Rank minimization via the γ_2 norm

Troy Lee
Columbia University

Adi Shraibman
Weizmann Institute

Rank Minimization Problem

- Consider the following problem

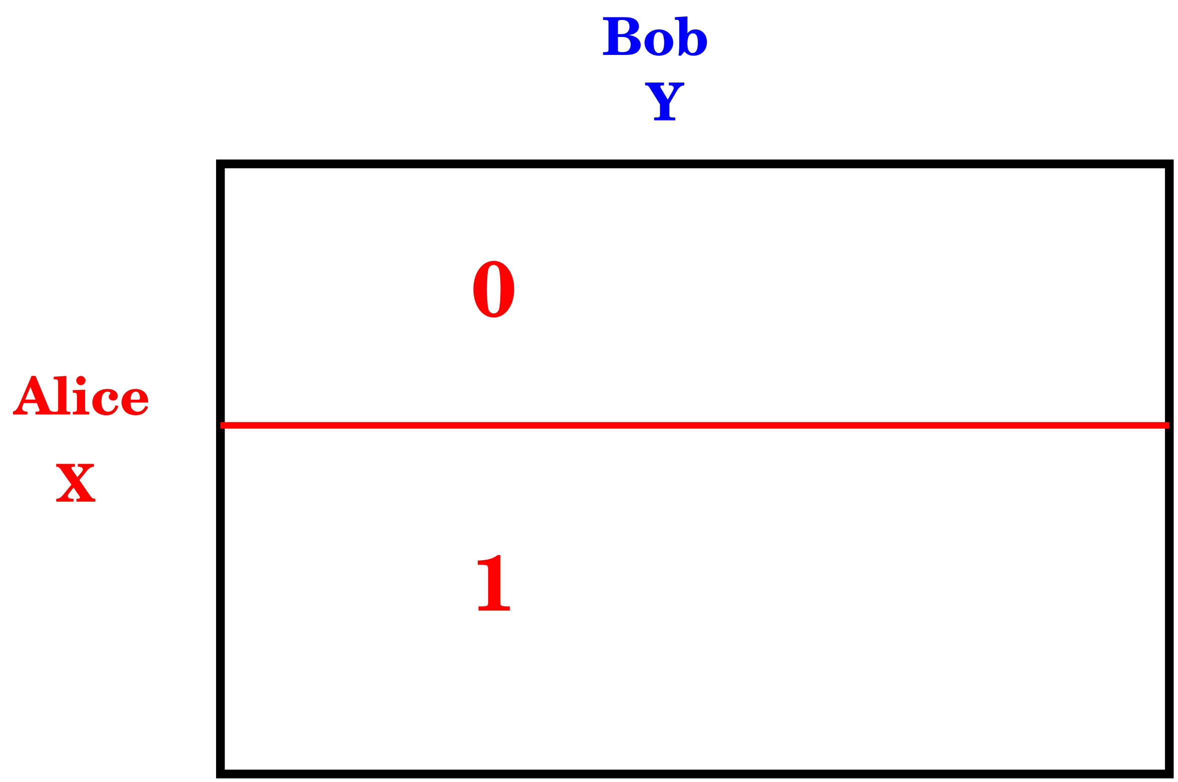
$$\min_X \operatorname{rank}(X)$$
$$\langle A_i, X\rangle \leq b_i \text{ for } i = 1, \ldots, k$$

- Arises in many contexts: complexity theory, recommendation systems, control theory
- Known to be NP-hard
- Optimization problem over a nonconvex function

