Direct product theorem for discrepancy

Troy Lee
Rutgers University

Joint work with: Robert Špalek
Direct product theorems

• Knowing how to compute $f$, how can you compute $f \oplus f \oplus \cdots \oplus f$?

• Obvious upper bounds:
  – If can compute $f$ with $t$ resources, can compute $\bigoplus_{i=1}^{k} f$ with $kt$ resources.
  – If can compute $f$ with success probability $1/2 + \epsilon/2$, then succeed on $\bigoplus_{i=1}^{k} f$ with probability $1/2 + \epsilon^k/2$.

• Question: is this the best one can do?
  – Direct sum theorem: Need $\Omega(kt)$ resources to achieve original advantage
  – Direct product theorem: advantage decreases exponentially
Applications

• Hardness amplification
  – Yao’s XOR lemma: if circuits of size $s$ err on $f$ with non-negligible probability, then any circuit of some smaller size $s' < s$ will have small advantage over random guessing on $\bigoplus_{i=1}^{k} f$.

• Soundness amplification
  – Parallel repetition: if Alice and Bob win game $G$ with probability $\epsilon < 1$ then win $k$ independent games with probability $\epsilon^{k'} < \epsilon$.

• Strong DPT for quantum query complexity of OR function: [A05, KSW07] Oracle where $\text{NP} \not\subseteq \text{BQP/qpoly}$, time-space tradeoffs for sorting.
Background

- Shaltiel [S03] started a systematic study of when direct product theorems might hold.

- Showed a general counter-example where strong direct product theorem does not hold.

- Looked at bounds proven by particular method: discrepancy method in communication complexity.

\[
\text{disc}_U(f^{\oplus k}) = O(\text{disc}_U(f))^{k/3}
\]
Discrepancy

• For a Boolean function $f : X \times Y \to \{0, 1\}$, let $M_f$ be sign matrix of $f$ 

$$M_f[x, y] = (-1)^{f(x,y)}.$$ 

Let $P$ be a probability distribution on entries.

$$\text{disc}_P(f) = \max_{x \in \{0,1\}^{|X|}} \max_{y \in \{0,1\}^{|Y|}} |x^T(M_f \circ P)y| = \|M_f \circ P\|_C$$

$$\text{disc}(f) = \min_P \|M_f \circ P\|_C.$$ 

• Discrepancy is one of most general techniques available:

$$D(f) \geq R_\epsilon(f) \geq Q_\epsilon^*(f) = \Omega \left( \log \frac{1}{\text{disc}(f)} \right)$$
Basic Orientation

- Identify a function $f(x, y)$ with its sign matrix

- $(f \oplus g)(x_1, x_2, y_1, y_2) = f(x_1, y_1) \oplus g(x_2, y_2)$

- Very nice in terms of sign matrices: sign matrix for $f \oplus g$ is $M_f \otimes M_g$

- **Shaltiel**: Does general discrepancy obey product theorem?
Results

- Yes!

\[ \text{disc}_P(A) \text{disc}_Q(B) \leq \text{disc}_{P \otimes Q}(A \otimes B) \leq 8 \text{disc}_P(A) \text{disc}_Q(B) \]

\[ \frac{1}{64} \text{disc}(A) \text{disc}(B) \leq \text{disc}(A \otimes B) \leq 8 \text{disc}(A) \text{disc}(B) \]

- Taken together this means that for tensor product matrices, a tensor product distribution is near optimal:

\[ \frac{1}{512} \text{disc}_{P \otimes Q}(A \otimes B) \leq \text{disc}(A \otimes B) \leq 8 \text{disc}_{P \otimes Q}(A \otimes B) \]
Optimality

• Discrepancy does not perfectly product

• Consider the 2-by-2 Hadamard matrix $H$ (inner product of one bit)

$$H = \begin{bmatrix}
1 & 1 \\
1 & -1
\end{bmatrix}$$

• Uniform distribution, $x = y = [1 \ 1]$, shows $\text{disc}(H) = 1/2$

• On the other hand, $\text{disc}(H^{\otimes k}) = \Theta(2^{-k/2})$. 
The proof: short answer

• [Linial and Shraibman 06] define a semidefinite programming quantity $\gamma_2$ which they show characterizes discrepancy up to a constant factor, using ideas from [Alon and Naor 06].

• Although not always the case, semidefinite programs tend to behave nicely under product: [L79, FL92, . . . , CSUU07].

