comput. complex. **27** (2018), 305 – 350

1016-3328/18/020305-46
published online September 1, 2016
DOI 10.1007/s00037-018-0144-9

computational complexity

MATRIX RIGIDITY OF RANDOM TOEPLITZ MATRICES

Oded Goldreich and Avishay Tal

Abstract. A matrix A is said to have rigidity s for rank r if A differs from any matrix of rank r on more than s entries. We prove that random n-by-n Toeplitz matrices over $\mathbb{F}_2$ (i.e., matrices of the form $A_{i,j} = a_{i-j}$ for random bits $a_{-(n-1)}, \ldots, a_{n-1}$) have rigidity $\Omega(n^3/(r^2 \log n))$ for rank $r \geq \sqrt{n}$, with high probability. This improves, for $r = o(n/\log n \log\log n)$, over the $\Omega(\frac{n^2}{r} \cdot \log(\frac{n}{r}))$ bound that is known for many explicit matrices.

Our result implies that the explicit trilinear $[n] \times [n] \times [2n]$ function defined by $F(x, y, z) = \sum_{i,j} x_i y_j z_{i+j}$ has complexity $\Omega(n^{3/5})$ in the multilinear circuit model suggested by Goldreich and Wigderson (Electron Colloq Comput Complex 20:43, 2013), which yields an $\exp(n^{3/5})$ lower bound on the size of the so-called *canonical* depth-three circuits for F. We also prove that F has complexity $\tilde{\Omega}(n^{2/3})$ if the multilinear circuits are further restricted to be of depth 2.

In addition, we show that a matrix whose entries are sampled from a 2^{-n}-biased distribution has complexity $\tilde{\Omega}(n^{2/3})$, regardless of depth restrictions, almost matching the known $O(n^{2/3})$ upper bound for any matrix. We turn this randomized construction into an explicit 4-linear construction with similar lower bounds, using the quadratic small-biased construction of Mossel et al. (Random Struct Algorithms 29(1):56–81, 2006).

Keywords. Matrix rigidity, multi-linear functions, multi-linear circuits

Subject classification. 68Q17

1. Introduction

This paper concerns the construction of rigid matrices, a central open problem posed by Valiant (1977), and its application to lower bounds on *canonical* depth-three Boolean circuits (a restricted model of depth-three circuits defined by Goldreich & Wigderson (2013)). In particular, we improve the known lower bound on matrix rigidity, but the improvement is for a range of parameters that is not the one motivated by Valiant's problem, but rather the one that arises from Goldreich & Wigderson (2013). Indeed, this improvement resolves open problems posed by Goldreich & Wigderson (2013).

1.1. Matrix rigidity. The "Matrix Rigidity Problem" (i.e., providing explicit matrices of high rigidity) is one of the most alluring problems in arithmetic circuits lower bounds. Introduced in 1977 by Valiant (1977), the problem was originally motivated by proving lower bounds for the computation of linear transformations. Loosely speaking, a matrix is called rigid if it cannot be written as a sum of a low rank matrix and a sparse matrix. Needless to say, the actual definition specifies both parameters.

Definition 1.1 (Matrix rigidity, Valiant 1977). *A matrix A over a field $\mathbb{F}$ has* rigidity *s* for rank *r if every matrix of rank at most r (over $\mathbb{F}$) differs from A on more than s entries.*

Valiant showed that any n-by-n matrix with rigidity $n^{1+\delta}$ for rank $\omega(n/\log\log n)$, where δ is some constant greater than 0, cannot be computed by a linear circuit of size $O(n)$ and depth $O(\log n)$. Valiant also proved that almost all n-by-n matrices, over a finite field $\mathbb{F}$ (e.g., the two-element field $\mathbb{F}_2$), have rigidity $\Omega((n-r)^2/\log n)$ for rank r. Since then, coming up with an explicit[1] rigid matrix has remained a challenge. The best techniques to date provide explicit n-by-n matrices of rigidity $\frac{n^2}{r}\log(\frac{n}{r})$ for rank r (see Friedman (1993) and Shokrollahi *et al.* (1997)). See Lokam (2009) for a survey on the subject.

[1] For an infinite $I \subseteq \mathbb{N}$, the sequence of matrices, $\{A_n\}_{n\in I}$ such that A_n is an $n \times n$ matrix, is called explicit if there exists a poly(n)-time algorithm that on input $n \in I$ outputs the matrix A_n (and outputs $\perp$ if $n \notin I$).

To the best of our knowledge, this state of affairs also holds for "simple" randomized constructions that use $O(n)$ random bits. The common belief is that rigidity bounds for such randomized constructions can be used for proving lower bounds for explicit computational problems that are related to the original ones. For example, an adequate rigidity lower bound for random Toeplitz matrices would yield a lower bound on the complexity of computing explicit bilinear transformations. Indeed, this is analogous to Andreev's proof of formula lower bounds (Andreev 1987), where a lower bound for a randomized function is transformed into a lower bound for an explicit function (which takes the $O(n)$ random bits of the construction as part of its input, increasing the input size only by a constant factor).[2]

Our main result shows that random Toeplitz/Hankel matrices are rigid with high probability. Recall that a Toeplitz matrix $T = (T_{i,j})$ has constant diagonals (i.e., $T_{i,j} = T_{i+1,j+1}$ for every i, j). Hankel matrices are obtained by turning Toeplitz matrices upside down; that is, a Hankel matrix $H = (H_{i,j})$ has constant skew-diagonals (i.e., $H_{i,j} = H_{i+1,j-1}$ for every i, j). Hence, any claim regarding one family translates to an equivalent claim regarding the other family.

THEOREM 1.2 (On the rigidity of random Toeplitz/Hankel matrices). *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Toeplitz/Hankel matrix. Then, for every $r \in [\sqrt{n}, n/32]$, with probability $1 - o(1)$, the matrix A has rigidity $\Omega(\frac{n^3}{r^2 \log n})$ for rank r.*

Our bounds are asymptotically better than $\Omega(\frac{n^2}{r}\log(\frac{n}{r}))$ for rank $r = o(\frac{n}{\log n \cdot \log\log n})$, alas Valiant's original motivation refers to $r > n/\log\log n$. For rank $r = n^{0.5+\varepsilon}$, where $\varepsilon \in (0, 0.5)$, our bound yields a significant improvement (i.e., $\frac{n^3}{r^2} = n^{2-2\varepsilon} \gg n^{1.5-\varepsilon} = \frac{n^2}{r}$), and this is actually the range that is relevant for the project of Goldreich & Wigderson (2013).

[2] Lower bounds for matrix multiplication and polynomial multiplication, in the model of arithmetic circuits over the reals with bounded constants, were previously achieved using this approach (Bürgisser & Lotz 2004; Raz 2003).

1.2. Goldreich–Wigderson's project. The work of Goldreich & Wigderson (2013) provides another motivation for the study of matrix rigidity. In fact, the problem of improving the rigidity bounds for random Toeplitz matrices was posed explicitly there. Specifically, proving a rigidity bound of $n^{1.5+\Omega(1)}$ for rank $n^{0.5+\Omega(1)}$ for random Toeplitz matrices was proposed there as a possible next step.

Lower Bounds for Depth Three Canonical Circuits. Håstad (1989) showed that any depth-three Boolean circuit[3] computing the n-way parity function must be of size at least $\exp(\sqrt{n})$. Though Håstad's bound was refined during the years (Paturi *et al.* 2005, 1999), to date, $\exp(\Omega(\sqrt{n}))$ is the best lower bound for an explicit function in the model of depth-three Boolean circuits. The work of Goldreich & Wigderson (2013) put forward a model of *depth three canonical circuits*, with the underlying long-term goal to exhibit better lower bounds for general depth-three Boolean circuits computing explicit *multi-linear* functions.

Canonical circuits are restricted type of Boolean depth-three circuits, which can be illustrated by considering the smallest known depth-three circuits for n-way parity. The latter $\widetilde{O}(2^{\sqrt{n}})$-size circuits are obtained by combining a CNF that computes a $\sqrt{n}$-way parity with $\sqrt{n}$ DNFs that compute $\sqrt{n}$-way parities of disjoint blocks of the input bits. The construction suggests the following scheme for obtaining Boolean circuits that compute multilinear functions. First, construct an arithmetic circuit that uses arbitrary multilinear gates of parameterized arity, and then convert it to a Boolean circuit whose size is exponential in the maximum between the arity and the number of gates in the arithmetic circuit. The arithmetic model is outlined next.

Lower Bounds for Multilinear Circuits. Suppose we wish to compute a t-linear function that depends on t blocks of inputs, $x^{(1)}, \ldots, x^{(t)}$, each of length n; that is, the function is linear in each of the $x^{(j)}$'s. We consider circuits that use arbitrary t-linear gates of parameterized arity. That is, the circuits are directed

[3] That is, a circuit of unbounded fan-in OR and AND gates with leaves that are variables or their negations.

acyclic graphs, where each internal node computes a t-linear function of its inputs. We further restrict our circuit such that each internal gate computes a multilinear formal polynomial in the inputs $x^{(1)} \ldots, x^{(t)}$. We say that such a multilinear circuit is of AN-complexity[4] m if m equals the maximum between the number of the circuit gates and the maximal arity of the gates. For a t-linear function F, we denote by $\mathsf{C}(F)$ the minimal AN-complexity of a multilinear circuit which compute the function F. (We will abuse notation and refer to the AN-complexity of a tensor/matrix as the AN-complexity of the corresponding t-linear function.)

In the example of parity, we have a bottom layer of $\sqrt{n}$ gates each taking $\sqrt{n}$ inputs and computing their parity. Above these gates, we have a gate which takes the $\sqrt{n}$ results and computes their parity. Overall, we got a (multi)-linear circuit of AN-complexity $\sqrt{n} + 1$.

Goldreich and Wigderson showed that any multilinear circuit of AN-complexity m yields a depth-three Boolean circuit of size $\exp(m)$ computing the same function (see Goldreich & Wigderson 2013, Prop. 2.9). In fact, these Boolean circuits have much more structure, and are referred to by Goldreich and Wigderson as *canonical circuits*. Thus, a preliminary step towards beating the $\exp(\Omega(\sqrt{n}))$ lower bound on the size of depth-three Boolean circuits for explicit $O(1)$-linear functions,[5] will be to beat the $\Omega(\sqrt{n})$ AN-complexity lower bound for such functions in the model of multilinear circuits.