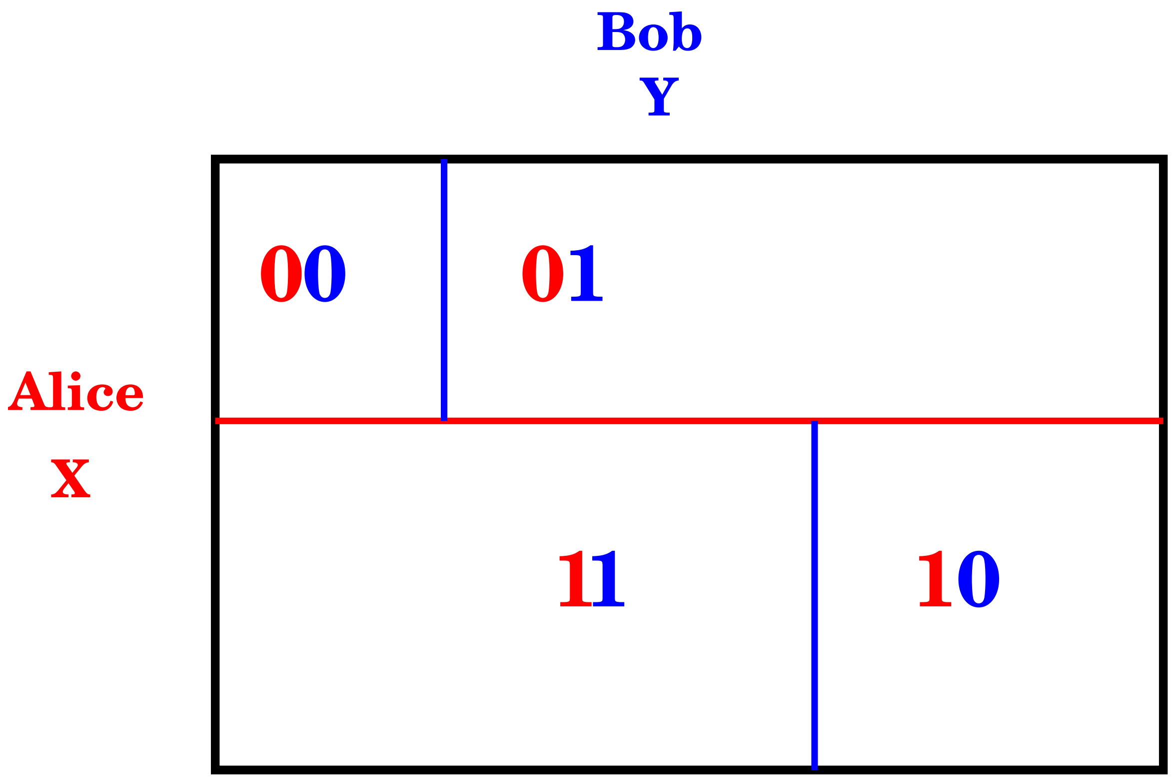
Communication complexity

- Two parties Alice and Bob wish to evaluate a function $f : X \times Y \to \{-1, +1\}$ where Alice holds $x \in X$ and Bob $y \in Y$.
- How much communication is needed? Can consider deterministic $D(f)$, randomized $R_\epsilon(f)$, and even quantum versions $Q_\epsilon(f)$
- Often convenient to work with $|X|$-by-$|Y|$ matrix A known as communication matrix where $A(x, y) = f(x, y)$. Allows tools from linear algebra to be applied.

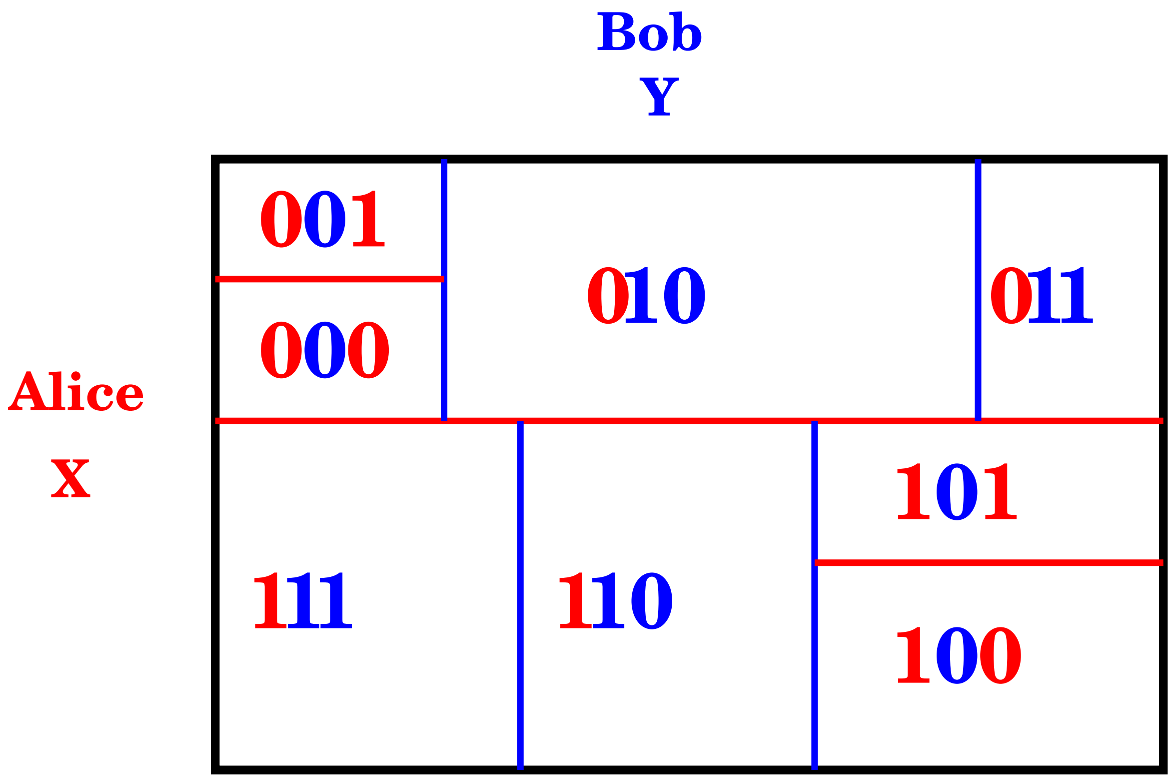
How a protocol partitions communication matrix



