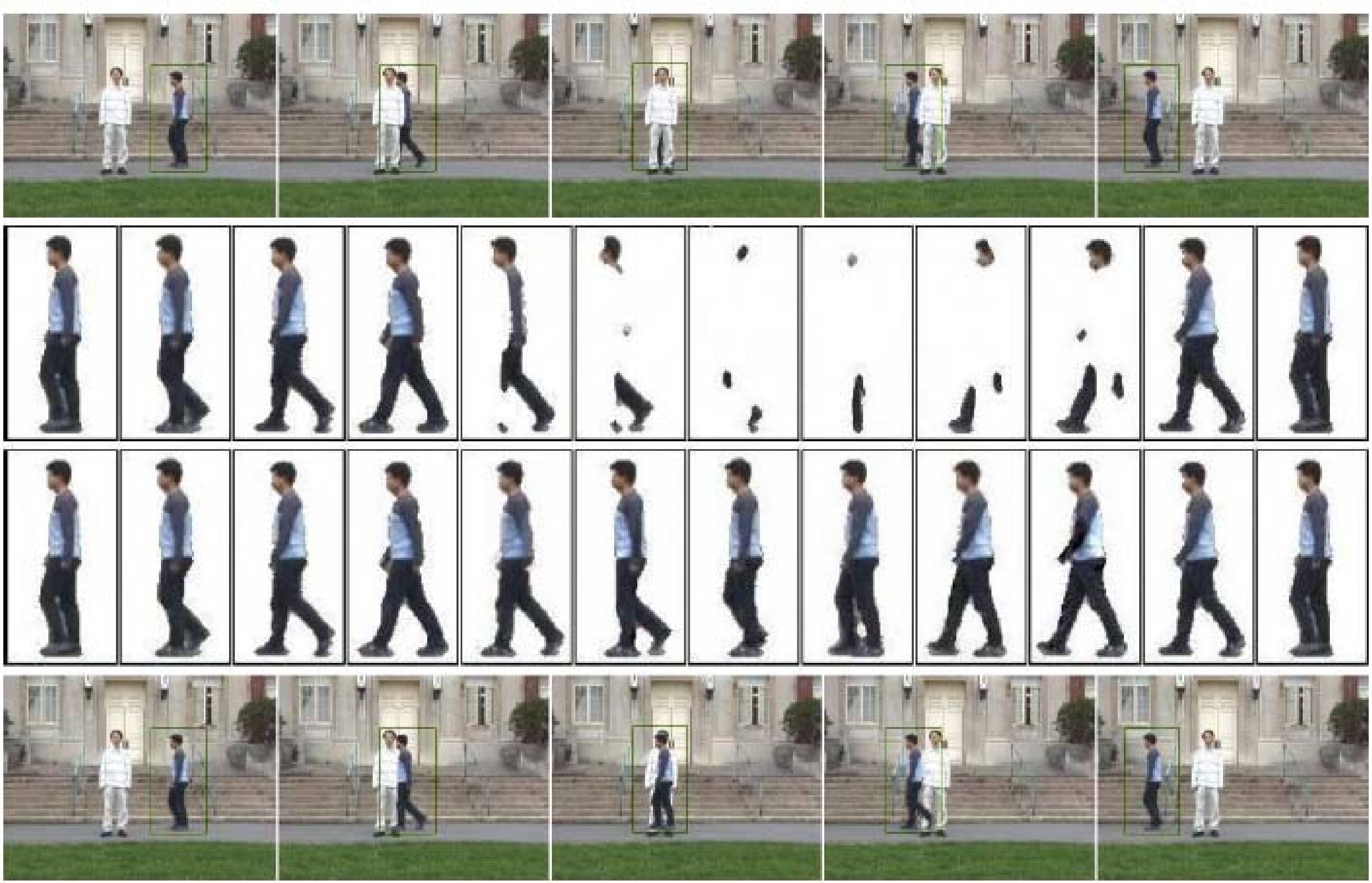



- Given video frames with missing portions


- Reconstruct / interpolate the missing data

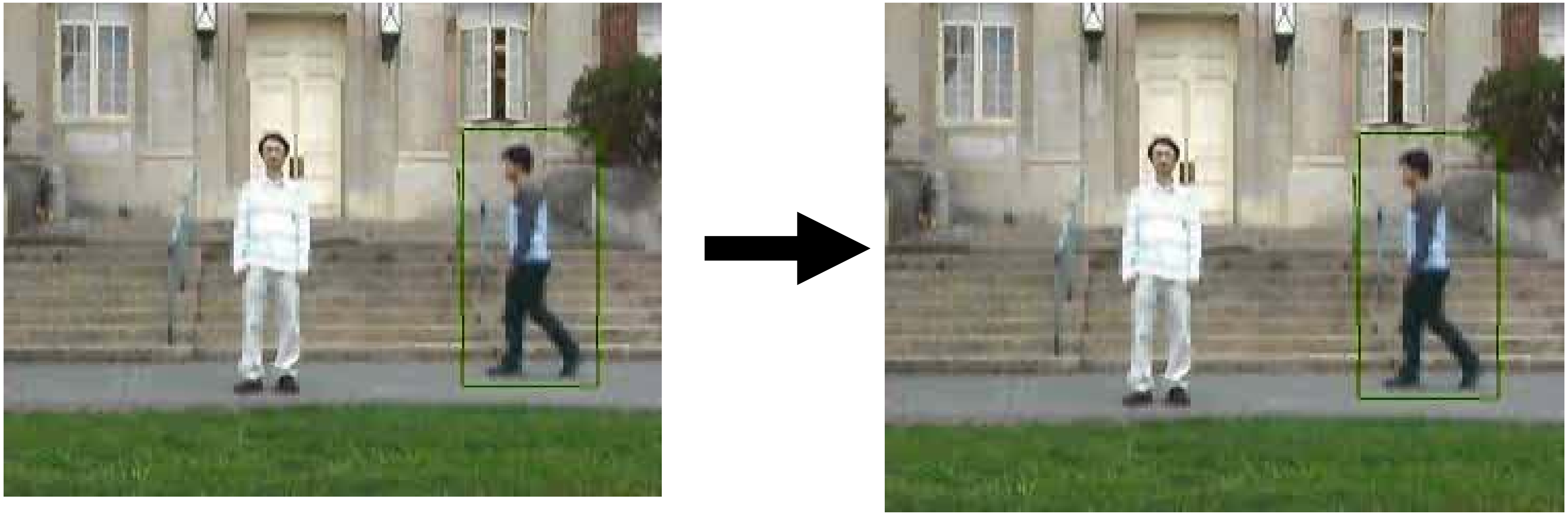
- “Simple” dynamics <-> Low rank (multi) Hankel

