Sparsity, Rank, and All That

Ben Recht
Center for the Mathematics of Information
Caltech

March 30, 2009

Undertermined Linear Systems

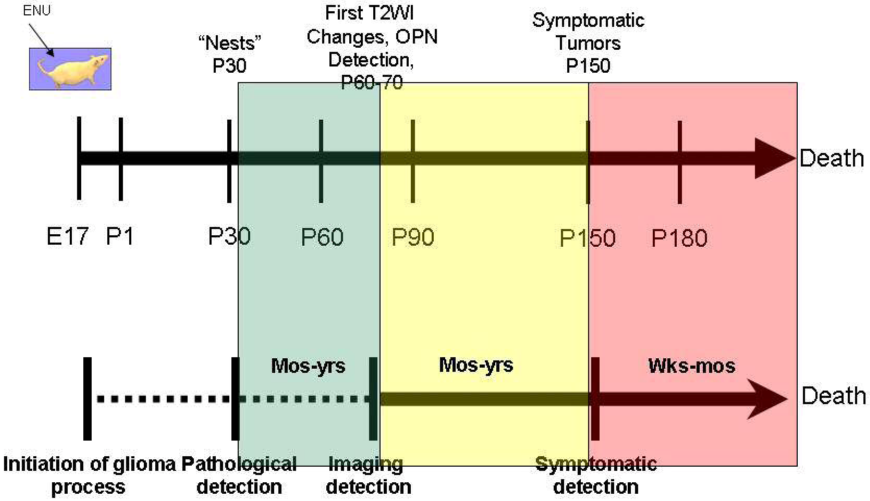
$$Ax = b$$

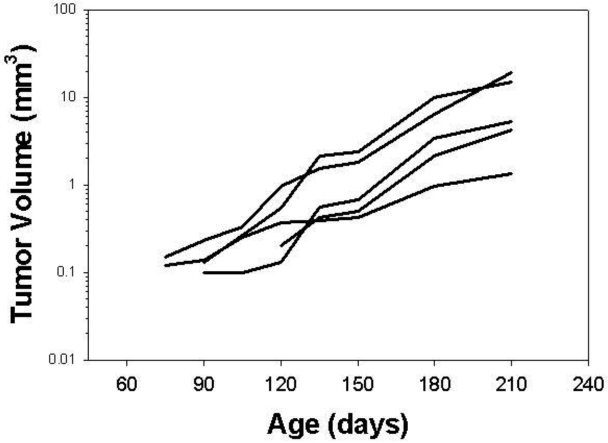
- When A has less rows than columns, there are an infinite number of solutions.
- Which one should be selected?

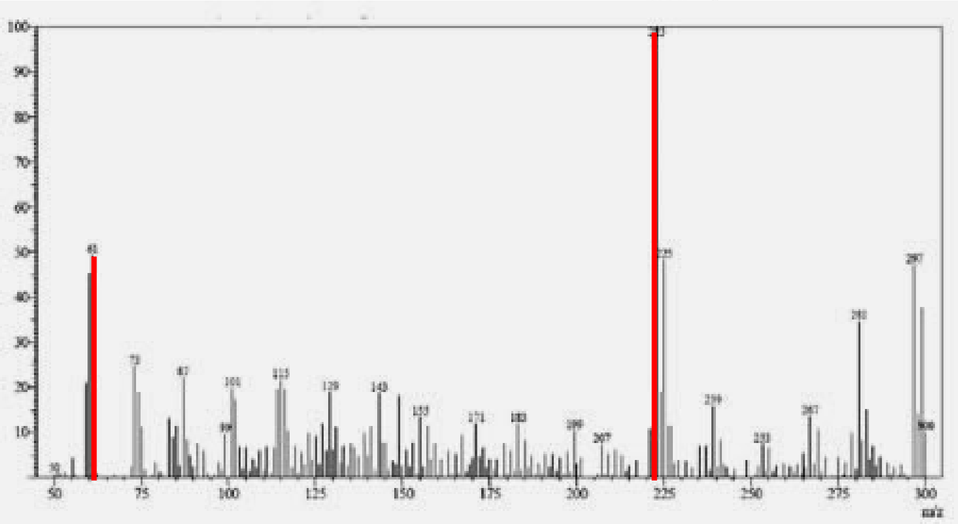
OR: $$\sum_{j=1}^{M} (x^* a_j - b_j)^2$$

$$M << \dim(x)$$

Mining for Biomarkers







- $n_{patients} << n_{peaks}$
- If very few are needed for diagnosis, search for a *sparse set of markers*
- l_1, LASSO, etc.

Recommender Systems

amazon.com

More Top Picks for You
All of Statistics
A Concise Course in Statistical Inference
DATA
All of Nonparametric Statistics

NETFLIX

Because you enjoyed:

2001: A Space Odyssey
Blue Velvet
Bottle Rocket

We think you'll enjoy:
Stalker

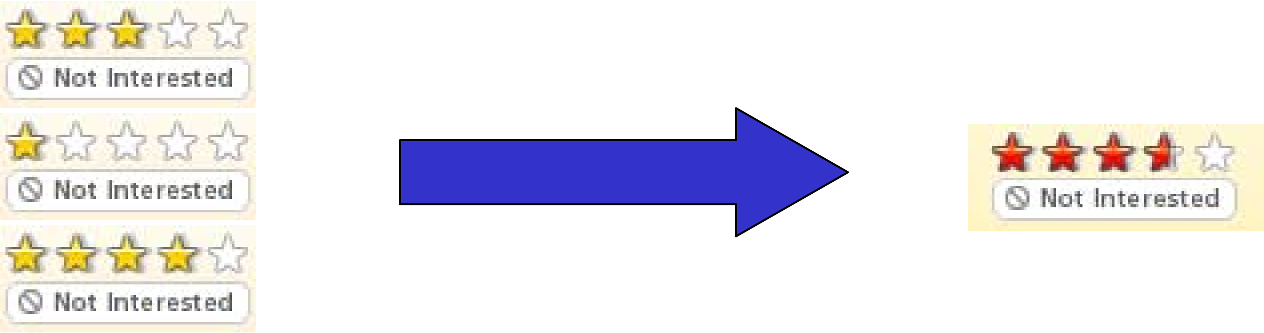
match.com

chemistry

eHarmony

Netflix Prize

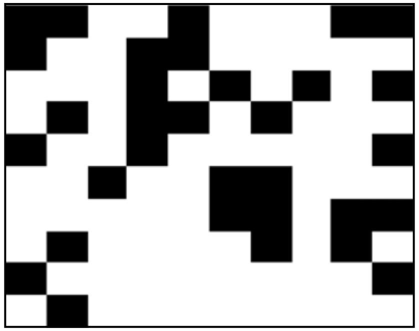
- One million big ones!
- Given 100 million ratings on a scale of 1 to 5, predict 3 million ratings to highest accuracy



- 17770 total movies x 480189 total users
- Over 8 billion total ratings
- How to fill in the blanks?

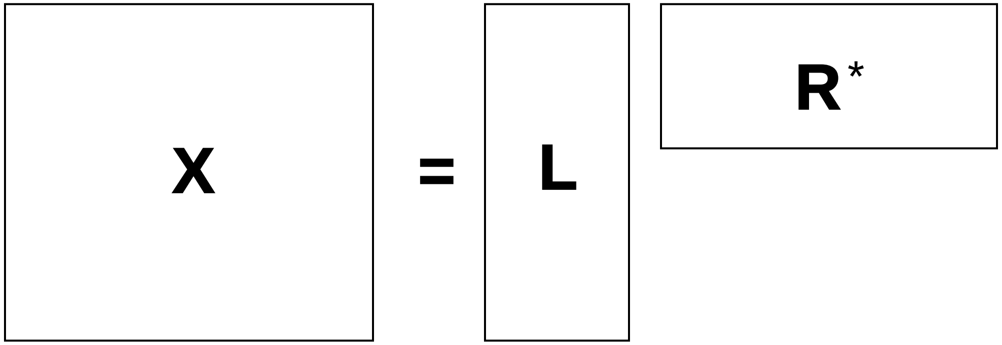
Abstract Setup: Matrix Completion

X =



X_{ij} known for black cells
X_{ij} unknown for white cells
Rows index movies
Columns index users

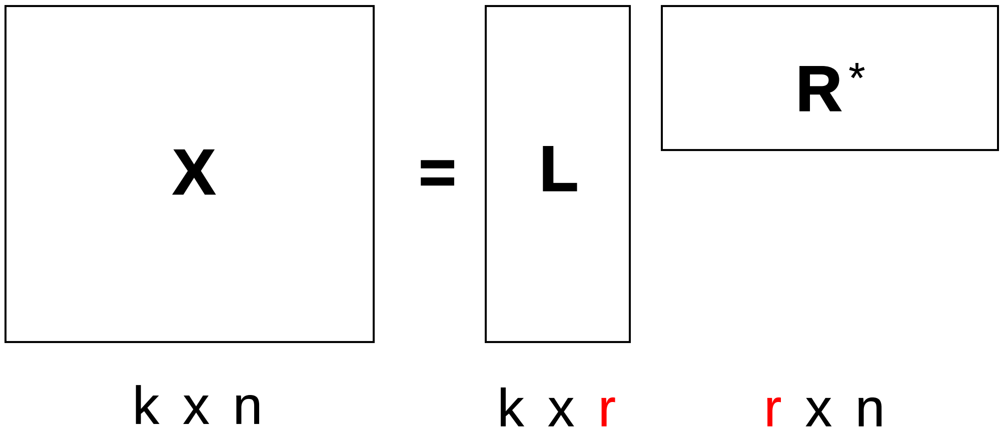
- How do you fill in the missing data?



k x n   k x r   r x n

kn entries   r(k+n) entries

Matrix Rank



$\mathbf{X} = \mathbf{L} \; \mathbf{R}^*$

$k \times n \qquad k \times r \qquad r \times n$

- The rank of **X** is...

  the dimension of the span of the rows

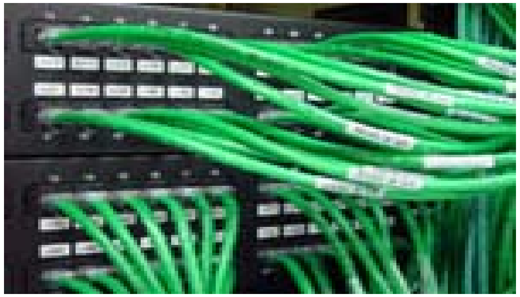
  the dimension of the span of the columns

  the smallest number r such that there exists an $k \times r$ matrix **L** and an $n \times r$ matrix **R** with $\mathbf{X}=\mathbf{L}\mathbf{R}^*$

