Approximation norms and duality for communication complexity lower bounds

Troy Lee
Columbia University

Adi Shraibman
Weizmann Institute

From min to max

- The cost of a “best” algorithm is naturally phrased as a minimization problem
- Dealing with this universal quantifier is one of the main challenges for lower bounders
- Norm based framework for showing communication complexity lower bounds
- Duality allows one to obtain lower bound expressions formulated as maximization problems

Example: Yao's principle

- One of the best known examples of this idea is Yao's minimax principle:

$$R_\epsilon(f) = \max_\mu D_\mu(f)$$

- To show lower bounds on randomized communication complexity, suffices to exhibit a hard distribution for deterministic protocols.

- The first step in many randomized lower bounds.

A few matrix norms

Let A be a matrix. The singular values of A are $\sigma_i(A) = \sqrt{\lambda_i(AA^T)}$.

Define

$$\|A\|_p = \ell_p(\sigma) = \left(\sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)^p \right)^{1/p}$$

A few matrix norms

Let A be a matrix. The singular values of A are $\sigma_i(A) = \sqrt{\lambda_i(AA^T)}$.

Define

$$\|A\|_p = \ell_p(\sigma) = \left(\sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)^p \right)^{1/p}$$

- Trace norm: $\|A\|_1$

A few matrix norms

Let A be a matrix. The singular values of A are $\sigma_i(A) = \sqrt{\lambda_i(AA^T)}$.

Define

$$\|A\|_p = \ell_p(\sigma) = \left(\sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)^p \right)^{1/p}$$

- Trace norm: $\|A\|_1$
- Spectral norm: $\|A\|_\infty$

A few matrix norms

Let A be a matrix. The singular values of A are $\sigma_i(A) = \sqrt{\lambda_i(AA^T)}$.

Define

$$\|A\|_p = \ell_p(\sigma) = \left(\sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)^p \right)^{1/p}$$

- Trace norm: $\|A\|_1$
- Spectral norm: $\|A\|_\infty$
- Frobenius norm $\|A\|_2 = (\sum_{i,j} |A_{ij}|^2)^{1/2}$

Example: trace norm

As ℓ_1 and ℓ_∞ are dual, so too are trace norm and spectral norm:

$$\|A\|_1 = \max_B \frac{|\langle A, B\rangle|}{\|B\|_\infty}$$

Example: trace norm

As ℓ_1 and ℓ_∞ are dual, so too are trace norm and spectral norm:

$$\|A\|_1 = \max_B \frac{|\langle A, B\rangle|}{\|B\|_\infty}$$

- Thus to show that the trace norm of A is large, it suffices to find B with non-negligible inner product with A and small spectral norm.

Example: trace norm

As ℓ_1 and ℓ_∞ are dual, so too are trace norm and spectral norm:

$$\|A\|_1 = \max_B \frac{|\langle A, B\rangle|}{\|B\|_\infty}$$

- Thus to show that the trace norm of A is large, it suffices to find B with non-negligible inner product with A and small spectral norm.

- We will refer to B as a *witness*.

Application to communication complexity

- For a function $f : X \times Y \to \{-1, +1\}$ we define the communication matrix $A_f[x, y] = f(x, y)$.

- For deterministic communication complexity, one of the best lower bounds available is log rank:
$$D(f) \geq \log \mathrm{rk}(A_f)$$

- The famous log rank conjecture states this lower bound is polynomially tight.

Application to communication complexity

- As the rank of a matrix is equal to the number of non-zero singular values, we have

$$\|A\|_1 = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)$$

Application to communication complexity

- As the rank of a matrix is equal to the number of non-zero singular values, we have

$$\|A\|_1 = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A) \leq \sqrt{\mathrm{rk}(A)} \left(\sum_i \sigma_i^2(A) \right)^{1/2}$$

Application to communication complexity

- As the rank of a matrix is equal to the number of non-zero singular values, we have

$$\|A\|_1 = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A) \leq \sqrt{\mathrm{rk}(A)} \left(\sum_i \sigma_i^2(A) \right)^{1/2} = \sqrt{\mathrm{rk}(A)} \|A\|_2$$

Application to communication complexity

- As the rank of a matrix is equal to the number of non-zero singular values, we have

$$\|A\|_1 = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A) \leq \sqrt{\mathrm{rk}(A)} \left(\sum_i \sigma_i^2(A) \right)^{1/2} = \sqrt{\mathrm{rk}(A)} \|A\|_2$$

- For a M-by-N sign matrix $\|A\|_2 = \sqrt{MN}$ so we have

$$2^{D(f)} \geq \mathrm{rk}(A_f) \geq \frac{(\|A_f\|_1)^2}{MN}$$

Call this the "trace norm method."

