# Relations between Communication Complexity, Linear Arrangements, and Computational Complexity

Jürgen Forster* Matthias Krause† Satyanarayana V. Lokam‡
Rustam Mubarakzjanov§ Niels Schmitt* Hans U. Simon*

**Abstract**

Recently, Forster [9] proved a new lower bound on probabilistic communication complexity in terms of the operator norm of the communication matrix. In this paper, we want to exploit the various relations between communication complexity of distributed Boolean functions, geometric questions related to half space representations of these functions, and the computational complexity of these functions in various restricted models of computation. In order to widen the range of applicability of Forster's bound, we start with the derivation of a generalized lower bound. We present a concrete family of distributed Boolean functions where the generalized bound leads to a linear lower bound on the probabilistic communication complexity (and thus to an exponential lower bound on the number of Euclidean dimensions needed for a successful half space representation), whereas the old bound fails. We move on to a geometric characterization of the well known communication complexity class C-PP in terms of half space representations achieving a large margin. Our characterization hints to a close connection between the bounded error model of probabilistic communication complexity and the area of large margin classification. In the final two sections of the paper, we describe how our techniques can be used to prove exponential lower bounds on the size of depth-2 threshold circuits (with still some technical restrictions) and on the size of Read-k-times Probabilistic Ordered Binary Decision Diagrams. The second bound easily generalizes to so-called weakly ordered Branching Programs.

## 1 Introduction

Linear algebraic techniques play a pervasive role in the study of computational complexity. In particular, several matrix functions related to rank and its robustness under various changes to the matrix have been extensively used in two-party communication complexity, many models of circuits, decision trees, branching programs, and span programs. Often such functions arise from or are equivalent to nice, natural geometric questions about the matrices. This paper studies certain geometric realizations of real matrices (with no zero entries) and applies results about such realizations of explicit matrices to derive lower bounds in communication complexity, threshold circuits, and ordered binary decision diagrams.

*Lehrstuhl Mathematik & Informatik, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany. Email: {forster, simon, nschmitt}@lmi.ruhr-uni-bochum.de

†Lehrstuhl für Theoretische Informatik, Institut für Informatik, Universität Mannheim, 68131 Mannheim, Germany. Email: krause@th.informatik.uni-mannheim.de

‡School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. Partially supported by NSF Grants CCR-9988359 and DMS-9729992. Email: satya@math.ias.edu

§Lehrstuhl für Theoretische Konzepte und neue Anwendungen der Informatik, Fakultät für Informatik, Universität Trier, 54286 Trier, Germany. Email: rustam@TI.uni-trier.de

The main mathematical problem we investigate is as follows. Let $M \in \mathbb{R}^{X \times Y}$ be a matrix with non-zero entries. We say $M$ can be *realized* by a $k$-dimensional linear arrangement if there are unit vectors $u_x, v_y \in \mathbb{R}^k$ for $x \in X$ and $y \in Y$ such that for all $(x, y) \in X \times Y$, $\mathrm{sign}(M_{xy}) = \mathrm{sign}(\langle u_x, v_y \rangle)$. The two functions of interest are **i)** the *minimal dimension* of $M$, defined to be the minimal $k$ such that $M$ can be realized by a $k$-dimensional linear arrangement, and **ii)** the *maximal margin* of $M$, defined to be the maximum over all realizations (of all dimensions) $\{u_x, v_y : x \in X, y \in Y\}$ of $\min\{|\langle u_x, v_y \rangle| : x \in X, y \in Y\}$.

The minimal dimension can be interpreted in terms of rank. It is easy to see that the minimal dimension of a matrix $M \in \mathbb{R}^{X \times Y}$ is the minimal rank of any real matrix $M' \in \mathbb{R}^{X \times Y}$ such that for all $(x, y) \in X \times Y$, $\mathrm{sign}(M_{x,y}) = \mathrm{sign}(M'_{xy})$. Note that the minimal dimension and minimal rank depend only on the sign-pattern of $M$ but not on its actual (non-zero) entries. Hence, we can ask the following question: given a sign-pattern (a matrix with entries $\pm 1$), what is the smallest rank of a matrix that obeys this sign-pattern (agrees with the given matrix in sign)? Stated differently, the minimal dimension of a $\pm 1$-matrix describes the robustness of its rank with respect to sign-preserving changes.

We review here, briefly, some history of the problem. Paturi and Simon [17] introduced the model of *unbounded error probabilistic communication complexity*. They showed that the problem of estimating the minimal dimension of matrices with $\pm 1$ entries is essentially equivalent to estimating the unbounded error probabilistic communication complexity of the corresponding Boolean function. Alon, Frankl, and Rödl [2] showed that for *almost all* $n \times n$ matrices with $\pm 1$ entries, the minimal dimension is at least $\Omega(n)$ implying that for most Boolean functions the unbounded error probabilistic communication complexity is asymptotically as large as it can be (linear in the length of the input). However, proving even a superlogarithmic lower bound on the minimal dimension of an explicit matrix remained a difficult open question. Recently, Forster [9] solved this long standing open question. He showed a general lower bound on the minimal dimension of a matrix in terms of its operator norm. As a corollary, he derived a lower bound of $\sqrt{n}$ on the minimal dimension of an $n \times n$ Hadamard matrix. This implies a linear lower bound on the unbounded error probabilistic communication complexity of the `inner product mod 2` function. Forster's results [9] also include upper bounds on maximal margins of explicit matrices. The problem of maximal margins is motivated by the excellent empirical performance of maximal margin classifiers [7] (these are learning algorithms that calculate the hyperplane with largest margin on a sample and use that hyperplane to classify new instances). As outlined in [4], the status of the problem for maximal margins was quite similar to the status of the problem for minimal dimensions: although the maximal margin of most matrices $M \in \{-1, +1\}^{X \times Y}$ is (provably) not substantially larger than the trivial margin (which can be shown to be $\max\{|X|^{-1/2}, |Y|^{-1/2}\}$), the question of proving such an upper bound for explicit quadratic matrices has been open. Forster's result gives an upper bound on the maximal margin of a matrix in terms of its operator norm. For the Hadamard matrix, this general bound yields the best possible upper bound on the maximal margin (matching the trivial lower bound).

In this paper, we first generalize Forster's bound [9] on the minimal dimension. These general bounds apply to any real matrix with non-zero entries, not just $\pm 1$ matrices. We demonstrate this generality by showing strong bounds for a class of matrices for which Forster's [9] old bound fails.

Next, we give a characterization of the communication complexity class C-PP (analog of the well-known Turing Machine complexity class PP) in terms of maximal margins. We show that C-PP is exactly the class of languages whose communication matrices (with $\pm 1$ entries) have a maximal margin of $2^{-\mathrm{polylog}(n)}$, where $n$ is the length of the input known to the two processors. An interesting ingredient of our proof is a random projection technique introduced by Arriaga and Vempala [3] in the context of learning algorithms. A result due to Halstenberg and Reischuk

[13] characterizes C-PP in terms of a probabilistic *one-way* protocol that uses at most $2^{\mathrm{polylog}(n)}$ messages and achieves an error bound of $2^{-\mathrm{polylog}(n)}$. Here, we give a characterization of C-PP in terms of a single parameter that is directly related to the communication matrix.

