Composition theorems in communication complexity

Troy Lee[1] and Shengyu Zhang[2]

[1] Rutgers University, troyjlee@gmail.com
[2] The Chinese University of Hong Kong, syzhang@cse.cuhk.edu.hk

Abstract. A well-studied class of functions in communication complexity are composed functions of the form $(f \circ g^n)(x, y) = f(g(x^1, y^1), \ldots, g(x^n, y^n))$. This is a rich family of functions which encompasses many of the important examples in the literature. It is thus of great interest to understand what properties of f and g affect the communication complexity of $(f \circ g^n)$, and in what way.

Recently, Sherstov [She09] and independently Shi-Zhu [SZ09b] developed conditions on the inner function g which imply that the quantum communication complexity of $f \circ g^n$ is at least the approximate polynomial degree of f. We generalize both of these frameworks. We show that the pattern matrix framework of Sherstov works whenever the inner function g is *strongly balanced*—we say that $g : X \times Y \to \{-1, +1\}$ is strongly balanced if all rows and columns in the matrix $M_g = [g(x, y)]_{x,y}$ sum to zero. This result strictly generalizes the pattern matrix framework of Sherstov [She09], which has been a very useful idea in a variety of settings [She08b,RS08,Cha07,LS09a,CA08,BHN09].

Shi-Zhu require that the inner function g has small *spectral discrepancy*, a somewhat awkward condition to verify. We relax this to the usual notion of discrepancy.

We also enhance the framework of composed functions studied so far by considering functions $F(x, y) = f(g(x, y))$, where the range of g is a group G. When G is Abelian, the analogue of the strongly balanced condition becomes a simple group invariance property of g. We are able to formulate a general lower bound on F whenever g satisfies this property.

1 Introduction

Communication complexity studies the minimum amount of communication needed to compute a function whose input variables are distributed between two or more parties. Since the introduction by Yao [Yao79] of an elegant mathematical model to study this question, communication complexity has grown into a rich field both because of its inherent mathematical interest and also its application to many other models of computation. See the textbook of Kushilevitz and Nisan [KN97] for a comprehensive introduction to the field.

In analogy with traditional computational complexity classes, one can consider different models of communication complexity based on the resources available to the parties. Besides the standard deterministic model, of greatest interest to us will be a randomized version of communication complexity, where the parties have access to a source of randomness and are allowed to err with some small constant probability, and a quantum model where the parties share a quantum channel and the cost is measured in qubits.

Several major open questions in communication complexity ask about how different complexity measures relate to each other. The log rank conjecture, formulated by Lovász and Saks [LS88], asks if the deterministic communication complexity of a Boolean function $F : X \times Y \to \{0, 1\}$ is upper bounded by a polynomial in the logarithm of the rank of the matrix $[F(x, y)]_{x,y}$. Another major open question is if randomized and quantum communication complexity are polynomially related for all total functions. We should mention here that the assumption of the function being total is crucial as an exponential separation is known for a partial function [Raz99].

One approach to these questions has been to study them for restricted classes of functions. Many functions of interest are *block composed* functions. For finite sets X, Y, and E, a function $f : E^n \to \{-1, +1\}$, and a function $g : X \times Y \to E$, the block composition of f and g is the function $f \circ g^n : X^n \times Y^n \to \{-1, +1\}$ defined by $(f \circ g^n)(x, y) = f(g(x^1, y^1), \ldots, g(x^n, y^n))$ where $(x^i, y^i) \in X \times Y$ for all $i = 1, \ldots, n$. For example, if $E = \{-1, +1\}$, the inner product function results when f is PARITY and g is AND, set-intersection when f is OR and g is AND, and the equality function when f is AND and g is the function IS-EQUAL, which is one if and only if $x = y$.

In a seminal paper, Razborov [Raz03] gave tight bounds for the bounded-error quantum communication complexity of block composed functions where the outer function is symmetric and the inner function is bitwise AND. In particular, this result showed that randomized and quantum communication complexity are polynomially related for such functions.

More recently, very nice frameworks have been developed by Sherstov [She07,She09] and independently by Shi and Zhu [SZ09b] to bound the quantum complexity of block composed functions that goes beyond the case of symmetric f to work for any f provided the inner function g satisfies certain technical conditions. When g satisfies these conditions, these frameworks allow one to lower bound the quantum communication complexity of $f \circ g^n$ in terms of $\deg_\epsilon(f)$, the approximate polynomial degree of f, a well-studied measure.

Shi and Zhu are able to get a bound on $f \circ g^n$ in terms of the approximate degree of f whenever g is sufficiently "hard"—unfortunately, the hardness condition they need is in terms of "spectral discrepancy," a quantity which is somewhat difficult to bound, and their bound requires that g is a function on at least $\Omega(\log(n/d))$ bits, where d is the approximate polynomial degree of f. Because of this, Shi-Zhu are only able to reproduce Razborov's results with a polynomially weaker bound.