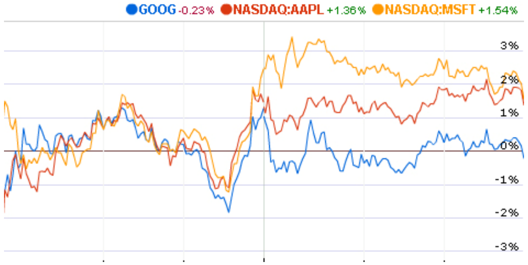
Complex Systems

Predictions

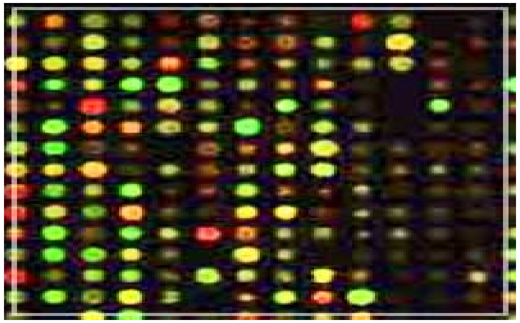
Structure



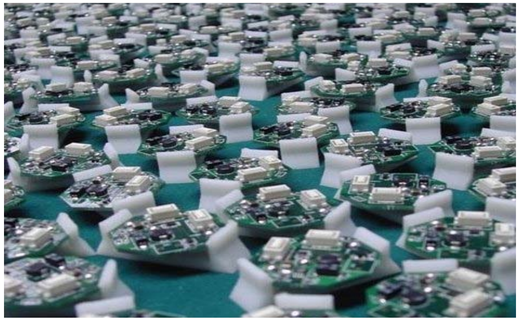
Rank

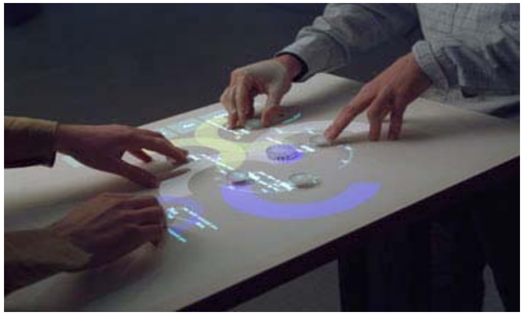


Smoothness



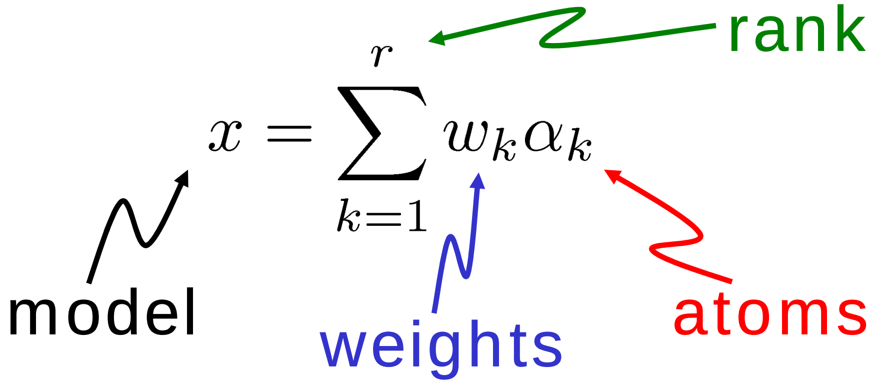
Sparsity





Dynamics

Parsimonious Models



$$x = \sum_{k=1}^{r} w_k \alpha_k$$

rank

model

weights

atoms

- Search for best linear combination of fewest atoms
- “rank” = fewest atoms needed to describe the model

- Suppose we want to solve

$$\begin{array}{ll} \text{minimize} & \text{rank}(x) \\ \text{subject to} & Ax = b \end{array}$$

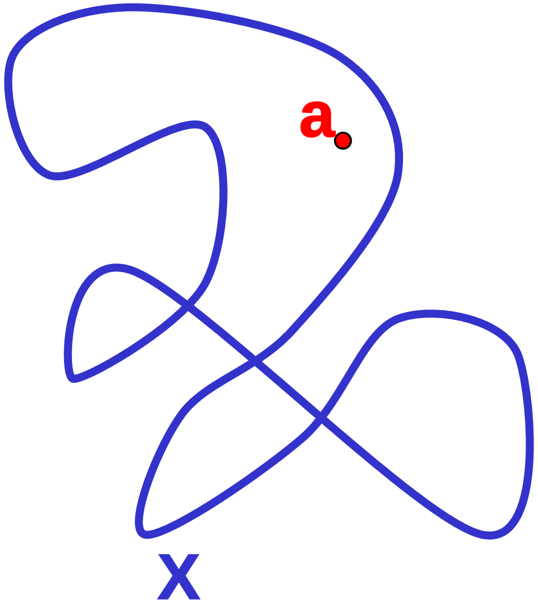
- **M** = {all rank r models}
- What happens when dimension(**M**) is smaller than the number of rows of A?

Plan of Attack

- Encoding parsimony
  - embeddings, projections, and the atomic norm
- Example 1: Sparse vectors
  - Atomic norm = l_1
  - Decoding via Restricted Isometry
  - Decoding via most encodings
- Example 2: Low rank matrices
  - Atomic norm = trace norm
  - Decoding via Restricted Isometry
  - Decoding via most encodings
- Other models and further directions

Whitney's Theorem

- Any random projection of a d-dimensional manifold into 2d+1 dimensions is en embedding!



- Let **X** = { t(x-y) : x,y$\in$ **M**, t $\in \mathbb{R}$} $\subset \mathbb{R}^{\mathbb{D}}$

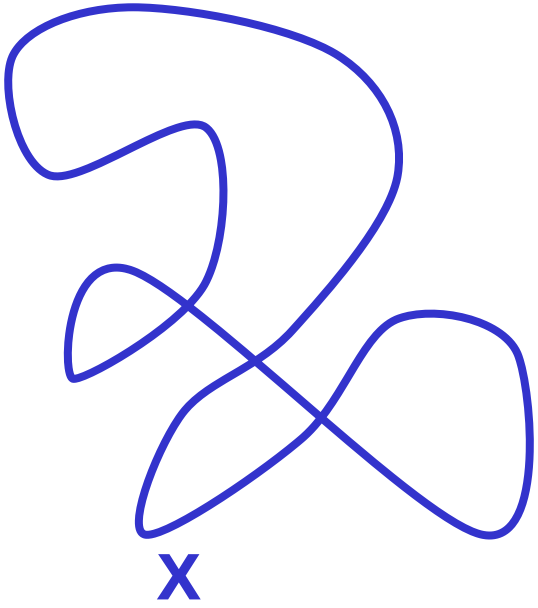
- If D>2d+1, any random **a** is not in **X**.

- Project orthogonal **a**.

- If there are x,y in **M** with $\pi_{\mathbf{a}}$(x) = $\pi_{\mathbf{a}}$(y), then there is a t with **a** = t(x-y) $\in$ **X** (*contradiction*).

Whitney's Theorem

- Any random projection of a d-dimensional manifold into 2d+1 dimensions is an embedding!



- If any random projection is an embedding, when can we reconstruct points in **X** from their projected values?

- Given a random **encoder**, when can we find a low-complexity **decoder**?

- **Answer**: need slightly more geometry

Parsimonious Models

rank

$$x = \sum_{k=1}^{r} w_k \alpha_k$$

model

weights

atoms

- Search for best linear combination of fewest atoms
- “rank” = fewest atoms needed to describe the model

- “natural” heuristic:

$$\|x\|_{\mathcal{A}} \equiv \inf\left\{ \sum_{k=1}^{r} |w_k| \;:\; x = \sum_{k=1}^{r} w_k \alpha_k \right\}$$

Cardinality

- Vector x has cardinality s if it has at most s nonzeros.

$$x = \sum_{k=1}^{s} w_k e_{i_k}$$

- Atoms are a discrete set of orthogonal points
- Typical Atoms:
  - standard basis
  - Fourier basis
  - Wavelet basis

Cardinality Minimization

- **PROBLEM:** Find the vector of lowest cardinality that satisfies/approximates the underdetermined linear system

$$Ax = b \qquad A : \mathbb{R}^n \to \mathbb{R}^m$$

- **NP-HARD:**
  - Reduce to EXACT-COVER [Natarajan 1995]
  - Hard to approximate
  - Known exact algorithms require enumeration

Proposed Heuristic

Cardinality Minimization:

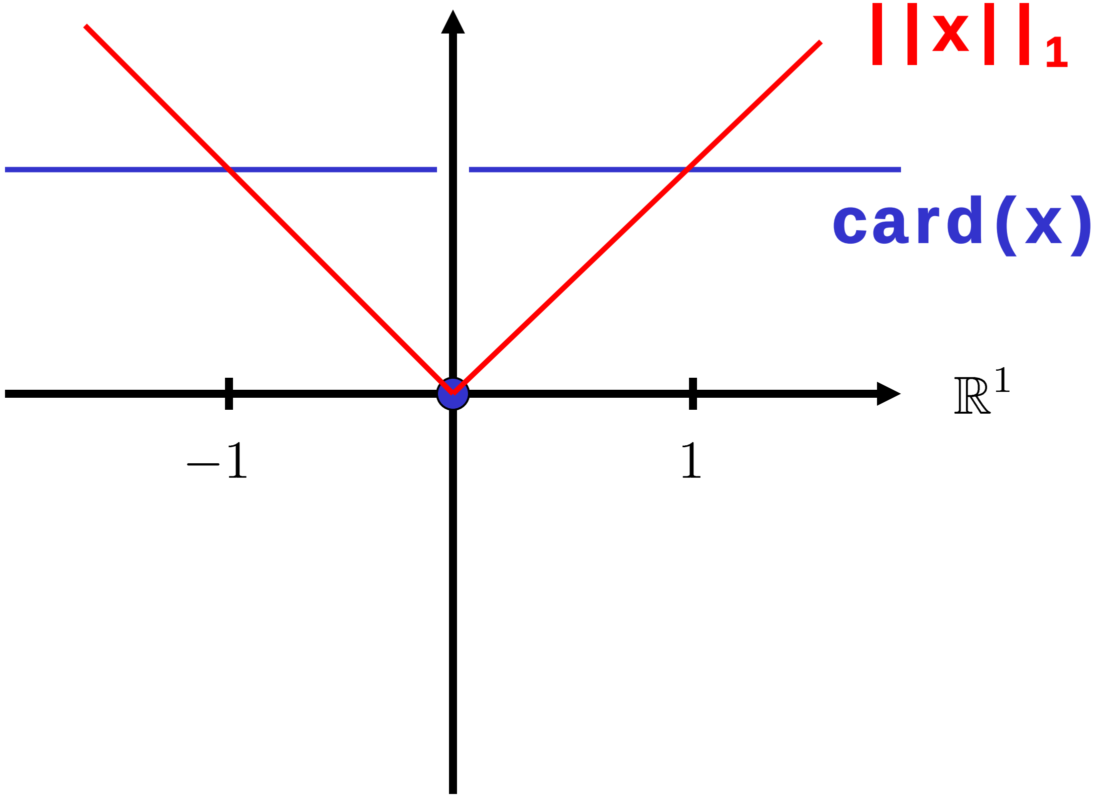
$$\begin{array}{ll} \text{minimize} & \operatorname{card}(x) \\ \text{subject to} & Ax = b \end{array}$$

