Exponential Separation of Quantum and Classical Communication Complexity

Ran Raz
ranraz@wisdom.weizmann.ac.il,
Department of Applied Mathematics,
Weizmann Institute,
Rehovot 76100, ISRAEL

Abstract

Communication complexity has become a central complexity model. In that model, we count the amount of communication bits needed between two parties in order to solve certain computational problems. We show that for certain communication complexity problems quantum communication protocols are exponentially faster than classical ones. More explicitly, we give an example for a communication complexity relation (or promise problem) $\mathcal{P}$ such that:

1. The quantum communication complexity of $\mathcal{P}$ is $O(\log m)$.
2. The classical probabilistic communication complexity of $\mathcal{P}$ is $\Omega(m^{1/4}/\log m)$.

(where m is the length of the inputs). This gives an exponential gap between quantum communication complexity and classical probabilistic communication complexity. Only a quadratic gap was previously known.

Our problem $\mathcal{P}$ is of geometrical nature, and is a finite precision variation of the following problem: Player I gets as input a unit vector $x \in R^n$ and two orthogonal subspaces $M_0, M_1 \subset R^n$. Player II gets as input an orthogonal matrix $T : R^n \to R^n$. Their goal is to answer 0 if $T(x) \in M_0$ and 1 if $T(x) \in M_1$, (and any answer in any other case). We give an almost tight analysis for the quantum communication complexity and for the classical-probabilistic communication complexity of this problem.

1 Introduction

Quantum computers, if ever built, will be able to factor numbers in polynomial time (as proved by Shor [Sho]). Since it is widely believed that numbers cannot be factored in polynomial time by any classical (deterministic or probabilistic) Turing machine, Shor's result gives a strong indication (although not a proof) that quantum computers will be able to solve certain computational problems significantly faster than classical ones.

In this paper, we are interested in the equivalent question for communication complexity: Can quantum communication channels significantly reduce the amount of communication needed to solve certain communication problems ? Are quantum communication protocols significantly stronger than classical ones ? We show that for certain communication complexity problems quantum communication protocols are exponentially faster than classical ones. Unlike the equivalent gap for complexity (that follows from Shor's result), our gap for communication complexity is completely proven.

A communication complexity problem is given by 3 finite sets X, Y, Z and a relation $\mathcal{R} \subset X \times Y \times Z$. We think of $\mathcal{R}$ also as a function $\mathcal{R} : X \times Y \times Z \to \{0, 1\}$. We have two players, (Player I and Player II). Player I is given an input $x \in X$. Player II is given an input $y \in Y$. Initially, non of the players has any information about the input of the other. Their goal is to come up with an answer z such that $\mathcal{R}(x, y, z) = 1$. The communication complexity of the problem is the number of bits the two players have to exchange between them in order to come up with such an answer z. Each player has an unlimited computational power and the players cooperate with each other. We count the amount of communication needed by the best protocol, for the worst case input.

There are several variations of communication complexity, depending on what are the allowed protocols. The classical deterministic communication complexity of a problem is the amount of communication bits needed by the best deterministic protocol. In the probabilistic communication complexity model, we allow the protocol to depend on random bits and we allow a small probability of error. We require that for every x, y the answer z will be correct with high probability (say with probability $\geq 1 - err$, where $err > 0$ is some small fixed constant). The classical probabilistic communication complexity of a problem is the amount of communication bits needed by the best such protocol. It is well known that probabilistic protocols are much stronger than deterministic ones.

In the quantum communication complexity model, each of the players has an infinitely powerful quantum computer and the two players can exchange between them quantum bits (qubits), rather than classical bits. Since the quantum model is probabilistic by its nature, we usually consider the probabilistic case and allow a small probability of error. As before, we require that for every x, y the answer z will be correct with high probability (say with probability $\geq 1 - err$, where $err > 0$ is some small fixed constant). The quantum communication complexity of a problem is the amount of communication qubits needed by the best such protocol. The 0-error quantum communication complexity is defined in the same way, but without allowing any errors. That is, for every x, y the answer z has to be correct with probability 1.

The model of communication complexity was introduced by Yao for functions [Yao1] (i.e., when there is only one correct answer z for every pair x, y) and was generalized to relations in [KW]. Besides being very interesting in its own right, the model was found out to be relevant to many other complexity issues and has become a central complexity model. The quantum communication complexity model, for both, functions and relations, was introduced by Yao in [Yao2]. We will discuss the different models of communication complexity with some more details in Section 2.

As mentioned above, it is well known that classical probabilistic communication protocols are much stronger than classical deterministic ones. In this paper, we show that quantum communication protocols are much stronger than classical probabilistic (or deterministic) ones.

Note, that by the quantum teleportation method of [BBCJPW], if the two players share EPR-pairs then the quantum communication between them can be done by classical channels ! That is, the two players can exchange between them qubits by transmitting classical bits, rather than by actually sending quantum particles. This is possible if the players share Einstein-Podolsky-Rosen pairs, that is, pairs of particles that interacted in the past and are in entangled quantum state. Note also that quantum communication and quantum teleportation have been successfully tested ! [BPMWZ, HADLMS, MWKZ]

For an excellent survey on communication complexity see [KN]. For excellent surveys on quantum complexity see [Aha] and [Pre].

1.1 Related Results

Holevo has proved that by communicating n qubits one cannot transmit more than n bits of information [Hol]. This, however, doesn't imply that quantum communication protocols are as strong as classical ones (as the answer for a communication complexity problem may be very short, e.g., one bit).

Previous to this work, the largest known gap between quantum protocols and classical probabilistic protocols was a quadratic gap, proved by Buhrman Cleve and Wigderson for the set disjointness function [BCW]. It is well known that the probabilistic communication complexity of the set disjointness function is $\Omega(n)$ [KS] (see also [Razb]). In [BCW] it was proved that the quantum communication complexity of that function is $O(\sqrt{n} \cdot \log n)$. (The proof is by a relatively simple reduction to Grover's database search algorithm [Gro]). This gives an (almost) quadratic gap between the two models. That gap is still the best known gap for the communication complexity of functions (the gap we prove here is for relations, promise problems or partial functions, but not for functions). The result of [BCW] is very interesting also because of the central role that the set disjointness function plays in communication complexity.

[BCW] also gave an example for a relation $\mathcal{R}$ with classical deterministic communication complexity of $\Theta(n)$ and 0-error quantum communication complexity of $\Theta(\log n)$. This gives an exponential gap between classical deterministic and 0-error quantum communication complexity. However, the classical probabilistic communication complexity of the same relation is $O(1)$ and hence that result is of less interest for us.

Another related result was recently proved by [ASTVW]. [ASTVW] initiates the study of the so called sampling problems. Consider for example the following problem: Let W be a set of size n. Player I and Player II want to generate a pair of subsets (U, V), such that $U \subset W$ is known to Player I, $V \subset W$ is known to Player II, and the pair (U, V) is a random variable uniformly distributed on the set of all pairs with $|U| = |V| = \sqrt{n}$ and $|U \cap V| = 0$. It was proved in [ASTVW] that this can be done (with a small probability of error) by a quantum protocol with communication of only $O(\log n)$ qubits, and cannot be done by any (private-coins) classical probabilistic protocol with communication of less than $\Omega(\sqrt{n})$ bits. This gives an exponential gap for sampling problems between the quantum model and the (private-coins) classical model.

Sampling problems, however, do not fall into the standard definitions of communication complexity and it is not clear if they are related to communication complexity problems. In particular, in a sampling problem the players do not have any inputs and the answers are required to be random variables with some specific distribution (up to a small probability of error). In addition, sampling problems make sense only in the private-coins model of probabilistic communication complexity, where the players do not share random bits. In the more general model, (the public-coins model), where the two players share a string of random bits, any sampling problem is trivial and can be solved with no communication at all. Note that for communication complexity problems it is well known that the private-coins model and the public-coins model have roughly the same power [New]. The model used in most discussions (of both upper and lower bounds) is the public-coins model.

1.2 Our Results

Our results are obtained by analyzing the communication complexity of a finite precision variation of the problem $\mathcal{P}_1(\vartheta)$ defined below. For completeness, we also define a related problem, $\mathcal{P}_0(\vartheta)$. Let ϑ be some constant $0 \leq \vartheta < 1/\sqrt{2}$, and let R^n be the n dimensional vector space over the real numbers.

The problem $\mathcal{P}_0(\vartheta)$:
Player I gets as input a unit vector $x \in R^n$. Player II gets as input two orthogonal vector-spaces $M_0, M_1 \subset R^n$ of dimension $n/2$ each (we assume for simplicity that n is even). Their goal is to answer 0 if x is of distance $\leq \vartheta$ from M_0 and 1 if x is of distance $\leq \vartheta$ from M_1, (and any answer in any other case).

The problem $\mathcal{P}_1(\vartheta)$:
Player I gets as input a unit vector $x \in R^n$, and two orthogonal vector-spaces $M_0, M_1 \subset R^n$ of dimension $n/2$ each. Player II gets as input an orthogonal matrix $T : R^n \to R^n$, (i.e., a matrix T, such that $TT^\dagger = Id$, where Id is the identity matrix). Their goal is to answer 0 if $T(x)$ is of distance $\leq \vartheta$ from M_0 and 1 if $T(x)$ is of distance $\leq \vartheta$ from M_1, (and any answer in any other case).

The problems $\mathcal{P}_0(\vartheta), \mathcal{P}_1(\vartheta)$ themselves are not standard communication complexity problems, as the set of possible inputs for each player is infinite. Nevertheless, their communication complexity can still be defined in the same way.

The problem $\mathcal{P}_0(\vartheta)$ was first defined in [Kre] as a complete problem for one round quantum communication complexity. The problem $\mathcal{P}_1(\vartheta)$ is, as far as we know, new and is a complete problem for two rounds quantum communication complexity. The quantum communication complexity of both problems is $\Theta(\log n)$. Moreover, the problem $\mathcal{P}_0(\vartheta)$ can be solved by a quantum protocol with communication complexity $O(\log n)$ with only one round of communication. The problem $\mathcal{P}_1(\vartheta)$ can be solved by a quantum protocol with communication complexity $O(\log n)$ with only two rounds of communication.

Our main result is obtained by proving a lower bound of $\Omega(\sqrt{n})$ for the classical probabilistic communication complexity of $\mathcal{P}_1(\vartheta)$ (for any $0 < \vartheta < 1/\sqrt{2}$). Since the quantum communication complexity of that problem is $\Theta(\log n)$, this gives an exponential gap between the two models. For $\vartheta = 0$, the same lower bound of $\Omega(\sqrt{n})$ still holds, but we will have to assume in addition that the entire communication between the two players is described by nice functions (say Borel functions, or Lebesgue measurable functions).

As for upper bounds, we give an upper bound of $O(n^{1/2})$ for the classical probabilistic communication complexity of $\mathcal{P}_0(\vartheta)$ and an upper bound of $O(n^{3/4})$ for the classical probabilistic communication complexity of $\mathcal{P}_1(\vartheta)$ (for any $0 \leq \vartheta < 1/\sqrt{2}$). This shows that our lower bound is almost tight.

As mentioned above, the sets of inputs for the problems $\mathcal{P}_0(\vartheta), \mathcal{P}_1(\vartheta)$ are infinite. We define the finite precision variations of these problems in order to get communication complexity problems with finite sets of inputs (as is required by the standard definition of communication complexity). Each input for $\mathcal{P}_0(\vartheta), \mathcal{P}_1(\vartheta)$ can be described by $O(n^2)$ real variables. We define the problems $\widetilde{\mathcal{P}_0(\vartheta)}, \widetilde{\mathcal{P}_1(\vartheta)}$ to be the same as $\mathcal{P}_0(\vartheta), \mathcal{P}_1(\vartheta)$, but where each of these $O(n^2)$ variables is described by $O(\log n)$ bits (i.e., with polynomially good precision). The length of the inputs for these problems is hence $m = O(n^2 \log n)$.

