A Composition Theorem for Conical Juntas

Mika Göös\textsuperscript{*1} and T. S. Jayram\textsuperscript{2}

1 University of Toronto, Toronto, Canada
mika.goos@mail.utoronto.ca
2 IBM Research – Almaden, San Jose, USA
jayram@us.ibm.com

Abstract
We describe a general method of proving degree lower bounds for conical juntas (nonnegative combinations of conjunctions) that compute recursively defined boolean functions. Such lower bounds are known to carry over to communication complexity. We give two applications:

- **AND-OR trees.** We show a near-optimal $\tilde{\Omega}(n^{0.753...})$ randomised communication lower bound for the recursive NAND function (a.k.a. AND-OR tree). This answers an open question posed by Beame and Lawry [6, 23].
- ** Majority trees.** We show an $\Omega(2^{0.59k})$ randomised communication lower bound for the 3-majority tree of height $k$. This improves over the state-of-the-art already in the context of randomised decision tree complexity.

1 Conical Juntas?

Conical juntas are nonnegative linear combinations of conjunctions. Here are two examples, one computing the two-bit OR function OR: $\{0,1\}^2 \rightarrow \{0,1\}$ and another computing the three-bit majority function $\text{Maj}_3$: $\{0,1\}^3 \rightarrow \{0,1\}$:

\[ h_1(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1x_2 + \frac{1}{2}x_1\bar{x}_2, \]
\[ h_2(y) = \frac{1}{3}y_1y_2 + \frac{1}{3}y_2y_3 + \frac{1}{3}y_1y_3 + \frac{2}{3}y_1y_2y_3 + \frac{2}{3}y_1\bar{y}_2\bar{y}_3 + \frac{2}{3}y_1y_2\bar{y}_3. \]  

(1)

The purpose of this work is to prove lower bounds on the degree $\deg(h)$ (maximum width of a conjunction in $h$) of any conical junta $h$ that computes – even approximately – a given boolean function $f: \{0,1\}^n \rightarrow \{0,1\}$. More precisely, we study the $\epsilon$-approximate conical junta degree of $f$, denoted $\deg_\epsilon(f)$, that is defined as the minimum degree of a conical junta $h$ satisfying

$$ \forall x : \ |h(x) - f(x)| \leq \epsilon. $$

**Communication complexity connection.** A major motivation for studying conical junta degree comes from the works [10, 13, 24] that connect conical juntas with nonnegative rank, a basic measure in communication complexity. Roughly speaking, lower bounds on approximate conical junta degree of $f$ can be translated into lower bounds on the approximate nonnegative rank of a certain two-party “lift” of $f$, and therefore into lower bounds against randomised protocols.

\textsuperscript{*} Work done while at IBM Research Almaden.

© Mika Göös and T.S. Jayram;
Licensed under Creative Commons License CC-BY
31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 5; pp. 5:1–5:16
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
Related models. Conical juntas have been studied under such names as the (one-sided) partition bound for query complexity [15] and query complexity in expectation [20]. Another closely related model is that of randomised subcube partitions [11, 17, 21]. Moreover, if we restrict the coefficients in a conical junta to be 0-1, we obtain the model of subcube partitions a.k.a. unambiguous DNFs [30, 7, 12, 14, 21].

Our Results

Our main technical result is a Composition Theorem that makes it easy to prove conical junta degree lower bounds for functions that are defined from simpler functions via composition. If \( f \) and \( g \) are boolean functions on \( n \) and \( m \) bits, respectively, their composition \( f \circ g \) is the function on \( nm \) bits that maps an input \( x = (x_1, \ldots, x_n) \in \{0,1\}^n \) to the output

\[
(f \circ g^n)(x) := f(g(x_1), \ldots, g(x_n)).
\]

Define also \( f^{\circ k} := f \circ (f^{\circ (k-1)})^n \) where \( f^{\circ 1} := f \). The exact statement of the Composition Theorem is deferred to Section 4 as it is somewhat technical. It is phrased in terms of dual solutions (or certificates) to a linear program that captures a certain average version of conical junta degree (defined in Section 3). The theorem splits the task of proving lower bounds into two steps: we first need to find dual certificates for \( f \) and \( g \) (e.g., by solving an LP, either by inspection, or by using a computer), and then we can let the Composition Theorem construct a dual certificate for \( f \circ g^n \) in a black-box fashion. We note that similar LP-based approaches have been extremely popular in analysing the degree of multivariate polynomials (see [31, 32, 9] for recent examples) – in short, this work develops such a framework for conical juntas, a nonnegative analogue of multivariate polynomials.

Setting these technical matters aside for a moment, let us illustrate the power the Composition Theorem by looking at some of its consequences.

2.1 Query complexity

We give applications for two well-studied recursively defined boolean functions; see Figure 1.

- **Theorem 2.1.** \( \deg_\epsilon(\text{NAND}^{\circ k}) \geq \Omega(n^{0.753\ldots}) \) for all \( \epsilon \leq 1/n \) where \( n := 2^k \).

- **Theorem 2.2.** \( \deg_\epsilon(\text{Maj}_3^{\circ k}) \geq \Omega(2^{0.59\ldots k}) \) for all \( \epsilon \leq 1/n \) where \( n := 3^k \).

Discussion of Theorem 2.1. The function \( \text{NAND}^{\circ k} \) is computed by a height-\( k \) binary tree consisting of \( \text{NAND} \) gates (a.k.a. \( \text{AND-OR} \) tree). A classical result [28, 29] states that any randomised decision tree needs to query \( \Omega(n^{0.753\ldots}) \) (here \( 0.753\ldots = \log(1 + \sqrt{33}) - 2 \)) many input bits in order to compute \( \text{NAND}^{\circ k} \) with high probability. This matches an upper bound due to Snir [33]. Our Theorem 2.1 shows that the same lower bound holds even for conical juntas that approximate \( \text{NAND}^{\circ k} \) sufficiently well. This is a qualitative strengthening of the classical results since conical juntas are relaxations of decision trees. Indeed, a randomised decision tree of depth \( d \) that computes a function \( f \) to within error \( \epsilon > 0 \) can be converted into a degree-\( d \) \( \epsilon \)-approximate conical junta for \( f \) – the reason is the same as for multivariate polynomials [8, Theorem 15]. Speaking of polynomials, Theorem 2.1 should be compared with the fact that the approximate polynomial degree of \( \text{NAND}^{\circ k} \) is only \( O(\sqrt{n}) \) (and this upper bound holds even for quantum algorithms [5]).

