Quantum One-Way Communication is Exponentially Stronger Than Classical Communication

Bo'az Klartag* Oded Regev†

Abstract

In STOC 1999, Raz presented a (partial) function for which there is a quantum protocol communicating only $O(\log n)$ qubits, but for which any classical (randomized, bounded-error) protocol requires $\mathrm{poly}(n)$ bits of communication. That quantum protocol requires two rounds of communication. Ever since Raz's paper it was open whether the same exponential separation can be achieved with a quantum protocol that uses only one round of communication. Here we settle this question in the affirmative.

1 Introduction

Communication complexity is one of the most basic models in computational complexity, with wide-ranging applications in computer science [KN97]. The typical question asked in this model is the following. Two remote players, call them Alice and Bob, are each given an input and are trying to compute some function of their inputs while using as little communication as possible. How much communication is needed in order to compute the function? The answer to this question often depends on what exactly we mean by "compute using as little communication as possible." One of the central models in this area is that of *randomized (bounded-error) communication*. Here we allow the players to toss coins, and require them to output the correct answer with probability at least (say) $2/3$ on *any* given input. This model is quite powerful and corresponds quite well to what is actually achievable in real-world communication. For instance, one of the most basic results in this area shows that the players can decide if their inputs are equal using only $O(\log n)$ bits of communication, where n is the size of their inputs in bits. Another well-established model of communication is that of *quantum communication* [Yao93]. Here, we allow the players to communicate quantum states, and to perform quantum operations on them. Although not nearly as common as classical (i.e., non-quantum) communication, this model is able to provide important insights into the power of quantum mechanics.

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Supported in part by the Israel Science Foundation and by a Marie Curie Reintegration Grant from the Commission of the European Communities.

†Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Supported by the Israel Science Foundation, by the Wolfson Family Charitable Trust, and by a European Research Council (ERC) Starting Grant. Part of the work done while a DIGITEO visitor in LRI, Orsay.

The focus of our work is on the relative power of these two central models, a question whose study started in the late 1990s [BCW98, AST+98]. Most notably, Raz [Raz99] presented a (partial) function for which there is a quantum protocol communicating only $O(\log n)$ qubits, but for which any classical (randomized, bounded-error) protocol requires $\text{poly}(n)$ bits of communication (i.e., $\Omega(n^c)$ for some constant $c > 0$). This result demonstrates that quantum communication is exponentially stronger than classical communication, and is one of the most fundamental results in quantum communication complexity.

However, although Raz's function can be computed using only $O(\log n)$ qubits, it seems to require at least two rounds of communication between Alice and Bob. This naturally leads to the following fundamental question, which has been open ever since Raz's paper: can a similar exponential separation be achieved with a quantum protocol that uses *only one round of communication*? In other words:

> *Can quantum one-way communication be exponentially stronger than classical two-way communication in computing a function*?

Such a result might be the strongest possible separation between quantum communication and classical communication.

There have been quite a few partial results in this direction. First, Bar-Yossef, Jayram, and Kerenidis [BJK04] presented a *relational problem* (i.e., one in which there is possibly more than one correct answer for a given input) that has a quantum one-way protocol using only $O(\log n)$ qubits of communication, but for which any classical protocol *using only one round of communication* must communicate $\text{poly}(n)$ bits. Classical two-way protocols, however, can easily solve their problem using $O(\log n)$ bits. Their result was improved by Gavinsky, Kempe, Kerenidis, Raz, and de Wolf [GKK+07] who proved the same separation, namely, $O(\log n)$ qubit protocol versus a $\text{poly}(n)$ lower bound for any classical one-way protocol, but in the standard setting of a *functional problem*. Again, classical two-way protocols can easily solve the problem using only $O(\log n)$ bits. See also [Mon10] for a similar separation. Another closely related result is by Gavinsky [Gav08], who improved on Bar-Yossef et al.'s [BJK04] result in the other direction: namely, he showed an exponential separation between one-way quantum communication and *two-way classical communication* (just as in the open question) but for a *relational problem*. Gavinsky's proof is quite involved, and it is not clear if his techniques can be used to attack the functional case.

It is important to note that there is a big difference between relational separations and functional ones, with the latter often being more interesting, involving deeper ideas, and having more profound implications. Indeed, the functional separation in [GKK+07] required the use of a hypercontractive inequality and also provided a surprising counterexample to a conjecture regarding extractors that are secure against quantum adversaries. Moreover, the existence of a relational separation often says little about the existence of a functional one; for instance, there are cases where relational separations provably have no functional counterpart [GRW08].

Here we settle the open question by exhibiting a (partial) function for which there exists a quantum one-way communication protocol using only $O(\log n)$ qubits, but for which any classical two-way communication protocol must communicate at least $\text{poly}(n)$ bits. The function we consider is actually the complete problem for one-way quantum communication [Kre95] and was

also described in [Raz99]. We call it the *Vector in Subspace Problem* (VSP). In this problem, Alice is given an n-dimensional unit vector $u \in S^{n-1}$ and Bob is given a subspace $H \subset \mathbb{R}^n$ of dimension $n/2$ with the promise that either $u \in H$ or $u \in H^\perp$. Their goal is to decide which is the case. (For a formal definition see Section 4.) The quantum protocol for the problem is almost immediate from the definition: Alice encodes the vector u as a quantum state of $\lceil \log_2 n \rceil$ qubits (by definition, the state of a quantum system with k qubits is a 2^k-dimensional unit vector) and sends it to Bob, who, after having received the quantum state, performs the projective measurement given by $(H, H^\perp)$. If $u \in H$, Bob is guaranteed to obtain the former outcome; if $u \in H^\perp$, Bob is guaranteed to obtain the latter outcome.

It is easy to see that VSP has a classical protocol using $O(n \log n)$ bits: Alice simply sends the vector u to Bob, by specifying each coordinate to within an additive $\pm 1/\mathrm{poly}(n)$ accuracy. As noted by Raz [Raz99], this protocol is not optimal, and the problem actually has an $O(\sqrt{n})$ protocol, which we will describe in Section 4.

But of course, our focus in this paper is on *lower bounds*. Our main result is an $\Omega(n^{1/3})$ lower bound on the (classical) communication complexity of VSP. Previously no lower bound better than logarithmic was known. Our proof involves some techniques that seem novel in the computer science literature. We use a hypercontractive inequality, applied in a fashion similar to that in Kahn, Kalai, and Linial [KKL88] and in other more recent papers, including the result by Gavinsky et al. mentioned above [GKK$^+$07] (see also [Wol08]). However, unlike previous work, our hypercontractive inequality is in the setting of functions defined on the sphere. We also use the Radon transform and some of its basic properties, as well as a rather delicate martingale argument. Finally, we feel that the proof, at least at a very high level, is conceptually simpler than some of the previous proofs in this line of work. We hope that our result and techniques will find other applications.

One obvious open question left by our work is to improve the lower bound to a tight $\Omega(n^{1/2})$; we will mention one possible approach below. Another open question is to strengthen our result by showing a separation between the quantum simultaneous message passing (SMP) model and the classical two-way model. This question was recently answered by Gavinsky [Gav09] for relational problems, but the question for functions seems quite challenging, and it is not even clear if such a separation can exist. A final important open question is to understand the power of quantum communication in computing *total* functions; so far the best known separation is polynomial.

2 Proof Sketch

Here we give an informal sketch of the main ideas in the proof of our lower bound, and include some remarks regarding the tightness and other aspects of our proofs. The proof starts in Section 4 with a more or less standard application of the rectangle bound which we do not describe here. This shows that in order to prove our communication lower bound, it suffices to prove the following sampling statement, which is our main technical theorem (see Figure 1 for an illustration). The formal statement will appear as Theorem 6.1.

Theorem 2.1 (Informal)**.** *Let A be an arbitrary (measurable) subset of the sphere S^{n-1} whose measure*

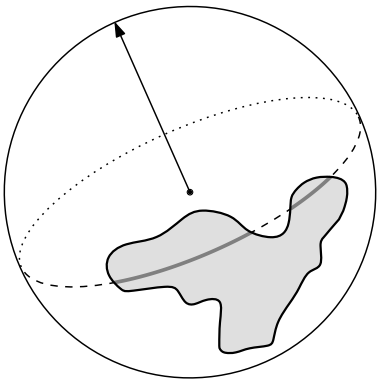

Figure 1: A subset of S^2 and an equator

$\sigma(A)$ (under the uniform probability measure on S^{n-1}) is at least $\exp(-n^{1/3})$. Assume we choose a uniformly random subspace $H \subset \mathbb{R}^n$ of dimension $n/2$. Consider the measure of the set $A \cap H$ under the uniform probability measure on the unit sphere $H \cap S^{n-1}$ of the $n/2$-dimensional subspace H. Then this measure is within a factor of (say) 1 ± 0.1 of $\sigma(A)$ except with probability at most $\exp(-n^{1/3})$.

Before we proceed to discuss the proof of this theorem, we make two remarks. First, it is interesting to note that this theorem is a considerable strengthening of Lemma 4.1 in [Raz99], which is a similar sampling statement, but one that applies only to sets A whose measure is constant (or slightly less). Raz proves that lemma using an elementary (but clever) use of Chernoff's concentration bound. See also the paper by Milman and Wagner [MW03] for a further discussion and applications of Raz's sampling lemma.

