A Composition Theorem for Conical Juntas

Mika Göös*[1] and T. S. Jayram[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 $\tilde{\Omega}(n^{0.753\ldots})$ 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 $\Omega(2.59^k)$ 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.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Composition theorems, conical juntas

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.5

1 Conical Juntas?

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

$$\begin{aligned} h_1(x) &= \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}\bar{x}_1 x_2 + \tfrac{1}{2}x_1\bar{x}_2, \\ h_2(y) &= \tfrac{1}{3}y_1y_2 + \tfrac{1}{3}y_2y_3 + \tfrac{1}{3}y_1y_3 + \tfrac{2}{3}\bar{y}_1y_2y_3 + \tfrac{2}{3}y_1\bar{y}_2y_3 + \tfrac{2}{3}y_1y_2\bar{y}_3. \end{aligned} \tag{1}$$

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 $f\colon \{0,1\}^n \to \{0,1\}$. More precisely, we study the ϵ-*approximate* conical junta degree of f, denoted $\deg_\epsilon(f)$, that is defined as the minimum degree of a conical junta h satisfying

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

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.

---

[*] Work done while at IBM Research Almaden.

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].

2 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* $f \circ g^n$ is the function on nm bits that maps an input $x = (x_1, \ldots, x_n) \in (\{0,1\}^m)^n$ to the output

$$(f \circ g^n)(x) \;\coloneqq\; f(g(x_1), \ldots, g(x_n)).$$

Define also $f^{\circ k} \coloneqq f \circ (f^{\circ(k-1)})^n$ where $f^{\circ 1} \coloneqq f$. 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.** $\deg_\epsilon(\mathsf{NAND}^{\circ k}) \geq \Omega(n^{0.753\ldots})$ *for all* $\epsilon \leq 1/n$ *where* $n \coloneqq 2^k$.

▶ **Theorem 2.2.** $\deg_\epsilon(\mathsf{Maj}_3^{\circ k}) \geq \Omega(2.59\ldots^k)$ *for all* $\epsilon \leq 1/n$ *where* $n \coloneqq 3^k$.

Discussion of Theorem 2.1. The function $\mathsf{NAND}^{\circ k}$ is computed by a height-k binary tree consisting of NAND gates (a.k.a. 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 $\mathsf{NAND}^{\circ k}$ 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 $\mathsf{NAND}^{\circ k}$ 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 ϵ-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 $\mathsf{NAND}^{\circ k}$ 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

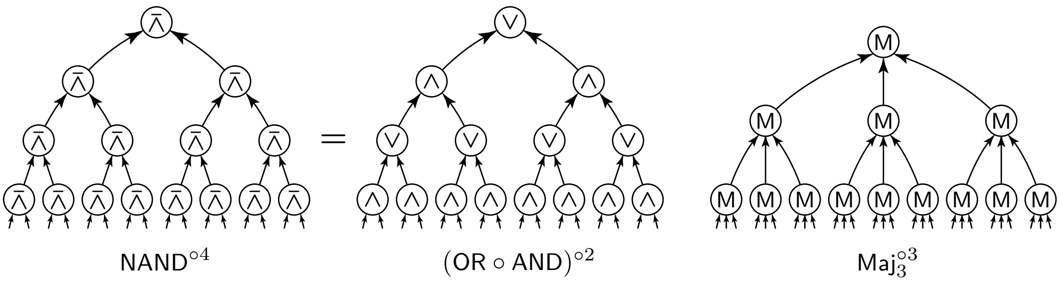


Figure 1 Examples of recursively defined boolean functions studied in this work.

well-known to be w.l.o.g., because the error can be reduced below any $\epsilon < 1/3$ with only a factor $O(\log(1/\epsilon))$ increase in query complexity. Interestingly, for conical juntas, it is known [13] that ϵ 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 $\epsilon = 1/3$: 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 $\tilde{\Omega}(2.59\ldots^k) \geq \Omega(2.59^k)$ (here $2.59\ldots = \sqrt[3]{35/2}$, and the $\tilde{\Omega}$-notation hides $\text{polylog}(n)$ factors) for the randomised query complexity of the recursive majority function $\text{Maj}_3^{\circ k}$. This slightly improves over the previous bound of $\Omega(2.57^k)$ 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 $O(2.65^k)$ 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 $\text{polylog}(n)$ factor loss in parameters, which is suppressed by the $\tilde{\Omega}$-notation used below. Here $\text{BPP}^{\text{cc}}(F)$ 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.** $\text{BPP}^{\text{cc}}(\text{NAND}^{\circ k}) \geq \tilde{\Omega}(n^{0.753\ldots})$.

▶ **Theorem 2.4.** $\text{BPP}^{\text{cc}}(\text{Maj}_3^{\circ k}) \geq \Omega(2.59^k)$.

