On Constant-Depth Canonical Boolean Circuits for Computing Multilinear Functions

Oded Goldreich and Avishay Tal

Abstract. We consider new complexity measures for the model of multilinear circuits with general multilinear gates introduced by Goldreich and Wigderson (*ECCC*, 2013). These complexity measures are related to the size of *canonical constant-depth Boolean circuits*, which extend the definition of canonical *depth-three* Boolean circuits. We obtain matching lower and upper bound on the size of canonical constant-depth Boolean circuits for almost all multilinear functions, and non-trivial lower bounds on the size of such circuits for some explicit multilinear functions.

An early version of this work appeared as TR17-193 of *ECCC*. The presentation was ellaborated in the current revision, and a summary of our complexity bounds was added (see Section 8). Furthermore, Corollary 6.6 is new.

1 Introduction

Goldreich and Wigderson [3] put forward a model of *depth-three canonical circuits*, with the underlying long-term goal of leading to better lower bounds for general depth-three Boolean circuits computing explicit *multi-linear* functions. Canonical circuits are restricted type of Boolean depth-three circuits, and their study is supposed to be a warm-up and/or a sanity check for the establishing of lower bound on the size of general depth-three Boolean circuits that compute explicit multi-linear functions.

The canonical circuits defined in [3] are *depth-three* Boolean circuits that are obtained by a two-stage process: First, one constructs arithmetic circuits that use arbitrary multilinear gates of parameterized arity (and number of gates), and next one converts these arithmetic circuits to Boolean circuits. As shown in [3], the size of the resulting depth-three Boolean circuits is exponential in the maximum between the arity and the number of gates in the arithmetic circuit.

Hence, a natural complexity measure of such arithmetic circuits arises; specifically, the AN-complexity of a multi-linear function is m if it can be computed by a multilinear circuit (see Section 2) having at most m gates of arity at most m, where 'A' stands for 'arity' and 'N' for 'number' (of gates). The immediate challenge posed by [3] is to present explicit t-linear functions on $t \cdot n$ variables that require AN-complexity significantly greater than $(tn)^{1/2}$. Note that a lower bound of $m = \omega(\sqrt{tn})$ on the AN-complexity of such a function f yields a lower bound of $\exp(m)$ on the size of depth-three canonical circuits computing f, whereas the best bound known on the size of a general depth-three Boolean circuit computing an explicit function over $\{0,1\}^n$ is $\exp(\sqrt{n})$. Hence, in the context of the AN-complexity measures of [3], a lower bound of $\omega(\sqrt{tn})$ is considered nontrivial.

In this context, a first nontrivial lower bound on an explicit function was obtained by Goldreich and Tal [4]. They exhibit explicit three-linear and four-linear functions having AN-complexities $\Omega(n^{0.6})$ and $\widetilde{\Omega}(n^{2/3})$, respectfully. Although there is still much to be understood about the foregoing model, which corresponds to depth-three canonical (Boolean) circuits, we dare take another speculative step and put forward a notion of constant-depth canonical (Boolean) circuits along with a corresponding model of arithmetic circuits. In particular:

- We define more permissive "AN-complexity" measures (for multilinear circuits) than those defined in [3] and show a partial correspondence between them and a notion of constant-depth canonical circuit.
  The more permissive "AN-complexity" are aimed to accommodate natural constructions of constant-depth (rather than depth-three) Boolean circuits for computing multi-linear functions.
- Extending the results of [3], we obtain matching lower and upper bound on the complexity of almost all multi-linear functions. Specifically, for most t-linear functions, the size of canonical circuits of depth d is $\exp(\Theta(tn)^{t/(t+d-2)})$.[1] (Indeed, the results of [3] refer to $d=3$, and assert size $\exp(\Theta(tn)^{t/(t+1)})$.)
- Extending the results of [3] and using the results of [4], we obtain a lower bound on the size of depth-four canonical circuits that compute an explicit trilinear function. The resulting lower bound of $\exp(\widetilde{\Omega}(n^{3/8}))$ should be compared to $\exp(\Omega(n^{1/3}))$, which is the best lower bound known on the size of a general depth-four Boolean circuit computing an explicit function over $\{0,1\}^n$.

The foregoing description is very vague: It does not say what are the "more permissive AN-complexity" measures that are suggested in the current work, let alone why they are defined in that way. These obvious gaps will be filled in Sections 2 and 3.

Organization

Our conceptual exposition (i.e., Sections 2 and 3) builds quite heavily on [3]. Familiarity with [3] may be useful also in the other sections. (Note that this volume contains a revised version of [3].) In contrast, the results of [4] are used as a black-box, and so familiarity with that paper is not needed here.

In Section 2 we recall the model of multi-linear circuits with general multi-linear gates, and present two complexity measures that refer to these circuits. These measures refine and generalize the AN-complexity measures introduced in [3], and are motivated by their relation to the size of canonical Boolean circuits

[1] In contrast to the notation used here, in the other sections of this paper, the depth of the canonical circuits is denoted $d+1$, whereas d corresponds to the depth of general multi-linear circuits.

of arbitrary constant-depth (rather than depth three). The latter relation, which actually defines the notion of canonical circuits, is presented in Section 3; that is, canonical circuits are constant-depth Boolean circuits that are derived (from multilinear circuits) by the transformation presented in Section 3.

In Section 4 we present matching lower and upper bounds on the foregoing complexity measures for almost all multilinear functions. These mark the lower bounds we should aim at for explicit functions. While we do not obtain these bounds, we do obtain non-trivial lower bounds in Sections 5–7. Specifically, in Section 5 we present bounds for an explicit trilinear function, and in Section 6 we present larger bounds for an explicit 4-linear function. In Section 7 we show that non-trivial lower bounds for one depth translate to non-trivial lower bounds for larger depths.

In light of the foregoing, a natural place to state our results is right after Section 3 (or, alternatively, right after Section 2). We chose not to do so, but rather provide a summary of our complexity bounds at the conclusion section (Section 8), which may be read right after reading Section 3.

2 Definitions

The basic definitions of multilinear circuits are as in [3,4]. Specifically, we focus on multi-linear functions and on multilinear circuits with general gates that compute them. The arity of these gates will serve as a main complexity measure, but we shall also refer to the depth of the circuit and/or to the number of gates in it. The focus on multilinear circuits with such general gates and the complexity measures associated with them are justified by the construction of canonical constant-depth circuits, which are presented in Section 3.

Multi-linear functions. For fixed $t, n \in \mathbb{N}$, we consider t-linear functions of the form $F : (\{0,1\}^n)^t \to \{0,1\}$, where F is linear in each of the t blocks of variables (which contain n variables each). Such a function F is associated with a t-dimensional array, called a tensor, $T \subseteq [n]^t$ such that

$$F(x^{(1)}, x^{(2)}, ..., x^{(t)}) = \sum_{(i_1,i_2,...,i_t)\in T} x^{(1)}_{i_1} x^{(2)}_{i_2} \cdots x^{(t)}_{i_t} \tag{1}$$

where here $x^{(j)} = (x^{(j)}_1, ..., x^{(j)}_n) \in \{0,1\}^n$ for every $j \in [t]$. That is, F is linear in the variables of each block.

Multi-linear circuits with general gates. We consider multilinear circuits with arbitrary multilinear gates, of bounded arity (where this bound will serve as a complexity measure). The *multilinear requirement* mandates that if two gates have directed paths to them from the same block of inputs, then the results of these two gates are not multiplied together by any other gate. The depth of a circuit is the distance between the input variable and the output gate (e.g., a circuit consisting of a top gate that computes the sum of multilinear gates that are fed by variables only has depth 2).

