Direct product theorem for discrepancy

Troy Lee
Rutgers University

Adi Shraibman
Weizmann Institute of Science

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 theory graduate students . . .

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 theory graduate students . . .

- 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 theory graduate students . . .

- 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 sum theorems

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

Direct sum theorems

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

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

Direct sum theorems

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

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

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

Direct sum theorems

- Let f, g be Boolean functions. Say you want to compute $h(x_1, x_2) = (f(x_1), g(x_2))$.
- Obviously can compute f and then compute g. Can you do better?
- Direct sum theorem: To compute h need sum of resources needed for f and g.
- "The shortest way to do many things is to do only one thing at once" – Samuel Smiles

Direct product theorems

- Study behavior of success probability: with obvious algorithm, if can compute f with success probability p, then succeed on $f^2(x_1, x_2) = (f(x_1), f(x_2))$ with probability p^2.

Direct product theorems

- Study behavior of success probability: with obvious algorithm, if can compute f with success probability p, then succeed on $f^2(x_1, x_2) = (f(x_1), f(x_2))$ with probability p^2.

- Direct product theorem: success probability decreases exponentially.

Direct product theorems

- Study behavior of success probability: with obvious algorithm, if can compute f with success probability p, then succeed on $f^2(x_1, x_2) = (f(x_1), f(x_2))$ with probability p^2.

- Direct product theorem: success probability decreases exponentially. Strong direct product theorem—this holds for f^k even with k times the resources.

Direct product theorems

- Study behavior of success probability: with obvious algorithm, if can compute f with success probability p, then succeed on $f^2(x_1, x_2) = (f(x_1), f(x_2))$ with probability p^2.

- Direct product theorem: success probability decreases exponentially. Strong direct product theorem—this holds for f^k even with k times the resources.

- Note: For us, more convenient to investigate $h(x_1, x_2) = f(x_1) \oplus g(x_2)$. By results of [VW07] showing bias of this problem decreases exponentially suffices to give direct product theorem.

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 $p < 1$ then win k independent games with probability $\bar{p}^{k'} < p$.

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 $p < 1$ then win k independent games with probability $\bar{p}^{k'} < p$.
- Time-space tradeoffs: Strong DPT for quantum query complexity of OR function [Aar05, KSW07] gives time-space tradeoffs for sorting with quantum computer.

Background

- Shaltiel [S03] began a systematic study of when strong direct product theorems might hold—in particular, in the context of communication complexity.
- 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] began a systematic study of when strong direct product theorems might hold—in particular, in the context of communication complexity.

- 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.

Background

- Shaltiel [S03] began a systematic study of when strong direct product theorems might hold—in particular, in the context of communication complexity.
- 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.
- Studied discrepancy method in communication complexity

Communication complexity

- Alice is given input x, Bob input y and wish to compute some distributed function $f(x, y)$.

- In classical case, complexity is number of bits of conversation needed to output $f(x, y)$ on worst case input.

- Identify f with its communication matrix $M_f[x, y] = (-1)^{f(x,y)}$.

- For functions f, g, notice that the sign matrix of

$$h(x_1, x_2) = f(x_1) \oplus g(x_2)$$

is simply $M_f \otimes M_g$

Discrepancy

- Discrepancy is one of most general techniques available:

$$D(f) \geq R_{1/3}(f) \geq Q^*_{1/3}(f) = \Omega\left(\log \frac{1}{\mathrm{disc}(f)}\right)$$

- Let $M_f[x,y] = (-1)^{f(x,y)}$ be sign matrix of f. Let P be a probability distribution on entries.

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

Discrepancy

- Discrepancy is one of most general techniques available:

$$D(f) \geq R_{1/3}(f) \geq Q^*_{1/3}(f) = \Omega\left(\log \frac{1}{\mathrm{disc}(f)}\right)$$

- Let $M_f[x,y] = (-1)^{f(x,y)}$ be sign matrix of f. Let P be a probability distribution on entries.