Discussion of Theorem 2.3. The question of proving a lower bound for the randomised communication complexity of the balanced alternating 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 $\Omega(n/2^{O(k)})$ (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 $\tilde{\Omega}(n^{0.753\ldots})$ bound for the particular case of $\mathsf{NAND}^{\circ k}$. 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 $\mathsf{Maj}_3^{\circ k}$ 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 $C\colon \{0,1\}^n \to \{0,1\}$ and $w_C \geq 0$ for each C. Note that $h\colon \{0,1\}^n \to \mathbb{R}_{\geq 0}$. 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 $\operatorname{adeg}(h)$, is defined as the maximum over all inputs x of

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

In particular, $\operatorname{adeg}(h) \leq \deg(h)$ 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 $f\colon \{0,1\}^n \to \{0,1\}$ we define

- *Degree:* $\deg(f)$ is the minimum $\deg(h)$ over all conical juntas h computing f.
- *Average degree:* $\operatorname{adeg}(f)$ is the minimum $\operatorname{adeg}(h)$ over all conical juntas h computing f.
- *Approximate degree:* $\deg_\epsilon(f)$ is the minimum $\deg(h)$ over all conical juntas h that compute f to within error ϵ, 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 $\operatorname{adeg}(h_1) = \operatorname{adeg}_{10}(h_1) = 3/2 < 2 = \deg(h_1)$ and $\operatorname{adeg}(h_2) = \operatorname{adeg}_{110}(h_2) = 8/3 < 3 = \deg(h_2)$. In fact, h_1 and h_2 are optimal:

$$\operatorname{adeg}(\mathsf{OR}) \;=\; 3/2 \qquad \text{and} \qquad \operatorname{adeg}(\mathsf{Maj}_3) \;=\; 8/3.$$

This can be seen by solving an LP whose value is $\operatorname{adeg}(f)$, 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 $\operatorname{adeg}(\neg\mathsf{OR}) = 2$ (observe that $\bar{x}_1\bar{x}_2$ is the only conical junta for $\neg\mathsf{OR}$) even though $\operatorname{adeg}(\mathsf{OR}) = 3/2$. (More dramatic gaps can be demonstrated using variations of a function introduced in [14].) By contrast, Maj_3 is *self-dual*, $\neg\mathsf{Maj}_3(x_1, x_2, x_3) = \mathsf{Maj}_3(\neg x_1, \neg x_2, \neg x_3)$, so we automatically have $\operatorname{adeg}(\mathsf{Maj}_3) = \operatorname{adeg}(\neg\mathsf{Maj}_3)$.

3.2 A wild example!

What is the average degree of $\mathsf{OR} \circ \mathsf{Maj}_3^2$? We can obtain a conical junta for this function starting with the optimal conical juntas $h_1(x)$, $h_2(y)$, $\bar{h}_2(y) \coloneqq h_2(\bar{y}_1, \bar{y}_2, \bar{y}_3)$ computing OR, Maj_3, $\neg\mathsf{Maj}_3$, respectively, as follows: Let $z^1 = (z^1_1, z^1_2, z^1_3)$ and $z^2 = (z^2_1, z^2_2, z^2_3)$ be fresh variables. Start with $h_1(x)$ and replace every positive literal x_i by $h_2(z^i)$ and every negative literal $\bar{x}_i$ by $\bar{h}_2(z^i)$. This construction shows that

$$\operatorname{adeg}(\mathsf{OR} \circ \mathsf{Maj}_3^2) \;\leq\; 3/2 \cdot 8/3 \;=\; 4.$$

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 $\operatorname{adeg}(f \circ g^n)$ solely in terms of $\operatorname{adeg}(f)$, $\operatorname{adeg}(g)$, and $\operatorname{adeg}(\neg g)$, even assuming $\operatorname{adeg}(g) = \operatorname{adeg}(\neg g)$. Consequently, for our LP-based Composition Theorem, we will have to introduce *some* technical assumptions: to enable the construction of a dual certificate for $\operatorname{adeg}(f \circ g^n)$, we assume we have dual certificates *of a special form* for $\operatorname{adeg}(f)$, $\operatorname{adeg}(g)$, $\operatorname{adeg}(\neg g)$. The rest of this section develops our LP formalism for average degree.

3.3 Generalised input costs

Let us first generalise the definition of $\operatorname{adeg}(h)$ 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 $|C|_{b_0,b_1} \coloneqq b_0 \cdot (\#\text{ of 0's read by } C) + b_1 \cdot (\#\text{ of 1's read by } C)$. In particular, $|C|_{1,1} = |C|$. Then $\operatorname{adeg}(h; b_0, b_1)$ is defined as the maximum over all inputs x of

$$\operatorname{adeg}_x(h; b_0, b_1) \;\coloneqq\; \sum w_C |C|_{b_0,b_1} C(x) \;=\; \sum w_C \operatorname{adeg}_x(C; b_0, b_1).$$

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

$$\operatorname{adeg}_{D_1}(h; b_0, b_1) \;\coloneqq\; \mathop{\mathbf{E}}_{x\sim D_1}\big[\operatorname{adeg}_x(h; b_0, b_1)\big].$$

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

- $\operatorname{adeg}(f; b_0, b_1)$ is the minimum of $\operatorname{adeg}(h; b_0, b_1)$ over all conical juntas h computing f.
- $\operatorname{adeg}_{D_1}(f; b_0, b_1)$ is the minimum of $\operatorname{adeg}_{D_1}(h; b_0, b_1)$ over all conical juntas h computing f.