• The semidefinite relaxation of discrepancy does as well.
Outline for rest of talk

• Try to convince you that $\gamma_2$ arises very naturally in communication complexity

• Sketch the proof of the product theorem, and try to convince you this is what you would do even if you didn't listen to first part

• Further extensions, open problems
Communication complexity

- For deterministic complexity, rank is all you need . . .
  - $\log \text{rk}(A) \leq D(A)$
  - Log rank conjecture: $\exists \ell : D(A) \leq (\log \text{rk}(A))^\ell$

- As $\text{rk}(A \otimes B) = \text{rk}(A)\text{rk}(B)$ log rank conjecture would give direct sum theorem for deterministic communication complexity, up to polynomial factors.
Bounded-error models

- Approximate rank: \( \tilde{\text{rk}}(A) = \min_B \{ \text{rk}(B) : \| A - B \|_\infty \leq \epsilon \} \).

- For randomized and quantum complexity

\[
R_\epsilon(A) \geq Q_\epsilon(A) \geq \frac{\log \tilde{\text{rk}}(A)}{2}
\]

- But these approximate ranks are very hard to work with . . . Borrow ideas from approximation algorithms.
Relaxation of rank

- Instead of working with rank, work with convex relaxation of rank

- For example, by Cauchy-Schwarz we have

\[
\frac{\|A\|_{tr}^2}{\|A\|_F^2} \leq \text{rk}(A)
\]

- Not a good complexity measure as can be too uniform.

\[
\max_{u,v: \|u\| = \|v\| = 1} \|A \circ uv^T\|_{tr}^2 \leq \text{rk}(A)
\]

for *sign* matrix A.
Also known as . . .

- Duality of spectral norm and trace norm . . .

\[ \|A\| = \max_{B: \|B\|_{tr} \leq 1} \langle A, B \rangle \]

- means

\[ \max_{u, v: \|u\| = \|v\| = 1} \|A \circ uv^T\|_{tr}^2 = \max_{B: \|B\|_{tr} \leq 1} \|A \circ B\|_{tr} \]

\[ = \max_{B: \|B\| \leq 1} \|A \circ B\| \]
Coming from learning theory, Linial and Shraibman define

\[ \gamma_2(A) = \min_{X,Y:XY = A} r(X)c(Y), \]

where \( r(X) \) is the largest \( \ell_2 \) norm of a row of \( X \), similarly \( c(Y) \) for column of \( Y \).

By duality of semidefinite programming

\[ \gamma_2(A) = \max_{u,v:\|u\| = \|v\| = 1} \| A \circ uv^* \|_{tr} \]
Different flavors of $\gamma_2$

- For deterministic complexity

$$\gamma_2(A) = \min_{X,Y:XY=A} r(X)c(Y) = \max_{Q:||Q||_{tr}\leq 1} ||A \circ Q||_{tr}$$

- For randomized, quantum complexity with entanglement

$$\gamma_2^\epsilon(A) = \min_{X,Y:1 \leq XY \circ A \leq 1+\epsilon} r(X)c(Y)$$

- For unbounded error

$$\gamma_2^\infty = \min_{X,Y:1 \leq XY \circ A} r(X)c(Y) = \max_{Q:||Q||_{tr}\leq 1,Q \circ A \geq 0} ||A \circ Q||_{tr}$$
**Product theorem:** \( \text{disc}_{P \otimes Q}(A \otimes B) \leq 8 \text{disc}_P(A)\text{disc}_Q(B) \)

- Let’s look at \( \text{disc}_P \) again:

\[
\text{disc}_P(A) = \|A \circ P\|_C
\]

- This is an example of a quadratic program, in general NP-hard to evaluate.

- In approximation algorithms, great success in looking at semidefinite relaxations of NP-hard problems.

- Semidefinite programs also tend to behave nicely under product!
Proof: first step

- Semidefinite relaxation of cut-norm studied by [Alon and Naor 06].

- First step: go from 0/1 vectors to ±1 vectors. Look at the norm

\[ \|A\|_{\infty \rightarrow 1} = \max_{x,y \in \{-1,1\}^n} x^T A y \]

- Simple lemma shows these are related.

\[ \|A\|_C \leq \|A\|_{\infty \rightarrow 1} \leq 4\|A\|_C \]
Proof: second step

- Now go to semidefinite relaxation:

\[
\|A\|_{\infty \to 1} \leq \max_{u_i, v_j} \sum_{i,j} A_{i,j} \langle u_i, v_j \rangle, \quad \|u_i\| = \|v_j\| = 1
\]

- Grothendieck’s Inequality says

\[
\max_{u_i, v_j} \sum_{i,j} A_{i,j} \langle u_i, v_j \rangle \leq K_G \|A\|_{\infty \to 1}
\]

where \(1.67 \leq K_G \leq 1.782 \ldots\)
Proof: last step

- Our approximating quantity is exactly the norm dual to $\gamma_2$:

$$
\gamma_2^*(A) = \max_{B: \gamma_2(B) \leq 1} \langle A, B \rangle
$$

$$
= \max_{u_i, v_j: \|u_i\|, \|v_j\| \leq 1} \sum_{i, j} A_{i, j} \langle u_i, v_j \rangle
$$

