Direct product theorem for discrepancy

Troy Lee
Rutgers University

Robert Špalek
Google

Direct product theorems: Why is Google interested?

Direct product theorems: Why should Google be interested?

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm,

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm, fight youtube copyright lawsuits,

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm, fight youtube copyright lawsuits, buy some promising new companies,

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm, fight youtube copyright lawsuits, buy some promising new companies, hire some Rutgers graduates . . .

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm, fight youtube copyright lawsuits, buy some promising new companies, hire some Rutgers graduates . . .

- What is the most effective way to distribute your limited resources to achieve these goals?

Direct product theorems: Why should Google be interested?

- Say you want to accomplish k independent tasks. . .
  improve search algorithm, fight youtube copyright lawsuits, buy some promising new companies, hire some Rutgers graduates . . .

- What is the most effective way to distribute your limited resources to achieve these goals?

- Is it possible to accomplish all of these tasks while spending less than the sum of the resources required for the individual tasks?

Direct product theorems

- Let f, g be Boolean functions. Say you want to compute $F(x_1, x_2) = f(x_1) \oplus g(x_2)$.

Direct product theorems

- Let f, g be Boolean functions. Say you want to compute $F(x_1, x_2) = f(x_1) \oplus g(x_2)$.

- Obviously can compute f and then compute g. Can you do better?

Direct product theorems

- Let f, g be Boolean functions. Say you want to compute $F(x_1, x_2) = f(x_1) \oplus g(x_2)$.
- Obviously can compute f and then compute g. Can you do better?
- Direct sum theorem: To compute F need sum of resources needed for f and g.

Direct product theorems

- Let f, g be Boolean functions. Say you want to compute $F(x_1, x_2) = f(x_1) \oplus g(x_2)$.

- Obviously can compute f and then compute g. Can you do better?

- Direct sum theorem: To compute F need sum of resources needed for f and g.

- With obvious algorithm, if can compute f, g with success probability $1/2 + \epsilon/2$, then succeed on F with probability $1/2 + \epsilon^2/2$.

Direct product theorems

- Let f, g be Boolean functions. Say you want to compute $F(x_1, x_2) = f(x_1) \oplus g(x_2)$.
- Obviously can compute f and then compute g. Can you do better?
- Direct sum theorem: To compute F need sum of resources needed for f and g.
- With obvious algorithm, if can compute f, g with success probability $1/2 + \epsilon/2$, then succeed on F with probability $1/2 + \epsilon^2/2$.
- 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 $\oplus_{i=1}^{k} f$.

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 $\oplus_{i=1}^{k} f$.
- Soundness amplification
  - Parallel repetition: if Alice and Bob win game G with probability $\epsilon < 1$ then win k independent games with probability $\bar{\epsilon}^{k'} < \epsilon$.

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 $\oplus_{i=1}^{k} f$.
- Soundness amplification
  - Parallel repetition: if Alice and Bob win game G with probability $\epsilon < 1$ then win k independent games with probability $\bar{\epsilon}^{k'} < \epsilon$.
- Time-space tradeoffs: Strong DPT for quantum query complexity of OR function [A05, KSW07] gives time-space tradeoffs for sorting with quantum computer.

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.
- In light of counter-example, we should look for direct product theorems under some assumptions

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.
- In light of counter-example, we should look for direct product theorems under some assumptions—say lower bound is shown by a particular method.

Discrepancy

- For a Boolean function $f : X \times Y \rightarrow \{0,1\}$, let M_f be sign matrix of f $M_f[x,y] = (-1)^{f(x,y)}$. Let P be a probability distribution on entries.

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

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.

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

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

Discrepancy

- For a Boolean function $f : X \times Y \rightarrow \{0,1\}$, let M_f be sign matrix of f $M_f[x,y] = (-1)^{f(x,y)}$. Let P be a probability distribution on entries.

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

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

- Discrepancy is one of most general techniques available:

$$D(f) \geq R_\epsilon(f) \geq Q^*_\epsilon(f) = \Omega\left(\log \frac{1}{\mathrm{disc}(f)}\right)$$

Distributional Complexity

- Let R be a deterministic c-bit protocol, and consider the correlation of R with M_f under distribution P. Say that R outputs R_i in the i^{th} rectangle:

$$\begin{aligned} \mathrm{cor}_P(R, M_f) &= \sum_{x,y} P[x,y]R[x,y]M_f[x,y] \\ &= \sum_{i=1}^{2^c} R_i \; \chi_i^T (M_f \circ P)\chi_i' \\ &\leq 2^c \mathrm{disc}_P(M_f) \end{aligned}$$

Results

- [Shaltiel 03] showed $\mathrm{disc}_{U^{\otimes k}}(M_f^{\otimes k}) = O(\mathrm{disc}_U(M_f))^{k/3}$

Results

- [Shaltiel 03] showed $\mathrm{disc}_{U^{\otimes k}}(M_f^{\otimes k}) = O(\mathrm{disc}_U(M_f))^{k/3}$
  Open question: does product theorem hold for general discrepancy?