It is clear that $\operatorname{adeg}(f; b_0, b_1) \geq \operatorname{adeg}_{D_1}(f; b_0, b_1)$ 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 $\mathrm{adeg}_{D_1}(f; b_0, b_1)$ 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 $C\colon \{0,1\}^n \to \{0,1\}$. Here is the LP:

$$\begin{array}{rlll} \min & \mathrm{adeg}_{D_1}\left(\sum w_C C; b_0, b_1\right) & & \\ \textit{subject to} & \sum w_C C(x) = f(x), & \forall x & \text{(Primal)} \\ & w_C \geq 0, & \forall C & \end{array}$$

Here is the LP dual; the free variables are packaged into a function $\Psi\colon \{0,1\}^n \to \mathbb{R}$.

$$\begin{array}{rlll} \max & \langle \Psi, f \rangle & & \\ \textit{subject to} & \langle \Psi, C \rangle \leq \mathrm{adeg}_{D_1}(C; b_0, b_1), & \forall C & \text{(Dual)} \\ & \Psi(x) \in \mathbb{R}, & \forall x & \end{array}$$

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 Ψ that is feasible for (Dual) witnesses a lower bound on $\mathrm{adeg}(f; b_0, b_1)$ like so:

$$\begin{aligned} \mathrm{adeg}(f; b_0, b_1) &\geq \mathrm{adeg}_{D_1}(f; b_0, b_1) \\ &= \mathrm{adeg}_{D_1}(h; b_0, b_1) \\ &= \textstyle\sum w_C \, \mathrm{adeg}_{D_1}(C; b_0, b_1) \\ &\geq \textstyle\sum w_C \langle \Psi, C \rangle \\ &= \langle \Psi, \textstyle\sum w_C C \rangle \\ &= \langle \Psi, f \rangle. \end{aligned} \tag{2}$$

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

$$\begin{aligned} \mathrm{adeg}(f; b_0, b_1) &\geq a_1, \\ \mathrm{adeg}(\neg f; b_0, b_1) &\geq a_0. \end{aligned} \tag{3}$$

▶ **Definition 4.1.** Call a function $\Psi\colon \{0,1\}^n \to \mathbb{R}$ *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 $f\colon \{0,1\}^n \to \{0,1\}$ consists of four balanced functions $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ mapping $\{0,1\}^n \to \mathbb{R}$ such that the following hold.

- **Special form:** Functions Ψ_0 and Ψ_1 have the form

$$\Psi_v = a_v(D_v - D_{1-v}), \tag{4}$$

where D_v is a distribution over $f^{-1}(v)$. Moreover, $\hat{\Psi}_v$ is supported on $f^{-1}(v)$.

- **Feasibility:** For all conjunctions C and $v \in \{0,1\}$,

$$\langle \Psi_v, C\rangle_{\geq 0} + \langle \hat{\Psi}_v, C\rangle \;\leq\; \operatorname{adeg}_{D_v}(C; b_0, b_1). \tag{5}$$

▶ **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.*

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

$$\begin{cases} \langle \Psi_v + \hat{\Psi}_v, C\rangle \;\leq\; \operatorname{adeg}_{D_v}(C; b_0, b_1), \\ \qquad\;\; \langle \hat{\Psi}_v, C\rangle \;\leq\; \operatorname{adeg}_{D_v}(C; b_0, b_1). \end{cases} \tag{5'}$$

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

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 $\operatorname{adeg}(\mathsf{Maj}_3) = 8/3$, while the best *balanced* solution to (Dual) only witnesses the lower bound $\operatorname{adeg}(\mathsf{Maj}_3) \geq 5/2$ (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 Ψ_1 and Ψ_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 $\mathsf{Maj}_3^{\circ k}$ 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 $\{\Upsilon_v, \hat{\Upsilon}_v\}_{v=0,1}$ for $f \circ g^n$. We use the following notation:

$$\underbrace{\Psi_v \;:=\; a_v(F_v - F_{1-v}),}_{\text{given}} \qquad \underbrace{\Phi_v \;:=\; b_v(G_v - G_{1-v}),}_{\text{given}} \qquad \underbrace{\Upsilon_v \;:=\; a_v(D_v - D_{1-v})}_{\text{want to construct}}.$$