Convex Relaxation:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 = \sum_{i=1}^n |x_i| \\ \text{subject to} & Ax = b \end{array}$$

- Long history (back to geophysics in the 70s)
- Flurry of recent work characterizing success of this heuristic: Candès, Donoho, Romberg, Tao, Tropp, etc., etc...
- “Compressed Sensing”

Why l_1 norm?



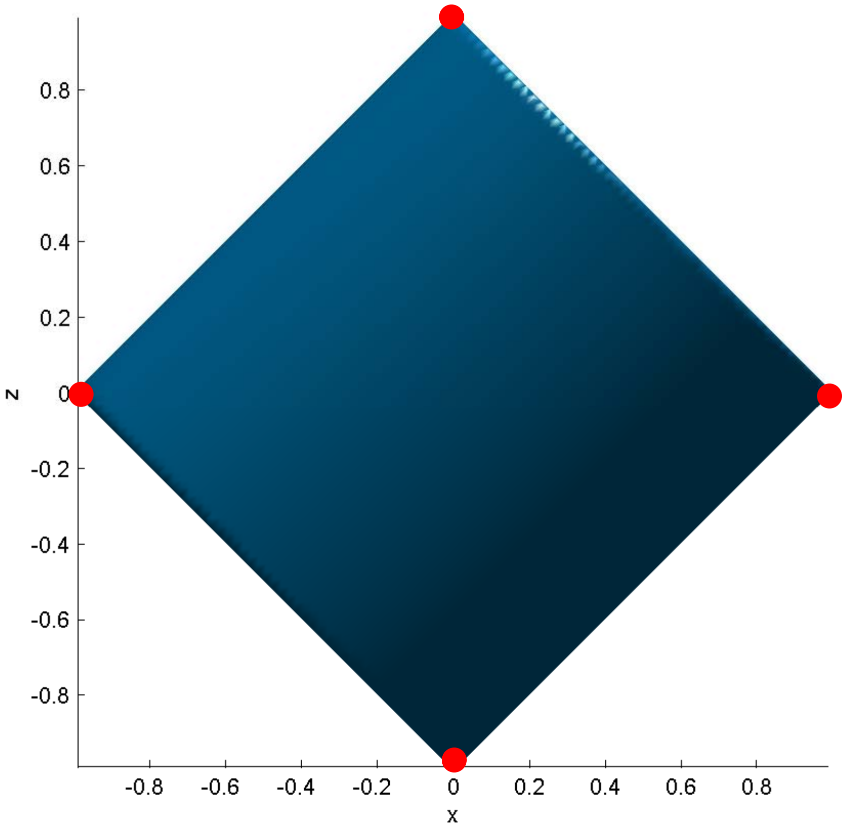
- 2d vectors

1 nonzero

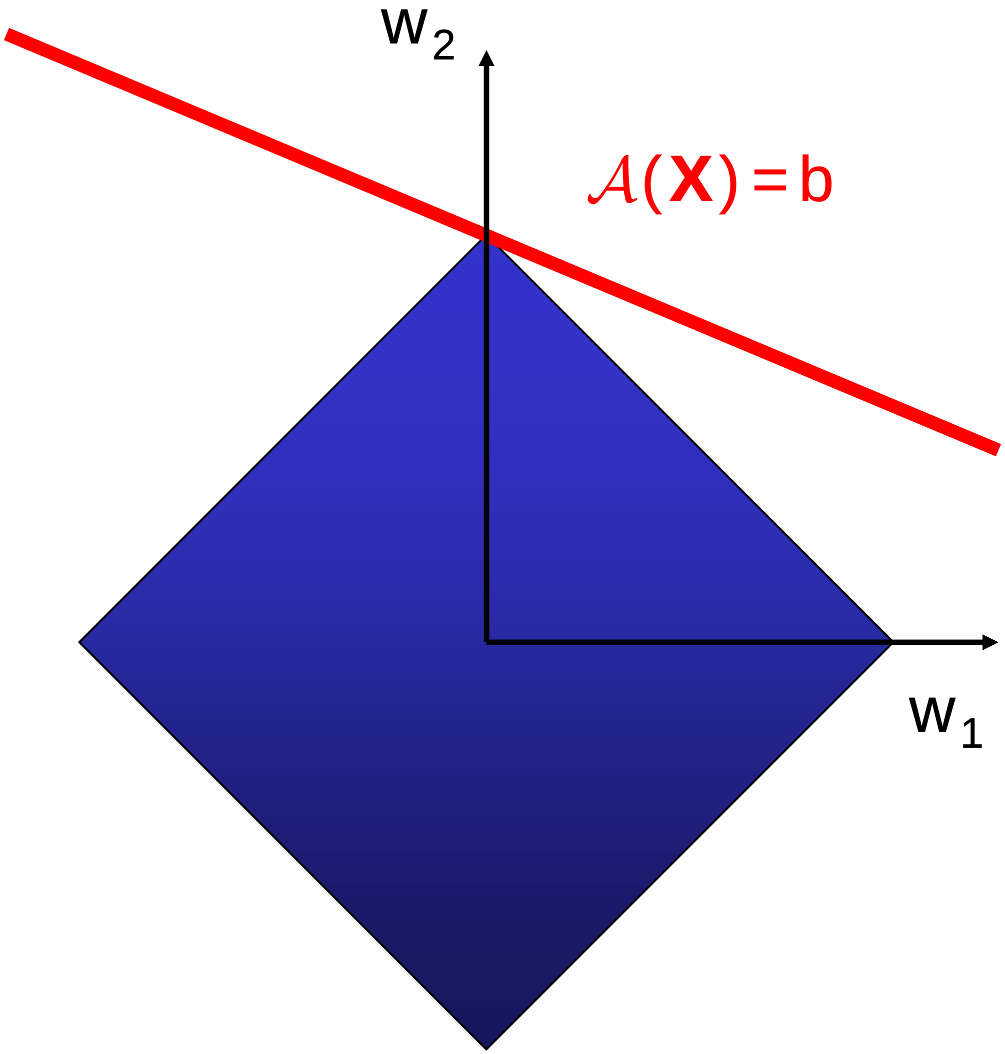
$x^2 + y^2 = 1$

Convex hull:

$$\{\mathbf{x} \ : \ \|\mathbf{x}\|_1 \leq 1\}$$



$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_1 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array}$$



When is this intuition precise?

Restricted Isometry Property (RIP)

- Let A:$\mathbb{R}^n \to \mathbb{R}^m$ be a linear map. For every positive integer s$\leq$m, define the s-restricted isometry constant to be the smallest number δ_s(A) such that

$$(1-\delta_s(A))\|x\| \leq \|Ax\| \leq (1+\delta_s(A))\|x\|$$

holds for all vectors x of cardinality at most s.

- Candès and Tao (2005).

RIP $\Rightarrow$ Unique Sparse Solution

- **Theorem** Suppose that $\delta_{2s}(A)<1$ for some integer $s\geq 1$. Then there can be at most one vector x with cardinality less than or equal to s satisfying Ax= b.

- **Proof:** Assume, on the contrary, that there exist two different vectors, x_1 and x_2, satisfying the matrix equation ($Ax_1=Ax_2=b$).
- Then $z:=x_1-x_2$ is a nonzero matrix of card at most 2s, and Az=0.
- But then we would have

$$0 = \|Az\| \geq (1-\delta_{2s}(A))\|z\| > 0$$

which is a contradiction.

RIP $\Rightarrow$ Heuristic Succeeds

- **Theorem:** Let x_0 be a vector of cardinality at most s. Let x_* be the solution of $Ax=Ax_0$ of smallest l_1 norm. Suppose that $\delta_{4s}(A) < 1/4$. Then $x_*=x_0$.

Independent of n,m,s

- Deterministic condition on A
- Current best bound: $\delta_{2s}(A) < 0.2$ suffices.