Again, as in Valiant's question, if we just ask about the existence of hard t-linear functions, then most t-linear functions cannot be computed by a multilinear circuit of AN-complexity smaller than $(nt)^{t/(t+1)}$: See Goldreich & Wigderson (2013, Thm. 4.1), which uses a counting argument. The more important and challenging problem is to came up with an explicit t-linear function for which such bounds, or even just $\omega(\sqrt{n})$ lower bounds, can be proved.

[4] Where AN stands for Arity and Number of gates.

[5] Indeed, this suggestion presumes that there exist $O(1)$-linear functions that require depth-three Boolean circuits of size $\exp(\omega(\sqrt{n}))$, which is also an open problem suggested in Goldreich & Wigderson (2013).

Reduction to (Structured) Rigidity. Goldreich and Wigderson reduce the problem of proving lower bounds for bilinear circuits to the problem of rigidity (Goldreich & Wigderson 2013, Sect. 4.2). They show that if a bilinear circuit is of AN-complexity $m/2$, then its corresponding matrix is not m^3 rigid for rank m (i.e., it can be expressed as a sum of an m^3-sparse matrix and a matrix of rank at most m). Hence, any matrix that has rigidity m^3 for rank m corresponds to a bilinear function that cannot be computed by a bilinear circuit of AN-complexity at most $m/2$. Furthermore, Goldreich and Wigderson show that the sparse matrix arising from their reduction has an additional structure (to be specified later). This leads to a weaker notion of rigidity (see Goldreich & Wigderson (2013, Thm. 4.12) which establishes a separation), called *structured rigidity*, for which it is potentially easier to prove lower bounds.

Open Problems in Goldreich-Wigderson. One open problem posed by Goldreich and Wigderson is proving that random Toeplitz matrices have rigidity m^3 (or just structured rigidity m^3) for rank $m = n^{0.5+\Omega(1)}$. This would yield an AN-complexity lower bound of m for the corresponding bilinear function (via the reduction in Goldreich & Wigderson 2013, Thm. 4.4)[6] as well as a similar lower bound for the following explicit trilinear function (via Goldreich & Wigderson 2013, Prop. 4.6):

$$F_{\mathrm{tet}}(x,y,z) = \sum_{\substack{i_1,i_2,i_3\in[n]:\\ \sum_{j=1}^{3}|i_j-n/2|\le n/2}} x_{i_1}y_{i_2}z_{i_3}\ . \tag{1.3}$$

1.3. Resolving the foregoing open problems. We resolve the aforementioned open problem (Goldreich & Wigderson 2013, Prob. 4.8) by proving that random Toeplitz matrices have rigidity m^3 for rank $m = \Theta(\frac{n^{3/5}}{\log^{1/5} n})$, with high probability. This follows from our main theorem (Theorem 1.2) by choosing $r = m$. Furthermore, we can remove the logarithmic factor in the $\widetilde{\Omega}$ notation, by proving a slightly better lower bound for structured rigidity.

[6] For structured rigidity, we use Goldreich & Wigderson (2013, Thm. 4.10).

THEOREM 1.4 (On the structured rigidity of random Toeplitz and Hankel matrices). *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Toeplitz/Hankel matrix. Then, for every $r \in [\sqrt{n}, n/32]$, the matrix A has structured rigidity $\Omega(n^3/r^2)$ for rank r.*

This implies (using Goldreich & Wigderson 2013, Thm. 4.10 and Goldreich & Wigderson 2013, Prop. 4.6) that the AN-complexity of a random Toeplitz matrix is $\Omega(n^{3/5})$, and ditto for the explicit trilinear function F_{tet} from Eq. (1.3). This resolves Problems 4.7 and 4.2 in Goldreich & Wigderson (2013), resp. In addition, we show that another explicit trilinear function has AN-complexity $\Omega(n^{3/5})$.

COROLLARY 1.5 (AN-complexity lower bound for an explicit trilinear function). *Let $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{2n} \to \{0,1\}$ be the trilinear function defined by $F(x,y,z) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j z_{i+j}$. Then, $\mathtt{C}(F) = \Omega(n^{3/5})$.*

New challenges. The most natural question that arises from the foregoing results is to tighten the lower bound; that is, to show that random Toeplitz matrices have AN-complexity $\Omega(n^{2/3})$ as conjectured by Goldreich & Wigderson (2013). This would be the best possible, since any bilinear function can be computed by a bilinear circuit of AN-complexity $O(n^{2/3})$; more generally, by Goldreich & Wigderson (2013, Thm. 3.1), for any $t \geq 2$, any t-linear function can be computed by a t-linear circuit of AN-complexity $O((tn)^{t/(t+1)})$. Another natural follow up question is to exhibit an explicit $O(1)$-linear function having AN-complexity $\Omega(n^\alpha)$ for some constant $\alpha > 3/5$; of course, the larger α, the better. Our progress on these open problems is captured by the following two results.

THEOREM 1.6 (Depth-two AN-complexity lower bound for random Toeplitz matrices). *Let F be a bilinear function that corresponds to a random Toeplitz matrix. Then, with probability $1 - o(1)$, the function F cannot be computed by multilinear circuits of* depth two *having AN-complexity $n^{2/3}/(\log n)^{1/3}$.*

Theorem 1.6 establishes the desired AN-complexity lower bound for random Toeplitz matrices, but only for depth-two multilinear

circuits. We note that the AN-complexity upper bound of Goldreich & Wigderson (2013, Thm. 3.1) holds via depth-two circuits, and so Theorem 1.6 is almost optimal with respect to depth-two multilinear circuits. Theorem 1.6 implies that the trilinear function $F(x,y,z) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j z_{i+j}$ cannot be computed by multilinear circuits of *depth two* and AN-complexity $n^{2/3}/(\log n)^{1/3}$.

THEOREM 1.7 (Improved AN-complexity lower bound for explicit 4-linear functions). *There exists an explicit 4-linear function having AN-complexity* $\Omega(n^{2/3}/(\log n)^{1/3})$.

Theorem 1.7 is proved by first showing that, with high probability, bilinear functions associated with matrices that are sampled from a 2^{-n}-biased sample space (over $\{0,1\}^{n^2}$) have AN-complexity $\widetilde{\Omega}(n^{2/3})$. Note that by the aforementioned upper bound, this lower bound is tight (up to logarithmic factors). Next, we note that sampling such matrices can be done using $O(n)$ random bits (Alon *et al.* 1992; Mossel *et al.* 2006; Naor & Naor 1993), which matches the amount of randomness used for sampling a random Toeplitz matrix. Furthermore, in the explicit small-biased construction of Mossel *et al.* (2006), each bit in the sampled string is a bilinear function of the random bits, allowing us to give an explicit 4-linear function with AN-complexity $\widetilde{\Omega}(n^{2/3})$.

1.4. Overview of the Proof of Theorem 1.2. We give an overview of the proof of Theorem 1.2 (for the case of Hankel matrices). Recall that we wish to show that a random Hankel matrix has rigidity $\Omega(n^3/(r^2\log n))$ for rank r, with high probability. Let A be a random n-by-n Hankel matrix, of the form $A_{i,j} = a_{i+j}$ for independent random bits $a_2,\ldots,a_{2n}$. We partition A into $(n/2r)^2$ submatrices each of size $2r\times 2r$ and show that with high probability each submatrix A' has rigidity $\Omega(n/\log n)$ for rank r. This easily implies that A has rigidity $\Omega((n/\log n)\cdot(n/2r)^2) = \Omega(\frac{n^3}{r^2\log n})$ for rank r, which will complete the proof.

Consider the partition of A into $(n/2r)\cdot(n/2r)$ submatrices, each of size $2r\times 2r$, such that a generic submatrix consists of $2r$ consecutive columns and $2r$ equally spaced rows (i.e., rows that

are at distance $n/2r$ apart). Next, we note that any of the above submatrices of A are of the form

$$A' = \begin{pmatrix} a_{i+1} & a_{i+2} & \dots & a_{i+2r} \\ a_{i+k+1} & a_{i+k+2} & \dots & a_{i+k+2r} \\ \dots & \dots & \dots & \dots \\ a_{i+(2r-1)k+1} & a_{i+(2r-1)k+2} & \dots & a_{i+(2r-1)k+2r} \end{pmatrix}$$

where $k = n/2r$ ($\leq 2r$, by the assumption $r \geq \sqrt{n}$), and i is determined by the location of A' in A (i.e., if A' is the (i', j)th submatrix, then $i = i' - 1 + (j-1) \cdot 2k$). Notice that A' is a $2r \times 2r$ submatrix that depends on $(2r-1)k + 2r = \Theta(n)$ random bits. This allows us to handle up to $\exp(n)$ bad events when applying a union bound.

In our main lemma, we show that for any fixed matrix S' (even if S' is not sparse) the submatrix matrix $A' - S'$ is of rank greater than r with probability at least $1 - 2^{-\Omega(n)}$, where the probability is taken over the choice of A' (equiv., over the choice of $a_{i+1}, \dots, a_{i+(2r-1)k+2r}$). As the number of $o(n/\log n)$-sparse matrices is $2^{o(n)}$, we may apply a union bound over all possible sparse submatrices and get that with high probability the submatrix A' has rigidity $\Omega(n/\log n)$ for rank r.

1.5. Organization. Our main results (i.e., Theorem 1.2, Theorem 1.4 and Corollary 1.5) are proved in Section 3, which follows a short preliminary section (Section 2). Next, Theorem 1.6 and 1.7 are proved, in two steps. In Section 4 we identify structural properties of matrices that correspond to bilinear functions of low AN (and AN2) complexity. These properties correspond to (even more) restricted notions of structured rigidity, and in Section 5 we show that (with high probability) matrices drawn from the two relevant distributions do not satisfy these properties. We conclude, with a technical digest (Section 6.1), a remark on the randomness-rigidity tradeoff (Section 6.2), and a list of some open problems (Section 6.3).