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

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

$$\Gamma^H \;:=\; \textstyle\sum_{y \in \{0,1\}^n} \Gamma(y) \cdot H_y.$$

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

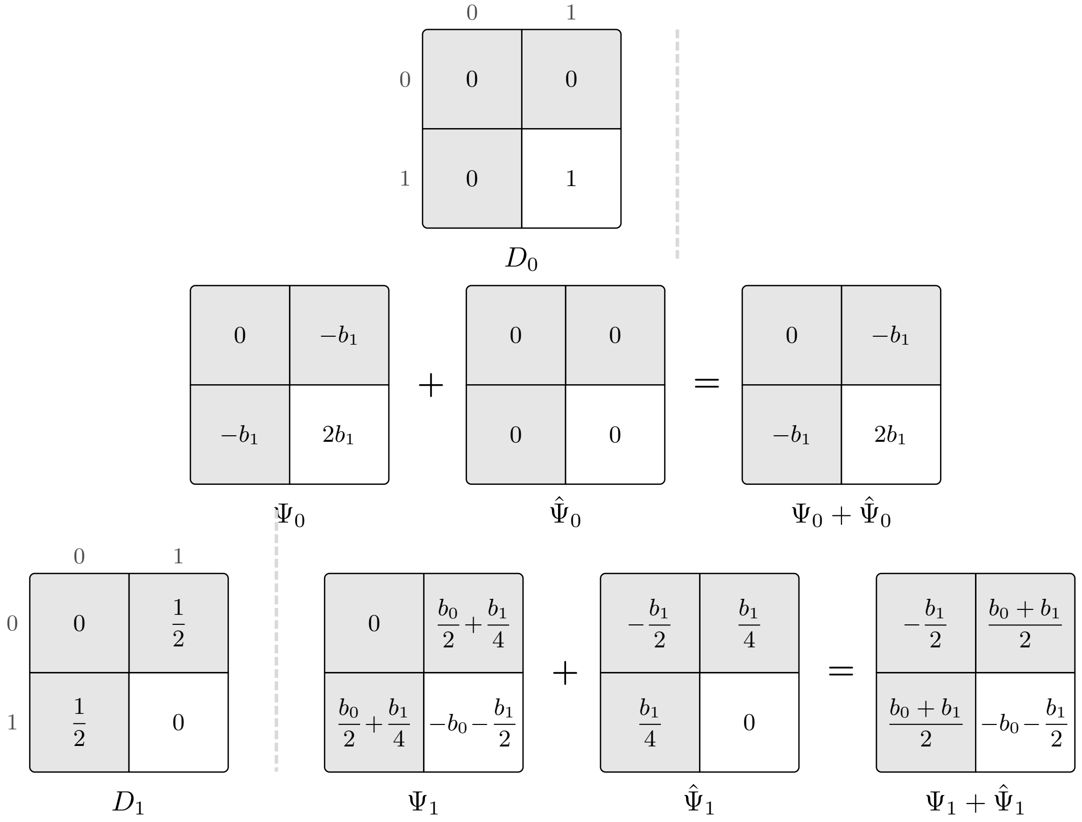


Figure 2 A $(2b_1, b_0+\frac{1}{2}b_1; b_0, b_1)$-certificate for $\mathsf{NAND}\colon \{0,1\}^2 \to \{0,1\}$ that is valid for all $b_0, b_1 \geq 0$. The 1-inputs $\mathsf{NAND}^{-1}(1)$ are highlighted in gray. For feasibility, there are 6 equivalence classes (see Section 6.2) of conjunctions to check: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

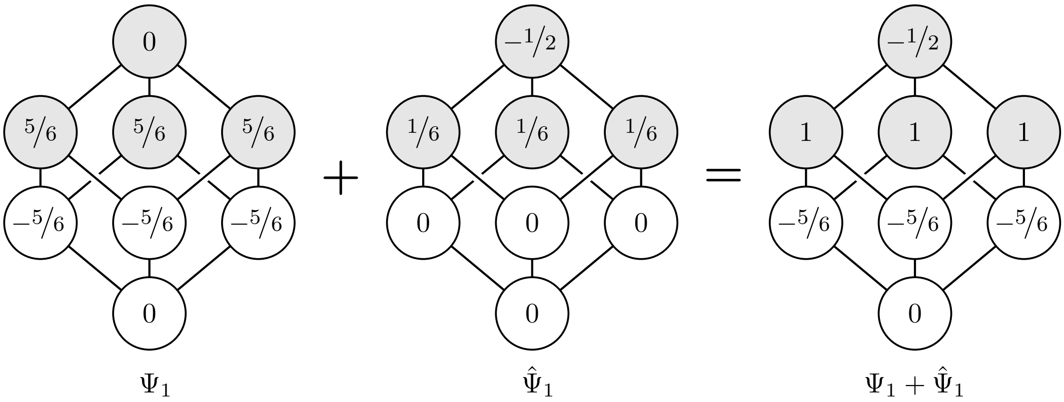


Figure 3 A $(\frac{5}{2}, \frac{5}{2}; 1, 1)$-certificate for $\mathsf{Maj}_3\colon \{0,1\}^3 \to \{0,1\}$. The 1-inputs $\mathsf{Maj}_3^{-1}(1)$ are highlighted in gray. Only $\Psi_1, \hat\Psi_1$ are shown as $\Psi_0, \hat\Psi_0$ are defined via self-duality. Here D_v is uniform on inputs of Hamming weight $v+1$. For feasibility, there are 10 equivalence classes of conjunctions to check: $\{{*}{*}{*}, {*}{*}1, {*}{*}0, {*}00, {*}10, {*}11, 000, 100, 110, 111\}$. 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 α.

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 $(G \overset{i}{\leftarrowtail} \hat{\Phi})_y$ where $i \in [n]$, that has a copy of $\hat{\Phi}_{y_i}$ in place of G_{y_i}; more formally

$$(G \overset{i}{\leftarrowtail} \hat{\Phi})_y(x) \;:=\; \hat{\Phi}_{y_i}(x_i) \cdot \textstyle\prod_{j \neq i} G_{y_j}(x_j).$$

Note that $(G \overset{i}{\leftarrowtail} \hat{\Phi})_y$ is a balanced function supported on $(g^n)^{-1}(y)$.

We now define $\{\Upsilon_v, \hat{\Upsilon}_v\}_{v=0,1}$ by

$$\begin{aligned} \Upsilon_v \;&:=\; \Psi_v^G, \\ \hat{\Upsilon}_v \;&:=\; \hat{\Psi}_v^G + \textstyle\sum_{i=1}^n F_v^{G \overset{i}{\leftarrowtail} \hat{\Phi}}. \end{aligned} \tag{6}$$

Since $\Psi_v^G = a_v(F_v^G - F_{1-v}^G)$, we have $D_v = F_v^G$. It is also easy to check that $\hat{\Upsilon}_v$ is a balanced function supported on $(f \circ g^n)^{-1}(v)$. Hence $\{\Upsilon_v, \hat{\Upsilon}_v\}_{v=0,1}$ 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

$$\langle \Psi_v^G, C \rangle_{\geq 0} + \langle \hat{\Psi}_v^G + \textstyle\sum_i F_v^{G \overset{i}{\leftarrowtail} \hat{\Phi}}, C \rangle \;\leq\; \operatorname{adeg}_{D_v}(C). \tag{7}$$

