A direct product theorem for discrepancy

Troy Lee
Department of Computer Science
Rutgers University [*]

Adi Shraibman
Department of Mathematics
Weizmann Institute of Science [†]

Robert Špalek
Google, Inc. [‡]

Abstract

Discrepancy is a versatile bound in communication complexity which can be used to show lower bounds in the distributional, randomized, quantum, and even unbounded error models of communication. We show an optimal product theorem for discrepancy, namely that for any two Boolean functions f, g, $\mathrm{disc}(f \oplus g) = \Theta(\mathrm{disc}(f)\mathrm{disc}(g))$. As a consequence we obtain a strong direct product theorem for distributional complexity, and direct sum theorems for worst-case complexity, for bounds shown by the discrepancy method. Our results resolve an open problem of Shaltiel (2003) who showed a weaker product theorem for discrepancy with respect to the uniform distribution, $\mathrm{disc}_{U^{\otimes k}}(f^{\otimes k}) = O(\mathrm{disc}_U(f))^{k/3}$. The main tool for our results is semidefinite programming, in particular a recent characterization of discrepancy in terms of a semidefinite programming quantity by Linial and Shraibman (2006).

1 Introduction

Say we know the complexity of a Boolean function f. How difficult is it to compute $F(x_1, x_2) = f(x_1) \oplus f(x_2)$, the parity of two independent instances of f? Theorems which address this situation are known as direct product and direct sum theorems. Perhaps the best known direct product theorem is Yao's XOR lemma, which states that if any circuit of size s errs with non-negligible probability when computing f, then any circuit of some smaller size $s' < s$ will have very small advantage over random guessing when computing $F(x_1, \ldots, x_k) = \bigoplus_i f(x_i)$. Notice here that

[*] Work supported in part by a National Science Foundation Mathematical Sciences Postdoctoral Fellowship, a Rubicon grant from the Netherlands Organization for Scientific Research, and by the European Commission under the Integrated Project Qubit Applications (AQP) funded by the IST directorate as Contract Number 015848. Part of this work conducted while at LRI, Université Paris-Sud, and while visiting the University of California, Berkeley. Email: troyjlee@gmail.com

[†] Email: adi.shraibman@weizmann.ac.il

[‡] Work conducted while at the University of California, Berkeley, supported by NSF Grant CCF-0524837 and ARO Grant DAAD 19-03-1-0082. Email: rspalek@gmail.com

while the error probability has increased, the amount of resources has actually decreased. This is known as a weak direct product theorem. On the other hand, a direct sum theorem aims to show that if it requires r resources to compute f with error ϵ, then computing $F(x_1, \ldots, x_k) = \oplus f(x_i)$ with error ϵ will require $\Omega(kr)$ resources. Here the error probability has not increased, but we allow the algorithm more resources.

The best of both lower bound worlds is a strong direct product theorem, which states that if computing f with success probability $1/2 + \epsilon/2$ requires r resources, then even with $\Omega(kr)$ resources any algorithm computing the parity of k independent copies of f will have success probability at most $1/2 + \epsilon^k/2$. While proving such a strong direct product result for Boolean circuits seems quite far off, a good testing grounds for our intuition about such theorems is communication complexity. Such a project was initiated in a systematic way by Shaltiel [Sha03], who showed a general counterexample where a strong direct product theorem does not hold. He further showed that bounds by the discrepancy method under the uniform distribution, a common way to show lower bounds on average-case communication complexity, do obey a product theorem. He left as an open question if discrepancy under arbitrary distributions also satisfies a direct product theorem.

We answer this question here and tighten Shaltiel's result to give a product theorem optimal up to a constant multiplicative factor. Namely, we show that $\mathrm{disc}(f \oplus g) = \Theta(\mathrm{disc}(f)\mathrm{disc}(g))$ for any Boolean functions f, g. Furthermore, we show that for functions of the form $f \oplus g$, the discrepancy bound is realized, up to a constant multiplicative factor, by a distribution of the form $P \otimes Q$, where P is a distribution over f and Q is a distribution over g, and $\otimes$ denotes tensor product.

As a consequence, we obtain a strong direct product theorem for distributional complexity bounds shown by the discrepancy method—If a c-bit protocol has correlation at most w with f, as shown by the discrepancy method, then a kc-bit protocol will have correlation at most $O(w^k)$ with the parity of k independent copies of f. Klauck [Kla01] has shown that the discrepancy bound characterizes the model of weakly-unbounded error complexity, a communication complexity version of the complexity class PP (formal definition given below in Section 2.2). As discrepancy characterizes this class, here we are able to obtain an unconditional direct sum theorem for this model of computation.

The main tool for our results is semidefinite programming, in particular a recent characterization of discrepancy in terms of a semidefinite quantity γ_2^∞ by Linial and Shraibman [LS07]. Linial and Shraibman also introduce a bounded-error version of the same semidefinite quantity, known as γ_2^α, which can be used to show lower bounds on bounded-error randomized and quantum communication complexity. It remains an interesting open question if a product theorem also holds for this quantity. As γ_2^α is able to prove an $\Omega(\sqrt{n})$ lower bound on the quantum communication complexity of disjointness, such a theorem would reprove a result of Klauck, Špalek, and de Wolf [KŠW07].

2 Preliminaries

In this section we will introduce some basic matrix notation, our main quantity of interest i.e. the discrepancy and its relation to communication complexity. We also introduce the γ_2 norm and its variants which we use to prove our main result.