In the appendices, we generalize Theorem 1.2 and 1.4 to general finite fields (Appendix A.1) and prove that AN-complexity and AN2-complexity are equivalent to restricted notions of structured rigidity (Appendix A.3).

2. Preliminaries

We denote by $[n] = \{1, \ldots, n\}$. For $n, k \in \mathbb{N}$, we denote by $\binom{n}{\leq k} = \sum_{i=0}^{k} \binom{n}{i}$, and use the following (crude) bound which suffices for our argument

$$\binom{n}{\leq k} \leq 2^k \binom{n}{k} \leq \min\{(2n)^k, (6n/k)^k\}. \tag{2.1}$$

For a matrix A, we denote its ith row by A_i, and its jth column by $A^{(j)}$. We denote by $\mathtt{wt}(A)$ the number of non-zero entries in the matrix A, and say that A is s-sparse if $\mathtt{wt}(A) \leq s$.

A Hankel matrix over a field $\mathbb{F}$ is a square matrix with constant skew-diagonals; that is, any matrix $A \in \mathbb{F}^{n \times n}$ of the form $A_{i,j} = a_{i+j}$ for some $a_2, \ldots, a_{2n} \in \mathbb{F}$. A Toeplitz matrix over a field $\mathbb{F}$ is a square matrix with constant diagonals, i.e. any matrix $A \in \mathbb{F}^{n \times n}$ of the form $A_{i,j} = a_{i-j}$ for some $a_{-(n-1)}, \ldots, a_{n-1} \in \mathbb{F}$. Note that a Hankel matrix is an "upside-down" Toeplitz matrix.

Throughout the paper, unless specified otherwise, we talk about matrices over the field $\mathbb{F}_2$, and matrix rank refers to the rank over $\mathbb{F}_2$.

Definition 2.2 (Structured rigidity, (Goldreich & Wigderson 2013, Def. 4.9)). *We say that a matrix A has* structured rigidity (m_1, m_2, m_3) for rank r *if for every matrix R of rank at most r and for every $X_1, \ldots X_{m_1}, Y_1, \ldots, Y_{m_1} \subseteq [n]$ such that $|X_1| = \cdots = |X_{m_1}| = m_2$ and $|Y_1| = \cdots = |Y_{m_1}| = m_3$ it holds that $A - R \not\subseteq \bigcup_{k=1}^{m_1} (X_k \times Y_k)$, where $M \subseteq S$ means that all non-zero entries of the matrix M reside in the set $S \subseteq [n] \times [n]$. We say that a matrix A has* structured rigidity m^3 for rank r *if A has structured rigidity (m, m, m) for rank r.*

Indeed, any matrix that has rigidity s for rank r, also has structured rigidity s for rank r, but the other direction does not hold (see Goldreich & Wigderson 2013, Thm. 4.12).

Definition 2.3 (Multilinear circuits). *A* multilinear circuit *on t blocks of inputs $x^{(1)}, \ldots, x^{(t)} \in \{0,1\}^n$ is a directed acyclic graph whose nodes are associated with arbitrary multilinear gates, such*

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 another gate.

Definition 2.4 (The AN-complexity of multilinear circuits with general gates, (Goldreich & Wigderson 2013, Def. 2.2)). *The* arity *of a multilinear circuit is the maximum arity of its (general) gates. The* AN-complexity of a multilinear circuit *is the maximum between its arity and its number of gates (where we count only the general gates and not the leaves, i.e., variables). The* AN-complexity of a multilinear function F, *denoted* $\mathtt{C}(F)$, *is the minimum AN-complexity of a multilinear circuit that computes* F. *The* AN2-complexity of a multilinear function F, *denoted* $\mathtt{C}_2(F)$, *is the minimum complexity of a depth-two multilinear circuit that computes* F.

Theorem 2.5 (Goldreich & Wigderson 2013, Thm. 4.10). *If* A *is an* n*-by-*n *matrix that has structured rigidity* m^3 *for rank* m, *then the corresponding bilinear function* F *satisfies* $\mathtt{C}(F) \geq m/2$.

3. Main results

We prove our results bottom-up, starting with the main lemma, as mentioned in the proof overview.

Lemma 3.1 (Main Lemma). *Let* $m, k \in \mathbb{N}$, $16 \leq k \leq m$. *Let* $A \in \mathbb{F}_2^{m \times m}$ *be the random matrix*

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_m \\ a_{k+1} & a_{k+2} & a_{k+3} & \dots & a_{k+m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{(m-1)k+1} & a_{(m-1)k+2} & a_{(m-1)k+3} & \dots & a_{(m-1)k+m} \end{pmatrix}$$

where $a_1, \dots, a_{(m-1)k+m}$ *are uniform independent random bits, and let* $S \in \mathbb{F}_2^{m \times m}$ *be some fixed matrix. Then,* $\mathbf{Pr}_A[\mathrm{rank}(S + A) \leq m/2] \leq 2^{-km/16}$.

Note that for $k = 1$ the matrix in Lemma 3.1 is a random Hankel matrix, and for $k = m$ it is a totally random matrix. The requirement $k \geq 16$ is not essential in the lemma; it is used to make expressions nicer. For $k \geq 1$ and rank $r \leq m/2$ the proof gives $\mathbf{Pr}_A[\text{rank}(S + A) \leq r] \leq \binom{m}{\leq r} \cdot 2^{-mk/8}$.

PROOF. For a fixed S and a random A as above, let $B = S + A$. If $r = \text{rank}(B) \leq m/2$, then one can construct a basis $B_{i_1}, B_{i_2}, \ldots, B_{i_r}$ of the row space of B by the following iterative process: Let i_1 be the first nonzero row of B, let $i_2 > i_1$ be the first row in B that is not spanned by row i_1, let $i_3 > i_2$ be the first row in B that is not spanned by rows i_1 and i_2, etc. We get that $i_1 < i_2 < \cdots < i_r$ and

1. For $j < i_1$ the jth row of B is the all zeroes row.
2. For $i_{t-1} < j < i_t$ the jth row of B is spanned by rows $i_1, \ldots, i_{t-1}$ of B.
3. For $i_r < j$ the jth row of B is spanned by rows $i_1, \ldots, i_r$ of B.

More concisely, denoting by $I = \{i_1, \ldots, i_r\}$, we get

$$\forall j \in [m] \backslash I : B_j \in \text{span}\{B_i : i \in I, i < j\} \; . \tag{3.2}$$

We bound the probability that such a sequence $I = \{i_1, \ldots, i_r\}$ exists, where $r \leq m/2$. We apply a union bound over all possible sequences I, and for any fixed sequence of length at most $m/2$, we shall show that (3.2) holds with very low probability. Given such a sequence I, let $J = [m] - I$ be its complement. Setting $\Delta = \lceil m/k \rceil$, we can select an increasing sequence of $|J|/\Delta$ indices in J such that each two indices differ by at least Δ.[7] Take $j_1 < j_2 < \cdots < j_t$ to be such a sequence of indices, where $t \geq \frac{|J|}{\Delta} \geq \frac{m/2}{\lceil m/k \rceil} \geq \frac{k}{4}$. For $\ell \in [t]$, let E_ℓ be the event that row j_ℓ is spanned by the rows indexed by $I \cap [j_\ell - 1]$. Then,

$$\begin{aligned} \mathbf{Pr}\left[\text{Eq. (3.2) holds for } I\right] &\leq \mathbf{Pr}[E_1, E_2, \ldots, E_t] \\ &= \mathbf{Pr}[E_1] \cdots \mathbf{Pr}[E_t | E_1, \ldots, E_{t-1}] \end{aligned} \tag{3.3}$$

[7] One can construct such a set greedily: choose the minimal index j in J, remove all indices in $J \cap [j, j + \Delta - 1]$. Repeat until J is empty.

Next, we show that for each $\ell \in [t]$, we have $\mathbf{Pr}[E_\ell | E_1, \ldots, E_{\ell-1}] \leq 2^{-m/2}$. However, instead of conditioning on $E_1, \ldots, E_{\ell-1}$, we shall condition on a set of the random bits, to be specified next, that determine rows $B_1, \ldots, B_{j_{\ell-1}}$ on one hand, but are independent from the random row B_{j_ℓ} on the other hand. Since $j_\ell \geq j_{\ell-1} + \lceil m/k \rceil$ by our design, we get $(j_\ell - 1)k \geq (j_{\ell-1} - 1)k + m$. Hence, the random bits $a_1, \ldots, a_{(j_\ell-1)k}$ determine $B_1, \ldots, B_{j_{\ell-1}}$, and leave the random row $B_{j_\ell} = (a_{(j_\ell-1)k+1}, \ldots, a_{(j_\ell-1)k+m})$ totally undetermined. Conditioning on the worst-case assignment for the former random variables (under which $E_1, \ldots, E_{\ell-1}$ holds) yields an upper bound on $\mathbf{Pr}[E_\ell | E_1, \ldots, E_{\ell-1}]$. Thus, it is enough to show that $\mathbf{Pr}[E_\ell | a_1, \ldots, a_{(j_\ell-1)k}] \leq 2^{-m/2}$ for any possible fixed choice of values to $a_1, \ldots, a_{(j_\ell-1)k}$.