The second remark is that our theorem is tight in the sense that there exists a set A of measure $\exp(-n^{1/3})$ such that the probability of the measure of $A \cap H$ deviating by more than 10% is essentially $\exp(-n^{1/3})$. This set A is simply a spherical cap, and the bad H's are those that are close to the center of the cap. We omit this standard calculation. One implication of this is that improving our $\Omega(n^{1/3})$ lower bound to a tight $\Omega(n^{1/2})$ is probably impossible using the rectangle bound, and one might have to use instead the *smooth rectangle bound* introduced in [Kla10, JK10] and used recently in [CR10]. For the interested reader, we note that the following reasonable sampling statement would imply the tight $\Omega(n^{1/2})$ bound. Let A be an arbitrary subset of the sphere S^{n-1} whose measure $\sigma(A)$ is at least $\exp(-n^{1/2})$, and assume we choose a uniformly random subspace $H \subset \mathbb{R}^n$ of dimension $n/2$. We now consider the measure of the set $A \cap H$ and that of the set $A \cap H^\perp$ (under the appropriate uniform probability measures). Then the goal would be to prove that the *average* of these two measures is at least $0.9\,\sigma(A)$ except with probability at most $\exp(-n^{1/2})$.

Theorem 2.1 is proven by a recursive application of the following core sampling statement for $(n-1)$-dimensional subspaces. Roughly speaking, it shows that sampling a set of measure at least $\exp(-n^{1/3})$ using a random $(n-1)$-dimensional subspace gives an error that is typically at most $1 \pm n^{-2/3}$ and has an exponential decay. The formal statement will appear as Theorem 5.1.

Theorem 2.2 (Informal)**.** *Let $A \subset S^{n-1}$ be of measure at least $\exp(-n^{1/3})$. Assume we choose a uniformly random subspace $H \subset \mathbb{R}^n$ of dimension $n-1$. Then, for any $0 < t < 1$, the measure of $A \cap H$*

(under the uniform measure on $H \cap S^{n-1}$) is within a factor of $1 \pm t$ of $\sigma(A)$ except with probability at most $\exp(-n^{2/3}t)$.

Section 6 will be dedicated to deriving Theorem 2.1 from the above theorem. This is done using a martingale argument and Bernstein-type inequalities; in the following we just give the rough idea. Consider the following equivalent way to choose a uniformly random subspace H of dimension $n/2$. First, let $H_0 = \mathbb{R}^n$. Then, choose a uniformly random subspace $H_1 \subset H_0 = \mathbb{R}^n$ of dimension $n-1$; then, choose a uniformly random subspace H_2 of H_1 of dimension $n-2$; continue in the same fashion until $H = H_{n/2}$ which is a uniformly random $n/2$-dimensional subspace of $H_{n/2-1}$. We now consider the sequence of measures of $A \cap H_i$ (with respect to the uniform measure in $S^{n-1} \cap H_i$) for $i = 0, \ldots, n/2$. By definition, this sequence starts with $\sigma(A)$. According to Theorem 2.2, at each step of the sequence we typically get an extra multiplicative error of $1 \pm n^{-2/3}$. After $n/2$ steps, the accumulated error becomes $1 \pm \sqrt{n} \cdot n^{-2/3} = 1 \pm n^{-1/6}$ (this of course requires a proof since, e.g., the steps are not independent). Hence, assuming the error has a Gaussian tail (which is also far from obvious), and recalling that the probability that a Gaussian variable deviates by more than t standard deviations is roughly $\exp(-t^2)$, we obtain that the probability of seeing a total deviation of more than 1 ± 0.1 is at most $\exp(-n^{1/3})$, as required.

We remark that we also have an alternative and direct proof of Theorem 2.1 that is similar in nature to the proof of Theorem 2.2 (as described below), except it uses the Grassmannian manifold; this proof, unfortunately, currently leads to a worse bound of $\exp(-n^{1/4})$ (instead of $\exp(-n^{1/3})$) and is therefore omitted. It is quite possible that this direct proof can be improved to obtain the tight $\exp(-n^{1/3})$ bound.

The proof of Theorem 2.2 will be given in Section 5. It uses the hypercontractive inequality for the sphere, applied in a fashion similar to the one done by Kahn, Kalai, and Linial [KKL88], as well as some basic properties of the Radon transform. In order to demonstrate these ideas in a setting that might be more familiar to some readers, we spend the remainder of this section on proving an analogous statement in the setting of the Boolean hypercube $\{0,1\}^n$, and for simplicity just consider the case $t = n^{-1/3}$ (the general case is similar).

Sampling statement for the Boolean cube. Let n be an even integer. For a vector $y \in \{0,1\}^n$ define $y^\perp = \{z \in \{0,1\}^n; \mathrm{HamDist}(y,z) = n/2\}$ as the "equator orthogonal to y". Let $A \subseteq \{0,1\}^n$ be of measure $\mu(A) := |A|/2^n$ at least $\exp(-n^{1/3})$. Assume we choose a uniform $y \in \{0,1\}^n$, and consider the fraction of points in $y^\perp$ that are contained in A. Then our goal is to show that this fraction is in $(1 \pm n^{-1/3})\mu(A)$ except with probability at most $\exp(-n^{1/3})$.

As stated, this statement is actually *false* due to a parity issue; this can be seen, e.g., by taking A to be all points of even Hamming weight, a set of measure $1/2$. Then the fraction of points in $y^\perp$ that are contained in A is either 0 or 1 depending on the parity of y. Although the statement can be easily mended, in the sequel we ignore this issue and proceed with an incomplete proof of the original incorrect statement. We allow ourselves to do this because this parity issue does not arise in the setting of the sphere, and the argument below becomes a valid proof there (with the necessary modifications, of course).

The above sampling statement can be stated in the following essentially equivalent way. For

any $A, B \subseteq \{0,1\}^n$ of measure at least $\exp(-n^{1/3})$,

$$\mathop{\mathbb{P}}_{y\sim B, x\sim y^\perp}[x \in A] \in (1 \pm n^{-1/3})\mu(A), \tag{1}$$

where the notation $x \sim E$ means that x is distributed uniformly in the set E, and the right hand side indicates the interval $[(1-n^{-1/3})\mu(A), (1+n^{-1/3})\mu(A)]$. For a function $f : \{0,1\}^n \to \mathbb{R}$, define its *Radon transform* $R(f) : \{0,1\}^n \to \mathbb{R}$ as

$$R(f)(y) := \mathbb{E}_{x\sim y^\perp}[f(x)].$$

Define $f = 1_A/\mu(A)$ and $g = 1_B/\mu(B)$ to be the indicator functions of A and B normalized so that their expectations over a uniform input are $\mathbb{E}_x[f(x)] = \mathbb{E}_x[g(x)] = 1$. With this notation, Eq. (1) becomes

$$\langle f, R(g)\rangle = \mathbb{E}_x[f(x)R(g)(x)] \in 1 \pm n^{-1/3}. \tag{2}$$

For a function $f : \{0,1\}^n \to \mathbb{R}$, define its Fourier transform $\hat{f} : \{0,1\}^n \to \mathbb{R}$ by $\hat{f}(w) := \mathbb{E}_x[(-1)^{w\cdot x} f(x)]$. Then by the orthogonality of the Fourier transform, Eq. (2) can be written equivalently as

$$\sum_w \hat{f}(w)\widehat{R(g)}(w) \in 1 \pm n^{-1/3}.$$

An easy direct calculation reveals that R is diagonal in the Fourier basis. (Alternatively, one can use Schur's lemma and the fact that R commutes with translations.) This calculation also reveals that the eigenvalue corresponding to $w \in \{0,1\}^n$ is 0 whenever the Hamming weight of w is odd, 1 when the Hamming weight of w is 0,

$$\frac{\binom{n-2}{n/2} - 2\binom{n-2}{n/2-1} + \binom{n-2}{n/2-2}}{\binom{n}{n/2}} \approx -\frac{1}{n}$$

when w is of Hamming weight 2, approximately $\frac{1}{n^2}$ when w is of Hamming weight 4, etc. We can therefore write

$$\sum_w \hat{f}(w)\widehat{R(g)}(w) \approx \hat{f}(0)\hat{g}(0) - \frac{1}{n}\sum_{|w|=2}\hat{f}(w)\hat{g}(w) + \frac{1}{n^2}\sum_{|w|=4}\hat{f}(w)\hat{g}(w) - \cdots.$$

The first term is $\hat{f}(0)\hat{g}(0) = \mathbb{E}_x[f(x)]\mathbb{E}_x[g(x)] = 1$. Hence our goal is to bound the remaining terms by $n^{-1/3}$. For simplicity, let us focus on the first term, and show that $\sum_{|w|=2}\hat{f}(w)\hat{g}(w)$ is at most $n^{2/3}$ in absolute value; one can similarly analyze the remaining terms and show that their total contribution is similar.[1] By using the Cauchy-Schwarz inequality, we can bound this sum by $(\sum_{|w|=2}\hat{f}(w)^2)^{1/2}(\sum_{|w|=2}\hat{g}(w)^2)^{1/2}$. The following lemma now completes the proof.

Lemma 2.3. *Let $A \subseteq \{0,1\}^n$ be of measure μ, and let $f = 1_A/\mu(A)$ be its (normalized) indicator function. Then, for some universal constant $C > 0$,*

$$\sum_{|w|=2}\hat{f}(w)^2 \le C(\log(1/\mu))^2.$$

[1] This is where we are cheating: the term $|w| = n$ can contribute a lot to this sum.