RIP $\Rightarrow$ Heuristic Succeeds

- **Theorem:** Let x_0 be a matrix of cardinality s. Let x_* be the solution of $Ax=Ax_0$ of smallest l_1 norm. Suppose that $s \geq 1$ is such that $\delta_{4s}(A) < 1/4$. Then $x_*=x_0$.
- **Proof Sketch:** Let $R:=x_*-x_0$ be the error.
- The majority of the mass of R is concentrated in the support of x_0:

$$\|x_0\|_1 \geq \|x_0 + R\|_1 = \|x_0 + R_0\|_1 + \left\| \sum_{j>1} R_j \right\|_1 \geq \|x_0\| - \|R_0\|_1 + \left\| \sum_{j>1} R_j \right\|_1$$

- We can decompose $R = R_0 + R_1 + R_2 + \ldots$
  - R_0 is projection on the support of x
  - R_i have cardinality at most 3s and disjoint support from x_0 for $i>0$

RIP $\Rightarrow$ Heuristic Succeeds (cont)

$$
\begin{aligned}
0 = \|AR\| &\geq \|A(R_0 + R_1)\| - \sum_{j\geq 2} \|AR_j\| \\
&\geq (1-\delta_{4s})\,\|R_0 + R_1\|_F - (1+\delta_{3s}) \sum_{j\geq 2} \|R_j\|_F \\
&\geq \underbrace{\left((1-\delta_{4s}) - \sqrt{\tfrac{1}{3}}(1+\delta_{3s})\right)}_{\textbf{Striclty positive for } \delta_{4s}<1/4} \|R_0\|_F
\end{aligned}
$$

- Using $\sum_{j\geq 2} \|R_j\| \leq \sqrt{\frac{1}{3}}\|R_0\|$ from CRT 06

- Proof of l_2 constrained version is similar

Nearly Isometric Random Variables

- Let A be a random variable that takes values in linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- We say that A is *nearly isometrically distributed* if

1. For all x $\in \mathbb{R}^n$,

$$\mathbf{E}[\|Ax\|^2] = \|x\|^2$$

Isometric in expectation

2. For all 0<ε<1 we have,

$$\mathbf{P}(|\|Ax\|^2 - \|x\|_F^2| \geq \epsilon\|x\|_F^2) \leq 2\exp\left(-\frac{m}{2}(\epsilon^2/2 - \epsilon^3/3)\right)$$

Large deviations unlikely

Nearly Isometric RVs obey RIP

- **Theorem:** Fix $0<\delta<1$. If A is a nearly isometric random variable, then for every $1\leq s\leq m$, there exist constants c_0, $c_1>0$ depending only on δ such that $\delta_s(A)\leq\delta$ whenever $m\geq c_0\, s \log(n/s)$ with probability at least $1-\exp(-c_1\, m)$.

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

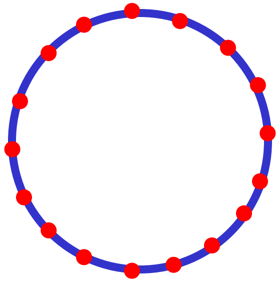
- Typical scaling for this type of result.

Examples of Restricted Isometries

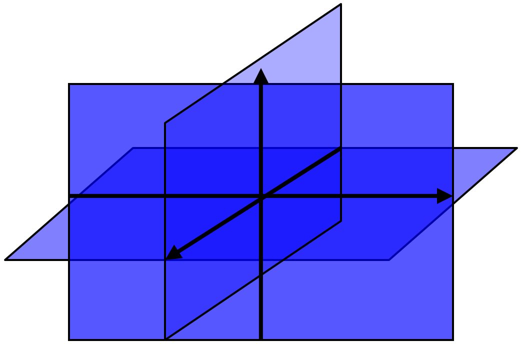
- A_{ij} Gaussian with variance $\frac{1}{m}$
- **A** a random projection
- $A_{ij} = \begin{cases} \sqrt{\frac{1}{m}} & \text{with probability } \frac{1}{2} \\ -\sqrt{\frac{1}{m}} & \text{with probability } \frac{1}{2} \end{cases}$
- $A_{ij} = \begin{cases} \sqrt{\frac{3}{m}} & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -\sqrt{\frac{3}{m}} & \text{with probability } \frac{1}{6} \end{cases} .$
- “Most” transformations when properly scaled

Proof of RIP:

- Probability x is distorted is at most $\exp\left(-\alpha_1(\varepsilon)m\right)$



- Can cover all x on the unit ball in $\mathbb{R}^s$ with at most $\alpha_2(\epsilon)^s$ points.



- Since nearby x's are distorted similarly, probability any s-sparse x is distorted is at most $O\left(\binom{n}{s}\alpha_2(\epsilon)^s\exp\left(-\alpha_2(\varepsilon)m\right)\right)$

- So no x is distorted with Prob at least 1-exp(-c_1m) if

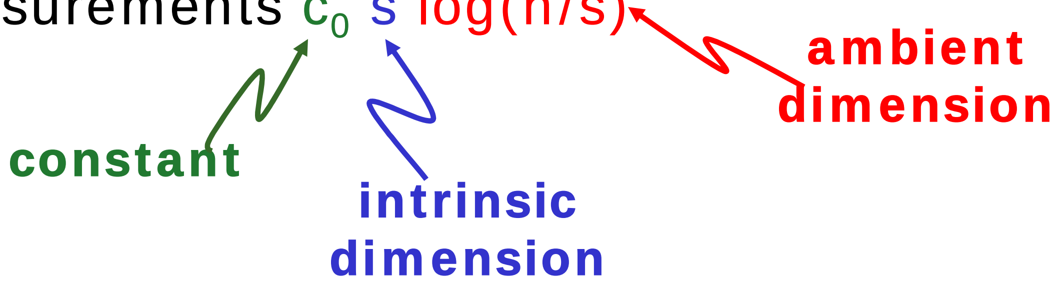
$$m > c_0 s \log\left(\frac{n}{s}\right)$$

The l_1 heuristic works!

$$Ax = b \qquad A : \mathbb{R}^n \to \mathbb{R}^m$$

- The l_1 heuristic succeeds (at sparsity level s) for most A with $m > c_0 s\log(n/s)$

- Number of measurements c_0 s log(n/s)



constant

intrinsic dimension

ambient dimension

- **Approach:** Show that a properly scaled random A is nearly an isometry on the set of 4s-sparse vectors.

(Matrix) Rank

- Matrix X has rank r if it has at most r nonzero singular values.

$$X = \sum_{j=1}^{r} \sigma_j u_j v_j^* = \sum_{j=1}^{r} \sigma_j A_j$$

- Atoms are the set of all rank one matrices
- Not a discrete set

Recommender Systems

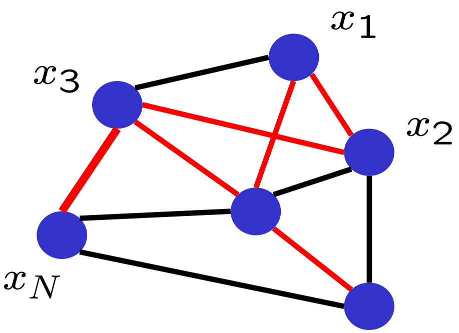
amazon.com

NETFLIX

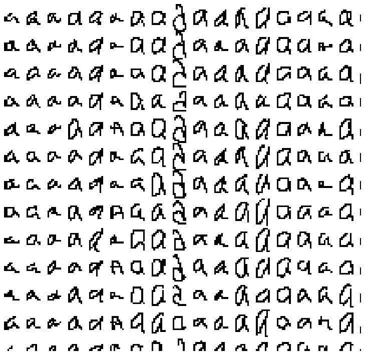
match.com

chemistry

Euclidean Embedding



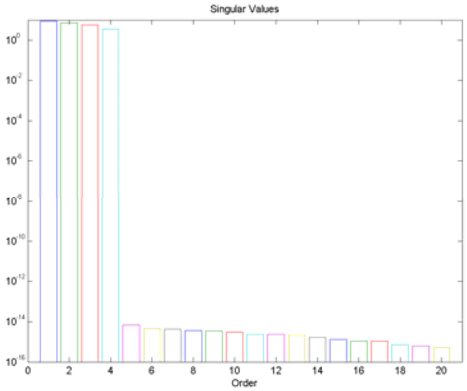
Multitask Learning



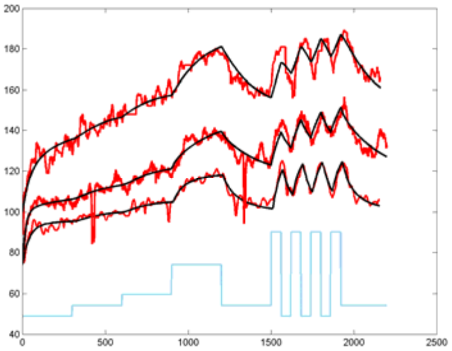
Rank of: Data Matrix

Gram Matrix

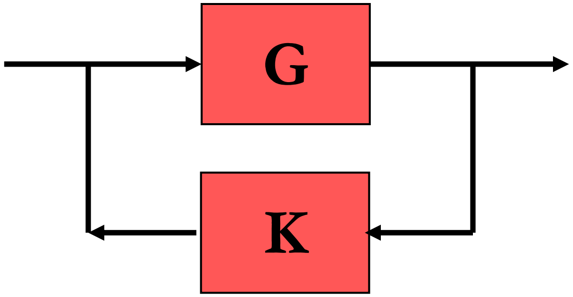
Matrix of Classifiers



Model Reduction



System Identification



Controller Design

Constraints involving the rank of the Hankel Operator, Matrix, or Singular Values

Affine Rank Minimization

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

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

- **NP-HARD:**
  - Reduce to finding solutions to polynomial systems
  - Hard to approximate
  - Exact algorithms are awful (doubly exponential)

Singular Value Decomposition (SVD)

- If **X** is a matrix of size k x n (k$\leq$n) then there matrices **U** (k x k) and **V** (n x k) such that

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$$

$$\mathbf{U}^*\mathbf{U} = I_k \qquad \mathbf{V}^*\mathbf{V} = I_k$$

- $\sum$ a diagonal matrix, $\sigma_1 \geq \ldots \geq \sigma_k \geq 0$

- **Fact:** If **X** has rank r, then **X** has only r non-zero singular values.