# Application 3: Video inpainting (cont)



(Ding-Sznaier-Camps, ICCV 07)

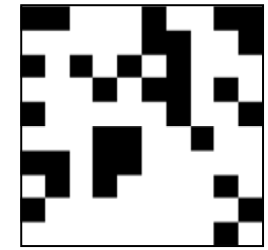
# Application 3: Video inpainting (cont)

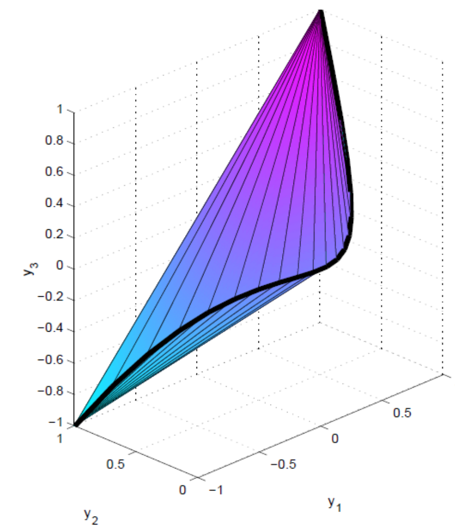


*(Ding-Sznaier-Camps, ICCV 07)*

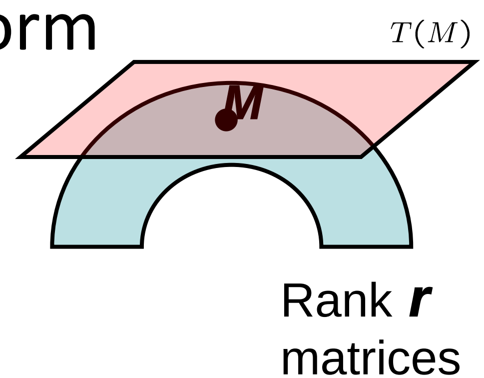
# Overview

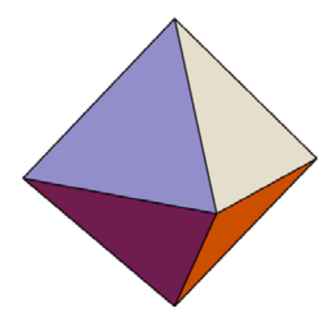
Rank optimization is ubiquitous in optimization, communications, and control. Difficult, even under linear constraints.






- Applications, formulation
- What is the underlying geometry?
- *Convex hulls of varieties*
- Geometry and nuclear norm



- Sparsity + Rank

# Rank minimization

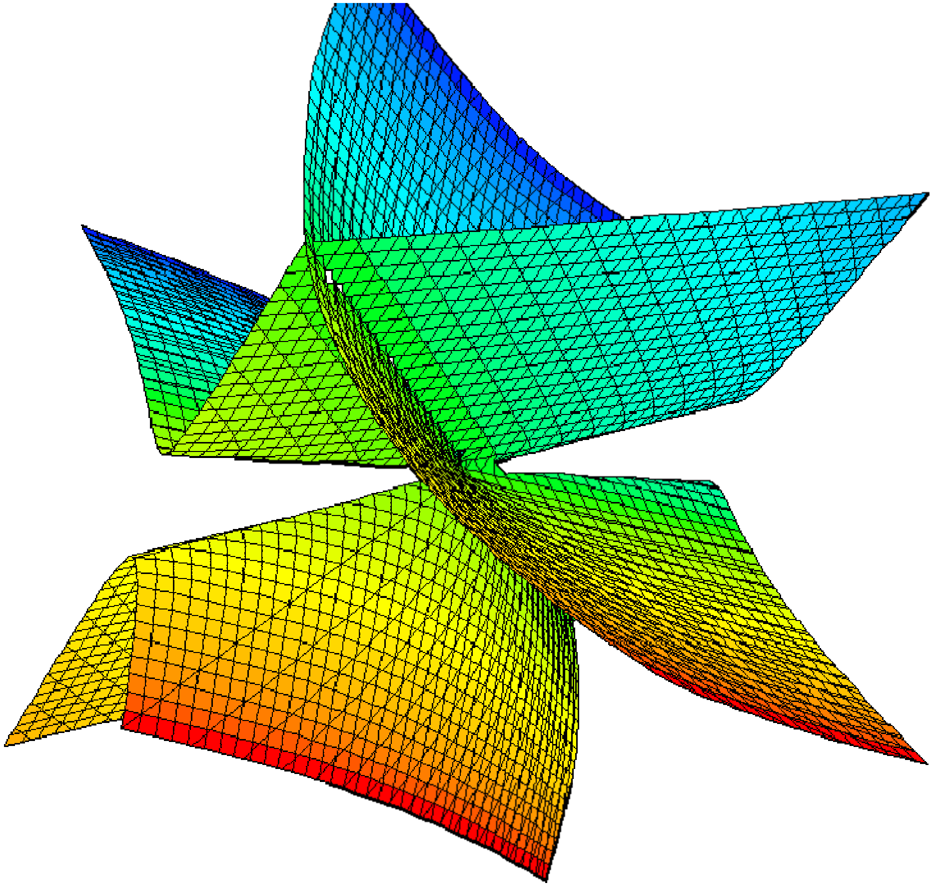
**PROBLEM:** Find the matrix of smallest rank that satisfies the underdetermined linear system:

$$\mathcal{A}(X) = b \qquad \mathcal{A} : \mathbb{R}^{n \times m} \to \mathbb{R}^p$$