Complexity measures. The main complexity measures are the *arity* of the general multilinear gates and the *number* of such gates, where we say that a multilinear circuit C has arity m if m equals the maximum arity of a general gate in C. Specifically, we denote by $\mathtt{AN}(F)$ the minimum m such that there exists a multilinear circuit that computes F with at most m gates that are each of arity at most m. This definition as well as its restriction to depth two multilinear circuits, denoted $\mathtt{AN}_2$, are taken from [3]. Specifically, $\mathtt{AN}_2(F) \leq m$ if there exists a *depth-two* multilinear circuit that computes F with at most m gates that are each of arity at most m.

The definitions of $\mathtt{AN}_2$ and $\mathtt{AN}$ are tailored to fit the emulation of the corresponding multilinear circuits by Boolean circuits of size that is exponential in these measures. The emulation of depth-two mulitlinear circuits by Boolean circuits is straightforward: Each gate is emulated by a CNF (or DNF) of size exponential in the gate's arity. This mimics the construction of depth-three Boolean circuits of n-way Parity in which one emulates $\sqrt{n}$-way Parity gates, and the Boolean circuits obtained in this way were called *canonical*.

The same reasoning motivates our generalized complexity measures for multilinear circuits, denoted $\mathtt{A}_d$ and $\mathtt{AN}^{(e)}$. In particular, in Section 3 we shall show how to emulate multilinear circuits by Boolean circuits of constant depth, where the size of the derived "canonical" circuits is related to the new complexity measures. The point is that the aim of deriving depth-three Boolean circuits is replaced by deriving constant-depth Boolean circuits, and this relaxation yields a relaxation of the complexity measures for multilinear circuits.

Definition 2.1 (The A-complexity of depth d multilinear circuits): *For a multilinear function F, we denote by $\mathtt{A}_d(F)$ the minimum arity of a multilinear circuit of depth d that computes F.*

We use the notation $\mathtt{A}_d$ (rather $\mathtt{AN}_d$) in order to stress the fact that the definition makes no reference to the number of gates in the circuit.[2] Still, such an upper bound is implied, because the number of gates in a circuit of depth d and arity m is at most $\sum_{i=0}^{d-1} m^i < (m+1)^{d-1}$, since there are at most m^i gates at distance $i \leq d-1$ from the output gate. (Note that gates in a depth d circuit are at distance at most $d-1$ from the output gate, whereas only variables may be at distance d from the output gate.) Hence, $\mathtt{A}_2(\cdot)$ matches the notion of AN2-complexity as used in [3, 4] (up to a slackness of one unit); that is, $\mathtt{A}_2(F) \leq \mathtt{AN}_2(F) \leq \mathtt{A}_2(F) + 1$.

Definition 2.2 (The AN-complexity of multilinear circuits wrt the exponent e): *For a multilinear function F, we denote by $\mathtt{AN}^{(e)}(F)$ the smallest m such that F can be computed by a circuit of arity at most m that has at most $(m+1)^e$ gates.*

[2] In contrast, the notation $\mathtt{AN}_d$ used in the revised version of [3] that appears in this volume refers to the maximum between the arity and the number of gates in the circuit; that is, $\mathcal{AN}_d(F) \leq m$ if there exists a multilinear circuit C of depth d that computes F such that C has arity at most m and at most m gates.

Definition 2.2 does look weird at first glance, but as hinted above it is justified by the emulation of such circuits by depth $e+2$ Boolean circuits (described in Section 3). At this point we observe that:

- $\mathtt{AN}^{(e)}(F) \leq \mathtt{A}_{e+1}(F)$, since the number of gates in a circuit of depth $e+1$ and arity m is at most $(m+1)^e$.
- $\mathtt{AN}^{(1)}(\cdot)$ matches the notion of AN-complexity as used in [3, 4] (again, up to a slackness of one unit); that is, $\mathtt{AN}^{(1)}(F) \leq \mathtt{AN}(F) \leq \mathtt{AN}^{(1)}(F) + 1$.

We stress that the definition of $\mathtt{AN}^{(e)}(F)$ makes no reference to the depth of multilinear circuits computing F; it only refers to the arity and number of gates in such circuits, while linking the gate count to the arity in a way that fits their relation in a circuit of depth $e+1$ (i.e., guaranteeing that $\mathtt{AN}^{(e)}(F) \leq \mathtt{A}_{e+1}(F)$ holds).

The definitions of $\mathtt{A}_d$ and $\mathtt{AN}^{(e)}$ are tailored to fit the emulation of the corresponding multilinear circuits by Boolean circuits of size that is exponential in these measures. These emulations are described in Section 3, and the circuits obtained by them are called canonical. This fact justifies the definitions of $\mathtt{A}_d$ and $\mathtt{AN}^{(e)}$, which may look weird at first glance.

3 Obtaining Boolean circuits

A direct implementation of the general multilinear gates in a multilinear circuits of depth d yields a Boolean circuit of depth $d+1$ and size $\exp(O(\mathtt{A}_d(\cdot)))$. Specifically, we replace each general gate of arity m by a CNF (resp., a DNF) of size 2^m, where we use CNFs (resp., DNFs) in all even (resp., odd) levels. This allows to combine neighboring levels in the resulting depth $2d$ Boolean circuit, yielding a circuit of depth $d+1$. Hence, we generalize the D-canonical circuits of [3, Cons. 2.6], which constitute the special case of $d=2$.

Proposition 3.1 (D-canonical circuits of depth $d+1$): *For every $d \geq 2$, every multilinear function F can be computed by a Boolean circuit of depth $d+1$ and size $\exp(O(\mathtt{A}_d(F)))$.*

For multilinear circuits of unbounded depth, a less direct emulation (i.e., using "Valiant method" [9]) yields depth-three Boolean circuits of size exponential in $\mathtt{AN}(\cdot)$, called ND-canonical circuits [3, Cons. 2.8]. Recalling that $\mathtt{AN}^{(1)}(F) \approx \mathtt{AN}(F)$, we wish to extend this construction to show that *for every $e \geq 1$, every multilinear function F can be computed by a Boolean circuit of depth $e+2$ and size* $\exp(O(\mathtt{AN}^{(e)}(F)))$, where the aforementioned result refers to the case of $e=1$. We were able to obtain such a result only in the special case that the multilinear circuit is "decompasble" in the following sense.

Specifically, we say that a circuit with N gates is (m,t)-decomposable *if omitting the outgoing edges of at most $t \cdot m$ of its gates yields sub-circuits that are each m-decomposable and has at most $N/(m+1)$ gates.* Note that a circuit with $m+1$ gates is trivially $(m,1)$-decomposable, and this fact underlies [3, Cons. 2.8].

Proposition 3.2 (ND-canonical circuits of depth $e+2$): *Let $e \geq 1$, and suppose that the multilinear function F has an $(m, O(1))$-decomposable multilinear circuit of arity at most m and at most $(m+1)^e$ gates. Then, F can be computed by a Boolean circuit of depth $e+2$ and size* $\exp(O(m))$.

Proof Sketch: The construction proceeds by induction on $e \geq 1$, where the case of $e = 1$ corresponds to [3, Cons. 2.8]. Let G_0 denote the output gate of the given circuit, denoted C.

- For $e > 1$, suppose that C can be decomposed by omitting the outgoing edges of the gates $G_1, ..., G_{O(m)}$ such that each G_i (including G_0) is the output gate of a sub-circuit that is $(m, O(1))$-decompasable and contains at most $(m+1)^{e-1}$ gates. Then, by the induction hypothesis, each of the corresponding sub-circuits can be computed by a Boolean circuit of depth $e+1$ and size $\exp(O(m))$.
- Consider a DNF that verifies the assertion *there exists $\alpha \in \{0,1\}^{O(m)}$ such that the outputs of $(G_0, G_1, ..., G_{O(m)})$ equal* 1α, where these $m+1$ outputs correspond to computations that use the values of the original variables and use α_i as the value that replaces the outcome of G_i that is fed to any other gate. Then, combining this $\exp(O(m))$-sized DNF with the aforementioned circuits of depth $e+1$, we obtain the desired circuit (of depth $e+2$).