Note: A caveat with Theorems 2.1–2.2 is that we only know how to prove them for \( \epsilon \leq 1/n \). By contrast, one usually takes \( \epsilon = 1/3 \) when studying decision trees, and this is
Figure 1 Examples of recursively defined boolean functions studied in this work.

well-known to be w.l.o.g., because the error can be reduced below any \( \epsilon < \frac{1}{3} \) with only a factor \( O(\log(1/\epsilon)) \) increase in query complexity. Interestingly, for conical juntas, it is known \([13]\) that \( \epsilon \) cannot always be efficiently reduced: for any constants \( \epsilon > \delta > 0 \) there exists a partial function \( f \) with \( \deg_\epsilon(f) = 1 \) but \( \deg_\delta(f) \geq \Omega(n) \). For total functions, it is still open whether efficient error reduction is possible (standard techniques \([8]\) at least show that \( \deg_\epsilon(f) \) is polynomially related to \( \deg_0(f) \)). In any case, Theorems 2.1–2.2 do indeed imply lower bounds for randomised decision trees with error \( \epsilon = \frac{1}{3} \): we simply have to reduce the error below \( 1/n \) first and only then convert the decision tree into a conical junta. This incurs a factor \( \Theta(\log n) \) loss in the value of the lower bound.

**Discussion of Theorem 2.2.** For the reasons discussed above, Theorem 2.2 implies a lower bound of \( \tilde{\Omega}(2^{0.59\ldots}) \geq \Omega(2^{0.59\ldots}) \) (here \( 2^{0.59\ldots} = \sqrt[3]{35}/2 \), and the \( \tilde{\Omega} \)-notation hides polylog(\( n \)) factors) for the randomised query complexity of the recursive majority function \( \text{Maj}_3^{\circ \circ k} \). This slightly improves over the previous bound of \( \Omega(2^{0.57\ldots}) \) that is the culmination of the line of work \([19, 22, 25, 27]\) wielding information theoretic tools. For comparison, a randomised zero-error decision tree of cost \( O(2^{0.65\ldots}) \) is known \([27]\). Even though our quantitative improvement in Theorem 2.2 is modest, the theorem nevertheless suggests that our new techniques are rather powerful: they are already competitive with highly optimised prior work, especially \([27]\).

**2.2 Communication complexity**

Using the machinery of \([13]\) we can now translate Theorems 2.1–2.2 into analogous communication results. The translation incurs some polylog(\( n \)) factor loss in parameters, which is suppressed by the \( \tilde{\Omega} \)-notation used below. Here \( \text{BPP}^{cc}(F) \) stands for the bounded-error communication complexity of \( F \) under a worst-case Alice–Bob bipartition of the input bits. For our functions, we may take the bipartition to be such that Alice gets the first bit of every bottom gate and Bob gets the rest.

- **Theorem 2.3.** \( \text{BPP}^{cc}(\text{NAND}^{\circ \circ k}) \geq \tilde{\Omega}(n^{0.753\ldots}) \).

- **Theorem 2.4.** \( \text{BPP}^{cc}(\text{Maj}_3^{\circ \circ k}) \geq \Omega(2^{0.59\ldots}) \).

**Discussion of Theorem 2.3.** The question of proving a lower bound for the randomised communication complexity of the balanced alternating \( \text{AND-OR} \) tree (with fan-in 2 gates next to the inputs) having \( n \) leaves was first posed by Beame and Lawry \([6, 23]\) to the best of our knowledge. They were interested in matching the randomised query complexity bound, towards separating randomized communication complexity from both nondeterministic and
co-nondeterministic communication complexity. Two independent works [18, 26] (building on [19]) arrived at a lower bound of $\Omega(n/2^{\Omega(k)})$ (or slightly worse $\Omega(n/k^{O(k)})$ in [18]) for the randomised communication complexity of any height-$k$ unbounded-fan-in alternating AND-OR tree (with fan-in 2 gates next to the inputs). While this lower bound is tight when $k = O(1)$, the bound becomes trivial in the setting of Theorem 2.3 where $k = \log n$. This shortcoming was partially addressed by [16] who showed, via a reduction from set-disjointness, a lower bound of $\Omega(\sqrt{n})$ for such AND-OR trees, independently of the height. Our Theorem 2.3 now gives an essentially optimal $\Omega(n^{0.753-})$ bound for the particular case of NAND$^k$. It remains open whether this lower bound holds for all AND-OR trees (with the appropriate gates next to the inputs). For query complexity, Amano [1] has come close to settling this question, known as the Saks–Wigderson conjecture [28] for the class of read-once formulas (a more general version of the conjecture was recently disproved [4]).

**Discussion of Theorem 2.4.** The function $\text{Maj}_k^2$ has not been studied in communication complexity previously – after all, even its randomised query complexity is not yet completely understood.

## 3 Definitions and Examples

We write $h = \sum w_C C$ for a generic conical junta, where the sum ranges over different conjunctions of literals $C: \{0,1\}^n \to \{0,1\}$ and $w_C \geq 0$ for each $C$. Note that $h: \{0,1\}^n \to \mathbb{R}_{\geq 0}$. Let $|C|$ denote the width of a conjunction $C$, i.e., the number of literals in $C$. The degree of $h$, denoted $\deg(h)$, is defined as the maximum width of a conjunction $C$ with $w_C > 0$. Here, it is helpful to work with a more robust notion of degree that we call *average degree*. The average degree of $h$, denoted $\text{adeg}(h)$, is defined as the maximum over all inputs $x$ of

$$\text{adeg}_x(h) := \sum w_C |C(x)| = \sum w_C \text{adeg}_x(C).$$

In particular, $\text{adeg}(h) \leq \deg(h)$ in the natural setting where $h(x) \leq 1$ for all $x$. Our definition of average degree is in perfect analogy to the usual definition of cost for randomised zero-error decision trees, namely, charging for the *expected* number of queries made on a given input. Indeed, it is not hard to see that any zero-error decision tree of cost $d$ gives rise to a conical junta of average degree $d$ computing exactly the same boolean function as the decision tree.

For a boolean function $f: \{0,1\}^n \to \{0,1\}$ we define

- **Degree**: $\deg(f)$ is the minimum $\deg(h)$ over all conical juntas $h$ computing $f$.
- **Average degree**: $\text{adeg}(f)$ is the minimum $\text{adeg}(h)$ over all conical juntas $h$ computing $f$.
- **Approximate degree**: $\text{deg}_\epsilon(f)$ is the minimum $\deg(h)$ over all conical juntas $h$ that compute $f$ to within error $\epsilon$, i.e., $h(x) \in f(x) \pm \epsilon$ for all $x$.