Results

- [Shaltiel 03] showed $\mathrm{disc}_{U^{\otimes k}}(M_f^{\otimes k}) = O(\mathrm{disc}_U(M_f))^{k/3}$
  Open question: does product theorem hold for general discrepancy?

- For any probability distributions P, Q:

$$\mathrm{disc}_{P \otimes Q}(A \otimes B) \leq 8 \ \mathrm{disc}_P(A)\mathrm{disc}_Q(B)$$

Results

- [Shaltiel 03] showed $\mathrm{disc}_{U^{\otimes k}}(M_f^{\otimes k}) = O(\mathrm{disc}_U(M_f))^{k/3}$
  Open question: does product theorem hold for general discrepancy?

- For any probability distributions P, Q:

$$\mathrm{disc}_{P \otimes Q}(A \otimes B) \leq 8\ \mathrm{disc}_P(A)\mathrm{disc}_Q(B)$$

- Product theorem also holds for $\mathrm{disc}(A) = \min_P \mathrm{disc}_P(A)$:

$$\frac{1}{64}\mathrm{disc}(A)\mathrm{disc}(B) \leq \mathrm{disc}(A \otimes B) \leq 8\ \mathrm{disc}(A)\mathrm{disc}(B)$$

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 $\mathrm{disc}(H) = 1/2$

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 $\mathrm{disc}(H) = 1/2$
- On the other hand, $\mathrm{disc}(H^{\otimes k}) = \Theta(2^{-k/2})$.

Some consequences

- Strong direct product theorem for average-case complexity: If correlation of M_f with c-bit protocols is at most $2^{-\ell}$, shown by discrepancy method, then correlation of $M_f^{\otimes k}$ with kc-bit protocols is at most $2^{k(-\ell+3)}$

Some consequences

- Strong direct product theorem for average-case complexity: If correlation of M_f with c-bit protocols is at most $2^{-\ell}$, shown by discrepancy method, then correlation of $M_f^{\otimes k}$ with kc-bit protocols is at most $2^{k(-\ell+3)}$
- Direct sum theorem for randomized, quantum bounds shown by discrepancy method

Some consequences

- Strong direct product theorem for average-case complexity: If correlation of M_f with c-bit protocols is at most $2^{-\ell}$, shown by discrepancy method, then correlation of $M_f^{\otimes k}$ with kc-bit protocols is at most $2^{k(-\ell+3)}$

- Direct sum theorem for randomized, quantum bounds shown by discrepancy method

- Direct sum theorem for weakly unbounded-error protocols: randomized model where
  - $\Pr[R[x,y] = f(x,y)] \geq 1/2$ for all x, y
  - If always succeed with probability $\geq 1/2 + \epsilon$, cost is number of bits communicated $+ \log(1/\epsilon)$.

Product theorem: $\mathrm{disc}_{P\otimes Q}(A \otimes B) \leq 8\ \mathrm{disc}_P(A)\mathrm{disc}_Q(B)$

- Let's look at disc_P again:

$$\mathrm{disc}_P(A) = \|A \circ P\|_C$$

- 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

$$\|A\|_{\infty \to 1} = \max_{x,y \in \{-1,1\}^n} x^T A y$$

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

$$\|A\|_{\infty \to 1} = \max_{x,y \in \{-1,1\}^n} x^T A y$$

- Simple lemma shows these are related.

$$\|A\|_C \leq \|A\|_{\infty \to 1} \leq 4\|A\|_C$$

- In fact, several discrepancy results proceed by bounding $\|A\|_{\infty \to 1}$ [Raz00, FG05, She07].

Proof: second step

- Now go to semidefinite relaxation:

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

Proof: second step

- Now go to semidefinite relaxation:

$$\|A\|_{\infty\to 1} \le \max_{\substack{u_i, v_j \\ \|u_i\|=\|v_j\|=1}} \sum_{i,j} A_{i,j}\langle u_i, v_j\rangle$$