(The induction hypothesis is that if a multilinear circuit is (m,t)-decomposable and has arity at most m and at most $(m+1)^{e-1}$ gates, then it can be computed by a Boolean circuit of depth $e+1$ and size $\exp(O((e-1)\cdot t\cdot m))$. The induction step starts with a multilinear circuit that is(m,t)-decomposable and has arity at most m and at most $(m+1)^e$ gates, derives $t\cdot m$ multilinear sub-circuits that are each (m,t)-decomposable with at most $(m+1)^{e-1}$ gates, which yields a Boolean circuit of depth $e+2$ and size $\exp(O((e-1)\cdot tm + tm))$.) ■

Discussion. Proposition 3.2 leaves open the general case in which we are given a multilinear circuit of arity m that has at most $(m+1)^e$ gates (where this circuit is not necessarily $(m, O(1))$-decomposable). Fortunately, the lower bounds (shown in the next sections) hold also for the general case, which means that we lost nothing by being potentially too permissive in defining $\mathtt{AN}^{(e)}(\cdot)$. Still, we wonder what is the "right" notion of the "AN-complexity of multilinear circuits wrt the exponent e". It is not inconceivable that a measure that requires decomposition is right, since it matches the natural application of the "Valiant method" [9].

4 Guiding Bounds

Analogously to [3], we have tight bounds on the complexities of almost all multilinear functions.

Theorem 4.1 (generic upper bound): *For every $d, t \in \mathbb{N}$, every t-linear function F satisfies $\mathtt{A}_d(F) = O(tn)^{t/(t+d-1)}$. In particular, $\mathtt{AN}^{(d-1)}(F) = O(tn)^{t/(t+d-1)}$.*

This generalizes [3, Thm. 3.1], which was stated for $d = 2$.

Proof Sketch: Let $m = t \cdot n^{t/(t+d-1)} \approx (tn)^{t/(t+d-1)}$. Consider a partition of $[n]^t$ into cubes of side-length m/t, and gates that compute the corresponding multilinear functions. We have $(n/(m/t))^t = (tn/m)^t$ such gates, each of arity m. By our setting $(tn/m)^t \approx (m^{\frac{t+d-1}{t}-1})^t = m^{d-1}$, whereas the sum of m^{d-1} values can be computed by a multilinear circuit of depth $d-1$ and arity m. Combining this circuit with the aforementioned m^{d-1} gates, we obtain the desired circuit. ■

Theorem 4.2 (non-explicit lower bound): *For every $d, t \in \mathbb{N}$, almost all t-linear functions F satisfy $\mathtt{AN}^{(d-1)}(F) = \Omega(tn)^{t/(t+d-1)}$. In particular, $\mathtt{A}_d(F) = \Omega(tn)^{t/(t+d-1)}$.*

This generalizes [3, Thm. 4.1], which was stated for $d = 2$.

Proof Sketch: Letting $e = d - 1$, we upper-bound the number of general multilinear circuits of arity m and size $(m+1)^e$. Ignoring the gates' functionalities, we note that the number of relevant DAGs is at most

$$\begin{aligned}\binom{tn + (m+1)^e}{m}^{(m+1)^e} &< (((tn+1)^e)^m)^{(m+1)^e} \\ &= \exp((m+1)^{e+1} \log(tn+1)^e),\end{aligned}$$

where the first expression represents the number of choices of variables and other gates that feed each gate, and the inequality uses $tn + (m+1)^e < (tn+1)^e$. But (for $t \geq 2$ and $m \gg t \log n$) this quantity is dominated by the number of possible gates' functionalities, which is

$$\left(2^{(m/t)^t}\right)^{(m+1)^e} = \exp(m^{t+e}/t^t),$$

since each gate corresponds to a tensor of volume at most $(m/t)^t$. The claim holds since $m^{t+e}/t^t \ll n^t$, provided that $m \ll (tn)^{t/(t+e)}$. ■

5 Lower Bounds on Explicit functions

The current section as well as the next section focus on the cases of $d = 3$ and $e = 2$ (i.e., $\mathtt{A}_3$ and $\mathtt{AN}^{(2)}$), which are the cases immediately above those studied in [3, 4] (which are $d = 2$ and $e = 1$ (equiv., $\mathtt{A}_2 = \mathtt{AN}_2$ and $\mathtt{AN}^{(1)} = \mathtt{AN}$)).

Using the rigidity results of [4], one can obtain non-trivial lower on the $\mathtt{A}_3$ and $\mathtt{AN}^{(2)}$ complexities of explicit trilinear functions, where by non-trivial we mean bounds significantly higher than $\Omega(n^{1/3})$. This relies on connections between the $\mathtt{A}_3$ and $\mathtt{AN}^{(2)}$ complexities of bilinear functions and the rigidity of the corresponding matrices, which adapt ideas of [3, Thm. 4.4]. We recall the relevant definition.

Definition 5.1 (matrix rigidity [8]): *A matrix A* (over a field $\mathcal{F}$) *has rigidity s for rank r if every matrix of rank at most r* (over $\mathcal{F}$) *differs from A on more than s entries.*

We shall consider bilinear functions in the variables $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, and trilinear functions in the variables x, y and $z = (z_1, ..., z_{2n-1})$.

5.1 The case of $\mathtt{A}_3$

Recall that $\mathtt{A}_3(F)$ refers to the arity of depth-three multi-linear circuits; that is, $\mathtt{A}_3(F) \leq m$ if F can be computed by a depth-three multi-linear circuit with gates of arity m. For perspective, recall that [3, Thm. 4.4] implies that *if the corresponding matrix corresponding to a bilinear function F has rigidity m^3 for rank m, then $\mathtt{A}_2(F) > m$.*

Lemma 5.2 (rigidity and $\mathtt{A}_3$): *Let F be a bilinear function and suppose that the corresponding matrix has rigidity m^5 for rank m. Then, $\mathtt{A}_3(F) > m$.*

The proof extends the warm-up of the proof of [3, Thm. 4.4], which referred to the case of $\mathtt{A}_2(F) \leq m$.

Proof: Suppose that $\mathtt{A}_3(F) \leq m$, and consider a depth-three multi-linear circuit C of arity m that computes F. Then, without loss of generality, C has the form

$$C(x, y) = G(L_1(x), ..., L_{m_0}, L'_1(y), ..., L'_{m'_0}(y), Q_1(x, y), ...Q_{m''_0}(x, y)),$$

where G is a quadratic gate, $m_0 + m'_0 + m''_0 \leq m$, the $L_i(x)$'s and $L'_j(y)$'s are linear functions computable by depth-two circuits and the $Q_i(x, y)$'s are bilinear functions that are computed by depth-two circuits. Hence, for some $P \subseteq [m_0] \times [m'_0]$ it holds that

$$C(x, y) = \sum_{(i,j)\in P} L_i(x)L'_j(y) + \sum_{i\in[m''_0]} Q_i(x, y),$$

and each Q_i has the form

$$Q_i(x, y) = \sum_{(j,k)\in P_i} L_{i,j}(x)L'_{i,k}(y) + \sum_{j\in[m''_i]} Q_{i,j}(x, y),$$

where $P_i \subseteq [m_i] \times [m'_i]$ and $m''_i \leq m - (m_i + m'_i)$, and the $L_{i,j}(x)$'s and $L'_{i,k}(y)$'s are linear functions computable by depth-one circuits and the $Q_{i,j}(x, y)$'s are bilinear functions that are computed by depth-one circuits. Hence, the $L_{i,j}(x)$'s and $L'_{i,k}(y)$'s are linear gates and the $Q_{i,j}(x, y)$'s are bilinear gates (each taking m variables). Consider the matrix that corresponds to the function computed by Q_i. It is the sum of $|P_i| \leq m_i \cdot m'_i$ matrices of rank one, each being an outer product of two vectors that each has at most m one-entries, and m''_i matrices each having at most m^2 one-entries. Hence, the matrix that corresponds to $\sum_{i\in[m''_0]} Q_i$ has sparsity at most $\sum_{i\in[m''_0]}(m_i m'_i \cdot m^2 + m''_i \cdot m^2) \leq m^5$, since