### 3.1 Tame examples

For our conical juntas $h_1$ and $h_2$ from (1), we have $\text{adeg}(h_1) = \text{adeg}_{10}(h_1) = 3/2 < 2 = \deg(h_1)$ and $\text{adeg}(h_2) = \text{adeg}_{110}(h_2) = 8/3 < 3 = \deg(h_2)$. In fact, $h_1$ and $h_2$ are optimal:

$$\text{adeg}(\text{OR}) = 3/2 \quad \text{and} \quad \text{adeg}(\text{Maj}_3) = 8/3.$$

This can be seen by solving an LP whose value is $\text{adeg}(f)$, as is discussed shortly. Note that our degree measures are inherently *one-sided*: $f$ and its negation $\neg f$ need not have the same
degree. For example, we have \( \text{adeg}(\neg \text{OR}) = 2 \) (observe that \( \bar{x}_1 \bar{x}_2 \) is the only conical junta for \( \neg \text{OR} \)) even though \( \text{adeg}(\text{OR}) = 3/2 \). (More dramatic gaps can be demonstrated using variations of a function introduced in [14].) By contrast, \( \text{Maj}_3 \) is self-dual, \( \neg \text{Maj}_3(x_1, x_2, x_3) = \text{Maj}_3(\neg x_1, \neg x_2, \neg x_3) \), so we automatically have \( \text{adeg}(\text{Maj}_3) = \text{adeg}(\neg \text{Maj}_3) \).

### 3.2 A wild example!

What is the average degree of \( \text{OR} \circ \text{Maj}_3^2 \)? We can obtain a conical junta for this function starting with the optimal conical juntas \( h_1(x) \), \( h_2(y) \), \( h_2(y) := h_2(\bar{y}_1, \bar{y}_2, \bar{y}_3) \) computing \text{OR}, \( \text{Maj}_3 \), \( \neg \text{Maj}_3 \), respectively, as follows: Let \( z^1 = (z_{11}^1, z_{12}^1, z_{13}^1) \) and \( z^2 = (z_{21}^2, z_{22}^2, z_{23}^2) \) be fresh variables. Start with \( h_1(x) \) and replace every positive literal \( x_i \) by \( h_2(z^1) \) and every negative literal \( \bar{x}_i \) by \( h_2(z^2) \). This construction shows that

\[
\text{adeg}(\text{OR} \circ \text{Maj}_3^2) \leq 3/2 \cdot 8/3 = 4.
\]

It would be natural to conjecture that this is tight — but this conjecture is false! There is in fact a more effective conical junta of average degree only \( 47/12 \approx 3.92 \). An analogous phenomenon is well-known in the context of zero-error decision trees: so-called directional decision trees need not be optimal for composed functions [28, 34, 2].

**What of it?** This example shows that we cannot hope for a perfect composition theorem for average degree that would determine \( \text{adeg}(f \circ g^m) \) solely in terms of \( \text{adeg}(f) \), \( \text{adeg}(g) \), and \( \text{adeg}(\neg g) \), even assuming \( \text{adeg}(g) = \text{adeg}(\neg g) \). Consequently, for our LP-based Composition Theorem, we will have to introduce some technical assumptions: to enable the construction of a dual certificate for \( \text{adeg}(f \circ g^m) \), we assume we have dual certificates of a special form for \( \text{adeg}(f) \), \( \text{adeg}(g) \), \( \text{adeg}(\neg g) \). The rest of this section develops our LP formalism for average degree.

### 3.3 Generalised input costs

Let us first generalise the definition of \( \text{adeg}(h) \) by allowing arbitrary costs \( b_0, b_1 \geq 0 \) to be assigned to reading the input bits. That is, for a conjunction \( C \), we set \( |C|_{b_0, b_1} := b_0 \cdot (\# \text{ of } 0\text{'s read by } C) + b_1 \cdot (\# \text{ of } 1\text{'s read by } C) \). In particular, \( |C|_{1, 1} = |C| \). Then \( \text{adeg}(h; b_0, b_1) \) is defined as the maximum over all inputs \( x \)

\[
\text{adeg}_x(h; b_0, b_1) := \sum w_C |C|_{b_0, b_1} C(x) = \sum w_C \text{adeg}_x(C; b_0, b_1).
\]

We also introduce some “distributional” notation: for a distribution \( D_1 \) over \( f^{-1}(1) \) we let

\[
\text{adeg}_{D_1}(h; b_0, b_1) := \mathbb{E}_{x \sim D_1} \left[ \text{adeg}_x(h; b_0, b_1) \right].
\]

For a boolean function \( f: \{0, 1\}^n \rightarrow \{0, 1\} \) we define

- \( \text{adeg}(f; b_0, b_1) \) is the minimum of \( \text{adeg}(h; b_0, b_1) \) over all conical juntas \( h \) computing \( f \).
- \( \text{adeg}_{D_1}(f; b_0, b_1) \) is the minimum of \( \text{adeg}_{D_1}(h; b_0, b_1) \) over all conical juntas \( h \) computing \( f \). It is clear that \( \text{adeg}(f; b_0, b_1) \geq \text{adeg}_{D_1}(f; b_0, b_1) \) for all distributions \( D_1 \). (In fact, it can be shown using the minimax theorem that this inequality can be turned into an equality if we maximise over \( D_1 \) on the right hand side – however, we do not use this fact.)
3.4 An LP for average degree

We formulate \( \text{adeg}_{D_1}(f; b_0, b_1) \) as the optimum value of an LP – here the data \( f, D_1, b_0, b_1, \) is thought of as fixed. We have a nonnegative variable \( w_C \geq 0 \) for each of the \( 3^n \) possible conjunctions \( C: \{0, 1\}^n \rightarrow \{0, 1\} \). Here is the LP:

\[
\begin{align*}
\text{min} & \quad \text{adeg}_{D_1}(\sum w_C C; b_0, b_1) \\
\text{subject to} & \quad \sum w_C C(x) = f(x), \quad \forall x \\
& \quad w_C \geq 0, \quad \forall C
\end{align*}
\]

(Primal)

Here is the LP dual; the free variables are packaged into a function \( \Psi: \{0, 1\}^n \rightarrow \mathbb{R} \).

\[
\begin{align*}
\text{max} & \quad \langle \Psi, f \rangle \\
\text{subject to} & \quad \langle \Psi, C \rangle \leq \text{adeg}_{D_1}(C; b_0, b_1), \quad \forall C \\
& \quad \Psi(x) \in \mathbb{R}, \quad \forall x
\end{align*}
\]

(Dual)