Our third result are lower bounds on depth-2 threshold circuits in terms of minimal dimension. Proving superpolynomial lower bounds for explicit functions when the threshold gates are allowed to use arbitrary weights is an interesting open question. Hajnal et al. [12] show an exponential lower bound for the `inner product mod 2` function when all the weights used in a depth-2 threshold circuit are polynomially bounded. Here, we strengthen their result by showing that the restriction on the weights of the top gate can be removed. We use lower bounds on minimal dimensions of explicit matrices to derive exponential lower bounds for some explicit functions including `inner product mod 2`. In fact, our lower bounds are exponential when the depth-2 circuit has a threshold gate (with unrestricted weights) at the top and either threshold gates with polynomial weights or gates computing arbitrary symmetric functions at the bottom. Our results also generalize and strengthen the results of Bruck and Smolensky [6], Krause [14], and Krause and Pudlák [15].

Finally, our fourth result is a lower bound on Probabilistic Ordered Binary Decision Diagrams (POBDD's). Using the lower bound on the minimal dimension of a Hadamard matrix, we construct an explicit function (containing the `inner product mod 2` as a subfunction) that requires exponential size POBDD's. An Ordered Binary Decision Diagram (OBDD) is a restricted model of a branching program that has been extensively studied both in complexity theory and in applications such as automatic verification.[1] While exponential lower bounds are already known on the size of OBDD's, this paper gives, to our knowledge, the first exponential lower bound on the size of *Probabilistic* OBDD's for an explicitly given function.

## 2 Definitions and Notations

A matrix $M \in \mathbb{R}^{X \times Y}$ induces the function $f(x,y) = M_{x,y}$ and vice versa. Since we often consider the situation that a function $f(x,y)$ is computed by two processors (one of them with $x$ as local input, the other one with $y$ as local input), we speak of a "distributed function". We often blur the distinction between matrices and distributed functions. In other words, we interpret a distributed function as matrix (or vice versa) whenever we find it convenient.

**Notions and Facts from Linear Algebra** We assume some familiarity with Linear Algebra and Matrix Theory. We use $L_p$ to denote the $L_p$-norm in $\mathbb{R}^n$. We will be mainly concerned with the $L_2$-norm (Euclidean norm) and the $L_1$-norm (Manhattan norm). $S^{n-1}$ denotes the $(n-1)$-dimensional unit sphere, i.e., $S^{n-1} := \{u \in \mathbb{R}^n : L_2(u) = 1\}$.
The *operator norm* of a matrix $M \in \mathbb{R}^{m \times n}$ is defined as follows:

$$\|M\| := \max\{L_2(Mu) : u \in S^{n-1}\} \tag{1}$$

It is well known that for every matrix $M$ the equalities $\|M\| = \|M^\top\|$ and $\|MM^\top\| = \|M\|^2$ hold. If $M \in \mathbb{R}^{X \times X}$ is symmetric, then $\|M\|$ coincides with the largest eigenvalue of $M$. If $M \in \mathbb{R}^{X \times X}$ is orthogonal, then $\|M\| = 1$. If $M \in \mathbb{R}^{X \times Y}$ has pairwise orthogonal columns, then $\|M\|$ coincides with the Euclidean length of the longest row vector of $M$. If $M \in \mathbb{R}^{X \times Y}$ has pairwise orthogonal rows, then $\|M\|$ coincides with the Euclidean length of the longest column vector of $M$.

[1] The reader interested in more information about OBDD's is referred to the book of Ingo Wegener [19], where also many pointers to the literature can be found.

**Linear Arrangements, Minimal Dimension, and Maximal Margin** We say that a matrix $M \in \mathbb{R}^{X \times Y}$ with no zero entries *can be realized by a k-dimensional linear arrangement* if there are vectors $u_x, v_y \in \mathbb{R}^k$ for $x \in X$, $y \in Y$ such that $\mathrm{sign}(M_{x,y}) = \mathrm{sign}(\langle u_x, v_y \rangle)$ for all $x \in X$, $y \in Y$. The *margin* achieved by this arrangement is defined as

$$\min_{x \in X, y \in Y} \frac{|\langle u_x, v_y \rangle|}{L_2(u_x) \cdot L_2(v_y)} \ . \tag{2}$$

Often we assume w.l.o.g. that the vectors $u_x$ and $v_y$ are normalized, i.e., $L_2(u_x) = L_2(v_y) = 1$. In this case, the margin is simply $\min_{x \in X, y \in Y} |\langle u_x, v_y \rangle|$. The *minimal dimension* of $M$, denoted as $d(M)$ in the sequel, is the smallest $k$ such that $M$ can be realized by a $k$-dimensional linear arrangement. The *maximal margin* of $M$, denoted as $\mu(M)$ in the sequel, is the largest $\mu > 0$ such that $M$ can be realized by a linear arrangement (of arbitrarily high dimension) with margin $\mu$.[2]

**Probabilistic Communication Complexity** A *two-way* probabilistic communication protocol is a probabilistic algorithm for two processors $\Pi_0$ and $\Pi_1$ that computes a distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Both processors have unbounded computational power. $\Pi_0$ sees only the first part, $x$, and $\Pi_1$ sees only the last part, $y$, of the input $(x, y) \in \{0,1\}^n \times \{0,1\}^n$. Obviously there has to be some communication between the two processors to calculate $f(x, y) \in \{0,1\}$. The processors can communicate by exchanging messages $b \in \{0,1\}^*$. The computation takes place in *rounds*. In each round one of the processors is active, in odd rounds it is $\Pi_0$ and in even rounds it is $\Pi_1$. The active processor probabilistically (depending on the part of the input it knows and on the past messages) chooses a message according to the communication protocol. In the final round the active processor probabilistically chooses the result of the computation. For a *one way* probabilistic communication protocol there are only two rounds: $\Pi_0$ randomly selects a message, sends it to $\Pi_1$, and $\Pi_1$ randomly performs the binary decision of either accepting the input or rejecting it.

We say that a protocol computes the distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ with unbounded error if for all inputs $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ the correct output is calculated with probability greater than $\frac{1}{2}$. (Since the slightest edge over random guessing is already sufficient, this model is sometimes called the *unbounded error* model.) The length of a communication protocol is $\lceil \log_2 N \rceil$, where $N$ is the number of distinct message sequences that can occur in computations that follow the protocol. The communication complexity $\mathrm{PComm}(f)$ of a distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ is the smallest length that any communication protocol that correctly computes $f$ can have.