2.1 Matrix preliminaries

We restrict ourselves to matrices over the real numbers. We use A^T to denote the transpose of the matrix A. For real matrices A, B we use $\leq$ to refer to entrywise comparison of matrices, that is $A \leq B$ iff $A[i,j] \leq B[i,j]$ for all (i,j). For a scalar c, we sometimes use the shorthand $A \geq c$ to indicate that all entries of A are at least as large as c. We denote tensor product by $\otimes$, Hadamard (entrywise) product by $\circ$ and inner product by $\langle \cdot, \cdot \rangle$. We let $\|A\|_1$ be the sum of the absolute values of the entries of A.

For a symmetric matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ denote the eigenvalues of A. Let $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$ be the i^{th} singular value of A. We make use of a few matrix norms. The Frobenius norm of A is the ℓ_2 norm of A thought of as a vector—that is

$$\|A\|_F = \sqrt{\sum_{i,j} A[i,j]^2}.$$

Notice also that $\|A\|_F^2 = \mathrm{Tr}(A^T A) = \sum_i \sigma_i^2(A)$. We also use the trace norm, $\|A\|_{tr} = \sum_i \sigma_i(A)$. Finally, we denote the spectral norm as $\|A\| = \sigma_1(A)$.

As the singular values of the matrix $A \otimes B$ are $\sigma_i(A)\sigma_j(B)$ where $\sigma_i(A), \sigma_j(B)$ range over the singular values of A and B respectively, all three of these matrix norms are multiplicative under tensor products.

Finally, we make use of the following simple fact

Fact 1 *For any matrices A, B, C, D, where A, C are of the same dimension and B, D are of the same dimension,*

$$(A \otimes B) \circ (C \otimes D) = (A \circ C) \otimes (B \circ D).$$

2.2 Communication complexity and discrepancy

Let X, Y be finite sets and $f : X \times Y \to \{0,1\}$ be a Boolean function. We associate with f a $|X|$-by-$|Y|$ sign matrix M_f known as the communication matrix. M_f is the $|X|$-by-$|Y|$ matrix where

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

We will identify the communication matrix with the function, and use them interchangeably.

Discrepancy is defined as follows:

Definition 2 (Discrepancy with respect to P) *Let P be a probability distribution on the entries of M_f. Discrepancy with respect to the distribution P is defined as:*

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

The maximum absolute value of a bilinear form over Boolean vectors is known as the cut norm, $\| \cdot \|_C$, thus it can be equivalently stated that $\mathrm{disc}_P(A) = \|A \circ P\|_C$. We will sometimes use this view in our proofs as our product results hold more generally for the cut norm, and not just discrepancy.

For showing lower bounds in communication complexity, one wishes to show that the discrepancy is small. We will let $\mathrm{disc}(A)$ without a subscript refer to $\mathrm{disc}_P(A)$ under the "hardest" distribution P.

Definition 3 (General discrepancy) *The discrepancy of a sign matrix M_f is defined as*

$$\mathrm{disc}(M_f) = \min_P \mathrm{disc}_P(M_f),$$

where the minimum is taken over all probability distributions P.

We will first see how discrepancy can be applied to communication complexity in the distributional model. The cost in this model is defined as follows:

Definition 4 (Distributional complexity) *Let $f : X \times Y \to \{0, 1\}$ be a Boolean function and P a probability distribution over the inputs $X \times Y$. For a fixed error $\epsilon \geq 0$, we define $D_P^\epsilon(f)$ to be the minimum communication of a deterministic protocol R where $\mathbb{E}_{(x,y) \leftarrow P}[R(x, y) \neq f(x, y)] \leq \epsilon$.*

The connection to discrepancy comes from the well known fact that a deterministic c-bit communication protocol partitions the communication matrix into 2^c many combinatorial rectangles. (See Kushilevitz and Nisan [KN97] for this and other background on communication complexity.) Let P be a probability distribution, R be a deterministic protocol, and let $R[x, y] \in \{-1, 1\}$ be the output of R on input (x, y). The correlation of R with f under the distribution P is

$$\mathrm{Corr}_P(M_f, R) = \mathbb{E}_{(x,y) \leftarrow P}[R[x, y] M_f[x, y]]$$

We then define the correlation with c-bit protocols as

$$\mathrm{Corr}_{c,P}(M_f) = \max_R \mathrm{Corr}_P(M_f, R)$$

where the max is taken over all deterministic c-bit protocols.

Fact 5

$$\mathrm{Corr}_{c,P}(M_f) \leq 2^c \mathrm{disc}_P(M_f)$$

Proof: Let R be a c-bit protocol which realizes the value $\mathrm{Corr}_{c,P}(M_f)$. A c-bit protocol partitions the communication matrix M_f into 2^c combinatorial rectangles, and on each such rectangle R reports the same answer for all the elements of the rectangle. We enumerate these rectangles by $i \in \{1, \ldots, 2^c\}$, and let R_i be the output of the protocol on elements of the i^{th} rectangle. Further, let $x_i \in \{0,1\}^{|X|}$ and $y_i \in \{0,1\}^{|Y|}$ be characteristic vectors of the respective rows and columns active in the i^{th} rectangle. Then we have

$$\begin{aligned}
\mathrm{Corr}_{c,P}(M_f) &= \langle R, M \circ P \rangle \\
&= \sum_{i=1}^{2^c} R_i \left(x_i^T (M_f \circ P) y_i \right) \\
&\leq \sum_{i=1}^{2^c} \left| x_i^T (M_f \circ P) y_i \right| \\
&\leq 2^c \mathrm{disc}_P(M_f).
\end{aligned}$$

□

We can turn this equation around to get a lower bound on $D_P^\epsilon(f)$. A protocol which has probability of error at most ϵ has correlation at least $1-2\epsilon$ with f, thus $D_P^\epsilon(f) \geq \log 1/((1-2\epsilon)\mathrm{disc}_P(M_f))$. This, in turn, shows how discrepancy can be used to lower bound randomized communication complexity. Let $R_\epsilon(f)$ be the minimum communication cost of a randomized protocol R such that $\Pr[R[x,y] \neq f(x,y)] \leq \epsilon$ for all x, y. Then, as by Yao's principle $R_\epsilon(f) = \max_P D_P^\epsilon(f)$, we find that $R_\epsilon(f) \geq \log 1/((1-2\epsilon)\mathrm{disc}(M_f))$.