$m_0'' \leq m$ and $m_i + m_i' + m_i'' \leq m$. On the other hand, the matrix that corresponds to $\sum_{(i,j)\in P} L_i(x)L_j'(y)$ has rank $\min(m_0, m_0') < m$. It follows that the matrix that corresponds to F does not have rigidity m^5 for rank m. ■

Corollary 5.3 (an $\mathtt{A}_3$ lower bound for random Toeplitz functions): *Almost all bilinear functions F that correspond to Toeplitz matrices satisfy $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{0.4})$.*

Proof: Using Lemma 5.2 it suffices to show that F has rigidity m^5 for rank $m = \widetilde{\Omega}(n^{0.4})$. This follows from special case of [4, Thm. 1.2], which asserts that a random Toeplitz matrix has rigidity $\Omega(n^2/\log n)$ for rank $\sqrt{n}$. ■

Corollary 5.4 (an $\mathtt{A}_3$ lower bound for an explicit trilinear function): *The trilinear function $F(x,y,z) = \sum_{i,j\in[n]} x_i y_j z_{n+i-j}$ satisfies $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{0.4})$.*

Proof: As in [3, 4], this follows from the existence of a bilinear function F' that corresponds to a Toeplitz matrix such that $\mathtt{A}_3(F') = \widetilde{\Omega}(n^{0.4})$, which is asserted in Corollary 5.3. ■

5.2 The case of $\mathtt{AN}^{(2)}$

Recall that $\mathtt{AN}^{(2)}(F) \leq m$ if F can be computed by a multi-linear circuit with at most $(m+1)^2$ gates, each having arity at most m. For perspective, recall that [3, Thm. 4.4] actually asserted that *if the corresponding matrix corresponding to a bilinear function F has rigidity m^3 for rank m, then $\mathtt{AN}^{(1)}(F) \geq m$.*

Lemma 5.5 (rigidity and $\mathtt{AN}^{(2)}$): *Let F be a bilinear function and suppose that the corresponding matrix has rigidity m^4 for rank m^2. Then, $\mathtt{AN}^{(2)}(F) \geq m$.*

Proof: Suppose that $\mathtt{AN}^{(2)}(F) \leq m-1$, and consider a multi-linear circuit C of arity $m-1$ that has at most m^2 gates and computes F. We call a bilinear gate mixed if it fed both by bilinear gates and by either linear gates or variables, and call it a terminal if it is fed by linear gates and/or variables only. We first get rid of mixed gates by introducing, for each mixed gate M_i, an auxiliary bilinear gate B_i that "take over" the linear gates and variables that feed M_i, and feeds the modified M_i; that is, suppose that M_i is fed by a sequence of bilinear gates $\overline{Q}$ and a sequence of linear gates and variables $\overline{L}$, then $M_i(\overline{Q}, \overline{L})$ is replaced by $M_i'(\overline{Q}, B_i)$ and $B_i(\overline{L})$. Note that M_i' computes the sum of other bilinear gates, whereas B_i is a terminal.

The resulting number of terminal gates is at most m^2, because each new terminal gate (i.e., the terminal gate B_i introduced by the foregoing process) can be charged to a non-terminal bilinear gate in the original circuit (i.e., to M_i). Hence, all the bilinear gates in C are either terminal gates or compute the sum of other bilinear gates (as the to gate and the modified gates M_i'), and so C is the sum of the terminal gates, denoted G_i for $i \in [m^2]$; that is,

$$C(x,y) = \sum_{i\in[m^2]} G_i(x,y),$$

where the each G_i is fed by $m-1$ linear gates and variables.

Considering the sets of linear gates that feed into each of the G_i's, we stress that *these sets are all subsets of a set of at most m^2 linear gates*, since C has at most this number of gates. That is, $G_i(x,y)$ takes the sum of some products of pairs of linear gates and variables; specifically, each product takes one element from $S_i \cup V_i$ and one element from $S'_i \cup V'_i$, where $S_i \subseteq [m^2]$ (resp., $S'_i \subseteq [m^2]$) represents the set of linear gates in x (resp., in y) that feed G_i, and $V_i \subseteq [n]$ (resp., $V'_i \subseteq [n]$) denotes the set of x-variables (resp., y-variables) that feed G_i. Recall that $|S_i| + |S'_i| + |V_i| + |V'_i| \leq m-1$. Hence, G_i has the form

$$G_i(x,y) = \sum_{j \in S_i} L_j(x) M'_{i,j}(y) + \sum_{j \in S'_i} M_{i,j}(x) L'_j(y) + \sum_{(j,k) \in P_i \subseteq V_i \times V'_i} x_j y_k,$$

where the $L_j(x)$'s and $L'_j(y)$'s are linear gates of C, and the $M_{i,j}(x)$'s and $M'_{i,j}(y)$'s are arbitrary linear functions (which may depend on i). Specifically, $M_{i,j}(x)$ (resp., $M'_{i,j}(y)$) is a partial sum of $\sum_{k \in S_i} L_k(x) + \sum_{k \in V_i} x_k$ (resp., $\sum_{k \in S'_i} L_k(y) + \sum_{k \in V'_i} y_k$), where these partial sums are determined by G_i. Denoting $S \stackrel{\text{def}}{=} \cup_{i \in [m^2]} S_i$ and $S' \stackrel{\text{def}}{=} \cup_{i \in [m^2]} S'_i$, we can express C as

$$\begin{aligned} C(x,y) &= \sum_{i \in [m^2]} \left(\sum_{j \in S_i} L_j(x) M'_{i,j}(y) + \sum_{j \in S'_i} M_{i,j}(x) L'_j(y) + \sum_{(j,k) \in P_i \subseteq V_i \times V'_i} x_j y_k \right) \\ &= \sum_{j \in S} L_j(x) M'_j(y) + \sum_{j \in S'} M_j(x) L'_j(y) + \sum_{(j,k) \in P} x_j y_k, \end{aligned}$$

where $M'_j(x) = \sum_{i \in [m^2]} M'_{i,j}(x)$ (resp., $M_j(y) = \sum_{i \in [m^2]} M_{i,j}(y)$) and P is the multi-set consisting of $\cup_{i \in [m^2]} P_i$. Recalling that $|P_i| \leq |V_i| \cdot |V'_i| \leq (m-1)^2$, it follows that the matrix corresponding to the function computed by C is the sum of two matrices of ranks $|S|$ and $|S'| \leq m^2 - |S|$, respectively, and a matrix of sparsity $m^2 \cdot (m-1)^2$. That is, this matrix does not have rigidity m^4 for rank m^2. ■

Corollary 5.6 (an $\mathtt{AN}^{(2)}$ lower bound for random Toeplitz functions): *Almost all bilinear functions F that correspond to Toeplitz matrices satisfy $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{3/8})$.*

Proof: Using Lemma 5.5 it suffices to show that F has rigidity m^4 for rank m^2, where $m = \widetilde{\Omega}(n^{3/8})$. This follows from [4, Thm. 1.2], which asserts that a random Toeplitz matrix has rigidity $\Omega(n^3/r^2 \log n)$ for rank $r > \sqrt{n}$. Specifically, using $r = m^2 = \widetilde{\Omega}(n^{6/8})$, we get rigidity $\Omega(n^3/r^2 \log n) \geq m^4$, provided that $\Omega(n^3/\log n) \geq m^8$. ■