We briefly note that Paturi and Simon [17] have shown that, in the unbounded error model, one-way and two-way protocols have almost the same power. In fact, any probabilistic two-way protocol of length $k$ can be converted into a one-way protocol with length at most $k + 1$.
Furthermore, the minimal dimension $d(f)$ of a distributed Boolean function $f(x, y)$ is closely related to its probabilistic communication complexity $\mathrm{PComm}(f)$. Paturi and Simon [17] have proven the following relation:

$$\lceil \log d(f) \rceil \le \mathrm{PComm}(f) \le \lceil \log d(f) \rceil + 1 \tag{3}$$

We will show in this paper that the following quantity is related to the margin: given a probabilistic protocol, let $\varepsilon(x, y)$ denote the difference between the probability that $(x, y)$ is accepted and the probability that $(x, y)$ is rejected. Thus $f(x, y) = \mathrm{sign}(\varepsilon(x, y))$ for any probabilistic protocol that

[2] A simple argument shows that the number of dimensions can always be reduced to $\min\{|X|, |Y|\}$ without decreasing the margin. A simple compactness argument shows that the maximal margin exists always.

correctly computes $f$. The *error bound* of the protocol is defined as $\min_{x\in X,y\in Y}|\varepsilon(x,y)|$. The class C-PP consists of all functions $f$ that can be correctly computed by a probabilistic communication protocol of length polylog($n$) with error bound $2^{-\mathrm{polylog}(n)}$. Halstenberg and Reischuk [13] have shown that the class C-PP does not change if we allow only one-way protocols. The class C-PP is one of the main issues in the bounded-error model for probabilistic communication complexity.

# 3 Bounds on the Minimal Dimension and on the Maximal Margin

In Subsection 3.1, we briefly review some of the main results in [9] and [10]. Subsection 3.2 presents a new lower bound on the minimal dimension that generalizes the old bound from [9]. The generalized bound does lead to exponential lower bounds in some cases where the old bound fails.

## 3.1 Known Bounds

Recently the following lower bound on the minimal dimension was proven:

**Theorem 1 (Forster [9]).** *For each matrix $M \in \{-1,+1\}^{X\times Y}$:*

$$d(M) \geq \frac{\sqrt{|X|\cdot|Y|}}{\|M\|}.$$

**Corollary 2 (Forster [9]).** Let $M \in \{-1,+1\}^{X\times Y}$ and $M' \in \mathbb{R}^{X\times Y}$ such that $\mathrm{sign}(M_{x,y}) = \mathrm{sign}(M'_{x,y})$ for all $x\in X, y\in Y$. Then the following holds:

$$\begin{aligned} \mathrm{PComm}(M) &\geq \log\left(\frac{\sqrt{|X|\cdot|Y|}}{\|M\|}\right), \\ \mathrm{rank}(M') &\geq \frac{\sqrt{|X|\cdot|Y|}}{\|M\|}. \end{aligned}$$

The lower bound on PComm($M$) follows directly from Theorem 1 and (3). The lower bound on rank($M'$) follows from the well known fact that each matrix of rank $k$ can be realized by a $k$-dimensional linear arrangement and vice versa.

As outlined in Forster [9], Theorem 1 allows to present concrete families of distributed Boolean functions whose minimal dimension grows exponentially in $n$. Thus, their probabilistic communication complexity grows linearly in $n$. Here is the example given in [9]:

**Example 3.** Let $\mathrm{ip}_n(x,y) := (-1)^{x^\top y}$, where $x,y\in\mathbb{Z}_2^n$, be the `inner product mod 2` function. It is well known that the corresponding matrix $H$ such that $H_{x,y} = \mathrm{ip}_n(x,y)$ is a Hadamard matrix. $H$ has operator norm $2^{n/2}$ because it has orthogonal columns and each row vector has Euclidean length $2^{n/2}$. According to Theorem 1 and Corollary 2:

$$d(\mathrm{ip}_n) \geq 2^{n/2} \tag{4}$$
$$\mathrm{PComm}(\mathrm{ip}_n) \geq n/2 \tag{5}$$

It was shown by Alon, Frankl and Rödl [2] that the communication complexity of almost all Boolean functions $f:\{0,1\}^n\times\{0,1\}^n\to\{-1,1\}$ is at least $n-5$. It is even the case that the lower bound of Corollary 2 on the communication complexity is linear in $n$ for the vast majority of all Boolean functions. This means that the lower bound of Corollary 2 is reasonably good for almost all Boolean functions. To show this, fix $n \in \mathbb{N}$, let $N := 2^n$ and let the entries of the matrix $M \in \{-1,+1\}^{N\times N}$ be i.i.d. with $\Pr(M_{ij} = 1) = \frac{1}{2}$. With high probability the operator norm $\|M\| = \sqrt{\|MM^\top\|}$ is small: For fixed $i \neq k$, the entry $(MM^\top)_{ik} = \sum_{j=1}^N M_{ij}M_{kj}$ is the sum of the $N$ i.i.d. $\pm 1$-valued random variables $X_j = M_{ij}M_{kj}$. From Hoeffding's Inequality (see [18], [11]) it follows that $\Pr\left(|(MM^\top)_{ik}| < \gamma\right) \geq 1 - \exp(-\frac{\gamma^2}{N})$ for all $\gamma > 0$. Thus with probability at least $1 - N^2\exp(-\frac{\gamma^2}{N})$ all entries $(MM^\top)_{ik}$, $i \neq k$, have absolute value at most $\gamma$. Because the entries on the diagonal of $MM^\top$ are all equal to $N$, it follows from Gerschgorin's Circle Theorem (see, e.g., [8]) that in this case the operator norm of $MM^\top$ is at most $N + N\gamma$. For a parameter $0 < c < \frac{1}{4}$ we choose $\gamma = N^{1-2c}$. Then $\|M\| \leq \sqrt{N+N\gamma} = \sqrt{N + N^{2(1-c)}} \leq \sqrt{2N^{2(1-c)}}$. From Corollary 2 it follows that that the probabilistic communication complexity is at least

$$n - \log\|M\| \geq n - \frac{1}{2}(\log(2) + 2n(1-c)) = cn - \frac{1}{2}$$

with probability at least $1 - N^2\exp(-N^{1-4c})$.

Furthermore, Forster, Schmitt, and Simon [10] proved the following upper bound on the maximal margin:

**Theorem 4 (Forster, Schmitt, Simon).** *For each matrix $M \in \mathbb{R}^{X\times Y}$ with no zero entries:*

$$\mu(M) \leq \frac{\sqrt{|X|}\cdot\|M\|}{\sqrt{\sum_{y\in Y}\left(\sum_{x\in X}|M_{x,y}|\right)^2}}.$$

Note that Theorem 4 implies the following upper bound on the maximal margin of the inner product mod 2 function:[3]

$$\mu(\mathrm{ip}_n) \leq 2^{-n/2}. \tag{6}$$

### 3.2 A Generalized Lower Bound on the Minimal Dimension

Theorem 1 allowed us to derive an exponential lower bound on $d(\mathrm{ip}_n)$ because the matrix induced by $\mathrm{ip}_n$, the Hadamard matrix, has $2^n$ orthogonal columns. However, Theorem 1 may also fail to certify valid exponential lower bounds. For this reason, we present a generalized lower bound which brings more functions in the reach of our machinery:

**Theorem 5.** *Let $M \in \mathbb{R}^{X\times Y}$ be a matrix with no zero entries. Then the following holds:*

$$d(M) \geq \frac{\sqrt{|X|\,|Y|}}{\|M\|}\min_{x\in X, y\in Y}|M_{x,y}|.$$

*Proof.* Assume that there are vectors $u_x, v_y \in \mathbb{R}^k$ such that $\mathrm{sign}(M_{x,y}) = \mathrm{sign}(\langle u_x, v_y\rangle)$ for all $x \in X$, $y \in Y$. Obviously we can assume that $\min_{x\in X,y\in Y}|M_{x,y}| = 1$. It has been shown by Forster [9] that a given linear arrangement $u_x, v_y$ can be normalized such that $u_x, v_y \in S^{k-1}$ and

$$\sum_{x\in X} u_x u_x^\top = \frac{|X|}{k} I_k \ . \tag{7}$$

[3]This upper bound had been shown already by Forster [9]. He derived it from a theorem that is slightly weaker than Theorem 4.