Since we are interested in proving lower bounds on average degree, we are only going to need the “weak” form of LP duality: Suppose \( h = \sum w_C C \) is an optimal solution to (Primal). Then any solution \( \Psi \) that is feasible for (Dual) witnesses a lower bound on \( \text{adeg}(f; b_0, b_1) \) like so:

\[
\begin{align*}
\text{adeg}(f; b_0, b_1) \geq & \quad \text{adeg}_{D_1}(f; b_0, b_1) \\
& = \text{adeg}_{D_1}(h; b_0, b_1) \\
& = \sum w_C \text{adeg}_{D_1}(C; b_0, b_1) \\
& \geq \sum w_C \langle \Psi, C \rangle \\
& = \langle \Psi, \sum w_C C \rangle \\
& = \langle \Psi, f \rangle. \\
\end{align*}
\]

(2)

4 Statement of the Composition Theorem

We start by defining an \((a_0, a_1; b_0, b_1)-certificate\) for \( f \) as a special collection of certificates witnessing

\[
\begin{align*}
\text{adeg}(f; b_0, b_1) \geq & \quad a_1, \\
\text{adeg}(-f; b_0, b_1) \geq & \quad a_0.
\end{align*}
\]

(3)

▶ Definition 4.1. Call a function \( \Psi: \{0, 1\}^n \rightarrow \mathbb{R} \) balanced if \( \sum x \Psi(x) = 0 \), and also write \( X_{\geq 0} := \max \{X, 0\} \) for short. An \((a_0, a_1; b_0, b_1)-certificate\) for a function \( f: \{0, 1\}^n \rightarrow \{0, 1\} \) consists of four balanced functions \( \{\Psi_v, \hat{\Psi}_v\}_{v=0,1} \) mapping \( \{0, 1\}^n \rightarrow \mathbb{R} \) such that the following hold.

= Special form: Functions \( \Psi_0 \) and \( \Psi_1 \) have the form

\[
\Psi_v = a_v (D_v - D_{1-v}),
\]

(4)

where \( D_v \) is a distribution over \( f^{-1}(v) \). Moreover, \( \hat{\Psi}_v \) is supported on \( f^{-1}(v) \).
Four Feasibility: For all conjunctions $C$ and $v \in \{0, 1\}$,
\[
\langle \Psi_v, C \rangle_{\geq 0} + \langle \hat{\Psi}_v, C \rangle \leq \text{adeg}_{D_v}(C; b_0, b_1).
\]  
\[
\text{(5)}
\]

\textbf{Theorem 4.2} (Composition Theorem). Suppose $f$ admits an $(a_0, a_1; b_0, b_1)$-certificate and $g$ admits a $(b_0, b_1; 1, 1)$-certificate. Then $f \circ g^n$ admits an $(a_0, a_1, 1, 1)$-certificate.

\textbf{Discussion.} First, we note that (5) actually packs together two linear inequalities; it would be equivalent to require that both $\Psi_v + \hat{\Psi}_v$ and $\hat{\Psi}_v$ are feasible for (Dual), namely that
\[
\begin{cases}
\langle \Psi_v + \hat{\Psi}_v, C \rangle \leq \text{adeg}_{D_v}(C; b_0, b_1), \\
\langle \hat{\Psi}_v, C \rangle \leq \text{adeg}_{D_v}(C; b_0, b_1).
\end{cases}
\]  
\[
\text{(5')}
\]

Here $\Psi_1 + \hat{\Psi}_1$ is the main attraction: it witnesses a lower bound of $(\Psi_1 + \hat{\Psi}_1, f) = (\Psi_1, f) + (\hat{\Psi}_1, f) = a_1 + 0 = a_1$ for $\text{adeg}(f; b_0, b_1)$ as promised above (3); similarly, $\Psi_0 + \hat{\Psi}_0$ witnesses the complementary lower bound $\text{adeg}(-f; b_0, b_1) \geq a_0$.

The requirement that $\Psi_1 + \hat{\Psi}_1$ must be balanced is perhaps our most critical assumption. We use it to manoeuvre around the counterexample of Section 3.2: we have $\text{adeg}(\text{Maj}_3) = 8/3$, while the best balanced solution to (Dual) only witnesses the lower bound $\text{adeg}(\text{Maj}_3) \geq 5/2$ (see also Figure 3). The requirement that $\hat{\Psi}_v$ is feasible for (Dual) is merely a technical assumption that helps us in the upcoming proof (akin to a “strengthened induction hypothesis”); we do not know whether the theorem is true without this condition. Another technical assumption is (4), which allows us to assume that $\Psi_1$ and $\Psi_0$ have opposite signs:
\[
\Psi_1 = -a_1/a_0 \cdot \Psi_0.
\]

Some simple certificates are illustrated in Figures 2–3. Their feasibility can be checked by hand. For more involved functions, certificates can in principle be found via a computer search (using computers is not uncommon even in “lower bounds” research [3]). We will in fact use this approach for $\text{Maj}_3^{ck}$ in Section 6.

5 Proof of the Composition Theorem

Let $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ and $\{\Phi_v, \hat{\Phi}_v\}_{v=0,1}$ be the certificates for $f$ and $g$, respectively. Our goal is to construct a certificate $\{T_v, \hat{T}_v\}_{v=0,1}$ for $f \circ g^n$. We use the following notation:
\[
\begin{align*}
\Psi_v &:= a_v(F_v - F_{1-v}), \\
\Phi_v &:= b_v(G_v - G_{1-v}), \\
T_v &:= a_v(D_v - D_{1-v}).
\end{align*}
\]

By assumption, the distribution $F_v$ is supported on $f^{-1}(v)$ and $G_v$ is supported on $g^{-1}(v)$. We will define $D_v$ to be supported on $(f \circ g^n)^{-1}(v)$.

5.1 Construction

\textbf{Lifts.} Let $\Gamma: \{0, 1\}^n \to \mathbb{R}$ and suppose that for each $y \in \{0, 1\}^n$ we have a function $H_y: \{0, 1\}^{mn} \to \mathbb{R}$ supported on $(g^n)^{-1}(y) = g^{-1}(y_1) \times \cdots \times g^{-1}(y_n)$. The lift of $\Gamma$ by $H$ is
\[
\Gamma^H := \sum_{y \in \{0, 1\}^n} \Gamma(y) \cdot H_y.
\]

In particular, if $\Gamma$ and the $H_y$’s are probability distributions, so is $\Gamma^H$. Note also that if $\Gamma$ is supported on $f^{-1}(v)$, then $\Gamma^H$ is supported on $(f \circ g^n)^{-1}(v)$. 

