Approximation norms and duality for communication complexity lower bounds

Troy Lee
Columbia University

Adi Shraibman
Weizmann Institute

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?

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? Many different models have been studied.

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? Many different models have been studied.

- Randomized complexity $R_\epsilon(f)$ with error probability ϵ.

- Quantum complexity $Q_\epsilon(f)$ without shared entanglement and $Q^*_\epsilon(f)$ with shared entanglement.

Open questions

- Many open questions remain relating the power of these different models.

Open questions

- Many open questions remain relating the power of these different models.

- Are $R_\epsilon(f)$ and $Q^*_\epsilon(f)$ polynomially related for all total functions f?

Open questions

- Many open questions remain relating the power of these different models.

- Are $R_\epsilon(f)$ and $Q^*_\epsilon(f)$ polynomially related for all total functions f? Largest gap known is a power of 2.

Open questions

- Many open questions remain relating the power of these different models.

- Are $R_\epsilon(f)$ and $Q^*_\epsilon(f)$ polynomially related for all total functions f? Largest gap known is a power of 2.

- How much can entanglement help? What is the largest gap between $Q_\epsilon(f)$ and $Q^*_\epsilon(f)$.

Open questions

- Many open questions remain relating the power of these different models.

- Are $R_\epsilon(f)$ and $Q^*_\epsilon(f)$ polynomially related for all total functions f? Largest gap known is a power of 2.

- How much can entanglement help? What is the largest gap between $Q_\epsilon(f)$ and $Q^*_\epsilon(f)$. Currently, the only uses of entanglement to save communication are as a source of shared randomness, and for superdense coding.

Lower bound techniques

- Nearly all lower bounds known for R_ϵ also work in the more powerful model Q^*_ϵ, up to small factors.

- Exceptions: "Corruption bound" which can show $\Omega(n)$ lower bound on randomized complexity of disjointness [KS87, Raz92].

Lower bound techniques

- Nearly all lower bounds known for R_ϵ also work in the more powerful model Q^*_ϵ, up to small factors.

- Exceptions: "Corruption bound" which can show $\Omega(n)$ lower bound on randomized complexity of disjointness [KS87, Raz92].

- "log rank bound" known to work for Q_ϵ [BW01] but not Q^*_ϵ.

Lower bound techniques

- Nearly all lower bounds known for R_ϵ also work in the more powerful model Q^*_ϵ, up to small factors.

- Exceptions: "Corruption bound" which can show $\Omega(n)$ lower bound on randomized complexity of disjointness [KS87, Raz92].

- "log rank bound" known to work for Q_ϵ [BW01] but not Q^*_ϵ.

- In this talk we focus on the log rank bound.