Equivalently, the lemma says that if $X = (x_1, \ldots, x_n)$ is uniformly chosen from A, then the sum over all pairs $\{i, j\}$ of the bias squared of $x_i \oplus x_j$ is at most $C(\log(1/\mu))^2$. This can be seen to be essentially tight by taking, e.g., $A = \{x \in \{0,1\}^n; x_1 = \cdots = x_{\log_2 1/\mu} = 0\}$. This lemma is proven by applying the Bonami-Gross-Beckner hypercontractive inequality [Bon70, Gro75, Bec75] in a way similar to that in [KKL88]. Essentially the exact same lemma appears in [GKK+07], and is also described in detail in the survey [Wol08]. We include a sketch of the proof, as later on we will have a similar proof in the spherical setting (in Lemma 5.3).

Proof. The hypercontractive inequality for the Boolean cube states that for any $f : \{0,1\}^n \to \mathbb{R}$, and $1 \leq p \leq 2$,

$$\left\| T_{\sqrt{p-1}} f \right\|_2 \leq \|f\|_p$$

where T_ρ is the noise operator with parameter ρ (which is the operator that is diagonal in the Fourier basis, and has eigenvalue ρ^k for each Fourier basis function of level k), and the pth norm is defined as $\|f\|_p = \mathbb{E}_x[|f(x)|^p]^{1/p}$. By plugging in our f we obtain

$$\begin{aligned}
\sum_{|w|=2} \hat{f}(w)^2 &\leq \frac{1}{(p-1)^2} \sum_w (p-1)^{|w|} \hat{f}(w)^2 \\
&= \frac{1}{(p-1)^2} \left\| T_{\sqrt{p-1}} f \right\|_2^2 \\
&\leq \frac{1}{(p-1)^2} \|f\|_p^2 = \frac{1}{(p-1)^2} \mu^{-2(1-1/p)}.
\end{aligned}$$

The lemma follows by optimizing over $1 \leq p \leq 2$. □

3 Preliminaries

General. Throughout the paper, by "measurable" we mean Borel measurable. All logarithms are natural logarithms unless otherwise specified. We adopt the following convention for denoting constants. The letters $c, \tilde{c}, C, \tilde{C}$, etc. stand for various positive universal constants, whose value may change from one line to the next. We usually use upper-case C to denote universal constants that we think of as "sufficiently large", and lower-case c to denote universal constants that are "sufficiently small".

Some manifolds and uniform distributions on them. Write $S^{n-1} = \{x \in \mathbb{R}^n; |x| = 1\}$ for the unit sphere in $\mathbb{R}^n$. We denote by σ the uniform probability measure on S^{n-1}, i.e., the unique rotationally-invariant probability measure on S^{n-1} (see, e.g., [MS86, Chapter I] for more information on Haar measures). We denote by $\mathcal{G}_{n,m}$ the Grassmannian manifold, i.e., the manifold of all m-dimensional subspaces in $\mathbb{R}^n$, and we let $\sigma_{\mathcal{G}}$ be the uniform distribution over it (or, more formally, the unique rotationally-invariant probability measure on $\mathcal{G}_{n,m}$). We also consider the incidence manifold

$$\mathcal{I}_{n,m} = \left\{ (x, H) \in S^{n-1} \times \mathcal{G}_{n,n-m} \,;\, x \in H \right\} \subset S^{n-1} \times \mathcal{G}_{n,n-m},$$

and let $\sigma_{\mathcal{I}}$ be the uniform probability measure on it (or more precisely, the unique rotationally-invariant probability measure on $\mathcal{I}_{n,m}$). We will implicitly use some basic properties of these manifolds and the uniform distributions on them; for a rigorous discussion of the topic, see, e.g., Helgason [Hel99, Chapter II].

4 Communication Complexity

In this section we give a formal definition of the VSP problem, and derive the main lower bound from the sampling statement. Our discussion in this section closely follows Raz's [Raz99], hence we will occasionally allow ourselves to be brief. We also assume some basic familiarity with randomized communication complexity [KN97].

We start with the formal definition of VSP. This is identical to the $\mathcal{P}_0$ problem defined in [Raz99].

Definition 4.1. Let $0 \leq \vartheta < 1/\sqrt{2}$ be a parameter. In the VSP_ϑ problem, Alice is given an n-dimensional unit vector $u \in S^{n-1}$ and Bob is given a subspace $H \subset \mathbb{R}^n$ of dimension $n/2$. They are promised that either the distance of u from H is at most ϑ or the distance of u from $H^\perp$ is at most ϑ. Their goal is to decide which is the case.

This problem was first defined by Kremer [Kre95] and was shown to be a complete problem for one-round quantum communication complexity. In particular, for any $0 \leq \vartheta < 1/\sqrt{2}$, VSP_ϑ has an (almost immediate) quantum protocol communicating only $O(\log n)$ qubits in a single message from Alice to Bob. (Moreover, there is a matching $\Omega(\log n)$ lower bound.)

In terms of its classical (randomized, bounded-error) communication complexity, Raz [Raz99] has shown that the problem has an $O(\sqrt{n})$ communication protocol, which we now briefly describe. Assume Alice and Bob use their shared randomness to pick a sequence of unit vectors chosen uniformly from S^{n-1}, $v_1, v_2, \ldots$. Alice looks for the vector v_i with the maximal inner product $v_i \cdot u$ among the first $2^{C\sqrt{n}}$ unit vectors, and sends the index i to Bob, who decides on the output based on which of H and $H^\perp$ is closer to v_i. The protocol clearly requires only $O(\sqrt{n})$ bits of communication, and moreover, the output produced by Bob is correct with high probability (essentially since the projection squared of v_i on H (or $H^\perp$) gets an addition of $n^{-1/2}$ due to the high inner product with u, which is sufficient to noticeably affect Bob's answer since the standard deviation of the projection squared is of order $n^{-1/2}$). Using Newman's theorem, the shared randomness can be replaced with private randomness by only communicating an extra $O(\log n)$ bits (which is negligible). For a more detailed proof, see Theorem 3.8 in [Raz99].

However, no lower bound better than logarithmic was previously known. Our main result is an $\Omega(n^{1/3})$ lower bound on the randomized communication complexity of the problem VSP_0 (which is the problem described in the introduction). One minor caveat here is that this lower bound holds only for protocols that are "measurable," in the sense that the functions describing the behavior of the players need to be measurable. Clearly, increasing ϑ can only make the problem harder, hence our lower bound also apply to any $0 < \vartheta < 1/\sqrt{2}$. Moreover, as we shall see below, there is no need to assume measurability in the case $\vartheta > 0$.

Another point to note is that the number of possible inputs to VSP is infinite. Although there is nothing terribly wrong with this, in the standard communication complexity model problems are

supposed to have inputs that are taken from a finite set. This can be easily achieved by specifying the inputs using an n-dimensional vector (for Alice) together with an $n \times n/2$ matrix (for Bob) each of whose entries is described by $O(\log n)$ bits. We denote this problem by $\widetilde{\mathsf{VSP}}$. Since this is a restriction of VSP, we clearly still have a one-way $O(\log n)$ qubit protocol. Next, notice that for any $0 < \vartheta < 1/\sqrt{2}$, we can convert any protocol for $\widetilde{\mathsf{VSP}}_\vartheta$ into a protocol for VSP_0 by simply rounding the coordinates of the inputs. Moreover, the resulting VSP_0 protocol is clearly measurable since its input space is partitioned into a finite number of simple sets, and the protocol's behavior is completely determined on each of these simple sets. We therefore obtain a lower bound of $\Omega(n^{1/3})$ on the randomized communication complexity of $\widetilde{\mathsf{VSP}}_\vartheta$ for any $0 < \vartheta < 1/\sqrt{2}$. Notice that the problem's input size is $m = O(n^2 \log n)$, and hence in terms of the input size, our lower bound is $\Omega((m/\log m)^{1/6})$. Finally, since $\widetilde{\mathsf{VSP}}_\vartheta$ is a restriction of VSP_ϑ, we also obtain a lower bound of $\Omega(n^{1/3})$ on the randomized communication complexity of VSP_ϑ for any $0 < \vartheta < 1/\sqrt{2}$, without the measurability assumption. We summarize this discussion in the following theorem, which we then proceed to prove.

Theorem 4.2. *Any measurable randomized (bounded-error) protocol for VSP_0 requires $\Omega(n^{1/3})$ bits of communication. As a result, we obtain that for all $0 < \vartheta < 1/\sqrt{2}$, the randomized communication complexity of both VSP_ϑ and $\widetilde{\mathsf{VSP}}_\vartheta$ is $\Omega(n^{1/3})$ (without any measurability assumption).*

Proof. As described above, it suffices to prove the lower bound on VSP_0. Fix an arbitrary randomized protocol communicating at most D bits, and assume that it solves VSP_0 with error probability at most $1/3$ on all legal inputs. (The argument applies to any error probability smaller than $1/2$ by a standard amplification technique.) Our goal is to lower bound D.

Recall the definition of $\mathcal{I}_{n,n/2}$ and the uniform distribution $\sigma_\mathcal{I}$ on it, given by a uniformly chosen subspace H of dimension $n/2$ and a uniformly chosen unit vector u in H. We also define the set $\bar{\mathcal{I}}_{n,n/2}$ as the set of all pairs $(x, H) \in S^{n-1} \times \mathcal{G}_{n,n/2}$ such that $x \in H^\perp$, and let $\bar{\sigma}_\mathcal{I}$ be the uniform distribution on it, given by a uniformly chosen subspace H of dimension $n/2$ and a uniformly chosen unit vector u in $H^\perp$.