- **Dimension of rank r matrices:** $\underline{r\,(k+n-r)} \leq 2\,n\,r$

Proposed Heuristic

Affine Rank Minimization:

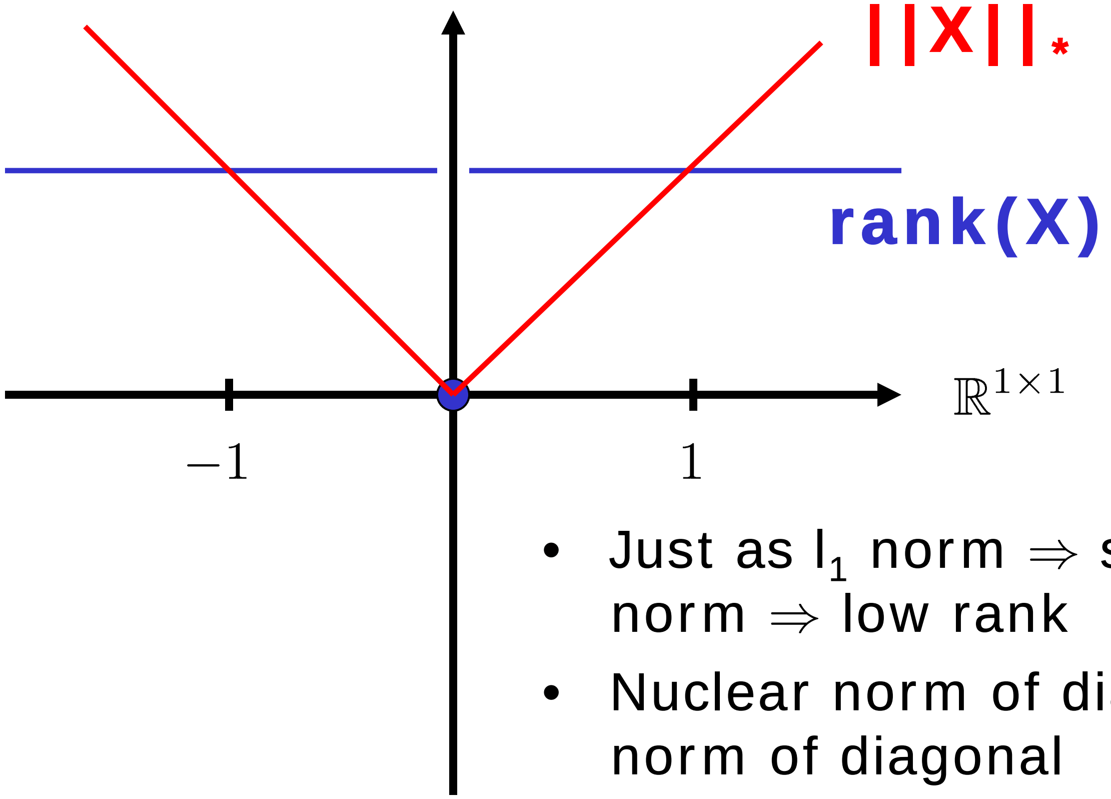
$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(\mathbf{X}) \\ \text{subject to} & \mathcal{A}(\mathbf{X}) = \mathbf{b} \end{array}$$

Convex Relaxation:

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

- Proposed by Fazel (2002).
- Nuclear norm is the “numerical rank” in numerical analysis
- The “trace heuristic” from controls if **X** is p.s.d.

Why nuclear norm?



- Just as l_1 norm ⇒ sparsity, nuclear norm ⇒ low rank
- Nuclear norm of diagonal matrix = l_1 norm of diagonal

Matrix and Vector Norms

- Vector $x \in \mathbb{R}^n$

$$\|x\| = \left(\sum_{t=1}^{n} x_t^2 \right)^{1/2}$$

$$\|x\|_\infty = \max_t |x_t|$$

$$\|x\|_1 = \sum_{t=1}^{n} |x_t|$$

- Matrix $X \in \mathbb{R}^{k \times n}$

- Singular Values
  $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$

$$\|X\|_F = \|\sigma\| = \left(\sum_{t=1}^{k} \sigma_t^2 \right)^{1/2}$$

$$\|X\| = \|\sigma\|_\infty = \max_t |\sigma_t|$$

$$\|X\|_* = \|\sigma\|_1 = \sum_{t=1}^{k} \sigma_t$$

- 2x2 matrices
- plotted in 3d

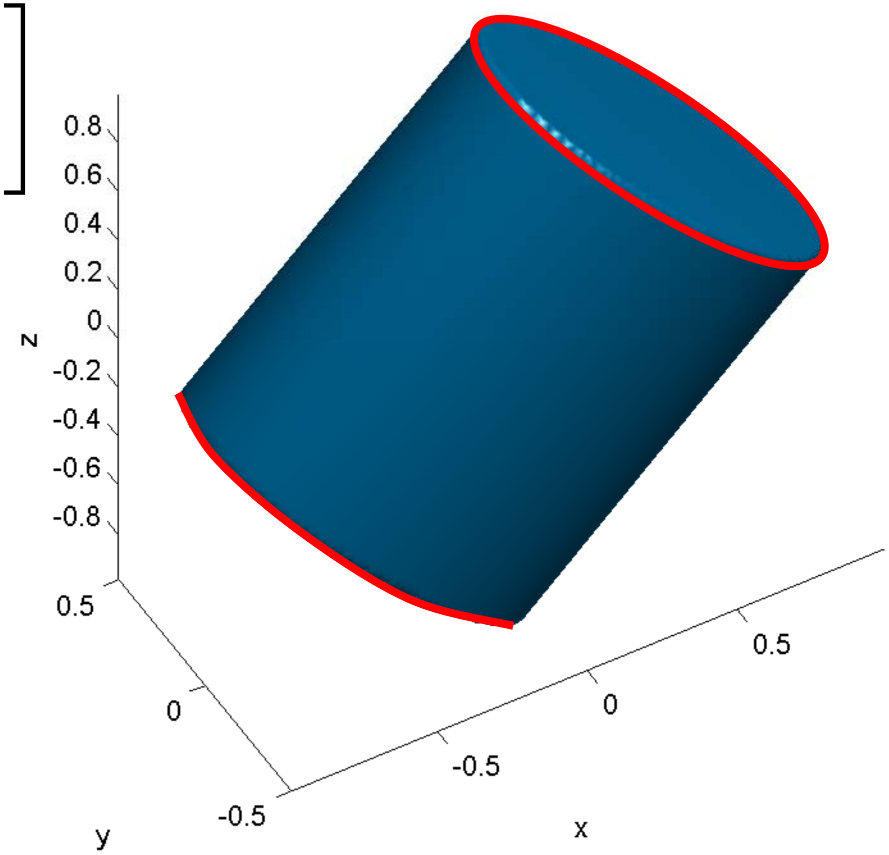
$$\begin{bmatrix} x & y \\ y & z \end{bmatrix}$$

— rank 1

$x^2 + z^2 + 2y^2 = 1$

Convex hull:

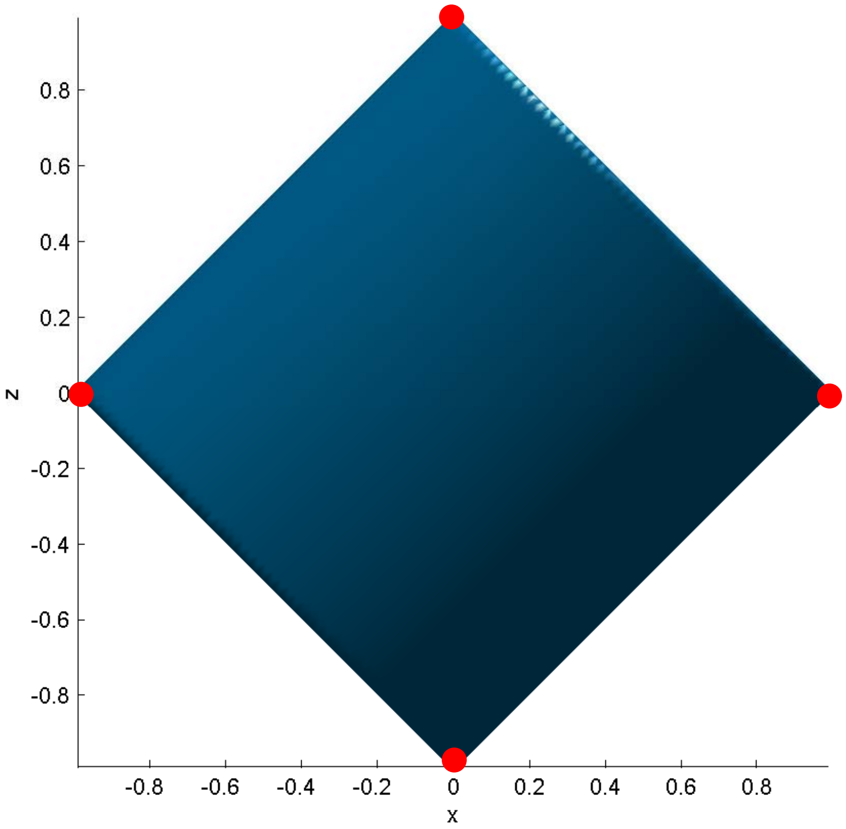
$$\{X \ : \ \|X\|_* \leq 1\}$$



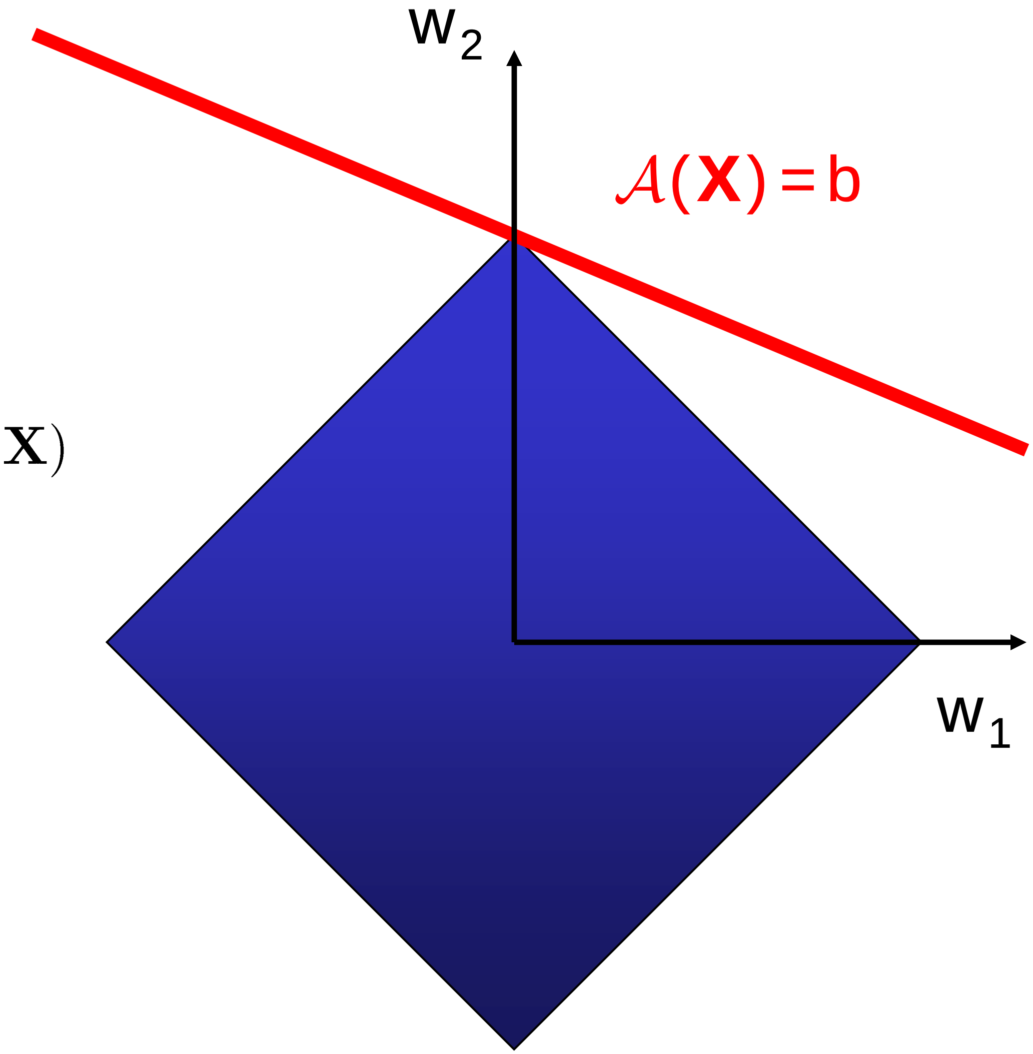
- 2x2 matrices
- plotted in 3d

$$\left\| \begin{bmatrix} x & 0 \\ 0 & z \end{bmatrix} \right\|_* \leq 1$$

- Projection onto x-z plane is l_1 ball



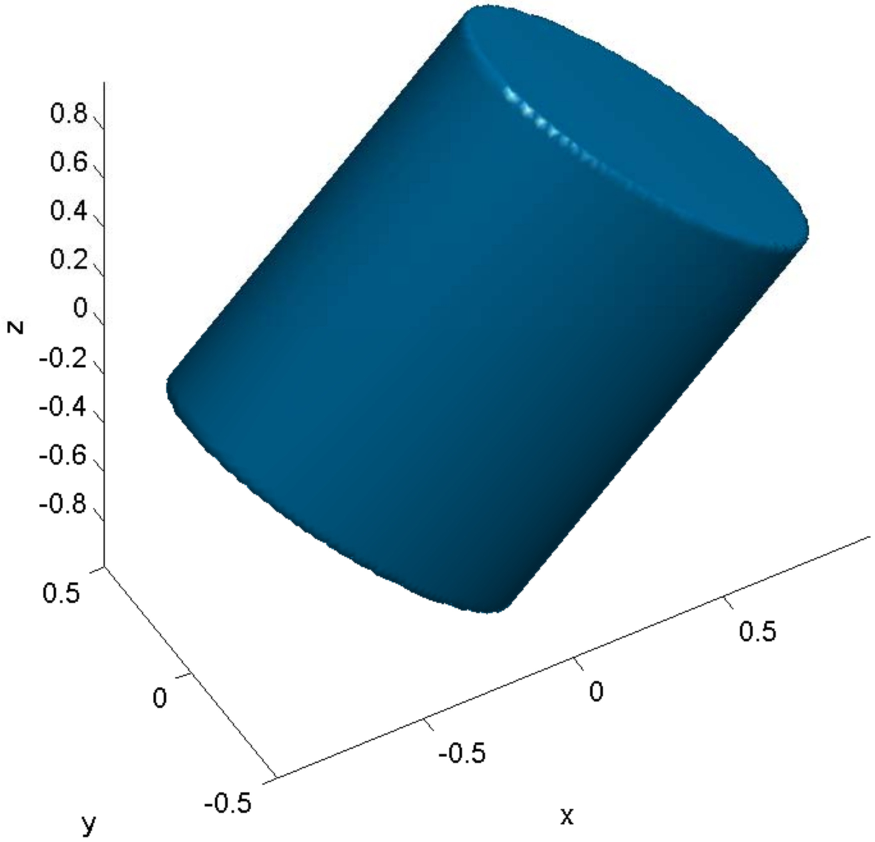
$$\begin{array}{ll} \text{minimize} & \|\mathbf{X}\|_* = \sum_{i=1}^k \sigma_i(\mathbf{X}) \\ \text{subject to} & \mathcal{A}(\mathbf{X}) = \mathbf{b} \end{array}$$



- 2x2 matrices
- plotted in 3d

$$\left\| \begin{bmatrix} x & y \\ y & z \end{bmatrix} \right\|_* \leq 1$$

- Not polyhedral...



So how do we compute it? And when does it work?

Equivalent Formulations

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

- Semidefinite embedding:

$$X = U\Sigma V^*$$

$$\left[\begin{array}{cc} W_1 & X \\ X^* & W_2 \end{array}\right] = \left[\begin{array}{c} U \\ V \end{array}\right] \Sigma \left[\begin{array}{c} U \\ V \end{array}\right]^*$$

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(\mathrm{Tr}(W_1) + \mathrm{Tr}(W_2)) \\ \text{subject to} & \left[\begin{array}{cc} W_1 & X \\ X^* & W_2 \end{array}\right] \succeq 0 \\ & \mathcal{A}(X) = b \end{array}$$

- Low rank parametrization:

$$L = U\Sigma^{1/2}$$

$$R = V\Sigma^{1/2}$$

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(\|L\|_F^2 + \|R\|_F^2) \\ \text{subject to} & \mathcal{A}(LR^*) = b \end{array}$$

Computationally: Gradient Descent

$$\mathcal{F}(\mathbf{L},\mathbf{R}) = \sum_{i=1}^{k}\sum_{j=1}^{r} L_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{r} R_{ij}^2 + \lambda \left\|\mathcal{A}(\mathbf{L}\mathbf{R}^*) - \mathbf{b}\right\|^2$$

- “Method of multipliers”
- Schedule for λ controls the noise in the data
- Same global minimum as nuclear norm
- Dual certificate for the optimal solution

- When will this fail and when it might succeed?

Restricted Isometry Property (RIP)

- Let $\mathcal{A}:\mathbb{R}^{k \times n} \to \mathbb{R}^m$ be a linear map. (Without loss of generality, assume k$\leq$ n throughout). For every positive integer r$\leq$k, define the r-restricted isometry constant to be the smallest number $\delta_r(\mathcal{A})$ 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.

- Directly adapted from RIP condition from Candès and Tao (2004).

RIP $\Rightarrow$ Unique Low-rank Solution

- **Theorem** Suppose that $\delta_{2r}(\mathcal{A})<1$ for some integer $r\geq 1$. Then there can be at most one matrix X with rank less than or equal to r satisfying $\mathcal{A}(X) = b$.

- **Proof:** Assume, on the contrary, that there exist two different matrices, X_1 and X_2, satisfying the matrix equation ($\mathcal{A}(X_1)=\mathcal{A}(X_2)=b$).
- Then $Z:=X_1-X_2$ is a nonzero matrix of rank at most 2r, and $\mathcal{A}(Z)=0$.
- But then we would have

$$0 = \|\mathcal{A}(Z)\| \geq (1-\delta_{2r}(\mathcal{A}))\|Z\|_F > 0$$

which is a contradiction.

RIP $\Rightarrow$ Heuristic Succeeds

- **Theorem:** Let X_0 be a matrix of rank r. Let X_* be the solution of $\mathcal{A}(X)=\mathcal{A}(X_0)$ of smallest nuclear norm. Suppose that $r \geq 1$ is such that $\delta_{5r}(\mathcal{A}) < 1/10$. Then $X_*=X_0$.

Independent of k,n,r,m

- Deterministic condition on $\mathcal{A}$
- No reason for estimate to be sharp

RIP $\Rightarrow$ Heuristic Succeeds

- **Theorem:** Let X_0 be a matrix of rank r. Let X_* be the solution of $\mathcal{A}(X)=\mathcal{A}(X_0)$ of smallest nuclear norm. Suppose that $r \geq 1$ is such that $\delta_{5r}(\mathcal{A}) < 1/10$. Then $X_*=X_0$.
- **Proof Sketch:** Let $R:=X_*-X_0$ be the error.
- The majority of the mass of R is concentrated in the row and column spaces of X_0.
- We can decompose $R = R_0 + R_1 + R_2 + \ldots$
  - R_0 is concentrated near the row and column space of X
  - R_i have rank at most 3r and orthogonal row/col spaces to X_0 for $i>0$
- Then we can show

$$\sum_{j\geq 2} \|R_j\|_F \leq \sqrt{\frac{2}{3}}\|R_0\|_F$$

RIP $\Rightarrow$ Heuristic Succeeds (cont)

$$\begin{aligned}
0 = \|\mathcal{A}(R)\| &\geq \|\mathcal{A}(R_0 + R_1)\| - \sum_{j\geq 2} \|\mathcal{A}(R_j)\| \\
&\geq (1-\delta_{5r})\, \|R_0 + R_1\|_F - (1+\delta_{3r}) \sum_{j\geq 2} \|R_j\|_F \\
&\geq \underbrace{\left((1-\delta_{5r}) - \sqrt{\tfrac{2}{3}}(1+\delta_{3r}) \right)}_{\textbf{Striclty positive for } \boldsymbol{\delta_{5r}<1/10}} \|R_0\|_F
\end{aligned}$$

Nearly Isometric RVs obey RIP

- **Theorem:** Fix $0<\delta<1$. If $\mathcal{A}$ is a nearly isometric random variable, then for every $1\leq r\leq k$, there exist constants c_0, $c_1>0$ depending only on δ such that $\delta_r(\mathcal{A})\leq\delta$ whenever $m\geq c_0\, r(k+n-r)\log(kn)$ with probability at least $1-\exp(-c_1 m)$.

- Number of measurements $\underbrace{c_0}_{\textbf{constant}}\ \underbrace{r(k+n-r)}_{\textbf{intrinsic dimension}}\ \underbrace{\log(kn)}_{\textbf{ambient dimension}}$

- Typical scaling for this type of result.