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

- $\mathrm{disc}(f) = \min_P \mathrm{disc}_P(f)$.

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 randomized lower bounds shown by the discrepancy method

Some consequences

- Strong direct product theorem for randomized lower bounds shown by the discrepancy method

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

Proof ideas

- Let's look at disc_P again:

$$\mathrm{disc}_P(A) = \max_{x,y \in \{0,1\}^N} |x^T (M_f \circ P) y|$$

- 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!

Enter γ_2 norm

- Looking at the natural semidefinite relaxation of cut norm one arrives at the γ_2 norm, or rather its dual [AN06, LS08].

$$(1/4K_G)\ \gamma_2^*(A \circ P) \leq \mathrm{disc}_P(A) \leq \gamma_2^*(A \circ P)$$

  where $1.67 < K_G < 1.783$ is Grothendieck's constant.

- Furthermore, for $\mathrm{disc}(A) = \min_P \mathrm{disc}_P(A)$ we have [LS08]

$$\gamma_2^\infty(A) \leq \frac{1}{\mathrm{disc}(A)} \leq 4K_G\ \gamma_2^\infty(A)$$

  where $\gamma_2^\infty(A) = \min_{A': 1 \leq A \circ A'} \gamma_2(A)$.

Proof: second step

- Thus for our results suffices to show

$$\begin{aligned} \gamma_2^*(A \otimes B) &= \gamma_2^*(A)\gamma_2^*(B) \\ \gamma_2^\infty(A \otimes B) &= \gamma_2^\infty(A)\gamma_2^\infty(B) \end{aligned}$$

- This is done in usual fashion—look at semidefinite formulations of $\gamma_2^*, \gamma_2^\infty$, and use min and max formulations to show upper and lower inequalities, respectively.

Proof: second step

- Thus for our results suffices to show

$$\begin{aligned} \gamma_2^*(A \otimes B) &= \gamma_2^*(A)\gamma_2^*(B) \\ \gamma_2^\infty(A \otimes B) &= \gamma_2^\infty(A)\gamma_2^\infty(B) \end{aligned}$$

- This is done in usual fashion—look at semidefinite formulations of $\gamma_2^*, \gamma_2^\infty$, and use min and max formulations to show upper and lower inequalities, respectively.

- First item actually shown for perfect parallel repetition for two-prover XOR games with entanglement in Complexity last year [CSUU07]

Open problems

- We have shown product theorem for γ_2^∞. How about bounded-error version $\gamma_2^\alpha(A) = \min_{A' : 1 \le A \circ A' \le \alpha} \gamma_2(A')$?

Open problems

- We have shown product theorem for γ_2^∞. How about bounded-error version $\gamma_2^\alpha(A) = \min_{A': 1 \leq A \circ A' \leq \alpha} \gamma_2(A')$?

- Optimal $\Omega(n)$ lower bound for disjointness can be shown by corruption bound or one-sided version of discrepancy. Does this obey product theorem? Known under product distributions [BPSW05].

Open problems

- We have shown product theorem for γ_2^∞. How about bounded-error version $\gamma_2^\alpha(A) = \min_{A':1\le A\circ A'\le\alpha} \gamma_2(A')$?

- Optimal $\Omega(n)$ lower bound for disjointness can be shown by corruption bound or one-sided version of discrepancy. Does this obey product theorem? Known under product distributions [BPSW05].

- Build on the general theory developed by [MS07, LM08] for classifying when semidefinite programs perfectly product.

Open problems

- We have shown product theorem for γ_2^∞. How about bounded-error version $\gamma_2^\alpha(A) = \min_{A':1 \leq A \circ A' \leq \alpha} \gamma_2(A')$?

- Optimal $\Omega(n)$ lower bound for disjointness can be shown by corruption bound or one-sided version of discrepancy. Does this obey product theorem? Known under product distributions [BPSW05].

- Build on the general theory developed by [MS07, LM08] for classifying when semidefinite programs perfectly product.

- More general composition theorems for operations other than tensor product. Recent work of [SZ07] has some results in this direction.