For the problems $\widetilde{\mathcal{P}_0(\vartheta)}, \widetilde{\mathcal{P}_1(\vartheta)}$, we will have the same upper and lower bounds as before (for any $0 < \vartheta < 1/\sqrt{2}$). In particular, the quantum communication complexity of $\widetilde{\mathcal{P}_1(\vartheta)}$ is $O(\log m)$, and the classical probabilistic communication complexity of the same problem is $\Omega(m^{1/4}/\log m)$. We hence get an exponential gap for standard communication complexity problems as well (i.e., problems defined on finite sets of inputs.)

1.3 Techniques and Other Results

In order to prove our results, we prove several lemmas about subsets of the sphere S^{n-1} and about sets of orthogonal matrices. These lemmas may be interesting in theirs own right and they may help to obtain other results in the spirit of our main theorem. In this subsection, we will shortly describe some of these lemmas. The exact lemmas and their full proofs appear in Section 4.

The subsection will be added in the full version of the paper.

2 Models of Communication Complexity

In this section, we discuss with some more details the different models of communication complexity that we will use.

As mentioned above, a communication complexity problem is given by 3 finite sets X, Y, Z

and a relation $\mathcal{R} \subset X \times Y \times Z$. Player I is given an input $x \in X$. Player II is given an input $y \in Y$. Initially, non of the players has any information about the input of the other. Their goal is to come up with an answer z such that $(x, y, z) \in \mathcal{R}$. The communication complexity of the problem is the number of bits the two players have to exchange between them in order to come up with such an answer z.

As mentioned above, we would also like to consider cases where the sets X, Y are infinite. Therefore, in all that comes below, we do not assume that X, Y are finite. The definitions will hold for both cases.

In all the communication complexity problems discussed in this paper, the set Z of possible answers is $\{0, 1\}$, that is, there are only two possible answers. In all that comes below, let us hence assume for simplicity that $Z = \{0, 1\}$.

Let $\mathcal{R} \subset X \times Y \times Z$ and $\mathcal{R}' \subset X' \times Y' \times Z$ be two communication complexity problems. A reduction from $\mathcal{R}'$ to $\mathcal{R}$ is a pair of functions

$$f : X' \to X \quad , \quad g : Y' \to Y,$$

such that for every $(x', y') \in X' \times Y'$ and every $z \in Z$,

$$(f(x'), g(y'), z) \in \mathcal{R} \Longrightarrow (x', y', z) \in \mathcal{R}'.$$

Obviously, this implies that any protocol for $\mathcal{R}$ gives also a protocol for $\mathcal{R}'$. If there is a reduction from $\mathcal{R}'$ to $\mathcal{R}$ we will say that $\mathcal{R}'$ can be reduced to $\mathcal{R}$ and we will denote

$$\mathcal{R}' \prec \mathcal{R}.$$

If $\mathcal{R}', \mathcal{R}$ are defined as functions of a natural parameter n (that is, $\mathcal{R}', \mathcal{R}$ denote a sequence of problems, one for each n), we will use asymptotical notation, and denote

$$\mathcal{R}' \prec \mathcal{R}$$

if this is the case for $n > n_0$ (for some constant n_0).

2.1 Classical Probabilistic Communication Complexity

In classical probabilistic communication complexity we allow the two players to use classical probabilistic protocols. In each step of the protocol, one of the players sends one bit of information (about his input) to the other player. In the end, they both have to know an answer z.

For simplicity, let us assume that it is known in advance which player speaks in each step of the protocol (e.g., Player I sends the first bit and then they alternate). In each step of the protocol, the bit sent by a player may depend on the player's input and on all the messages already exchanged between the two players. The bit may also depend on a random string s, shared by both players.

The answer z is therefore a random variable (depending on the random string s). We require that for every input pair (x, y) the answer z obtained by the protocol satisfies

$$\mathrm{PROB}_s[(x, y, z) \in \mathcal{R}] \geq 1 - err,$$

where $err > 0$ is some small constant (the probability of error). The exact value of the constant err is of less importance (as long as that value is less than $1/2$) and it may change the communication complexity of a problem by only a multiplicative constant. This is true, because the probability of error can be efficiently reduced by repetition.

The maximum number of bits sent by the players in such a protocol is called the communication complexity of the protocol (where the maximum is taken over all the possible inputs). The probabilistic communication complexity of a problem is the communication complexity of the best such protocol for that problem. We identify the problem with the relation $\mathcal{R}$, and we denote the probabilistic communication complexity of $\mathcal{R}$ by $PCC(\mathcal{R})$ (for simplicity we remove the constant err from the notation, as it is of less importance).

Note that the two players share a random string s, i.e., they can both read s. That model is hence called the public-coins model. An alternative model is the private-coins model, where each of the two players has his own random string that cannot be read by the other player. Obviously, the private-coins model is weaker than the public-coins model. It is well known, however, that the private-coins model is only slightly weaker. More precisely, any problem that can be solved with communication complexity k by a public-coins protocol can also be solved with communication complexity $k + O(\log m)$ by a private-coins protocol (where m is the length of the inputs) [New]. Hence, the two models have roughly the same power. The results proved in this paper will be correct for both models.

2.2 Quantum Communication Complexity

In the quantum communication complexity model, the two players exchange between them quantum bits (qubits), rather than classical bits. For simplicity, let us assume that the protocol has k steps (rounds), and that in each step one of the players sends d qubits to the other player. We will assume for simplicity that Player I sends the d qubits in the first round and then they alternate.

Mathematically, we will use the following model: Let d, l be two integers. We think of l as the size of the quantum memory of each player, and we think of d as the size of a "quantum blackboard", shared by both players. The blackboard is used as a communication channel between the two players. The size of the memories is not limited, that is, l can be arbitrarily large.

We will have the following 3 sets:

$$L_1 = \{0,1\}^l \ \ , \ \ L_2 = \{0,1\}^l \ \ , \ \ D = \{0,1\}^d.$$

We think of L_1 as the set of classical assignments to Player I's memory, and we think of L_2 as the set of classical assignments to Player II's memory. We think of D as the set of classical assignments to the blackboard. The set $L_1 \times D \times L_2$ is hence the set of classical assignments to the entire system. Denote the elements of $L_1 \times D \times L_2$ by $e_0, ..., e_{2^{2l+d}-1}$. The element e_0, for example, will be the all 0 assignment to $L_1 \times D \times L_2$.

Denote by $\mathcal{C}$ the complex numbers. We will work with the 2^{2l+d} dimensional vector space

$$\Lambda = \mathcal{C}^{L_1 \times D \times L_2}.$$

We can define Λ also by

$$\Lambda = \mathcal{C}^{L_1} \otimes \mathcal{C}^{D} \otimes \mathcal{C}^{L_2},$$

where $\otimes$ denotes the tensor product. We think of Λ as the set of all pure quantum states of the system. Each element $e_i \in L_1 \times D \times L_2$ corresponds to a unit vector in Λ, with 1 in the i^{th} coordinate and 0 in all the other coordinates. Each such element e_i can hence be viewed also as a vector in Λ. The set $\{e_0, ..., e_{2^{2l+d}-1}\}$ is then the standard basis for Λ.

Let $\mathcal{U}$ be the set of all unitary operators on Λ. That is, the set of all linear operators $U : \Lambda \to \Lambda$, such that $UU^\dagger = Id$, (where Id is the identity operator). Each operator $U \in \mathcal{U}$ can be described by its action on the elements of the standard basis $\{e_0, ..., e_{2^{2l+d}-1}\}$.

Let $\mathcal{U}_1 \subset \mathcal{U}$ be the set of all such operators that act only on $\mathcal{C}^{L_1} \otimes \mathcal{C}^{D}$. Formally, we can define $\mathcal{U}_1$ by: $U \in \mathcal{U}_1$ iff

$$U = U' \otimes Id,$$

where U' is a unitary operator on $\mathcal{C}^{L_1} \otimes \mathcal{C}^{D}$ and Id is the identity operator on $\mathcal{C}^{L_2}$. We can hence think of $\mathcal{U}_1$ also as the set of unitary operators on $\mathcal{C}^{L_1 \times D}$. In the same way, let $\mathcal{U}_2 \subset \mathcal{U}$ be the set of all operators in $\mathcal{U}$ that act only on $\mathcal{C}^{D} \otimes \mathcal{C}^{L_2}$. We can think of $\mathcal{U}_2$ also as the set of unitary operators on $\mathcal{C}^{D \times L_2}$.

$\mathcal{U}_1$ will be the set of allowed operators for Player I. $\mathcal{U}_2$ will be the set of allowed operators for Player II. The players start from the initial vector $e_0 \in \Lambda$ (recall that e_0 corresponds to the all 0 assignment to $L_1 \times D \times L_2$). In each step of the protocol, one of the players apply a unitary operator from $\mathcal{U}_1$ or $\mathcal{U}_2$ respectively. In each step of the protocol, the operator used by a player may depend on the player's input (x or y respectively), but cannot depend on anything else. We will assume for simplicity that Player I applies the first operator and then they alternate.

Denote by U_1 the operator used by Player I in the first step and by U_2 the operator used by Player II in the second step, and so on (recall that U_1 is chosen as a function of x and U_2 is chosen as a function of y, and so on). The final state F is then defined by

$$F = U_k U_{k-1} \cdots U_2 U_1 e_0.$$

Note that since e_0 is a unit vector and since all operators are unitary, the final vector F is a unit vector as well.

The answer z is now determined by a measurement applied on F by one of the players. We assume for simplicity that the player that applies the measurement is the one that didn't apply the last operator U_k. That is, if k is odd then the measurement is applied by Player II, and if k is even then the measurement is applied by Player I.

Formally, the measurement is described by two orthogonal linear subspaces $M_0, M_1 \subset \Lambda$, of dimension 2^{2l+d-1} each. If the measurement is applied by Player I we require that the measurement is applied on $L_1 \times D$ only. Formally, this means that

$$M_0 = M'_0 \otimes \mathcal{C}^{L_2},$$

and

$$M_1 = M'_1 \otimes \mathcal{C}^{L_2},$$

where M'_0, M'_1 are two orthogonal subspaces of $\mathcal{C}^{L_1} \otimes \mathcal{C}^{D}$, of dimension 2^{l+d-1} each.

In the same way, if the measurement is applied by Player II we require that the measurement is applied on $D \times L_2$ only. Thus, if the measurement is applied by Player I we can think of M_0, M_1 as subspaces of $\mathcal{C}^{L_1 \times D}$ and we can think of the measurement as applied on $L_1 \times D$. If the measurement is applied by Player II we can think of M_0, M_1 as subspaces of $\mathcal{C}^{D \times L_2}$ and we can think of the measurement as applied on $D \times L_2$.

Now denote by λ_0 the length of the projection of the final state F on M_0 and denote by λ_1 the length of the projection of F on M_1. The answer z, given by the protocol, is defined to be 0 with probability λ_0^2 and 1 with probability λ_1^2. Note, that since F is a unit vector, we have

$$\lambda_0^2 + \lambda_1^2 = 1.$$

The answer z is hence a random variable. As before, we require that for every input pair (x, y),

$$\text{PROB}[(x, y, z) \in \mathcal{R}] \geq 1 - err,$$

where $err > 0$ is some small constant. (As before, the exact value of the constant err is of less importance, as it can be reduced efficiently). The communication complexity of the protocol is defined to be $k \cdot d$. The number of steps k is also called the number of rounds in the protocol.

The quantum communication complexity of a problem is the communication complexity of the best such protocol for that problem. As before, we identify the problem with the relation $\mathcal{R}$, and we denote the quantum communication complexity of $\mathcal{R}$ by $QCC(\mathcal{R})$.