Now we have for all $y \in Y$ that

$$\sum_{x\in X} \underbrace{M_{x,y}\langle u_x, v_y\rangle}_{\geq 0} \overset{|M_{x,y}|\geq 1 \geq |\langle u_x,v_y\rangle|}{\geq} \sum_{x\in X} \langle u_x, v_y\rangle^2 = v_y^\top \left(\sum_{x\in X} u_x u_x^\top\right) v_y = \frac{|X|}{k} \ . \tag{8}$$

Lemma 17 (which is given in Appendix A) gives a corresponding upper bound in terms of the operator norm $\|M\|$. It follows that

$$|Y| \left(\frac{|X|}{k}\right)^2 \overset{(8)}{\leq} \sum_{y\in Y} \left(\sum_{x\in X} M_{x,y}\langle u_x, v_y\rangle\right)^2 \overset{\text{Lemma 17}}{\leq} |X| \, \|M\|^2 \ .$$

□

## 4 An Application of the Generalized Lower Bound

We would like to demonstrate (by means of an example) the greater flexibility provided by Theorem 5. Let $\mathbb{F}_p$ be the prime field of characteristic $p$, let $\mathbb{F}_p^n$ be the $n$-dimensional vector space over $\mathbb{F}_p$, and let $\mathbb{P}_{n-1}(\mathbb{F}_p)$ be the $(n-1)$-dimensional projective space. Remember that the elements (projective points) in $\mathbb{P}_{n-1}(\mathbb{F}_p)$ are the 1-dimensional linear subspaces of $\mathbb{F}_p^n$. We assume that the projective points are given by their homogeneous coordinates. We say that a projective point $Q = (q_1, \ldots, q_n)$ is *orthogonal* to a projective point $Q' = (q_1', \ldots, q_n')$, denoted as $Q \perp Q'$, if $\sum_{i=1}^n q_i q_i' = 0$. The Boolean function $\mathrm{ORT}_{p,n} : \mathbb{P}_{n-1}(\mathbb{F}_p) \times \mathbb{P}_{n-1}(\mathbb{F}_p) \to \{-1, 1\}$

$$\mathrm{ORT}_{p,n}(Q, Q') := \begin{cases} 1, & \text{if } Q \perp Q' \\ -1, & \text{otherwise} \end{cases} \tag{9}$$

is the indicator function of the orthogonality relation. In the full paper, we show that the operator norm of $\mathrm{ORT}_{p,n}$ is too small for certifying an exponential lower bound on the minimal dimension with Theorem 1. In contrast to this, the following can be shown by means of Theorem 5 and Theorem 4 (and by means of a nice trick from [16]):

**Theorem 6.** *Asymptotically, the minimal dimension and the maximal margin of* $\mathrm{ORT}_{p,n}$ *are bounded as follows:*

$$d(\mathrm{ORT}_{p,n}) \geq p^{n/2-1}(1 - o(1))$$

$$\frac{1}{p^{(n-1)/2}}(1 - o(1)) \leq \mu(\mathrm{ORT}_{p,n}) \leq \frac{\sqrt{p}}{2p^{(n-1)/2}}(1 + o(1))$$

*Proof.* Our proof strategy (borrowed from [16]) is to define a function (matrix) $\mathrm{ORT}_{p,n}'$ such that $\mathrm{ORT}_{p,n}$ and $\mathrm{ORT}_{p,n}'$ satisfy the requirements of Theorem 5, and such that the columns of $\mathrm{ORT}_{p,n}'$ are orthogonal (which makes $\|\mathrm{ORT}_{p,n}'\|$ small and, therefore, the lower bound on $d(\mathrm{ORT}_{p,n})$ large). Details follow.

For

$$w_+ := p - 1 \text{ and } w_- := \frac{1 + p^{(n-2)/2}}{p^{(n-2)/2}},$$

$\mathrm{ORT}'_{p,n}$ is given by

$$\mathrm{ORT}'_{p,n}(Q,Q') := \begin{cases} w_+, & \text{if } Q \perp Q' \\ -w_-, & \text{otherwise} \end{cases} . \tag{10}$$

Obviously, $|\mathrm{ORT}'_{p,n}(Q,Q')| \geq 1$ and $\mathrm{sign}(\mathrm{ORT}'_{p,n}(Q,Q')) = \mathrm{sign}(\mathrm{ORT}_{p,n}(Q,Q'))$ for all $Q, Q'$. We may therefore apply Theorem 5 and Theorem 4 to $\mathrm{ORT}'_{p,n}$ (instead of $\mathrm{ORT}_{p,n}$). It has been shown by Krause and Waack [16] that $\mathrm{ORT}'_{p,n}$ (viewed as matrix) has orthogonal columns. Its operator norm coincides therefore with the Euclidean length of one of its row vectors (all of which have the same length). The following facts are well known. $N := |\mathbb{P}_{n-1}(\mathbb{F}_p)| = (p^n-1)/(p-1)$. Furthermore, each projective point is orthogonal to $N_+ = (p^{n-1}-1)/(p-1)$ projective points and non-orthogonal to the remaining $N_- = N - N_+ = p^{n-1}$ projective points. We conclude that

$$\|\mathrm{ORT}'_{p,n}\| = \sqrt{N_+ w_+^2 + N_- w_-^2} = p^{n/2} + 1.$$

According to Theorem 5:

$$d(\mathrm{ORT}_{p,n}) \geq \frac{\sqrt{N \cdot N}}{\|\mathrm{ORT}'_{p,n}\|} = \frac{p^{n/2}-1}{p-1}.$$

According to Theorem 4 and by straightforward computation,

$$\mu(\mathrm{ORT}_{p,n}) \leq \frac{\|\mathrm{ORT}'_{p,n}\| \cdot \sqrt{N}}{\sqrt{\sum_Q \left(\sum_{Q'} |\mathrm{ORT}'_{p,n}(Q,Q')|\right)^2}} = \frac{p^{n/2}+1}{2p^{n-1}+p^{n/2}-1}.$$

The trivial margin is $N^{-1/2} = ((p-1)/(p^n-1))^{1/2}$. □

# 5 Large Margins and Bounded Error

Recall that a family $(f_n)$ of Boolean functions $f_n : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ belongs to C-PP if there exists a probabilistic one-way protocol that transmits at most polylog$(n)$ bits (uses at most $2^{\mathrm{polylog}(n)}$ messages) and achieves error bound $2^{-\mathrm{polylog}(n)}$.