Discrepancy is even more widely applicable to proving lower bounds on worst-case complexity. Kremer [Kre95] shows that discrepancy can be used to lower bound quantum communication with bounded-error, and Linial and Shraibman [LS07] extend this to show the discrepancy bound is valid even when the communicating parties share entanglement. Klauck [Kla01] shows that discrepancy characterizes, up to a small multiplicative factor, the communication cost of weakly unbounded-error protocols. We state this latter result for future use.

Definition 6 (Weakly unbounded-error) *Let R be a c-bit randomized protocol for f, and denote $\epsilon(R) = \min_{x,y} (\Pr[R(x,y) = f(x,y)] - 1/2)$. The weakly unbounded-error cost of R is $\mathrm{UPC}_R(f) = c + \log(1/\epsilon(R))$. The weakly unbounded-error cost of f, denoted $\mathrm{UPC}(f)$, is the minimal weakly unbounded-error cost of a randomized protocol for f.*

Theorem 7 (Klauck) *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then*

$$\begin{aligned}
\mathrm{UPC}(f) &\geq \log(1/\mathrm{disc}(f)) - O(1) \\
\mathrm{UPC}(f) &\leq 3\log(1/\mathrm{disc}(f)) + \log n + O(1).
\end{aligned}$$

The lower bound can be seen immediately from Fact 5, while the upper bound requires more work. Forster et al. [FKL$^+$01] show a similar result characterizing UPC complexity in terms of a notion from learning theory known as the maximal margin complexity. Linial and Shraibman later showed that discrepancy and maximal margin complexity are equivalent up to a constant factor.

2.3 Definitions of γ_2

The quantity γ_2 was introduced in [LMSS07] in a study of complexity measures of sign matrices. We give here a leisurely introduction to this quantity, its relatives, and their many equivalent forms.

2.3.1 Motivation

Matrix rank plays a fundamental role in communication complexity. Many different models of communication complexity have an associated rank bound which is usually the best technique available for showing lower bounds. For deterministic complexity, $D(f) \geq \log \mathrm{rk}(M_f)$, and the long-standing log rank conjecture asserts that this bound is tight up to polynomial factors. For randomized and quantum communication complexity, one becomes concerned not with the rank of the communication matrix, but of matrices close to the communication matrix in ℓ_∞ norm. Namely, if let the approximate rank be defined as $\widetilde{\mathrm{rk}}(M_f) = \min\{\mathrm{rk}(M) : \|M - M_f\|_\infty \leq \epsilon\}$, then one has $R_\epsilon(f) \geq Q_\epsilon(f) \geq (1/2) \log \widetilde{\mathrm{rk}}(M_f)$. As $\epsilon \rightarrow 1/2$ one obtains unbounded-error complexity, where one simply has to obtain the correct answer with probability strictly greater than $1/2$. This class is characterized up to one bit by the log of sign rank, the minimum rank of a matrix which agrees in sign everywhere with M_f.

In the case of approximate rank and sign rank, a difficulty arises as such rank minimization problems are in general NP-hard to compute. A (now) common approach to deal with NP-hard problems is to consider a semidefinite programming relaxation of the problem. The quantity $\gamma_2(M_f)$ can very naturally be viewed as a semidefinite relaxation of rank.

As the rank of a matrix is equal to the number of non-zero singular values, it follows from the Cauchy-Schwarz inequality that

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

A problem with this bound as a complexity measure is that it is not monotone—the bound can be larger on a submatrix of A than on A itself. As taking the Hadamard product of a matrix with a rank one matrix does not increase its rank, a way to fix this problem is to consider instead:

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

When A is a sign matrix, this bound simplifies nicely—for then, $\|A \circ vu^T\|_F = \|u\|\|v\| = 1$, and we are left with

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

This quantity turns out to be exactly $\gamma_2(A)$, as we shall now see.

2.3.2 The many faces of γ_2

The primary definition of γ_2 given in [LMSS07] is

Definition 8

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

where $r(X)$ is the largest ℓ_2 norm of a row of X and similarly $c(Y)$ is the largest ℓ_2 norm of a column of Y.

We now see that this quantity is the same as the one just discussed. Note that this equivalence holds for *any* matrix A, not just a sign matrix.