Corollary 5.7 (an $\mathtt{AN}^{(2)}$ lower bound for an explicit trilinear function): *The trilinear function $F(x,y,z) = \sum_{i,j \in [n]} x_i y_j z_{n+i-j}$ satisfies $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{3/8})$.*

Proof: As in [3, 4] (and Corollary 5.4), this follows from the existence of a bilinear function F' that corresponds to Toeplitz matrices such that $\mathtt{A}_3(F') = \widetilde{\Omega}(n^{3/8})$, which is asserted in Corollary 5.6. ■

6 Better Lower Bounds on other Explicit Functions

Recall that Corollaries 5.3 and 5.6 establish that *almost all bilinear functions F that correspond to Toeplitz matrices satisfy $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{0.4})$ and $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{3/8})$*. In this section we get improved bounds for function that belong to any set of $\exp(-n)$-biased space: Specifically, *a*lmost all bilinear functions F whose coefficients are taken from an 2^{-n}-biased space satisfy $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{4/9})$ as well as $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{0.4})$. Recall that these results yield similar lower bounds for an explicit 4-linear function [4]. (We shall consider bilinear functions in the variables $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, and 4-linear functions in the variables x, y and $(s', s'') \in \{0,1\}^{O(n)+O(n)}$.)

Preliminaries. We recall the definition of an ε-biased distribution (introduced by Naor and Naor [7]).

Definition 6.1 (small-biased distribution): *A distribution Z over $\{0,1\}^N$ is said to be ε-*biased *if for every non-empty set $S \subseteq [N]$, it holds that*

$$\left|\mathrm{E}_{z\sim Z}[(-1)^{\sum_{i\in S} z_i}]\right| \leq \varepsilon\ .$$

We shall use the following property of ε-biased distributions (which is implicit in [7]).

Claim 6.2 (upper-bounding the probability of hitting a linear space [1, Lem. 1]): *Let Z be an ε-biased distribution over $\{0,1\}^N$. Let $\ell_1, \ldots, \ell_t$ be linearly independent linear functions on $z_1, \ldots, z_N$. Then, the probability that all linear functions evaluate to 0 on $z \sim Z$ is at most $\varepsilon + 2^{-t}$; that is,*

$$\mathrm{Pr}_{z\sim Z}[(\forall i \in [t])\ \ell_i(z)=0] \leq \varepsilon + 2^{-t}.$$

6.1 The case of $\mathtt{A}_3$

Here we use techniques that that are similar to those used in [4], but the actual argument is different. We call the reader's attention to an argument at the end of Step 2 of the proof, where a union bound on too many values is avoided and the (linear equations satisfied by the) linear span of these values is considered instead.[3]

Theorem 6.3 (a $\mathtt{A}_3$ lower bound for bilinear functions selected from a small-biased sample space): *Almost all bilinear functions F that correspond to matrices drawn from a 2^{-n}-biased distribution on $\mathbb{F}_2^{n\times n}$ satisfy $\mathtt{A}_3(F) \geq \widetilde{\Omega}(n^{4/9})$.*

[3] This technique was used in [4].

Proof: Let m and r be non-negative integer parameters smaller than n, which we will set later. Along the way, we shall assume a few inequalities on m and r, which we will eventually satisfy by appropriately choosing m and r.

Our proof will show that the matrices associated with bilinear circuits of arity at most m and depth 3 can be partitioned into at most $\widetilde{O}(2^{n/2})$ families such that, for each family of matrices, there exists a system of $r^2/2$ (linearly independent) linear equations in the matrix entries that all matrices in the family satisfy. We will finish the proof by showing that most matrices drawn from a 2^{-n}-biased distribution on $\{0,1\}^{n^2}$ do not belong to any of these families, and hence cannot be computed by a bilinear depth-3 circuits of arity at most m.

Step 1: Classifying matrices to families. We start by classifying all matrices associated with bilinear functions F that satisfy $\mathtt{A}_3(F) \leq m$ into $O(2^{n/2})$ families of matrices such that in each family all entries in some r-by-r submatrix are linear combinations of $r^2/2$ values. Actually, the current step only identifies the families based on properties that will be useful towards identifying the linear combinations (in Step 2). Consider a depth-three multi-linear circuit C of arity m that computes F. As in Lemma 5.2, a generic C has the form

$$C(x,y) = \sum_{(i,j)\in[m]\times[m]} p_{i,j} \cdot \left(\sum_{\ell\in L_i} x_\ell\right) \cdot \left(\sum_{\ell\in L'_j} y_\ell\right) + \sum_{i\in[m]} Q_i(x,y),$$

where $P^{(0)} = (p_{i,j})_{i,j\in[m]} \in \{0,1\}^{m\times m}$, the L_i's and L'_j's are subsets of size at most m^2 of $[n]$, and

$$Q_i(x,y) = \sum_{(j,k)\in[m]\times[m]} p^{(i)}_{j,k} \cdot \left(\sum_{\ell\in L_{i,j}} x_\ell\right) \cdot \left(\sum_{\ell\in L'_{i,k}} y_\ell\right) + \sum_{j\in[m]} Q_{i,j}(x,y),$$

where $P^{(i)} = (p^{(i)}_{j,k})_{j,k\in[m]} \in \{0,1\}^{m\times m}$, the $L_{i,j}$'s and $L'_{i,k}$'s are subsets of size at most m of $[n]$, and the $Q_{i,j}(x,y)$'s are bilinear gates (each taking m variables).

To be more precise, for each $Q_{i,j}$, we associate two subsets $S^{(i,j)}, T^{(i,j)} \subseteq [n]$ corresponding to the indices of the x and y input variables of $Q_{i,j}$, respectively. We require $|S^{(i,j)}| + |T^{(i,j)}| \leq m$ and write $Q_{i,j}$ as

$$Q_{i,j}(x,y) = \sum_{k\in S^{(i,j)}} \sum_{\ell\in T^{(i,j)}} c_{i,j,k,\ell} \cdot x_k \cdot y_\ell \tag{2}$$

where $c_{i,j,k,\ell}$ are coefficients in $\{0,1\}$ (defined for any $k \in S^{(i,j)}$ and $\ell \in T^{(i,j)}$).

Hence, a concrete depth-three (multi-linear) circuit C of arity m is specified in terms of the foregoing generic description by specifying the sets $L_i, L'_i, L_{i,j}, L'_{i,j}$ and $S^{(i,j)}, T^{(i,j)}$, hereafter called the variable wiring (or wiring), as well as the $m+1$ matrices $P^{(i)}$'s (for $i = 0,1,...,m$) and the coefficients $c_{i,j,k,\ell}$'s, hereafter called the bilinear forms. Without loss of generality, we may envision C as a *formula* (i.e., a tree), and the elements in the foreging sequence of sets as its leaves; that is, each leaf corresponds to one of the elements in one of the sets, and each such element is an index of a variable from $x_1, \ldots, x_n, y_1, \ldots, y_n$. This formula has at most $5m^3$ leaves (i.e., $\sum_{i\in[m]}(|L_i|+|L'_i|)+\sum_{i,j\in[m]}(|L_{i,j}|+|L'_{i,j}|+|S^{(i,j)}|+|T^{(i,j)}|) \leq 2m \cdot m^2 + m^2 \cdot (2m+m) = 5m^3$), each labeled with a variable from $x_1, \ldots, x_n, y_1, \ldots, y_n$.