How a protocol partitions communication matrix



How a protocol partitions communication matrix



Log rank bound

- As a successful protocol partitions the communication matrix into rank one matrices we find

$$D(f) \geq \log \operatorname{rank}(A_f)$$

- One of the greatest open problems in communication complexity is the log rank conjecture [LS88], which states that $D(f) \leq (\log \operatorname{rank}(A_f))^k$ for some constant k.

Approximation rank

- As rank lower bounds deterministic communication complexity, the relevant quantity for randomized (and quantum) models is *approximation rank*.

- Given a target matrix A, find a low rank matrix entrywise close to A:

$$\mathrm{rank}_\epsilon(A) = \min_X \mathrm{rank}(X)$$
$$|X(i,j) - A(i,j)| \leq \epsilon \text{ for all } i,j$$

- Krause [Kra96] shows that

$$R_\epsilon(A) \geq \log \mathrm{rank}_\epsilon(A).$$

Approximation rank

- We do not know if approximation rank remains NP-hard to compute, but can be difficult in practice.

- For the "disjointness" matrix A with rows and columns labeled by n-bit strings where
$$A(x,y) = \begin{cases} -1 \text{ if } |x \cap y| = 0 \\ 1 \text{ otherwise.} \end{cases}$$
The $\epsilon = 1/3$ approximation rank is $2^{\Theta(\sqrt{n})}$ [Raz03, AA05].

Example 2: Matrix completion

- Popularly known as the "Netflix problem." Think of a M-by-N matrix where rows are labeled by users, columns are labeled by movies, and entries are "ratings."

- From a partial filling of this matrix—ratings supplied by some users—would like to make predictions for other users, i.e. fill out the rest of the matrix.

- A useful assumption: a user's rating depends on only on a few factors, thus the completed matrix should have low rank.

Example 2: Matrix completion

- Given a set $\Omega \subseteq M \times N$ of constraints $X(i,j) = a_{i,j}$ for $(i,j) \in \Omega$, find the lowest rank completion of X

$$\min_X \operatorname{rank}(X)$$
$$X(i,j) = a_{i,j} \text{ for all } (i,j) \in \Omega$$

- For applications, can think of a "hidden" low rank matrix A in the background. The goal is to exactly recover this matrix.

- Basic observations: rank r matrix has $O((M+N)r)$ degrees of freedom. In general, will need some assumptions for interesting results—think of matrix with only one nonzero entry.

Convex relaxations

- Part of the difficulty of the rank minimization problem is that it is an optimization problem over a nonconvex function.
- Much work has looked at substituting the rank function by a convex function.
- We will look at substituting rank function by different norms.