Extracting a conical junta from C. Our analysis will be centered around a conical junta $h(y)$, defined below, that computes the acceptance probability $\mathbf{Pr}_{x \sim G_y}[C(x) = 1] = \mathbf{E}_{x \sim G_y}[C(x)] = \langle G_y, C \rangle$. 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,

$$\langle G_y, C \rangle \;=\; \textstyle\prod_i \langle G_{y_i}, C_i \rangle \;=:\; \prod_i p_{i, y_i},$$

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 $\langle G_y, C \rangle$:

$$h(y) \;:=\; \textstyle\prod_{i=1}^n \big(p_{i,1-y_i^*} + \underbrace{(p_{i,y_i^*} - p_{i,1-y_i^*})}_{\geq 0} \cdot \ell_i \big) \quad \text{where literal } \ell_i \text{ is } \begin{cases} y_i & \text{if } y_i^* = 1, \\ \bar{y}_i & \text{if } y_i^* = 0. \end{cases} \tag{8}$$

This product expression can be expanded into a conical combination of conjunctions, $h = \sum w_T T$, 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.** $\operatorname{adeg}_y(h; b_0, b_1) = \sum_i \langle \Phi_{y_i}, C_i \rangle_{\geq 0} \prod_{j \neq i} \langle G_{y_j}, C_j \rangle$.

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

$$\operatorname{adeg}_y(h; b_0, b_1) \;=\; \textstyle\sum_i |\ell_i|_{b_0, b_1} \cdot \sum_{T \ni \ell_i} w_T T(y).$$

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 ℓ_i contributes no weight in this case. Suppose then that i is such that $y_i = y_i^*$; here we can write

$$h(y) \;=\; p_{i,1-y_i} \textstyle\prod_{j \neq i} p_{j,y_j} \;+\; \ell_i \cdot (p_{i,y_i} - p_{i,1-y_i}) \prod_{j \neq i} p_{j,y_j}.$$

The conjunctions T underlying the first term do not involve ℓ_i, so they contribute no weight for ℓ_i. The conjunctions T underlying the second term all involve ℓ_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

$$\begin{aligned}
\operatorname{adeg}_y(h; b_0, b_1) &= \textstyle\sum_i |\ell_i|_{b_0,b_1} \cdot \sum_{T \ni \ell_i} w_T T(y) \\
&= \textstyle\sum_{i : y_i = y_i^*} b_{y_i} \cdot (p_{i,y_i} - p_{i,1-y_i}) \prod_{j \neq i} p_{j,y_j} \\
&= \textstyle\sum_i b_{y_i} (p_{i,y_i} - p_{i,1-y_i})_{\geq 0} \prod_{j \neq i} p_{j,y_j} \\
&= \textstyle\sum_i b_{y_i} (\langle G_{y_i}, C_i \rangle - \langle G_{1-y_i}, C_i \rangle)_{\geq 0} \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \\
&= \textstyle\sum_i \langle b_{y_i} (G_{y_i} - G_{1-y_i}), C_i \rangle_{\geq 0} \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \\
&= \textstyle\sum_i \langle \Phi_{y_i}, C_i \rangle_{\geq 0} \prod_{j \neq i} \langle G_{y_j}, C_j \rangle.
\end{aligned}$$

◀

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

Proof. We calculate

$$\begin{aligned}
\langle \Gamma, h \rangle &= \textstyle\sum_y \Gamma(y) h(y) = \sum_y \Gamma(y) \langle G_y, C \rangle = \sum_y \Gamma(y) \big[\sum_x G_y(x) C(x) \big] \\
&= \textstyle\sum_x \big[\sum_y \Gamma(y) G_y(x) \big] C(x) = \sum_x \Gamma^G(x) C(x) = \langle \Gamma^G, C \rangle.
\end{aligned}$$

◀

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

$$\begin{aligned}
\operatorname{adeg}_{D_v}(C) &= |C| \cdot \langle F_v^G, C \rangle \\
&= \mathbf{E}_{y \sim F_v} \big[|C| \cdot \langle G_y, C \rangle \big] \\
&= \mathbf{E}_{y \sim F_v} \big[\big(\textstyle\sum_i |C_i| \big) \cdot \prod_i \langle G_{y_i}, C_i \rangle \big] \\
&= \mathbf{E}_{y \sim F_v} \big[\textstyle\sum_i |C_i| \langle G_{y_i}, C_i \rangle \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \big] \\
&= \mathbf{E}_{y \sim F_v} \big[\textstyle\sum_i \operatorname{adeg}_{G_{y_i}}(C_i) \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \big].
\end{aligned}$$

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

$$\operatorname{adeg}_{D_v}(C) \geq \underbrace{\mathop{\mathbf{E}}_{y \sim F_v} \big[\sum_i \langle \Phi_{y_i}, C_i \rangle_{\geq 0} \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \big]}_{\textbf{(I)}} + \underbrace{\mathop{\mathbf{E}}_{y \sim F_v} \big[\sum_i \langle \hat{\Phi}_{y_i}, C_i \rangle \prod_{j \neq i} \langle G_{y_j}, C_j \rangle \big]}_{\textbf{(II)}}.$$

For the first term,

$$\begin{aligned}
\textbf{(I)} &= \mathbf{E}_{y \sim F_v} \big[\operatorname{adeg}_y(h; b_0, b_1) \big] && \text{(Lemma 5.1)} \\
&= \operatorname{adeg}_{F_v}(h; b_0, b_1) \\
&\geq \langle \Psi_v, h \rangle_{\geq 0} + \langle \hat{\Psi}_v, h \rangle && \text{(Feasibility of } \{\Psi_v, \hat{\Psi}_v\} \text{ and (2))} \\
&= \langle \Psi_v^G, C \rangle_{\geq 0} + \langle \hat{\Psi}_v^G, C \rangle. && \text{(Lemma 5.2)}
\end{aligned}$$

For the second term,

$$\begin{aligned}
\textbf{(II)} &= \mathbf{E}_{y \sim F_v} \big[\textstyle\sum_i \langle (G \xleftarrow{i} \hat{\Phi})_y, C \rangle \big] \\
&= \big\langle \textstyle\sum_i F_v^{G \xleftarrow{i} \hat{\Phi}}, C \big\rangle.
\end{aligned}$$

Combining these yields (7). This concludes the proof of Theorem 4.2.