The goal of this section is to express membership in C-PP in terms of only one parameter: the maximal margin. Here is the main result:

**Theorem 7.** $(f_n) \in$ C-PP $\Leftrightarrow \mu(f_n) \geq 2^{-\mathrm{polylog}(n)}$.

Theorem 7 follows directly from Lemmas 8, 9, and 10 below. Lemma 9 makes use of the random projection technique that was introduced by Arriaga and Vempala [3]. Lemmas 8 and 10 are implicitly proven in the paper of Paturi and Simon [17] on probabilistic communication complexity with unbounded error. Since they do not explicitly keep track of error bounds (not to speak of margins), we sketch the proofs for these lemmas in Appendix B.

**Lemma 8.** *Each probabilistic one-way protocol for $f$ that uses at most $N$ messages and achieves error-bound $\varepsilon$ can be converted into an $N$-dimensional linear arrangement that realizes $f$ with margin $\mu \geq \varepsilon/\sqrt{N}$.*

**Lemma 9.** *Each linear arrangement (of arbitrarily high dimension) that realizes $f$ with margin $\mu$ can be converted into an $\lceil (12(n+1)/\mu)^2 \rceil$-dimensional linear arrangement that realizes $f$ with margin $\mu/2$.*

*Proof.* The following result (whose proof can be found in [4]), is based on the technique of random projections (from Arriaga and Vempala [3]):
*Let $w, x \in \mathbb{R}^r$ be arbitrary but fixed. Let $R = (R_{i,j})$ be a random $(k \times r)$-matrix such that the entries $R_{i,j}$ are i.i.d. according to the normal distribution $N(0,1)$. Consider the random projection $u_R := \frac{1}{\sqrt{k}}(Ru) \in \mathbb{R}^k$ for all $u \in \mathbb{R}^k$. Then the following holds for every constant $\mu > 0$:*

$$\Pr_R \left[ |\langle w_R, x_R \rangle - \langle w, x \rangle| \geq \frac{\mu}{2} \left( L_2(w)^2 + L_2(x)^2 \right) \right] \leq 4e^{-\mu^2 k/8}.$$

This result can be used in the obvious way to guarantee the existence of a random projection that maps an $r$-dimensional linear arrangement that realizes $f$ with margin $\mu$ to a $k$-dimensional linear arrangement that realizes $f$ with margin $\mu/2$. The dimension $k$ must exceed a critical threshold that depends on $\mu$ and $n$ (but does not depend on $r$). If the computations are carried out carefully, it turns out that $k := \lceil (12(n+1)/\mu)^2 \rceil$ is sufficiently large. We omit the (somewhat tedious) computations in this abstract. □

**Lemma 10.** *Each $N$-dimensional linear arrangement that realizes $f$ with margin $\mu$ can be converted into a probabilistic one-way protocol that uses at most $2N$ messages and achieves error bound $\varepsilon \geq \mu/\sqrt{N}$.*

# 6 A Lower Bound on the Size of Depth-2 Threshold Circuits

The following theorem, which is the main result of this section, states that (loosely speaking) Boolean functions with "high" minimal dimension cannot be computed by "small" (somewhat technically restricted) depth-2 threshold circuits. Part a) of this theorem strengthens the lower bound of Hajnal et al. [12] and Part b) generalizes and strengthens the results of Bruck and Smolensky [6], Krause [14], and Krause and Pudlák [15]. Note that for technical reasons we assume that the top threshold gate is $\pm 1$-valued. The threshold gates on the bottom level are $\{0,1\}$-valued.

**Theorem 11.** *Let $(f_n)$ be a family of distributed Boolean functions $f_n : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$. Suppose $(f_n)$ is computed by depth-2 threshold circuits in which the top gate is a linear threshold gate (with* unrestricted *weights). Then the following holds:*

a) *If the bottom level has $s$ linear threshold gates using integer weights of absolute value at most $W$, then $s = \Omega\left(\frac{d(f_n)}{nW}\right)$. In particular, $s = \Omega\left(\frac{2^{n/2}}{nW}\right)$ for $f_n = \mathrm{ip}_n$, and $s = \Omega\left(\frac{p^{n/2-1}}{nW}\right)$ for $f_{n\lceil \log p \rceil} = \mathrm{ORT}_{p,n}$.*

b) *If the bottom level has $s$ gates computing symmetric functions, then $s = \Omega\left(\frac{d(f_n)}{n}\right)$. In particular, $s = \Omega\left(\frac{2^{n/2}}{n}\right)$ for $f_n = \mathrm{ip}_n$, and $s = \Omega\left(\frac{p^{n/2-1}}{n}\right)$ for $f_{n\lceil \log p \rceil} = \mathrm{ORT}_{p,n}$.*

Note that the specific bounds for $\mathrm{ip}_n$ and $\mathrm{ORT}_{p,n}$ follow immediately from the general bound for $f_n$ by applying inequality (4) and Theorem 6.

The proof of this theorem is based on Theorem 1 and on ideas from the results cited above. We start with two lemmas.

**Lemma 12.** *Let $G : \{0,1\}^n \times \{0,1\}^n \longrightarrow \{0,1\}$ be a threshold function where for $x, y \in \{0,1\}^n$, $G(x,y) = 1$ iff $\sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \beta_i y_i \geq \mu$ for weights $\alpha_i, \beta_i, \mu \in \mathbb{Z}$. Then, $G$ (viewed as a matrix) has rank at most* $\min\{\sum_{i=1}^n |\alpha_i|, \sum_{i=1}^n |\beta_i|\} + 1$.

*Proof.* W.l.o.g. let $\sum_{i=1}^n |\alpha_i| \leq \sum_{i=1}^n |\beta_i|$. Let $\alpha_{\min}$ ($\alpha_{\max}$) be the minimal (maximal) value taken by $\sum_{i=1}^n \alpha_i x_i$ as $x$ ranges over all possible inputs in $\{0,1\}^n$. As the weights are integers, this sum takes at most $\alpha_{\max} - \alpha_{\min} + 1$ distinct values. We partition the rows of $G$ according to the weight contributed by $x$ and, within each block of these rows, we partition the columns into two groups depending on whether or not the weight contributed by $y$ together with that of any row of that block exceeds the threshold $\mu$ or not. Specifically, define the following sets of entries of $G$ for all $\alpha$ such that $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$:

$$S_{\alpha,0} := \{(x,y) : \sum_{i=1}^n \alpha_i x_i = \alpha \text{ and } \sum_{i=1}^n \beta_i y_i < \mu - \alpha\},$$
$$S_{\alpha,1} := \{(x,y) : \sum_{i=1}^n \alpha_i x_i = \alpha \text{ and } \sum_{i=1}^n \beta_i y_i \geq \mu - \alpha\}.$$