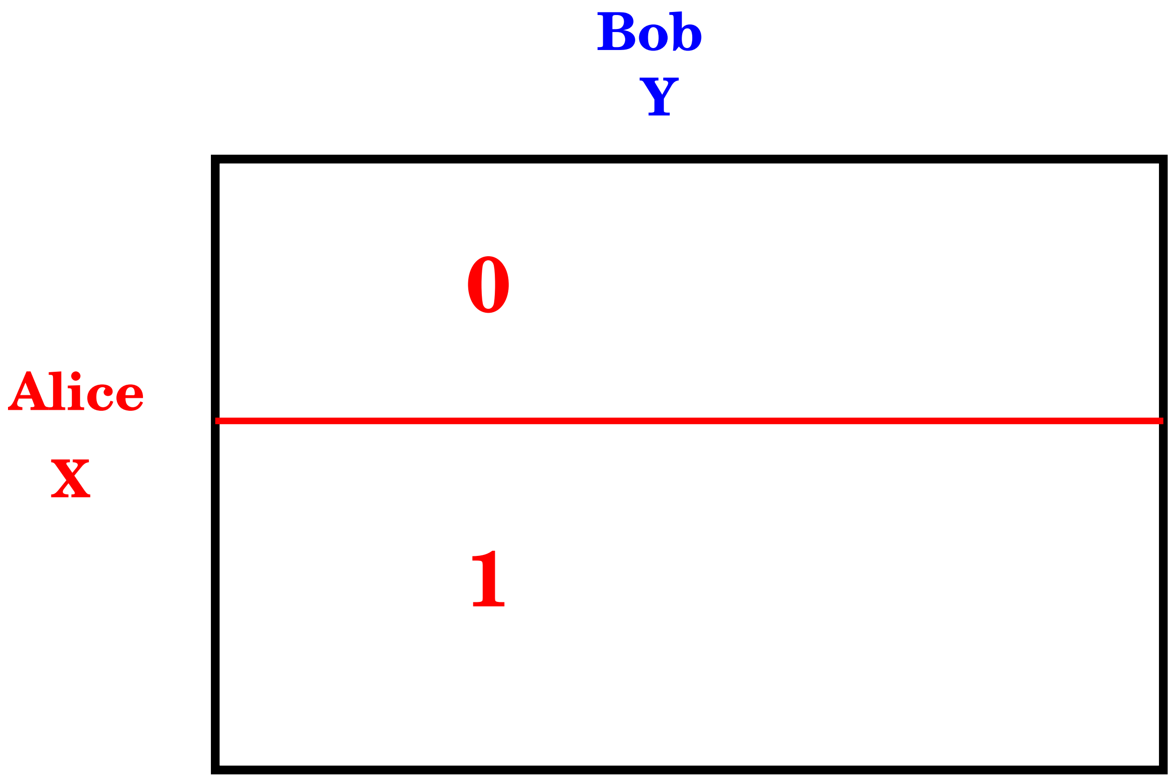
Log rank lower bound

- To a function $f : X \times Y \to \{-1, +1\}$ we associate a X-by-Y communication matrix M_f, where $M_f[x, y] = f(x, y)$.

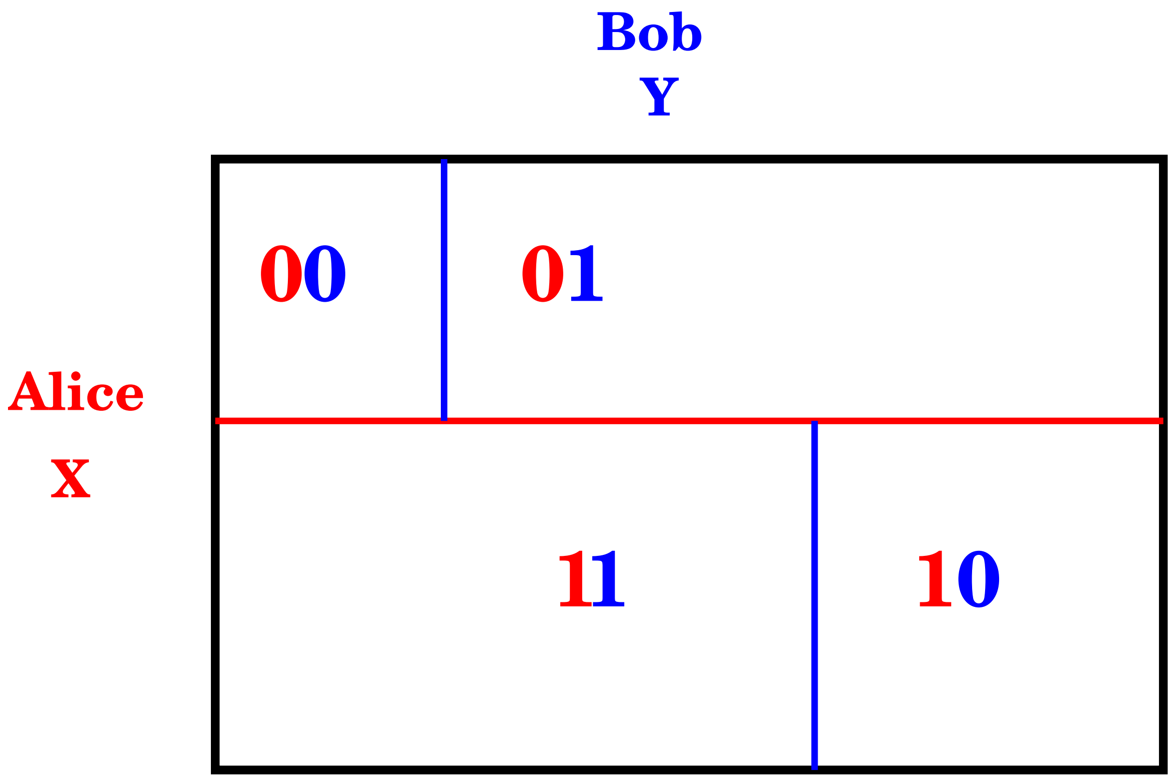
- The log rank bound states $D(f) \geq \log \mathrm{rk}(M_f)$ [MS82].