Sherstov developed so-called *pattern matrices* which are the matrix representation of a block composed function when g is a fixed function of a particularly nice form. Namely, in a pattern matrix the inner function $g : \{-1, +1\}^k \times ([k] \times \{-1, +1\}) \to \{-1, +1\}$ is parameterized by a positive integer k and defined by $g(x, (i, b)) = x_i \cdot b$, where x_i denotes the i^{th} bit of x. With this g, Sherstov shows that $\deg_\epsilon(f)$ is a lower bound on the quantum communication complexity of $f \circ g^n$, for any function f. Though seemingly quite special, pattern matrices have proven to be an extremely useful concept. First, they give a simple proof of Razborov's tight lower bounds for $f(x \wedge y)$ for symmetric f. Second, they have also found many other applications in unbounded-error communication complexity [She08b,RS08] and have been successfully extended to multiparty communication complexity [Cha07,LS09a,CA08,BHN09].

A key step in both the works of Sherstov and Shi-Zhu is to bound the spectral norm of a sum of matrices $\|\sum_i B_i\|$. This is the major step where these works differ. Shi-Zhu apply the triangle inequality to bound this as $\|\sum_i B_i\| \leq \sum_i \|B_i\|$. On the other hand, Sherstov observes that, in the case of pattern matrices, the terms of this sum are mutually orthogonal, *i.e.* $B_i^\dagger B_j = B_i B_j^\dagger = 0$ for all $i \neq j$. In this case, one has a stronger bound on the spectral norm $\|\sum_i B_i\| = \max_i \|B_i\|$.

We extend both the frameworks of Sherstov and Shi-Zhu. In the case of Shi-Zhu, we are able to reprove their theorem with the usual notion of discrepancy instead of the somewhat awkward spectral discrepancy. The main observation we make is that as all Shi-Zhu use in this step is the triangle inequality, we can repeat the argument with any norm here, including discrepancy itself.

In the case of pattern matrices, special properties of the spectral norm are used, namely the fact about the spectral norm of a sum of orthogonal matrices. We step back to see what key features of a pattern matrix lead to this orthogonality property. We begin with the Boolean case, that is, where the intermediate set E is taken to be $\{-1, +1\}$. In this case, a crucial concept is the notion of a *strongly balanced* function. We say that $g : X \times Y \to \{-1, +1\}$ is strongly balanced if in the sign matrix $M_g[x, y] = g(x, y)$ all rows and all columns

sum to zero. We show that whenever the inner function g is strongly balanced, the key orthogonality condition holds; this implies that whenever g is strongly balanced and has discrepancy under the uniform distribution bounded away from one, the approximate degree of the outer function f is a lower bound on the quantum communication complexity of $f \circ g^n$.

We also consider the general case where the intermediate set is any group G. That is, we consider functions $F(x,y) = f(g(x,y))$, where $g : X \times Y \to G$ for a group G and $f : G \to \{-1,+1\}$ is a class function on G. The case $E = \{-1,+1\}$ discussed above corresponds to taking the group $G = \mathbb{Z}_2^n$. When G is a general Abelian group, the key orthogonality condition requires more than that the matrix $M_g[x,y] = g(x,y)$ is strongly balanced; still, it admits a nice characterization in terms of group invariance. A multiset $T \in G \times G$ is said to be G-invariant if $(s,s)T = T$ for all $s \in G$. The orthogonality condition will hold if and only if all pairs of rows and all pairs of columns of M_g (when viewed as multisets) are G- invariant. One can generalize the results discussed above to this general setting with appropriate modifications. In the case that $G = \mathbb{Z}_2^n$, the G-invariant condition degenerates to the strongly balanced requirement of M_g.

2 Preliminaries

All logarithms are base two. For a complex number $z = a + ib$ we let $\bar{z} = a - ib$ denote the complex conjugate of z and $|z| = \sqrt{a^2 + b^2}$ and $\mathrm{Re}(z) = a$.

2.1 Complexity measures

We will make use of several complexity measures of functions and matrices. Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be a function. For $T \subseteq \{0,1\}^n$, the Fourier coefficient of f corresponding to the character χ_T is $\hat{f}_T = \frac{1}{2^n} \sum_x f(x)\chi_T(x) = \frac{1}{2^n} \sum_x f(x) \prod_{i \in T} x_i$. The *degree* of f as a polynomial, denoted $\deg(f)$, is the size of a largest set T for which $\hat{f}_T \neq 0$.

We reserve J for the all ones matrix, whose size will be determined by the context. For a matrix A let $A^\dagger$ denote the conjugate transpose of A. We use $A * B$ for the entrywise product of A, B, and $A \otimes B$ for the tensor product. If A is an m-by-n matrix then we say that $\mathrm{size}(A) = mn$. We use $\langle A, B \rangle = \mathrm{Tr}(AB^\dagger)$ for the inner product of A and B.

Let $\|A\|_1$ be the ℓ_1 norm of A, i.e. sum of the absolute values of entries of A, and $\|A\|_\infty$ the maximum absolute value of an entry. For a positive semidefinite matrix M let $\lambda_1(M) \geq \cdots \geq \lambda_n(M) \geq 0$ be the eigenvalues of M. We define the i^{th} singular value of A, denoted $\sigma_i(A)$, as $\sigma_i(A) = \sqrt{\lambda_i(AA^\dagger)}$. The rank of A, denoted $\mathrm{rk}(A)$ is the number of nonzero singular values of A. We will use several matrix norms. The spectral or operator norm is the largest singular value $\|A\| = \sigma_1(A)$, the trace norm is the summation of all singular values $\|A\|_{tr} = \sum_i \sigma_i(A)$, and the Frobenius norm is the ℓ_2 norm of the singular values $\|A\|_F = \sqrt{\sum_i \sigma_i(A)^2}$. When $AB^\dagger = A^\dagger B = 0$ we will say that A, B are *orthogonal*. Please note the difference with the common use of this term, which usually means $\langle A, B \rangle = 0$.