Theorem 9 *Let A be a m-by-n matrix. Then*

$$\gamma_2(A) = \max_{Q:\|Q\|\leq 1} \|A \circ Q\| = \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \|A \circ vu^T\|_{tr}$$

Proof: We obtain this by writing γ_2 as a semidefinite program and dualizing. Let $J_{m,n}$ be the m-by-n matrix all whose entries are equal to one. It will be convenient to work with a $(m+n)$-by-$(m+n)$ matrix A' which is a square and Hermitian "bipartite version" of A, and an auxiliary matrix F defined as follows:

$$A' = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}, \ F = \begin{bmatrix} 0 & J_{m,n} \\ J_{n,m} & 0 \end{bmatrix}$$

With these definitions in hand, one can see that γ_2 is equivalent to the following program:

$$\begin{aligned} \min\ & \eta \\ X[i,i] &\leq \eta \text{ for all } i \\ X &\succeq 0 \\ X \circ A' &= F \end{aligned}$$

Here $X \succeq 0$ means the X is positive semidefinite. Dualizing this program we obtain:

$$\max \langle Q, A' \rangle \tag{1}$$
$$\|\alpha\|_1 = 1 \tag{2}$$
$$\mathrm{diag}(\alpha) \succeq Q \tag{3}$$
$$Q \circ F = Q \tag{4}$$
$$\alpha \geq 0. \tag{5}$$

We can bring this program into a particularly nice form by letting $\beta[i] = 1/\sqrt{\alpha[i]}$, and $Q' = Q \circ \beta\beta^T$. Then the condition $\alpha \succeq Q$ can be rewritten as $I \succeq Q'$, or in other words $\|Q'\| \leq 1$. Letting $\gamma[i] = \sqrt{\alpha[i]}$, the objective function then becomes

$$\langle Q, A' \rangle = \langle Q' \circ \gamma\gamma^T, A' \rangle = \gamma^T (Q' \circ A')\gamma.$$

The condition $\mathrm{Tr}(\alpha) = 1$ means that γ is a unit vector. As γ is otherwise unconstrained, we obtain the first equivalence of the theorem:

$$\gamma_2(A) = \max_Q \frac{\|Q \circ A\|}{\|Q\|}$$

This shows that γ_2 is equivalent to a quantity known in the matrix analysis literature as the *Hadamard product operator norm* [Mat93]. The duality of the spectral norm and trace norm easily gives that this is equivalent to the Hadamard product trace norm (see [Mat93] for a proof):

$$\gamma_2(A) = \max_Q \frac{\|Q \circ A\|_{tr}}{\|Q\|_{tr}} = \max_{u,v:\|u\|=\|v\|=1} \|A \circ uv^T\|_{tr} \tag{6}$$

□

The fact that $(\gamma_2(A))^2 \leq \mathrm{rk}(A)$ implies its usefulness for communication complexity:

Theorem 10 (Linial-Shraibman [LS07]) *Let f be a Boolean function and $M_f[x,y] = (-1)^{f(x,y)}$. Then*

$$2 \log \gamma_2(M_f) \leq D(f).$$

2.3.3 Dual norm of γ_2

The norm dual to γ_2 will also play a key role in our study of discrepancy. By definition of a dual norm, we have

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

Since the dual norm is uniquely defined, we can read off the conditions for $\gamma_2^*(B) \leq 1$ from Equations (2)–(5) in the formulation of $\gamma_2(A)$. This tells us

$$\gamma_2^*(B) = \min_\alpha \left\{ \frac{1}{2}(1^T\alpha) : \mathrm{diag}(\alpha) - B' \succeq 0 \right\} \tag{7}$$

We can interpret the value of this program as follows:

Theorem 11

$$\gamma_2^*(B) = \min_{\substack{X,Y \\ X^TY=B}} \frac{1}{2}\left(\|X\|_F^2 + \|Y\|_F^2\right) = \min_{\substack{X,Y \\ X^TY=B}} \|X\|_F\|Y\|_F$$

where the min is taken over X, Y with orthogonal columns.

Proof: Let α be the optimal solution to (7). As $\mathrm{diag}(\alpha) - B' \succeq 0$, we have a factorization $\mathrm{diag}(\alpha) - B' = M^TM$. Write M as

$$M = \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Then we see that $X^TY = B$ and the columns of X, Y are orthogonal as B' is block anti-diagonal. The value of the program is simply $(1/2)(\|X\|_F^2 + \|Y\|_F^2)$.

In the other direction, for X, Y such that $X^TY = B$, we define the vector α as $\alpha[i] = \|X_i^T\|^2$ if $i \leq m$ and $\alpha[i] = \|Y_i\|^2$ otherwise. A similar argument to the above shows that $\mathrm{diag}(\alpha) - B' \succeq 0$, and the objective function is $\frac{1}{2}\left(\|X\|_F^2 + \|Y\|_F^2\right)$.

To see the equivalence between the additive and multiplicative forms of the bound, notice that if X, Y is a feasible solution, then so is $cX, (1/c)Y$ for a constant c. Thus we see that in the additive form of the bound, the optimum can be achieved with $\|X\|_F^2 = \|Y\|_F^2$, and similarly for the multiplicative form. The equivalence follows. □

2.3.4 Approximate versions of γ_2

To talk about randomized communication models, we need to go to an approximate version of γ_2. Linial and Shraibman [LS07] define

Definition 12 *Let A be a sign matrix, and $\alpha \geq 1$.*

$$\gamma_2^\alpha = \min_{X,Y:\alpha \geq (XY \circ A) \geq 1} r(X)c(Y).$$

An interesting limiting case is where XY simply has everywhere the same sign as A.

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

As we did with γ_2, we can represent γ_2^α and γ_2^∞ as semidefinite programs and dualize to obtain equivalent max formulations, which are more useful for proving lower bounds. We start with γ_2^∞ as it is simpler.

Theorem 13 *Let A be a sign matrix.*

$$\gamma_2^\infty(A) = \max_{Q:Q \circ A \geq 0} \frac{\|A \circ Q\|}{\|Q\|}.$$

Notice that this is the same as the definition of $\gamma_2(A)$ except for the restriction that $Q \circ A \geq 0$. We similarly obtain the following max formulation of γ_2^α.

Theorem 14 *Let A be a sign matrix and $\epsilon \geq 0$.*

$$\gamma_2^{1+\epsilon}(A) = \max_Q \frac{\|(1+\epsilon/2)Q \circ A - (\epsilon/2)|Q|\|}{\|Q\|} \tag{8}$$

where $|Q|$ denotes the matrix whose (x,y) entry is $|Q[x,y]|$.