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

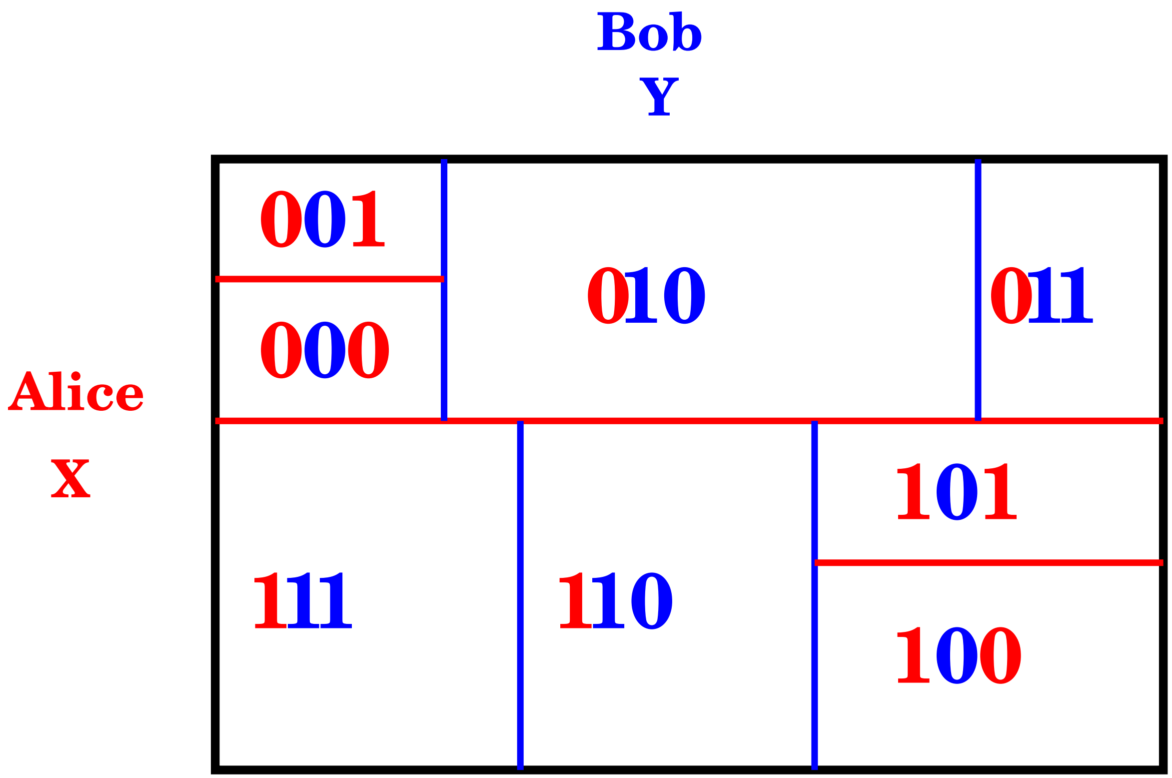
How a protocol partitions communication matrix



How a protocol partitions communication matrix



How a protocol partitions communication matrix



Approximation rank

- For randomized and quantum models, the relevant quantity is no longer rank, but approximation rank. For a sign matrix A:

$$\mathrm{rk}_\alpha(A) = \min_B \{\mathrm{rk}(B) : 1 \leq A[i,j]B[i,j] \leq \alpha\}$$

Approximation rank

- For randomized and quantum models, the relevant quantity is no longer rank, but approximation rank. For a sign matrix A:

$$\mathrm{rk}_\alpha(A) = \min_B \{\mathrm{rk}(B) : 1 \leq A[i,j]B[i,j] \leq \alpha\}$$

- Buhrman and de Wolf show

$$R_\epsilon(f) \geq Q_\epsilon(f) \geq \frac{\log \mathrm{rk}_\alpha(M_f)}{2}$$

for $\alpha = 1/(1-2\epsilon)$.