Fact 1 *For two matrices A, B of the same dimensions, if $AB^\dagger = A^\dagger B = 0$, then*

$$\mathrm{rk}(A+B) = \mathrm{rk}(A) + \mathrm{rk}(B), \;\; \|A+B\|_{tr} = \|A\|_{tr} + \|B\|_{tr}, \;\; \|A+B\| = \max\{\|A\|, \|B\|\}.$$

Another norm we will use is the γ_2 norm, introduced to complexity theory in [LMSS07], and familiar in matrix analysis as the Schur product operator norm. The γ_2 norm can be viewed as a weighted version of the trace norm.

Definition 1.

$$\gamma_2(A) = \max_{u,v : \|u\| = \|v\| = 1} \|A * uv^\dagger\|_{tr}.$$

It is clear from this definition that $\gamma_2(A) \geq \|A\|_{tr}/\sqrt{mn}$ for an m-by-n matrix A.

For a norm Φ, the dual norm Φ^* is defined as $\Phi^*(v) = \max_{u:\Phi(u)\leq 1} |\langle u, v\rangle|$. The norm γ_2^*, dual to the γ_2 norm, looks as follows.

Definition 2.

$$\gamma_2^*(A) = \max_{\substack{u_i, v_j: \\ \|u_i\|=\|v_j\|=1}} \sum_{i,j} A[i,j]\langle u_i, v_j\rangle.$$

Another complexity measure we will make use of is discrepancy.

Definition 3. *Let A be an m-by-n sign matrix and let P be a probability distribution on the entries of A. The discrepancy of A with respect to P, denoted $\mathrm{disc}_P(A)$, is defined as*

$$\mathrm{disc}_P(A) = \max_{x\in\{0,1\}^m, y\in\{0,1\}^n} |x^\dagger (A * P) y|.$$

We will write $\mathrm{disc}_U(A)$ for the special case where P is the uniform distribution. It is easy to see from this definition that $\mathrm{disc}_U(A) \leq \frac{\|A\|}{\sqrt{\mathrm{size}(A)}}$. Shaltiel [Sha03] has shown the deeper result that this bound is in fact polynomially tight:

Theorem 2 (Shaltiel). *Let A be a sign matrix. Then*

$$\frac{1}{108}\left(\frac{\|A\|}{\sqrt{\mathrm{size}(A)}}\right)^3 \leq \mathrm{disc}_U(A).$$

Discrepancy and the γ_2^* norm are very closely related. Linial and Shraibman [LS09c] observed that Grothendieck's inequality gives the following.

Theorem 3 (Linial-Shraibman). *For any sign matrix A and probability distribution P*

$$\mathrm{disc}_P(A) \leq \gamma_2^*(A * P) \leq K_G \,\mathrm{disc}_P(A)$$

where $1.67\ldots \leq K_G \leq 1.78\ldots$ is Grothendieck's constant.

Approximate measures We will also use approximate versions of these complexity measures which come in handy when working with bounded-error models. Say that a function g gives an ϵ-approximation to f if $|f(x) - g(x)| \leq \epsilon$ for all $x \in \{-1,+1\}^n$. The ϵ-approximate polynomial degree of f, denoted $\deg_\epsilon(f)$, is the minimum degree of a function g which gives an ϵ-approximation to f. We will similarly look at the ϵ-approximate version of the trace and γ_2 norms. We give the general definition with respect to any norm.

Definition 4 (approximation norm). *Let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary norm. Let $v \in \mathbb{R}^n$ be a sign vector. For $0 \leq \epsilon < 1$ we define the approximation norm Φ^ϵ as*

$$\Phi^\epsilon(v) = \min_{u:\|v-u\|_\infty \leq \epsilon} \Phi(u).$$

Notice that an approximation norm Φ^ϵ is not itself a norm—we have only defined it for sign vectors, and it will in general not satisfy the triangle inequality.

As a norm is a convex function, using the separating hyperplane theorem one can quite generally give the following equivalent dual formulation of an approximation norm.

Proposition 1. *Let $v \in \mathbb{R}^n$ be a sign vector, and $0 \leq \epsilon < 1$*

$$\Phi^\epsilon(v) = \max_u \frac{|\langle v, u\rangle| - \epsilon\|u\|_1}{\Phi^*(u)}$$

A proof of this can be found in the survey [LS09b].

2.2 Communication complexity

Let X, Y, S be finite sets and $f : X \times Y \to S$ be a function. We will let $D(f)$ be the deterministic communication complexity of f, and $R_\epsilon(f)$ denote the randomized public coin complexity of f with error probability at most ϵ. We refer to the reader to [KN97] for a formal definition of these models. We will also study $Q_\epsilon(f)$ and $Q^*_\epsilon(f)$, the ϵ-error quantum communication complexity of f without and with shared entanglement, respectively. We refer the reader to [Raz03] for a nice description of these models.

For notational convenience, we will identify a function $f : X \times Y \to \{-1, +1\}$ with its sign matrix $M_f = [f(x, y)]_{x,y}$. Thus, for example, $\|f\|$ refers to the spectral norm of the sign matrix representation of f.