Proof: The theorem is obtained by writing the definition of γ_2^α as a semidefinite programming and dualizing. The primal problem can be written as

$$\begin{aligned} &\min \eta \\ X[i,i] &\leq \eta \\ X &\succeq 0 \\ \alpha F \geq X &\circ A' \geq F \end{aligned}$$

Again in a straightforward way we can form the dual of this program:

$$\begin{aligned} \max \langle Q_1 - Q_2, F\rangle - (\alpha-1)\langle Q_2, F\rangle \\ \mathrm{Tr}(\beta) = 1 \\ \beta \succeq (Q_1 - Q_2)\circ A' \\ \beta, Q_1, Q_2 \geq 0, \end{aligned}$$

where β is a diagonal matrix. Notice that as $\alpha \to \infty$ in the optimal solution $Q_2 \to 0$ and so we recover the dual program for γ_2^∞.

We can argue that in the optimal solution to this program, Q_1, Q_2 will be disjoint. For if $Q_1[x,y] - Q_2[x,y] = a \geq 0$ then we set $Q_1'[x,y] = a$ and $Q_2'[x,y] = 0$ and increase the objective function. Similarly, if $Q_1[x,y] - Q_2[x,y] = a < 0$ we set $Q_1'[x,y] = 0$ and $Q_2'[x,y] = -a \leq Q_2[x,y]$ and increase the objective function.

Let $\epsilon = \alpha - 1$. In light of this observation, we can let $Q = Q_1 - Q_2$ be unconstrained and our objective function becomes $\langle (1+\epsilon/2)Q - \epsilon/2|Q|, F\rangle$, as the entrywise absolute value of Q in our case is $|Q| = Q_1 + Q_2$. As with γ_2 above, we can reformulate $\gamma_2^\alpha(A)$ in terms of spectral norms. □

Linial and Shraibman [LS07] show that γ_2^α can be used to lower bound quantum communication complexity with entanglement.

Theorem 15 (Linial and Shraibman) *Let A be a sign matrix, and $\epsilon \geq 0$. Then*

$$Q_\epsilon^*(A) \geq \log \gamma_2^{\alpha_\epsilon} - \log \alpha_\epsilon - 2,$$

where $\alpha_\epsilon = \frac{1}{1-2\epsilon}$

In his seminal result showing a $\Omega(\sqrt{n})$ lower bound on the quantum communication complexity of disjointness, Razborov [Raz03] essentially used a "uniform" version of γ_2^α. Namely, if A is a $|X|$-by-$|Y|$ matrix, we can in particular lower bound the spectral norm in the numerator of Equation (8) by considering uniform unit vectors x of length $|X|$ and y of length $|Y|$ where $x[i] = 1/\sqrt{|X|}$ and $y[i] = 1/\sqrt{|Y|}$. Then we have

$$\begin{aligned} \|(1+\epsilon/2)Q\circ A - (\epsilon/2)|Q|\| &\geq x^T((1+\epsilon/2)Q\circ A - (\epsilon/2)|Q|)y \\ &= \frac{\langle (1+\epsilon/2)Q, A\rangle - \epsilon/2\|Q\|_1}{\sqrt{|X||Y|}}, \end{aligned}$$

and so

$$\gamma_2^{1+\epsilon}(A) \geq \max_{Q:\|Q\|_1=1} \frac{\langle (1+\epsilon/2)Q, A\rangle - \epsilon/2}{\|Q\|\sqrt{|X||Y|}}$$

Sherstov [She07a] also uses this bound in simplifying Razborov's proof, giving an extremely elegant way to choose the matrix Q for a wide class of sign matrices A.

3 Relation of γ_2 to discrepancy

In looking at the definition of $\mathrm{disc}_P(A)$, we see that it is a quadratic program with quadratic constraints. Such problems are in general NP-hard to compute. A (now) common approach for dealing with NP-hard problems is to consider a semidefinite relaxation of the problem. In fact, Alon and Naor [AN06] do exactly this in developing a constant factor approximation algorithm for the cut norm. While we do not need the fact that semidefinite programs can be solved in polynomial time, we do want to take advantage of the fact that semidefinite programs often have the property of behaving nicely under product of instances. While not always the case, this property has been used many times in computer science, for example [Lov79, FL92, CSUU07].

As shown by Linial and Shraibman [LS06], it turns out that the natural semidefinite relaxations of $\mathrm{disc}_P(A)$ and $\mathrm{disc}(A)$ are given by $\gamma_2^*(A \circ P)$ and $\gamma_2^\infty(A)$, respectively.

Theorem 16 (Linial and Shraibman) *Let A, B be sign matrices. Then*

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

4 Product theorems for γ_2

In this section, we show that γ_2, γ_2^*, and γ_2^∞ all behave nicely under the tensor product of their arguments. This, together with Theorem 16, will immediately give our main results.

Theorem 17 *Let A, B be real matrices. Then*

1. $\gamma_2(A \otimes B) = \gamma_2(A)\gamma_2(B)$
2. $\gamma_2^\infty(A \otimes B) = \gamma_2^\infty(A)\gamma_2^\infty(B)$
3. $\gamma_2^*(A \otimes B) = \gamma_2^*(A)\gamma_2^*(B)$.

Item (3) has been previously shown by [CSUU07]. The following easy lemma will be useful in the proof of the theorem.