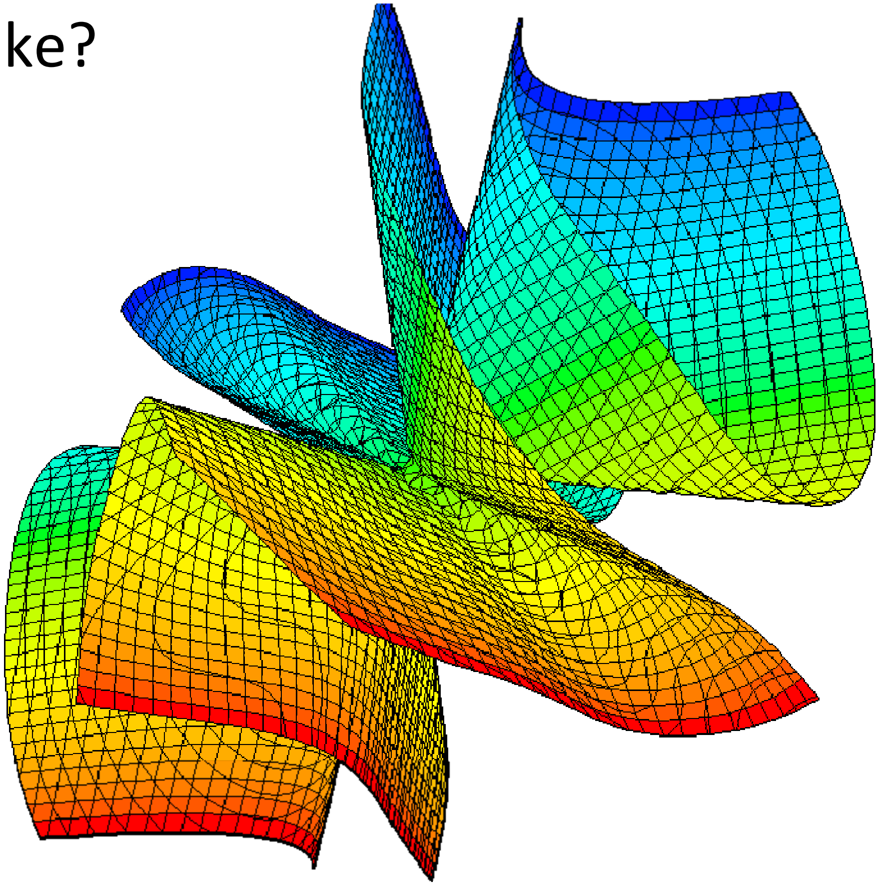
- Given an affine subspace of matrices, find one of lowest rank

# Rank can be complicated...

What does a rank constraint look like?



(section of)
3x3 matrices, rank 2

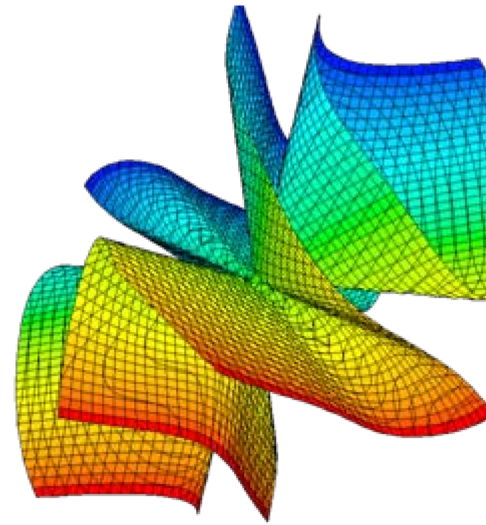


(section of)
7x7 matrices, rank 6

# How to solve this?

*Many* methods have been proposed (after all, it's an NLP!)

- Newton-like local methods
- Manifold optimization
- Alternating projections
- Augmented Lagrangian
- Etc...



Sometimes (often?) work very well.

But, very hard to prove results on *global* performance.

# Geometry of rank varieties

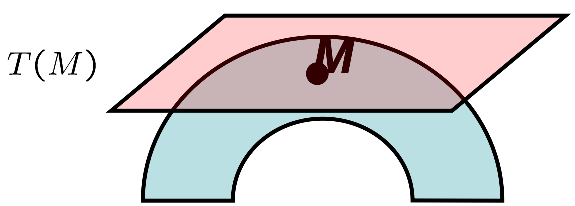
What is the structure of the set of matrices of fixed rank?

$$\mathcal{P}(k) \triangleq \{M \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(M) \leq k\}.$$

An *algebraic variety* defined by the vanishing of all (k+1) ≀ (k+1) minors of *M*.

Its dimension is *k* x *(2n-k),* and is nonsingular, except on those matrices of rank less than or equal to *k-1*.

At smooth points $M = U\Sigma V^T$ well-defined tangent space:




Rank ***k*** matrices

$$T(M) = \{UX^T + YV^T \mid X, Y \in \mathbb{R}^{n\times k}\}.$$

# Nuclear norm

- *Sum of the singular values* of a matrix

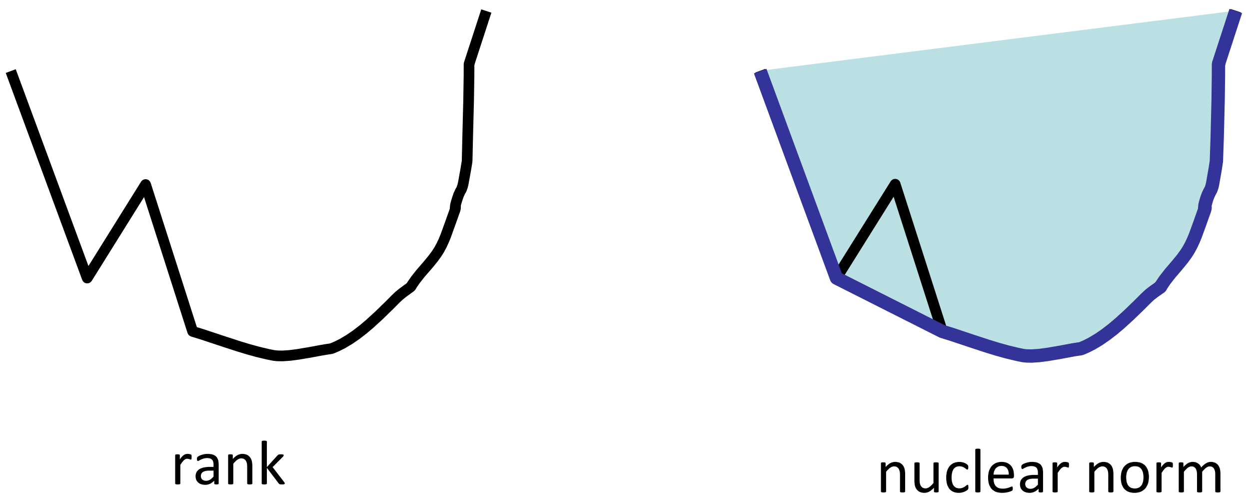
$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X)\,, \qquad \sigma_i(X) := \sqrt{\lambda_i(X^T X)}$$

- Unitarily invariant $\quad ||UXV||_* = ||X||_*$

- Also known as *Schatten 1-norm*, *Ky-Fan r-norm*, *trace norm*, etc...