Main result

- Approximation rank is essentially the strongest technique available to show lower bounds on quantum communication complexity.

Main result

- Approximation rank is essentially the strongest technique available to show lower bounds on quantum communication complexity. But it suffers from two drawbacks: it is not known to be a lower bound on complexity with entanglement, and it can be quite difficult to compute in practice.

Main result

- Approximation rank is essentially the strongest technique available to show lower bounds on quantum communication complexity. But it suffers from two drawbacks: it is not known to be a lower bound on complexity with entanglement, and it can be quite difficult to compute in practice.

- We show
$$Q^*_\epsilon(f) = \Omega(\log \mathrm{rk}_\alpha(M_f))$$
for $\alpha = 1/(1-2\epsilon)$.

Main result

- Approximation rank is essentially the strongest technique available to show lower bounds on quantum communication complexity. But it suffers from two drawbacks: it is not known to be a lower bound on complexity with entanglement, and it can be quite difficult to compute in practice.

- We show
$$Q^*_\epsilon(f) = \Omega(\log \mathrm{rk}_\alpha(M_f))$$
for $\alpha = 1/(1-2\epsilon)$.

- We further give a (randomized) polynomial time approximation algorithm for $\log \mathrm{rk}_\alpha(A)$.

γ_2 norm

- Both results will be obtained by relating approximation rank to a norm known as γ_2 introduced to quantum communication complexity by Linial and Shraibman [LS07].

- Linial and Shraibman show that γ_2 gives a lower bound on quantum communication complexity with entanglement, and that it generalizes many other bounds in the literature, including discrepancy [Kre95], Fourier bounds [Kla01], trace norm method [Raz03].

γ_2 norm

- Both results will be obtained by relating approximation rank to a norm known as γ_2 introduced to quantum communication complexity by Linial and Shraibman [LS07].

- Linial and Shraibman show that γ_2 gives a lower bound on quantum communication complexity with entanglement, and that it generalizes many other bounds in the literature, including discrepancy [Kre95], Fourier bounds [Kla01], trace norm method [Raz03].

- On the other hand, $\mathrm{rk}(A) \geq \gamma_2(A)^2$.

γ_2 norm definition

- For a matrix A, define

$$\gamma_2(A) = \min_{X^T Y = A} c(X)c(Y)$$

where $c(X)$ is the largest ℓ_2 norm of a column of X.

γ_2 norm definition

- For a matrix A, define

$$\gamma_2(A) = \min_{X^T Y = A} c(X)c(Y)$$

where $c(X)$ is the largest ℓ_2 norm of a column of X.

- As with rank, we also consider an approximation version: for a sign matrix A

$$\gamma_2^\alpha(A) = \min_B \{\gamma_2(B) : 1 \leq A[i,j]B[i,j] \leq \alpha\}.$$

γ_2 norm remarks

- In matrix analysis known as "Schur/Hadamard product operator/trace norm,"

γ_2 norm remarks

- 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.

γ_2 norm remarks

- 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.
- We will also use the dual norm:

$$\gamma_2^*(A) = \max_B \frac{\langle A, B \rangle}{\gamma_2(B)}$$

γ_2 norm remarks

- 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.

- We will also use the dual norm:

$$\begin{aligned}
\gamma_2^*(A) &= \max_B \frac{\langle A, B \rangle}{\gamma_2(B)} \\
&= \max_{\substack{u_i, v_j: \\ \|u_i\| = \|v_j\| = 1}} \sum_{i,j} A[i,j] \langle u_i, v_j \rangle
\end{aligned}$$

Dual norm

- The dual norm γ_2^* shows up in XOR games with entanglement.

- This is a game between a verifier and two provers Alice and Bob. Alice and Bob share an entangled state. Verifier wants to compute some function $f : X \times Y \to \{-1, +1\}$.

- Verifier sends questions x to Alice, y to Bob with probability $\pi(x, y)$.

- Alice/Bob respond with $a_x, b_y \in \{-1, +1\}$ with the aim that $a_x b_y = f(x, y)$.

Tsirelson's characterization

- Look at the correlation, under π between the function f and the output of the protocol.