To avoid multiple subscripts, we set for the rest of the proof $j \triangleq j_\ell$. Let us remark that after fixing $a_1, \ldots, a_{(j-1)k}$, rows $1, \ldots, j - \lceil m/k \rceil$ are completely fixed, rows $j - \lceil m/k \rceil + 1, \ldots, j - 1$ are partially fixed, and row j is entirely undetermined. Based on that, we shall show that

$$\mathbf{Pr}[E_\ell | a_1, \ldots, a_{(j-1)k}] \leq 2^{-m/2} \,. \tag{3.4}$$

Let $I' := I \cap [j-1]$, and fix a linear combination of the rows indexed by I', i.e., $\sum_{i \in I'} c_i B_i$, among all $2^{|I'|}$ such linear combinations. We show that the probability that

$$B_j = \sum_{i \in I'} c_i B_i \tag{3.5}$$

is 2^{-m}. (This is similar, up to minor differences, to the folklore result that any fixed linear combination of rows in a random Toeplitz matrix is distributed uniformly over $\mathbb{F}_2^m$ – see Goldreich (2008, Prop. 8.25). We give the details for completeness.) The probability that the first bit of B_j equals the first bit of the linear combination in (3.5) is exactly $1/2$, since $B_{j,1} = S_{j,1} + a_{(j-1)k+1}$, and all entries $\{B_{i,1}\}_{i \in I'}$ involve only bits from $a_1, \ldots, a_{(j-2)k+1}$, which were already fixed (since $(j-2)k + 1 \leq (j-1)k$). Fixing $a_{(j-1)k+1}$ such that equality on the first bit holds, the second bit $B_{j,2}$ equals the resulting linear combination with probability $1/2$ as well. This happens since $B_{j,2}$ equals $S_{j,2} + a_{(j-1)k+2}$, where

$a_{(j-1)k+2}$ wasn't already fixed, and all entries $\{B_{i,2}\}_{i\in I'}$ involve only bits from $a_2,\ldots,a_{(j-2)k+2}$, which were already fixed (since $(j-2)k+2 \leq (j-1)k+1$). And so on, every bit in the jth row of B equals the resulting linear combination with probability $1/2$, conditioned on the fixing of the previous bits. Overall, $B_j = \sum_{i\in I'} c_i B_i$ with probability 2^{-m} for a fixed choice of coefficients $\{c_i\}_{i\in I'}$.[8] Taking a union bound over all possible coefficients $\{c_i\}_{i\in I'}$ gives $\mathbf{Pr}[E_\ell|E_1,\ldots,E_{\ell-1}] \leq 2^{|I'|}\cdot 2^{-m} \leq 2^{-m/2}$. Plugging this bound into Eq. (3.3) we get

$$\begin{aligned}\mathbf{Pr}\left[\text{Eq. (3.2) holds for } I\right] &\leq \mathbf{Pr}[E_1]\cdots\mathbf{Pr}[E_t|E_1,\ldots,E_{t-1}]\\ &\leq \left(2^{-m/2}\right)^t \leq 2^{-mk/8}\,,\end{aligned}$$

where in the last inequality we used $t \geq k/4$. Taking a union bound over all possible sequences I of length at most $m/2$, whose number is definitely less than 2^m, and using $k \geq 16$, we get

$$\mathbf{Pr}[\mathrm{rank}(S+A) \leq m/2] \leq 2^m \cdot 2^{-mk/8} \leq 2^{-mk/16}\,. \qquad \square$$

We continue with our main theorem.

THEOREM 3.6 (Random Hankel matrices are rigid). *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Hankel matrix $A_{i,j} = a_{i+j}$ where $a_2,\ldots,a_{2n}$ are uniform independent random bits. Then, for every $\sqrt{n} \leq r \leq n/32$, with probability $1-o(1)$, the matrix A has rigidity $\frac{n^3}{160r^2\log(960r^2/n)}$ for rank r.*

Before proving Theorem 3.6, we state an immediate corollary of it.

COROLLARY 3.7. *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Hankel matrix. Then, there exists a universal constant $c>0$ such that for every $\varepsilon > 0$*

(i) With probability $1-o(1)$, the matrix A has rigidity cn^2 for rank $\sqrt{n}$.

[8] Alternatively, conditioned on $a_1,\ldots,a_{(j-1)k}$ and the choice of the linear combination, there exist exactly one choice for $a_{(j-1)k+1},\ldots,a_{(j-1)k+m}$ that satisfies Eq. (3.5).

(ii) With probability $1 - o(1)$*, the matrix* A *has rigidity* $cn^{2-2\varepsilon}/\log n$ *for rank* $n^{1/2+\varepsilon}$*.*

(iii) With probability $1 - o(1)$*, the matrix* A *has rigidity* m^3 *for rank* $m = c \cdot \frac{n^{3/5}}{\log^{1/5} n}$*.*

(iv) With probability $1 - o(1)$*, the matrix* A *has rigidity* $cn^{1+2\varepsilon}/\log n$ *for rank* $n^{1-\varepsilon}$*.*

Proof of Theorem 3.6. Suppose towards contradiction that A can be represented as a sum of a matrix R of rank at most r, and an s-sparse matrix S, where $s \leq n^3/(160r^2 \log(960r^2/n))$. Let $m = 2r$, and assume for convenience that $k = n/m$ is an integer. Consider the following partition of A's entries into $(n/m)^2$ submatrices, each of dimension $m \times m$. For $i \in [n/m]$ and $j \in [n/m]$, let

$$\begin{aligned} I_i &= \{i, i+k, \ldots, i+(m-1)k\} \\ J_j &= \{(j-1)m+1, (j-1)m+2, \ldots, jm\}\,. \end{aligned} \tag{3.8}$$

Denote by $A^{i,j}$ ($R^{i,j}$, $S^{i,j}$, resp.) the matrix A (R, S, resp.) restricted to rows I_i and columns J_j. See Figure 3.1 for an example of such a submatrix. The main observation is that for each $(i,j) \in [n/m]^2$, the matrix $A^{i,j}$ is of the form needed by the main lemma. Another observation is that since the submatrices $S^{i,j}$ partitions the sparse matrix S, one of them has sparsity at most

$$\begin{pmatrix}
\boxed{a_2} & \boxed{a_3} & \boxed{a_4} & \boxed{a_5} & a_6 & a_7 & a_8 & a_9 \\
a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} \\
\boxed{a_4} & \boxed{a_5} & \boxed{a_6} & \boxed{a_7} & a_8 & a_9 & a_{10} & a_{11} \\
a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} \\
\boxed{a_6} & \boxed{a_7} & \boxed{a_8} & \boxed{a_9} & a_{10} & a_{11} & a_{12} & a_{13} \\
a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} \\
\boxed{a_8} & \boxed{a_9} & \boxed{a_{10}} & \boxed{a_{11}} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16}
\end{pmatrix}$$

Figure 3.1: A submatrix $A^{1,1}$ of the matrix A, for $m = 4$ and $k = 2$.

$s' \triangleq \frac{s \cdot m^2}{n^2}$. In addition, since rank of a submatrix may only decrease, for every i, j, it holds that $\mathrm{rank}(R^{i,j}) \leq \mathrm{rank}(R) \leq r$.

We say that $A^{i,j}$ is *simple* if it can be represented as a sum of an s'-sparse matrix and a matrix of rank at most r. By the above discussion, A can be represented as $S + R$ where S is s-sparse and R is of rank at most r, only if there exists a submatrix $A^{i,j}$ that is simple. We shall show that the latter occurs with very low probability:

(Union Bound)
$$\mathbf{Pr}\left[\exists i,j \ : \ A^{i,j} \text{ is simple}\right] \leq \sum_{i,j} \mathbf{Pr}[A^{i,j} \text{ is simple}]$$

(Union Bound)
$$\leq \sum_{i,j} \sum_{\substack{S \in \mathbb{F}_2^{m \times m}: \\ \mathtt{wt}(S) \leq s'}} \mathbf{Pr}[\mathrm{rank}(A^{i,j} + S) \leq \tfrac{m}{2}]$$

(Lemma 3.1)
$$\leq \binom{n}{m}^2 \cdot \binom{m^2}{\leq s'} \cdot 2^{-mk/16}$$

($n = km$ and (2.1))
$$< n^2 \cdot (6m^2/s')^{s'} \cdot 2^{-n/16} \ .$$

Using $s' \leq \frac{n}{40 \log(240m^2/n)}$, which follows from $s \leq \frac{n^3}{160r^2 \log(960r^2/n)}$, we get that

$$\begin{aligned}
&\mathbf{Pr}\left[\exists i,j \ : \ A^{i,j} \text{ is simple}\right] \\
&< n^2 \cdot \left(6m^2 \cdot \frac{40 \log(240m^2/n)}{n}\right)^{n/40\log(240m^2/n)} \cdot 2^{-n/16} \\
&= n^2 \cdot \left((240m^2/n) \cdot \log(240m^2/n)\right)^{n/40\log(240m^2/n)} \cdot 2^{-n/16} \\
&\leq n^2 \cdot \left((240m^2/n)^2\right)^{n/40\log(240m^2/n)} \cdot 2^{-n/16} \\
&= n^2 \cdot 2^{n/20} \cdot 2^{-n/16} \ = \ o(1).
\end{aligned}$$

□

Note that the proof works as long as the number of possibilities for an s'-sparse matrix $S^{i,j}$ is smaller than $2^{n/16}/n^2$. Our next theorem exploits the fact that there is a smaller number of possibilities for submatrices of *structured* sparse matrices (as in Definition 2.2). In fact, this is the only property of S that the foregoing proof uses. This yields the following improved bound.

THEOREM 3.9 (Random Hankel matrices are structured rigid). *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Hankel matrix. Then, for every $\sqrt{n} \le r \le n/32$, and $s \le n^3/1000r^2$, with probability $1-o(1)$, the matrix A has structured rigidity s for rank r.*

Before proving Theorem 3.9 we state three corollaries of it. The first corollary is immediate by choosing $r = n^{3/5}$.

COROLLARY 3.10. *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Hankel matrix. Then, there exists a universal constant $c > 0$ such that with probability $1-o(1)$, the matrix A has structured rigidity $cn^{9/5}$ for rank $n^{3/5}$.*

The second corollary follows from the first corollary and Theorem 2.5.

COROLLARY 3.11. *Let $A \in \mathbb{F}_2^{n\times n}$ be a random Hankel matrix, and let $F(x,y) = \sum_{i=1}^n \sum_{j=1}^n A_{i,j}x_iy_j$. Then, with probability $1-o(1)$, it holds that $\mathtt{C}(F) = \Omega(n^{3/5})$.*

The last corollary shows that there exists an explicit trilinear form with AN-complexity $\Omega(n^{3/5})$. This is the first improvement over the trivial $\Omega(\sqrt{n})$ lower bound for explicit tensors, and in doing so it solves Problem 4.2 from Goldreich & Wigderson (2013) in the affirmative. Goldreich & Wigderson (2013, Prop. 4.6) show that if some Toeplitz matrix has AN-complexity $\Omega(m)$, then F_{tet} defined in Eq. (1.3) has AN-complexity $\Omega(m)$ has well. We follow their method, but present a simpler argument for a different trilinear function.