Trace norm method (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 method (drawback)

- As a complexity measure, the trace norm method 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 (drawback)

- As a complexity measure, the trace norm method 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

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

worse bound on whole than on H_N submatrix!

Trace norm method (a fix)

- We can fix this by considering

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

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 $\mathrm{rk}(A \circ uv^T) \leq \mathrm{rk}(A)$ we still have

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

The γ_2 norm

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

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

- By our previous discussion, for a sign matrix A

$$\mathrm{rk}(A) \geq \max_{\substack{u,v: \\ \|u\|_2 = \|v\|_2 = 1}} \left(\frac{\|A \circ uv^T\|_1}{\|A \circ uv^T\|_2} \right)^2 = \gamma_2(A)^2$$

γ_2 norm: Surprising usefulness

- In matrix analysis known as “Schur product norm,” among others

γ_2 norm: Surprising usefulness

- In matrix analysis known as "Schur product norm," among others
- Schur (1911) showed that $\gamma_2(A) = \max_i A_{ii}$ if A psd.

γ_2 norm: Surprising usefulness

- In matrix analysis known as "Schur product norm," among others
- Schur (1911) showed that $\gamma_2(A) = \max_i A_{ii}$ if A psd.
- The dual norm $\gamma_2^*(A) = \max_B \langle A, B\rangle / \gamma_2(B)$ satisfies

$$\mathrm{sdpval}(G) = \frac{1}{2} + \frac{\gamma_2^*(G)}{2}$$

for XOR game G. Also, $\mathrm{sdpval}(G)$ equals value of game where Alice and Bob share entanglement.

γ_2 norm: Surprising usefulness

- In matrix analysis known as "Schur product norm," among others
- Schur (1911) showed that $\gamma_2(A) = \max_i A_{ii}$ if A psd.
- The dual norm $\gamma_2^*(A) = \max_B \langle A, B \rangle / \gamma_2(B)$ satisfies

$$\mathrm{sdpval}(G) = \frac{1}{2} + \frac{\gamma_2^*(G)}{2}$$

  for XOR game G. Also, $\mathrm{sdpval}(G)$ equals value of game where Alice and Bob share entanglement.

- $\mathrm{disc}_P(A) = \Theta(\gamma_2^*(A \circ P))$ [Linial, Shraibman 08]

Randomized and quantum communication complexity

- So far it is not clear what we have gained. Many techniques available to bound matrix rank.

- But for randomized and quantum communication complexity the relevant measure is no longer rank, but *approximation rank*. For a sign matrix A:

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

- NP-hard? Can be difficult even for basic matrices. Disjointness was longstanding open problem resolved by [Razborov 03] who showed optimal bound $2^{\Omega(\sqrt{n})}$ using approximate version of trace norm method.

Approximation rank

- By [Buhrman, de Wolf 01]

$$R_\epsilon(A_f) \geq Q_\epsilon(A_f) \geq (1/2) \log \mathrm{rk}_{\alpha_\epsilon}(A_f)$$

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

- Perhaps more plausible than the log rank conjecture: there exists c

$$Q_\epsilon(f) \leq (\log \mathrm{rk}_{\alpha_\epsilon}(A_f))^c$$

Approximation norms

- We have seen how trace norm and γ_2 lower bound rank.

- In a similar fashion to approximation rank, we can define approximation norms. For an arbitrary norm $|||\cdot|||$ let

$$|||A|||^{\alpha} = \min_{B} \{|||B||| : 1 \leq A[x,y] \cdot B[x,y] \leq \alpha\}$$

- Note that an approximation norm is not itself necessarily a norm

- However, we we can still use duality to obtain a max expression

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

Approximate γ_2

- From our discussion, for a sign matrix A

$$\mathrm{rk}_\alpha(A) \geq \frac{\gamma_2^\alpha(A)^2}{\alpha^2} \geq \frac{(\|A\|_1^\alpha)^2}{\alpha^2 MN}$$

Approximate γ_2

- From our discussion, for a sign matrix A

$$\mathrm{rk}_\alpha(A) \geq \frac{\gamma_2^\alpha(A)^2}{\alpha^2} \geq \frac{(\|A\|_1^\alpha)^2}{\alpha^2 MN}$$

- We show that for any sign matrix A and constant $\alpha > 1$

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

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.

- This 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

- Look at the min 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.