We consider the following two quantities. The first is the probability that the protocol incorrectly outputs "u not in H" when the inputs are chosen from $\sigma_\mathcal{I}$. The second is the probability that the protocol incorrectly outputs "u in H" when the inputs are chosen from $\bar{\sigma}_\mathcal{I}$. By our assumption, each of these quantities is at most $1/3$, and hence their sum is at most $2/3$. By linearity there exists a way to fix the random string used by the protocol such that the resulting deterministic protocol also satisfies that the sum of these two quantities is at most $2/3$. From now on we consider that deterministic protocol.

As is well known, such a deterministic protocol induces a partition of $S^{n-1} \times \mathcal{G}_{n,n/2}$ into 2^D rectangles, i.e., measurable sets of the form $A \times B$ where $A \subseteq S^{n-1}$ and $B \subseteq \mathcal{G}_{n,n/2}$, where each rectangle is labelled with "in" or "not in", corresponding to the protocol's output on inputs from this rectangle. In order to analyze this partition, we use the following lemma, which follows easily from our main sampling theorem, as will be shown in Section 6.

Lemma 4.3. *Suppose that $A \subseteq S^{n-1}$ and $B \subseteq \mathcal{G}_{n,n/2}$ are measurable sets with*

$$\sigma(A) \geq C \exp(-cn^{1/3}), \qquad \sigma_\mathcal{G}(B) \geq C \exp(-cn^{1/3})$$

for some universal constants $c, C > 0$. Then,

$$\sigma_{\mathcal{I}}\left((A \times B) \cap \mathcal{I}_{n,n/2}\right) \geq 0.8\,\sigma(A)\sigma_{\mathcal{G}}(B).$$

As a result, we obtain that for *all* measurable sets $A \subseteq S^{n-1}$ and $B \subseteq \mathcal{G}_{n,n/2}$,

$$\sigma_{\mathcal{I}}\left((A \times B) \cap \mathcal{I}_{n,n/2}\right) \geq 0.8\,\sigma(A)\sigma_{\mathcal{G}}(B) - C\exp(-cn^{1/3}). \tag{3}$$

By simply replacing H with $H^{\perp}$ we also obtain that

$$\bar{\sigma}_{\mathcal{I}}\left((A \times B) \cap \bar{\mathcal{I}}_{n,n/2}\right) \geq 0.8\,\sigma(A)\sigma_{\mathcal{G}}(B) - C\exp(-cn^{1/3}). \tag{4}$$

We now sum the inequalities (3) over all rectangles $A \times B$ that are labelled with "not in" and the inequalities (4) over all rectangles labelled with "in". Our assumption above says precisely that the left hand side is at most $2/3$. The right hand side is exactly $0.8 - 2^D \cdot C\exp(-cn^{1/3})$. Rearranging, we obtain that $2^D \geq c\exp(cn^{1/3})$, as required. □

5 Sampling Sets by Equators

In this section we prove one of the main components of our proof, namely, a sampling theorem using equators: we show that any (not too small) subset A of the sphere S^{n-1} is sampled well by a randomly chosen equator (where an equator is the intersection of S^{n-1} with an $(n-1)$-dimensional subspace). See Figure 1.

Theorem 5.1. *Let $A \subseteq S^{n-1}$ be a measurable set. Assume H is a uniformly chosen $(n-1)$-dimensional subspace. Then, for any $0 < t < 1$, the probability that*

$$\left|\frac{\sigma_H(A \cap H)}{\sigma(A)} - 1\right| \geq t$$

is at most $C\exp(-cnt/\log(2/\sigma(A)))$ for some universal constants $C, c > 0$, where σ_H denotes the uniform probability measure on the sphere $H \cap S^{n-1}$.

In the rest of this section, we actually prove the following more symmetric statement, from which Theorem 5.1 follows as described below. Here we denote by $\mathcal{V}_n$ the manifold of all pairs of orthogonal vectors,

$$\mathcal{V}_n = \left\{(x, y) \in S^{n-1} \times S^{n-1}\,;\, x \cdot y = 0\right\}$$

and we let $\sigma_{\mathcal{V}}$ denote the uniform probability measure over $\mathcal{V}_n$.

Theorem 5.2. *Suppose $f, g : S^{n-1} \to [0, \infty)$ are bounded measurable functions with $\int_{S^{n-1}} f\mathrm{d}\sigma = \int_{S^{n-1}} g\mathrm{d}\sigma = 1$ and set*

$$s = \log(2\|f\|_\infty) \cdot \log(2\|g\|_\infty).$$

Then, when $s \leq cn$,

$$\left|\int_{\mathcal{V}_n} f(x)g(y)\mathrm{d}\sigma_{\mathcal{V}}(x, y) - 1\right| \leq \frac{Cs}{n},$$

where $C, c > 0$ are universal constants.

In order to derive Theorem 5.1, let E be the set of all $y \in S^{n-1}$ for which the subspace $y^\perp \subset \mathbb{R}^n$ orthogonal to y satisfies

$$\frac{\sigma_{y^\perp}(A \cap y^\perp)}{\sigma(A)} \geq 1 + t.$$

Let $f = 1_A/\sigma(A)$ and $g = 1_E/\sigma(E)$ be the normalized indicator functions of A and B, respectively. Then it follows that

$$\int_{\mathcal{V}_n} f(x)g(y)\mathrm{d}\sigma_{\mathcal{V}}(x,y) \geq 1 + t$$

since the left hand side is exactly the average of $\sigma_{y^\perp}(A \cap y^\perp)/\sigma(A)$ over y chosen uniformly from E. Hence by Theorem 5.2,

$$t \leq \frac{C \log(2/\sigma(A)) \log(2/\sigma(E))}{n}.$$

Rearranging, we obtain that

$$\sigma(E) < C \exp(-cnt/\log(2/\sigma(A))).$$

Repeating a similar argument for the lower bound, we obtain Theorem 5.1.

Our proof of Theorem 5.2 resembles a small jigsaw puzzle, in which all of the pieces are known mathematical constructions that have to be put in place in order to yield a proof. Therefore most of this section is devoted to a brief summary of standard mathematical material, such as some basic features of spherical harmonics, the Laplacian, log-Sobolev inequalities, hypercontractivity, growth of L^p norms of eigenfunctions, and the Radon transform and its eigenvalues.

Spherical harmonics. We write $L^2(S^{n-1})$ for the space of all square-integrable functions on S^{n-1}. For $U \in SO(n)$ and $f \in L^2(S^{n-1})$ denote

$$U(f)(x) = f(U^{-1}x) \qquad (x \in S^{n-1}).$$

We say that $U(f)$ is the rotation of f by U. For any integer $k \geq 0$, there is a special finite-dimensional subspace $\mathcal{S}_k \subset L^2(S^{n-1})$ of smooth functions called the space of "spherical harmonics of degree k." For instance, $\mathcal{S}_0$ is the one-dimensional space of constant functions. More generally, $\mathcal{S}_k$ is defined as the restriction to the sphere of all harmonic, homogenous polynomials of degree k in $\mathbb{R}^n$. See, e.g., Müller [Mül66] or Stein and Weiss [SW71] for a quick introduction and for more information on spherical harmonics. The space $\mathcal{S}_k$ is invariant under rotations and hence provides a representation of $SO(n)$. This representation is known to be irreducible, that is, for any subspace $E \subseteq \mathcal{S}_k$ that is invariant under rotations, we necessarily have

$$E = \{0\} \qquad \text{or} \qquad E = \mathcal{S}_k.$$

Moreover, these representations in $\mathcal{S}_k$ for $k = 0, 1, \ldots$ are known to be inequivalent; this follows, e.g., from the fact that their dimensions (given by $\binom{n+k-1}{n-1} - \binom{n+k-3}{n-1}$) are all different (assuming $n \geq 3$). Elements of $\mathcal{S}_k$ are orthogonal to elements of $\mathcal{S}_\ell$ for $k \neq \ell$. We denote by $Proj_{\mathcal{S}_k}$ the

orthogonal projection operator onto $\mathcal{S}_k$ in $L^2(S^{n-1})$. Then any function $f \in L^2(S^{n-1})$ may be decomposed as

$$f = \sum_{k=0}^{\infty} Proj_{\mathcal{S}_k} f$$

where the sum converges in $L^2(S^{n-1})$. This decomposition of a function on S^{n-1} is analogous to the decomposition of a function on the Boolean hypercube into Fourier levels.

Laplacian. Write $C^\infty(S^{n-1})$ for the space of infinitely differentiable functions on S^{n-1}. For a function $f \in C^\infty(S^{n-1})$ and $x \in S^{n-1}$, we define

$$(\triangle f)(x) = \sum_{i=1}^{n-1} \frac{d}{dt^2} f\left((\cos t)x + (\sin t)e_i\right)\Big|_{t=0}, \tag{5}$$

where $e_1, \ldots, e_{n-1}$ is an orthonormal basis of $x^\perp$. Notice that for any orthogonal $x, y \in S^{n-1}$, the curve $t \mapsto (\cos t)x + (\sin t)y$ draws a great circle on S^{n-1}, that visits x at $t = 0$, and its tangent vector at $t = 0$ is the vector y. The right hand side of (5) does not depend on the choice of the orthonormal basis $e_1, \ldots, e_{n-1}$. The operator $\triangle$, acting from $C^\infty(S^{n-1})$ to itself, is called the *spherical Laplacian*.

One computes (see, e.g., [SW71]) that for any $k \geq 0$ and $\varphi_k \in \mathcal{S}_k$,

$$\triangle \varphi_k = -\lambda_k \varphi_k \tag{6}$$

where

$$\lambda_k = k(k + n - 2).$$

The Laplacian thus has a complete system of orthonormal eigenfunctions in $L^2(S^{n-1})$ (even though the Laplacian is defined only for smooth functions and not in the entire space $L^2(S^{n-1})$).