# Why is the nuclear norm relevant?

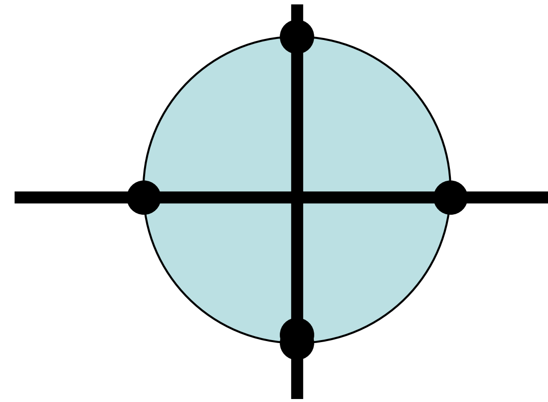
- Bad nonconvex problem -> Convexify!
- Nuclear norm is “best” convex approximation of rank



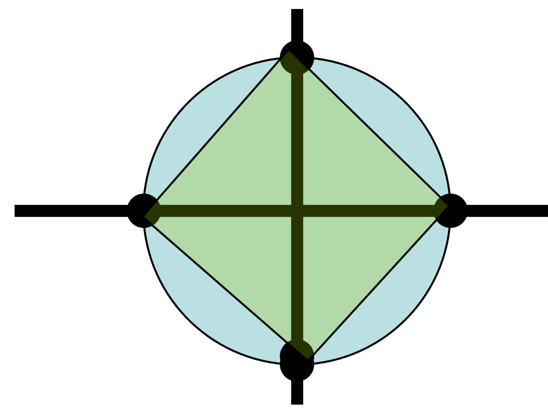
# Comparison with sparsity

Consider sparsity minimization

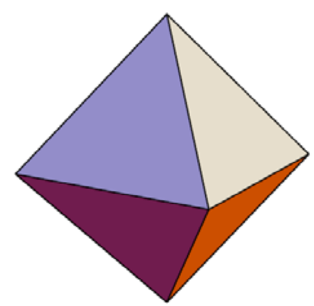
- Geometric interpretation



- Take "sparsity 1" variety
- Intersect with unit ball
- Take convex hull

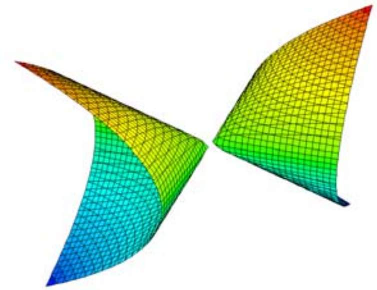


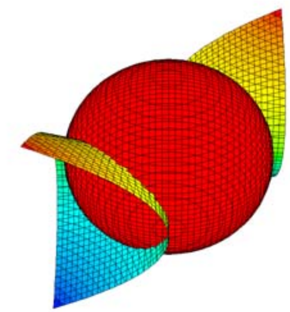
**L1 ball! (crosspolytope)**



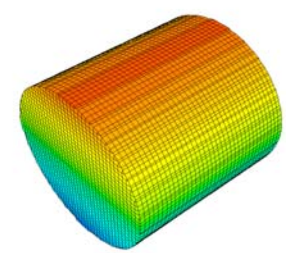
# Nuclear norm

- Same idea!





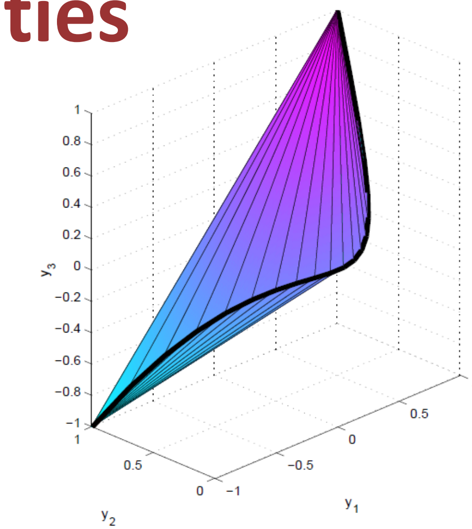
- Take “rank 1” variety
- Intersect with unit ball
- Take convex hull



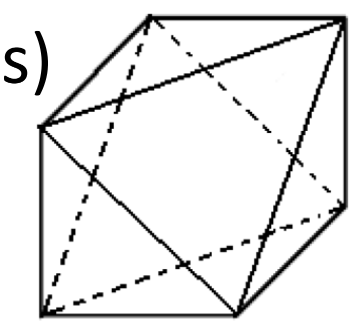
$$\mathrm{conv}\{uv^T \;:\; u \in \mathbb{R}^n, v \in \mathbb{R}^m, ||u|| = 1, ||v|| = 1\}$$

**Nuclear ball!**

# Convex hulls of algebraic varieties




- Systematic methods to produce (exact or approximate) SDP representations of convex hulls
- Based on sums of squares (Shor, Nesterov, Lasserre, P., Nie, Helton)
- Parallels/generalizations from combinatorial optimization (*theta bodies*, e.g., Gouveia-Laurent-P.-Thomas 09, *orbitopes* Sanyal-Sottile-Sturmfels)



# Nuclear norm and SDP

- The nuclear norm is SDP-representable!

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X)\,,$$

- Semidefinite programming characterization:

$$\begin{array}{ll} \max\limits_{Y} & \mathrm{Tr}(X'Y) \\ \text{s.t.} & \begin{bmatrix} I_m & Y \\ Y' & I_n \end{bmatrix} \succeq 0. \end{array} \qquad \begin{array}{ll} \min\limits_{W_1, W_2} & \frac{1}{2}(\mathrm{Tr}(W_1) + \mathrm{Tr}(W_2)) \\ \text{s.t.} & \begin{bmatrix} W_1 & X \\ X' & W_2 \end{bmatrix} \succeq 0. \end{array}$$

# A convex heuristic

- **PROBLEM:** Find the matrix of lowest rank that satisfies the underdetermined linear system

$$\mathcal{A}(X) = b \qquad \mathcal{A} : \mathbb{R}^{n \times m} \to \mathbb{R}^p$$

- Convex optimization heuristic
  - Minimize *nuclear norm* (sum of singular values) of X
  - This is a *convex* function of the matrix X
  - Equivalent to a SDP problem

$$\begin{array}{ll} \text{minimize} & \|X\|_* = \sum_{i=1}^{m} \sigma_i(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

# Nuclear norm heuristic

**Affine Rank Minimization:**

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

**Relaxation:**

$$\begin{array}{ll} \text{minimize} & \|X\|_* = \sum_{i=1}^m \sigma_i(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

- Proposed in Maryam Fazel's PhD thesis (2002).
- Nuclear norm is the *convex envelope* of rank
- Convex, can be solved efficiently
- Seems to work well in practice

# Nice, but will it work?

**Affine Rank Minimization:**

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

**Relaxation:**

$$\begin{array}{ll} \text{minimize} & \|X\|_* = \sum_{i=1}^{m} \sigma_i(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