Let $G_{\alpha,0}$ and $G_{\alpha,1}$ be (disjoint) submatrices of $G$ defined by the entries $S_{\alpha,0}$ and $S_{\alpha,1}$ respectively. It is clear that $G_{\alpha,0}$ is an all-0 matrix and $G_{\alpha,1}$ is an all-1 matrix for any $\alpha$. Furthermore $G = \sum_\alpha (G_{\alpha,0} + G_{\alpha,1})$. Hence, by subadditivity of rank we see that the rank of $G$ is at most the number of distinct values taken by $\alpha$. The latter is bounded above by $\alpha_{\max} - \alpha_{\min} + 1 \leq \sum_{i=1}^n |\alpha_i| + 1$. $\square$

Note that the same proof goes through even if we generalize the definition of $G$ by setting $G(x,y) = 1$ iff $\sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \beta_i y_i \in T$ for an *arbitrary* subset $T$ of $\mathbb{Z}$. Specifically, we have the following corollary of the proof:

**Corollary 13.** Let $G : \{0,1\}^n \times \{0,1\}^n \longrightarrow \{0,1\}$ be a symmetric function in the sense that its value depends only on $\sum_{i=1}^n (x_i + y_i)$. Then its matrix $G$ as defined above has rank at most $n+1$.

**Lemma 14.** *Let $f : \{0,1\}^n \times \{0,1\}^n \longrightarrow \{-1,+1\}$ be a Boolean function computed by a depth-2 threshold circuit $C$ with the top gate using* unrestricted weights *and each of the bottom gates using weights of absolute value at most $W$. Then there is a real matrix $F$ such that* $\mathrm{sign}(F(x,y)) = f(x,y)$ *for all $x, y \in \{0,1\}^n$ and* $\mathrm{rank}(F) = O(snW)$*, where $s$ is the number of bottom gates.*

*Proof.* Let the top gate of $C$ have weights $\phi_1, \ldots \phi_s$ and threshold $\phi_0$. Hence we can write $f(x,y) = \mathrm{sign}(\sum_{i=1}^s \phi_i G_i(x,y) - \phi_0)$, where $G_i$ are the functions computed by the bottom gates. Define the matrix

$$F := \phi_1 G_1 + \cdots + \phi_s G_s - \phi_0 J,$$

where $J$ is the all 1's $2^n \times 2^n$ matrix. It is clear that $f(x,y) = \mathrm{sign}(F(x,y))$ for all $x, y \in \{0,1\}^n$. Moreover, $\mathrm{rank}(F) \leq 1 + \sum_{i=1}^s \mathrm{rank}(G_i) \leq 1 + s(1 + nW)$ using Lemma 12. $\square$

Using Corollary 13, one can similarly prove that if $f$ is computed by a depth-2 circuit with a threshold gate at the top and $s$ symmetric gates at the bottom, then there is a matrix $F$ of rank $O(sn)$ that sign-represents $f$.

*Proof of Theorem 11.* By Lemma 14, if a depth-2 threshold circuit computes $f_n(x,y)$, then there is a matrix $F_n$ such that $\mathrm{sign}(F_n(x,y)) = \mathrm{sign}(f_n(x,y))$ and $\mathrm{rank}(F_n) = O(snW)$. On the other hand, $\mathrm{rank}(F_n) \geq d(f_n)$ (because there is always a $\mathrm{rank}(F_n)$-dimensional linear arrangement for $F_n$, which is then also a linear arrangement for $f_n$). Comparing the upper and lower bounds on $\mathrm{rank}(F)$, we get $snW = \Omega(d(f_n))$. This proves part a) of the theorem. Part b) is proved similarly by means of Corollary 13. $\square$

# 7 A Lower Bound on the Size of Probabilistic OBDD's

We briefly recall the technical definitions concerning OBDD's. An *ordered binary decision diagram (OBDD)* over Boolean variables $x_1, \ldots, x_n$ is given by a directed acyclic graph $G = (V, E)$. $G$ contains exactly one source (node without proper predecessor) and two sinks (nodes without proper successor). One of the sinks (the $(-1)$-sink) is labeled $-1$, the other-one (the $(+1)$-sink) is labeled 1. Each non-sink is labeled by a Boolean variable such that, on each path from the source to a sink, each variable occurs exactly once and they occur always in the same order.[4] Each non-sink has two outgoing edges. One of them is labeled 0, the other-one is labeled 1. Each OBDD $G$ represents a Boolean function $f_G$ which is evaluated on $a \in \{0,1\}^n$ as follows. Vector $a$ enters $G$ at the source and is then routed through $G$ until it reaches a sink. More specifically, at a node labeled $x_i$, vector $a$ is routed through the outgoing edge labeled $a_i$. According to this rule, $a$ will be routed either to the $(-1)$-sink or to the $(+1)$-sink. In the former case, we have $f_G(a) = -1$; in the latter case $f_G(a) = 1$. The *size* of an OBDD is the number of its nodes.

A *probabilistic ordered binary decision diagram (POBDD) G* is defined analogously, except that the routing procedure becomes probabilistic. Loosely speaking the deterministic transitions along edges labeled either 0 or 1 are replaced by random transitions. Formally, this can be achieved as follows. With each node $v$ on layer $l$ of $G$, we associate two probability distributions $D_0$ and $D_1$ that assign probabilities to all nodes on layer $l+1$. If $v$ is labeled $x_i$ and vector $a$ is currently located at $v$, then $a$ is routed to node $w$ on layer $l+1$ with probability $D_{a_i}(w)$. Each POBDD $G$ represents a collection $(B_G(x))$ of Bernoulli random variables in the obvious sense. We say that $G$ *realizes* Boolean function $f : \{0,1\}^n \rightarrow \{-1,+1\}$ if $f(x) = 1$ implies that $\Pr(B_G(x) = 1) > 1/2$, whereas $f(x) = -1$ implies that $\Pr(B_G(x) = 1) < 1/2$. The *size* of a POBDD is the number of its nodes.[5]

POBDD's are quite powerful because they need only a slight edge over random guessing. To the best of our knowledge, exponential lower bounds on the size of POBDD's are not yet known for concrete functions. In the sequel, we present a function $\mathrm{IP}_n$ that is provably hard to realize with POBDD's. Our function $\mathrm{IP}_n$ embodies $\mathrm{ip}_n$ (the inner product modulo 2) as a subfunction. Note however that $\mathrm{ip}_n$ itself is easy to compute for OBDD's (even for deterministic ones) given that they do not accidentally use a "weird" variable order. We will define $\mathrm{IP}_n$ in such a way, that, no matter how the variable order in the OBDD is chosen, there will be a subfunction of the form $\mathrm{ip}_n$ such that the chosen variable order is actually "weird". A similar technique has been used in [1].

We move on to the definition of $\mathrm{IP}_n$. We use $2n$ Boolean variables $z_1, \ldots, z_{2n}$ and, in addition, $2n$ auxiliary Boolean variables $h_1, \ldots, h_{2n}$. Assume that the Hamming weight of $h$ is $n$ and that the 0-entries occur in positions $i(1) < \cdots < i(n)$, whereas the 1-entries occur in positions $j(1) <$

[4]Since we say that each variable occurs exactly once (as opposed to at most once) on each path, we implicitly make the assumption that $G$ is *layered*, where all nodes in the same layer are labeled by the same variable and edges go from one layer to the next-one according to the fixed order of the variables. This assumption is made only for the sake of simple technical exposition.