Noise operator. The noise operators on S^{n-1} are

$$U_\rho = \rho^{-\triangle} \qquad (0 \leq \rho \leq 1).$$

A priori, these operators are defined, say, on the dense space of finite linear combinations of spherical harmonics. Since the norm of U_ρ does not exceed one, we may uniquely extend U_ρ to a self-adjoint operator $U_\rho : L^2(S^{n-1}) \to L^2(S^{n-1})$ of norm one. From (6) it follows that for any $k \geq 0$ and $\varphi_k \in \mathcal{S}_k$,

$$U_\rho \varphi_k = \rho^{\lambda_k} \varphi_k.$$

Hypercontractivity. We proceed with a short review of hypercontractivity, a subject going back to Nelson [Nel66]. For $p \geq 1$ and for a measurable function $f : S^{n-1} \to \mathbb{R}$ we write $\|f\|_p = \left(\int_{S^{n-1}} |f|^p d\sigma\right)^{1/p}$ for the L^p-norm of f. The hypercontractive inequality states that for any $1 \leq p \leq q$, and any function $f \in L^p(S^{n-1})$,

$$\|U_\rho f\|_q \leq \|f\|_p \qquad \text{for } 0 \leq \rho \leq \left(\frac{p-1}{q-1}\right)^{1/(2n-2)}. \tag{7}$$

We now briefly describe how one proves such an inequality. By differentiating with respect to p and q, Gross [Gro75] showed that hypercontractive inequalities such as the one above are directly equivalent to so-called *log-Sobolev inequalities*. Indeed, a common technique for proving hypercontractive inequalities is by proving the analogous log-Sobolev inequality (as the latter is often cleaner and easier to work with). More specifically, for our hypercontractive inequality (7), the equivalent log-Sobolev inequality turns out to be

$$\int_{S^{n-1}} f^2(x) \log \frac{f^2(x)}{\int f^2(y) \mathrm{d}\sigma(y)} d\sigma(x) \leq \frac{1}{n-1} \int_{S^{n-1}} |\nabla f(x)|^2 \mathrm{d}\sigma(x) \tag{8}$$

for any smooth $f : S^{n-1} \to \mathbb{R}$ where ∇f denotes the gradient of f. Finally, this (tight) inequality was proven by Rothaus [Rot86].

We note that a slightly weaker inequality, in which $\frac{1}{n-1}$ is replaced by $\frac{1}{n-2}$ (leading to a corresponding worsening of the exponent in (7) from $1/(2n-2)$ to $1/(2n-4)$), follows from the elegant *Bakry-Émery criterion* (see [BÉ85], or e.g., [BL06]). This criterion states that a log-Sobolev inequality holds for any connected manifold whose Ricci curvature is uniformly bounded from below by some positive constant. In our very special case, the manifold is S^{n-1}, whose Ricci curvature is constantly $n-2$, leading to (8) with the slightly weaker constant $\frac{1}{n-2}$. This slightly weaker version certainly suffices for all of our needs in this paper.

Kahn, Kalai, and Linial [KKL88] realized that hypercontractive inequalities such as (7) imply certain bounds on the growth of L^p norms of the Laplacian eigenfunctions. Although they focused on the Boolean hypercube, their idea can be applied in much greater generality, and in particular to the sphere. Indeed, suppose $\varphi_k \in \mathcal{S}_k$ for some $k \geq 0$. Then $U_\rho \varphi_k = \rho^{\lambda_k} \varphi_k$. From (7), for any $1 \leq p \leq q$,

$$\|\varphi_k\|_q \leq \left(\frac{q-1}{p-1}\right)^{\lambda_k/(2n-2)} \|\varphi_k\|_p. \tag{9}$$

For large n and fixed k, we have $\lambda_k/(2n-2) \approx k/2$. In this case, the bound (9) roughly says that for any t, the set of points $x \in S^{n-1}$ where $|\varphi_k| \geq t\|\varphi_k\|_1$ has measure at most $C\exp(-ct^{2/k})$. The following lemma runs in a similar vein, and provides an upper bound on the mass that the indicator function of a set can have on each level of the spherical harmonics decomposition.

Lemma 5.3. *Suppose $f : S^{n-1} \to \mathbb{R}$ satisfies $\|f\|_1 = 1$ and $\|f\|_\infty \leq M$. Then, for any $k \geq 1$,*

$$\left\|Proj_{\mathcal{S}_k} f\right\|_2 \leq \left(e \cdot \max\left(1, \frac{\log M}{\lambda_k/(2n-2)}\right)\right)^{\lambda_k/(2n-2)}.$$

Proof. First, note that for any $p \geq 1$,

$$\|f\|_p = \left(\int_{S^{n-1}} |f|^p \mathrm{d}\sigma\right)^{1/p} \leq \left(M^{p-1} \int_{S^{n-1}} |f| \mathrm{d}\sigma\right)^{1/p} = M^{(p-1)/p} \leq M^{p-1}.$$

In particular, since $\|Proj_{\mathcal{S}_k} f\|_2 \leq \|f\|_2 \leq M$, we obtain that the lemma holds whenever $\lambda_k > (2n-2)\log M$. So assume from now on that $\lambda_k \leq (2n-2)\log M$. We use (7) for $q = 2$ and obtain that for any $1 \leq p \leq 2$,

$$\|U_\rho f\|_2 \leq \|f\|_p \leq M^{p-1} \qquad \text{for } \rho = (p-1)^{1/(2n-2)}.$$

Projecting to $\mathcal{S}_k$, we see that for any $1 \le p \le 2$,

$$(p-1)^{\frac{\lambda_k}{2n-2}} \left\| Proj_{\mathcal{S}_k} f \right\|_2 = \| Proj_{\mathcal{S}_k}(U_\rho f) \|_2 \le \| U_\rho f \|_2 \le M^{p-1}.$$

We complete the proof by choosing $p = 1 + \frac{\lambda_k}{(2n-2)\log M} \le 2$. □

Radon transform. Recall that for $\theta \in S^{n-1}$ we write $\sigma_{\theta^\perp}$ for the uniform probability measure on the sphere $S^{n-1} \cap \theta^\perp$. Then the spherical Radon transform $R(f)$ of an integrable function $f : S^{n-1} \to \mathbb{R}$ is defined as

$$R(f)(\theta) = \int_{S^{n-1} \cap \theta^\perp} f(x) \mathrm{d}\sigma_{\theta^\perp}(x), \qquad (\theta \in S^{n-1}).$$

So $R(f)$ is simply the average of f on the equator of vectors orthogonal to θ. Observe that for functions $f, g \in L^2(S^{n-1})$, we have

$$\int_{\mathcal{V}_n} f(x) g(y) \mathrm{d}\sigma_{\mathcal{V}}(x, y) = \int_{S^{n-1}} f(x) R g(x) \mathrm{d}\sigma(x). \tag{10}$$

This equality describes the intuitive fact that integrating uniformly over all orthogonal pairs (x, y) is the same as integrating uniformly over x, and then uniformly over all y in the orthogonal complement of x. See, e.g., Helgason [Hel99, Chapter II] for a more formal derivation.

Define a sequence of numbers $(\mu_k)_{k=0,1,\ldots}$ as follows. Suppose $X = (X_1, \ldots, X_{n-1})$ is a random vector that is uniformly distributed in S^{n-2}. For an even $k \ge 0$ denote

$$\mu_k = (-1)^{k/2} \mathbb{E}[X_1^k],$$

and for odd k set $\mu_k = 0$. We now show that $\mathcal{S}_k$ are the eigenspaces of R with μ_k being the corresponding eigenvalues.

Lemma 5.4. *For any $k \ge 0$ and $\varphi_k \in \mathcal{S}_k$,*

$$R(\varphi_k) = \mu_k \varphi_k.$$

Proof. The Radon transform clearly commutes with rotations. Therefore, because the $\mathcal{S}_k$'s give rise to inequivalent irreducible representations, Schur's lemma implies that R must have the $\mathcal{S}_k$'s as its eigenspaces. We briefly recall the proof of this standard representation-theoretic fact. Consider the restriction $R_{k,j}$ of $Proj_{\mathcal{S}_j} R$ to an operator from $\mathcal{S}_k$ to $\mathcal{S}_j$ for some $k, j \ge 0$. Our goal is to show that $R_{k,j}$ is zero whenever $k \ne j$ and a multiple of the identity otherwise. Since $R_{k,j}$ commutes with the action of $SO(n)$, and $\mathcal{S}_k$ is irreducible, we have that $\ker R_{k,j}$ is either all of $\mathcal{S}_k$ or $\{0\}$. In the former case $R_{k,j} = 0$ and we are done, so assume the latter case. By the same argument the image of $R_{k,j}$ is either all of $\mathcal{S}_j$ or $\{0\}$, and since we assumed $R_{k,j} \ne 0$, it must be the former. Hence $R_{k,j}$ is an isomorphism between the representation on $\mathcal{S}_k$ and on $\mathcal{S}_j$, which is impossible when $k \ne j$ since we know that $\mathcal{S}_k$ and $\mathcal{S}_j$ are inequivalent representations. So assume $k = j$, and let $\lambda \in \mathbb{R}$ be an arbitrary eigenvalue of $R_{k,k}$ (there exists such an eigenvalue since $R_{k,k}$ is a symmetric operator). Then the kernel of $\lambda I - R_{k,k}$ must also be either all of $\mathcal{S}_k$ or $\{0\}$; the latter is impossible since λ is an eigenvalue, hence we necessarily have $R_{k,k} = \lambda I$.