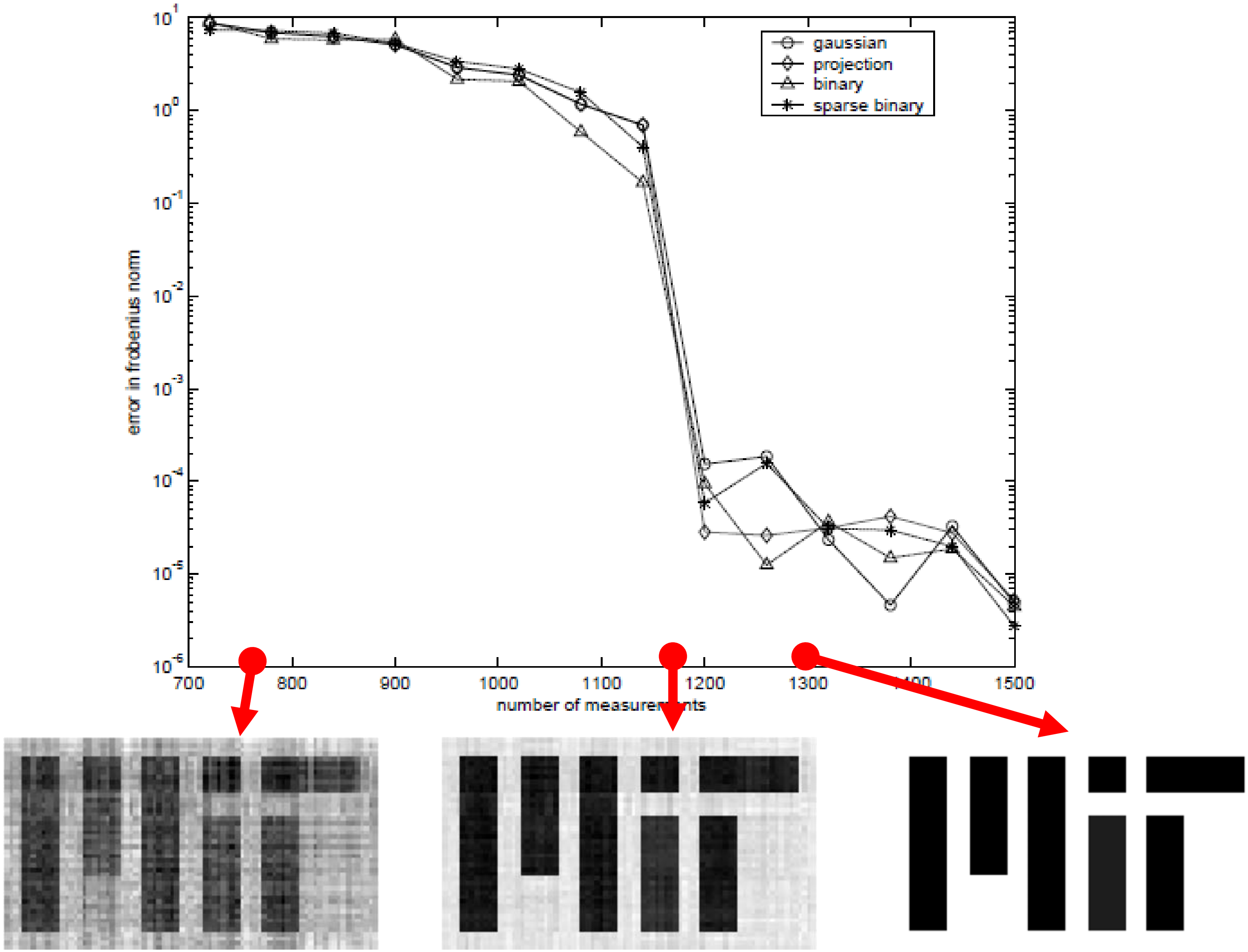
Let's see...

# Numerical experiments

- Test matrix arbitrarily chosen ;)

- Rank 5 matrix, 46x81 pixels
- Generate random equations, Gaussian coeffs.
- Nuclear norm minimization via SDP (SeDuMi)

# Phase transition



# How to explain this?

Apparently, under certain conditions, nuclear norm minimization "works".

How to formalize this?

# How to *prove* that relaxation will work?

**Affine Rank Minimization:**

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

**Relaxation:**

$$\begin{array}{ll} \text{minimize} & \|X\|_* = \sum_{i=1}^m \sigma_i(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

General recipe:

- Find a *deterministic* condition that ensures success
- Sometimes, condition may be hard to check
- If so, invoke *randomness* of problem data, to show condition *holds with high probability (concentration of measure)*
- Generalizes L1 “compressed sensing” results (Donoho, Candès-Tao, etc...)

# Compressed sensing - Overview

- {Compressed|compressive}{sensing|sampling}
- New paradigm for data acquisition/estimation
- Influential recent work of Donoho/Tanner, Candès/Romberg/Tao, Baraniuk,...

- Relies on sparsity (on some domain, e.g., Fourier, wavelet, etc)
- “Few” random measurements + smart decoding
- Many applications, particularly MRI, geophysics, radar, etc.

# Nice, but will it work?

**Affine Rank Minimization:**

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

**Relaxation:**

$$\begin{array}{ll} \text{minimize} & \|X\|_* = \sum_{i=1}^{m} \sigma_i(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

- Use a “restricted isometry property” (RIP)
- Then, this heuristic *provably* works.
- For “random” operators, RIP holds with overwhelming probability

# Restricted Isometry Property (RIP)

- Let :R$^{m \times n}$ $\mathcal{X}$ R$^{p}$ be a linear map. For every positive integer $r \leq m$, define the $r$-restricted isometry constant to be the smallest number $\delta_r$( ) such that

$$(1-\delta_r(\mathcal{A}))\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1+\delta_r(\mathcal{A}))\|X\|_F$$

holds for all matrices $X$ of rank at most $r$.

- Similar to RIP condition for sparsity studied by Candès and Tao (2004).

- Implies “transverse intersection” betwen Ker $\langle$and the low rank variety

# Nuclear norm works!

**Theorem (Recht-Fazel-P.):** If is "random" and p∋ $c_0\, r(2n-r)\log(n^2)$, the nuclear norm heuristic succeeds with high probability.

- Number of measurements $\underbrace{c_0}_{\text{constant}}\ \underbrace{r(2n-r)}_{\text{intrinsic dimension}}\ \underbrace{\log(n^2)}_{\text{ambient dimension}}$

- Typical scaling for this type of result.

- Extensions (e.g., matrix completion: Candès-Recht, Candès-Tao, Keshavan-Montanari-Oh, etc.)