For all of our lower bound results we will actually lower bound the approximate trace norm or γ_2 norm of the function. Razborov showed that the approximate trace norm can be used to lower bound on quantum communication complexity, and Linial and Shraibman generalized this to the γ_2 norm.

Theorem 4 (Linial-Shraibman [LS09d]). *Let A be a sign matrix and $0 \leq \epsilon < 1/2$. Then*

$$Q^*_\epsilon(A) \geq \log\left(\gamma_2^{2\epsilon}(A)\right) - 2.$$

Composed functions Before discussing lower bounds on a block composed function $f \circ g^n$, let us see what we expect the complexity of such a function to be. A fundamental idea going back to Nisan [Nis94] and Buhrman, Cleve, and Wigderson [BCW98], is that the complexity of $f \circ g^n$ can be related to the *query* complexity of f and the communication complexity of g. This holds true for deterministic, randomized, and quantum models of communication complexity and query complexity. For formal definitions of these measures and a survey of query complexity we recommend [BW02].

One advantage of working with block composed functions in light of this is that query complexity is in general better understood than communication complexity. In particular, a polynomial relationship between deterministic query complexity and degree, and randomized and quantum query complexities and approximate degree is known. Putting these two facts together gives the following corollary:

Corollary 1.

$$D(f \circ g^n) = O(\deg(f)^4 D(g)), \quad R_{1/4}(f \circ g^n) = O(\deg_{1/4}(f)^6 R_{1/4}(g) \log \deg_{1/4}(f))$$

Our goal, then, in showing lower bounds on the complexity of a block composed function $f \circ g^n$ is to get something at least in the ballpark of this upper bound. Of course, this is not always possible. For example, when f is the PARITY function on n bits, and $g(x, y) = \oplus(x, y)$ this protocol just gives an upper bound of n bits, when the true complexity is constant. See recent results by Zhang [Zha09] and Sherstov [She10] for discussions on the tightness of the bounds in Corollary 1.

3 Rank of block composed functions

We begin by analyzing the rank of a block composed function $f \circ g^n$ when the inner function g is strongly balanced. This case will illustrate the use of the strongly balanced assumption, and is simpler to understand than the bounded-error situation treated in the next section.

Let us first formally state the definition of strongly balanced.

Definition 5 (strongly balanced). *Let A be a sign matrix, and J be the all ones matrix of the same dimensions as A. We say that A is* balanced *if $\mathrm{Tr}(AJ^\dagger) = 0$. We further say that A is* strongly balanced *if $AJ^\dagger = A^\dagger J = 0$. We will say that a two-variable Boolean function is balanced or strongly balanced if its sign matrix representation is.*

Theorem 5. *Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be an arbitrary function, and let g be a strongly balanced function. Then*

$$\mathrm{rk}(M_{f\circ g^n}) = \sum_{T\subseteq[n],\ \hat{f}_T\neq 0} \mathrm{rk}(M_g)^{|T|}.$$

Proof. Let us write out the sign matrix for $\chi_T \circ g^n$ explicitly. If we let $M_g^0 = J$ be the all ones matrix and $M_g^1 = M_g$, then we can nicely write the sign matrix representing $\chi_T(g(x^1,y^1),\ldots,g(x^n,y^n))$ as $M_{\chi_T\circ g^n} = \bigotimes_i M_g^{T[i]}$ where $T[i] = 1$ if $i \in T$ and 0 otherwise.

We see that the condition on g implies $M_{\chi_T\circ g^n} M^\dagger_{\chi_S\circ g^n} = 0$ if $S \neq T$. Indeed,

$$M_{\chi_T\circ g^n} M^\dagger_{\chi_S\circ g^n} = \left(\bigotimes_i M_g^{T[i]}\right)\left(\bigotimes_i M_g^{S[i]}\right)^\dagger = \bigotimes_i \left(M_g^{T[i]}(M_g^{S[i]})^\dagger\right) = 0.$$

This follows since, by the assumption $S \neq T$, there is some i for which $S[i] \neq T[i]$ which means that this term is either $M_g J^\dagger = 0$ or $J M_g^\dagger = 0$ because g is strongly balanced. The other case follows similarly.

Now that we have established this property, we can use Fact 1 to obtain

$$\mathrm{rk}(M_{f\circ g^n}) = \mathrm{rk}\Big(\sum_{T\subseteq[n]} \hat{f}_T \chi_T(g(x^1,y^1),\ldots,g(x^n,y^n))\Big) = \sum_{\substack{T\subseteq[n]\\ \hat{f}_T\neq 0}} \mathrm{rk}(M_{\chi_T\circ g^n}) = \sum_{\substack{T\subseteq[n]\\ \hat{f}_T\neq 0}} \mathrm{rk}(M_g)^{|T|}$$

In the last step we used the fact that rank is multiplicative under tensor product.

In particular, this theorem means that whenever g is a strongly balanced function on a constant number of bits and $\mathrm{rk}(M_g) > 1$, then the log rank conjecture holds for $f \circ g^n$.