Lemma 18 *Let $\|\cdot\|$ be a norm on Euclidean space. If for every $x \in \mathbb{R}^m$,$y \in \mathbb{R}^n$*

$$\|x \otimes y\| \leq \|x\|\|y\|,$$

then, for every $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^n$

$$\|\alpha \otimes \beta\|^* \geq \|\alpha\|^*\|\beta\|^*,$$

where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$.

Proof: For a vector γ denote by x_γ a vector satisfying $\|x_\gamma\| = 1$ and

$$\langle \gamma, x_\gamma \rangle = \max_{x \in \mathbb{R}^n, \|x\|=1} \langle \gamma, x \rangle = \|\gamma\|^*.$$

Then, for every $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^n$

$$\begin{aligned} \|\alpha \otimes \beta\|^* &= \max_{x \in \mathbb{R}^{mn}, \|x\|=1} \langle \alpha \otimes \beta, x \rangle \\ &\geq \langle \alpha \otimes \beta, x_\alpha \otimes x_\beta \rangle \\ &= \langle \alpha, x_\alpha \rangle \langle \beta, x_\beta \rangle \\ &= \|\alpha\|^* \|\beta\|^*. \end{aligned}$$

For the first inequality recall that $\|x_\alpha \otimes x_\beta\| \leq \|x_\alpha\| \|x_\beta\| = 1$. □

Now we are ready for the proof of Theorem 17

Proof of Theorem 17: We will first show items 1 and 2 .

To see $\gamma_2(A \otimes B) \geq \gamma_2(A)\gamma_2(B)$, let Q_A be a matrix with $\|Q_A\| = 1$, such that $\gamma_2(A) = \|A \circ Q_A\|$, and similarly let Q_B satisfy $\|Q_B\| = 1$ and $\gamma_2(B) = \|B \circ Q_B\|$. Now consider the matrix $Q_A \otimes Q_B$. Notice that $\|Q_A \otimes Q_B\| = 1$. Thus

$$\gamma_2(A \otimes B) \geq \|(A \otimes B) \circ (Q_A \otimes Q_B)\| \quad = \|(A \circ Q_A) \otimes (B \circ Q_B)\| = \|A \circ Q_A\| \|B \circ Q_B\|.$$

The same proof shows that $\gamma_2^\infty(A \otimes B) \geq \gamma_2^\infty(A)\gamma_2^\infty(B)$ with the additional observation that if $Q_A \circ A \geq 0$ and $Q_B \circ B \geq 0$ then $(Q_A \otimes Q_B) \circ (A \otimes B) \geq 0$.

For the other direction, we use the min formulation of γ_2. Let X_A, Y_A be such that $X_A Y_A = A$ and $\gamma_2(A) = r(X_A)c(Y_A)$ and similarly let X_B, Y_B be such that $X_B Y_B = B$ and $\gamma_2(B) = r(X_B)c(Y_B)$. Then

$$(X_A \otimes X_B)(Y_A \otimes Y_B) = A \otimes B$$

gives a factorization of $A \otimes B$, and $r(X_A \otimes X_B) = r(X_A)r(X_B)$ and similarly $c(Y_A \otimes Y_B) = c(Y_A)c(Y_B)$.

The same proof shows that $\gamma_2^\infty(A \otimes B) \leq \gamma_2^\infty(A)\gamma_2^\infty(B)$ with the additional observation that if $X_A Y_A \circ A \geq 1$ and $X_B Y_B \circ B \geq 1$ then $(X_A \otimes X_B)(Y_A \otimes Y_B) \circ (A \otimes B) \geq 1$.

We now turn to item 3. As we have already shown $\gamma_2(A \otimes B) \leq \gamma_2(A)\gamma_2(B)$, thus by Lemma 18 it suffices to show that $\gamma_2^*(A \otimes B) \leq \gamma_2^*(A)\gamma_2^*(B)$.

To this end, let X_A, Y_A be an optimal factorization for A and similarly X_B, Y_B for B. That is, $X_A^T Y_A = A, X_B^T Y_B = B$, the columns of X_A, Y_A, X_B, Y_B are orthogonal, and $\gamma_2^*(A) = \|X_A\|_F \|Y_A\|_F$ and $\gamma_2^*(B) = \|X_B\|_F \|Y_B\|_F$.

Now consider the factorization $(X_A^T \otimes X_B^T)(Y_A \otimes Y_B) = A \otimes B$. It is easy to check that the columns of $X_A \otimes X_B$ and $Y_A \otimes Y_B$ remain orthogonal, and so

$$\begin{aligned} \gamma_2^*(A \otimes B) &\leq \|X_A \otimes X_B\|_F \|Y_A \otimes Y_B\|_F \\ &= \|X_A\|_F \|Y_A\|_F \|X_B\|_F \|Y_B\|_F \\ &= \gamma_2^*(A)\gamma_2^*(B). \end{aligned}$$

□

5 Direct product theorem for discrepancy

Shaltiel showed a direct product theorem for discrepancy under the uniform distribution as follows:

$$\mathrm{disc}_{U^{\otimes k}}(A^{\otimes k}) = O(\mathrm{disc}_U(A)^{k/3})$$

Our first result generalizes and improves Shaltiel's result to give an optimal product theorem, up to constant factors.