# Experiment 2

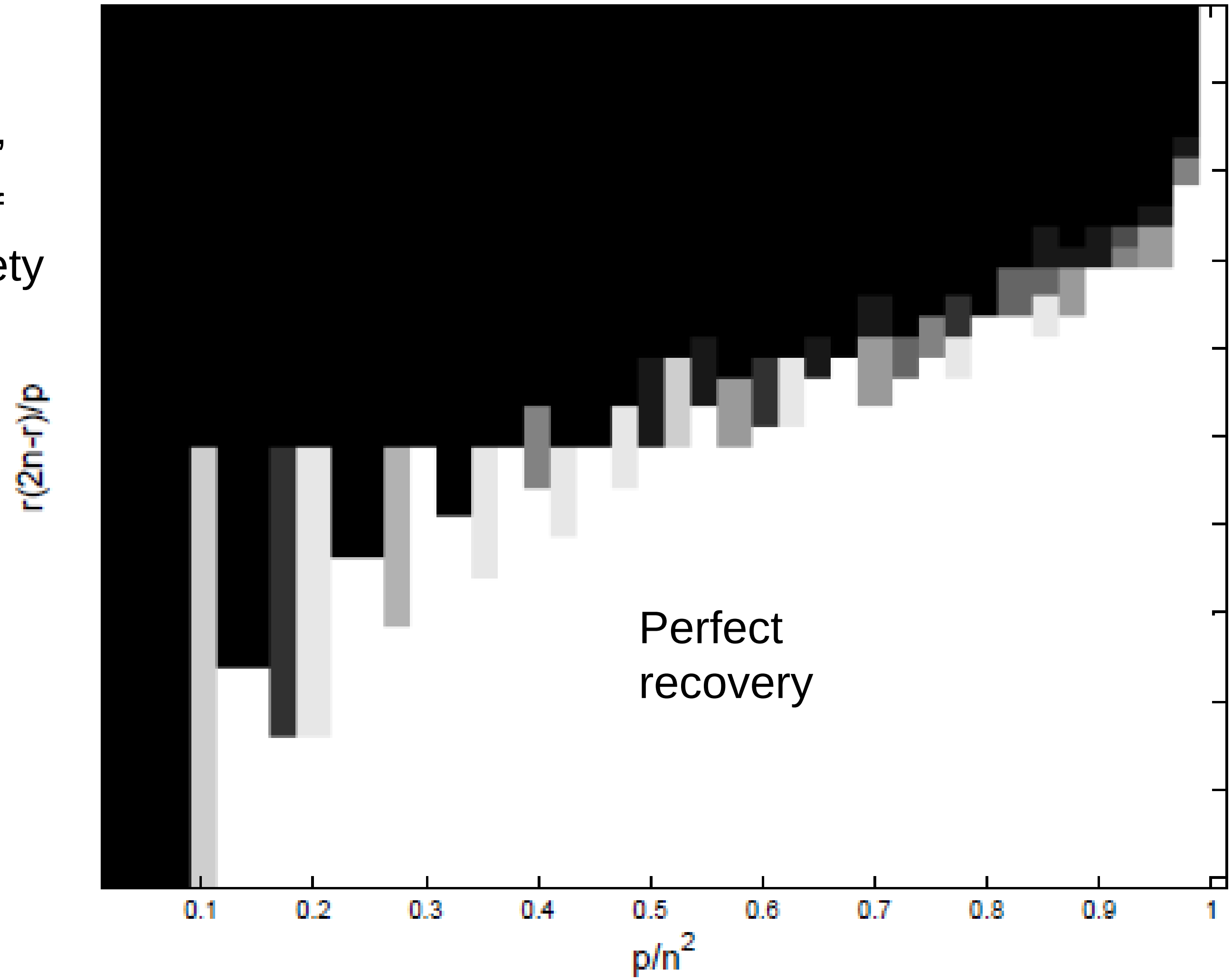
- For a triple (n,r,p), generate random n x n matrix Y of rank r and create p Gaussian random measurements
- For each instance, we solved the SDP

$$\begin{array}{ll} \min_X & \|X\|_* \\ \text{s.t.} & A(X) = A(Y) \end{array}$$

- A matrix Y is “recovered” if the output of SeDuMi, X, satisfied $||X-Y||_2/||Y||_2 < 1e\text{-}5$.

# Phase transition

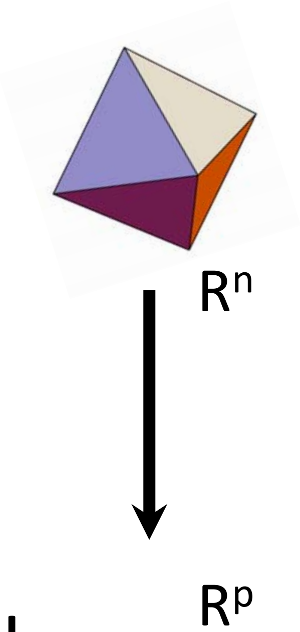
“Normalized” dimension of the rank *r* variety




Fraction of measurements

# The “neighborly” intuition

(Donoho, Tanner)



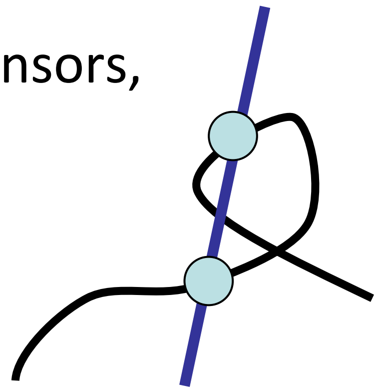

Behavior of neighborly polytopes under “random” projections

**Fact:** All (or “most”) faces remain extremal

For nuclear norm, similar situation, but more complicated (faces are not polyhedral)

# Secant varieties

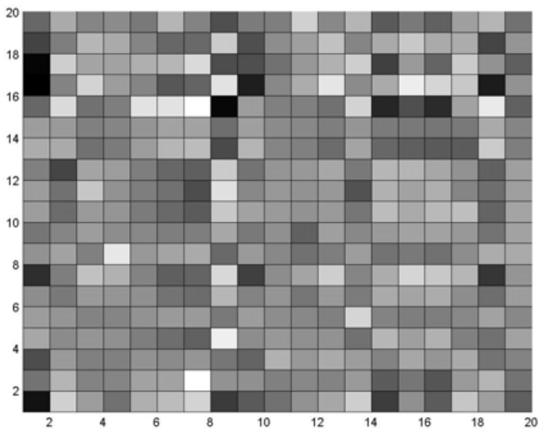
- What is common to the two cases? Can this be further extended?
- A natural notion: *secant* varieties
- Generalize notions of rank to other objects (e.g., tensors, nonnegative matrices, etc.)