\textbf{CCC 2016}
A Composition Theorem for Conical Juntas

Figure 2 A \((2b_1, b_0 + \frac{1}{2}b_1; b_0, b_1)\)-certificate for NAND: \(\{0, 1\}^2 \rightarrow \{0, 1\}\) that is valid for all \(b_0, b_1 \geq 0\). The 1-inputs NAND\(^{-1}\)(1) are highlighted in gray. For feasibility, there are 6 equivalence classes (see Section 6.2) of conjunctions to check: \(\{***,*,+1,00,10,11\}\).

Figure 3 A \((\frac{5}{2}, \frac{5}{2}; 1,1)\)-certificate for Maj\(_3\): \(\{0, 1\}^3 \rightarrow \{0, 1\}\). The 1-inputs Maj\(_3\)^{-1}(1) are highlighted in gray. Only \(\Psi_1, \hat{\Psi}_1\) are shown as \(\Psi_0, \hat{\Psi}_0\) are defined via self-duality. Here \(D_v\) is uniform on inputs of Hamming weight \(v + 1\). For feasibility, there are 10 equivalence classes of conjunctions to check: \(\{***, **+, **0, *00, *11, 111, 000, 100, 110, 111\}\). Note that for any \(\alpha \geq 0\), we can obtain an \((\frac{5}{2}\alpha, \frac{5}{2}\alpha; \alpha, \alpha)\)-certificate by simply scaling the functions \(\Psi_v, \hat{\Psi}_v\) by \(\alpha\).
New certificate. Write $G_{y} := G_{y_{1}} \times \cdots \times G_{y_{n}}$ for the canonical product distribution on $(g^{n})^{-1}(y)$. We also need a modified version of $G_{y}$, denoted $(G^{c} \Phi)_{y}$ where $i \in [n]$, that has a copy of $\Phi_{y_{i}}$ in place of $G_{y_{i}}$; more formally

$$(G^{c} \Phi)_{y}(x) := \Phi_{y_{i}}(x_{i}) \cdot \prod_{j \neq i} G_{y_{j}}(x_{j}).$$

Note that $(G^{c} \Phi)_{y}$ is a balanced function supported on $(g^{n})^{-1}(y)$.

We now define $\{\Psi_{v}, \hat{\Psi}_{v}\}_{v=0,1}$ by

$$\Psi_{v} := \Psi_{v}^{G}, \quad \hat{\Psi}_{v} := \hat{\Psi}_{v}^{G} + \sum_{i=1}^{p} F_{v}^{G^{c} \Phi}. \quad (6)$$

Since $\Psi_{v}^{G} = a_{v}(F_{v}^{G} - F_{v}^{G^{c} \Phi})$, we have $D_{v} = F_{v}^{G^{c} \Phi}$. It is also easy to check that $\hat{\Psi}_{v}$ is a balanced function supported on $(f \circ g^{n})^{-1}(v)$. Hence $\{\Psi_{v}, \hat{\Psi}_{v}\}_{v=0,1}$ is of the special form required of an $(a_{0}, a_{1}; 1, 1)$-certificate for $f \circ g^{n}$. The interesting part is to verify the feasibility condition (5).

5.2 Feasibility

Fix a conjunction $C$ in the domain of $f \circ g^{n}$. Our goal is to show

$$\langle \Psi_{v}^{G}, C \rangle \geq 0 \implies \langle \hat{\Psi}_{v}^{G} + \sum_{i=1}^{p} F_{v}^{G^{c} \Phi}, C \rangle \leq \text{adeg}_{D_{v}}(C). \quad (7)$$

Extracting a conical junta from $C$. Our analysis will be centered around a conical junta $h(y)$, defined below, that computes the acceptance probability $\Pr_{x \sim G_{y}}[C(x) = 1] = \mathbb{E}_{x \sim G_{y}}[C(x)] = (G_{y}, C)$. In a certain sense, $h$ serves as a projection of $C$ to the domain of $f$. Write $C(x) = \prod_{i=1}^{n} C_{i}(x_{i})$ where $C_{i}$ is a conjunction depending only on $x_{i}$. Since $G_{y}$ is a product distribution,

$$\langle G_{y}, C \rangle = \prod_{i=1}^{n} \langle G_{y_{i}}, C_{i} \rangle =: \prod_{i} p_{i,y_{i}},$$

where we wrote $p_{i,v} := \langle G_{v}, C_{i} \rangle \in \mathbb{R}_{\geq 0}$ for short. Fix $y^{*} \in \{0, 1\}^{n}$ such that $p_{i,y_{i}^{*}} \geq p_{i,1-y_{i}^{*}}$ for all $i$. We now define $h(y)$ that computes $\langle G_{y}, C \rangle$:

$$h(y) := \prod_{i=1}^{n} \left( p_{i,1-y_{i}^{*}} + (p_{i,y_{i}^{*}} - p_{i,1-y_{i}^{*}}) \cdot \ell_{i} \right) \geq 0,$$  \quad (8)

where literal $\ell_{i}$ is $\begin{cases} y_{i} & \text{if } y_{i}^{*} = 1, \\ \bar{y}_{i} & \text{if } y_{i}^{*} = 0. \end{cases}$

This product expression can be expanded into a conical combination of conjunctions, $h = \sum_{w \in T} T$, in the natural way, but the above “implicit” form is more concise.

Next, we record two properties of $h$ that will suffice for the remaining analysis.

Lemma 5.1. $\text{adeg}_{y}(h; b_{0}, b_{1}) = \sum_{i} \langle \Phi_{y_{i}}, C_{i} \rangle \geq 0 \prod_{j \neq i} \langle G_{y_{j}}, C_{j} \rangle$.

Proof. Write $h = \sum_{w \in T} T$. We compute the average degree by summing together the weights $\sum_{T \ni \ell_{i}} w_{T}(y)$ contributed by each of the $n$ literals $\ell_{i}$, i.e.,

$$\text{adeg}_{y}(h; b_{0}, b_{1}) = \sum_{i} |\ell_{i}| b_{0} b_{1} \cdot \sum_{T \ni \ell_{i}} w_{T}(y).$$