6 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 $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ also yield lower bounds for approximate degree, if the 1-norm $\|\hat{\Psi}_1\|_1$ is not too large. We call $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ an $(a_0, a_1; b_0, b_1; c)$-*certificate* if $\max_v \|\hat{\Psi}_v\|_1 \le c$.

▶ **Lemma 6.1.** *Suppose f admits an $(a_0, a_1; 1, 1; c)$-certificate. If $\epsilon \le 1/4$ and $c \cdot \epsilon \le a_1/4$, then $\deg_\epsilon(f) \ge \Omega(a_1)$.*

Proof. Fix a certificate $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ for f and suppose $\deg_\epsilon(f) = \deg(h)$ where h is a conical junta with $\|h - f\|_\infty \le \epsilon$. Since $h(x) \le 1 + \epsilon$ for all x, we have $\deg(h) \ge (1+\epsilon)^{-1} \operatorname{adeg}(h) \ge \Omega(\operatorname{adeg}(h))$. Now we calculate

$$\begin{aligned}
\operatorname{adeg}(h) &\ge \langle \Psi_1 + \hat{\Psi}_1, h\rangle && \text{(as in (2))}\\
&= \langle \Psi_1 + \hat{\Psi}_1, f\rangle + \langle \Psi_1 + \hat{\Psi}_1, h - f\rangle\\
&\ge a_1 - |\langle \Psi_1 + \hat{\Psi}_1, h - f\rangle|\\
&\ge a_1 - \|\Psi_1 + \hat{\Psi}_1\|_1 \cdot \|h - f\|_\infty\\
&\ge a_1 - (\|\Psi_1\|_1 + \|\hat{\Psi}_1\|_1) \cdot \epsilon\\
&\ge a_1 - (2a_1 + c) \cdot \epsilon\\
&\ge a_1/4.
\end{aligned}$$

◀

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

▶ **Theorem 6.2.** *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 $f \circ g^n$ admits an $(a_0, a_1; 1, 1; c + nd)$-certificate.*

6.1 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 $\mathsf{NAND}^{\circ k}$ where $\gamma_k \ge \alpha_k, \beta_k$.
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 $\mathsf{NAND}^{\circ(k+1)}$ where

$$\begin{aligned}
\alpha_{k+1} &\coloneqq 2\beta_k,\\
\beta_{k+1} &\coloneqq \alpha_k + \beta_k/2,\\
\gamma_{k+1} &\coloneqq \beta_k + 2\gamma_k.
\end{aligned}$$

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

6.2 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 $\Omega(2.5^k)$ lower bound for $\mathsf{Maj}_3^{\circ k}$. This is the best possible for our approach if we were to just compose certificates for individual Maj_3 functions. To obtain a better lower

Table 1 Certificates for $\mathsf{Maj}_3^{\circ\ell}$ for heights $\ell = 1, 2, 3$. The table lists $(\alpha_\ell, \alpha_\ell; 1, 1)$-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 Maj_3 gates by their Hamming weight, e.g., $001 \rightsquigarrow 1$, $011 \rightsquigarrow 2$, etc.

Function	Class representative	Class size	Ψ_1	$\hat{\Psi}_1$	$\Psi_1 + \hat{\Psi}_1$
$\mathsf{Maj}_3^{\circ 1}$	$(0,0,1)$	3	$-5/2$	0	$-5/2$
	$(0,1,1)$	3	$5/2$	$1/2$	3
	$(1,1,1)$	1	0	$-1/2$	$-1/2$
	All others		0	0	0
$\mathsf{Maj}_3^{\circ 2}$	$(001, 001, 011)$	81	$-20/3$	0	$-20/3$
	$(001, 011, 011)$	81	$20/3$	$7/3$	9
	$(000, 011, 011)$	27	0	$-1/3$	$-1/3$
	$(001, 011, 111)$	54	0	$-2/3$	$-2/3$
	$(011, 011, 011)$	27	0	$-4/3$	$-4/3$
	All others		0	0	0
$\mathsf{Maj}_3^{\circ 3}$	$(112, 112, 122)$	1594323	$-35/2$	0	-35/2
	$(112, 122, 122)$	1594323	$35/2$	$19/2$	27
	$(122, 122, 122)$	531441	0	$-7/2$	$-7/2$
	$(112, 122, 222)$	1062882	0	-2	-2
	$(112, 122, 123)$	2125764	0	$-4/3$	$-4/3$
	$(112, 122, 022)$	1062882	0	$-2/3$	$-2/3$
	$(111, 122, 122)$	531441	0	$-5/6$	$-5/6$
	$(113, 122, 122)$	531441	0	$-1/2$	$-1/2$
	$(012, 122, 122)$	1062882	0	$-2/3$	$-2/3$
	All others		0	0	0