This defines the mathematical model of quantum communication complexity. Let us add two comments about that model:

First comment: As mentioned above, we do not limit the memory size l. It is not hard to see, however, that w.l.o.g. we can assume that l is at most $k \cdot d$. To show that Player I can use memory size of at most $k \cdot d$, we can use the following argument: Fix x (the input for Player I). Assume by induction that after i steps, a memory of size $i \cdot d$ is enough. If i is odd then Player II applies the next operator U_{i+1}, and hence the same memory size (for Player I) is enough after $i+1$ steps as well. If i is even then Player I applies the operator U_{i+1}. Obviously, we only care about the action of U_{i+1} on the set of assignments to the previous memory and to the set D. Since that set of assignments is of size $\leq 2^{(i+1) \cdot d}$, the image of U_{i+1} on that set of assignments is of dimension $\leq 2^{(i+1) \cdot d}$, and hence it can be condensed into a memory of size $(i+1) \cdot d$.

Second comment: The above definition uses complex numbers and require all operators to be unitary operators (over the complex numbers). However, we can assume w.l.o.g. that all numbers used are real numbers and that all the operators used are orthogonal operators over the real numbers. This was first observed by [BV] for quantum computation in general (see also [Aha]) and is obviously true for quantum communication complexity as well. It can be proved simply by representing each complex number by two real numbers (one for the real part and one for the imaginary part). This requires adding only one qubit to the size of the blackboard.

3 Communication Complexity of $\mathcal{P}_0(\vartheta), \mathcal{P}_1(\vartheta), \widetilde{\mathcal{P}_0(\vartheta)}, \widetilde{\mathcal{P}_1(\vartheta)}$

First note that for any $0 \leq \vartheta < \vartheta' < 1/\sqrt{2}$, we have

$$\mathcal{P}_0(\vartheta) \prec \mathcal{P}_0(\vartheta')$$

and

$$\mathcal{P}_1(\vartheta) \prec \mathcal{P}_1(\vartheta'),$$

and also

$$\widetilde{\mathcal{P}_0(\vartheta)} \prec \widetilde{\mathcal{P}_0(\vartheta')}$$

and

$$\widetilde{\mathcal{P}_1(\vartheta)} \prec \widetilde{\mathcal{P}_1(\vartheta')}.$$

Since $\widetilde{\mathcal{P}_0(\vartheta)}$ and $\widetilde{\mathcal{P}_1(\vartheta)}$ are sub-problems of $\mathcal{P}_0(\vartheta)$ and $\mathcal{P}_1(\vartheta)$ respectively (i.e., they are the same problems defined on a subset of inputs), we have for any $0 < \vartheta < 1/\sqrt{2}$,

$$\widetilde{\mathcal{P}_0(\vartheta)} \prec \mathcal{P}_0(\vartheta)$$

and

$$\widetilde{\mathcal{P}_1(\vartheta)} \prec \mathcal{P}_1(\vartheta)$$

(simply by matching every input to itself). As for the other direction, for any $0 \leq \vartheta < \vartheta' < 1/\sqrt{2}$, we have

$$\mathcal{P}_0(\vartheta) \prec \widetilde{\mathcal{P}_0(\vartheta')}$$

and

$$\mathcal{P}_1(\vartheta) \prec \widetilde{\mathcal{P}_1(\vartheta')}$$

(by matching every input to the closest input in the finite precision variation of the problem. If there are several such elements, we will match the input to the smallest of them (say, according to the lexicographic order)).

The communication complexity of $\widetilde{\mathcal{P}_0(\vartheta)}$ and $\widetilde{\mathcal{P}_1(\vartheta)}$ is hence basically the same as the one of $\mathcal{P}_0(\vartheta)$ and $\mathcal{P}_1(\vartheta)$ respectively (in both models). Let us hence prove our results for the problems $\mathcal{P}_0(\vartheta)$ and $\mathcal{P}_1(\vartheta)$, and then conclude the same results for $\widetilde{\mathcal{P}_0(\vartheta)}$ and $\widetilde{\mathcal{P}_1(\vartheta)}$.

3.1 The Quantum Case

The problems $\mathcal{P}_0(\vartheta)$ and $\mathcal{P}_1(\vartheta)$ can be solved by quantum protocols with communication complexity $O(\log n)$ and with only one and two rounds of communication, respectively. This is formally stated by the following two propositions.

Proposition 3.1 *[Kre] For any $0 \leq \vartheta < 1/\sqrt{2}$, the problem $\mathcal{P}_0(\vartheta)$ can be solved by a quantum protocol with quantum communication complexity $O(\log n)$ and with only one round of communication (that is, with $k = 1$ in the above discussion).*

Proof:
Intuitively, Player I encodes the unit vector $x \in R^n$ by $\log_2 n$ qubits and sends these qubits to Player II. Player II then measures according to the vector spaces M_0, M_1.

Formally, the two players have a blackboard of size $d = \log_2 n$. The vector $x \in R^n$ is interpreted as $x \in \mathcal{C}^D$, where $D = \{0,1\}^d$. Player I applies some unitary operator U_1 with $U_1(e_0) = x$, where e_0 is the unit vector corresponding to the all 0 assignment to the blackboard. Player II then measures according to M_0, M_1, extended to be subspaces of $\mathcal{C}^D$. By the definition of the quantum communication complexity model, if x is of distance $\leq \vartheta$ from M_0 the answer will be 0 with probability $\geq 1 - \vartheta^2$, and if x is of distance $\leq \vartheta$ from M_1 the answer will be 1 with probability $\geq 1 - \vartheta^2$. □

Proposition 3.2 *For any $0 \leq \vartheta < 1/\sqrt{2}$, the problem $\mathcal{P}_1(\vartheta)$ can be solved by a quantum protocol with quantum communication complexity $O(\log n)$ and with only two rounds of communication (that is, with $k = 2$ in the above discussion).*

Proof:
As above, Player I applies some U_1 with $U_1(e_0) = x$. Player II extends T to be a unitary operator $U_2 : \mathcal{C}^D \to \mathcal{C}^D$ and apply U_2. Player I then measures according to M_0, M_1 (extended to be subspaces of $\mathcal{C}^D$). □

Moreover, for any $0 < \vartheta < 1/\sqrt{2}$, the problem $\mathcal{P}_0(\vartheta)$ is a complete problem for one round quantum communication complexity and the problem $\mathcal{P}_1(\vartheta)$ is a complete problem for two rounds quantum communication complexity. This is formally stated by the following two propositions.

Proposition 3.3 *[Kre] Let $\mathcal{R} \subset X \times Y \times Z$ be a communication complexity problem, and assume that there exists a quantum protocol for $\mathcal{R}$ with communication complexity d and with only one round of communication. Then, for any $0 < \vartheta < 1/\sqrt{2}$, there is a reduction from $\mathcal{R}$ to the problem $\mathcal{P}_0(\vartheta)$ with $n = 2^{O(d)}$.*

Proof:
By the second comment given at the end of Subsection 2.2, we can assume that the quantum protocol for $\mathcal{R}$ uses only real numbers (i.e., everything is over R^n). In that protocol, Player I applies some matrix U_1 (chosen as a function of x), and Player II measures according to some vector spaces M_0, M_1 (chosen as a function of y). Let $x' = U_1(e_0)$. Then x' is a unit vector

and we can use x' and the spaces M_0, M_1 as inputs for the problem $\mathcal{P}_0(\vartheta)$. Note that by the first comment given at the end of Subsection 2.2, we can assume that $l = O(d)$ and hence the parameter n of the problem $\mathcal{P}_0(\vartheta)$ satisfies $n = 2^{O(d)}$. $\square$

Proposition 3.4 *Let $\mathcal{R} \subset X \times Y \times Z$ be a communication complexity problem, and assume that there exists a quantum protocol for $\mathcal{R}$ with communication complexity d and with only two rounds of communication. Then, for any $0 < \vartheta < 1/\sqrt{2}$, there is a reduction from $\mathcal{R}$ to the problem $\mathcal{P}_1(\vartheta)$ with $n = 2^{O(d)}$.*

Proof:
As above, we assume that the quantum protocol for $\mathcal{R}$ uses real numbers. In that protocol, Player I applies some matrix U_1 (chosen as a function of x). Player II applies some matrix U_2 (chosen as a function of y). Player I then measures according to some vector spaces M_0, M_1 (chosen as a function of x). We can use the unit vector $x' = U_1(e_0)$, the operator $T = U_2$, and the spaces M_0, M_1 as inputs for the problem $\mathcal{P}_1(\vartheta)$. $\square$

Note that in Proposition 3.3 and Proposition 3.4, n is exponential in d. Recall, however, that by Proposition 3.1 and Proposition 3.2, the problems $\mathcal{P}_0(\vartheta)$ and $\mathcal{P}_1(\vartheta)$ have quantum protocols with communication complexity logarithmic in n.

To show that $\mathcal{P}_0(\vartheta)$, $\mathcal{P}_1(\vartheta)$ cannot be solved by better quantum protocols, we can use (for example) Proposition 3.3 and Proposition 3.4.

Proposition 3.5 *For any $0 < \vartheta < 1/\sqrt{2}$,*

$$QCC(\mathcal{P}_0(\vartheta)) \ , \ QCC(\mathcal{P}_1(\vartheta)) = \Omega(\log n).$$

Proof:
Otherwise, by Proposition 3.3 and Proposition 3.4, all problems with input size $d = \log n$ can be solved by quantum protocols with communication complexity $o(\log n)$. It is well known, however, that this is not the case (e.g., for the inner product problem, as proved in [Kre] (the origin of the proof is attributed there to Yao)). $\square$

3.2 Lower Bound for the Classical Probabilistic Communication Complexity of $\mathcal{P}_1(\vartheta)$

In this subsection we prove our main theorem, a lower bound of $\Omega(\sqrt{n})$ for the classical probabilistic communication complexity of $\mathcal{P}_1(\vartheta)$. Together with Proposition 3.5, the theorem gives the exponential separation between quantum and classical probabilistic communication complexity. The proof of the main lemma is deferred to Section 4.

Theorem 3.6 *For any $0 < \vartheta < 1/\sqrt{2}$,*

$$PCC(\mathcal{P}_1(\vartheta)) = \Omega(\sqrt{n}).$$

Proof:
Fix n. Denote by X the set of inputs for Player I and by Y the set of inputs for Player II. Denote $W = X \times Y$. The sets X, Y, W are compact manifolds (and each of them is a compact metric space with a transitive group of isometries). On each of them we have a standard notion of *the uniform measure*, which is the Haar's measure (see for example [MS] Chapter 1). We assume that the uniform measure is normalized to be a probability measure.

Let $Prot$ be some classical probabilistic communication protocol for $\mathcal{P}_1(\vartheta)$, and denote by k its communication complexity (i.e., the maximal number of bits exchanged between the two players by this protocol). Assume for a contradiction that $k = o(\sqrt{n})$. Recall that for any universal constant err, we can assume (by repetition) that the probability of error of the protocol $Prot$ is smaller than err. We will get a contradiction by showing that the probability of error of $Prot$ is larger than $\epsilon/2 - o(1)$, where ϵ is some universal constant.

For any input pair $(x, y) \in W$ and any assignment s to the random string of $Prot$, the string of communication bits exchanged by the two players on the inputs (x, y), using the random string s, is called the history of (x, y, s). For any $h \in \{0, 1\}^k$ and any assignment s to the random string of $Prot$, we denote by $W_{s,h} \subset W$ the set of all input pairs $(x, y) \in W$ such that the history of (x, y, s) is h.