[5]Our definition of probabilistic OBDD's makes them computationally more powerful than the traditional definition. Note, however, that lower bounds on the size of POBDD's according to our definition are (à-fortiori) valid for POBDD's according to the traditional definition.

$\dots < j(n)$. Then, we define

$$\mathrm{IP}_n(z,h) := (-1)^{\sum_{k=1}^n z_{i(k)} z_{j(k)}} \ . \tag{11}$$

If the Hamming weight of $h$ differs from $n$ we define $\mathrm{IP}_n(z,h) := 1$ by default.

**Theorem 15.** *Each POBDD $G$ that computes $\mathrm{IP}_n$ has size at least $2^{n/2}$.*

*Proof.* Assume that the variables $z_1, \dots, z_{2n}$ appear in $G$ in the order $z_{\sigma(1)}, \dots, z_{\sigma(2n)}$. Let $i(1) < \dots < i(n)$ be the sorted sequence of the elements of $\{\sigma(1), \dots, \sigma(n)\}$. Likewise $j(1) < \dots < j(n)$ denotes the sorted sequence of all the elements of $\{\sigma(n+1), \dots, \sigma(2n)\}$. Note that in $G$ all $z_{i(k)}$-variables occur before all $z_{j(k)}$-variables. Fix a vector $h$ of Hamming weight $n$ such that $h_{i(k)} = 0$ and $h_{j(k)} = 1$ for $k = 1, \dots, n$. This leads to a subfunction of the form (11). Let $V'$ be the set of nodes labeled $z_{j(1)}$. In what follows, we write input $z$ in the form $(x,y)$, where $x = (z_{i(1)}, \dots, z_{i(n)})$ and $y = (z_{j(1)}, \dots, z_{j(n)})$. For each $v \in V'$, denote by $p_v(x)$ the probability that $z = (x,y)$ is routed from the source to $v$. Denote by $q_v(y)$ the probability that $z$ is routed from $v$ to the $(+1)$-sink. We conclude that

$$\mathrm{IP}_n(x,y,h) = 1 \Leftrightarrow \sum_{v \in V'} p_v(x) q_v(y) > 1/2 \ . \tag{12}$$

Note that, for our fixed choice of $h$, the subfunction $\mathrm{IP}_n(x,y,h)$ is the inner product modulo 2 of $x$ and $y$. Equation (12) realizes this subfunction by a $|V'|$-dimensional arrangement that uses half spaces induced by inhomogeneous hyper-planes. Adding one dimension, the hyper-planes can be made homogeneous (and we arrive at a linear arrangement). According to (4), $|V| \geq |V'| + 1 \geq 2^{n/2}$. $\square$

It follows from Example 3 that every probabilistic two-way protocol for $\mathrm{IP}_n$ has length at least $n/2 - 1$. Our lower bound can be extended to $k$-POBDD's: $k$-POBDD's generalize POBDD's in the following way. On each path from the source to a sink the variables appear at most $k$ times in an order that is the same permutation of variables repeated $k$ times, i.e. a $k$-POBDD can be partitioned into $k$ layers such that each layer is read-once ordered according to the same variable ordering. It holds

**Corollary 16.** Each $k$-POBDD, $k \geq 2$, computing $\mathrm{IP}_n$ has size at least $2^{\frac{n-2}{4k-2} - 1}$.

Indeed, if there exists a $k$-POBDD $P$ of size $S$ then one can find a probabilistic communication protocol of $2k$ rounds that computes the same function as $P$ and has length $(2k-1)\lceil \log S \rceil$. In this communication game the input variables of processor $\Pi_0$ are the variables that are read in the first half of each layer of $P$ and the input variables of processor $\Pi_1$ are the variables read in the second half of each layer, respectively. The computation can be performed in the way that the processors follow a path in $P$ in turn such that each processor sends to the other one the number of the first node on the path not belonging to its sublayer. This well known technique (e.g. [5]) yields the desired lower bound.

In the full paper we generalize Theorem 15 and Corollary 16 to the case of weakly ordered probabilistic branching programs.

## References


[1] F. Ablayev and M. Karpinski. On the power of randomized ordered branching programs, ECCC 98-004.

[2] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*. New York: IEEE Computer Society, 1985

[3] Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. In *Proceedings of the 40'th Annual Symposium on the Foundations of Computer Science*, pages 616–623, 1999.

[4] Shai Ben-David, Nadav Eiron, and Hans Ulrich Simon. Limitations of learning via embeddings in Euclidean half-spaces. In *Proceedings of the 14th Annual Workshop on Computational Learning Theory*. Springer Verlag, 2001. To appear.

[5] B. Bollig, M. Sauerhoff, D. Sieling, and I. Wegener. On the Power of Different Types of Branching Programs, ECCC TR94-026.

[6] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, $AC^0$ functions and spectral norms. In *Proceedings of the 31st Symposium on Foundations of Computer Science*, pages 632–641, 1990.

[7] N. Christianini, J. Shawe-Taylor. *An Introduction to Support Vector Machines,* Cambridge University Press, 2000.

[8] D. K. Faddejew, and W. N. Faddejewa. *Numerische Methoden der linearen Algebra.* München: R. Oldenbourg Verlag, 1973.

[9] Jürgen Forster. A linear bound on the unbounded error probabilistic communication complexity. In *Proceedings of the 16th Annual Conference on Computational Complexity.* IEEE Computer Society Press, 2001. To appear.

[10] Jürgen Forster, Niels Schmitt, and Hans Ulrich Simon. Estimating the optimal margins of embeddings in Euclidean half spaces. In *Proceedings of the 14th Annual Workshop on Computational Learning Theory.* Springer Verlag, 2001. To appear.

[11] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30.

[12] A. Hajnal, W. Maass, P. Pudlàk, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46:129–154, 1993.

[13] B. Halstenberg, R. Reischuk: Different Modes of Communication. *SIAM Journal on Computing,* Vol. 22, No. 5, pp. 913-934, 1993.

[14] Matthias Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theoretical Computer Science*, 156:99–117, 1996.

[15] Matthias Krause and Pavel Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. In *Proceedings of the 25th Annual Symposium on Theory of Computing*, pages 48–57, 1993.

[16] Matthias Krause and Stephan Waack. Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in. In *Proceedings of the 32nd Symposium on Foundations of Computer Science*, pages 777–782, 1991.

[17] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986.

[18] David Pollard. *Convergence of Stochastic Processes.* New York: Springer Verlag, 1984.

[19] Ingo Wegener. *Branching Programs and Binary Decision Diagrams - Theory and Applications.* SIAM Monographs on Discrete Mathematics and Applications, 2000.