COROLLARY 3.12. *Let $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{2n} \to \{0,1\}$ be the trilinear function defined by $F(x,y,z) = \sum_{i=1}^n \sum_{j=1}^n z_{i+j}x_iy_j$. Then, $\mathtt{C}(F) = \Omega(n^{3/5})$.*

PROOF. According to Corollary 3.11, there exists a Hankel matrix A, defined by some diagonal values $a_2, \ldots, a_{2n}$, such that the bilinear form $\sum_{i,j} a_{i+j}x_iy_j$ has AN-complexity $\Omega(n^{3/5})$.

Let $\mathcal{C}$ be a trilinear circuit computing F with minimal AN-complexity, and denote its complexity by m. Fixing the values of

the variables z_i to a_i, for all $i \in \{2, \ldots, 2n\}$, we get a bilinear circuit in x and y of AN-complexity at most m. Thus, $m = \Omega(n^{3/5})$. □

We return to prove Theorem 3.9.

Proof of Theorem 3.9. The proof follows the lines of the proof of Theorem 3.6. We let $m = 2r$, $k = n/m$, and $t = s^{1/3}$. We assume towards contradiction that $A = S + R$, where R is of rank at most r, and S is a sum of t matrices $S_1, \ldots, S_t \in \mathbb{F}_2^{n \times n}$, such that the ones in each matrix S_ℓ are a subset of some $X_\ell \times Y_\ell$, where $|X_\ell|, |Y_\ell| \leq t$. Denote by T the n-by-n matrix over $\mathbb{F}_2$ with $T_{i,j} = 1$ iff (i, j) is contained in at least one $X_\ell \times Y_\ell$. It is clear from T's definition that the ones in S are a subset of the ones in T. As in Theorem 3.6, we partition A, R, S, and also T, to $(n/m)^2$ submatrices, according to the partition of row indices $I_1, \ldots, I_{n/m}$ and column indices $J_1, \ldots, J_{n/m}$, defined as in the proof of Theorem 3.6 (see Eq. (3.8)). For a random $(i, j) \in [n/m]^2$, it holds that

$$\begin{aligned}
\mathop{\mathbf{E}}_{i,j}\left[\mathtt{wt}(T^{i,j})\right] &\leq t^3 \cdot \frac{m^2}{n^2}, \\
\mathop{\mathbf{E}}_{i,j}\left[\sum_{\ell=1}^{t} |X_\ell \cap I_i|\right] &\leq t \cdot t \cdot \frac{m}{n}, \\
\mathop{\mathbf{E}}_{i,j}\left[\sum_{\ell=1}^{t} |Y_\ell \cap J_j|\right] &\leq t \cdot t \cdot \frac{m}{n}\,.
\end{aligned}$$

We say that a submatrix $T^{i,j}$ is *good* if

$$\begin{aligned}
\mathtt{wt}(T^{i,j}) &\leq 4t^3 \cdot \frac{m^2}{n^2}, \\
\sum_{\ell=1}^{t} |X_\ell \cap I_i| &\leq 4t \cdot t \cdot \frac{m}{n}, \\
\sum_{\ell=1}^{t} |Y_\ell \cap J_j| &\leq 4t \cdot t \cdot \frac{m}{n}\,.
\end{aligned}$$

Using Markov's inequality, each of the above three events happen with probability at least 3/4. Using union bound (on the

complement events) with probability at least 1/4 all events occur simultaneously, making $T^{i,j}$ good.

Next, we count the number of possible good submatrices $T^{i,j}$. Each such submatrix is uniquely determined by the sets $X'_1, \ldots, X'_t$ and $Y'_1, \ldots, Y'_t$, where $X'_\ell = X_\ell \cap I_i$ and $Y'_\ell = Y_\ell \cap J_j$. Furthermore, a collection $(X'_1, ..., X'_t)$ such that $\sum_\ell |X'_\ell| \leq \frac{4t^2m}{n}$ corresponds to a set $X' \subseteq I_i \times [t]$ of size at most $\frac{4t^2m}{n}$ such that $(p, \ell) \in X'$ iff $p \in X'_\ell$ (and similarly for $(Y'_1, \ldots, Y'_t)$). Hence, the number of possible good submatrices is at most

$$\left|\left\{X' \subseteq I_i \times [t] : |X'| \leq \frac{4t^2m}{n}\right\}\right|^2 = \binom{mt}{\leq 4t^2m/n}^2 \leq \left((2mt)^{4t^2m/n}\right)^2 \leq n^{16t^2m/n} .$$

We say that $S^{i,j}$ is *good* if $T^{i,j}$ is good, and we say that $A^{i,j}$ is *simple* if it is the sum of a good $S^{i,j}$ and a matrix of rank at most r. Next, we count the number of possible good submatrices $S^{i,j}$. Since the ones of $S^{i,j}$ are a subset of the ones in $T^{i,j}$, the number of possibilities for $S^{i,j}$ is at most

$$n^{16t^2m/n} \cdot 2^{\mathtt{wt}(T^{i,j})} 2^{4t^3m^2/n^2} .$$

Using the bound on the number of possible good submatrices $S^{i,j}$, we may bound the probability that some $A^{i,j}$ is simple:

(Union Bound)
$$\mathbf{Pr}\left[\exists i,j \ : \ A^{i,j} \text{ is simple}\right] \leq \sum_{i,j} \sum_{S^{i,j} \text{ good}} \mathbf{Pr}[\mathrm{rank}(A^{i,j} + S^{i,j}) \leq \tfrac{m}{2}]$$

(Lemma 3.1)
$$\leq \left(\frac{n}{m}\right)^2 \cdot n^{16t^2m/n} \cdot 2^{4t^3m^2/n^2} \cdot 2^{-mk/16}$$

Recall that $m = 2r$ and $k = n/m$ to get

$$\mathbf{Pr}\left[\exists i,j \ : \ A^{i,j} \text{ is simple}\right] \leq 2^{2\log n + 32\log n \cdot t^2 r/n + 16t^3r^2/n^2 - n/16} ,$$

which is $o(1)$ for $t^3 \leq \frac{n^3}{1000r^2}$ and $r \geq \sqrt{n}$. □

Generalization to Larger Fields. The choice of field $\mathbb{F}_2$ was not crucial in the proofs of Lemma 3.1, Theorem 3.6 and Theorem 3.9. One can syntactically replace the field size 2 by any prime power q, keeping the proofs intact. Furthermore, in Theorem 3.6, we slightly benefit from taking a larger field. For details see Appendix A.1.

4. The structure of matrices of small bilinear circuits

In this section we shall further refine the structure of matrices associated with small bilinear circuits, beyond the structure captured by Definition 2.2 and Theorem 2.5. We begin by explicitly stating structural results that are implicit in the proof of Goldreich & Wigderson (2013, Thm. 4.4): Section 4.1 refers to the structure of bilinear functions that are computed by depth-2 bilinear circuits of small AN-complexity, whereas Section 4.2 refers to general bilinear circuits. These statements can be viewed as relating AN-complexity to finer notions of structured rigidity (than the one of Definition 2.2). In Section 4.3 we go beyond Goldreich & Wigderson (2013), and analyze the structure of the submatrices of matrices associated with small bilinear circuits, by starting with the foregoing structural results (of Goldreich & Wigderson (2013)) and proceeding analogously to the first part of the proof of Theorem 3.9. The results of Section 4.3 will play a pivotal role in the improved lower bounds proved in Section 5.

4.1. The structure of matrices associated with depth two bilinear circuits.

We say a row/column in a matrix is m-sparse if it contains at most m non-zero entries. Likewise, a linear function $\ell(x)$ (resp. $\ell'(y)$) is m-sparse if it depends on at most m entries in x (resp. y). Lastly, recall that by Definition 2.4, $\mathtt{C}_2(F)$ is the minimal AN-complexity of a depth-two bilinear circuit computing F.

Proposition 4.1 (Structure of functions computed by depth two bilinear circuits (Goldreich & Wigderson 2013, Thm. 4.4)). *Let F be a bilinear function over $x, y \in \{0,1\}^n$ with $\mathtt{C}_2(F) \leq m$. Then,*

F can be expressed as

$$\sum_{(i,j)\in P} L_i(x)L'_j(y) + \sum_{\ell=1}^{m} Q_\ell(x,y) \tag{4.2}$$

where P is a subset of $[m]\times[m]$, $L_1,\ldots,L_m$ and $L'_1,\ldots,L'_m$ are m-sparse linear functions, and each Q_ℓ is a bilinear function of at most m variables from x and at most m variables from y. The matrix associated with F has the form

$$A = L_{\text{col}}\cdot P\cdot L_{\text{row}} + \sum_{\ell=1}^{m} S_\ell \tag{4.3}$$

where L_{col} is an $n\times m$ matrix with m-sparse columns, P is a general $m\times m$ matrix, L_{row} is an $m\times n$ matrix with m-sparse rows, and each S_ℓ is an $n\times n$ matrix whose ones reside in an $m\times m$ rectangle.

Proposition 4.1 is proved explicitly in the warm-up part of the proof of Goldreich & Wigderson (2013, Thm. 4.4). The following proposition asserts that the converse holds as well. This implies that the characterization of Proposition 4.1 captures $\mathtt{C}_2$ completely.

Proposition 4.4. *Any bilinear form F that can be written as in Eq. (4.2), has $\mathtt{C}_2(F) = O(m)$.*

We defer the proof of this proposition to Appendix A.3.