- Similarly rank can be phrased as

$$\mathrm{rk}(A) = \min_{\substack{X,Y: \\ X^T Y = A}} \min\{d(X), d(Y)\}$$

where $d(X)$ is the number of rows of X.

First step: dimension reduction

- Look at $X^T Y = A'$ factorization realizing $\gamma_2^\alpha(A)$. Say X, Y are K-by-N.

First step: dimension reduction

- Look at $X^T Y = A'$ factorization realizing $\gamma_2^\alpha(A)$. Say X, Y are K-by-N.

- 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 $\gamma_2^\alpha(A)$. Say X, Y are K-by-N.

- 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...

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

First step: dimension reduction

- Look at $X^T Y = A'$ factorization realizing $\gamma_2^\alpha(A)$. Say X, Y are K-by-N.

- 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...

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

- This shows there is a matrix $A'' = (RX)^T (RY)$ which is, say, a 2α approximation to A and has rank $O(\gamma_2^\alpha(A)^2 \log N)$.

Second step: Error reduction

- Now we have a matrix A'' which is of the desired rank, but is only a 2α approximation to A, whereas we wanted an α 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 $\mathrm{rk}_\alpha(A) = O\left(\gamma_2^\alpha(A)^2 \log(MN)\right)^3$

Norms for multiparty complexity

- In multiparty complexity, have a function $f : X_1 \times \ldots \times X_k \rightarrow \{-1, +1\}$. Instead of communication matrix, have communication tensor $A_f[x_1, \ldots, x_k] = f(x_1, \ldots, x_k)$.

- One difficulty about proving lower bounds is that linear algebraic concepts like rank, trace norm, spectral norm, either become very difficult to use or have no analog with tensors.

- Only method known for general model of number-on-the-forehead is discrepancy method. While can show bounds of $n/2^{2k}$ for generalized inner product [BNS89] for other functions like disjointness can only show $O(\log n)$ bounds.

Norms for multiparty complexity

- Basic fact: A successful c-bit NOF protocol partitions the communication tensor into at most 2^c many monochromatic cylinder intersections.

- This allows us to define our norm

$$\mu(A) = \min\{\sum |\gamma_i| : A = \sum \gamma_i C_i\}$$

  C_i is a cylinder intersection.

- We have $D(A) \geq \log \mu(A)$. For matrices $\mu(A) = \Theta(\gamma_2(A))$

- Also by usual arguments get $R_\epsilon(A) \geq \mu^\alpha(A)$.

The dual norm

- The dual norm is

$$\mu^*(A) = \max_{B:\mu(B)\leq 1} |\langle A, B\rangle|$$

The dual norm

- The dual norm is

$$\mu^*(A) = \max_{B:\mu(B)\leq 1} |\langle A, B\rangle|$$

- So we see $\mu^*(A) = \max_C |\langle A, C\rangle|$ where C is a cylinder intersection.

The dual norm

- The dual norm is
$$\mu^*(A) = \max_{B:\mu(B)\leq 1} |\langle A, B\rangle|$$
- So we see $\mu^*(A) = \max_C \ |\langle A, C\rangle|$ where C is a cylinder intersection.
- $\mathrm{disc}_P(A) = \mu^*(A \circ P)$

The dual norm

- The dual norm is
$$\mu^*(A) = \max_{B:\mu(B)\leq 1} |\langle A, B\rangle|$$

- So we see $\mu^*(A) = \max_C |\langle A, C\rangle|$ where C is a cylinder intersection.

- $\mathrm{disc}_P(A) = \mu^*(A \circ P)$

- Bound $\mu^\alpha(A)$ in the following form. Standard discrepancy method is exactly μ^∞
$$\mu^\alpha(A) = \max_B \frac{(1+\alpha)\langle A, B\rangle + (1-\alpha)\ell_1(B)}{2\mu^*(B)}$$

Choosing a witness

- Use framework of pattern matrices [Sherstov 07, 08] and generalization to pattern tensors in multiparty case [Chattopadhyay 07]: Choose witness derived from dual polynomial witnessing that f has high approximate degree.

- Degree/Discrepancy Theorem [Sherstov 07,08 Chattopadhyay 08]: Pattern tensor derived from function with pure high degree will have small discrepancy. In multiparty case, this uses [BNS 89] technique of bounding discrepancy.

Final result

- Final result: Randomized k-party complexity of disjointness

$$\Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$$

- Independently shown by Chattopadhyay and Ada
- Beame and Huynh-Ngoc have recently shown non-trivial lower bounds on disjointness for up to $\log^{1/3} n$ players (though not as strong as ours for small k).