- Grothendieck's Inequality says

$$\max_{\substack{u_i, v_j \\ \|u_i\|=\|v_j\|=1}} \sum_{i,j} A_{i,j}\langle u_i, v_j\rangle \le K_G \|A\|_{\infty\to 1}$$

where $1.67 \le K_G \le 1.782\ldots$

Proof: last step

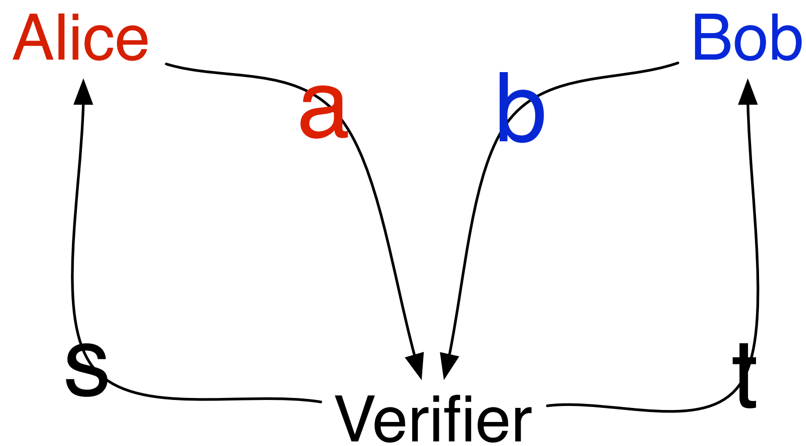
- Let
$$\sigma(A) = \max_{\substack{u_i, v_j \\ \|u_i\| = \|v_j\| = 1}} \sum_{i,j} A_{i,j} \langle u_i, v_j \rangle$$

- We now have $(1/4K_G)\ \sigma(A \circ P) \leq \mathrm{disc}_P(A) \leq \sigma(A \circ P)$

- All that remains is to show $\sigma(A_1 \otimes A_2) = \sigma(A_1)\sigma(A_2)$.

- In fact, this has already been shown in the literature [FL92, CSUU07, MS07]

Connection to XOR games



P[s,t] chooses (s,t), desires ab=V(s,t)

Connection to XOR games

- Let $P[s,t]$ be the probability the verifier asks questions s,t, and $V[s,t] \in \{-1,1\}$ be the desired response. Provers send $a,b \in \{-1,1\}$ trying to achieve $ab = V[s,t]$.

Connection to XOR games

- Let $P[s,t]$ be the probability the verifier asks questions s,t, and $V[s,t] \in \{-1,1\}$ be the desired response. Provers send $a,b \in \{-1,1\}$ trying to achieve $ab = V[s,t]$.

- Best correlation provers can achieve with V is $\|V \circ P\|_{\infty \to 1}$

Connection to XOR games

- Let $P[s,t]$ be the probability the verifier asks questions s,t, and $V[s,t] \in \{-1,1\}$ be the desired response. Provers send $a,b \in \{-1,1\}$ trying to achieve $ab = V[s,t]$.

- Best correlation provers can achieve with V is $\|V \circ P\|_{\infty \to 1}$

- By characterization of Tsirelson, best correlation of entangled provers is $\sigma(V \circ P)$ [Tsirelson80, CHTW04]

Connection to XOR games

- Let $P[s,t]$ be the probability the verifier asks questions s,t, and $V[s,t] \in \{-1,1\}$ be the desired response. Provers send $a,b \in \{-1,1\}$ trying to achieve $ab = V[s,t]$.

- Best correlation provers can achieve with V is $\|V \circ P\|_{\infty \to 1}$

- By characterization of Tsirelson, best correlation of entangled provers is $\sigma(V \circ P)$ [Tsirelson80, CHTW04]

- Product theorem for σ gives parallel repetition theorem for classical or entangled games.

Next theorem: Product for $\mathrm{disc}(A)$

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

Next theorem: Product for $\mathrm{disc}(A)$

- $\mathrm{disc}(A) = \min_P \|A \circ P\|_C$
- Linial and Shraibman 07 introduce a quantity γ_2^∞, and show

$$\frac{1}{8\gamma_2^\infty(A)} \leq \mathrm{disc}(A) \leq \frac{1}{\gamma_2^\infty(A)}$$

Next theorem: Product for $\mathrm{disc}(A)$

- $\mathrm{disc}(A) = \min_P \|A \circ P\|_C$
- Linial and Shraibman 07 introduce a quantity γ_2^∞, and show
$$\frac{1}{8\gamma_2^\infty(A)} \leq \mathrm{disc}(A) \leq \frac{1}{\gamma_2^\infty(A)}$$
- Taking this as a black box, just need to show $\gamma_2^\infty(A \otimes B) = \gamma_2^\infty(A)\gamma_2^\infty(B)$

Next theorem: Product for $\mathrm{disc}(A)$

- $\mathrm{disc}(A) = \min_P \|A \circ P\|_C$
- Linial and Shraibman 07 introduce a quantity γ_2^∞, and show

$$\frac{1}{8\gamma_2^\infty(A)} \le \mathrm{disc}(A) \le \frac{1}{\gamma_2^\infty(A)}$$

- Taking this as a black box, just need to show $\gamma_2^\infty(A \otimes B) = \gamma_2^\infty(A)\gamma_2^\infty(B)$
- In fact, $\frac{1}{\gamma_2^\infty(A)} = \min_P \sigma(A \circ P)$.