Generic Proof:

- Probability X is distorted is at most $\exp\left(-\alpha_1(\varepsilon)m\right)$

- I can cover all X with O(D^d) points where d is the intrinsic dimension and D is the embedded/ambient dimension

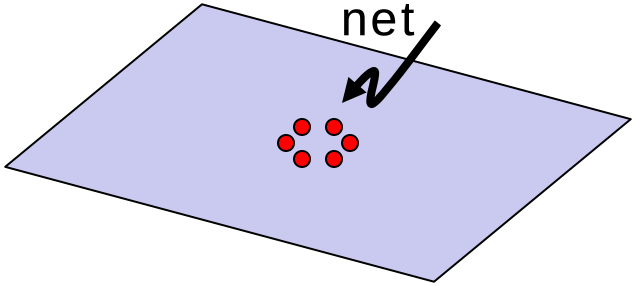
- Since nearby X's are distorted similarly, probability any X is distorted is at most $O\left(D^d \exp\left(-\alpha_2(\varepsilon)m\right)\right)$

- So no X is distorted with Prob at least 1-exp(-c_1m) if

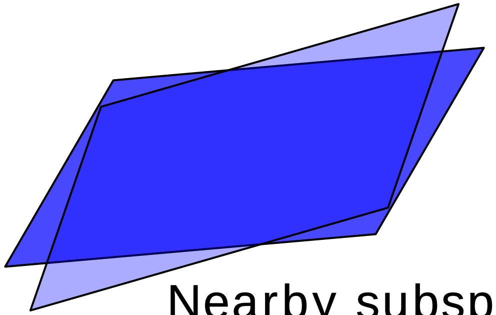
$$m > c_0 d \log D$$

Proof Sketch

- Show concentration holds for all matrices with same row and column space. (large deviations unlikely)
- Show that the distortion of a subspace of matrices by a linear map is robust to perturbations of the subspace. (maps have bounded norm)
- Provide an ε-net over the set of all subspaces of low-rank matrices (a Grassmann manifold). Show RIP holds at all points in the net with overwhelming probability and hence holds everywhere.



Apply large deviations property at an ε-net



Nearby subspaces have same distortion

The trace-norm heuristic succeeds!

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

- If m > c_0r(k+n-r)log(kn), the heuristic succeeds for most $\mathcal{A}$

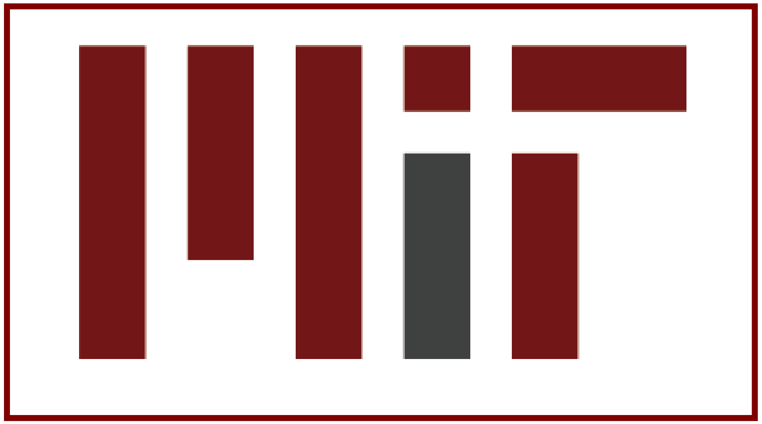
Recht, Fazel, and Parrilo. 2007.

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

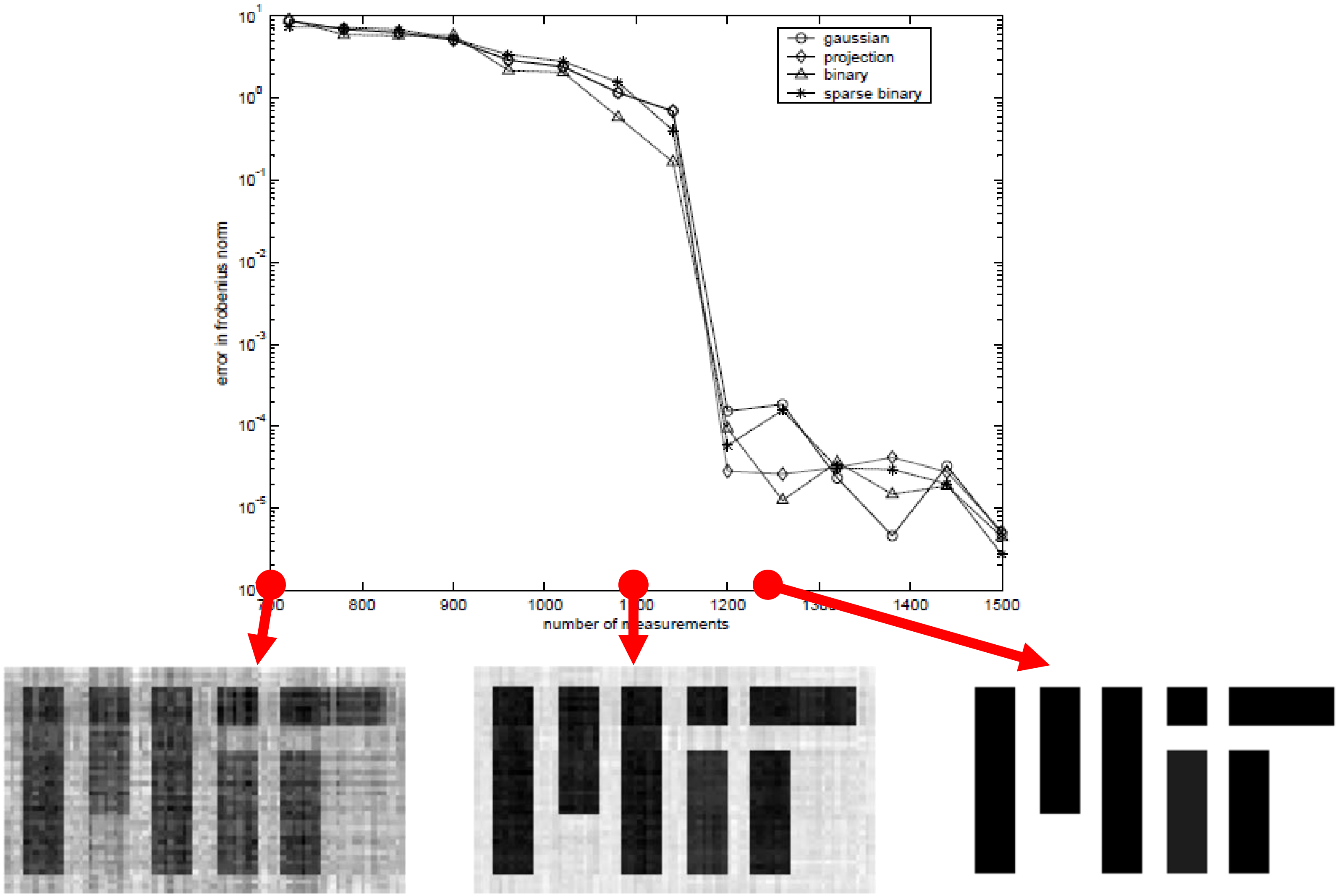
- **Approach:** Show that a random $\mathcal{A}$ is nearly an isometry on the manifold of rank 5r matrices.

Numerical Experiments

- Test “image”
- Rank 5 matrix, 46x81 pixels
- Random Gaussian measurements
- Nuclear norm minimization via SDP (sedumi)