- Tsirelson's characterization of XOR games gives

$$\max_{\text{strategies}} \sum_{x,y} \pi(x,y) f(x,y) a_x b_y = \max_{\substack{u_x, v_y: \\ \|u_x\|=\|v_y\|=1}} \sum_{x,y} \pi(x,y) M_f[x,y] \langle u_x, v_y \rangle$$
$$= \gamma_2^*(M_f \circ \pi).$$

γ_2 communication complexity lower bound

- Tsirelson's characterization can give an alternative proof that γ_2 lower bounds quantum communication complexity with entanglement (observed by Harry Buhrman).

γ_2 communication complexity lower bound

- Tsirelson's characterization can give an alternative proof that γ_2 lower bounds quantum communication complexity with entanglement (observed by Harry Buhrman).

- Recall
$$\gamma_2(M_f) = \max_{g,\pi} \frac{\langle M_f, M_g \circ \pi \rangle}{\gamma_2^*(M_g \circ \pi)}$$

- Consider a c-qubit protocol for f. Using teleportation, we may transform this into a protocol that uses at most $2c$ classical bits.

- We will now show that $\gamma_2^*(M_g \circ \pi)$ is large by designing an XOR strategy for the provers.

XOR strategy for provers

- We design an XOR strategy P. Alice and Bob share a random $2c$ bit string r. Alice and Bob simulate actions of the protocol for f, assuming i^{th} message sent is r_i.
- If Alice/Bob notices inconsistency with protocol outputs a random bit.
- If Alice consistent outputs $f(x,y)$. If Bob consistent outputs 1.
- Then

$$\gamma_2(M_g \circ \pi) \geq \sum_{x,y} \pi(x,y) g(x,y) P(x,y) = \frac{1}{2^{2c}} \sum_{x,y} \pi(x,y) g(x,y) f(x,y)$$

XOR strategy for provers

- From the last slide we have

$$\gamma_2(M_g \circ \pi) \geq \sum_{x,y} \pi(x,y) g(x,y) P(x,y) = \frac{1}{2^{2c}} \sum_{x,y} \pi(x,y) g(x,y) f(x,y)$$

- As g, π were arbitrary this gives

$$\max_{g,\pi} \frac{\langle M_f, M_g \circ \pi \rangle}{\gamma_2^*(M_g \circ \pi)} \leq 2^{2c}$$

which implies $Q^*(f) = \Omega(\log \gamma_2(M_f))$.

XOR strategy for provers

- From the last slide we have

$$\gamma_2(M_g \circ \pi) \geq \sum_{x,y} \pi(x,y)g(x,y)P(x,y) = \frac{1}{2^{2c}} \sum_{x,y} \pi(x,y)g(x,y)f(x,y)$$

- As g, π were arbitrary this gives

$$\max_{g,\pi} \frac{\langle M_f, M_g \circ \pi \rangle}{\gamma_2^*(M_g \circ \pi)} \leq 2^{2c}$$

which implies $Q^*(f) = \Omega(\log \gamma_2(M_f))$.The proof for bounded-error complexity follows similarly.

Relating γ_2 and rank

- Now that we have introduced γ_2, we can state our main theorem.
- For any M-by-N sign matrix A and constant $\alpha > 1$

$$\frac{\gamma_2^\alpha(A)^2}{\alpha^2} \leq \mathrm{rk}_\alpha(A) = O\left(\gamma_2^\alpha(A)^2 \log(MN)\right)^3$$

Remarks

- When $\alpha = 1$ theorem does not hold. For equality function (sign matrix) $\mathrm{rk}(2I_N - 1_N) \geq N - 1$, but

$$\gamma_2(2I_N - 1_N) \leq 2\gamma_2(I_N) + \gamma_2(1_N) = 3,$$

  by Schur's theorem.

- Equality example also shows that the $\log N$ factor is necessary, as approximation rank of identity matrix is $\Omega(\log N)$ [Alon 08].

Advantages of γ_2^α

- γ_2^α can be formulated as a max expression

$$\gamma_2^\alpha(A) = \max_B \frac{(1+\alpha)\langle A, B\rangle + (1-\alpha)\ell_1(B)}{2\gamma_2^*(B)}$$

- γ_2^α is polynomial time computable by semidefinite programming

- γ_2^α is also known to lower bound quantum communication with shared entanglement, which was not known for approximation rank.