4.2. The structure of matrices associated with general bilinear circuits

Proposition 4.5 (Structure of functions computed by general bilinear circuits). *Let F be a bilinear function over $x,y\in\{0,1\}^n$ with $\mathtt{C}(F)\le m$. Then, F can be expressed as*

$$\sum_{i=1}^{m} L_i(x)L'_i(y) + \sum_{i=1}^{m} M'_i(x)M_i(y) + \sum_{\ell=1}^{m} Q_\ell(x,y) \tag{4.6}$$

where $L_1,\ldots,L_m$ and $M_1,\ldots,M_m$ are m-sparse linear functions, $L'_1,\ldots,L'_m$ and $M'_1,\ldots,M'_m$ are general linear functions, and each

Q_ℓ is a bilinear function of at most m variables from x and at most m variables from y. The matrix associated with F has the form

$$A = L_{\text{col}} B + C L_{\text{row}} + \sum_{\ell=1}^{m} S_\ell \tag{4.7}$$

where L_{col} is an $n \times m$ matrix with m-sparse columns, B is a general $m \times n$ matrix, C is a general $n \times m$ matrix, L_{row} is an $m \times n$ matrix with m-sparse rows, and each S_ℓ is an $n \times n$ matrix whose ones reside in an $m \times m$ rectangle.

Proposition 4.5 is only implicit in the proof of Goldreich & Wigderson (2013, Thm. 4.4), and we include its proof in Appendix A.2. The following proposition asserts that the converse holds as well for $m \geq \sqrt{n}$. This implies that the characterization of Proposition 4.5 captures $\mathtt{C}$.

Proposition 4.8. *Any bilinear form F that can be written as in Eq. (4.6), has $\mathtt{C}(F) = O(m + \sqrt{n})$.*

We defer the proof of this proposition to Appendix A.3.

4.3. Substructures. In this subsection, similarly to the first part of the proof of Theorem 3.9, we find a structured submatrix of the matrix associated with any bilinear function with low AN-complexity. In Section 5, we prove that random Toeplitz matrices and small-biased matrices do not have these structured submatrices, with high probability. This ultimately proves AN-complexity lower bounds for such random matrices.

Let F be a bilinear function over $x, y \in \{0,1\}^n$ with $\mathtt{C}(F) \leq m$. Starting with Proposition 4.5, we write the matrix A associated with F as

$$A = L_{\text{col}} B + C L_{\text{row}} + \sum_{\ell=1}^{m} S_\ell$$

such that the non-zero entries of S_ℓ are a subset of $X_\ell \times Y_\ell$, where $|X_\ell|, |Y_\ell| \leq m$. Denote by $T = \bigcup_{\ell=1}^{m} X_\ell \times Y_\ell$, and note that $|T| \leq m^3$.

Let $I_1, \ldots, I_{n/2m}$ and $J_1, \ldots, J_{n/2m}$ be some fixed equipartition of the row indices and column indices of A, respectively, where each

I_i and J_j is of size $2m$. This partition naturally defines $(n/2m)^2$ submatrices as follows. For any (i,j) we denote by $A^{i,j}$ (resp. $S_\ell^{i,j}$) the matrix A (resp. S_ℓ) restricted to rows I_i and columns J_j. For any i (resp. j) we denote by L^i_{col} and C^i (resp. B^j and L^j_{row}) the matrices L_{col} and C (resp,. B and L_{row}) restricted to I_i (resp. J_j). Then, one can write

$$A^{i,j} = L^i_{\text{col}} B^j + C^i L^j_{\text{row}} + \sum_{\ell=1}^{m} S_\ell^{i,j} \,, \tag{4.9}$$

where $S_\ell^{i,j} \subseteq T \cap (I_i \times J_j)$. Next, we show that there exists a choice of (i,j) with favorable properties of the submatrices of $L^i_{\text{col}}, L^j_{\text{row}}$ and of the subsets $\{X_\ell \cap I_i\}_\ell$, $\{Y_\ell \cap J_j\}_\ell$, and $T \cap (I_i \times J_j)$.

Proposition 4.10. (Structure of submatrix of matrices associated with small bilinear circuits). *For every $\ell \in [m]$, let $S_\ell \subseteq X_\ell \times Y_\ell$, where $|X_\ell|, |Y_\ell| \leq m$, and $T = \bigcup_{\ell=1}^m X_\ell \times Y_\ell$. Let $I_1, \ldots, I_{n/2m} \subseteq [n]$ and $J_1, \ldots, J_{n/2m} \subseteq [n]$ be two partitions of $[n]$ where each I_i and J_j is of size $2m$. Let $A^{i,j}$, L^i_{col} and L^j_{row} be as in (4.9). Then, there exists an $(i,j) \in [n/2m]^2$ such that: (1) $|T \cap (I_i \times J_j)| \leq \frac{24m^5}{n^2}$, (2) $\sum_{\ell=1}^m |X_\ell \cap I_i| \leq \frac{12m^3}{n}$, (3) $\sum_{\ell=1}^m |Y_\ell \cap J_j| \leq \frac{12m^3}{n}$, (4) $\mathtt{wt}(L^i_{\text{col}}) \leq \frac{12m^3}{n}$, and (5) $\mathtt{wt}(L^j_{\text{row}}) \leq \frac{12m^3}{n}$.*

If $\mathtt{C}_2(F) \leq m$, then starting with Proposition 4.1, we can express $A^{i,j}$ as $L^i_{\text{col}} \cdot P \cdot L^j_{\text{row}} + \sum_{\ell=1}^m S_\ell^{i,j}$ and Proposition 4.10 holds as well in this case.

Proof. For a uniformly random $(i,j) \in [n/2m]^2$, it holds that

$$\begin{aligned}
\mathop{\mathbf{E}}_{i,j}\left[|T \cap (I_i \times J_j)|\right] &\leq m^3 \cdot \tfrac{(2m)^2}{n^2} \\
\mathop{\mathbf{E}}_{i,j}\left[\sum_{\ell=1}^m |X_\ell \cap I_i|\right] &\leq m \cdot m \cdot \tfrac{2m}{n} \\
\mathop{\mathbf{E}}_{i,j}\left[\sum_{\ell=1}^m |Y_\ell \cap J_j|\right] &\leq m \cdot m \cdot \tfrac{2m}{n} \\
\mathop{\mathbf{E}}_{i,j}[\mathtt{wt}(L^i_{\text{col}})] &\leq m \cdot m \cdot \tfrac{2m}{n} \\
\mathop{\mathbf{E}}_{i,j}[\mathtt{wt}(L^j_{\text{row}})] &\leq m \cdot m \cdot \tfrac{2m}{n}
\end{aligned}$$

Using Markov's inequality, each of the following "bad" events occur with probability at most 1/6

$$\begin{aligned}
|T \cap (I_i \times J_j)| &\geq 6m^3 \cdot \tfrac{(2m)^2}{n^2} \\
\sum_{\ell=1}^{m} |X_\ell \cap I_i| &\geq 6m \cdot m \cdot \tfrac{2m}{n} \\
\sum_{\ell=1}^{m} |Y_\ell \cap J_j| &\geq 6m \cdot m \cdot \tfrac{2m}{n} \\
\mathtt{wt}(L^i_{\mathrm{col}}) &\geq 6m \cdot m \cdot \tfrac{2m}{n} \\
\mathtt{wt}(L^j_{\mathrm{row}}) &\geq 6m \cdot m \cdot \tfrac{2m}{n}
\end{aligned}$$

By union bound, with probability at least $1-5/6$ over the choice of (i,j), none of the "bad" events occur, which completes the proof. □

We wish to express the structure captured by Eq. (4.9) in terms of linear equations on the entries of the matrix $A^{i,j} - \sum_\ell S^{i,j}_\ell$. To do so we need the following definition.

Definition 4.11 (Orthogonal complement of a matrix). *Let $m \leq n$. If A is an $n \times m$ matrix, and B is a $(n-m) \times n$ matrix of rank $n-m$ such that $BA = 0$ then we say that B is a* left orthogonal complement *of A. If A is an $m \times n$ matrix, and B is a $n \times (n-m)$ matrix of rank $n-m$ such that $AB = 0$ then we say that B is a* right orthogonal complement *of A.*

Remark 4.12. *Note that there are many possible choices of an orthogonal complement of a given matrix. Therefore, we shall refer to* the left *(right, resp.)* orthogonal complement *of A as some canonical choice of a left (right, resp.) orthogonal complement of A, say the first such matrix according to lexicographical order (over a finite field $\mathbb{F}$).*

It is well known that any matrix over a field has an orthogonal complement. Now, suppose that $A^{i,j} - \sum_\ell S_\ell{}^{i,j} = L^i_{\mathrm{col}} B^j + C^i L^j_{\mathrm{row}}$ (as in Eq. (4.9)). Let D be an $m \times 2m$ matrix which is the left

orthogonal complement of L^i_{col}, and let E be a $2m \times m$ matrix which the right orthogonal complement of L^j_{row}. Then,

$$D \cdot (A^{i,j} - \sum_\ell {S_\ell}^{i,j}) \cdot E = 0^{m \times m} \ . \tag{4.13}$$

In the case of depth-2 circuits we have $A^{i,j} - \sum_\ell {S_\ell}^{i,j} = L^i_{\text{col}} \cdot P \cdot L^j_{\text{row}}$. Using E, the right orthogonal complement of L^j_{row} as above, we can write

$$(A^{i,j} - \sum_\ell {S_\ell}^{i,j}) \cdot E = 0^{2m \times m} \ . \tag{4.14}$$

In the next section, we shall design tests based on Equations (4.13) and (4.14).

5. Tests for AN complexity and AN2 complexity

Having identified (in Section 4.3) structural properties that are satisfied by any matrix associated with any bilinear function with low AN-complexity, we prove lower bounds on the AN-complexity of explicit distributions of matrices by showing that (with high probability) these distributions do not satisfy these properties. We do so by designing a test that always accepts matrices that have these properties, but rejects (with high probability) matrices drawn from certain explicit distributions. (Since the test is merely a mental experiment, i.e., we do not intend to actually run it, the test could be inefficient.) Specifically, for a complexity bound m, any matrix rejected by the corresponding test must have complexity greater than m. We will show that a random Toeplitz matrix, as well as a matrix whose entries are sampled from an 2^{-n}-biased distribution, are rejected by the corresponding test with overwhelming probability, thus proving complexity lower bounds for such matrices.