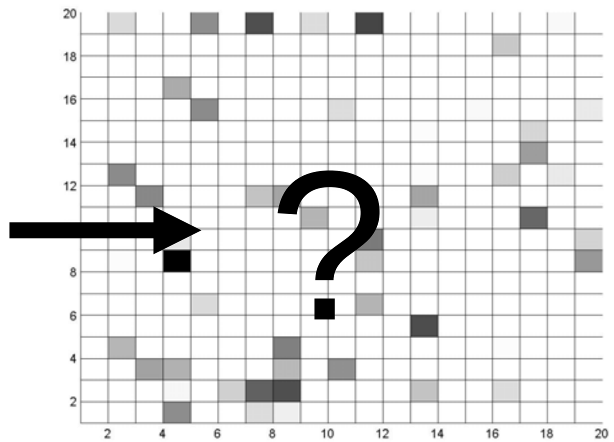
- However, technical difficulties
  - In general, these varieties may not be closed
  - In general, associated norms are not polytime computable

# What if sparsity pattern is not known?

$$C = A^* + B^*$$

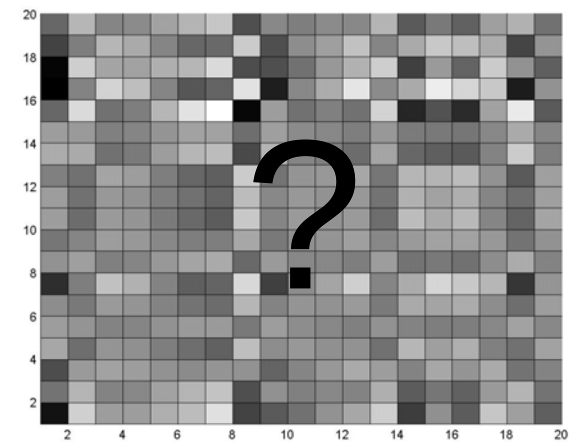


Given Composite matrix



Unknown Sparse Matrix

Unknown support, values



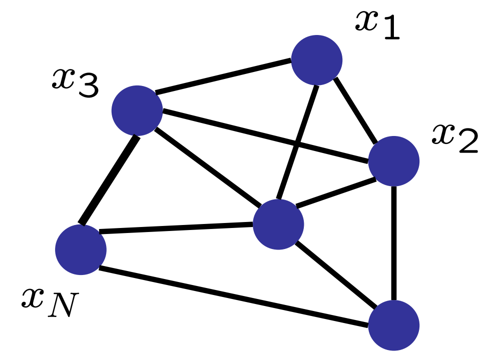
Unknown Low-rank Matrix

Unknown rank, eigenvectors

**Task: given *C*, recover *A** and *B***

# Application: Graphical models

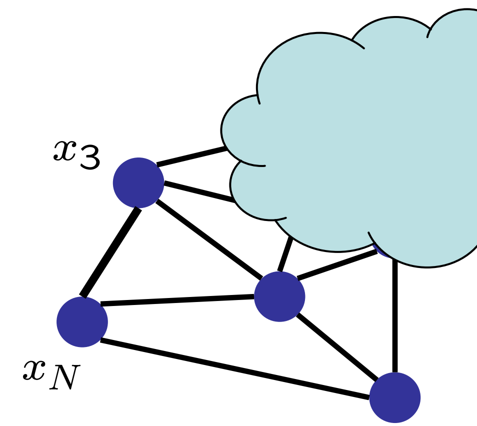
A probabilistic model given by a graph.




- Then, the inverse covariance $\Sigma^{-1}$ is sparse.

No longer true if graphical model has *hidden variables.*

But, it is the sum of a sparse
and a low-rank matrix (Schur complement)




$$\hat{K}_o = \Sigma_o^{-1} = K_o - K_{o,h} K_h^{-1} K_{h,o}.$$

# Application: Matrix rigidity

**Smallest** number of changes to reduce rank [Valiant 77]

Rigidity of a matrix NP-hard to compute

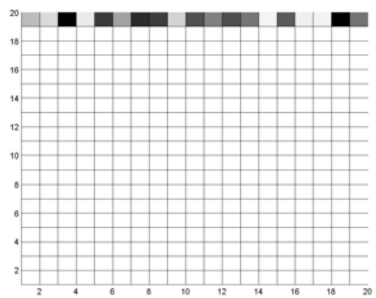
Rigidity bounds have a number of applications

- Complexity of linear transforms [Valiant 77]
- Communication complexity [Lokam 95]
- Cryptography

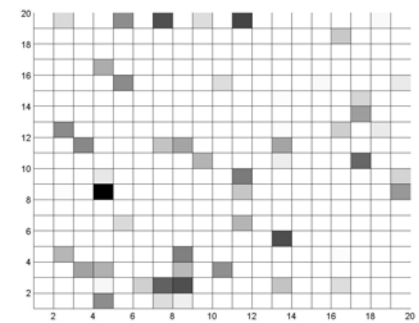
# Identifiability issues

Problem can be *ill-posed.* Need to ensure that terms cannot be *simultaneously* sparse and low-rank.

Bad



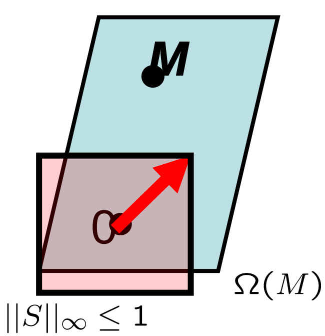
Benign

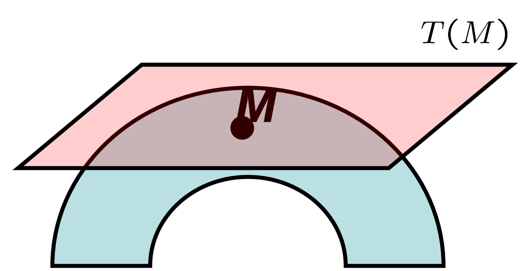


Define two geometric quantities, related to tangent spaces to sparse and low-rank varieties:

$$\mu_1(M) = \max_{S \in \Omega(M)} \frac{||S||}{||S||_\infty}$$

$$\mu_2(M) = \max_{S \in T(M)} \frac{||S||_\infty}{||S||}$$

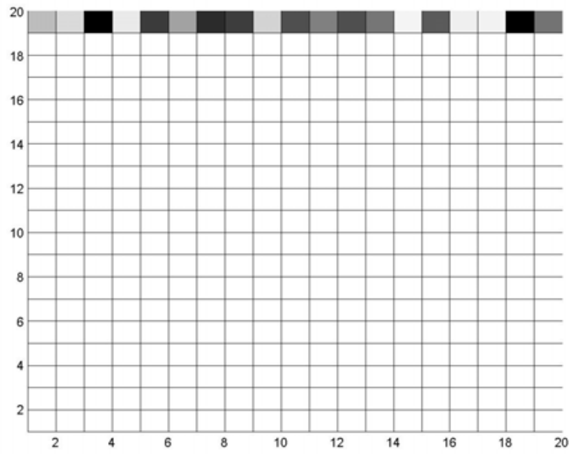


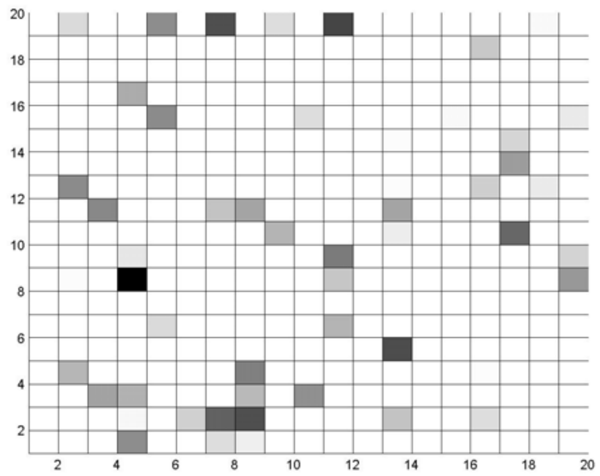

Rank ***r*** matrices

# "Benign" Sparse Matrices

For matrix decomposition, not all sparse matrices are equally benign.