Theorem 19 *For any sign matrices A, B and probability distributions on their entries P, Q*

$$\mathrm{disc}_P(A)\mathrm{disc}_Q(B) \leq \mathrm{disc}_{P\otimes Q}(A \otimes B) \leq 64\, \mathrm{disc}_P(A)\mathrm{disc}_Q(B)$$

Proof: It follows directly from the definition of discrepancy that

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

For the other inequality, we have

$$\begin{aligned}\mathrm{disc}_{P\otimes Q}(A \otimes B) &\leq \gamma_2^*((A \otimes B) \circ (P \otimes Q)) \\ &= \gamma_2^*((A \circ P) \otimes (B \circ Q)) \\ &= \gamma_2^*(A \circ P)\gamma_2^*(B \circ Q) \\ &\leq 64\, \mathrm{disc}_P(A)\mathrm{disc}_Q(B)\end{aligned}$$

□

A simple example shows that we cannot expect a perfect product theorem. Let H be the 2-by-2 Hadamard matrix

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

which also represents the communication problem inner product on one bit. It is not too difficult to verify $\mathrm{disc}(H) = \mathrm{disc}_U(H) = 1/2$, where U represents the uniform distribution. On the other hand $\mathrm{disc}_{U\otimes U}(H \otimes H) \geq 5/16$ as witnessed by the vector $x = [1, 1, 1, 0]$.

Shaltiel also asked whether a direct product theorem holds for general discrepancy $\mathrm{disc}(A) = \min_P \mathrm{disc}_P(A)$. The function inner product can also be used here to show we cannot expect a perfect product theorem. As stated above, for the inner product function on one bit, $\mathrm{disc}(H) = 1/2$. Thus if discrepancy obeyed a perfect product theorem, then, $\mathrm{disc}(H^{\otimes k}) = 2^{-k}$. On the other hand, $\gamma_2^\infty(H^{\otimes k}) = 2^{k/2}$—for the upper bound look at the trivial factorization $IH^{\otimes k}$, and for the lower bound take the matrix Q to be $H^{\otimes k}$ itself. Thus we obtain a contradiction for sufficiently large k as $\gamma_2^\infty(A)$ and $1/\mathrm{disc}(A)$ differ by at most a multiplicative factor of 8.

Our next theorem shows that this example is nearly the largest violation possible.

Theorem 20 *Let A, B be sign matrices. Then*

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

Proof: By Theorem 16 and Theorem 17 we have

$$\mathrm{disc}(A \otimes B) \leq \frac{1}{\gamma_2^\infty(A \otimes B)} = \frac{1}{\gamma_2^\infty(A)\gamma_2^\infty(B)} \leq 64\,\mathrm{disc}(A)\mathrm{disc}(B).$$

Similarly,

$$\mathrm{disc}(A \otimes B) \geq \frac{1}{8}\frac{1}{\gamma_2^\infty(A \otimes B)} = \frac{1}{8}\frac{1}{\gamma_2^\infty(A)\gamma_2^\infty(B)} \geq \frac{1}{8}\mathrm{disc}(A)\mathrm{disc}(B)$$

□

These two theorems taken together mean that for a tensor product $A \otimes B$ there is a tensor product distribution $P \otimes Q$ that gives a nearly optimal bound for discrepancy. We state this as a corollary:

Corollary 21 *Let A, B be sign matrices. Then*

$$\frac{1}{512}\mathrm{disc}_{P\otimes Q}(A \otimes B) \leq \mathrm{disc}(A \otimes B) \leq 64\,\mathrm{disc}_{P\otimes Q}(A \otimes B),$$

where P is the optimal distribution for $\mathrm{disc}(A)$ *and Q is the optimal distribution for* $\mathrm{disc}(B)$.

5.1 Applications

Now we discuss some applications of our product theorem for discrepancy. We first show how our results give a strong direct product theorem in distributional complexity, for bounds shown by the discrepancy method.

Theorem 22 *Let $f : X \times Y \to \{0,1\}^n$ be a Boolean function and P a probability distribution over $X \times Y$. If $\mathrm{Corr}_{c,P}(M_f) \leq w$ is proved by the discrepancy method (Fact 5), then*

$$\mathrm{Corr}_{kc,P^{\otimes k}}(M_f^{\otimes k}) \leq (8w)^k$$

Proof: By generalizing Theorem 19 to tensor products of more matrices,

$$\begin{aligned}\mathrm{Corr}_{kc,P^{\otimes k}}(M_f^{\otimes k}) &\leq 2^{kc}\mathrm{disc}_{P^{\otimes k}}(M_f^{\otimes k})\\ &\leq 2^{kc}(8 \cdot \mathrm{disc}_P(M_f))^k\\ &\leq (8 \cdot 2^c \mathrm{disc}_P(M_f))^k\end{aligned}$$

□

This is a strong direct product theorem as even with k times the original amount c of communication, the correlation still decreases exponentially. Note, however, that we can only show this for bounds shown by the discrepancy method—it remains an interesting open problem if a direct product theorem holds for distributional complexity in general.

As results of Klauck (stated in our Theorem 7) show that discrepancy captures the complexity of weakly-unbounded error protocols, we can show an unconditional direct sum theorem for this entire class.

Theorem 23 *Let $f_i : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be Boolean functions, for $1 \le i \le k$. Then*

$$\mathrm{UPC}\left(\bigoplus_{i=1}^{k} f_i\right) \ge \frac{1}{3}\left(\sum_{i=1}^{k} \mathrm{UPC}(f_i)\right) - \frac{k}{3}\log n - O(1).$$

Similarly one also obtains direct sum results for lower bounds on randomized or quantum communication complexity with entanglement shown via the discrepancy method.

5.2 Connections to recent work

There have been several recent papers which discuss issues related to those here. We now explain some of the connections between our work and these results.

Viola and Wigderson [VW07] study direct product theorems for, among other things, multiparty communication complexity. For the two-party case, they are able to recover Shaltiel's result, with a slightly worse constant in the exponent. The quantity which they bound is correlation with two-bit protocols, which they remark is equal to discrepancy, up to a constant factor. Indeed, in our language, the maximum correlation of a sign matrix A with a two-bit protocol under a distribution P is exactly $\|A \circ P\|_{\infty\to 1}$. This is because a two-bit protocol in the ± 1 representation is described by a rank one sign matrix.