It is well known (and easy to show) that for any s, h, the set $W_{s,h}$ is a product set, that is

$$W_{s,h} = X_{s,h} \times Y_{s,h},$$

where $X_{s,h} \subset X$ and $Y_{s,h} \subset Y$. Also, it is well known (and easy to show) that for any fixed s, the family $\{W_{s,h}\}_{h \in \{0,1\}^k}$ is a partition of W. That is,

1. For any s and any $h \neq h'$,
$$W_{s,h} \cap W_{s,h'} = \emptyset.$$
2. For any s,
$$\bigcup_h W_{s,h} = W.$$

The answer given by the protocol $Prot$ on (x, y, s) depends only on s, h. That is, for all input pairs in $W_{s,h}$ the answer given by the protocol $Prot$ on (x, y, s) will be the same. Let us denote that answer by $Prot(s, h)$.

In general, $X_{s,h} \subset X$ and $Y_{s,h} \subset Y$ can be arbitrary sets. In particular, they are not necessarily measurable. For our analysis, we will need to assume that for any s, h, the sets $X_{s,h}, Y_{s,h}$ are Borel sets (the entire argument seems to work for Lebesgue measurable sets as well, but in some places one has to be more careful with the measurability of the intersection of a measurable set and a sub-manifold of measure 0).

Claim 1 *W.l.o.g., we can assume that for any s, h, the sets $X_{s,h}, Y_{s,h}$ are Borel sets.*

Proof:
For the proof, it will help to use the finite precision variation $\widetilde{\mathcal{P}_1(\vartheta')}$. As discussed above, for

any $\vartheta'' < \vartheta' < \vartheta$, $\mathcal{P}_1(\vartheta'')$ can be reduced to $\widetilde{\mathcal{P}_1(\vartheta')}$ which, in its turn, can be reduced to $\mathcal{P}_1(\vartheta)$. Using the two reductions (as described above), we can translate $Prot$ into a protocol $Prot'$ for $\widetilde{\mathcal{P}_1(\vartheta')}$ and then translate it back into a protocol $Prot''$ for $\mathcal{P}_1(\vartheta'')$. For the definition of the new protocols $Prot', Prot''$, we only need to know the behaviour of $Prot$ on a finite set of inputs (the inputs for $\widetilde{\mathcal{P}_1(\vartheta')}$).

It is not hard to see that in every step of the protocol $Prot''$, the bit sent by a player can be described as a nice function (say Borel function) of the inputs and of the messages already exchanged between the two players. The sets $X_{s,h}, Y_{s,h}$ of the protocol $Prot''$ will hence be Borel sets. $\square$

Let us hence assume that for any s, h, the sets $X_{s,h}, Y_{s,h}$ are Borel sets.

Define $H_0 \subset W$ to be the set of all input pairs $((x, M_0, M_1), T) \in W$ such that $T(x) \in M_0$. Define $H_1 \subset W$ to be the set of all input pairs $((x, M_0, M_1), T) \in W$ such that $T(x) \in M_1$. The sets H_0, H_1 are compact manifolds as well (and each of them is a compact metric space with a transitive group of isometries), and (as before) on each of them we have a standard notion of *uniform measure*. As before, we assume that the uniform measure is normalized to be a probability measure. The uniform measures on H_0, H_1 are actually the ones induced on H_0, H_1 from the uniform measure on W. Obviously, $Prot$ has to answer 0 with probability $\geq 1 - err$, on all inputs in H_0, and 1 with probability $\geq 1 - err$, on all inputs in H_1.

For a Borel set $W' \subset W$, we denote by $\alpha(W')$ the measure of W' in W. We denote by $\beta_0(W')$ the measure of $W' \cap H_0$ in H_0, and we denote by $\beta_1(W')$ the measure of $W' \cap H_1$ in H_1. Theorem 3.6 will follow easily from the following lemma.

Lemma 3.7 *There exists a universal constant $\epsilon > 0$, s,t,. for any two Borel sets $X' \subset X$, $Y' \subset Y$,*

$$\beta_0(X' \times Y') \ , \ \beta_1(X' \times Y') \geq \epsilon \cdot \alpha(X' \times Y') - O(2^{-\sqrt{n}}).$$

Proof:
The lemma is proved in Section 4. It is restated there (in a slightly more exact form) as Corollary 4.8. $\square$

For any s, denote by $A_0(s) \subset W$ the union of all sets $W_{s,h}$, s.t., $Prot(s, h) = 0$ (i.e., the answer of the protocol is 0). For any s, denote by $A_1(s) \subset W$ the union of all sets $W_{s,h}$, s.t., $Prot(s, h) = 1$ (i.e., the answer of the protocol is 1). Then, for any s, the sets $A_0(s)$ and $A_1(s)$ are disjoint, and their union is W.

Since each of $A_0(s), A_1(s)$ is a union of at most 2^k of the sets $X_{s,h} \times Y_{s,h}$, we have by Lemma 3.7 for any s,

$$\beta_1(A_0(s)) \geq \epsilon \cdot \alpha(A_0(s)) - O(2^k \cdot 2^{-\sqrt{n}}),$$

and

$$\beta_0(A_1(s)) \geq \epsilon \cdot \alpha(A_1(s)) - O(2^k \cdot 2^{-\sqrt{n}}).$$

Hence,

$$\beta_1(A_0(s)) + \beta_0(A_1(s)) \geq \epsilon \cdot [\alpha(A_0(s)) + \alpha(A_1(s))] - o(1) = \epsilon - o(1).$$

But $\beta_1(A_0(s))$ is the fraction of inputs in H_1, s.t., the answer of $Prot$ on (x, y, s) is 0, and $\beta_0(A_1(s))$ is the fraction of inputs in H_0, s.t., the answer of $Prot$ on (x, y, s) is 1. Hence, for any s, the protocol makes an error on a fraction of at least $\epsilon/2 - o(1)$ of inputs in $H_0 \cup H_1$. When averaging over all the possibilities for s, we get that for at least one input pair $(x, y) \in H_0 \cup H_1$, the probability (over s) that the protocol makes an error on (x, y) is $\geq \epsilon/2 - o(1)$. □

3.3 Upper Bounds for the Classical Probabilistic Communication Complexity of $\mathcal{P}_0(\vartheta)$, $\mathcal{P}_1(\vartheta)$

We will now prove upper bounds for the classical probabilistic communication complexity of the two problems. The upper bounds are not needed for our main result and are added here for completness (and to show that our lower bound is almost tight). Let us start with the upper bound for $\mathcal{P}_0(\vartheta)$. For that problem we will prove an upper bound of $O(\sqrt{n})$.

Theorem 3.8 *For any* $0 < \vartheta < 1/\sqrt{2}$,

$$PCC(\mathcal{P}_0(\vartheta)) = O(\sqrt{n}).$$

Proof:
Let $z_1, ..., z_n$ be n mutually independent random elements in R, each chosen according to the Normal distribution $N(0, n^{-1/2})$, that is, each z_i has the Normal distribution with expectation 0 and variance n^{-1}. The distribution of $z = (z_1, ..., z_n)$ in R^n is hence multi-normal. Denote that distribution by ψ. Note that (for large n) with very high probability z is very close to a unit vector. We can hence think of z almost as a uniformly distributed random element of the unit sphere in R^n. (we prefer to use the distribution ψ because it is easier to analyze).

Let $k = 2^{const \cdot n^{1/2}}$, where $const$ is a large enough constant. Let $z^1, ..., z^k$ be k mutually independent random elements of R^n, each chosen according to the distribution ψ. We think of these variables as common for both players. That is, we assume that both players can see $z^1, ..., z^k$ (this can be assumed since the two players can interpret their common random string s as the k random elements $z^1, ..., z^k$, with arbitrarily good precision).

The protocol will be the following: Let $y = z^j$ be the element in $\{z^1, ..., z^k\}$ with the largest scalar product with the input vector x. Player I knows the value of the index j and sends that index to Player II. Now Player II knows the vector y. Player II will answer 0 if y is closer to M_0 and 1 if y is closer to M_1.

The complete analysis of the protocol will be given in the full version of the paper.

□

For the problem $\mathcal{P}_1(\vartheta)$ we will prove an upper bound of $O(n^{3/4})$.

Theorem 3.9 *For any* $0 < \vartheta < 1/\sqrt{2}$,

$$PCC(\mathcal{P}_1(\vartheta)) = O(n^{3/4}).$$

Proof:
The proof is by a generalization of the previous protocol. Let ψ be the same distribution as above.

Let $k = 2^{const \cdot n^{3/4}}$, where *const* is a large enough constant. Let $z^1, ..., z^k$, $\hat{z}^1, ..., \hat{z}^k$ be $2k$ mutually independent random elements of R^n, each chosen according to the distribution ψ, (and assume that both players can see $z^1, ..., z^k$ and $\hat{z}^1, ..., \hat{z}^k$).

The protocol will be the following: Let $y = z^j$ be the element in $\{z^1, ..., z^k\}$ with the largest scalar product with the input vector x. Player I knows the value of the index j and sends that index to Player II. Now Player II knows the vector y and computes $\hat{x} = T(y)$. Let $\hat{y} = \hat{z}^{\hat{j}}$ be the element in $\{\hat{z}^1, ..., \hat{z}^k\}$ with the largest scalar product with $\hat{x}$. Player II knows the value of the index $\hat{j}$ and sends that index to Player I. Now Player I knows the vector $\hat{y}$. Player I will answer 0 if $\hat{y}$ is closer to M_0 and 1 if $\hat{y}$ is closer to M_1.

The complete analysis of the protocol will be given in the full version of the paper.

$\square$

3.4 Communication Complexity of $\widetilde{\mathcal{P}_0(\vartheta)}, \widetilde{\mathcal{P}_1(\vartheta)}$

We can now conclude the same results for the problems $\widetilde{\mathcal{P}_0(\vartheta)}, \widetilde{\mathcal{P}_1(\vartheta)}$. Recall that the length of inputs for these problems is $m = O(n^2 \cdot \log n)$.

In particular, the quantum communication complexity of $\widetilde{\mathcal{P}_1(\vartheta)}$ is $O(\log n) = O(\log m)$, and the classical probabilistic communication complexity of the same problem is $\Omega(n^{1/2}) = \Omega(m^{1/4}/\log m)$. We hence get an exponential gap for standard communication complexity problems as well (i.e., problems defined on finite sets of inputs.)

4 Proof of the Main Lemma

In this section, we give the proof of Lemma 3.7 (restated here as Corollary 4.8). In order to prove Corollary 4.8, we prove several other lemmas that may be of independent interest.

4.1 Notations

Recall that we denote by R^n the n dimensional vector space over the real numbers R. We denote by S^{n-1} the unit sphere in R^n, that is,

$$S^{n-1} = \left\{ (x_1, ..., x_n) \in R^n \;\middle|\; \sum_{i=1}^{n} x_i^2 = 1 \right\}.$$

By $AF_{k,r}$ we denote the set of all affine subspaces (of R^n) of dimension k and of distance r from the 0 vector (for simplicity we remove n from the notation). The set $AF_{k,0}$ is the set of all linear subspaces (of R^n) of dimension k (that is, the Grassman manifold, which is sometimes denoted by $G_{n,k}$). By O_n we denote the set of all orthogonal matrices of size $n \times n$.

All these sets are compact manifolds (and each of them is a compact metric space with a transitive group of isometries). On each of them we have, as before, a standard notion of *the uniform measure* (which is the Haar's measure). We assume that the uniform measure is normalized to be a probability measure. We denote the uniform measure by μ.

If B is such a manifold and $A \subset B$ is a Borel set, we denote by $\mu(A|B)$ the measure of A in B. If S is such a manifold and $A \subset B \subset S$ are Borel sets, $\mu(A|B)$ still denotes the measure of A in B (when the measure on B is the one induced from S). In particular, if $\mu(B|S) > 0$ we will have $\mu(A|B) = \mu(A|S)/\mu(B|S)$. Sometimes B will be such a manifold by itself and then the two notations will agree. (In general, the uniform measure on a manifold will always agree with the one on a sub-manifold). Also, sometimes A will not be a subset of B and then $\mu(A|B)$ denotes the measure of $A \cap B$ in B, that is, $\mu(A|B) = \mu(A \cap B|B)$.