Bad




Benign

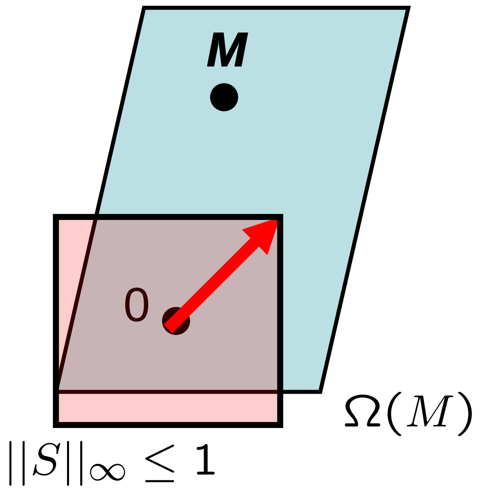
Sparsity pattern matters, not just the number of non-zeros…

# Measure of Sparsity Pattern

Given matrix $M$, define

$$\Omega(M) = \{S \mid \text{support}(S) \subset \text{support}(M)\}$$

$$\mu_1(M) = \max_{S \in \Omega(M)} \frac{||S||}{||S||_\infty}$$ ⟵ Spectral norm




Small $\mu_1 \Leftrightarrow$ sparse and benign

Concrete example -- “Bounded degree” matrices have small $\mu_1$

# “Benign” Low-rank Matrices

Similarly, not all low-rank matrices equally benign…

Benign

Bad

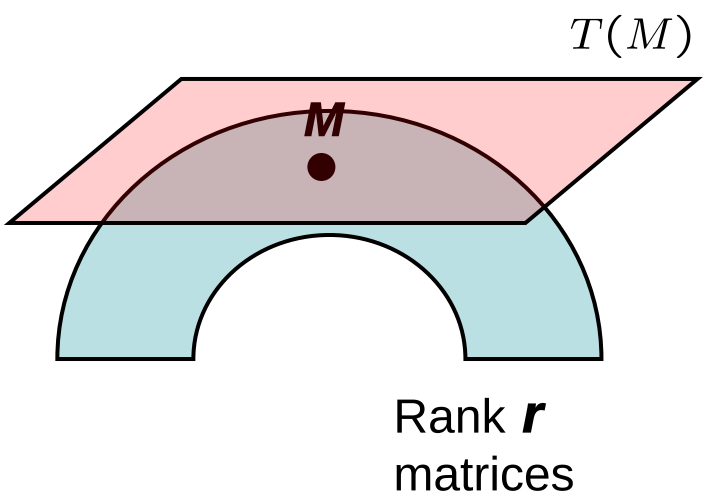
Row and column spaces “diffuse” w.r.t standard basis

“concentrated” row and column spaces

# Measure of Row/Column Spaces

Given matrix ***M***, define

$$T(M) \;=\; \text{span}\left\{ S : \begin{array}{l} RowSpace(S) \subset RowSpace(M) \text{ or} \\ ColSpace(S) \subset ColSpace(M) \end{array} \right\}$$




$$\mu_2(M) = \max_{S \in T(M)} \frac{||S||_\infty}{||S||}$$

Small $\mu_2 \Leftrightarrow$ low-rank and benign

Concrete example -- Low-rank matrices with "incoherent" row/column spaces (w.r.t. standard basis) have small $\mu_2$

# Tangent space identifiability

Necessary and sufficient condition for exact decomposition

**Tangent spaces must intersect transversally**

$$\Omega(A^*) \cap T(B^*) = \{0\}$$

Sufficient condition for transverse intersection

$$\mu_1(A^*)\mu_2(B^*) < 1 \quad \Rightarrow \quad \Omega(A^*) \cap T(B^*) = \{0\}$$

# Natural convex relaxation

$$||A||_0 \longrightarrow ||A||_1 = \sum_{i,j} |a_{ij}|$$

$$rank(B) = ||\underline{\sigma}(B)||_0 \longrightarrow ||B||_* = \sum_i \sigma_i(B)$$

Propose:

$$\boxed{\begin{aligned} (\hat{A}, \hat{B}) = \arg\min_{A,B} \quad & \gamma ||A||_1 + ||B||_* \\ \text{s.t.} \quad & A + B = C \end{aligned}}$$

Convex program (in fact, an SDP)

# Matrix decomposition

**Theorem (CSPW09):** For any ***A**** and ***B****

$$\mu_1(A^*)\mu_2(B^*) < \frac{1}{8} \quad \Rightarrow \quad (\hat{A}, \hat{B}) = (A^*, B^*)$$ is **unique optimum** of convex program for a **range** of $\gamma$

Essentially a **refinement** of tangent space transversality conditions

| | |
|---|---|
| Transverse intersection: | $\mu_1(A^*)\mu_2(B^*) < 1$ |
| Convex recovery: | $\mu_1(A^*)\mu_2(B^*) < \frac{1}{8}$ |

Under "natural" random assumptions, conditions holds w.h.p. (see also Candès-Li-Ma-Wright, Dec. 2009).

# Summary

- Sparsity, rank, and beyond
  - Many applications
  - Common geometric formulation: secant varieties
  - Convex hulls of these varieties give "good" proxies for optimization
  - Algebraic and geometric aspects

- Theoretical challenges
  - Efficient descriptions
  - Approximate recovery (correct rank and sparsity)
  - Other formulations (e.g., Ames-Vavasis on planted cliques)
  - Finite fields?

- Algorithmic issues
  - Reliable, large scale methods

# Thank you!

Want to know more? Details below, and in references therein:


- B. Recht, M. Fazel, P.A. Parrilo, *Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization*, arXiv:0706.4138. *SIAM Review*, to appear.
- V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, A. Willsky, *Rank-Sparsity Incoherence for Matrix Decomposition*, arXiv:0906.2220, 2009.


Thanks for your attention!