- Thus we have

$$
disc_P(A) \leq \gamma_2^*(A \circ P) \leq 4K_G \; disc_P(A)
$$
Connection to XOR games

- Let $P[x, y]$ be the probability the verifier asks questions $x, y$, and $A[x, y] = (-1)^{f(x,y)}$ be the desired response. Provers send $a, b \in \{-1, 1\}$ trying to achieve $ab = A[x, y]$. 

- Value of classical game is $1/2 + \frac{\|A \circ P\|_{\infty \rightarrow 1}}{2}$

- Value of entanglement game is $1/2 + \frac{\gamma^*_2(A \circ P)}{2}$ [Tsirelson80, CHTW04]

- A product theorem for $\gamma^*_2$ has been shown twice before in the literature [FL92, CSUU07]
**Product theorem:** \( \text{disc}(A \otimes B) \leq 8 \text{disc}(A)\text{disc}(B) \)

- \( \text{disc}(A) = \min_P \| A \circ P \|_C \)

- \( \frac{1}{4K_G} \min_P \gamma_2^*(A \circ P) \leq \text{disc}(A) \leq \min_P \gamma_2^*(A \circ P) \)

- Now need to show product theorem for

\[
\min_{P: \| P \|_1 = 1, P \geq 0} \gamma_2^*(A \circ P) = \min_{P: \| P \|_1 = 1, P \geq 0} \frac{\gamma_2^*(A \circ P)}{\langle A, A \circ P \rangle} = \min_{Q: Q \circ A \geq 0} \frac{\gamma_2^*(Q)}{\langle A, Q \rangle}
\]
Direct product for \( \text{disc}(A) \): Last step

- Quantity from last slide:
  \[
  \min_{Q: Q \circ A \geq 0} \frac{\gamma_2^*(Q)}{\langle A, Q \rangle}
  \]

- Reciprocal looks like \( \gamma_2(A) \), except for non-negativity restriction.

- Reciprocal equals \( \gamma_2^\infty(A) \):
  \[
  \gamma_2^\infty(A) = \max_{Q: \|Q\|_{tr} \leq 1} \|A \circ Q\|_{tr} = \min_{X, Y} r(X)c(Y)
  \]
**Direct product for** \( \text{disc}(A) \): **Final step**

- [Linial and Shraibman 06] \( \gamma_2^\infty(A) \leq 1/\text{disc}(A) \leq 8 \gamma_2^\infty \)

- If \( Q_A, Q_B \) are optimal witnesses for \( A, B \) respectively, then

  \[
  \gamma_2^\infty(A \otimes B) \geq \|(A \otimes B) \circ (Q_A \otimes Q_B)\|_{tr} = \|(A \circ Q_A) \otimes (B \circ Q_B)\|_{tr}
  \]

  and \( Q_A \otimes Q_B \) agrees in sign everywhere with \( A \otimes B \)

- If \( A = X_A Y_A \) and \( B = X_B Y_B \) are optimal factorizations, then

  \[
  \gamma_2^\infty(A \otimes B) \leq r(X_A \otimes X_B)c(Y_A \otimes Y_B) = r(X_A)c(Y_A)r(X_B)c(Y_B)
  \]
Future directions

• Bounded-error version of $\gamma_2$

$$\gamma_2^\epsilon(A) = \min_{B: \|A - B\|_\infty \leq \epsilon} \max_{u, v} \|B \circ vu^T\|_{tr}$$

• Lower bounds quantum communication complexity with entanglement [LS07]. Strong enough to reprove Razborov’s optimal results for symmetric functions.

• Does $\gamma_2^\epsilon$ obey product theorem? Would generalize some results of [KSW06]
Composition theorem

• What about functions of the form \( f(g(x_1, y_1), g(x_2, y_2), \ldots, g(x_n, y_n)) \)?

• When \( f \neq \oplus \) lose the tensor product structure . . .

• Recent paper of [Shi and Zhu 07] show some results in this direction—use bound like \( \gamma_2^\xi \) on \( f \) but need \( g \) to be hard.
Open problems

• Optimal $\Omega(n)$ lower bound for disjointness can be shown by one-sided version of discrepancy. Does this obey product theorem?

• [Mittal and Szegedy 07] have begun a systematic theory of when a product theorem holds for a general semidefinite program. All of $\gamma_2, \gamma_2^*, \gamma_2^\infty$ fit in their framework.