When we have such a manifold S (e.g., S^{n-1}) and we say that x is a random element uniformly distributed on S, we mean that x is chosen according to the uniform measure μ. Sometimes, we just say that x is a random element of S and we mean to the same thing. Sometimes, it is clear what the manifold S is, and we just say that x is *uniformly chosen among all the possibilities*.

By *measurable set*, we mean Borel set. The entire argument seems to work for Lebesgue measurable sets as well (but in some places one has to be a little bit more careful with the measurability of certain sets). When we say *subspace*, we mean linear subspace (i.e., vector space). When we mean to affine subspace we say explicitly *affine subspace*. By *distance*, we usually mean to the Euclidean distance (in R^n). Sometimes we mean to the Geodesic distance (usually on the sphere), and then we say explicitly *Geodesic distance*.

Given k vectors $y_1, ..., y_k \in R^n$, we denote by $SP(y_1, ..., y_k)$ or $SP(\bar{y})$ the span of $y_1, ..., y_k$, that is, the smallest linear subspace containing $y_1, ..., y_k$. By $Af(y_1, ..., y_k)$ or $Af(\bar{y})$ we denote the smallest affine subspace containing $y_1, ..., y_k$.

For $y_1, ..., y_k \in S^{n-1}$, we say that $y_1, ..., y_k$ are r-orthogonal if for every $1 \leq i \leq k-1$, the projection of y_{i+1} on $SP(y_1, ..., y_i)$ is of length $\leq r$.

4.2 Some Properties of Subsets of S^{n-1}

Recall that $\mu(A|B)$ denotes the measure of $A \cap B$ in B, that is, $\mu(A|B) = \mu(A \cap B|B)$.

The following lemma shows that if $C \subset S^{n-1}$ is a measurable set of size c, and V is a random subspace (of R^n) with large enough dimension, then with very high probability the size of $C \cap V$ in $S^{n-1} \cap V$ is more or less c.

Lemma 4.1 *Let $C \subset S^{n-1}$ be a measurable set of size c, (i.e., with $\mu(C|S^{n-1}) = c$). Let $V \subset R^n$ be a random subspace of dimension k, (uniformly chosen among all such subspaces). Then, for any $\epsilon > 0$,*

$$PROB_V[\, |\mu(C|S^{n-1} \cap V) - c| \geq \epsilon \,] \; < \; (4/\epsilon) \cdot e^{-(\epsilon^2/2) \cdot k}.$$

Proof:
For every subspace V' of dimension k, denote,

$$\gamma(V') = \mu(C | S^{n-1} \cap V'),$$

that is, $\gamma(V')$ is the relative size of C in the $k-1$ dimensional unit sphere $V' \cap S^{n-1}$. We will prove that,

$$\mathrm{PROB}_V[\gamma(V) \geq c+\epsilon], \mathrm{PROB}_V[\gamma(V) \leq c-\epsilon] \ < \ (2/\epsilon) \cdot e^{-(\epsilon^2/2)\cdot k}.$$

The main idea of the proof is to choose the random subspace V by choosing k random elements $y_1, ..., y_k \in S^{n-1}$ and taking V to be their span. The proof will follow from the fact that with very high probability the number of elements $y_1, ..., y_k$ that are contained in C is very close to $c \cdot k$.

Let $y_1, ..., y_k$ be k random elements of S^{n-1}, chosen independently according to the uniform distribution. Then for every i, $\mathrm{PROB}[y_i \in C] = c$, and the events $y_i \in C$ are mutually independent. Denote by Y the number of indices $i \in \{1, ..., k\}$ with $y_i \in C$. Then the random variable Y has the binomial distribution $B(k, c)$. By the standard Chernoff's bounds (see for example [AS] Appendix A, Theorem A.4), for any $a > 0$

$$\mathrm{PROB}[|Y - c \cdot k| > a \cdot k] < e^{-2a^2 \cdot k}.$$

Recall that $SP(\bar{y})$ is the span of $y_1, ..., y_k$, that is, the smallest vector space containing $y_1, ..., y_k$. With probability 1 the dimension of $SP(\bar{y})$ is k. Let us hence assume that the dimension of $SP(\bar{y})$ is indeed k. Then, by a symmetry argument it is clear that the random variable $SP(\bar{y})$ has the uniform distribution over all subspaces of dimension k.

Let V' be some specific subspace of dimension k. Given the event $SP(\bar{y}) = V'$, it is not true that $y_1, ..., y_k$ are mutually independent random variables over $S^{n-1} \cap V'$ (unless $k = n$) — e.g., $y_1, ..., y_k$ have a tendency to be more orthogonal than k mutually independent uniformly distributed random elements of $S^{n-1} \cap V'$. Nevertheless, it is true that each y_i is uniformly distributed over $S^{n-1} \cap V'$. Hence by the additivity of expectation,

$$\mathbf{E}[Y \mid SP(\bar{y}) = V'] = \gamma(V') \cdot k.$$

Therefore, since Y is always at most k, for any $b > 0$ we have by Markov's inequality

$$\mathrm{PROB}[Y > (\gamma(V') - b) \cdot k \mid SP(\bar{y}) = V'] > b.$$

Hence, for any $a > 0$ and any space V' with $\gamma(V') \geq c + 2a$, we can fix $b = \gamma(V') - (c+a) \geq a$ and get

$$\mathrm{PROB}[Y > (c+a) \cdot k \mid SP(\bar{y}) = V'] > a.$$

Since this is true for any such space V', we get for any $a > 0$,

$$\mathrm{PROB}[Y > (c+a) \cdot k \mid \gamma(SP(\bar{y})) \geq c + 2a] > a.$$

We can now bound

$$\text{PROB}[Y > (c+a)\cdot k] \geq$$
$$\text{PROB}[Y > (c+a)\cdot k \mid \gamma(SP(\bar{y})) \geq c+2a] \;\cdot\; \text{PROB}[\gamma(SP(\bar{y})) \geq c+2a]$$
$$> a \cdot \text{PROB}[\gamma(SP(\bar{y})) \geq c+2a].$$

But we have proved that

$$\text{PROB}[|Y - c\cdot k| > a\cdot k] < e^{-2a^2\cdot k}.$$

Therefore, for any $a > 0$,

$$\text{PROB}[\gamma(SP(\bar{y})) \geq c+2a] < (1/a)\cdot e^{-2a^2\cdot k}.$$

Since $SP(\bar{y})$ is a random subspace of dimension k, uniformly chosen among all such subspaces (i.e., $SP(\bar{y})$ is a random variable with the same distribution as V), we can conclude that for any $a > 0$,

$$\text{PROB}_V[\gamma(V) \geq c+2a] < (1/a)\cdot e^{-2a^2\cdot k}.$$

In the same way, since Y is always at least 0, for any $b > 0$ we have by Markov's inequality

$$\text{PROB}[Y < (\gamma(V')+b)\cdot k \mid SP(\bar{y}) = V'] > b,$$

and by the same argument we have for any $a > 0$,

$$\text{PROB}_V[\gamma(V) \leq c-2a] < (1/a)\cdot e^{-2a^2\cdot k}.$$

The lemma is now proved by fixing $\epsilon = 2a$. □

Recall that $AF_{k,r}$ is the set of all affine subspaces (of R^n) of dimension k and of distance r from the 0 vector. The following lemma generalizes the previous one to the case of affine subspaces of a short distance from the 0 vector.

Lemma 4.2 *Let $C \subset S^{n-1}$ be a measurable set of size c, (i.e., with $\mu(C|S^{n-1}) = c$). Let V be a (uniformly distributed) random affine subspace in $AF_{k,r}$. Then, for any $\epsilon > 0$,*

$$PROB_V[\,|\mu(C|S^{n-1}\cap V) - c| \geq \epsilon\,] \;<\; (8/\epsilon)\cdot e^{-(\epsilon^2/32)\cdot k'} + (128\cdot k'/\epsilon^2)\cdot e^{-(\epsilon^2/64)\cdot n},$$

where $k' = MIN[\lceil n/2\rceil, k, \lceil 1/r^2\rceil] - 1$.

Proof:
The proof goes according to the same lines as the proof of Lemma 4.1. The main difference will be that here we will choose only k' random vectors, $y_1, ..., y_{k'}$, and we will require $y_1, ..., y_{k'}$ to be mutually orthogonal. We will then choose the random affine subspace V by choosing a random affine subspace containing $y_1, ..., y_{k'}$.

For every affine subspace $V' \in AF_{k,r}$, denote,

$$\gamma(V') = \mu(C|S^{n-1}\cap V'),$$

that is, $\gamma(V')$ is the relative size of C in $S^{n-1} \cap V'$. We will prove that,

$$\text{PROB}[\gamma(V) \geq c + \epsilon], \text{PROB}[\gamma(V) \leq c - \epsilon] < (4/\epsilon) \cdot e^{-(\epsilon^2/32) \cdot k'} + (64 \cdot k'/\epsilon^2) \cdot e^{-(\epsilon^2/64) \cdot n}.$$

Let $y_1, ..., y_{k'}$ be k' mutually orthogonal random elements of S^{n-1}. We will choose $y_1, ..., y_{k'}$ by the following inductive procedure: First choose a (uniformly distributed) random element $y_1 \in S^{n-1}$. Then, given $y_1, ..., y_i$, denote by U_i the $n - i$ dimensional vector space orthogonal to $y_1, ..., y_i$ and choose y_{i+1} to be a (uniformly distributed) random element of the $n - i - 1$ dimensional sphere $S^{n-1} \cap U_i$.

Denote by Y the number of indices $i \in \{1, ..., k'\}$ with $y_i \in C$. Using Lemma 4.1, it will not be hard to show that, as before, with very high probability Y is very close to $c \cdot k'$.

Claim 2 *For any $a > 0$,*

$$PROB[|Y - c \cdot k'| > a \cdot k'] < 2 \cdot e^{-(a^2/8) \cdot k'} + (16 \cdot k'/a) \cdot e^{-(a^2/16) \cdot n}.$$

Proof:
Let us prove that

$$\text{PROB}[Y - c \cdot k' > a \cdot k'] < e^{-(a^2/8) \cdot k'} + k' \cdot (8/a) \cdot e^{-(a^2/16) \cdot n}.$$

The inequality

$$\text{PROB}[Y - c \cdot k' < -a \cdot k'] < e^{-(a^2/8) \cdot k'} + k' \cdot (8/a) \cdot e^{-(a^2/16) \cdot n}$$

follows in the same way.

Define z_i by $z_i = 1$ if $y_i \in C$ and $z_i = 0$ otherwise. For any i, U_i is a random vector space of dimension $n - i \geq n - k' \geq n/2$. Hence by Lemma 4.1, for any ϵ_0, with probability larger than

$$1 - (4/\epsilon_0) \cdot e^{-(\epsilon_0^2/4) \cdot n}$$

(where the probability is over all the possibilities for fixing $y_1, ..., y_i$), we have

$$\mu(C | S^{n-1} \cap U_i) < c + \epsilon_0$$

and hence after fixing $y_1, ..., y_i$,

$$\text{PROB}_{y_{i+1}}[y_{i+1} \in C \mid y_1, ..., y_i] < c + \epsilon_0.$$

With probability larger than

$$1 - k' \cdot (4/\epsilon_0) \cdot e^{-(\epsilon_0^2/4) \cdot n}$$

the same is satisfied for every $1 \leq i \leq k'$.