Proof sketch

- For the proof, we will use the primal formulation of γ_2:

$$\gamma_2(A) = \min_{\substack{X,Y: \\ X^T Y = A}} c(X)c(Y)$$

where $c(X)$ is the maximum ℓ_2 norm of a column of X.

Proof sketch

- For the proof, we will use the primal formulation of γ_2:

$$\gamma_2(A) = \min_{\substack{X,Y: \\ X^T Y = A}} c(X)c(Y)$$

where $c(X)$ is the maximum ℓ_2 norm of a column of X.

- Rank can also be phrased as optimizing over factorizations: the minimum K such that $A = X^T Y$ where X, Y are K-by-N matrices.

First step: dimension reduction

- Look at $X^TY = A'$ factorization realizing $\gamma_2^{1+\epsilon}(A)$. Say X, Y are K-by-N matrices.

First step: dimension reduction

- Look at $X^T Y = A'$ factorization realizing $\gamma_2^{1+\epsilon}(A)$. Say X, Y are K-by-N matrices.
- Know that the columns of X, Y have squared ℓ_2 norm at most $\gamma_2(A')$, but X, Y might still have many rows...

First step: dimension reduction

- Look at $X^T Y = A'$ factorization realizing $\gamma_2^{1+\epsilon}(A)$. Say X, Y are K-by-N matrices.

- Know that the columns of X, Y have squared ℓ_2 norm at most $\gamma_2(A')$, but X, Y might still have many rows...

- Johnson-Lindenstrauss lemma: let R be a random K'-by-K matrix

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

First step: dimension reduction

- Consider RX and RY where R is random matrix of size K'-by-K for $K' = O(\gamma_2^{1+\epsilon}(A)^2 \log N)$. By Johnson-Lindenstrauss lemma whp all the inner products $(RX)_i^T(RY)_j \approx X_i^T Y_j$ will be approximately preserved, up to additive factor of ϵ.

First step: dimension reduction

- Consider RX and RY where R is random matrix of size K'-by-K for $K' = O(\gamma_2^{1+\epsilon}(A)^2 \log N)$. By Johnson-Lindenstrauss lemma whp all the inner products $(RX)_i^T (RY)_j \approx X_i^T Y_j$ will be approximately preserved, up to additive factor of ϵ.

- This shows there is a matrix $A'' = (RX)^T (RY)$ which is a $1 + 2\epsilon$ approximation to A and has rank $O(\gamma_2^{1+\epsilon}(A)^2 \log N)$.

Second step: Error reduction

- Now we have a matrix $A'' = (RX)^T(RY)$ which is of the desired rank, but is only a $1+2\epsilon$ approximation to A, whereas we wanted an $1+\epsilon$ approximation of A.