A communication complexity short story

- For deterministic complexity, rank is all you need . . .

A communication complexity short story

- For deterministic complexity, rank is all you need . . .
- $\log \mathrm{rk}(M_f) \leq D(f)$

A communication complexity short story

- For deterministic complexity, rank is all you need . . .
- $\log \mathrm{rk}(M_f) \leq D(f)$
- $\mathrm{rk}(M_f)$ polynomial time computable in length of truth table of f

A communication complexity short story

- For deterministic complexity, rank is all you need . . .
- $\log \mathrm{rk}(M_f) \leq D(f)$
- $\mathrm{rk}(M_f)$ polynomial time computable in length of truth table of f
- Log rank conjecture: $\exists \ell : D(f) \leq (\log \mathrm{rk}(M_f))^{\ell}$

Bounded-error models

- Approximate rank: $\widetilde{\mathrm{rk}}(A) = \min_B\{\mathrm{rk}(B) : \|A - B\|_\infty \leq \epsilon\}$.
- For randomized and quantum complexity

$$R_\epsilon(A) \geq Q_\epsilon(A) \geq \frac{\log \widetilde{\mathrm{rk}}(A)}{2}$$

- 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
- Let i^{th} singular value be $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$

Relaxation of rank

- Instead of working with rank, work with convex relaxation of rank
- Let i^{th} singular value be $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$
- Remember, $\|A\|_{tr} = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A)$, $\|A\|_F^2 = \sum_i \sigma_i(A)^2$

Relaxation of rank

- Instead of working with rank, work with convex relaxation of rank
- Let i^{th} singular value be $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$
- Remember, $\|A\|_{tr} = \sum_{i=1}^{\mathrm{rk}(A)} \sigma_i(A), \; \|A\|_F^2 = \sum_i \sigma_i(A)^2$
- By Cauchy-Schwarz inequality we have

$$\frac{\|A\|_{tr}^2}{\|A\|_F^2} \leq \mathrm{rk}(A)$$

Relaxation of rank

- Not a good complexity measure as too uniform.

- Since $\mathrm{rk}(A \circ uv^T) \leq \mathrm{rk}(A)$ can remedy this as follows

$$\max_{u,v:\|u\|=\|v\|=1} \frac{\|A \circ uv^T\|_{tr}^2}{\|A \circ uv^T\|_F^2} \leq \mathrm{rk}(A)$$

- Simplifies nicely for a *sign* matrix A

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

Also known as . . .

- This bound has many equivalent forms.

Also known as . . .

- This bound has many equivalent forms.
- As $\|A\| = \max_{u,v} \mathrm{Tr}(Avu^T) = \max_{B:\|B\|_{tr}\leq 1} \mathrm{Tr}(AB)$ one can show

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

Also known as . . .

- This bound has many equivalent forms.

- As $\|A\| = \max_{u,v} \mathrm{Tr}(Avu^T) = \max_{B:\|B\|_{tr}\leq 1} \mathrm{Tr}(AB)$ one can show

$$\begin{aligned}
\max_{u,v:\|u\|=\|v\|=1} \|A \circ uv^T\|_{tr}^2 &= \max_{B:\|B\|_{tr}\leq 1} \|A \circ B\|_{tr} \\
&= \max_{B:\|B\|\leq 1} \|A \circ B\|
\end{aligned}$$

aka . . . Linial and Shraibman's γ_2

- Coming from learning theory, Linial and Shraibman define

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

$r(X)$ is largest ℓ_2 norm of a row of X, similarly $c(Y)$ for column of Y

aka . . . Linial and Shraibman's γ_2

- Coming from learning theory, Linial and Shraibman define

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

$r(X)$ is largest ℓ_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 γ_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

$$\gamma_2^{\infty} = \min_{X,Y:1\leq XY\circ A} r(X)c(Y) = \max_{Q:\|Q\|_{tr}\leq 1, Q\circ A\geq 0} \|A \circ Q\|_{tr}$$

Direct product for $\mathrm{disc}(A)$: Final step

- Using max and min formulations of γ_2^∞ easy to show product theorem

Direct product for $\mathrm{disc}(A)$: Final step

- Using max and min formulations of γ_2^∞ easy to show product theorem
- 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 $A \otimes B$

Direct product for $\mathrm{disc}(A)$: Final step

- Using max and min formulations of γ_2^∞ easy to show product theorem
- 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 γ_2

$$\gamma_2^\epsilon(A) = \min_{\substack{B \\ 1 \leq A \circ B[i,j] \leq 1+\epsilon}} \max_{u,v} \|B \circ vu^T\|_{tr}$$

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

- Does γ_2^ϵ 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 γ_2^ϵ 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. γ_2, σ fit in their framework, but γ_2^∞ does not seem to. Can we extend this theory to handle such cases?