Hence, ignoring a bad event that occurs with probability smaller than $k' \cdot (4/\epsilon_0) \cdot e^{-(\epsilon_0^2/4) \cdot n}$, we have that for every i, after fixing $z_1, ..., z_i$,

$$\text{PROB}_{z_{i+1}}[z_{i+1} = 1 \mid z_1, ..., z_i] < c + \epsilon_0.$$

Now define

$$Y_j = \sum_{i=1}^{j} [z_i - (c + \epsilon_0)].$$

Then the sequence $Y_1, ..., Y_{k'}$ is a submartingale, and hence satisfies the same large deviation inequalities as the sum of independent variables. In particular, for any $\epsilon_1 > 0$, the variable $Y = Y_{k'} + (c + \epsilon_0) \cdot k'$ satisfies

$$\text{PROB}[Y > (c + \epsilon_0) \cdot k' + \epsilon_1 \cdot k'] < e^{-({\epsilon_1}^2/2) \cdot k'}$$

(see for example [AS] Chapter 7, Theorem 2.1. The bound is stated there for martingales, but obviously the same is true for submartingales as well).

Thus, without ignoring the bad event (that occurs with small probability) we have

$$\text{PROB}[Y > (c + \epsilon_0) \cdot k' + \epsilon_1 \cdot k'] < e^{-({\epsilon_1}^2/2) \cdot k'} + k' \cdot (4/\epsilon_0) \cdot e^{-(\epsilon_0^2/4) \cdot n}.$$

We can now fix $\epsilon_0 = \epsilon_1 = a/2$ and get

$$\text{PROB}[Y > (c + a) \cdot k'] < e^{-(a^2/8) \cdot k'} + k' \cdot (8/a) \cdot e^{-(a^2/16) \cdot n}.$$

$\square$

Recall that $Af(\bar{y})$ is the smallest affine subspace containing $y_1, ..., y_{k'}$. Then $Af(\bar{y})$ is of dimension $k' < k$. Since $y_1, ..., y_{k'}$ were required to be orthogonal, the distance of $Af(\bar{y})$ from the 0 vector is $1/\sqrt{k'} \geq r$. Hence, $Af(\bar{y})$ is included in at least one affine subspace in $AF_{k,r}$. Choose V to be one such affine subspace (uniformly chosen among all the possibilities). That is, $V \in AF_{k,r}$ is a random affine subspace with $Af(\bar{y}) \subset V$. By a symmetry argument it is clear that the random variable V has the uniform distribution over $AF_{k,r}$.

The rest of the proof is as in the proof of Lemma 4.1. Let V' be some specific affine subspace in $AF_{k,r}$. Given the event $V = V'$, each y_i is uniformly distributed over $S^{n-1} \cap V'$. Hence by the additivity of expectation,

$$\mathbf{E}[Y \mid V = V'] = \gamma(V') \cdot k'.$$

Therefore, since Y is always at most k', for any $b > 0$ we have by Markov's inequality

$$\text{PROB}[Y > (\gamma(V') - b) \cdot k' \mid V = V'] > b.$$

Hence, for any $a > 0$ and any space V' with $\gamma(V') \geq c + 2a$, we can fix $b = \gamma(V') - (c + a) \geq a$ and get

$$\text{PROB}[Y > (c + a) \cdot k' \mid V = V'] > a.$$

Since this is true for any such space V', we get for any $a > 0$,

$$\text{PROB}[Y > (c + a) \cdot k' \mid \gamma(V) \geq c + 2a] > a.$$

We can now bound

$$\text{PROB}[Y > (c+a)\cdot k'] \geq$$
$$\text{PROB}[Y > (c+a)\cdot k' \mid \gamma(V) \geq c+2a] \;\cdot\; \text{PROB}[\gamma(V) \geq c+2a]$$
$$> a \cdot \text{PROB}[\gamma(V) \geq c+2a].$$

But we have proved that

$$\text{PROB}[|Y - c\cdot k'| > a\cdot k'] < 2\cdot e^{-(a^2/8)\cdot k'} + (16\cdot k'/a)\cdot e^{-(a^2/16)\cdot n}.$$

Therefore, for any $a > 0$,

$$\text{PROB}[\gamma(V) \geq c+2a] < (2/a)\cdot e^{-(a^2/8)\cdot k'} + (16\cdot k'/a^2)\cdot e^{-(a^2/16)\cdot n}.$$

In the same way, since Y is always at least 0, for any $b > 0$ we have by Markov's inequality

$$\text{PROB}[Y < (\gamma(V') + b)\cdot k' \mid V = V'] > b,$$

and by the same argument we have for any $a > 0$,

$$\text{PROB}[\gamma(V) \leq c-2a] < (2/a)\cdot e^{-(a^2/8)\cdot k'} + (16\cdot k'/a^2)\cdot e^{-(a^2/16)\cdot n}.$$

The lemma is now proved by fixing $\epsilon = 2a$. □

An alternative proof for Lemma 4.2 goes according to the following lines: Let $y_1, ..., y_{k'}$ be k' random elements of S^{n-1} chosen independently according to the uniform distribution, (as in the proof of Lemma 4.1). Recall that $Af(\bar{y})$ is the smallest affine subspace containing $y_1, ..., y_{k'}$, and show that with very high probability $Af(\bar{y})$ is of dimension k' and of distance more than r from the 0 vector. Hence, with very high probability $Af(\bar{y})$ is included in at least one affine subspace in $AF_{k,r}$. Choose V to be one such affine subspace (uniformly chosen among all the possibilities). As before, the random variable V has the uniform distribution over $AF_{k,r}$. The rest is the same as in the proofs of Lemma 4.1, Lemma 4.2.

Let us now restate Lemma 4.2 in a slightly weaker form that will be easier to use. The proof is immediate from Lemma 4.2.

Corollary 4.3 *Let $C \subset S^{n-1}$ be a measurable set of size c, (i.e., with $\mu(C|S^{n-1}) = c$). Let V be a (uniformly distributed) random affine subspace in $AF_{k,r}$. Then, for any $\epsilon > 0$ there exist $n_0, const > 0$ (depending only on ϵ) such that if $n > n_0$ then*

$$PROB_V[\;|\mu(C|S^{n-1}\cap V) - c| \geq \epsilon\;] \;<\; const\cdot e^{-(k'/const)},$$

where $k' = MIN[k, 1/r^2]$.

Recall that for $x_1, ..., x_m \in S^{n-1}$, and for $r > 0$, we say that $x_1, ..., x_m$ are r-orthogonal if for every $1 \leq i \leq m-1$, the projection of x_{i+1} on $SP(x_1, ..., x_i)$ is of length $\leq r$. The following lemma shows that if $C \subset S^{n-1}$ is not too small then C contains $\lfloor n/2 \rfloor$ almost orthogonal elements.

Lemma 4.4 *Let $C \subset S^{n-1}$ be a measurable set of size $\geq e^{-(\delta^2/4)\cdot n}$, i.e., with*

$$\mu(C|S^{n-1}) \geq e^{-(\delta^2/4)\cdot n},$$

where $\delta > 0$ and $(\delta^2/2) \cdot n \geq \ln n$. Then, there exist $x_1, ..., x_m \in C$, such that $x_1, ..., x_m$ are 3δ-orthogonal, and $m = \lfloor n/2 \rfloor$.

Proof:
For a set $C' \subset S^{n'-1}$, and for any $\delta' > 0$, denote by $C'_{\delta'} \subset S^{n'-1}$ the set of all elements of geodesic distance $\leq \delta'$ from the set C'.

Claim 3 *Let $C' \subset S^{n'-1}$ be a measurable set of size*

$$\mu(C'|S^{n'-1}) \geq \sqrt{\pi/8} \cdot e^{-(\delta'^2/2)\cdot n'}$$

(for some $\delta' > 0$). Then,

$$\mu(C'_{3\delta'}|S^{n'-1}) \geq 1 - \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'}.$$

Proof:
Assume w.l.o.g. that

$$\mu(C'|S^{n'-1}) = \sqrt{\pi/8} \cdot e^{-(\delta'^2/2)\cdot n'}.$$

By the classical isoperimetric inequality on the sphere (see for example [MS] Chapter 2, Theorem 2.1), the smallest possible $\mu(C'_{3\delta'}|S^{n'-1})$ is obtained when C' is a cap of a suitable radius. Hence, w.l.o.g. we can assume that $C', C'_{3\delta'}$ are caps.

Now, we can just use the well known analysis for caps. For example, we can use a double application of the following claim that can be found in [MS] Chapter 2, Corollary 2.2: If $C_0 \subset S^{n'-1}$ is a cap with

$$\mu(C_0|S^{n'-1}) \geq \sqrt{\pi/8} \cdot e^{-(\delta_0^2/2)\cdot n'},$$

and $C_1 \subset S^{n'-1}$ is a cap with $\mu(C_1|S^{n'-1}) \geq 1/2$, then the geodesic distance between C_0 and C_1 is at most δ_0. We need to use that claim twice: The first time with $C_0 = C'$ and $\delta_0 = \delta'$, and the second time with $C_0 = S^{n'-1} \setminus C'_{3\delta'}$ and $\delta_0 = 2\delta'$. □

For an element $x \in R^{n'}$, denote by $U_x \subset R^{n'}$ the $n' - 1$ dimensional subspace orthogonal to x.

Claim 4 *Let $C' \subset S^{n'-1}$ be a measurable set of size*

$$\mu(C'|S^{n'-1}) \geq \sqrt{\pi/8} \cdot e^{-(\delta'^2/2)\cdot n'}$$

(for some $\delta' > 0$). Then, there exists $x \in C'_{3\delta'}$ such that

$$\mu(C'|S^{n'-1} \cap U_x) \geq \mu(C'|S^{n'-1}) - \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'}.$$

Proof:
By a symmetry argument we have

$$\mu(C'|S^{n'-1}) = \mathbf{E}_x[\mu(C'|S^{n'-1} \cap U_x)],$$

where x is a (uniformly distributed) random element of $S^{n'-1}$. But if the claim is false then (using also Claim 3) we can derive the following contradiction

$$\mathbf{E}_x[\mu(C'|S^{n'-1} \cap U_x)] \leq$$

$$\text{PROB}[x \in C'_{3\delta'}] \cdot \left(\mu(C'|S^{n'-1}) - \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'}\right) \; + \; \text{PROB}[x \in S^{n'-1} \setminus C'_{3\delta'}] \cdot 1 =$$

$$\mu(C'_{3\delta'}|S^{n'-1}) \cdot \left(\mu(C'|S^{n'-1}) - \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'}\right) \; + \; \mu(S^{n'-1} \setminus C'_{3\delta'}|S^{n'-1}) \cdot 1 <$$

$$\left(\mu(C'|S^{n'-1}) - \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'}\right) + \sqrt{\pi/8} \cdot e^{-(2\delta'^2)\cdot n'} \; = \; \mu(C'|S^{n'-1}).$$

□

Lemma 4.4 is now proved by a repetitive application of Claim 4. $x_m, ..., x_1$ will be defined in the following way: Define $x \in C_{3\delta}$ to be an element with the maximal $\mu(C|S^{n-1} \cap U_x)$. By Claim 4,

$$\mu(C|S^{n-1} \cap U_x) \geq \mu(C|S^{n-1}) - \sqrt{\pi/8} \cdot e^{-(2\delta^2)\cdot n}.$$

Define $x_m \in C$ to be an element in C, closest to x, (hence the geodesic distance between x_m and x is at most 3δ). Define $\hat{C} = C \cap S^{n-1} \cap U_x$. Now, to define $x_{m-1}, ..., x_1$, repeat recursively the same construction for the set $\hat{C}$ in the $n-2$ dimensional sphere $S^{n-1} \cap U_x$.

The elements $x_{m-1}, ..., x_1$ are all contained in $\hat{C}$ and hence also in U_x. Therefore, x is orthogonal to all of them. Hence, since x_m is of geodesic distance of at most 3δ from x, its projection on $SP(x_1, ..., x_{m-1})$ is of length $< 3\delta$. Therefore, if we are able to apply the recursive construction for m steps we will get that $x_1, ..., x_m$ are 3δ-orthogonal. Thus we just have to check that this is possible. More precisely, we have to show that after applying the recursive construction up to m times, the obtained set $\hat{C}$ is still not empty (and hence we are able to continue the recursive construction).