Our next goal is to show that the μ_k's are the corresponding eigenvalues. Fix some arbitrary $e \in S^{n-1}$. For $k \geq 0$, we define the function $f_k : S^{n-1} \to \mathbb{R}$ by

$$f_k(x) = G_k(x \cdot e) \qquad (x \in S^{n-1})$$

where $G_k : [-1, 1] \to \mathbb{R}$ is the Gegenbauer polynomial (see, e.g., Müller [Mül66]),

$$G_k(t) = \mathbb{E}\left(t + iX_1\sqrt{1-t^2}\right)^k.$$

Here, $i^2 = -1$ and $X = (X_1, \ldots, X_{n-1})$ is a random vector that is distributed uniformly over the sphere S^{n-2}. The function f_k is known to be a spherical harmonic of degree k, i.e., in $\mathcal{S}_k$ [Mül66], and by our above discussion, must be an eigenfunction of R, i.e., Rf is proportional to f. From the definition of the Radon transform,

$$(Rf)(e) = G_k(0) \qquad \text{and} \qquad f(e) = G_k(1) = 1.$$

We conclude that $G_k(0)$ is the eigenvalue corresponding to $\mathcal{S}_k$. It remains to notice that $G_k(0)$ vanishes for odd k and equals $(-1)^{k/2}\mathbb{E}X_1^k$ for even k, and hence equals μ_k for all k. □

The next technical lemma gives upper bounds on the eigenvalues μ_k.

Lemma 5.5. *Suppose $n \geq 10$. Then, the sequence $|\mu_0|, |\mu_2|, |\mu_4|, \ldots$ is non-increasing, and moreover, for all $k \geq 1$,*

$$|\mu_k| \leq \left(C\frac{k}{n}\right)^{k/2}.$$

Proof. The first claim follows immediately from the fact that $|X_1| \leq 1$ and $|\mu_{2k}| = \mathbb{E}[|X_1|^{2k}]$. For the second claim, notice that the density of X_1 is proportional to $(1-x^2)^{(n-4)/2}$ for $x \in [-1, 1]$, and vanishes outside this interval. Hence, our goal is to prove that for all even $k \geq 2$,

$$\int_{-1}^{1} x^k (1-x^2)^{(n-4)/2} \mathrm{d}x \leq \left(C\frac{k}{n}\right)^{k/2} \int_{-1}^{1} (1-x^2)^{(n-4)/2} \mathrm{d}x.$$

The integral on the right hand side is at least $c/\sqrt{n}$ (this is true even for the integral from $-1/\sqrt{n}$ to $1/\sqrt{n}$). The integral on the left hand side may be estimated as follows:

$$\int_{-1}^{1} x^k (1-x^2)^{\frac{n-4}{2}} \mathrm{d}x \leq \int_{-1}^{1} x^k e^{-\frac{n-4}{2}x^2} \mathrm{d}x \leq \int_{-\infty}^{\infty} x^k e^{-\frac{n-4}{2}x^2} \mathrm{d}x.$$

The latter integral is exactly the kth moment of a normal variable with mean 0 and variance $1/(n-4)$, times the missing normalization factor of $\sqrt{2\pi/(n-4)}$. A standard fact is that for even k this moment is

$$(n-4)^{-k/2} \cdot (k-1)!! \leq \left(\frac{k}{n-4}\right)^{k/2}$$

where $(k-1)!! = (k-1)(k-3)\cdots 1$. The lemma follows. □

Proof of Theorem 5.2. It suffices to prove the theorem under the assumption that $n \geq 10$ (otherwise there is no $s \leq cn$, for a sufficiently small universal constant $c > 0$). By Lemma 5.4 and (10),

$$\int_{\mathcal{V}_n} f(x)g(y)\mathrm{d}\sigma_{\mathcal{V}}(x,y) = \int_{S^{n-1}} fR\,(g)\,\mathrm{d}\sigma = \sum_{k=0}^{\infty} \mu_k \int_{S^{n-1}} Proj_{\mathcal{S}_k}\,(f)\,Proj_{\mathcal{S}_k}\,(g)\,\mathrm{d}\sigma.$$

Note that $\mu_0 = 1$ and $Proj_{\mathcal{S}_0}\,(f) \equiv Proj_{\mathcal{S}_0}\,(g) \equiv 1$. Therefore, by the Cauchy-Schwarz inequality,

$$\left| \int_{\mathcal{V}_n} f(x)g(y)\mathrm{d}\sigma_{\mathcal{V}}(x,y) \, - \, 1 \right| \leq \sum_{k=1}^{\infty} |\mu_{2k}| \|Proj_{\mathcal{S}_{2k}} f\|_2 \|Proj_{\mathcal{S}_{2k}} g\|_2.$$

We will prove the theorem by showing that the latter sum is at most $C\alpha\beta/n$, where $\alpha = \log(2\|f\|_\infty)$ and $\beta = \log(2\|g\|_\infty)$. Observe that $\alpha, \beta \geq 1/2$ and recall our assumption that $\alpha\beta$ is at most cn. We start by analyzing the part of the sum in which k runs from 1 to $T-1$ where $T = \lfloor \delta n \rfloor$ for some sufficiently small constant $\delta > 0$. Using Lemmas 5.3 and 5.5, we have the bounds

$$|\mu_{2k}| \leq \left(\frac{Ck}{n}\right)^k,$$
$$\|Proj_{\mathcal{S}_{2k}} f\|_2 \leq \left(C \max\left(1, \frac{\alpha}{k}\right)\right)^{\lambda_{2k}/(2n-2)},$$

and similarly for g with β. Therefore,

$$\sum_{k=1}^{T-1} |\mu_{2k}| \|Proj_{\mathcal{S}_{2k}} f\|_2 \|Proj_{\mathcal{S}_{2k}} g\|_2 \leq \sum_{k=1}^{T-1} \left(\frac{Ck}{n}\right)^k \left(C \max\left(1, \frac{\alpha}{k}\right)\right)^{\lambda_{2k}/(2n-2)} \left(C \max\left(1, \frac{\beta}{k}\right)\right)^{\lambda_{2k}/(2n-2)}.$$

The term $k = 1$ is at most

$$\frac{C\alpha\beta}{n}.$$

We will now show that the terms in the latter sum decay geometrically, and hence we can also bound the sum by $C\alpha\beta/n$. To this end, first notice that

$$\left(\frac{C(k+1)}{n}\right)^{k+1} \Big/ \left(\frac{Ck}{n}\right)^k = \frac{C(k+1)}{n} \cdot \left(\frac{k+1}{k}\right)^k \leq \frac{\tilde{C}k}{n}.$$

Second,

$$\begin{aligned}
\left(C \max\left(1, \frac{\alpha}{k+1}\right)\right)^{\lambda_{2k+2}/(2n-2)} \Big/ \left(C \max\left(1, \frac{\alpha}{k}\right)\right)^{\lambda_{2k}/(2n-2)} &\leq \left(C \max\left(1, \frac{\alpha}{k}\right)\right)^{(\lambda_{2k+2}-\lambda_{2k})/(2n-2)} \\
&= \left(C \max\left(1, \frac{\alpha}{k}\right)\right)^{1+\frac{4k+1}{n-1}} \\
&\leq \tilde{C} \max\left(1, \frac{\alpha}{k}\right).
\end{aligned}$$

Hence the ratio between the term for $k+1$ and that for k is at most

$$C\frac{k}{n} \max\left(1, \frac{\alpha}{k}\right) \max\left(1, \frac{\beta}{k}\right) \leq \frac{1}{2},$$

as $k \leq \delta n$, once we choose δ to be a sufficiently small positive universal constant. This implies that we can upper bound the sum from 1 to $T-1$ by $C\alpha\beta/n$, as required.

It remains to analyze the less significant part of the sum, in which k runs from $T = \lfloor \delta n \rfloor$ to infinity. Then, by Lemma 5.5 and another application of Cauchy-Schwarz,

$$\begin{aligned}\sum_{k=T}^{\infty} |\mu_{2k}| \|Proj_{S_{2k}} f\|_2 \|Proj_{S_{2k}} g\|_2 &\leq |\mu_{2T}| \sum_{k=T}^{\infty} \|Proj_{S_{2k}} f\|_2 \|Proj_{S_{2k}} g\|_2 \\ &\leq |\mu_{2T}| \|f\|_2 \|g\|_2 \\ &\leq \exp(\alpha + \beta - cn) \\ &\leq \frac{C}{n} \leq \frac{\tilde{C}\alpha\beta}{n},\end{aligned}$$

under the legitimate assumption that $\alpha\beta \leq \tilde{c}n$. We conclude that the entire sum is bounded by $C\alpha\beta/n$. □

6 Sampling Sets by Lower Dimensional Subspaces

Our next step is to iterate Theorem 5.2, using a certain martingale process, in order to obtain a corresponding theorem for the Grassmannian. The constants 0.1 and 9/10 appearing below do not play any special role and can be replaced with any other constants (as long as the former is positive and the latter is smaller than 1).

Theorem 6.1. *Let $1 \leq m \leq 9n/10$. Suppose that $A \subseteq S^{n-1}$ is a measurable set with $\sigma(A) \geq C\exp(-cn^{1/3})$. Assume that H is a uniformly chosen $(n-m)$-dimensional subspace. Then,*

$$\left| \frac{\sigma_H(A \cap H)}{\sigma(A)} - 1 \right| < 0.1$$

except with probability at most $C\exp(-cn^{1/3})$. Here, $c, C > 0$ are universal constants.