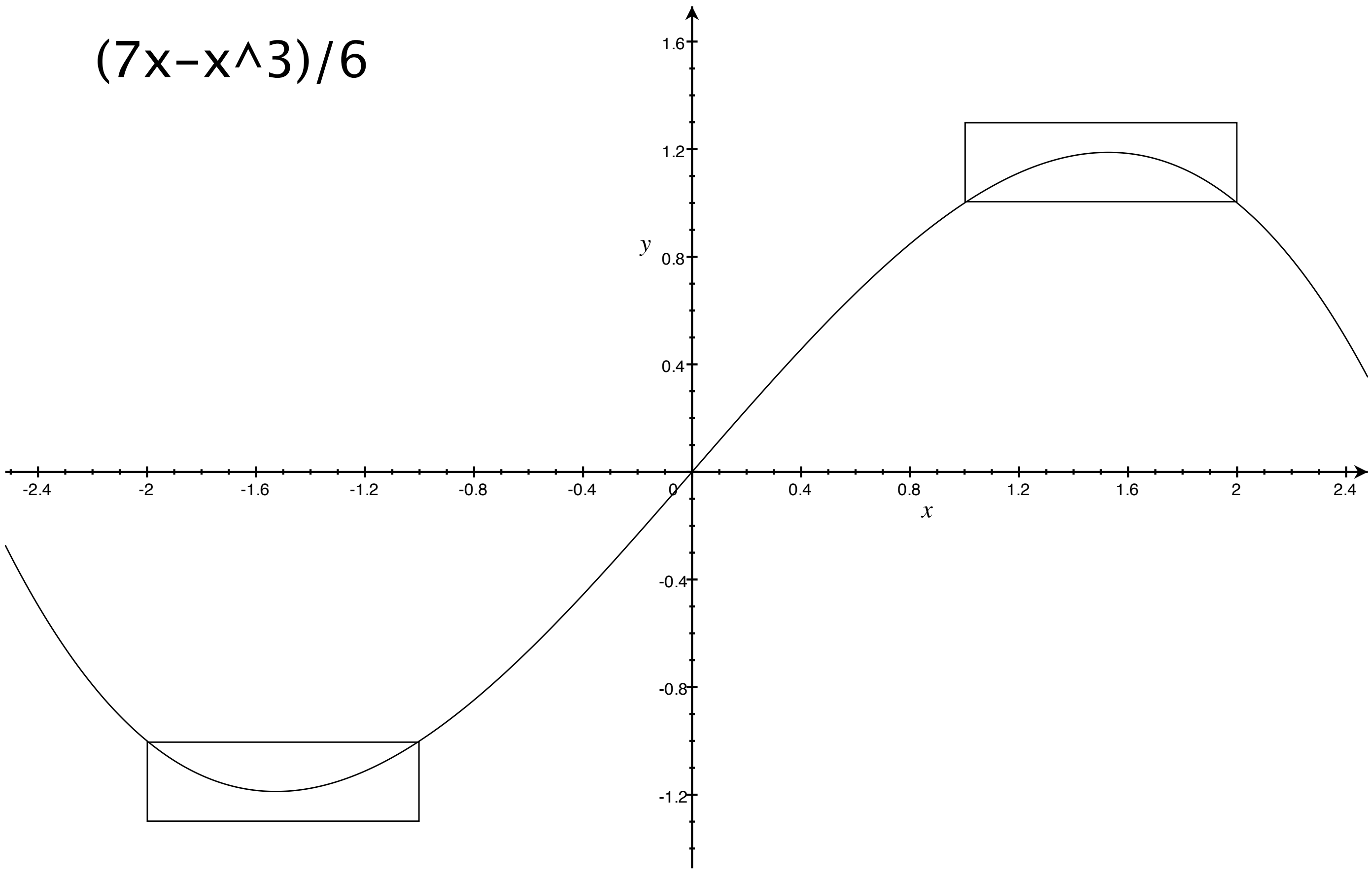
- Idea [Alon 08, Klivans Sherstov 07]: apply a polynomial to the entries of the matrix. Can show $\mathrm{rk}(p(A)) \leq (d+1)\mathrm{rk}(A)^d$ for degree d polynomial.

- Taking p to be low degree approximation of sign function makes $p(A'')$ better approximation of A. For our purposes, can get by with degree 3 polynomial.

- Completes the proof $\mathrm{rk}_\alpha(A) = O\left(\gamma_2^\alpha(A)^2 \log(N)\right)^3$

Polynomial for Error Reduction

(7x-x^3)/6



Open questions

- We have shown a polynomial time algorithm to approximate $\mathrm{rk}_\alpha(A)$, but ratio deteriorates as $\alpha \to \infty$.

$$\frac{\gamma_2^\alpha(A)^2}{\alpha^2} \leq \mathrm{rk}_\alpha(A) \leq O\left(\gamma_2^\alpha(A)^2 \log(N)\right)^3$$

- For the case of sign rank, lower bound fails! In fact, exponential gaps are known [BVW07, Sherstov07]

- Polynomial time algorithm to approximate sign rank?

Open questions

- Upper bound in terms of γ_2^{α}?

Open questions

- Upper bound in terms of γ_2^α? Linial and Shraibman show $R_\epsilon(f) = O(\gamma_2^\infty(M_f)^2)$.

Open questions

- Upper bound in terms of γ_2^α? Linial and Shraibman show $R_\epsilon(f) = O(\gamma_2^\infty(M_f)^2)$.

- By showing a relation between γ_2^α and approximation rank, we have simplified the picture of lower bound techniques. What is relationship between $\log \gamma_2^\alpha$ and corruption bound?