In the i^{th} step (of the recursive construction), we have to apply the construction for a set $\hat{C}$ in an $n-i$ dimensional sphere. Denote by S_i the $n-i$ dimensional sphere in the i^{th} step, and by C^i the set $\hat{C}$ in that step. We will prove by induction that in all the steps the conditions of Claim 4 are satisfied. That is, we will prove that for every i,

$$\mu(C^i|S_i) \geq \sqrt{\pi/8} \cdot e^{-(\delta^2/2)\cdot(n/2)}$$

(recall that $n-i > n/2$). This ensures that the set C^i is not empty, (and hence that we are able to continue the recursive construction).

Assuming the inductive hypothesis that the last inequality is satisfied for every $j \leq i$, we have by Claim 4 that for every $j \leq i$,

$$\mu(C^{j+1}|S_{j+1}) \geq \mu(C^j|S_j) - \sqrt{\pi/8} \cdot e^{-(2\delta^2)\cdot(n/2)}.$$

Since

$$\mu(C^1|S_1) = \mu(C|S^{n-1}) \geq e^{-(\delta^2/4)\cdot n},$$

and since we assumed that $(\delta^2/2) \cdot n \geq \ln n$, we can conclude that

$$\mu(C^{i+1}|S_{i+1}) \geq e^{-(\delta^2/4)\cdot n} - (n/2) \cdot \sqrt{\pi/8} \cdot e^{-(2\delta^2/2)\cdot n} \geq$$

$$e^{-(\delta^2/4)\cdot n} - (n/2) \cdot \sqrt{\pi/8} \cdot e^{-(\delta^2/2)\cdot n} \cdot e^{-\ln n} >$$

$$\left(1 - (1/2) \cdot \sqrt{\pi/8}\right) \cdot e^{-(\delta^2/4)\cdot n} > \sqrt{\pi/8} \cdot e^{-(\delta^2/4)\cdot n},$$

that is, the inductive hypothesis is satisfied for $i + 1$ as well. □

4.3 Some Properties of Subsets of O_n

Recall that O_n is the set of all orthogonal matrices of size $n \times n$. The following lemma shows that if the image of a set T of orthogonal matrices is small on each of m almost orthogonal elements of S^{n-1} then T is exponentially small in m.

Lemma 4.5 *For $m \leq n/2$, let $x_1, ..., x_m \in S^{n-1}$ be r-orthogonal. For every $1 \leq i \leq m$, let $C_i \subset S^{n-1}$ be a measurable set of size $1/2$, i.e.,*

$$\mu(C_i|S^{n-1}) = 1/2.$$

Let $T \subset O_n$ be the set of all orthogonal matrices $t \in O_n$, such that for every $1 \leq i \leq m$, $t(x_i) \in C_i$. Then, for some universal constants $n_0, const > 0$, if $n > n_0$ then

$$\mu(T|O_n) < const \cdot e^{-(k/const)},$$

where $k = MIN[m, 1/r^2]$.

Proof:
The lemma will follow by Corollary 4.3. For every $2 \leq j \leq m$, define r_j to be the length of the projection of x_j on $SP(x_1, ..., x_{j-1})$. Then for every j, $r_j \leq r$. For every $1 \leq j \leq m$, define $T_j \subset O_n$ to be the set of all orthogonal matrices t, such that for every $1 \leq i \leq j$, $t(x_i) \in C_i$. Obviously,

$$T = T_m \subset T_{m-1} \subset \cdots \subset T_2 \subset T_1.$$

Let t be a (uniformly distributed) random element of O_n. It is easy to see that t can be chosen in the following way: First choose a random element $y_1 \in S^{n-1}$ and fix $t(x_1) = y_1$.

Then, inductively, after $t(x_1),...,t(x_j)$ were already fixed to $y_1,...,y_j$, denote by $U_{j+1} \subset R^n$ the affine space of all $y \in R^n$ such that for every $1 \leq i \leq j$, the angle between y and y_i is the same as the one between x_{j+1} and x_i. (The set $U_{j+1} \cap S^{n-1}$ is therefore the set of all the possibilities to fix $t(x_{j+1})$, in a way that still keeps t orthogonal). Now choose y_{j+1} to be a random element of $U_{j+1} \cap S^{n-1}$, and fix $t(x_{j+1}) = y_{j+1}$. Finally, after $t(x_1),...,t(x_m)$ were fixed to $y_1,...,y_m$ (and hence the linear transformation t is already defined on $SP(x_1,...x_m)$), extend t to R^n in a random way that keeps t orthogonal (uniformly chosen among all the possibilities).

By a symmetry argument, for every j, the affine space U_{j+1} is a random element of $AF_{n-j,r_{j+1}}$. Hence, by Corollary 4.3 for some $n_0, const_0$, if $n > n_0$ then

$$\text{PROB}_{y_1,...,y_j}[\mu(C_{j+1}|S^{n-1} \cap U_{j+1}) \geq 3/4] \;<\; const_0 \cdot e^{-(k'/const_0)},$$

where $k' \geq \text{MIN}[n/2, 1/r^2]$. That is, for all but a $const_0 \cdot e^{-(k'/const_0)}$ fraction of the possibilities to fix $t(x_1),...,t(x_j)$ to $y_1,...,y_j$ we will have

$$\text{PROB}_{y_{j+1}}[y_{j+1} \in C_{j+1}|t(x_1) = y_1,...,t(x_j) = y_j] < 3/4.$$

Therefore, unless

$$\mu(T_j|O_n) < 2 \cdot const_0 \cdot e^{-(k'/const_0)},$$

we will have

$$\mu(T_{j+1}|O_n) < (7/8) \cdot \mu(T_j|O_n).$$

If for some j,

$$\mu(T_j|O_n) < 2 \cdot const_0 \cdot e^{-(k'/const_0)}$$

the lemma follows by the fact that $T = T_m \subset T_j$. Otherwise, we have by induction

$$\mu(T_m|O_n) < (7/8)^m,$$

and the lemma follows as well. □

The following lemma generalizes the previous one to the case where the image of T on each of the m almost orthogonal elements is of small entropy.

Lemma 4.6 *For $m \leq n/2$, let $x_1,...,x_m \in S^{n-1}$ be r-orthogonal. For every $1 \leq i \leq m$, let $C_i \subset S^{n-1}$ be a measurable set of size $1/2$, i.e.,*

$$\mu(C_i|S^{n-1}) = 1/2.$$

Let $T \subset O_n$ be a set of orthogonal matrices, and let t be a (uniformly distributed) random element of T. Assume that for every i,

$$PROB_t[t(x_i) \in C_i] > 1 - \epsilon,$$

(for some $\epsilon \geq 0$). Then, for some universal constants $n_0, \epsilon_0, const > 0$, if $n > n_0$ and $\epsilon < \epsilon_0$ then

$$\mu(T|O_n) < const \cdot e^{-(k/const)},$$

where $k = MIN[m, 1/r^2]$.

Proof:
W.l.o.g., assume that $1/r^2$ is integer and that $m = 1/r^2$, and hence $k = m$. (Otherwise, we can just consider the first k elements $x_1, ..., x_k$.) For every i, define D_i to be the complement to C_i in S^{n-1}, and define t_i to be 1 if $t(x_i) \in C_i$ and 0 if $t(x_i) \in D_i$. Denote by $\mathbf{H}$ the entropy function for random variables, and by $h(\epsilon)$ the entropy of the distribution $(\epsilon, 1-\epsilon)$, that is,

$$h(\epsilon) = -[\epsilon \cdot \log_2 \epsilon + (1-\epsilon) \cdot \log_2 (1-\epsilon)].$$

Then,

$$\mathbf{H}(t_1, ..., t_m) \leq \sum_{i=1}^{m} \mathbf{H}(t_i) \leq m \cdot h(\epsilon) \leq m \cdot h(\epsilon_0).$$

Hence, for at least one assignment $a_1, ..., a_m$ to $t_1, ..., t_m$,

$$\text{PROB}_t[t_1 = a_1, ..., t_m = a_m] \geq 2^{-m \cdot h(\epsilon_0)}.$$

W.l.o.g., assume that that assignment is the all 0 assignment (i.e., for every i, $a_i = 0$), and hence,

$$\text{PROB}_t[\forall i, t(x_i) \in C_i] \geq 2^{-m \cdot h(\epsilon_0)}.$$

Define $\hat{T} \subset T$ to be the set of all $\hat{t} \in T$ such that $\forall i, \hat{t}(x_i) \in C_i$. Then,

$$\text{PROB}_t[t \in \hat{T}] \geq 2^{-m \cdot h(\epsilon_0)},$$

and hence

$$\mu(T|O_n) \leq \mu(\hat{T}|O_n) \cdot 2^{m \cdot h(\epsilon_0)}.$$

We can now apply Lemma 4.5 for the set $\hat{T}$, and get

$$\mu(\hat{T}|O_n) < const_0 \cdot e^{-(m/const_0)},$$

where $const_0$ is the universal constant $const$ from Lemma 4.5 (recall that we assume $k = m$). Therefore, we have

$$\mu(T|O_n) < const_0 \cdot e^{-(m/const_0)} \cdot 2^{m \cdot h(\epsilon_0)}.$$

Hence, for small enough constants ϵ_0, $const$ we get

$$\mu(T|O_n) < const \cdot e^{-(k/const)}.$$

□

4.4 The Main Lemma

Let G_n be the manifold of all vector spaces of dimension $n/2$ in R^n (we assume for simplicity that n is even, and remove for simplicity the dimension $n/2$ from the notation). For every $u \in G_n$, denote $\hat{u} = u \cap S^{n-1}$. Let $W_n = S^{n-1} \times G_n \times O_n$. Let $H_n \subset W_n$ be the manifold of all $(x, u, t) \in W_n$ such that $t(x) \in u$. For a set $C \subset S^{n-1}$, let $I_C : S^{n-1} \to \{0, 1\}$ be 1 on C and 0 outside C. The main lemma will follow easily from the following lemma.

Lemma 4.7 *Let $Q \subset S^{n-1} \times G_n$, and $T \subset O_n$ be measurable sets. Then, for some universal constants $n_0, const, \epsilon > 0$, if $n > n_0$ and*

$$\mu(Q|S^{n-1} \times G_n), \mu(T|O_n) \geq const \cdot e^{-(\sqrt{n}/const)}$$

then

$$\mu(Q \times T|H_n) \geq \epsilon \cdot \mu(Q \times T|W_n).$$

Proof:
For every $x \in S^{n-1}$, denote

$$Q(x) = \{u \in G_n \mid (x, u) \in Q\}.$$

Denote by Q_X the set of all $x \in S^{n-1}$ such that $Q(x)$ is not empty.

Claim 5 *W.l.o.g., we can assume that for every $x \in Q_X$,*

$$\mu(Q(x)|G_n) > (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)}.$$

Proof:
Denote,

$$B_1 = \left\{x \in S^{n-1} \;\middle|\; \mu(Q(x)|G_n) \leq (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)}\right\},$$

and

$$Z_1 = \{(x, u) \in Q \mid x \in B_1\}.$$

Then, obviously

$$\mu(Z_1|S^{n-1} \times G_n) \leq (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)}.$$

Now denote

$$Q_1 = Q \setminus Z_1,$$

and for every $x \in S^{n-1}$, denote

$$Q_1(x) = \{u \in G_n \mid (x, u) \in Q_1\}.$$

Then we have

$$\mu(Q_1|S^{n-1} \times G_n) \geq (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)},$$

and for every $x \in S^{n-1}$, either $Q_1(x)$ is empty or

$$\mu(Q_1(x)|G_n) > (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)}.$$

We can now prove the lemma for the sets Q_1, T, and get

$$\mu(Q \times T|H_n) \geq \mu(Q_1 \times T|H_n) \geq \epsilon \cdot \mu(Q_1 \times T|W_n) \geq (1/2) \cdot \epsilon \cdot \mu(Q \times T|W_n).$$

(note that since we just have to prove the existence of constants $n_0, const, \epsilon$, we have the freedom to change *const* to $2 \cdot const$, and ϵ to $\epsilon/2$). □

Let t be a (uniformly distributed) random element of T. For every $x \in S^{n-1}$, $t(x)$ is a random variable with some distribution over S^{n-1}. Since $\mu(T|O_n) > 0$, there exists a (distribution density) function $\tau_x : S^{n-1} \to R$, such that for every measurable subset $C \subset S^{n-1}$,

$$\mathrm{PROB}_t[t(x) \in C] = \mathbf{E}_{y \in S^{n-1}}[I_C(y) \cdot \tau_x(y)],$$

where $y \in S^{n-1}$ means that y is a (uniformly distributed) random element of S^{n-1}, and $I_C : S^{n-1} \to \{0,1\}$ is 1 on C and 0 outside C.