Let r be an integer and assume (for simplicity) that r divides n. We partition the x variables into n/r buckets, and similarly we partition the y variables. Specifically, for $a, b \in [n/r]$, let $X_a := \{x_{(a-1)\cdot r+1}, x_{(a-1)\cdot r+2}, \ldots, x_{i\cdot r}\}$ be the a^{th} bucket of the x variables, and let $Y_b := \{y_{(b-1)\cdot r+1}, y_{(b-1)\cdot r+2}, \ldots, y_{b\cdot r}\}$ be the b^{th} bucket of the y variables. For a fixed variable wiring, we call a bucket-pair (X_a, Y_b) typical if the following three conditions (or properties) hold:

1. At most $10 \cdot \frac{5m^3}{n/r} = \frac{50m^3r}{n}$ of the leaves in the formula are labeled with variables from X_a.
   (That is, the number of leaves labeled with a variable in X_a is at most ten times the expectation (for a random pair).)
2. At most $10 \cdot \frac{5m^3}{n/r}$ of the leaves in the formula are labeled with variables from Y_b.
3. There are at most $10 \cdot \frac{m^4}{(n/r)^2}$ quadruples (i, j, k, ℓ) such that $(x_k, y_\ell) \in X_a \times Y_b$ and x_k and y_ℓ are inputs to $Q_{i,j}$. (i.e., $k \in S^{(i,j)}$ and $\ell \in T^{(i,j)}$).

Observing that a random bucket-pair (X_a, Y_b) satisfies each condition (individually) with probability at least 0.9, it follows that most bucket-pairs satisfy all conditions simultaneously. Hence, for each wiring, most bucket-pairs (X_a, Y_b) are typical.

For each pair $(a, b) \in [n/r] \times [n/r]$, we consider all wirings for which (X_a, Y_b) is typical. Actually, it suffices to consider a partial wiring that specifies only the placing/wiring of variables in $X_a \cup Y_b$. To specify such a partial wiring it suffices to specify which of these variables appears in which leaf of the formula; that is, assign a variable of X_a (resp., Y_b) to at most $50m^3r/n$ of the leaves. Hence, we have at most

$$\binom{5m^3}{50m^3r/n} \cdot (|X_a|+1)^{50m^3r/n} < (n^4)^{50m^3r/n}$$

possibilities for wiring of variables in X_a, where the first factor corresponds to the choice of leaves and the second factor corresponds to the choice of a variable in X_a for each chosen leaf. Ditto for Y_b. Thus, there are at most $(n^4)^{100\cdot m^3 \cdot r/n}$ possible wirings for all variables in $X_a \cup Y_b$ to the gates that read them. We shall assume

$$100 \cdot m^3 \cdot \frac{r}{n} \leq \frac{n}{10 \cdot \log n} \tag{3}$$

giving us at most $(n^4)^{n/10\log n} = 2^{0.4\cdot n}$ possible wirings.

We partition all bilinear functions F with $\mathtt{A}_3(F) \leq m$ to families according to a choice of a bucket-pair (X_a, Y_b) and a partial wiring of $X_a \cup Y_b$ such that (X_b, Y_b) is typical for this wiring. This gives us an upper bound of $(n/r)^2 \cdot 2^{0.4n} < 2^{n/2}$

on the number of families. (Note that we have used Properties 1-2 of a typical bucket-pair in order to derive an upper bound on the number of families; we shall use Property 3 in the next step.)

Step 2: Associating a system of linear equations with each family of matrices. We consider a fixed family of matrices; that is, we fix a choice of a bucket-pair (X_a, Y_b) and a choice of wirings of $X_a \cup Y_b$ for which the said pair is typical. We focus on the r-by-r submatrices of the matrices in the family whose rows correspond to variables in X_a and columns correspond to variables in Y_b.

For every (k, ℓ) such that $x_k \in X_a$ and $y_\ell \in Y_b$, we consider how the (k, ℓ)-th entry of the matrices in the family looks like. Note that the (k, ℓ)-th entry in the matrix corresponding to the bilinear function equals the value of the bilinear function on the input $e_{k,\ell} \stackrel{\text{def}}{=} (0^{k-1}10^{n-k}, 0^{\ell-1}10^{n-\ell})$ (i.e., the input with all zeros except for x_k and y_ℓ). Now, for a fixed family, since the wirings of X_a and Y_b are fixed, the (k, ℓ)-th entry is a *fixed* linear combination in the entries that correspond to the $P^{(i)}$'s (with $i \in \{0, 1, ..., m\}$) and the relevant coefficients $c_{i,j,k,\ell}$ with $i, j \in [m]$, where the relevant coefficients $c_{i,j,k,\ell}$ are those for which $k \in S^{(i,j)}$ and $\ell \in T^{(i,j)}$. Specifically, letting $\chi_e(A) = 1$ if $e \in A$ and $\chi_e(A) = 0$ otherwise, we have

$$\begin{aligned} C(e_{k,\ell}) = & \sum_{j,j' \in [m] : \chi_k(L_j) = \chi_\ell(L'_{j'}) = 1} p_{j,j'} \\ & + \sum_{i,j,j' \in [m] : \chi_k(L_{i,j}) = \chi_\ell(L'_{i,j'}) = 1} p^{(i)}_{j,j'} \\ & + \sum_{i,j \in [m] : \chi_k(S^{(i,j)}) = \chi_\ell(T^{(i,j)}) = 1} c_{i,j,k,\ell}. \end{aligned}$$

Thus, each entry in the r-by-r submatrix corresponding to $X_a \times Y_b$ is a fixed linear combinations in the entries of $P^{(i)}$'s and the relevant coefficients $c_{i,j,k,\ell}$. There are at most $(m+1) \cdot m^2$ entries in the $P^{(i)}$'s, and at most $10 \cdot m^4 \cdot r^2/n^2$ relevant coefficients $c_{i,j,k,\ell}$ for $(k, \ell) \in X_a \times Y_b$ (by Property 3 of a typical bucket-pair). Assuming that

$$(m+1) \cdot m^2 + 10 \cdot m^4 \cdot \frac{r^2}{n^2} \le \frac{r^2}{2} \tag{4}$$

this means that the r^2 entries of a submatrix in a generic matrix in the family are fixed linear combinations of at most $r^2/2$ values (i.e., the entries of $P^{(i)}$'s and the relevant coefficients $c_{i,j,k,\ell}$). Hence, these r^2 entries must satisfy a fixed system of at least $r^2/2$ independent linear equations, since each entry is a fixed linear combination of at most $r^2/2$ values.[4]

[4] Formally, we can write each of the r^2 entries as a fixed linear combination of at most $r^2/2$ symbolic variables. Viewing these r^2 entries as an r^2-dimensional vector, we note that this vector must resided in a fixed vector space of dimension at most $r^2/2$ over $\mathbb{F}_2$, which in turn can be characterized by a fixed system of at least $r^2/2$ independent linear equations.

Step 3: Showing that, w.h.p., small-biased matrices do not belong to any of the families. To finish the proof, we show that a matrix drawn from a 2^{-n}-biased distribution is unlikely to be a member in any of these $2^{n/2} \cdot (n/r)^2$ families of matrices. For a fixed family, we upper-bound the probability that a matrix B drawn from a 2^{-n}-biased distribution belongs to this family, and then take a union bound over all families.

To be included in a fixed family, the matrix B should satisfy at least $r^2/2$ specific independent linear equations. By Claim 6.2, this happens with probability at most $2^{-n} + 2^{-r^2/2} \leq 2 \cdot 2^{-n}$ assuming $r \geq \sqrt{2n}$. Recalling that the number of families is smaller than $2^{n/2}$, it follows that, with very high probability, a random matrix B drawn from a 2^{-n}-biased distribution corresponds to a bilinear function F that satisfies $\mathtt{A}_3(F) \geq m$.