bound, we can instead directly find a certificate for $\mathsf{Maj}_3^{\circ\ell}$ where ℓ is a small constant, and then compose that certificate. Table 1 gives certificates for $\mathsf{Maj}_3^{\circ\ell}$ for height up to $\ell = 3$. We used a computer to solve the dual LP (Dual), with the additional restriction that Ψ $(= \Psi_1 + \hat{\Psi}_1)$ should be balanced. The best balanced Ψ 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 $\mathsf{Maj}_3^{\circ\ell}$ (permutations of input coordinates that do not change the value of the function) are the symmetries of the underlying height-ℓ 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 Ψ that is invariant, i.e., uniform on each equivalence class. (Indeed, if Ψ is any feasible solution, we obtain an invariant solution by averaging Ψ over all the symmetries.) Thus we need only maintain one variable in the LP per equivalence class $\mathcal{X} \subseteq \{0,1\}^n$ recording the *total weight* $\sum_{x \in \mathcal{X}} \Psi(x)$ of that class. Also, for such invariant Ψ, we need only check the LP feasibility constraint $\langle \Psi, C \rangle \leq \operatorname{adeg}_{D_1}(C; b_0, b_1)$ 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 Ψ 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 $\mathsf{Maj}_3^{\circ 3}$ 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 $((35/2)\alpha, (35/2)\alpha; \alpha, \alpha; 19\alpha)$. Using Theorem 6.2 iteratively as in Section 6.1, we get a certificate for $\mathsf{Maj}_3^{\circ k}$ 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 ϵ-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 $\mathsf{IP}_b\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be the two-party (Alice has x, Bob has y) inner-product function given by

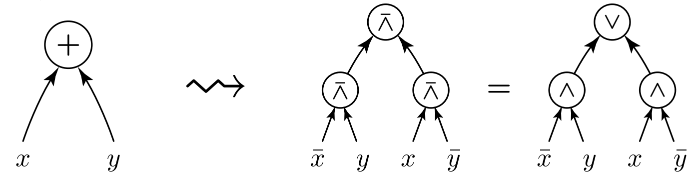
$$\mathsf{IP}_b(x, y) \;:=\; \langle x, y\rangle \bmod 2.$$

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

▶ **Theorem 7.1** ([13]). *Let* $\epsilon := 1/n$ *and* $b := \Theta(\log n)$ *(with a large enough implicit constant). For any* $f\colon \{0,1\}^n \to \{0,1\}$ *we have*

$$\mathsf{BPP}^{\mathsf{cc}}_{\epsilon/2}(f \circ \mathsf{IP}_b^n) \;\geq\; \Omega(\deg_{\epsilon}(f) \cdot b).$$

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



In the binary XOR tree, replace the top XOR gate with this NAND tree (this involves making copies of some subtrees), push the negations to inputs, and repeat recursively. This gives us a height-$(2\log b)$ NAND tree. Moreover, the bottom layer of AND gates in IP_b is also easily simulated by NAND gates. Consequently, for some $N := \Theta(nb^2)$, the communication matrix of $\mathsf{NAND}^{\circ \log n} \circ \mathsf{IP}_b^n$ appears as a submatrix of $\mathsf{NAND}^{\circ \log N}$ (relative to some bipartition of the input given by the reduction).

We can now derive Theorem 2.3 – here ϵ and b are defined as in Theorem 7.1, and $\gtrsim$ means that we ignore polylog(N) factors.

$$\begin{aligned}
\mathsf{BPP}^{\mathsf{cc}}_{1/3}(\mathsf{NAND}^{\circ \log N}) &\gtrsim \mathsf{BPP}^{\mathsf{cc}}_{\epsilon/2}(\mathsf{NAND}^{\circ \log N}) && \text{(Error reduction)}\\
&\gtrsim \mathsf{BPP}^{\mathsf{cc}}_{\epsilon/2}(\mathsf{NAND}^{\circ \log n} \circ \mathsf{IP}^n_b) && \text{(Key observation)}\\
&\gtrsim \deg_\epsilon(\mathsf{NAND}^{\circ \log n}) && \text{(Theorem 7.1)}\\
&\gtrsim n^{0.753\ldots} && \text{(Theorem 2.1)}\\
&= \tilde{\Theta}(N^{0.753\ldots}).
\end{aligned}$$

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

1 Kazuyuki Amano. Bounding the randomized decision tree complexity of read-once boolean functions. In *Proceedings of the 22nd Symposium on Discrete Algorithms (SODA)*, pages 1729–1744. SIAM, 2011.

2 Kazuyuki Amano. On directional vs. general randomized decision tree complexity for read-once formulas. *Chicago Journal of Theoretical Computer Science*, 2011(3), 2011. `doi:10.4086/cjtcs.2011.003`.

3 Kazuyuki Amano. Researching the complexity of boolean functions with computers. *Bulletin of EATCS*, 101:64–91, 2014. URL: `http://bulletin.eatcs.org/index.php/beatcs/article/view/181`.

4 Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. Technical Report TR15-098, Electronic Colloquium on Computational Complexity (ECCC), 2015. URL: `http://eccc.hpi-web.de/report/2015/098/`.

5 Andris Ambainis, Andrew Childs, Ben Reichardt, Robert Špalek, and Shengyu Zhang. Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. *SIAM Journal on Computing*, 39(6):2513–2530, 2010. `doi:10.1137/080712167`.

6 Paul Beame and Joan Lawry. Randomized versus nondeterministic communication complexity. In *Proceedings of the 24th Symposium on Theory of Computing (STOC)*, pages 188–199. ACM, 1992. `doi:10.1145/129712.129732`.