If $i$ is such that $y_{i} \neq y_{i}^{*}$, we have $\ell_{i}(y) = 0$ and so $T(y) = 0$ for all $T \ni \ell_{i}$; hence $\ell_{i}$ contributes no weight in this case. Suppose then that $i$ is such that $y_{i} = y_{i}^{*}$; here we can write

$$h(y) = p_{i,1-y_{i}} \prod_{j \neq i} p_{j,y_{j}} + \ell_{i} \cdot (p_{i,y_{i}} - p_{i,1-y_{i}}) \prod_{j \neq i} p_{j,y_{j}},$$

where $\ell_{i} = p_{i,y_{i}} - p_{i,1-y_{i}}$. We now define $\Psi_{v}^{G}, \hat{\Psi}_{v}^{G}$ by

$$\Psi_{v}^{G} := \Psi_{v}^{G}, \quad \hat{\Psi}_{v}^{G} := \hat{\Psi}_{v}^{G} + \sum_{i=1}^{p} F_{v}^{G^{c} \Phi}. \quad (6)$$
The conjunctions $T$ underlying the first term do not involve $\ell_i$, so they contribute no weight for $\ell_i$. The conjunctions $T$ underlying the second term all involve $\ell_i$ and contribute a total weight of $(p_{i,y_i} - p_{i,1-y_i}) \prod_{j \neq i} p_{j,y_j}$. Altogether we get

$$\text{adeg}_y(h; b_0, b_1) = \sum_i |\ell_i| b_0 b_1 \cdot \sum_{T \subseteq \ell_i} w_T T(y) = \sum_{\ell_i=y_i} b_y (p_{i,y_i} - p_{i,1-y_i}) \prod_{j \neq i} p_{j,y_j} = \sum_i b_y (p_{i,y_i} - p_{i,1-y_i}) \prod_{j \neq i} p_{j,y_j} = \sum_i (b_y (G_{y_i} - G_{1-y_i})) \prod_{j \neq i} (G_{y_j}, C_j) = \sum_i (\Phi_{y_i}, C_i) \prod_{j \neq i} (G_{y_j}, C_j).$$

\[ \square \]

**Lemma 5.2.** $\langle \Gamma, h \rangle = \langle \Gamma^G, C \rangle$ for all $\Gamma : \{0,1\}^n \to \mathbb{R}$.

**Proof.** We calculate

$$\langle \Gamma, h \rangle = \sum_y \Gamma(y) h(y) = \sum_y \Gamma(y) (G_y, C) = \sum_y \Gamma(y) [\sum_x G_y(x) C(x)] = \sum_x [\sum_y \Gamma(y) G_y(x) C(x)] = \sum_x \Gamma^G(x) C(x) = \langle \Gamma^G, C \rangle.$$ 

\[ \square \]

**Analysis.** Let us expand the right hand side of the desired inequality (7):

$$\text{adeg}_{D_x}(C) = |C| \cdot \langle F^G_{\Psi}, C \rangle = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ |C| \cdot \langle G_y, C \rangle \right] = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \left( \sum_i |C_i| \right) \prod_i \langle G_{y_i}, C_i \rangle \right] = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \sum_i |C_i| \langle G_{y_i}, C_i \rangle \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \right] = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \sum_i \text{adeg}_{G_{y_i}}(C_i) \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \right].$$

Substituting our hypothesis $\text{adeg}_{G_{y_i}}(C_i) \geq \langle \Phi_{y_i}, C_i \rangle \geq 0 + \langle \hat{\Phi}_{y_i}, C_i \rangle$ into the above, we obtain

$$\text{adeg}_{D_x}(C) \geq \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \sum_i \langle \Phi_{y_i}, C_i \rangle \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \right] + \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \sum_i \langle \hat{\Phi}_{y_i}, C_i \rangle \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \right].$$

For the first term,

$$\text{(I)} = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \text{adeg}_y(h; b_0, b_1) \right] \quad \text{(Lemma 5.1)}$$

$$= \mathbf{E}_{y \sim F^G_{\Psi}} (h; b_0, b_1) \geq \langle \Psi_v, h \rangle \geq 0 + \langle \hat{\Psi}_v, h \rangle \quad \text{(Feasibility of $\{\Psi_v, \hat{\Psi}_v\}$ and (2))}$$

$$= \langle \Psi^G_v, C \rangle \geq 0 + \langle \hat{\Psi}^G_v, C \rangle.$$  

For the second term,

$$\text{(II)} = \mathbf{E}_{y \sim F^G_{\Psi}} \left[ \sum_i \langle G_{\ell \Psi}^G(h), C \rangle \right]$$

$$= \langle \sum_i F^G_{\Psi} \hat{\Phi}, C \rangle.$$ 

Combining these yields (7). This concludes the proof of Theorem 4.2.
\section{Approximate Degree Lower Bounds}

In this section we prove Theorems 2.1–2.2 using the Composition Theorem. We begin by observing that \((a_0, a_1; b_0, b_1)-certificates\) \(\{\Psi_v, \tilde{\Psi}_v\}_{v=0,1}\) also yield lower bounds for approximate degree, if the 1-norm \(\|\tilde{\Psi}_1\|\) is not too large. We call \(\{\Psi_v, \tilde{\Psi}_v\}_{v=0,1}\) an \((a_0, a_1; b_0, b_1; c)-certificate\) if \(\max_v \|\tilde{\Psi}_v\|_1 \leq c\).

\begin{lemma}
Suppose \(f\) admits an \((a_0, a_1; 1, 1; c)-certificate.\) If \(\epsilon \leq 1/4\) and \(c \cdot \epsilon \leq a_1/4,\) then \(\deg_c(f) \geq \Omega(a_1).\)
\end{lemma}

\begin{proof}
Fix a certificate \(\{\Psi_v, \tilde{\Psi}_v\}_{v=0,1}\) for \(f\) and suppose \(\deg_c(f) = \deg(h)\) where \(h\) is a conical junta with \(\|h - f\|_\infty \leq c.\) Since \(b(x) \leq 1 + \epsilon\) for all \(x,\) we have \(\deg(h) \geq (1 + \epsilon)^{-1} \deg(h) \geq \Omega(\deg(h)).\) Now we calculate
\[
\deg(h) \geq \langle \Psi_1 + \tilde{\Psi}_1, h \rangle \quad \text{(as in (2))}
\]
\[
= \langle \Psi_1 + \tilde{\Psi}_1, f \rangle + \langle \Psi_1 + \tilde{\Psi}_1, h - f \rangle
\]
\[
\geq a_1 - \|\langle \Psi_1 + \tilde{\Psi}_1, h - f \rangle\|
\]
\[
\geq a_1 - \|\tilde{\Psi}_1 + \Psi_1\| \cdot \|h - f\|_\infty
\]
\[
\geq a_1 - (\|\Psi_1\| + \|\tilde{\Psi}_1\|) \cdot \epsilon
\]
\[
\geq a_1 - (2a_1 + c) \cdot \epsilon
\]
\[
\geq a_1/4.
\]
\end{proof}

We use the following version of the Composition Theorem where the bounds on 1-norms (following immediately from the definition (6)) are made explicit.

\begin{theorem}
Suppose \(f\) admits an \((a_0, a_1; b_0, b_1; c)-certificate\) and \(g\) admits a \((b_0, b_1; 1, 1; d)-certificate.\) Then \(f \circ g^a\) admits an \((a_0, a_1; 1, 1; c + nd)-certificate.\)
\end{theorem}

\subsection{Proof of Theorem 2.1}
We iteratively apply Theorem 6.2 as follows.