4 A bound in terms of approximate degree

In this section, we will address the frameworks of Sherstov and Shi-Zhu. We extend both of these frameworks to give more general conditions on the inner function g which still imply that the approximate degree of f is a lower bound on the quantum query complexity of the composed function $f \circ g^n$. In outline, both of these frameworks follow the same plan. By Theorem 4 it suffices to lower bound the approximate γ_2 norm (or approximate trace norm) of $f \circ g^n$. To do this, they use the dual formulation given by Proposition 1 and construct a witness matrix B which has non-negligible correlation with the target function and small γ_2^* (or spectral) norm.

A very nice way to construct this witness, used by both Sherstov and Shi-Zhu, is to use the *dual polynomial* of f. This is a polynomial v which certifies that the approximate polynomial degree of f is at least a certain value. More precisely, duality theory of linear programming gives the following lemma.

Lemma 1 (Sherstov [She09], Shi-Zhu [SZ09b]). *Let $f : \{-1,+1\}^n \to \{-1,+1\}$ and let $d = \deg_\epsilon(f)$. Then there exists a function $v : \{-1,+1\}^n \to \mathbb{R}$ such that*

1. *$\langle v, \chi_T\rangle = 0$ for every character χ_T with $|T| < d$.*
2. *$\langle v, f\rangle \geq \epsilon$.*
3. *$\|v\|_1 = 1$.*

Items (2),(3) are used to lower bound the correlation of the witness matrix with the target matrix and to upper bound the ℓ_1 norm of the witness matrix. In the most difficult step, and where these works diverge, Item (1) is used to upper bound the γ_2^* (or spectral) norm of the witness matrix. We treat each of these frameworks separately in the next two sections.

4.1 Sherstov's framework

The proof of the next theorem follows the same steps as Sherstov's proof for pattern matrices (Theorem 5.1 [She09]). Our main contribution is to identify the strongly balanced condition as the key property of pattern matrices which enables the proof to work.

Theorem 6. *Let X, Y be finite sets, $g : X \times Y \to \{-1, +1\}$ be a strongly balanced function, and $M_g[x, y] = g(x, y)$ be the corresponding sign matrix. Let $f : \{-1, +1\}^n \to \{-1, +1\}$ be an arbitrary function. Then for any $\epsilon > 0$ and $\epsilon_0 > 2\epsilon$, we have*

$$Q^*_\epsilon(f \circ g^n) \geq \deg_{\epsilon_0}(f) \log_2 \Big(\frac{\sqrt{|X||Y|}}{\|M_g\|} \Big) - O(1).$$

Proof (Sketch). Let $d = \deg_{\epsilon_0}(f)$ and let v be a dual polynomial for f with properties as in Lemma 1. We define a witness matrix as

$$B[x, y] = \frac{2^n}{\text{size}(M_g)^n} v(g(x^1, y^1), \ldots, g(x^n, y^n))$$

We will use the matrix B to witness that $\|M_{f \circ g^n}\|^{\epsilon_0}_{tr}$ is large via Proposition 1. The theorem will then follow from Theorem 4.

There are three quantities to evaluate. One can compute $\langle M_{f \circ g^n}, B \rangle$ and $\|B\|_1$ using properties 2 and 3 of Lemma 1 respectively, together with the fact that as M_g is strongly balanced, it is in particular balanced.

The more interesting step is to bound $\|B\|$. As shown above, the strongly balanced property of g implies that the matrices $\chi_T \circ g^n$ and $\chi_S \circ g^n$ are orthogonal for distinct sets S, T and so Fact 1 can be used to greatly simplify this computation. Details are given in the full version.

Using the theorem of Shaltiel relating discrepancy to the spectral norm (Theorem 2), we get the following corollary:

Corollary 2. *Let the quantities be defined as in Theorem 6.*

$$Q^*_{1/8}(f \circ g^n) \geq \frac{1}{3} \deg_{1/3}(f) \big(\log \big(\frac{1}{\text{disc}_U(M_g)} \big) - 7 \big) - O(1).$$

Comparison to Sherstov's pattern matrix: As mentioned in [She09], Sherstov's pattern matrix method can prove a quantum lower bound of $\Omega(\deg_\epsilon(f))$ for block composed functions $f \circ g^n$ if the matrix M_g contains the following 4×4 matrix S_4 as a submatrix:

$$S_4 = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}, \qquad S_6 = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 \end{bmatrix}$$

In this paper we show that the same lower bound holds as long as M_g contains a strongly balanced submatrix with discrepancy bounded away from one. Are there strongly balanced matrices not containing S_4 as a submatrix? It turns out that the answer is yes: we give the above 6×6 matrix S_6 as one example.

4.2 Shi-Zhu framework

The method of Shi-Zhu does not restrict the form of the inner function g, but rather works for any g which is sufficiently "hard." The hardness condition they require is phrased in terms of a somewhat awkward measure they term spectral discrepancy.

Chattopadhyay [Cha08] extended the technique of Shi-Zhu to the case of multiparty communication complexity, answering an open question of Sherstov [She08a]. In doing so, he gave a more natural condition on the hardness of g in terms of an upper bound on discrepancy frequently used in the multiparty setting and originally due to Babai, Nisan, and Szegedy [BNS92]. As all that is crucially needed is subadditivity, we do the argument here with γ_2^*, which is essentially equal to the discrepancy.