Actually, we will present two tests: One for AN-complexity, rejecting most matrices taken from a small-biased space, and one for AN2-complexity, rejecting most Toeplitz matrices, see Section 5.1 and 5.3, respectively. In Section 5.2 we show that the lower bound for matrices taken from a small-biased space yields a similar lower bound for an explicit 4-linear function.

5.1. Lower bounds for the AN-complexity of small-biased matrices. For $i \in [n/2m]$ and $j \in [n/2m]$, let[9]

(5.1)
$$\begin{aligned} I_i &= \{i, i+(n/2m), \ldots, i+(2m-1)\cdot(n/2m)\}\,, \\ J_j &= \{(j-1)\cdot(2m)+1, (j-1)\cdot(2m)+2, \ldots, j\cdot(2m)\}\,. \end{aligned}$$

and denote by $A^{i,j}$ the $2m$-by-$2m$ sub-matrix of A obtained by restricting A to rows I_i and columns J_j. Consider the following test, where $A^{i,j}$ is viewed as indexed by $[2m]\times[2m]$ rather than by $I_i \times J_j$.

Test 1 AN-Complexity Test

Input: Matrix $A \in \mathbb{F}_2^{n\times n}$ and parameter $m \in [n]$

for $i = 1, \ldots, n/2m$ and $j = 1, \ldots, n/2m$ **do**
  for all subsets $\{X_\ell^i\}_{\ell=1}^m$ of $[2m]$ with $\sum_\ell |X_\ell^i| \leq \frac{12m^3}{n}$ **do**
    for all subsets $\{Y_\ell^j\}_{\ell=1}^m$ of $[2m]$ with $\sum_\ell |Y_\ell^j| \leq \frac{12m^3}{n}$ **do**
      Let $T := \bigcup_{\ell=1}^m X_\ell^i \times Y_\ell^j$.
      if $|T| \leq \frac{24m^5}{n^2}$ **then**
        for all matrices L_{col}^i of dimension $2m\times m$ and sparsity at most $\frac{12m^3}{n}$ **do**
          Let D be the left orthogonal complement of L_{col}^i (recall Remark 4.12).
          for all matrices L_{row}^j of dimension $m \times 2m$ and sparsity at most $\frac{12m^3}{n}$ **do**
            Let E be the right orthogonal complement of L_{row}^j.
            if there exists $N \in \mathbb{F}_2^{2m\times 2m}$ such that $N \subseteq T$, and $D(A^{i,j} - N)E = 0^{m\times m}$ **then**
              return "Pass".
return "Fail".

The following is an immediate corollary of Proposition 4.10 and Eq. (4.13).

COROLLARY 5.2. *Every matrix associated with a bilinear circuit of AN-complexity at most m passes Test 1 with parameter m.*

[9] The specific choice for I_i and J_j is not crucial for our argument in this subsection, however it will be important in the next subsection. Hence, since we need to pick some partition, we might as well choose this one.

We consider a distribution of matrices whose entries are chosen from a small biased sample space. Specifically, we shall use a sample space over strings of length $N = n^2$ in order to define n-by-n matrices. We shall show that almost all such matrices are rejected by Test 1 with parameter m. But we need a few preliminaries first.

Preliminaries. Recall the definition of an ε-biased distribution from Naor & Naor (1993).

Definition 5.3 (Small-biased distribution). *A distribution X over $\{0,1\}^N$ is said to be ε-*biased *if for every non-empty set $S \subseteq [N]$, it holds that*

$$\left| \mathop{\mathbf{E}}_{x \sim X} [(-1)^{\sum_{i \in S} x_i}] \right| \leq \varepsilon.$$

We shall use the following property of ε-biased distributions (implicit in Naor & Naor (1993)).

Lemma 5.4 (Alon *et al.* 1992, Lem. 1). *Let X be an ε-biased distribution over $\{0,1\}^N$. Let $\ell_1, \ldots, \ell_t$ be linearly independent linear functions on $x_1, \ldots, x_N$. Then, the probability that all linear functions evaluate to 0 on $x \sim X$ is at most $\varepsilon + 2^{-t}$. Then, the probability that all linear functions equal 0 simultaneously is at most $\varepsilon + 2^{-t}$.*

We shall also use the following simple fact from linear algebra.

Fact 5.5. *Let $t, n, m \in \mathbb{N}$ such that $t \leq m \leq n$. Let $\ell_1, \ldots, \ell_t$ be a sequence of linearly independent linear functions (over $\mathbb{F}$) on $x_1, \ldots, x_n$. Then, $\ell_1, \ldots, \ell_t$ span at least $t - m$ linearly independent functions that involve only the variables $x_{m+1}, \ldots, x_n$.*

Proof. Think of the linear functions as vectors in $\mathbb{F}^n$, and let $V = \mathrm{span}\{\ell_1, \ldots, \ell_t\}$. Consider the subspace $U = \mathrm{span}\{e_{m+1}, \ldots, e_n\}$, where $e_i \in \mathbb{F}^n$ is the unit vector with 1 in the ith coordinate and 0 elsewhere. Then, $\dim(U \cap V) \geq \dim(U) + \dim(V) - n = (n - m) + t - n = t - m$, whereas $U \cap V$ is the span of $\ell_1, \ldots, \ell_t$ that is supported only on the last $n - m$ coordinates. □

Actual Results. We are now ready to analyze the probability that a matrix sampled from a small biased space passes Test 1. The core of the analysis refers to a single application of Step 10, which refers to a specific choice of $i, j, \{X^i_\ell\}_{\ell=1}^m, \{Y^j_\ell\}_{\ell=1}^m$ as well as $L^i_{\text{col}}, L^j_{\text{row}}$ (which in turn, fixes D and E as well).

LEMMA 5.6 (Core of the analysis of Test 1). *Fix $i, j, \{X^i_\ell\}_{\ell=1}^m$ and $\{Y^j_\ell\}_{\ell=1}^m$ that pass the check of Step 5, and fix L^i_{col} and L^j_{row} (which in turn, fixes D and E as well). Then, a matrix A whose entries are sampled from an ε-biased distribution satisfies the condition in Step 10 with probability at most*

$$\varepsilon + 2^{-m^2+24m^5/n^2} .$$

PROOF. For a fixed choice of $i, j, \{X^i_\ell\}_{\ell=1}^m, \{Y^j_\ell\}_{\ell=1}^m, L^i_{\text{col}}$ and L^j_{row} as above, we consider a specific submatrix of dimension $2m \times 2m$ of A, denoted $A^{i,j}$. Note that the corresponding left (resp. right) orthogonal complement of L^i_{col} (resp. L^j_{row}) is a m-by-$2m$ (resp. $2m$-by-m) matrix of rank m, denoted by D (resp. E). Recall that $A^{i,j}$ is a submatrix whose entries are sampled according to an ε-biased distribution. Our goal is to show that the equation $D(A^{i,j} - N)E = 0$ (checked in Step 10) implies a lot of linearly independent linear equations on the entries of $A^{i,j}$.

Let Z be a $2m \times 2m$ matrix of $(2m)^2$ Boolean variables, where we will later take Z to be $A^{i,j} - N$. Interpret the equations $DZE = 0^{m\times m}$ as m^2 linear equations on the $(2m)^2$ variables in Z. For $i \in [m]$ and $j \in [m]$, we have an equation of the form $D_i Z E^{(j)} = 0$, where D_i is the ith row of D and $E^{(j)}$ is the jth column of E. We can write

$$D_i Z E^{(j)} = \sum_{k=1}^{2m}\sum_{\ell=1}^{2m} D_{i,k} Z_{k,\ell} E_{\ell,j} = \sum_{k,\ell} (D_i \otimes E^{(j)})_{k,\ell} Z_{k,\ell};$$

that is, the coefficients of the equation are the tensor product of the vector D_i with the vector $E^{(j)}$. Thinking of these m^2 linear equations on $(2m)^2$ variables as a big matrix of dimension $m^2 \times (2m)^2$, we note that this matrix of linear equations is the tensor product of D and $E^\top$, since the (i, j) row equals to $D_i \otimes E^{(j)}$ (viewed as a $(2m)^2$-bit long vector).

It is a known fact that the rank of the tensor product of any two matrices is the product of their rank; hence, we get $\operatorname{rank}(D \otimes E^\top) = \operatorname{rank}(D) \cdot \operatorname{rank}(E^\top) = m^2$. In other words, we have a linearly independent set of m^2 linear equations on the variables Z. However, we want to get linear equations over the variables of A, where $Z = A - N$. Say that $Z_{k,\ell}$ is a *noisy variable* if $(k,\ell) \in T$. It will be enough to show that there are many independent linear equations which involve only non-noisy variables of the matrix. Since the number of noisy variables is $|T|$, by Fact 5.5 we can find at least $m^2 - |T|$ independent linear equations that do not involve noisy variables.

Overall, we got $m^2 - |T|$ independent linear equations on $A^{i,j}$. By Lemma 5.4, a submatrix $A^{i,j}$ whose entries are sampled according to an ε-biased distribution satisfies all $m^2 - |T|$ equations with probability at most $\varepsilon + 2^{-m^2+|T|}$. Lastly, the fact that $\{X^i_\ell\}_{\ell=1}^m$ and $\{Y^j_\ell\}_{\ell=1}^m$ passed the check of Step 5 means that $|T| \le 24m^5/n^2$, which finishes the proof. □

THEOREM 5.7 (Almost all ε-biased matrices have high AN-complexity). *A matrix A whose entries are sampled from an ε biased distribution is rejected by Test 1 with parameter m (which implies that the corresponding bilinear function has AN-complexity greater than m), with probability at least*

$$1 - \left(\frac{n}{2m}\right)^2 \cdot \binom{2m^2}{\le 12m^3/n}^4 \cdot \left(\varepsilon + 2^{-m^2+24m^5/n^2}\right).$$

In particular, for $\varepsilon = 2^{-n}$ and $m = \frac{n^{2/3}}{10(\log n)^{1/3}}$, this probability is at least $1 - 2^{-n/2}$, for sufficiently large n.