1. Assume we have an \((\alpha_k, \beta_k; 1, 1; \gamma_k)-certificate\) for NAND\(^{\alpha_k}\) where \(\gamma_k \geq \alpha_k, \beta_k.\)
2. Obtain a \((2\beta_k, \alpha_k + \frac{1}{2}\beta_k; \alpha_k, \beta_k; \beta_k)-certificate\) for NAND from Figure 2.
3. Compose the above to get an \((\alpha_{k+1}, \beta_{k+1}; 1, 1; \gamma_{k+1})-certificate\) for NAND\(^{\alpha_{(k+1)}}\) where
\[
\alpha_{k+1} := 2\beta_k,
\]
\[
\beta_{k+1} := \alpha_k + \beta_k/2,
\]
\[
\gamma_{k+1} := \beta_k + 2\gamma_k.
\]

Note that \(\alpha_{k+1}, \beta_{k+1} \leq \gamma_{k+1} \leq 3\gamma_k.\) Starting with \(\alpha_0 = \beta_0 = \gamma_0 = 1\) these recurrences (famously [28]) evaluate to \(\alpha_k, \beta_k = \Theta(n^{0.753-})\) where \(n := 2^k.\) In addition, \(\gamma_k \leq 3^k \leq n^{1.6}\) Now take \(\epsilon \leq 1/n\) in Lemma 6.1 to prove Theorem 2.1.

\subsection{Computer search for certificates}
Iteratively composing (scaled versions of) the \((5/2, 5/2; 1, 1)-certificate\) given in Figure 3 would yield only an \(\Omega(2^{5k})\) lower bound for \(\text{Maj}^k.\) This is the best possible for our approach if we were to just compose certificates for individual \(\text{Maj}_j\) functions. To obtain a better lower
A Composition Theorem for Conical Juntaς

Table 1 Certificates for $\text{Maj}_3^\ell$ for heights $\ell = 1, 2, 3$. The table lists $(\alpha_1, \alpha_3; 1, 1)$-certificates with values $\alpha_1 = 5/2$ (also illustrated in Figure 3), $\alpha_2 = 20/3$, and $\alpha_3 = 35/2$. Only $\Psi_1, \hat{\Psi}_1$ are shown as $\Psi_0, \hat{\Psi}_0$ are defined dually. We give the total weight for each equivalence class of inputs; the functions are uniform on each class. For height $\ell = 3$ we represent the inputs to the bottom-most $\text{Maj}_3$ gates by their Hamming weight, e.g., 001 $\sim$ 1, 011 $\sim$ 2, etc.

<table>
<thead>
<tr>
<th>Function</th>
<th>Class representative</th>
<th>Class size</th>
<th>$\Psi_1$</th>
<th>$\hat{\Psi}_1$</th>
<th>$\Psi_1 + \hat{\Psi}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Maj}_3^1$</td>
<td>(0, 0, 1)</td>
<td>3</td>
<td>$-5/2$</td>
<td>0</td>
<td>$-5/2$</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1)</td>
<td>3</td>
<td>$5/2$</td>
<td>$1/2$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(1, 1, 1)</td>
<td>1</td>
<td>0</td>
<td>$-1/2$</td>
<td>$-1/2$</td>
</tr>
<tr>
<td></td>
<td>All others</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\text{Maj}_3^2$</td>
<td>(001, 001, 011)</td>
<td>81</td>
<td>$-20/3$</td>
<td>0</td>
<td>$-20/3$</td>
</tr>
<tr>
<td></td>
<td>(001, 011, 011)</td>
<td>81</td>
<td>$20/3$</td>
<td>$7/3$</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>(000, 011, 011)</td>
<td>27</td>
<td>0</td>
<td>$-1/3$</td>
<td>$-1/3$</td>
</tr>
<tr>
<td></td>
<td>(001, 011, 111)</td>
<td>54</td>
<td>0</td>
<td>$-2/3$</td>
<td>$-2/3$</td>
</tr>
<tr>
<td></td>
<td>(011, 011, 011)</td>
<td>27</td>
<td>0</td>
<td>$-4/3$</td>
<td>$-4/3$</td>
</tr>
<tr>
<td></td>
<td>All others</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\text{Maj}_3^3$</td>
<td>(112, 112, 122)</td>
<td>1594323</td>
<td>$-35/2$</td>
<td>0</td>
<td>$-35/2$</td>
</tr>
<tr>
<td></td>
<td>(112, 122, 122)</td>
<td>1594323</td>
<td>$35/2$</td>
<td>$19/2$</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>(122, 122, 122)</td>
<td>531441</td>
<td>0</td>
<td>$-7/2$</td>
<td>$-7/2$</td>
</tr>
<tr>
<td></td>
<td>(112, 122, 222)</td>
<td>1062882</td>
<td>0</td>
<td>$-2/3$</td>
<td>$-2/3$</td>
</tr>
<tr>
<td></td>
<td>(112, 122, 123)</td>
<td>2125764</td>
<td>0</td>
<td>$-4/3$</td>
<td>$-4/3$</td>
</tr>
<tr>
<td></td>
<td>(112, 122, 022)</td>
<td>1062882</td>
<td>0</td>
<td>$-2/3$</td>
<td>$-2/3$</td>
</tr>
<tr>
<td></td>
<td>(111, 122, 122)</td>
<td>531441</td>
<td>0</td>
<td>$-5/6$</td>
<td>$-5/6$</td>
</tr>
<tr>
<td></td>
<td>(113, 122, 122)</td>
<td>531441</td>
<td>0</td>
<td>$-1/2$</td>
<td>$-1/2$</td>
</tr>
<tr>
<td></td>
<td>(012, 122, 122)</td>
<td>1062882</td>
<td>0</td>
<td>$-2/3$</td>
<td>$-2/3$</td>
</tr>
<tr>
<td></td>
<td>All others</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

bound, we can instead directly find a certificate for $\text{Maj}_3^\ell$ where $\ell$ is a small constant, and then compose that certificate. Table 1 gives certificates for $\text{Maj}_3^\ell$ for height up to $\ell = 3$. We used a computer to solve the dual LP (Dual), with the additional restriction that $\Psi = \Psi_1 + \hat{\Psi}_1$ should be balanced. The best balanced $\Psi$ happened to satisfy the other conditions required by our Definition 4.1.