Theorem 7. *Let $f : \{-1,+1\}^n \to \{-1,+1\}$, and $g : X \times Y \to \{-1,+1\}$. Fix $0 < \epsilon < 1/2$, and let $\epsilon_0 > 2\epsilon$. Then*

$$Q_\epsilon^*(f \circ g^n) \geq \deg_{\epsilon_0}(f) - O(1).$$

provided there is a distribution μ which is balanced with respect to g and for which $\gamma_2^*(M_g * \mu) \leq \frac{\deg_{\epsilon_0}(f)}{2en}$.

Proof. We again use Proposition 1, this time with the γ_2 norm instead of the trace norm. To prove a lower bound we choose a witness matrix B as follows

$$B[x,y] = 2^n \cdot v(g(x^1,y^1),\ldots,g(x^n,y^n)) \cdot \prod_{i=1}^n \mu(x^i,y^i).$$

where v witnesses that f has approximate degree at least $d = \deg_{\epsilon_0}(f)$. This definition is the same as in the previous section where μ was simply the uniform distribution. As argued before, we have $\langle M_{f\circ g^n}, B\rangle \geq \epsilon_0$ and $\|B\|_1 = 1$ because $M_g * \mu$ is balanced.

The more interesting step is to upper bound $\gamma_2^*(B)$. We again expand B in terms of the Fourier coefficients of v. As we do not have special knowledge of the function g, we simply bound the resulting sum by the triangle inequality. Details are given in the full version.

5 A general framework for functions composed through a group

In this section we begin the study of general function compositions through a group G. In this case the inner function $g : X \times Y \to G$ has a group G as its range, and the outer function $f : G \to \{-1,+1\}$ is a class function, i.e. one that is invariant on conjugacy classes. Previous sections deal with the special case that $G = \mathbb{Z}_2^n$.

Let us recall the basic idea of the proof of Theorem 6. Following the work of Sherstov and Shi-Zhu [She09,SZ09b], to prove a lower bound on the quantum communication complexity for a composed function $f \circ g$, we constructed a witness matrix B which had non-negligible correlation with $f \circ g$ and small spectral norm. We used the dual polynomial p (of f) which has two important properties, first that p has non-negligible correlation with f and second that p has no support on low degree polynomials. We can then use the first property to show that the composed function $p \circ g$ will give non-negligible inner product with $f \circ g$ and the second to upper bound the spectral norm of $p \circ g$. The second of these tasks is the more difficult. In the case of $G = \{-1,+1\}^n$, the degree of a character χ_T is a natural measure of how "hard" the character is — the larger T is, the smaller the spectral norm of $\chi_T \circ g$ will be. In the general group case, however, it is less clear what the corresponding "hard" and "easy" characters should be. In Section 5.1, we will show that this framework actually works for an arbitrary partition of the basis functions into Easy and Hard.

In carrying out this plan, one is still left with upper bounding $\|M_{p\circ g}\|$. Here, as in the Boolean case, it is again very convenient to have an orthogonality condition which can greatly

simplify the computation of $\|M_{p\circ g}\|$ and give good bounds. In the Boolean case we have shown that M_g being strongly balanced implies this key orthogonality condition. In Section 5.2 and 5.3, we will show that for the general group, the condition is not only about each row and column of matrix M_g, but all pairs of rows and pairs of columns. In the Abelian group case, this reduces to a nice group invariance condition.

Even after applying the orthogonality condition to use the maximum bound instead of the triangle inequality for $\|M_{p\circ g}\|$, the remaining term $\|M_{\chi_i\circ g}\|$ (where χ_i is a "hard" character) is still not easy to upper bound. For block composed functions, fortunately, the tensor structure makes it feasible to compute. Section 5.4 gives a generalized version of Theorem 6.

5.1 General framework

For a multiset T, $x \in T$ means x running over T. Thus $T = \{a(s) : s \in S\}$ means the multiset formed by collecting $a(s)$ with s running over S.

For a set S, denote by $L_{\mathbb{C}}(S)$ the $|S|$-dimensional vector space over the field $\mathbb{C}$ (of complex numbers) consisting of all linear functions from S to $\mathbb{C}$, endowed with inner product $\langle \psi, \phi \rangle = \frac{1}{|S|}\sum_{s\in S} \psi(s)\overline{\phi(s)}$. The distance of a function $f \in L_{\mathbb{C}}(S)$ to a subspace Φ of $L_{\mathbb{C}}(S)$, denoted by $d(f, \Phi)$, is defined as $\min\{\delta : \|f' - f\|_\infty \le \delta, f' \in \Phi\}$, *i.e.* the magnitude of the least entrywise perturbation to turn f into Φ.

In the above setting, Theorem 6 generalizes to the following. The proof is along the line of that of Theorem 6 and is omitted in this conference version.