Matrix norms

- Define the i^{th} singular value as $\sigma_i(X) = \sqrt{\lambda_i(XX^t)}$
- Many useful matrix norms expressed in terms of vector of singular values $\sigma(X) = (\sigma_1(X), \ldots, \sigma_n(X))$.

$$\|X\|_1 = \ell_1(\sigma(X)) \text{ “trace norm”}$$
$$\|X\|_\infty = \ell_\infty(\sigma(X)) \text{ “spectral norm”}$$
$$\|X\|_2 = \ell_2(\sigma(X)) = \sqrt{\mathrm{Tr}(XX^t)} \text{ “Frobenius norm”}$$

Trace norm heuristic

- Popular heuristic is to replace rank by the trace norm

$$\min_X \|X\|_1$$
$$\langle A_i, X\rangle \le b_i \text{ for } i = 1, \ldots, k$$

- Motivation: rank is equal to the number of nonzero singular values, thus

$$\frac{\|X\|_1}{\|X\|_\infty} \le \operatorname{rank}(X).$$

- Over matrices satisfying $\|X\|_\infty \le 1$, trace norm is the largest convex lower bound on rank [Faz02].

Trace norm heuristic for matrix completion

- There has recently been a lot of work on the trace norm heuristic for the matrix completion problem [Faz02, RFP07, CR08, CT09].
- These results are of the form: Say that X is generated by taking N-by-r random Gaussian matrices Y and Z and setting $X = YZ^t$.
- Let $|\Omega| \geq Nr \log^7 n$ consist of entries of X sampled uniformly at random. Then with high probability the trace norm heuristic will *exactly* recover X.

Trace norm heuristic for approximation rank

- In the context of approximation rank and communication complexity, often work with sign matrices.

- Here the trace norm heuristic is better motivated by another simple inequality:
$$\|X\|_1 = \sum_i \sigma_i(X) \leq \sqrt{\mathrm{rank}(X)}\|X\|_2.$$

- For a M-by-N *sign* matrix A this simplifies nicely:

$$\mathrm{rank}(A) \geq \frac{\|A\|_1^2}{MN}$$

Trace norm bound on rank (example)

- Let H_N be a N-by-N Hadamard matrix (entries from $\{-1, +1\}$).
- Then $\|H_N\|_1 = N^{3/2}$.
- Trace norm method gives bound on rank of $N^3/N^2 = N$

Trace norm bound (drawback)

- As a complexity measure, the trace norm bound suffers one drawback—it is not monotone.

$$\begin{pmatrix} H_N & 1_N \\ 1_N & 1_N \end{pmatrix}$$

- Trace norm at most $N^{3/2} + 3N$

- Trace norm method gives

$$\frac{(N^{3/2} + 3N)^2}{4N^2} = \frac{N}{4} + O(\sqrt{N})$$

worse bound on whole than on H_N submatrix!

Trace norm method (a fix)

- We can fix this by considering

$$\max_{\substack{u,v: \\ \|u\|_2 = \|v\|_2 = 1}} \|A \circ uv^t\|_1$$

- As $\mathrm{rank}(A \circ uv^t) \leq \mathrm{rank}(A)$ we still have

$$\mathrm{rank}(A) \geq \left(\frac{\|A \circ uv^t\|_1}{\|A \circ uv^t\|_2} \right)^2$$

The γ_2 norm

- This bound simplifies nicely for a sign matrix A

$$\mathrm{rank}(A) \geq \max_{\substack{u,v: \\ \|u\|_2=\|v\|_2=1}} \left(\frac{\|A \circ uv^t\|_1}{\|A \circ uv^t\|_2} \right)^2 = \max_{\substack{u,v: \\ \|u\|_2=\|v\|_2=1}} \|A \circ uv^t\|_1^2$$

- We have arrived at the γ_2 norm introduced to communication complexity by [LMSS07, LS07]