For every $x \in S^{n-1}$, denote by $C_x \subset S^{n-1}$ the set of all $y \in S^{n-1}$ such that $\tau_x(y) \geq \epsilon_0$, where ϵ_0 is the universal constant ϵ_0 from Lemma 4.6. For every $x \in S^{n-1}$, denote by D_x the complement to C_x in S^{n-1}, that is, $D_x = S^{n-1} \setminus C_x$. Obviously, for every x,

$$\mathrm{PROB}_t[t(x) \in D_x] = \mathbf{E}_{y \in S^{n-1}}[I_{D_x}(y) \cdot \tau_x(y)] < \epsilon_0.$$

Hence, for every x,

$$\mathrm{PROB}_t[t(x) \in C_x] > 1 - \epsilon_0.$$

Let $B \subset S^{n-1}$ be the set of all $x \in S^{n-1}$ with $\mu(C_x|S^{n-1}) < 1/2$.

Claim 6 *If the constants $n_0, const$ are large enough then*

$$\mu(B|S^{n-1}) < e^{-(\sqrt{n}/4)}.$$

Proof:
Let $\delta = n^{-1/4}$. Assume for a contradiction that

$$\mu(B|S^{n-1}) \geq e^{-(\delta^2/4)\cdot n}.$$

By Lemma 4.4, there exist $x_1, ..., x_m \in B$, such that $x_1, ..., x_m$ are r-orthogonal, where $r = 3\delta$ and $m = \lfloor n/2 \rfloor$.

For every $1 \leq i \leq m$,

$$\mu(C_{x_i}|S^{n-1}) < 1/2.$$

For every such i, let $C_i \subset S^{n-1}$ be some measurable set of size exactly $1/2$, such that $C_{x_i} \subset C_i$. Then, for every i,

$$\mathrm{PROB}_t[t(x_i) \in C_i] > \mathrm{PROB}_t[t(x_i) \in C_{x_i}] > 1 - \epsilon_0.$$

By Lemma 4.6, for some universal constants $n_1, const_0 > 0$, if $n > n_1$ then

$$\mu(T|O_n) < const_0 \cdot e^{-(\sqrt{n}/(9 \cdot const_0))}.$$

If $const \geq 9 \cdot const_0$ we get a contradiction to the inequality

$$\mu(T|O_n) \geq const \cdot e^{-(\sqrt{n}/const)}.$$

$\square$

Claim 7 *W.l.o.g., we can assume that $B \cap Q_X$ is empty, and hence for every $x \in Q_X$,*

$$\mu(C_x | S^{n-1}) \geq 1/2.$$

Proof:
Otherwise denote,

$$Q_1 = Q \setminus (B \times G_n).$$

By Claim 6, if $const \geq 4$,

$$\mu(Q_1 | S^{n-1} \times G_n) \geq (1/2) \cdot \mu(Q | S^{n-1} \times G_n).$$

As in the proof of Claim 5, we can now prove the lemma for Q_1, T, and get

$$\mu(Q \times T | H_n) \geq \mu(Q_1 \times T | H_n) \geq \epsilon \cdot \mu(Q_1 \times T | W_n) \geq (1/2) \cdot \epsilon \cdot \mu(Q \times T | W_n).$$

□

For the rest of the proof let (x, u) be a (uniformly distributed) random element of Q. Obviously, $x \in Q_X$ with probability 1. The ratio $\mu(Q \times T | H_n)/\mu(Q \times T | W_n)$ can be written as

$$\frac{\mu(Q \times T | H_n)}{\mu(Q \times T | W_n)} = \mathbf{E}_{(x,u)} \mathbf{E}_{y \in \hat{u}} [\tau_x(y)],$$

where $y \in \hat{u}$ means that y is a uniformly distributed random element of $u \cap S^{n-1}$. We will prove that for small enough ϵ,

$$\mathbf{E}_{(x,u)} \mathbf{E}_{y \in \hat{u}} [\tau_x(y)] \geq \epsilon.$$

This is proved in the following way:

$$\mathbf{E}_{(x,u)} \mathbf{E}_{y \in \hat{u}} [\tau_x(y)] \geq \mathbf{E}_{(x,u)} \mathbf{E}_{y \in \hat{u}} [\epsilon_0 \cdot I_{C_x}(y)] = \epsilon_0 \cdot \mathbf{E}_{(x,u)} \mathbf{E}_{y \in \hat{u}} [I_{C_x}(y)]$$
$$= \epsilon_0 \cdot \mathbf{E}_{(x,u)} [\mu(C_x | \hat{u})] = \epsilon_0 \cdot \mathbf{E}_x \mathbf{E}_u [\mu(C_x | \hat{u})].$$

Since with probability 1 we have $x \in Q_X$, it is enough to prove that for every fixed $x \in Q_X$,

$$\mathbf{E}_u [\mu(C_x | \hat{u})] \geq 1/4.$$

Since by Claim 5 for every $x \in Q_X$

$$\mu(Q(x) | G_n) > (1/2) \cdot const \cdot e^{-(\sqrt{n}/const)},$$

and since for fixed x, the space u is a uniformly distributed random element of $Q(x)$, the required bound on $\mathbf{E}_u [\mu(C_x | \hat{u})]$ follows immediately from Lemma 4.1. □

Let $X \subset S^{n-1} \times G_n \times G_n$ be the manifold of all triples $(x, u_0, u_1) \in S^{n-1} \times G_n \times G_n$, such that u_0 is orthogonal to u_1. (X is the set of inputs for Player I). Let $Y = O_n$. (Y is the set of inputs for Player II). Let $W = X \times Y$. Let $H_0 \subset W$ be the manifold of all $((x, u_0, u_1), t) \in W$ such that $t(x) \in u_0$. Let $H_1 \subset W$ be the manifold of all $((x, u_0, u_1), t) \in W$ such that $t(x) \in u_1$. Lemma 3.7 can now be restated by the following corollary.

Corollary 4.8 *Let $X' \subset X$, and $Y' \subset Y$ be measurable sets. Then, for some universal constants $n_0, const, \epsilon > 0$, if $n > n_0$ then*

$$\mu(X' \times Y' | H_0) \geq \epsilon \cdot \mu(X' \times Y' | W) - const \cdot e^{-(\sqrt{n}/const)}$$

and in the same way,

$$\mu(X' \times Y' | H_1) \geq \epsilon \cdot \mu(X' \times Y' | W) - const \cdot e^{-(\sqrt{n}/const)}.$$

Proof:
The proof is immediate by Lemma 4.7. Note that for every subspace u_0 of dimension $n/2$ there is exactly one orthogonal subspace u_1 of dimension $n/2$. Hence, the manifold X is equivalent to $S^{n-1} \times G_n$, as in Lemma 4.7. Note that the requirements

$$\mu(X'|X), \mu(Y'|Y) \geq const \cdot e^{-(\sqrt{n}/const)}$$

(from Lemma 4.7) are not needed here, as we reduced (from the right hand side) a term of $const \cdot e^{-(\sqrt{n}/const)}$. □

Acknowledgments

I am grateful to Avi Wigderson for numerous of very helpful conversations in all stages of this work.

References

[Aha] D. Aharonov, "Quantum Computation", to appear in *Annual Reviews of Computational Physics*, ed. Dietrich Stauffer, World Scientific, vol VI, 1998. (Also in quant-ph/9812037).

[AS] N. Alon and J.H. Spencer, *The Probabilistic Method*, John Wiley & Sons Inc., 1992.

[ASTVW] A. Ambainis, L. Schulman, A. Ta-Shma, U. Vazirani and A. Wigderson, "The quantum communication complexity of sampling", *Proc. of the 39th IEEE Symp. Found. of Computer Science*, 1998.

[BBCJPW] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channells", *Phys. Rev. Lett.*, 70, pp.1895–1898, 1993.

[BCW] H. Buhrman, R. Cleve and A. Wigderson, "Quantum vs. classical communication and computation", *Proc. 30th Ann. ACM Symp. Theor. Comput.*, 1998. (Also in quant-ph/9802040).

[BPMWZ] D. Bouwmeester, J-W. Pan, K. Mattle, H. Weinfurter and A. Zeilinger, "Experimental quantum teleportation", *Nature*, 390, pp.575–579, 1997.

[BV] E. Bernstein and U. Vazirani, "Quantum complexity theory", *SIAM J. of Comput.*, 26 (5), pp.1411–1473, 1997. (Also in STOC 1993).

[Gro] L.K. Grover, "Quantum mechanics helps in searching for a needle in a haystack", *Phys. Rev. Lett.*, 79, pp.325–328, 1997. (Also in quant-ph/9706033 and STOC 1996).

[Hol] A.S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel", *Problemy Peredachi Informatsii*, 9 (3) pp.3–11, 1973. English translation: *Problems of Information Transmission*, 9, pp.177–183, 1973.

[HADLMS] R.J. Hughes, D.M. Alde, P. Dyer, G.G. Luther, G.L. Morgan and M. Schauer, "Quantum cryptography", *Contemp. Phys.*, 36, pp.149–163, 1995.

[Kre] I. Kremer, *Quantum Communication*, MSc Thesis, Computer Science Department, The Hebrew University, 1995.

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

[KS] B. Kalyanasundaram and G. Schnitger, "The probabilistic communication complexity of set intersection", *SIAM J. on Disc. Math.*, 5 (4), pp. 545–557, 1992. (Also in Structures in complexity theory 1987).

[KW] M. Karchmer and A. Wigderson, "Monotone circuits for connectivity require super-logarithmic depth", *SIAM J. on Disc. Math.*, 3 (2), pp.255–265, 1990. (Also in STOC 1998).

[MS] V.D. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Springer-Verlag, 1986.

[MWKZ] K. Mattle, H. Weinfurter, P.G. Kwiat and A. Zeilinger, "Dense coding in experimental quantum communication", *Phys. Rev. Lett.*, 76, pp.4656–4659, 1996.

[New] I. Newman, "Private vs. common random bits in communication complexity", *Information Processing Letters*, 39, pp.67–71, 1991.

[Pre] J. Preskill, Lecture notes on quantum information and quantum computation, http://www.theory.caltech.edu/people/preskill/ph229/

[Razb] A. A. Razborov, "On the Distributional Complexity of Disjointness", *Theoretical Computer science*, 106 (2), pp.385–390, 1992. (Also in ICALP 1990).

[Sho] P.W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", *SIAM J. of Comput.*, 26 (5), pp.1484–1509, 1997. (Also in FOCS 1994).

[Yao1] A. C. C. Yao, "Some complexity questions related to distributive computing," *Proc. 11th Ann. ACM Symp. Theor. Comput.*, 209–213, 1979.

[Yao2] A.C.C. Yao, "Quantum circuit complexity", *Proc. of the 34th IEEE Symp. Found. of Computer Science*, pp. 352–361, 1993.