Theorem 8. *Consider a sign matrix $A = [f(g(x, y))]_{x,y}$ where $g : X \times Y \to S$ for a set S, and $f : S \to \{-1, +1\}$. Suppose that there are orthogonal basis functions $\Psi = \{\psi_i : i \in [|S|]\}$ for $L_{\mathbb{C}}(S)$. For any partition $\Psi = \Psi_{Hard} \uplus \Psi_{Easy}$, let $\delta = d(f, span(\Psi_{Easy}))$. If*

1. **(regularity)** *the multiset $\{g(x, y) : x \in X, y \in Y\}$ is a multiple of S, i.e. S repeated for some number of times.*
2. **(orthogonality)** *for all x, x', y, y' and all distinct $\psi_i, \psi_j \in \Psi_{Hard}$,*

$$\sum_y \psi_i(g(x, y))\overline{\psi_j(g(x', y))} = \sum_x \psi_i(g(x, y))\overline{\psi_j(g(x, y'))} = 0,$$

then

$$Q_\epsilon(A) \ge \log_2 \frac{\sqrt{MN}\cdot(\delta - 2\epsilon)}{\max_{\psi_i\in\Psi_{Hard}}(\max_g |\psi_i(g)| \cdot \|[\psi_i(g(x, y))]_{x,y}\|)} - O(1).$$

In the Boolean block composed function case, the regularity condition reduces to the matrix $[g(x, y)]$ being balanced, and later we will prove that the orthogonality condition reduces to the strongly balanced property.

5.2 Functions with group symmetry

For a general finite group G, two elements s and t are conjugate, denoted by $s \sim t$, if there exists an element $r \in G$ *s.t.* $rsr^{-1} = t$. Define H as the set of all class functions, *i.e.* functions f *s.t.* $f(s) = f(t)$ if $s \sim t$. Then H is an h-dimensional subspace of $L_{\mathbb{C}}(G)$, where h is the number of conjugacy classes. The irreducible characters $\{\chi_i : i \in [h]\}$ form an orthogonal basis of H. For a class function f and irreducible characters χ_i, let $\hat{f}_i = \langle \chi_i, f \rangle = \frac{1}{|G|}\sum_{g\in G} \chi_i(g)\overline{f(g)}$. An easy fact is that

$$|\hat{f}_i| = \frac{1}{|G|}\left|\sum_{g\in G} \chi_i(g)\overline{f(g)}\right| \le \frac{1}{|G|}\sum_{g\in G} |f(g)||\chi_i(g)| \le \Big(\frac{1}{|G|}\sum_{g\in G} |f(g)|\Big)\cdot \max_g |\chi_i(g)|. \quad (1)$$

In this section we consider the setting that S is a finite group G. In particular, we hope to have a better understanding of the orthogonality condition and the matrix operator norm $\|[\psi_(g(x,y))]_{x,y}\|$ in this setting.

The standard orthogonality of irreducible characters says that $\sum_{s\in G}\chi_i(s)\overline{\chi_j(s)}=0$. The second condition in Theorem 8 is concerned with a more general case: For a multiset T with elements in $G\times G$, we need

$$\sum_{(s,t)\in T}\chi_i(s)\overline{\chi_j(t)}=0, \qquad \forall i\neq j. \tag{2}$$

The standard orthogonality relation corresponds to the special that $T=\{(s,s):s\in G\}$. We hope to have a characterization of a multiset T to make Eq. (2) hold.

We may think of the a multiset T with elements in set S as a function on S, with the value on $s\in S$ being the multiplicity of s in T. Since characters are class functions, for each pair (C_k,C_l) of conjugacy classes, only the value $\sum_{g_1\in C_k,t\in C_l}T(g_1,t)$ matters for the sake of Eq. (2). We thus make T a class function by taking average within each class pair (C_k,C_l). That is, define a new function T' as

$$T'(s,t)=\sum_{s\in C_k,t\in C_l}T(s,t)/(|C_k||C_l|),\ \forall s\in C_k,\ \forall t\in C_l.$$

Proposition 2. *For a finite group G and a multiset T with elements in $G\times G$, the following three statements are equivalent:*

1. $\sum_{(s,t)\in T}\chi_i(s)\overline{\chi_j(t)}=0,\ \forall i\neq j$
2. *T', as a function, is in span$\{\chi_i\otimes\overline{\chi_i}:i\in[h]\}$*
3. *$[T'(s,t)]_{s,t}=C^\dagger DC$ where D is a diagonal matrix and $C=[\chi_i(s)]_{i,s}$. That is, T', as a matrix, is normal and diagonalized exactly by the irreducible characters.*

5.3 Abelian group

When G is Abelian, we have further properties to use. The first one is that $|\chi_i(g)|=1$ for all i. The second one is that the irreducible characters are homomorphisms of G; that is, $\chi_i(st)=\chi_i(s)\chi_i(t)$. This gives a clean characterization of the orthogonality condition by group invariance. For a multiset T, denote by sT another multiset obtained by collecting all st where t runs over T. A multiset T with elements in $G\times G$ is *G-invariant* if it satisfies $(g,g)T=T$ for all $g\in G$. We can also call a function $T:G\times G\to\mathbb{C}$ G-invariant if $T(s,t)=T(rs,rt)$ for all $r,s,t\in G$. The overloading of the name is consistent when we view a multiset T as a function (counting the multiplicity of elements).

Proposition 3. *For a finite Abelian group G and a multiset T with elements in $G\times G$,*

$$T \text{ is } G\text{-invariant} \Leftrightarrow \sum_{(s,t)\in T}\chi_i(s)\overline{\chi_j(t)}=0, \quad \forall i\neq j. \tag{3}$$

Another nice property of Abelian groups is that the orthogonality condition condition implies the regularity condition; see the full version for a proof. So what we finally get for Abelian groups is the following.