$$\gamma_2(A) = \max_{\substack{u,v: \\ \|u\|_2=\|v\|_2=1}} \|A \circ uv^t\|_1$$

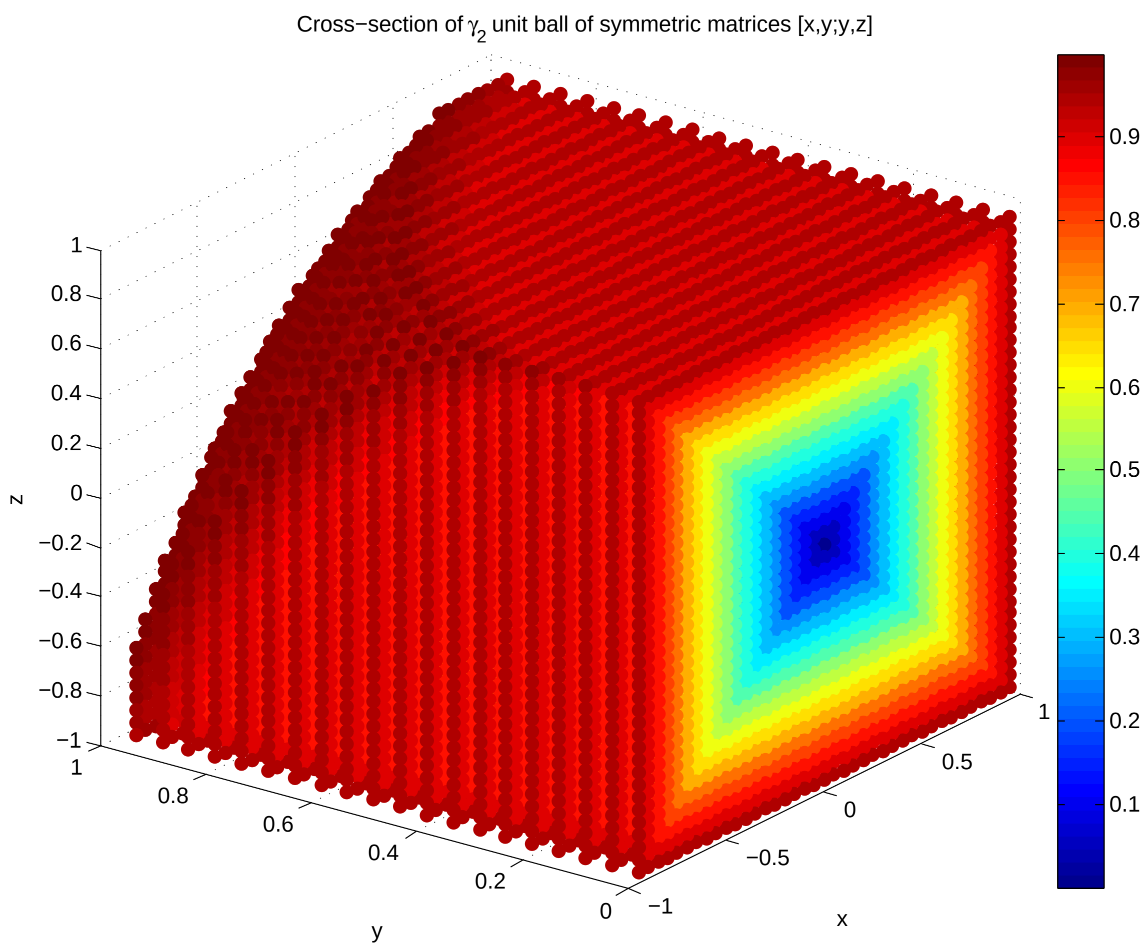
γ_2 norm: Surprising usefulness

- γ_2 is a norm, though not a matrix norm. In matrix analysis known as "Schur/Hadamard product operator/trace norm."

- Schur (1911) showed that $\gamma_2(A) = \max_i A_{ii}$ if A positive semidefinite.

- $\gamma_2(A)$ can be written as a semidefinite program and so can be well approximated in time polynomial in the size of A.

- The dual norm $\gamma_2^*(A) = \max_B \langle A, B \rangle / \gamma_2(B)$ turns up in semidefinite programming relaxation of MAX-CUT of Goemans and Williamson, and is closely related to the discrepancy method in communication complexity.


Cross−section of γ_2 unit ball of symmetric matrices [x,y;y,z]
z
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
y
1
0.8
0.6
0.4
0.2
0
x
−1
−0.5
0
0.5
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

Approximation rank with γ_2

- Substitute rank by γ_2 in the approximation rank optimization problem:

$$\gamma_2^\epsilon(A) = \min_X \gamma_2(X)$$

$$|A(i,j) - X(i,j)| \leq \epsilon \text{ for all } (i,j).$$

- Main theorem: For any M-by-N sign matrix A and constant $0 < \epsilon < 1/2$

$$\frac{\gamma_2^\epsilon(A)^2}{(1+\epsilon)^2} \leq \mathrm{rank}_\epsilon(A) = O\left(\gamma_2^\epsilon(A)^2 \log(MN)\right)^3$$

Proof sketch

- We introduced γ_2 as a maximization problem.
- For the proof, we use an alternative characterization of γ_2 in terms of a minimization problem.
- Trace norm can be written as

$$\|X\|_1 = \min_{Y,Z:X=YZ^t} \|Y\|_2 \|Z\|_2$$

  This follows from singular value decomposition: $X = U\Sigma V$ where U, V unitary.