Conclusion. All that is left is picking r and m while satisfying Eq. (3), Eq. (4), and $r \geq \sqrt{2n}$. The choice

$$r \overset{\text{def}}{=} \frac{n^{2/3}}{8 \cdot \log^{1/3}(n)} \qquad \text{and} \qquad m \overset{\text{def}}{=} \frac{n^{4/9}}{8 \cdot \log^{2/9}(n)} = \frac{r^{2/3}}{2}$$

satisfies all of the above, assuming n is large enough. ■

Corollary 6.4 (an $\mathtt{A}_3$ lower bound for an explicit 4-linear function): *There exists an explicit bilinear function $G : \{0,1\}^{O(n)+O(n)} \to \{0,1\}^{n^2}$ such that the 4-linear function $F(x,y,s',s'') = \sum_{i,j\in[n]} G(s',s'')_{i,j} \cdot x_i y_j$ satisfies $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{4/9})$.*

Proof: As in [4], this follows by combining Theorem 6.3 with a construction of a small-biased generator $G : \{0,1\}^{O(n)+O(n)} \to \{0,1\}^{n^2}$ that is a bilinear function (see [6]). By Theorem 6.3, for most settings of $s = (s', s'')$, it holds that the resulting bilinear function $F_s(x,y) = \sum_{i,j\in[n]} G(s)_{i,j} \cdot x_i y_j$ satisfies $\mathtt{A}_3(F_s) = \widetilde{\Omega}(n^{4/9})$, whereas $\mathtt{A}_3(F) \geq \mathtt{A}_3(F_s)$ (for every s). ■

6.2 The case of $\mathtt{AN}^{(2)}$

We mention that following the proof in [4], one can get $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{0.4})$ for F's as in Theorem 6.3 and Corollary 6.4. We do not present the proof here, since it amounts to reproducing large portions of [4] (i.e., [4, Sec. 4] and [4, Sec. 5.1]), without any new ideas or techniques. The only difference would have been decoupling the number of gates from the arity, and using these two parameters rather than one. Specifically, we have

Theorem 6.5 ([4, Thm. 5.6], revised by decoupling size and arity):[5] *Let A be an n-by-n matrix A whose entries are sampled from an ε-biased distribution.*

[5] Indeed, in [4, Thm. 5.6], $s = m$.

Then, the corresponding bilinear function can be computed by a bilinear circuit of arity r having s gates with probability at most

$$\left(\frac{n}{2s}\right)^2 \cdot \binom{2s^2}{\leq 12s^2r/n}^4 \cdot \left(\varepsilon + 2^{-s^2+24s^3r^2/n^2}\right).$$

In particular, using $s = r^e > \sqrt{2n}$ (for any constant $e \in \mathbb{N}$) and $\varepsilon = 2^{-n}$, we get a probability bound of

$$\exp(\widetilde{O}(s^2r/n) - \min(n, s^2 - O(s^3r^2/n^2))) = \exp(\widetilde{O}(r^{2e+1}/n) - n),$$

assuming $r = o(n^{2/(e+2)})$. Hence, with high probability, the bilinear function F_A associated with a matrix A whose entries are sampled from an 2^{-n}-biased distribution satisfies $\mathtt{AN}^{(e)}(F_A) = \widetilde{\Omega}(n^{2/(2e+1)})$, since for a sufficiently small $r = \widetilde{\Omega}(n^{2/(2e+1)})$ it holds that $\widetilde{O}(r^{2e+1}/n) < n$.

Corollary 6.6 (an $\mathtt{AN}^{(e)}$ lower bound for an explicit 4-linear function): *There exists an explicit bilinear function $G : \{0,1\}^{O(n)+O(n)} \to \{0,1\}^{n^2}$ such that for every constant $e \in \mathbb{N}$ the 4-linear function $F(x, y, s', s'') = \sum_{i,j\in[n]} G(s', s'')_{i,j} \cdot x_iy_j$ satisfies $\mathtt{AN}^{(e)}(F) = \widetilde{\Omega}(n^{2/(2e+1)})$.*

For $e = 1$ this reproduces the $\widetilde{\Omega}(n^{2/3})$ lower bound of [4], but for $e \geq 2$ we get new bounds; for example, $\mathtt{AN}^{(2)}(F) = \widetilde{\Omega}(n^{0.4})$ and $\mathtt{AN}^{(3)}(F) = \widetilde{\Omega}(n^{2/7})$.

7 Depth Reductions

In this section, we show connections between $\mathtt{A}_d(\cdot)$ for different depths d. First, we show a simple connection between $\mathtt{A}_{kd}(\cdot)$ and $\mathtt{A}_d(\cdot)$ for any $k \in \mathbb{N}$. As a special case, we get $\mathtt{A}_{2k}(F) \geq \mathtt{A}_2(F)^{1/k}$. Next, we show a less clean connection between $\mathtt{A}_{2k+1}(F)$ and $\mathtt{A}_2(F)$. We note that establishing connections between $\mathtt{AN}^{(e)}(\cdot)$ for different values of e remains open.

Lemma 7.1 (depth reduction – simple case): *For any multilinear function F and $d, k \in \mathbb{N}$, it holds that*

$$\mathtt{A}_d(F) \leq \mathtt{A}_{kd}(F)^k.$$

As a special case, we get $\mathtt{A}_d(F) \geq \mathtt{A}_2(F)^{2/d}$ for every even depth d. Hence, *any non-trivial lower bound for depth 2 implies a non-trivial lower bound for every even depth d*, where a non-trivial lower bound for depth d refers to any lower bound of the form $\mathtt{A}_d(F) = \omega((tn)^{1/d})$ for a t-linear function F. This terminology is justified by the fact that a lower bound of the form $\mathtt{A}_d(F) = \Omega((tn)^{1/d})$ holds trivially for any t-linear function F that depends on all its tn input variables (because otherwise the multilinear circuit cannot even read all the input bits).

Proof Sketch: Starting with any multilinear circuit for F having depth kd and arity $m = \mathtt{A}_{kd}(F)$, collapse every k consecutive layers into one layer, resulting in a t-linear circuit of depth d and arity m^k. Hence, $\mathtt{A}_d(F) \leq \mathtt{A}_{kd}(F)^k$. ■

Since we have non-trivial lower bounds for depth 2, we get from Lemma 7.1 non-trivial lower bounds on $\mathtt{A}_d(\cdot)$ for any even d (see the discussion after Lemma 7.2 for specific details). We would like to get a similar result for odd depths, but the straightforward approach gives $\mathtt{A}_d(F) \geq \mathtt{A}_{d+1}(F) \geq \mathtt{A}_2(F)^{2/(d+1)}$ for every odd d. While this implies non-trivial lower bounds on $\mathtt{A}_d(F)$ for all sufficiently large odd d, at the time of performing this work, it yielded only trivial bounds for small d (e.g., $d = 3$).[6] Specifically, the best lower bound known (at the time) on an explicit function F asserts $\mathtt{A}_2(F) = \widetilde{\Omega}(n^{2/3})$, which implies only the trivial bound of $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{1/3})$.