We start with a few technical lemmas. The first one below bounds the moments of a random variable that has an exponentially decaying tail around 1. We will apply it with random variables whose expectation is very close to 1.

Lemma 6.2. *Let $R, \delta > 0$ and let Z be a non-negative random variable satisfying that for any $t \geq 0$,*

$$\mathbb{P}(|Z-1| \geq t) \leq R\exp(-t/\delta).$$

Then, for any $2 \leq \ell \leq (2\delta)^{-1}$,

$$\mathbb{E}[Z^\ell] \leq 1 + \ell\,\mathbb{E}[Z-1] + 2R(\ell\delta)^2.$$

Proof. Using the Taylor expansion, we have that for any $x \geq -1$,

$$\begin{aligned}(1+x)^\ell &= 1 + \ell x + \sum_{k=2}^{\lfloor \ell \rfloor - 1} \binom{\ell}{k} x^k + \binom{\ell}{\lfloor \ell \rfloor} (1+\xi)^{\ell - \lfloor \ell \rfloor} x^{\lfloor \ell \rfloor} \\ &\leq 1 + \ell x + \sum_{k=2}^{\lfloor \ell \rfloor} \frac{\ell^k}{k!} |x|^k + \frac{\ell^{\lfloor \ell \rfloor}}{\lfloor \ell \rfloor !} |x|^{\lfloor \ell \rfloor + 1},\end{aligned}$$

where ξ is some real number between x and 0. Next, for any $k \geq 1$,

$$\begin{aligned}\mathbb{E}[|Z-1|^k] &= \int_0^\infty \mathbb{P}[|Z-1|^k \geq t]\mathrm{d}t \\ &= \int_0^\infty kt^{k-1}\,\mathbb{P}[|Z-1| \geq t]\mathrm{d}t \\ &\leq R\,k\int_0^\infty t^{k-1}\exp(-t/\delta)\mathrm{d}t = R\cdot k!\cdot\delta^k. \end{aligned} \tag{11}$$

Combining the two inequalities, we obtain

$$\begin{aligned}\mathbb{E}[Z^\ell] &\leq 1+\ell\,\mathbb{E}[Z-1] + R\sum_{k=2}^{\lfloor \ell\rfloor}(\ell\delta)^k + R(\ell\delta)^{\lfloor\ell\rfloor}(\lfloor\ell\rfloor+1)\delta \\ &\leq 1+\ell\,\mathbb{E}[Z-1] + 2R(\ell\delta)^2.\end{aligned}$$

□

Our second lemma bounds the upper tail of a certain martingale-like product and is based on a Bernstein-type inequality. We then derive as a corollary a similar bound on the lower tail.

Lemma 6.3. *Let $R,\delta > 0$ and let $Z_1,\ldots,Z_k$ be non-negative random variables where $k \leq 1/(320R\delta^2)$. Assume that for all $1 \leq i \leq k$, when conditioning on any values of $Z_1,\ldots,Z_{i-1}$, we almost surely have*

$$\mathbb{E}[Z_i \mid Z_1,\ldots,Z_{i-1}] \leq 1+\frac{1}{20k}, \tag{12}$$

$$\mathbb{P}[|Z_i-1| \geq t \mid Z_1,\ldots,Z_{i-1}] \leq R\exp(-t/\delta) \quad \textit{for all } t \geq 0. \tag{13}$$

Then,

$$\mathbb{P}\left[\prod_{i=1}^k Z_i \geq 1.1\right] \leq \begin{cases} \exp(-1/(80\delta)+Rk/2), & k < 1/(80R\delta) \\ \exp(-1/(12800Rk\delta^2)), & \textit{otherwise}. \end{cases}$$

Proof. Let $2 \leq \ell \leq (2\delta)^{-1}$ be a real number to be determined later on. Then, by Lemma 6.2,

$$\begin{aligned}\mathbb{E}\left[\Big(\prod_{i=1}^k Z_i\Big)^\ell\right] &= \mathbb{E}_{Z_1,\ldots,Z_{k-1}}\left[\Big(\prod_{i=1}^{k-1} Z_i\Big)^\ell \mathbb{E}[Z_k^\ell \mid Z_1,\ldots,Z_{k-1}]\right] \\ &\leq \left(1+\frac{\ell}{20k}+2R(\ell\delta)^2\right)\mathbb{E}_{Z_1,\ldots,Z_{k-1}}\left[\Big(\prod_{i=1}^{k-1} Z_i\Big)^\ell\right] \\ &\leq \cdots \leq \left(1+\frac{\ell}{20k}+2R(\ell\delta)^2\right)^k \leq \exp\left(\frac{\ell}{20}+2Rk(\ell\delta)^2\right).\end{aligned}$$

Therefore,

$$\mathbb{P}\left[\prod_{i=1}^k Z_i \geq 1.1\right] \leq 1.1^{-\ell}\exp\left(\frac{\ell}{20}+2Rk(\ell\delta)^2\right) \leq \exp\left(-\frac{\ell}{40}+2Rk(\ell\delta)^2\right).$$

The minimum of the right hand side over ℓ is $\exp(-1/(12800Rk\delta^2))$ and is obtained for $\ell = 1/(160Rk\delta^2)$. We set ℓ to this value, unless it is greater than $1/(2\delta)$, in which case we set $\ell = 1/(2\delta)$. The lemma follows. □

Corollary 6.4. *Let $R, \delta > 0$ and let $Z_1, \ldots, Z_k$ be random variables taking values in $(1/2, \infty)$ where $k \leq 1/(1280R\delta^2)$. Assume that for all $1 \leq i \leq k$, conditioning on any values of $Z_1, \ldots, Z_{i-1}$, we almost surely have*

$$\mathbb{E}[Z_i \mid Z_1, \ldots, Z_{i-1}] \geq 1 - \frac{1}{40k},$$
$$\mathbb{P}[|Z_i - 1| \geq t \mid Z_1, \ldots, Z_{i-1}] \leq R\exp(-t/\delta) \quad \textit{for all } t \geq 0.$$

Then,

$$\mathbb{P}\left[\prod_{i=1}^{k} Z_i \leq 0.9\right] \leq \begin{cases} \exp(-1/(160\delta) + Rk/2), & k < 1/(160R\delta) \\ \exp(-1/(51200Rk\delta^2)), & \textit{otherwise}. \end{cases}$$

Proof. We simply apply Lemma 6.3 to the random variables $Z_1^{-1}, \ldots, Z_k^{-1}$ with R and 2δ. Eq. (13) holds because for all $t \geq 0$ and $x \geq 1/2$ if $|x^{-1} - 1| \geq t$ then also $|x - 1| \geq t/2$. For Eq. (12), we use the inequality $x^{-1} \leq 1 - (x-1) + 2(x-1)^2$, valid for all $x \geq 1/2$. This implies that

$$\begin{aligned} \mathbb{E}[Z_i^{-1} \mid Z_1, \ldots, Z_{i-1}] &\leq 1 + \frac{1}{40k} + 2\mathbb{E}[(Z_i - 1)^2 \mid Z_1, \ldots, Z_{i-1}] \\ &\leq 1 + \frac{1}{40k} + 4R\delta^2 \leq 1 + \frac{1}{20k}, \end{aligned}$$

where the next-to-last inequality follows from the calculation in (11). □

Proof of Theorem 6.1. Fix $1 \leq m \leq 9n/10$ and a set $A \subseteq S^{n-1}$. Consider a sequence of random subspaces in $\mathbb{R}^n$,

$$\mathbb{R}^n = H_0 \supset H_1 \supset H_2 \supset \cdots \supset H_m$$

in which $\dim(H_i) = n - i$, defined as follows. For each $i \geq 1$, the subspace H_i is chosen uniformly in the Grassmannian of all $(n-i)$-dimensional subspaces of H_{i-1}. An important observation, which follows from the uniqueness of the Haar measure, is that the subspace H_i is distributed uniformly over $\mathcal{G}_{n,n-i}$, and in particular, H_m is a uniform $(n-m)$-dimensional subspace.

For $k = 1, \ldots, m$ define the random variable

$$X_k = \frac{\sigma_{H_k}(A \cap H_k)}{\sigma_{H_{k-1}}(A \cap H_{k-1})},$$

where σ_{H_k} is the uniform measure on the sphere $S^{n-1} \cap H_k$. If the denominator vanishes, we set the random variable to 1. Notice that

$$\prod_{k=1}^{m} X_k = \frac{\sigma_{H_m}(A \cap H_m)}{\sigma(A)}$$

and hence our goal is to show that this product is in 1 ± 0.1 except with probability at most $C\exp(-cn^{1/3})$. We will do this by applying Lemma 6.3 and Corollary 6.4 to a regularized version of $X_1, \ldots, X_m$ defined below.

We note three properties of the random variables X_k. First, we have that for any $1 \leq k \leq m$, conditioned on any values of $H_1, \ldots, H_{k-1}$,

$$\mathbb{E}\left(X_k | H_1, \ldots, H_{k-1}\right) = 1.$$

This holds since H_k is distributed uniformly over the Grassmannian of subspaces of H_{k-1}. Second, by definition, X_k is bounded from above by $1/(\sigma_{H_{k-1}}(A \cap H_{k-1}))$. Finally, by Theorem 5.1, for any $0 < t < 1$,

$$\begin{aligned}\mathbb{P}(|X_k - 1| \geq t \mid H_1, \ldots, H_{k-1}) &\leq C\exp(-c(n-k+1)t/\log(2/\sigma_{H_{k-1}}(A \cap H_{k-1}))) \\ &\leq C\exp(-\tilde{c}nt/\log(2/\sigma_{H_{k-1}}(A \cap H_{k-1}))),\end{aligned}$$

where we used the fact that $k \leq m \leq 9n/10$. Because this tail bound holds only for $t < 1$, we cannot apply Lemma 6.3 and Corollary 6.4 directly, and instead proceed below to define the regularized random variables $Z_1, \ldots, Z_m$.