The infinity-to-one norm also plays an important role in a special class of two-prover games known as XOR games. Here the verifier wants to evaluate some function $f : X \times Y \to \{-1,1\}$, and with probability $P[x,y]$, sends question x to Alice and question y to Bob. The provers Alice and Bob are all powerful, but cannot communicate. Alice and Bob send responses $a_x, b_y \in \{-1,1\}$ back to the verifier who checks if $a_x \cdot b_y = f(x,y)$. Here we see that a strategy of Alice is given by a sign vector $\mathbf{a}$ of length $|S|$, and similarly for Bob. Thus the maximum correlation the provers can achieve with f is

$$\max_{\mathbf{a}\in\{-1,1\}^{|S|},\mathbf{b}\in\{-1,1\}^{|T|}} \mathbf{a}^T(M_f \circ P)\mathbf{b},$$

which is exactly $\|M_f \circ P\|_{\infty\to 1}$.

Two-prover XOR games have also been studied where the provers are allowed to share entanglement. In this case, results of Tsirelson [Tsi87] show that the best correlation achievable can be described by a semidefinite program [CHTW04]. In fact, the best correlation achievable by entangled provers under distribution P turns out to be given exactly by $\gamma_2^*(M_f \circ P)$. In studying a parallel repetition theorem for XOR games with entanglement, [CSUU07] have already shown, in our language, that $\gamma_2^*(A \otimes B) = \gamma_2^*(A)\gamma_2^*(B)$.

This connection to XOR games also gives another possible interpretation of the quantity $\gamma_2^\infty(A)$. The best correlation the provers can achieve with M_f under the "hardest" probability distribution P is given by $1/\gamma_2^\infty(A)$.

Finally, inspired by the work of [CSUU07], Mittal and Szegedy [MS07] have begun to develop a general theory of when semidefinite programs obey a product theorem. While γ_2 and γ_2^* fit into their framework, interestingly γ_2^∞ does not.

6 Conclusion

We have shown a tight product theorem for discrepancy by looking at semidefinite relaxation of discrepancy which gives a constant factor approximation, and which composes perfectly under tensor product. With the great success of semidefinite programming in approximation algorithms we feel that such an approach should find further applications.

Many open questions remain. Can one show a product theorem for γ_2^ϵ? We have only been able to show a very weak result in this direction:

$$\gamma_2^{\epsilon^2/2(1+\epsilon)}(A \otimes A) \geq \gamma_2^\epsilon(A)\gamma_2^\epsilon(A)$$

It would be nice to continue in the line of work of Mittal and Szegedy [MS07] to understand what conditions are necessary and sufficient for a semidefinite program to obey a product rule. While their sufficient condition captures γ_2, γ_2^*, it does not yet work for programs like γ_2^∞, or the semidefinite relaxation of two-prover games studied by Feige and Lovasz [FL92].

Finally, an outstanding open question which remains is if a direct product theorem holds for the randomized communication complexity of disjointness. Razborov's [Raz92] proof of the $\Omega(n)$ lower bound for disjointness uses a one-sided version of discrepancy under a non-product distribution. Could a similar proof technique apply by first characterizing one sided discrepancy as a semidefinite program?

References

[AN06] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM Journal on Computing*, 35:787–803, 2006.

[CHTW04] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Proceedings of the 19th IEEE Conference on Computational Complexity*, pages 236–249. IEEE, 2004.

[CSUU07] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*. IEEE, 2007.

[FG05] J. Ford and A. Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *Proceedings of the 32th International Colloquium On Automata, Languages and Programming*, pages 1163–1175, 2005.

[FKL$^+$01] J. Forster, M. Krause, S. Lokam, R. Mubarakzjanov, N. Schmitt, and H. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Foundations of Software Technology and Theoretical Computer Science*, pages 171–182, 2001.

[FL92] U. Feige and L. Lovász. Two-prover one-round proof systems: their power and their problems. In *Proceedings of the 24th ACM Symposium on the Theory of Computing*, pages 733–744. ACM, 1992.

[Kla01] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[Kre95] I. Kremer. Quantum communication. Technical report, Hebrew University of Jerusalem, 1995.

[Kri79] J. Krivine. Constantes de Grothendieck et fonctions de type positif sur les sphères. *Adv. Math.*, 31:16–30, 1979.

[KŠW07] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007.

[LMSS07] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 2007. To appear.

[Lov79] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, IT-25:1–7, 1979.

[LS06] N. Linial and A. Shraibman. Learning complexity versus communication complexity. Available at `http://www.cs.huji.ac.il/~nati/`, 2006.

[LS07] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*. ACM, 2007.

[Mat93] R. Mathias. The Hadamard operator norm of a circulant and applications. *SIAM journal on matrix analysis and applications*, 14(4):1152–1167, 1993.

[MS07] R. Mittal and M. Szegedy. Product rules in semidefinite programming. In *16th International Symposium on Fundamentals of Computation Theory*, 2007.

[Raz92] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

[Raz00] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[Ree91] J. Reeds. A new lower bound on the real Grothendieck constant. Available at `http://www.dtc.umn.edu/~reedsj`, 1991.

[Sha03] R. Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1–2):1–22, 2003.

[She07a] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. Technical report, ECCC TR07-100, 2007.

[She07b] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*. ACM, 2007.

[Tsi87] B. Tsirelson. Quantum analouges of the Bell inequalities: the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

[VW07] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*. IEEE, 2007.