Corollary 3. *For a sign matrix $A=[f(g(x,y))]_{x,y}$ and an Abelian group G, if $d(f,span(Ch_{Easy}))=\Omega(1)$, and the multisets $S^{x,x'}=\{(g(x,y),g(x',y)):y\in Y\}$ and $T^{y,y'}=\{(g(x,y),g(x,y')):x\in X\}$ are G-invariant for any (x,x') and any (y,y'), then*

$$Q(A)\geq\log_2\frac{\sqrt{MN}}{\max_{i\in Hard}\|[\chi_i(g(x,y))]_{x,y}\|}-O(1).$$

5.4 Block composed functions

We now consider a special class of functions g: block composed functions. Suppose the group G is a product group $G = G_1 \times \cdots \times G_t$, and $g(x, y) = (g_1(x^1, y^1), \cdots, g_t(x^t, y^t))$ where $x = (x^1, \cdots, x^t)$ and $y = (y^1, \cdots, y^t)$. That is, both x and y are decomposed into t components and the i-th coordinate of $g(x, y)$ only depends on the i-th components of x and y. The tensor structure makes all the computation easy. Theorem 6 can be generalized to the general product group case for arbitrary groups G_i.

Definition 6. *The ϵ-approximate degree of a class function f on product group $G_1 \times \cdots \times G_t$, denoted by $d_\epsilon(f)$, is the minimum d s.t. $\|f - f'\|_\infty \leq \epsilon$, where f' can be represented as a linear combination of irreducible characters with at most d non-identity component characters.*

Theorem 9. *For sign matrix $A = [f(g_1(x^1, y^1), \cdots, g_t(x^t, y^t)]_{x,y}$ where all g_i satisfy their orthogonality conditions, we have*

$$Q(A) \geq \min_{\{\chi_i\},S} \sum_{i \in S} \log_2 \frac{\sqrt{\text{size}(M_{g_i})}}{\deg(\chi_i)\|M_{\chi_i \circ g_i}\|} - O(1)$$

where the minimum is over all $S \subseteq [n]$ with $|S| > \deg_{1/3}(f)$, and all non-identity irreducible characters χ_i of G_i.

Previous sections, and [SZ09b], consider the case where all g_i's are the same and all G_i's are $\mathbb{Z}_2$. In this case, the above bound is equal to the one in Theorem 6, and the following proposition says that the group invariance condition degenerates to the strongly balanced property.

Proposition 4. *For $G = \mathbb{Z}_2^{\times t}$, the following two conditions for $g = (g_1, \cdots, g_t)$ are equivalent:*

1. *The multisets $S^{x,x'} = \{(g(x, y), g(x', y)) : y \in Y\}$ and $T^{y,y'} = \{(g(x, y), g(x, y')) : x \in X\}$ are G-invariant for any (x, x') and any (y, y'),*
2. *Each matrix $[g_i(x^i, y^i)]_{x^i,y^i}$ is strongly balanced.*

This equivalence does not in general hold if any group G_i has size larger than two.

Acknowledgments

We would like to thank Adi Shraibman for many enlightening conversations about these topics. TL is supported in part by a NSF postdoctoral research fellowship and by the grants CCF-0728937 and CCF-0832787. SZ is supported in part by Hong Kong grant RGC-419309.

References

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th ACM Symposium on the Theory of Computing*, pages 63–68, 1998.

[BHN09] P. Beame and D. Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of AC^0. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, pages 53–62, 2009.

[BNS92] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for Logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45:204–232, 1992.

[BW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288:21–43, 2002.

[CA08] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[Cha07] A. Chattopadhyay. Discrepancy and the power of bottom fan-in depth-three circuits. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 449–458, 2007.

[Cha08] A. Chattopadhyay. *Circuits, Communication, and Polynomials*. PhD thesis, McGill University, 2008.

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

[LMSS07] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[LS88] L. Lovász and M. Saks. Möbius functions and communication complexity. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 81–90, 1988.

[LS09a] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[LS09b] T. Lee and A. Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3, 2009.

[LS09c] N. Linial and A. Shraibman. Learning complexity versus communication complexity. *Combinatorics, Probability, and Computing*, 18:227–245, 2009.

[LS09d] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 34:368–394, 2009.

[LSŠ08] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 71–80. IEEE, 2008.

[Nis94] Noam Nisan. The communication complexity of threshold gates. In *In Proceedings of Combinatorics, Paul Erdos is Eighty*, pages 301–315, 1994.

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[Raz99] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the 31st ACM Symposium on the Theory of Computing*, pages 358–367, 1999.

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

[RS08] A. Razborov and A. Sherstov. The sign rank of AC^0. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 57–66, 2008.

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

[She07] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*, pages 294–301. ACM, 2007.

[She08a] A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[She08b] A. Sherstov. The unbounded-error communication complexity of symmetric functions. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008.

[She09] A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 2009.

[She10] A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Information and Computation*, 10(5-6):435–455, 2010.

[SZ09a] Y. Shi and Z. Zhang. Communication complexities of XOR functions. *Quantum information and computation*, 9(3-4):255–263, 2009.

[SZ09b] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum information and computation*, 9(5,6):444–460, 2009.

[Yao79] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pages 209–213, 1979.

[Zha09] S. Zhang. On the tightness of the Buhrman-Cleve-Wigderson simulation. In *Proceedings of the 20th International Symposium on Algorithms and Computation*, pages 434–440, 2009.