Next, for $0 \leq k \leq m$, we define the "bad" event B_k as the event that $X_1 X_2 \cdots X_k \leq 1/2$ and for $1 \leq k \leq m$, the "bad" event C_k as the event that $|X_k - 1| \geq 1/2$. Condition on any $H_1, \ldots, H_{k-1}$ such that B_{k-1} does not occur. In this case, $\sigma_{H_{k-1}}(A \cap H_{k-1}) \geq \sigma(A)/2$. Hence, X_k is upper bounded by $2/\sigma(A) \leq C\exp(cn^{1/3})$. Moreover, for any $0 < t < 1$ the probability that $|X_k - 1| \geq t$ is at most $C\exp(-cnt/\log(4/\sigma(A))) \leq C\exp(-\tilde{c}n^{2/3}t)$, and in particular the probability that C_k occurs is at most $C\exp(-cn^{2/3})$. For $1 \leq k \leq m$, we define the random variable Z_k as follows: if either B_{k-1} or C_k occurs, Z_k is 1. Otherwise, $Z_k = X_k$.

We can now finally apply Lemma 6.3 and Corollary 6.4: for each $1 \leq k \leq m$, we apply them to the sequence $Z_1, \ldots, Z_k$ with $R = C$ and $\delta = \tilde{C}n^{-2/3}$. To see why the conditions there hold, condition on any $H_1, \ldots, H_{k-1}$, and assume first that B_{k-1} does not occur. Then

$$\begin{aligned}|\mathbb{E}[Z_k|H_1, \ldots, H_{k-1}] - 1| &= |\mathbb{E}[Z_k - X_k|H_1, \ldots, H_{k-1}]| \\ &\leq \mathbb{P}[C_k|H_1, \ldots, H_{k-1}] \cdot C\exp(cn^{1/3}) \\ &\leq \tilde{C}\exp(-\tilde{c}n^{2/3}).\end{aligned}$$

Moreover, for *all* non-negative t, the probability that $|Z_k - 1| \geq t$ is at most $C\exp(-cn^{2/3}t)$. Finally, these two statements are obviously true even when B_{k-1} does occur (since in this case Z_k is simply 1), hence we obtain that the two statements hold conditioned on any $H_1, \ldots, H_{k-1}$ (and in particular, on any $Z_1, \ldots, Z_{k-1}$). As a result, the lemma and the corollary imply that for each $1 \leq k \leq m$, $|Z_1 \cdots Z_k - 1| \geq 0.1$ with probability at most $C\exp(-cn^{1/3})$. Moreover, by a union bound, the probability that there exists a k for which $|Z_1 \cdots Z_k - 1| \geq 0.1$, an event which we denote by D, is at most

$$\mathbb{P}[D] \leq m \cdot C\exp(-cn^{1/3}) \leq \tilde{C}\exp(-\tilde{c}n^{1/3}). \tag{14}$$

Next, we claim that for any $1 \leq k \leq m$,

$$\mathbb{P}[\neg D \wedge \neg C_1 \wedge \cdots \wedge \neg C_{k-1} \wedge C_k] \leq C\exp(-cn^{2/3}). \tag{15}$$

To see why, notice that $\neg C_1$ implies that $X_1 = Z_1$, which together with $\neg D$ implies that $\neg B_1$; the latter, in turn, implies that $X_2 = Z_2$ (since neither B_1 nor C_2 happens), which implies that B_2 does not happen either; etc. As a result, we obtain that $\neg B_{k-1}$, which implies that the probability of C_k is at most $C\exp(-cn^{2/3})$, as desired.

By summing all the probabilities in (14) and (15), we obtain that

$$\mathbb{P}[\neg D \wedge \neg C_1 \wedge \cdots \wedge \neg C_m] \geq 1 - C\exp(-cn^{1/3}).$$

It remains to notice using the same argument as above that this event implies that for all k, $Z_k = X_k$ and therefore also that $|X_1 \cdots X_m - 1| < 0.1$. □

The only thing remaining is to derive Lemma 4.3 from Theorem 6.1. We restate it here in a slightly more general form.

Lemma 6.5. *Let $1 \leq m \leq 9n/10$. Suppose that $A \subseteq S^{n-1}$ and $B \subseteq \mathcal{G}_{n,n-m}$ are measurable sets with*

$$\sigma(A) \geq C\exp(-cn^{1/3}), \qquad \sigma_{\mathcal{G}}(B) \geq C\exp(-cn^{1/3})$$

for some universal constants $c, C > 0$. Then,

$$\sigma_{\mathcal{I}}\left((A \times B) \cap \mathcal{I}_{n,m}\right) \geq 0.8\,\sigma(A)\sigma_{\mathcal{G}}(B).$$

Proof. Notice that $\sigma_{\mathcal{I}}((A \times B) \cap \mathcal{I}_{n,m})/\sigma_{\mathcal{G}}(B)$ may be interpreted as the probability that when choosing a subspace H uniformly from B and a uniform vector x in $H \cap S^{n-1}$, we have $x \in A$. To analyze this probability, denote by $E \subseteq \mathcal{G}_{n,n-m}$ the set of all $(n-m)$-dimensional subspaces H for which

$$\frac{\sigma_H(A \cap H)}{\sigma(A)} \leq 0.9.$$

Then, by Theorem 6.1, $\sigma_{\mathcal{G}}(E) \leq C\exp(-cn^{1/3})$. Next, observe that the probability that $H \in E$ is at most $\sigma_{\mathcal{G}}(E)/\sigma_{\mathcal{G}}(B)$. Moreover, if $H \notin E$, then by definition, the probability that $x \in A$ is at least $0.9\,\sigma(A)$. Hence,

$$\frac{\sigma_{\mathcal{I}}((A \times B) \cap \mathcal{I}_{n,m})}{\sigma_{\mathcal{G}}(B)} \geq \left(1 - \frac{\sigma_{\mathcal{G}}(E)}{\sigma_{\mathcal{G}}(B)}\right) 0.9\,\sigma(A) > 0.8\,\sigma(A),$$

assuming the universal constants are chosen properly. □

Acknowledgments

We thank Ronald de Wolf for comments on an earlier draft.

References

[AST+98] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson. The quantum communication complexity of sampling. *SIAM Journal on Computing*, 32(6):1570–1585, 2003. Preliminary version in FOCS 1998.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proc. of 30th Annual ACM Symposium on the Theory of Computing*, pages 63–68. 1998. Quant-ph/9802040.

[BÉ85] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[Bec75] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[BJK04] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proc. of 36th Annual ACM Symposium on the Theory of Computing*, pages 128–137. 2004.

[BL06] D. Bakry and M. Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.*, 22(2):683–702, 2006.

[Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.

[CR10] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of gap Hamming distance, 2010. To appear.

[Gav08] D. Gavinsky. Classical interaction cannot replace a quantum message. In *Proc. of 40th Annual ACM Symposium on the Theory of Computing*, pages 95–102. 2008. quant-ph/0703215.

[Gav09] D. Gavinsky. Classical interaction cannot replace quantum nonlocality, 2009. Arxiv:0901.0956.

[GKK+07] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. d. Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal on Computing*, 38(5):1695–1708, 2008. quant-ph/0611209.

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[GRW08] D. Gavinsky, O. Regev, and R. d. Wolf. Simultaneous communication protocols with quantum and classical messages. *Chicago Journal of Theoretical Computer Science*, 2008(7), December 2008.

[Hel99] S. Helgason. *The Radon transform*, volume 5 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, second edition, 1999.

[JK10] R. Jain and H. Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th Annual IEEE Conference on Computational Complexity*, pages 247–258. 2010.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. of 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80. 1988.

[Kla10] H. Klauck. A strong direct product theorem for disjointness. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 77–86. 2010.

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

[Kre95] I. Kremer. *Quantum Communication*. Master's thesis, Hebrew University, Computer Science Department, 1995.

[Mon10] A. Montanaro. A new exponential separation between quantum and classical one-way communication complexity, 2010. arxiv:1007.3587.

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[Mül66] C. Müller. *Spherical harmonics*, volume 17 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1966.

[MW03] V. Milman and R. Wagner. Some remarks on a lemma of Ran Raz. *Lecture Notes in Mathematics*, 1807:158–168, 2003.

[Nel66] E. Nelson. A quartic interaction in two dimensions. In *Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965)*, pages 69–73. M.I.T. Press, Cambridge, Mass., 1966.

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

[Rot86] O. S. Rothaus. Hypercontractivity and the Bakry-Emery criterion for compact Lie groups. *J. Funct. Anal.*, 65(3):358–367, 1986.

[SW71] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces*. Princeton University Press, Princeton, N.J., 1971. Princeton Mathematical Series, No. 32.

[Wol08] R. d. Wolf. *A Brief Introduction to Fourier Analysis on the Boolean Cube*. Number 1 in Graduate Surveys. Theory of Computing Library, 2008.

[Yao93] A. C.-C. Yao. Quantum circuit complexity. In *34th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 352–361. 1993.