7 Aleksandrs Belovs. Non-intersecting complexity. In *Proceedings of the 32nd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, pages 158–165. Springer, 2006. `doi:10.1007/11611257_13`.

8 Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002. `doi:10.1016/S0304-3975(01)00144-X`.

9 Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov–Bernstein inequalities. *Information and Computation*, 243:2–25, 2015. `doi:10.1016/j.ic.2014.12.003`.

10 Siu On Chan, James Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *Proceedings of the 54th Symposium on Foundations of Computer Science (FOCS)*, pages 350–359. IEEE, 2013. `doi:10.1109/FOCS.2013.45`.

11 Ehud Friedgut, Jeff Kahn, and Avi Wigderson. Computing graph properties by randomized subcube partitions. In *Proceedings of the 6th International Workshop on Randomization and Computation (RANDOM)*, pages 105–113. Springer, 2002. `doi:10.1007/3-540-45726-7_9`.

12 Mika Göös. Lower bounds for clique vs. independent set. In *Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS)*, pages 1066–1076. IEEE, 2015. `doi:10.1109/FOCS.2015.69`.

13 Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the 47th Symposium on Theory of Computing (STOC)*, pages 257–266. ACM, 2015. (Full version: `http://eccc.hpi-web.de/report/2014/147/`). `doi:10.1145/2746539.2746596`.

14 Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS)*, pages 1077–1088. IEEE, 2015. `doi:10.1109/FOCS.2015.70`.

15 Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th Conference on Computational Complexity (CCC)*, pages 247–258. IEEE, 2010. `doi:10.1109/CCC.2010.31`.

16 Rahul Jain, Hartmut Klauck, and Shengyu Zhang. Depth-independent lower bounds on the communication complexity of read-once boolean formulas. In *Proceedings of the 16th International Computing and Combinatorics Conference (COCOON)*, pages 54–59. Springer, 2010. `doi:10.1007/978-3-642-14031-0_8`.

17 Rahul Jain, Troy Lee, and Nisheeth Vishnoi. A quadratically tight partition bound for classical communication complexity and query complexity. Technical report, arXiv, 2014. `arXiv:1401.4512`.

18 T.S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once AC^0 formulae. In *Proceedings of the 24th Conference on Computational Complexity (CCC)*, pages 329–340, 2009. `doi:10.1109/CCC.2009.39`.

19 T.S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th Symposium on Theory of Computing (STOC)*, pages 673–682. ACM, 2003. `doi:10.1145/780542.780640`.

20 Jedrzej Kaniewski, Troy Lee, and Ronald de Wolf. Query complexity in expectation. In *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 761–772. Springer, 2015. `doi:10.1007/978-3-662-47672-7_62`.

21 Robin Kothari, David Racicot-Desloges, and Miklos Santha. Separating decision tree complexity from subcube partition complexity. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, Leibniz International Proceedings in Informatics (LIPIcs), pages 915–930. Schloss Dagstuhl, 2015. `doi:10.4230/LIPIcs.APPROX-RANDOM.2015.915`.

22 Itamar Landau, Asaf Nachmias, Yuval Peres, and Sithparran Vanniasegaram. The lower bound for evaluating a recursive ternary majority function: an entropy-free proof. Technical report, U.C. Berkeley, 2006.

23 Joan Lawry. *Communication complexity: Iterative techniques for lower bounds.* PhD thesis, University of Washington, 1993. UW-CSE-93-3-04.

24 James Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Symposium on Theory of Computing (STOC)*, pages 567–576. ACM, 2015. `doi:10.1145/2746539.2746599`.

25 Nikos Leonardos. An improved lower bound for the randomized decision tree complexity of recursive majority. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 696–708. Springer, 2013. `doi:10.1007/978-3-642-39206-1_59`.

26 Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity*, 19(2):153–181, 2010. `doi:10.1007/s00037-010-0292-2`.

27 Frédéric Magniez, Ashwin Nayak, Miklos Santha, Jonah Sherman, Gábor Tardos, and David Xiao. Improved bounds for the randomized decision tree complexity of recursive majority. *Random Structures & Algorithms*, 2015. In press. `doi:10.1002/rsa.20598`.

28 Michael Saks and Avi Wigderson. Probabilistic boolean decision trees and the complexity of evaluating game trees. In *Proceedings of the 27th Symposium on Foundations of Computer Science (FOCS)*, pages 29–38. IEEE, 1986. `doi:10.1109/SFCS.1986.44`.

29 Miklos Santha. On the Monte Carlo boolean decision tree complexity of read-once formulae. *Random Structures & Algorithms*, 6(1):75–87, 1995. `doi:10.1002/rsa.3240060108`.

30 Petr Savický. On determinism versus unambiguous nondeterminism for decision trees. Technical Report TR02-009, Electronic Colloquium on Computational Complexity (ECCC), 2002. URL: `http://eccc.hpi-web.de/report/2002/009/`.

31 Alexander Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013. `doi:10.4086/toc.2013.v009a020`.

32 Alexander Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *Proceedings of the 46th Symposium on Theory of Computing (STOC)*, pages 223–232. ACM, 2014. `doi:10.1145/2591796.2591871`.

33 Marc Snir. Lower bounds on probabilistic linear decision trees. *Theoretical Computer Science*, 38:69–82, 1985. `doi:10.1016/0304-3975(85)90210-5`.

34 Nikolai Vereshchagin. Randomized boolean decision trees: Several remarks. *Theoretical Computer Science*, 207(2):329–342, 1998. `doi:10.1016/S0304-3975(98)00071-1`.