Notes on implementation. For computational efficiency, it is useful to prune the search space by eliminating symmetries. The symmetries of $\text{Maj}_3^\ell$ (permutations of input coordinates that do not change the value of the function) are the symmetries of the underlying height-$\ell$ ternary tree. These symmetries partition the set of inputs and the set of conjunctions into equivalence classes: two inputs/conjunctions are “equivalent” if one can be mapped to the other by a symmetry. The set of feasible solutions to the LP is also invariant under these symmetries. It follows that we may look w.l.o.g. for a $\Psi$ that is invariant, i.e., uniform on each equivalence class. (Indeed, if $\Psi$ is any feasible solution, we obtain an invariant solution by averaging $\Psi$ over all the symmetries.) Thus we need only maintain one variable in the LP per equivalence class $X \subseteq \{0, 1\}^n$ recording the total weight $\sum_{x \in X} \Psi(x)$ of that class. Also, for such invariant $\Psi$, we need only check the LP feasibility constraint $\langle \Psi, C \rangle \leq \text{adeg}_{D_1}(C; b_0, b_1)$ for a single representative $C$ from each class of conjunctions.

The optimal height-2 certificate happens to have the same support as the certificate produced by our Composition Theorem starting with two height-1 certificates. Inspired
by this, in order to speed up the search for height 3, we only optimised over those \( \Psi \) whose support coincides with that coming from the Composition Theorem – this LP has only 9 variables (i.e., equivalence classes of inputs), but well over 100,000 constraints (i.e., equivalence classes of conjunctions).

It is open to analyse height 4. Is there an efficient separation oracle for (Dual)?

### 6.3 Proof of Theorem 2.2

Table 1 defines a certificate for \( \text{Maj}_3^3 \) with parameters \((35/2, 35/2; 1, 1; 19)\) and we may scale the certificate by any scalar \( \alpha \geq 0 \) to obtain one with parameters \((\alpha 35/2, \alpha 35/2; \alpha, \alpha; 19\alpha)\).

Using Theorem 6.2 iteratively as in Section 6.1, we get a certificate for \( \text{Maj}_3^3 \) with parameters

\[
((35/2)^{k/3}, (35/2)^{k/3}; 1, 1; 28^{k/3} \cdot 19).
\]

Here \((35/2)^{k/3} \geq n^{0.8} \) and \(28^{k/3} \cdot 19 \leq n^{1.1}\) where \( n = 3^k \). Hence we may apply Lemma 6.1 with \( \epsilon \leq 1/n \) to conclude an \( \epsilon \)-approximate degree lower bound of \( \Omega((35/2)^{k/3}) = \Omega(2.59 \ldots ^k) \).

### 7 Communication Lower Bounds

In this section we prove Theorems 2.3–2.4 by applying the main result of [13]: a simulation of randomised communication protocols by conical juntas. To this end, let \( \text{IP}_b : \{0, 1\}^b \times \{0, 1\}^b \rightarrow \{0, 1\} \) be the two-party (Alice has \( x \), Bob has \( y \)) inner-product function given by

\[
\text{IP}_b(x, y) := \langle x, y \rangle \mod 2.
\]

Let \( \text{BPP}_\epsilon^c(F) \) denote the randomised \( \epsilon \)-error communication complexity of \( F : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\} \). The following is a corollary of [13, Theorem 31] (the original formulation there talks about \( \text{WAPP}_{\epsilon, b}^c(f) \) which is the same as \( \text{deg}_c(f) \); moreover, the result is stated for \( \epsilon = \Theta(1) \), but the theorem is true more generally for \( \epsilon = 2^{-\Omega(b)} \).

**Theorem 7.1** ([13]). Let \( \epsilon = 1/n \) and \( b := \Theta(\log n) \) (with a large enough implicit constant).

For any \( f : \{0, 1\}^n \rightarrow \{0, 1\} \) we have

\[
\text{BPP}_\epsilon^c(f \circ \text{IP}_b^b) \geq \Omega(\text{deg}_c(f) \cdot b).
\]

Let us prove Theorem 2.3 (a similar argument works for Theorem 2.4). A key observation (also made in [16, §3]) is that \( \text{IP}_b = \text{XOR}_0 \circ \text{AND}^b \) reduces to computing a binary \( \text{NAND} \) tree on \( O(b^2) \) bits. To see this, think of the \( b \)-bit parity function \( \text{XOR}_b \) as a height-\((\log b) \) binary tree of \( \text{XOR} \) gates. Each such \( \text{XOR} \) gate can be rewritten as a height-2 \( \text{NAND} \) tree (with some negations on inputs):

\[
\begin{align*}
\text{XOR} & \sim \text{NAND} \\
\overline{x} & \overline{y} \quad \overline{x} \quad \overline{y}
\end{align*}
\]

In the binary \( \text{XOR} \) tree, replace the top \( \text{XOR} \) gate with this \( \text{NAND} \) tree (this involves making copies of some subtrees), push the negations to inputs, and repeat recursively. This gives us a height-\((2 \log b) \) \( \text{NAND} \) tree. Moreover, the bottom layer of \( \text{AND} \) gates in \( \text{IP}_b \) is also easily simulated by \( \text{NAND} \) gates. Consequently, for some \( N := \Theta(nb^2) \), the communication matrix of \( \text{NAND}^{\log n} \circ \text{IP}_b^b \) appears as a submatrix of \( \text{NAND}^{\log N} \) (relative to some bipartition of the input given by the reduction).
We can now derive Theorem 2.3 – here \( \epsilon \) and \( b \) are defined as in Theorem 7.1, and \( \gtrsim \) means that we ignore \( \text{polylog}(N) \) factors.

\[
\begin{align*}
\text{BPP}^{cc}_{1/3}(\text{NAND}^{\text{log}N}) & \gtrsim \text{BPP}^{cc}_{\epsilon/2}(\text{NAND}^{\text{log}N}) & \text{(Error reduction)} \\
& \gtrsim \text{BPP}^{cc}_{\epsilon/3}(\text{NAND}^{\text{log}N} \circ \text{IP}_b) & \text{(Key observation)} \\
& \gtrsim \deg_{\epsilon}(\text{NAND}^{\text{log}N}) & \text{(Theorem 7.1)} \\
& \gtrsim n^{0.753...} & \text{(Theorem 2.1)} \\
& = \tilde{\Theta}(N^{0.753...}).
\end{align*}
\]

Acknowledgements. Many thanks to Thomas Watson for a careful perusal of the manuscript, to Li-Yang Tan for correspondence, and to anonymous referees for comments.

References