Lemma 7.2 (depth reduction – odd depths to depth 2): *Let $k \in \mathbb{N}$. Then, for any t-linear function F, it holds that*

$$\mathtt{A}_2(F) \leq O(\mathtt{A}_{2k+1}(F)^{k+(t/(t+1))}).$$

Proof Sketch: The main idea is to first split the middle layer into two layers of smaller arity using [3, Thm. 3.1], and then collapse the top $k+1$ (resp., the bottom $k+1$) layers into one layer. Specifically, using [3, Thm. 3.1] (alternatively Theorem 4.1), split each gate in layer $k+1$ to an equivalent sub-circuit with two layers and arity $O(m)^{t/(t+1)}$. After the split, the circuit has $2k+2$ layers, where the first k layers have gates of arity at most m, the next two layers have gates of arity at most $O(m)^{t/(t+1)}$, and the last k layers have of gates with arity at most m. Collapsing the first $k+1$ layers and the last $k+1$ layers, results in a multilinear circuit of depth 2 and arity $O(m^{k+(t/(t+1))})$ computing F. Thus, $\mathtt{A}_2(F) = O(\mathtt{A}_d(F)^{k+(t/(t+1))})$ as required. ■

Corollaries. We use the lower bound from [4, Thm. 1.5], which asserts that the bilinear function associated with a random Toeplitz matrix has $\mathtt{A}_2(F) = \widetilde{\Omega}(n^{2/3})$, with high probability (over the random choices of the $2n-1$ values along the diagonals). Using Lemma 7.1, we get the non-trivial lower bound $\mathtt{A}_d(F) = \widetilde{\Omega}(n^{4/(3d)})$ for even depths d. For odd depths $d = 2k+1$, we use Lemma 7.2 to get the non-trivial lower bound

$$\mathtt{A}_d(F) = \widetilde{\Omega}\left(n^{\frac{2/3}{k+(t/(t+1))}}\right) = \widetilde{\Omega}\left(n^{4/(3d+1)}\right),$$

where the second equality uses the fact that $t = 2$ and $d = 2k+1$. As in [3, 4] (and Corollary 5.4), these lower bounds for random Toeplitz matrices imply a similar lower bound for an explicit *tri-linear* function.

Corollary 7.3 (an $\mathtt{A}_d$ lower bound for an explicit trilinear function): *The trilinear function $F(x,y,z) = \sum_{i,j\in[n]} x_i y_j z_{n+i-j}$ satisfies $\mathtt{A}_d(F) = \widetilde{\Omega}(n^{4/(3d)})$ for even d and $\mathtt{A}_d(F) = \widetilde{\Omega}(n^{4/(3d+1)})$ for odd d.*

In particular, we get $\mathtt{A}_3(F) = \widetilde{\Omega}(n^{0.4})$, just as in Corollary 5.4.

[6] Added in revision: This is no longer the case, since [2] presented an explicit poly$(1/\epsilon)$-linear function F_ϵ such that $\mathtt{A}_2(F_\epsilon) \geq n^{1-\epsilon}$ for every constant $\epsilon > 0$. Hence, $\mathtt{A}_d(F_\epsilon) \geq \mathtt{A}_{d+1}(F_\epsilon) \geq n^{(2-2\epsilon)/(d+1)}$ holds for every odd d.

Remark: In light of the above, it may seem that Section 5.1 is redundant. However, on top of serving as a warmup for Sections 5.2 and 6.1, the contents Section 5.1 is not exhausted by Corollary 5.4, since it offers a structural result for matrices associated with low-complexity depth-3 bilinear circuits (i.e., Lemma 5.2). Furthermore, the proof in Section 5.1 relies on a rigidity lower bound of [4, Thm. 1.2], whereas Corollary 7.3 relies on a higher lower bound on "structured rigidity" provided by [4, Thm. 1.5] via a more complex proof.

8 Summary of bounds on the generalized AN-complexity

Our study of the two complexity measures (i.e., $\mathtt{A}_d$ and $\mathtt{AN}^{(d-1)}$) is guided by the following two facts:

- A generic upper bound that assert that *for every $d, t \in \mathbb{N}$, every t-linear function F satisfies* $\mathtt{AN}^{(d-1)}(F) \leq \mathtt{A}_d(F) = O(tn)^{t/(t+d-1)}$ (see Theorem 4.1).
- A matching lower bound that holds for almost all t-linear functions, which actually asserts that *for every $d, t \in \mathbb{N}$, almost every t-linear function F satisfies* $\mathtt{A}_d(F) \geq \mathtt{AN}^{(d-1)}(F) = \Omega(tn)^{t/(t+d-1)}$ (see Theorem 4.2).

Recall that it is trivial to show that the n-bit partity function, denoted $\mathtt{PAR}_n$, satisfies $\mathtt{A}_d(F) \geq \mathtt{AN}^{(d-1)}(\mathtt{PAR}_n) = \Omega(n^{1/d})$. Any lower bound that is greater than that is considered non-trivial.

Our lower bounds for explicit functions are non-trivial, but do not meet the foregoing goal. Specifically, we focus on the case of $d = 3$ (i.e., $\mathtt{A}_3$ and $\mathtt{AN}^{(2)}$), whereas the case of $d = 2$ (i.e., $\mathtt{A}_2 = \mathtt{AN}_2$ and $\mathtt{AN}^{(1)} = \mathtt{AN}$) was studied in [3, 4]. Our results include:

- A non-trivial $\mathtt{A}_3$ lower bound for an explicit trilinear function: The trilinear function $F_3(x, y, z) = \sum_{i,j\in[n]} x_i y_j z_{n+i-j}$ satisfies $\mathtt{A}_3(F_3) = \widetilde{\Omega}(n^{0.4})$ (see Corollary 5.4).
- A non-trivial $\mathtt{AN}^{(2)}$ lower bound for an explicit trilinear function: The foregoing trilinear function F_3 satisfies $\mathtt{AN}^{(2)}(F_3) = \widetilde{\Omega}(n^{3/8})$ (see Corollary 5.7)
- A non-trivial $\mathtt{A}_3$ lower bound for an explicit 4-linear function: There exists an explicit 4-linear function F_4 that satisfies $\mathtt{A}_3(F_4) = \widetilde{\Omega}(n^{4/9})$ (see Corollary 6.4).
- A non-trivial $\mathtt{AN}^{(2)}$ lower bound for an explicit 4-linear function: The foregoing 4-linear function F_4 satisfies $\mathtt{AN}^{(2)}(F_4) = \widetilde{\Omega}(n^{0.4})$ (see Corollary 6.6).

Actually, Corollary 6.6 yields a lower bound for every $e \geq 2$ asserting that *the foregoing 4-linear function F_4 satisfies* $\mathtt{AN}^{(e)}(F_4) = \widetilde{\Omega}(n^{2/(2e+1)})$ *for every $e \geq 2$.* This implies $\mathtt{A}_d(F_4) = \widetilde{\Omega}(n^{2/(2d-1)})$ for every $d \geq 3$. For $d \geq 4$, a stronger lower bound for $\mathtt{A}_d$ asserts that *the foregoing trilinear function F_3 satisfies* $\mathtt{A}_d(F_3) = \widetilde{\Omega}(n^{4/(3d)})$ *for even d and* $\mathtt{A}_d(F_3) = \widetilde{\Omega}(n^{4/(3d+1)})$ *for odd d* (see Corollary 7.3). While all these lower bounds are non-trivial for every $d \geq 2$, only the last bound does not approach the trivial bound when d is large enough (but rather approaches a 4/3-power of the trivial bound).

References

1. Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.
2. Oded Goldreich. Improved bounds on the AN-complexity of multilinear functions. *ECCC*, TR19-171, November 2019.
3. Oded Goldreich and Avi Wigderson. On the Size of Depth-Three Boolean Circuits for Computing Multilinear Functions. *ECCC*, TR13-043, March 2013. See revised version in this volume.
4. Oded Goldreich and Avishay Tal. Matrix Rigidity of Random Toeplitz Matrices. *ECCC*, TR15-079, May 2015.
5. Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*. Algorithms and Combinatorics, Vol. 27, Springer, 2012.
6. Elchanan Mossel, Amir Shpilka, and Luca Trevisan. On ϵ-biased generators in NC^0. In 44th FOCS, pages 136–145, 2003.
7. Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.
8. Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. Mathematical Foundations of Computer Science, Springer, Lecture Notes in Computer Science (Vol. 53), pages 162–176, 1977.
9. Leslie G. Valiant. Exponential lower bounds for restricted monotone circuits. In *15th STOC*, pages 110–117, 1983.