Phase transition

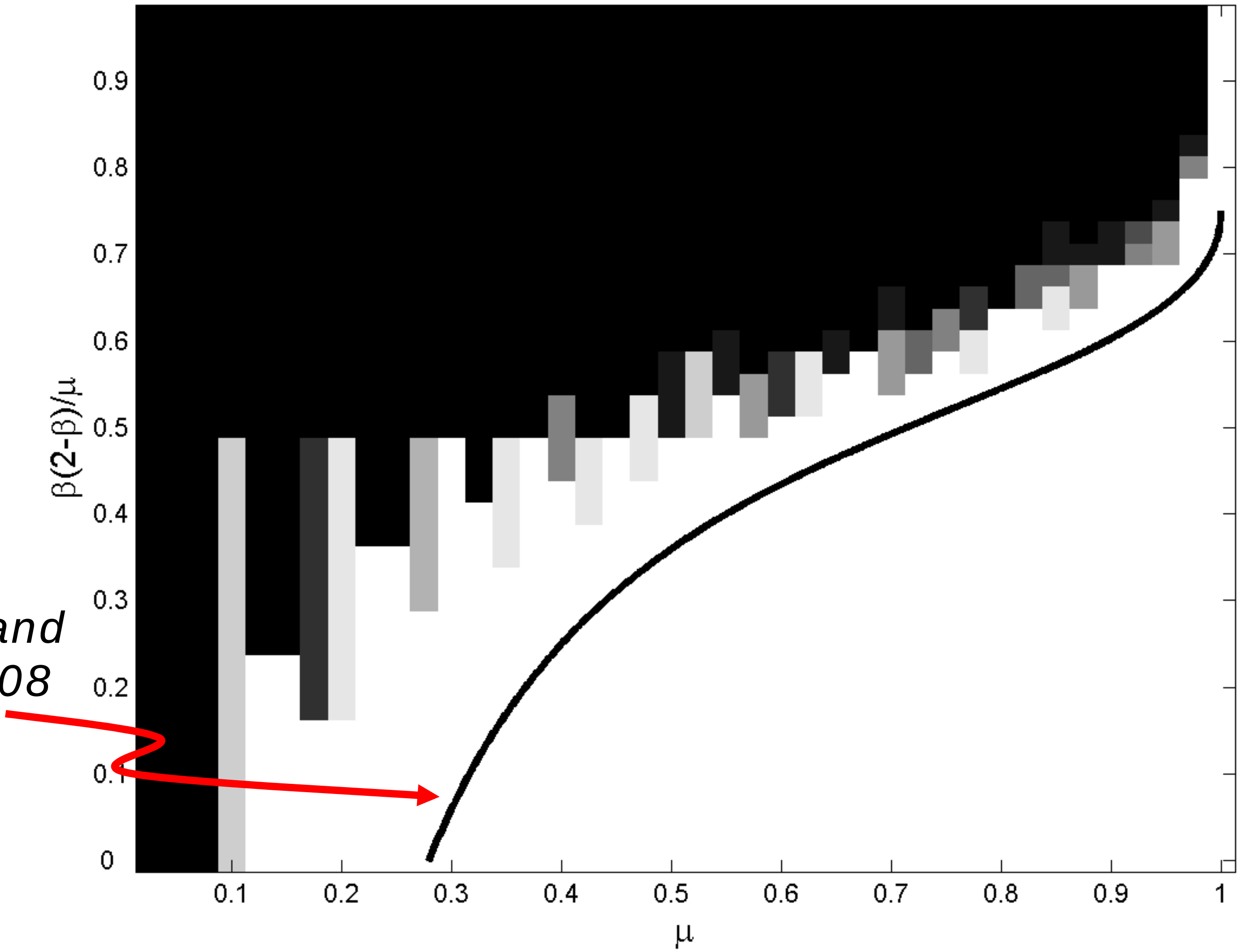


Phase transition

“Normalized” dimension of the rank r matrices
$\beta = r/n$

model-size vs measurements

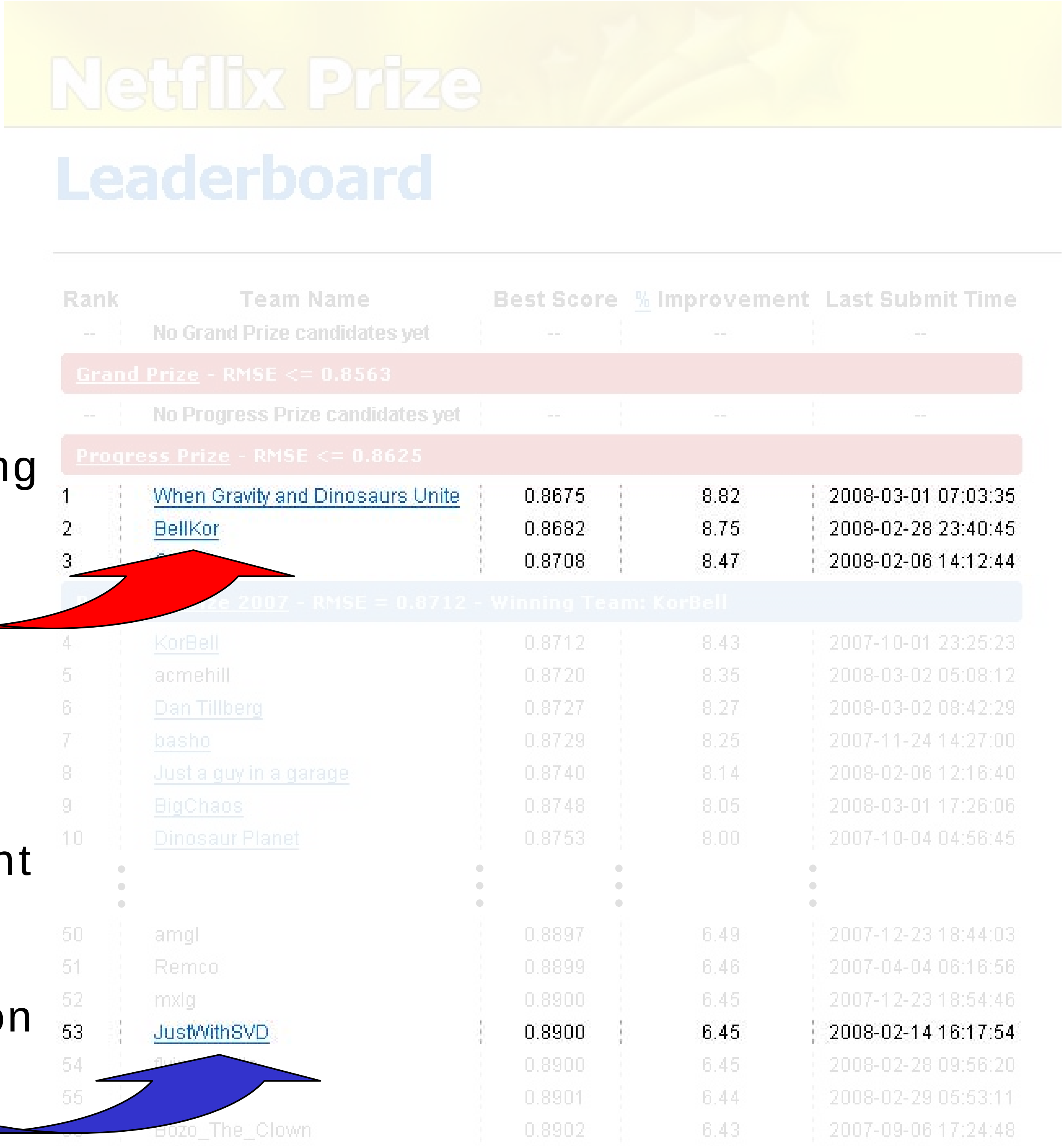
Recht, Xu, and Hassibi, 2008



measurements vs parameters: $\mu = m/n^2$

Mixture of hundreds of models, including nuclear norm

Gradient descent on low-rank nuclear norm parameterization



Netflix Prize

Leaderboard

Rank	Team Name	Best Score	% Improvement	Last Submit Time
--	No Grand Prize candidates yet	--	--	--
Grand Prize - RMSE <= 0.8563				
--	No Progress Prize candidates yet	--	--	--
Progress Prize - RMSE <= 0.8625				
1	When Gravity and Dinosaurs Unite	0.8675	8.82	2008-03-01 07:03:35
2	BellKor	0.8682	8.75	2008-02-28 23:40:45
3	[illegible]	0.8708	8.47	2008-02-06 14:12:44
[illegible]ize 2007 - RMSE = 0.8712 - Winning Team: KorBell				
4	KorBell	0.8712	8.43	2007-10-01 23:25:23
5	acmehill	0.8720	8.35	2008-03-02 05:08:12
6	Dan Tillberg	0.8727	8.27	2008-03-02 08:42:29
7	basho	0.8729	8.25	2007-11-24 14:27:00
8	Just a guy in a garage	0.8740	8.14	2008-02-06 12:16:40
9	BigChaos	0.8748	8.05	2008-03-01 17:26:06
10	Dinosaur Planet	0.8753	8.00	2007-10-04 04:56:45
⋮		⋮	⋮	⋮
50	amgl	0.8897	6.49	2007-12-23 18:44:03
51	Remco	0.8899	6.46	2007-04-04 06:16:56
52	mxlg	0.8900	6.45	2007-12-23 18:54:46
53	JustWithSVD	0.8900	6.45	2008-02-14 16:17:54
54	[illegible]	0.8900	6.45	2008-02-28 09:56:20
55	[illegible]	0.8901	6.44	2008-02-29 05:53:11
56	Bozo_The_Clown	0.8902	6.43	2007-09-06 17:24:48

Parsimonious Models

rank

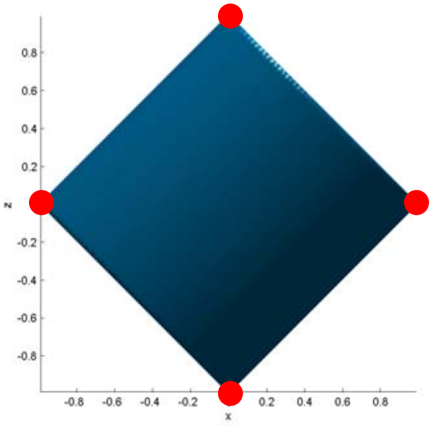
$$x = \sum_{k=1}^{r} w_k \alpha_k$$

model

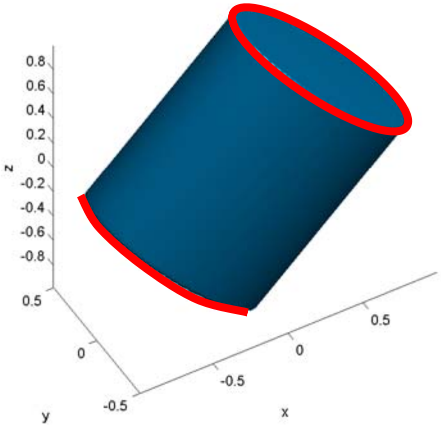
weights

atoms

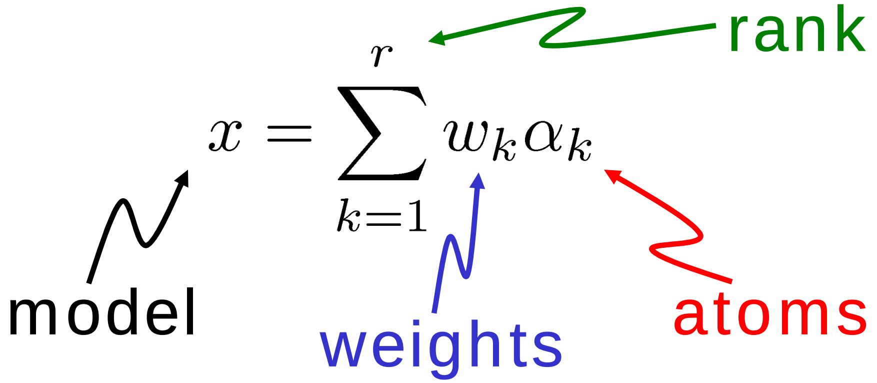
- Search for best linear combination of fewest atoms
- “rank” = fewest atoms needed to describe the model



$$\|x\|_{\mathcal{A}} \equiv \inf_{(w,\alpha)} \sum_{k=1}^{r} |w_k|$$



Other Directions



$$x = \sum_{k=1}^{r} w_k \alpha_k$$

rank

model

weights

atoms

- Random Features for Learning (Rahimi & Recht 07-08)
  - Atomic norm on basis functions
- Dynamical Systems
  - Atomic norm on filter banks
- Multivariate Tensors
  - Applications in genetics and vision
- Jordan Algebras, Polynomial Varieties, nonlinear models, completely positive matrices, ...

References

- "Some remarks on greedy algorithms." Ron DeVore and Vladimir Temlyakov. *Advances in Computational Mathematics*. **5**, pp. 173-187, 1996.
- "Decoding by Linear Programming." Emmanuel Candes and Terence Tao. *IEEE Transactions on Information Theory.* **51** (12), pp. 4203-4215, 2005.
- "Stable Signal Recovery from Incomplete and Inaccurate Measurements." Emmanuel Candes, Justin Romberg, and Terence Tao. **59** (8), pp. 1207 – 1223, 2006.
- "A Simple Proof of the Restricted Isometry Property for Random Matrices." R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. *Constructive Approximation*, **28**(3), pp. 253-263, 2008.
- "Guaranteed Minimum Rank Solutions to Linear Matrix Equations via Nuclear Norm Minimization." Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Submitted to *SIAM Review*. 2007.
- "Necessary and Sufficient Condtions for Success of the Nuclear Norm Heuristic for Rank Minimization." Benjamin Recht, Weiyu Xu, and Babak Hassibi. Submitted to *IEEE Transactions on Information Theory*. 2008.
- More extensions on my website: http://www.ist.caltech.edu/~brecht/publications.html