Min formulation of γ_2

$$\gamma_2(X) = \max_{u,v : \|u\|_2 = \|v\|_2 = 1} \|X \circ uv^t\|_1$$

$$= \max_{u,v} \min_{\substack{Y,Z \\ X = YZ^t}} \|D(u)Y\|_2 \, \|Z^t D(v)\|_2$$

$$\text{“}=\text{”} \min_{\substack{Y,Z \\ X = YZ^t}} \max_{u,v} \|D(u)Y\|_2 \, \|Z^t D(v)\|_2$$

$$= \min_{\substack{Y,Z \\ X = YZ^t}} \|Y\|_r \|Z\|_r$$

where $\|Y\|_r$ is the largest ℓ_2 norm of a row of Y.

First step: dimension reduction

- Thus γ_2 looks for factorization $X = YZ^t$ where Y, Z have short rows in terms of ℓ_2 norm.

- Similarly rank looks for factorization $X = YZ^t$ where Y, Z have short rows in terms of *dimension*.

- Use Johnson-Lindenstrauss lemma to project rows of Y, Z to dimension about equal to $\gamma_2(X)^2$. Let R be a random K'-by-K matrix

$$\Pr_R \left[\langle Ru, Rv \rangle - \langle u, v \rangle \geq \frac{\delta}{2}(\|u\|^2 + \|v\|^2) \right] \leq 4e^{-\delta^2 K'/8}$$

Second step: error reduction

- After the first step, we obtain a new matrix X' of rank $O(\gamma_2^\epsilon(A)^2 \log N)$ but the approximation factor has worsened—X' is only 2ϵ close to A.

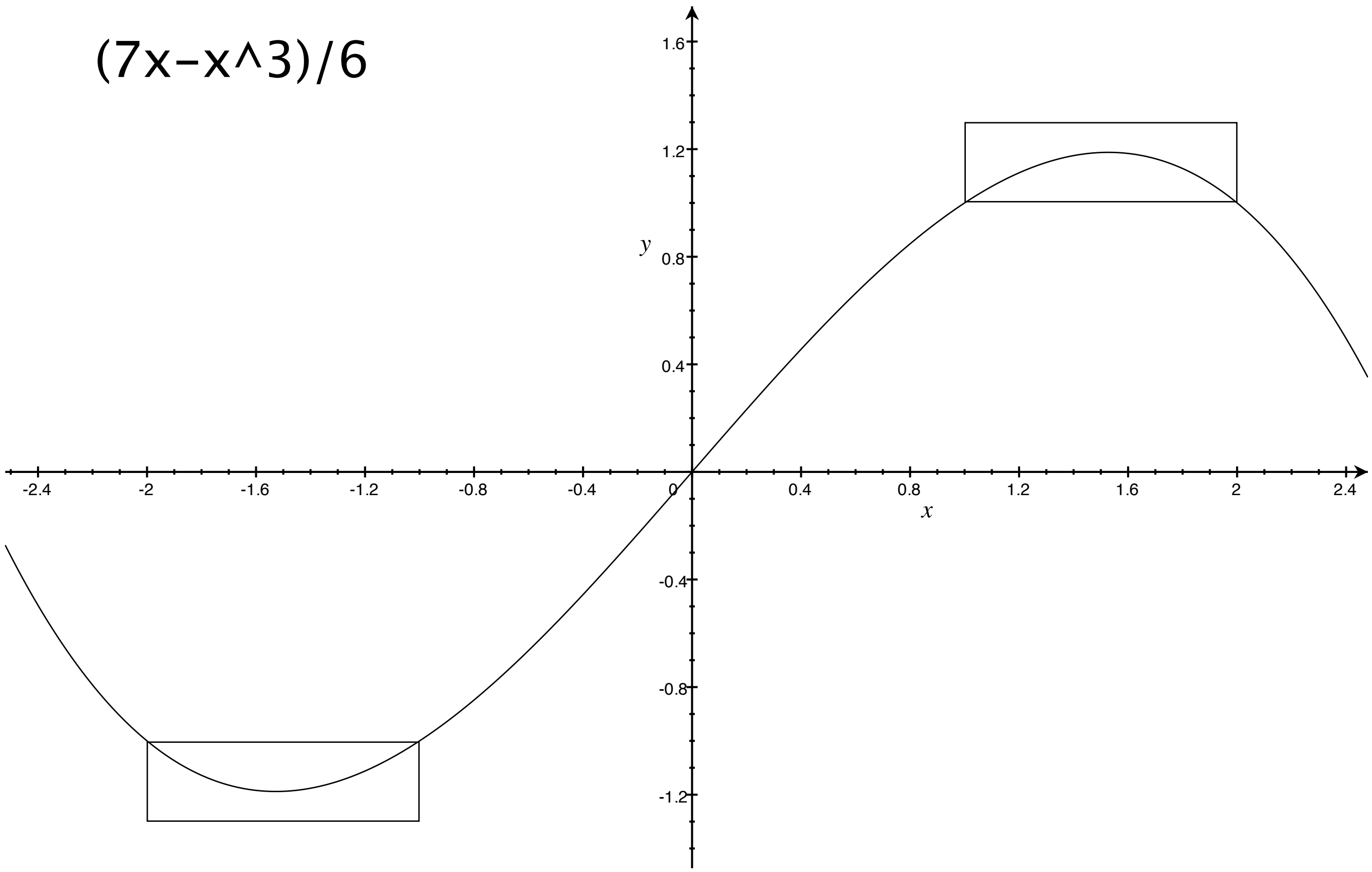
- Trick going back to Krivine: Apply a low degree polynomial entrywise to the matrix X'.