## A Proof of Lemma 17

In the proof of Lemma 17 we will use the following simple fact: For any two symmetric positive semi-definite matrices $A, B \in \mathbb{R}^{X\times X}$ the sum of the products of the entries is nonnegative. This can be seen as follows. Because of the Spectral Theorem there are eigenvectors $d_1, \ldots, d_{|X|} \in \mathbb{R}^X$ and eigenvalues $\lambda_1, \ldots, \lambda_{|X|} \geq 0$ of $A$ such that $A = \sum_{i=1}^{|X|} \lambda_i d_i d_i^\top$. It follows that

$$\sum_{x,\tilde{x}\in X} A_{x,\tilde{x}} B_{x,\tilde{x}} = \sum_{x,\tilde{x}\in X} \sum_{i=1}^{|X|} \lambda_i (d_i)_x (d_i)_{\tilde{x}} B_{x,\tilde{x}} = \sum_{i=1}^{|X|} \underbrace{\lambda_i}_{\geq 0} \underbrace{d_i^\top B d_i}_{\geq 0} \geq 0 \ .$$

**Lemma 17.** *Let $M \in \mathbb{R}^{X\times Y}$ be an arbitrary matrix. Then*

$$\sum_{y\in Y} \left( \sum_{x\in X} M_{x,y} \langle u_x, v_y\rangle \right)^2 \leq |X| \, \|M\|^2$$

*for any vectors $u_x, v_y \in S^{k-1}$ with $\operatorname{sign}(\langle u_x, v_y\rangle) = \operatorname{sign}(M_{x,y})$ for all $x \in X$, $y \in Y$.*

*Proof.* For every $y \in Y$ we have that

$$\sum_{x\in X} M_{x,y} \langle u_x, v_y\rangle = \left\langle \sum_{x\in X} M_{x,y} u_x, v_y \right\rangle \overset{L_2(v_y)=1}{\leq} L_2\left( \sum_{x\in X} M_{x,y} u_x \right) \ . \tag{13}$$

We square this inequality and sum over $y \in Y$:

$$\begin{aligned} \sum_{y\in Y} \left( \sum_{x\in X} M_{x,y} \langle u_x, v_y\rangle \right)^2 &\overset{(13)}{\leq} \sum_{y\in Y} L_2 \left( \sum_{x\in X} M_{x,y} u_x \right)^2 \\ &= \sum_{y\in Y} \sum_{x,\tilde{x}\in X} M_{x,y} \cdot M_{\tilde{x},y} \cdot \langle u_x, u_{\tilde{x}}\rangle = \sum_{x,\tilde{x}\in X} \left( MM^\top \right)_{x,\tilde{x}} \langle u_x, u_{\tilde{x}}\rangle \\ &\overset{(*)}{\leq} \sum_{x,\tilde{x}\in X} \left( \|M\|^2 I_X \right)_{x,\tilde{x}} \langle u_x, u_{\tilde{x}}\rangle = \|M\|^2 \sum_{x\in X} L_2(u_x)^2 = |X| \, \|M\|^2 \ . \end{aligned}$$

Here inequality $(*)$ holds because $A := \|M\|^2 I_X - MM^\top$ and $B := (\langle u_x, u_{\tilde{x}}\rangle)_{x,\tilde{x}\in X}$ are symmetric positive semi-definite matrices, and therefore $\sum_{x,\tilde{x}\in X} A_{x,\tilde{x}} B_{x,\tilde{x}} \geq 0$. $\square$

## B Proofs of Lemmas 8 and 10

**Lemma 8.** *Each probabilistic one-way protocol for $f$ that uses at most $N$ messages and achieves error-bound $\varepsilon$ can be converted into an $N$-dimensional linear arrangement that realizes $f$ with margin $\mu \geq \varepsilon/\sqrt{N}$.*

*Proof.* Let $p_i(x)$ denote the probability that processor $\Pi_0$ with local input $x$ sends message $M_i$ to processor $\Pi_1$. Let $q_i(y)$ denote the probability that processor $\Pi_1$ accepts after it has received message $M_i$. Recall that $\varepsilon(x,y)$ is the difference between the probability of acceptance and the probability of rejection of input $(x,y)$ and the error bound achieved by the protocol is $\min_{x,y}|\varepsilon(x,y)|$. Thus:

$$\varepsilon(x,y) = \sum_{i=1}^{N} p_i(x)q_i(y) - \sum_{i=1}^{N} p_i(x)(1-q_i(y)) = \sum_{i=1}^{N} p_i(x)(2q_i(y)-1) \tag{14}$$

Since the protocol is correct and achieves error bound $\varepsilon$, $f(x,y)=1$ implies that $\varepsilon(x,y) \geq \varepsilon$ and $f(x,y)=-1$ implies that $\varepsilon(x,y) \leq -\varepsilon$. Mapping $x$ to vector $u_x := (p_i(x))_{i=1,\ldots,N}$ and $y$ to vector $v_y := (2q_i(y)-1)_{i=1,\ldots,N}$, we arrive at an $N$-dimensional linear arrangement that realizes $f$. Clearly, the margin $\mu$ achieved by this arrangement satisfies

$$\mu \geq \frac{\varepsilon}{\max_{x\in X} L_2(u_x)\cdot \max_{y\in Y} L_2(v_y)} \ .$$

Recall that $L_2$ is upper-bounded by $L_1$. Since $0 \leq p_i(x), q_i(y) \leq 1$ and $\sum_{i=1}^{N} p_i(x) = 1$, we get $L_2(u_x) \leq L_1(u_x) = 1$ and $L_2(v_y) \leq L_2(1,\ldots,1) = \sqrt{N}$ for all $x,y$. Thus, $\mu \geq \varepsilon/\sqrt{N}$. □

**Lemma 10.** *Each $N$-dimensional linear arrangement that realizes $f$ with margin $\mu$ can be converted into a probabilistic one-way protocol that uses at most $2N$ messages and achieves error bound $\varepsilon \geq \mu/\sqrt{N}$.*

*Proof.* Let $u(x), v(y) \in S^{N-1}$ be the vectors of the $N$-dimensional linear arrangement that realizes $f$ with margin $\mu$. Consider the following probabilistic one-way protocol. Processor $\Pi_0$ with local input $x$ sends the message $(i, \mathrm{sign}(u_i(x)))$ with probability

$$p_i(x) := \frac{|u_i(x)|}{L_1(u(x))} \ .$$

If processor $\Pi_1$ receives message $M_i$, it accepts with probability

$$q_i(y) := \frac{\mathrm{sign}(u_i(x))\cdot v_i(y)+1}{2} \ .$$

For any $x \in X$ and any $y \in Y$:

$$\varepsilon(x,y) = \sum_{i=1}^{N} p_i(x)\cdot(2q_i(y)-1) = \frac{1}{L_1(u(x))}\cdot\sum_{i=1}^{N}|u_i(x)|\cdot \mathrm{sign}(u_i(x))\cdot v_i(y) = \frac{\langle u(x), v(y)\rangle}{L_1(u(x))} \ .$$

Thus, the protocol correctly computes $f$. Since $L_1(u(x)) \leq \sqrt{N}\cdot L_2(u(x)) = \sqrt{N}$ and since the arrangement realizes $f$ with margin $\mu$, the error bound $\varepsilon$ achieved by the protocol satisfies $\varepsilon \geq \mu/\sqrt{N}$. □