$$p(M) = a_0 J + a_1 M + \ldots + a_d M^{\circ d}.$$

- See that $\mathrm{rank}(p(M)) \leq (d+1)\mathrm{rank}(M)^d$. Taking p to be approximation to the sign function reduces error.

Polynomial for Error Reduction

$(7x-x^3)/6$



Final result

- For any M-by-N sign matrix A and constant $0 < \epsilon < 1/2$

$$\frac{\gamma_2^{\epsilon}(A)^2}{(1+\epsilon)^2} \leq \mathrm{rank}_{\epsilon}(A) = O\left(\gamma_2^{\epsilon}(A)^2 \log(MN)\right)^3$$

- Logarithmic factor is necessary as (sign version of) identity matrix has approximation rank $\log(N)$ [Alo03] but constant γ_2.

- [BES02] used dimension reduction to upper bound sign rank by $\gamma_2^{\infty}(A)^2$. Interestingly, here the *lower bound* fails.

Extension to general rank minimization problem

- Consider again the general rank minimization problem

$$\alpha(\mathbf{A}, \mathbf{b}) = \min_{X} \ \mathrm{rank}(X)$$
$$\langle A_i, X \rangle \le b_i \text{ for } i = 1, \ldots, k$$

- Let $\mathcal{C} = \{X : \langle A_i, X \rangle \le b_i\}$ be the feasible set.

- Let $\ell_\infty(\mathcal{C}) = \max_{X \in \mathcal{C}} \ \ell_\infty(X)$.

Extension to general rank minimization problem

- As argued before we have

$$\alpha(\mathbf{A}, \mathbf{b}) \geq \min_{X \in \mathcal{C}} \frac{\gamma_2(X)^2}{\ell_\infty(\mathcal{C})^2}.$$

- Say we solve via semidefinite programming the program $\min_{X \in \mathcal{C}} \ \gamma_2(X)$

- Then by doing dimension reduction on an optimal X^* we obtain a matrix Y of rank about $\gamma_2(X^*)^2 \log(N)$ which is ϵ close to X^*.

- This matrix will satisfy the i^{th} constraint up to a factor $\epsilon \ \ell_1(A_i)$.

Application to the matrix completion problem

- In matrix completion, often is a natural bound on $\ell_\infty(\mathcal{C})$. For example, Netflix uses ratings $\{1, 2, 3, 4, 5\}$ so matrix will be bounded.

- Also in the matrix completion problem each constraint matrix A_i consists of a single entry so $\ell_1(A_i) = 1$.

- Thus if the lowest rank completion has rank d, via γ_2 we can find a rank $O(d \log(N))$ matrix which is ϵ-close on the specified entries.

Open questions

- Approximation algorithm for the limiting case of sign rank?
- For matrix completion, can one show similar unconditional results for trace norm heuristic?
- Practical implementations of γ_2 for large matrices.