PROOF. We use a union bound over all possible $i, j, \{X^i_\ell\}_{\ell=1}^m$, $\{Y^j_\ell\}_{\ell=1}^m, L^i_{\text{col}}$ and L^j_{row} that can be selected by the test, and employ Lemma 5.6 for each possibility. The number of options for choosing (i,j) is $(n/2m)^2$; the number of options for choosing $\{X^i_\ell\}_{\ell=1}^m$ (resp., $\{Y^j_\ell\}_{\ell=1}^m$) is at most $\binom{2m^2}{\le 12m^3/n}$; the number of options for choosing L^i_{col} (resp., L^j_{row}) is at most $\binom{2m^2}{\le 12m^3/n}$. □

5.2. Explicit 4-Linear Functions with AN-Complexity $\tilde{\Omega}(n^{2/3})$. We show that based on the ε-biased generator of Mossel *et al.* (2006) (described next), the AN-complexity lower bound for the randomized bilinear function in Theorem 5.7 yields a similar lower bound on an explicit 4-linear function.

To describe Mossel et al.'s construction, we begin with some preliminaries. Let N be a natural number, denote by $\mathbb{F} = GF(2^N)$, and suppose we have an explicit representation of $\mathbb{F}$ as the quotient $\mathbb{F}_2[x]/(p(x))$ where $p(x)$ is an irreducible polynomial over $\mathbb{F}_2$ of degree N. We remark that for $N = 2 \cdot 3^k$, the polynomial $p(x)$ may be chosen to be $x^{2 \cdot 3^k} + x^{3^k} + 1$ (cf. Lidl & Niederreiter 1997, Ex. 3.96).[10] Then, $1, x, x^2, \ldots, x^{N-1}$ is a basis for $GF(2^N)$ over $\mathbb{F}_2$. The map $\phi : \mathbb{F} \to \mathbb{F}$ defined by $\phi : z \mapsto z \cdot x$ is a linear transformation over $\mathbb{F}_2$, thus may be represented by a matrix $A \in \mathbb{F}_2^{N \times N}$. The Frobenius transformation $\varphi : \mathbb{F} \to \mathbb{F}$ defined by $\varphi : z \mapsto z^2$ is also a linear transformation over $\mathbb{F}_2$, thus may be represented by a matrix $B \in \mathbb{F}_2^{N \times N}$. Given the polynomial $p(x)$, the matrix A (resp. B) can be computed in $\mathrm{poly}(N)$ time by writing the images of the basis elements $\phi(1), \phi(x), \ldots, \phi(x^{N-1})$ (resp. $\varphi(1), \varphi(x), \ldots, \varphi(x^{N-1})$) as polynomials modulo $p(x)$.

The generator of Mossel et al. is given $2N$ input bits $c_1, \ldots, c_N$ and $d_1, \ldots, d_N$, and outputs $N \cdot n$ bits where $1 \leq n \leq N$, such that each output bit is a bilinear function in $c = (c_1, \ldots, c_N)$ and $d = (d_1, \ldots, d_N)$. The output of the generator on vectors $c, d \in \mathbb{F}_2^N$ is the $N \cdot n$ bits $g_{i,j} = c^\top (A^i B^j) d$ for $i \in \{0, \ldots, N-1\}$ and $j \in \{0, \ldots, n-1\}$. Mossel *et al.* (2006) proved that this is an ε-bias generator.

THEOREM 5.8 (Mossel *et al.* 2006, Thm. 6). *The bias of any non-trivial linear combination of the $g_{i,j}$s is at most $2^{-\frac{N-n}{2}}$.*

COROLLARY 5.9. *Let $N = 2 \cdot 3^k$ and let A and B be the explicit matrices as above for the field $GF(2^N)$. Let $n = N/3$ and let $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^N \times \{0,1\}^N$ be the 4-linear func-*

[10] For general N, it is not known how to find such a polynomial $p(x)$ without advice or randomness (Kopparty *et al.* 2014).

tion defined by $F(a,b,c,d) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} a_i b_j \cdot (c^T A^i B^j d)$. *Then,* $\mathtt{C}(F) = \Omega(n^{2/3}/\log^{1/3} n)$.

PROOF. For a fixed value of c and d, denote by $F_{c,d}$ the bilinear function defined by $F_{c,d}(a,b) = F(a,b,c,d)$. By Theorem 5.8 for a random c,d the random matrix $F_{c,d}$ is an n-by-n matrix whose entries are drawn from an ε-bias distribution, for $\varepsilon = 2^{-(N-n)/2} = 2^{-(3n-n)/2} = 2^{-n}$. By Theorem 5.7, this means that there exists a choice for c and d under which $F_{c,d}$ satisfies $\mathtt{C}(F_{c,d}) \geq \Omega(n^{2/3}/\log^{1/3} n)$, for a large enough n (in fact, at least $1 - 2^{-n/2}$ fraction of the choices have this property). By the fact that the AN-complexity of F is at least as large as the AN-complexity of $F_{c,d}$ (see the proof of Corollary 3.12), we get $\mathtt{C}(F) \geq \Omega(n^{2/3}/\log^{1/3} n)$. □

5.3. Lower bounds for the AN2-complexity of random Toeplitz matrices. The following is a degenerate version of Test 1. Recall the definition of I_i and J_j from Eq. (5.1), and the definition of $A^{i,j}$.

Test 2 AN-2-Complexity Test

Input: Matrix $A \in \mathbb{F}_2^{n\times n}$ and parameter $m \in [n]$

for $i = 1,\ldots,n/2m$ and $j = 1,\ldots,n/2m$ **do**
  for all subsets $\{X_\ell^i\}_{\ell=1}^m$ of $[2m]$ with $\sum_\ell |X_\ell^i| \leq \frac{12m^3}{n}$ **do**
    for all subsets $\{Y_\ell^j\}_{\ell=1}^m$ of $[2m]$ with $\sum_\ell |Y_\ell^j| \leq \frac{12m^3}{n}$ **do**
      Let $T := \bigcup_{\ell=1}^m X_\ell^i \times Y_\ell^j$.
      if $|T| \leq \frac{24m^5}{n^2}$ **then**
        for all matrices L_{row}^j of dimension $m \times 2m$ and sparsity at most $\frac{12m^3}{n}$ **do**
          Let E be the right orthogonal complement of L_{row}^j.
          if there exists $N \in \mathbb{F}_2^{2m\times 2m}$ such that $N \subseteq T$, and $(A^{i,j} - N)E = 0^{2m\times m}$ **then**
            return "Pass".
return "Fail".

The following is an immediate corollary of Proposition 4.10 and Eq. (4.14).

COROLLARY 5.10. *Every matrix associated with a bilinear circuit of AN2-complexity at most m passes Test 2 with parameter m.*

LEMMA 5.11 (Core of the analysis of Test 2). *Fix $i, j, \{X^i_\ell\}_{\ell=1}^m$ and $\{Y^j_\ell\}_{\ell=1}^m$ that pass the check of Step 5, and fix L^j_{row} (which in turn, fixes E as well). Then, a random Hankel matrix A satisfies the condition in Step 8 with probability at most*

$$2^{-n/2+6m^3/n}$$

PROOF. For a fixed choice of $i, j, \{X^i_\ell\}_{\ell=1}^m, \{Y^j_\ell\}_{\ell=1}^m$ and L^j_{row}, we consider a specific submatrix of dimension $2m \times 2m$ of A, denoted $A^{i,j}$. Note that the corresponding right orthogonal complement of L^j_{row} is a $2m$-by-m matrix of rank m, denoted by E. By the definition of I_i and J_j in Eq. (5.1), $A^{i,j}$ is of the form

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{2m} \\ a_{k+1} & a_{k+2} & a_{k+3} & \cdots & a_{k+2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{(2m-1)k+1} & a_{(2m-1)k+2} & a_{(2m-1)k+3} & \cdots & a_{(2m-1)k+2m} \end{pmatrix}$$

where $k = n/(2m)$ and $a_1, \ldots, a_{(2m-1)k+2m}$ are uniform independent random bits. Our goal will be to show that the equation $(A^{i,j} - N) \cdot E = 0^{2m \times m}$ implies a lot of linearly independent linear equations on the random variables $a_1, \ldots, a_{(2m-1)k+2m}$.

First think of a generic $2m \times 2m$ matrix Z as a matrix of $(2m)^2$ variables, and interpret the equations $ZE = 0^{2m \times m}$ as linear equations on Z. For each row $\ell \in [2m]$, we have m equations corresponding to $Z_\ell E = 0^{1 \times m}$, which are linearly independent. Denote by T_ℓ the intersection of T with the indices corresponding to the ℓth row of the submatrix, i.e. $T_\ell = T \cap (\{\ell\} \times [2m])$. Say that $Z_{\ell,\ell'}$ is a *noisy variable* if $(\ell, \ell') \in T$. By Fact 5.5, we can get at least $m - |T_\ell|$ independent linear equations on the ℓth row of Z that do not involve noisy variables. Summing over all ℓ's we have at least $\sum_{\ell=1}^{2m} (m - |T_\ell|) = 2m^2 - |T|$ independent linear equations that do not involve the noisy entries of the matrix, and such that each equation involves only variables from one row of Z. Take $Z = A^{i,j} - N$; since we got equations on Z that do not involve noisy entries, these are actually equations on $A^{i,j}$ as well.

The main difficulty is that we want to exhibit linearly independent linear equations on the variables $a_1, \ldots a_{(2m-1)k+2m}$, but the equations we got may not be linearly independent once we identify multiple entries in the matrix $A^{i,j}$ with the same variable.[11] To solve this issue, we shall look for a set of equations which remains linearly independent after this identification. Let $n_\ell = m - |T_\ell|$ be number of linearly independent equations we got on the ℓth row. Let $s = \lceil (2m)^2/n \rceil$, and consider all rows starting from some index $r \in [s]$, and taking jumps of s. Then, by the pigeon-hole principle there exists a $r \in [s]$ such that

$$\sum_{\ell : \ell \equiv r \mod s} n_\ell \geq (2m^2 - |T|)/s \; .$$

A key point is that by our choice of s, the ℓth row and the $(\ell + s)$th row of $A^{i,j}$ depend on disjoint sets of random variables, since $s \cdot k \geq \frac{(2m)^2}{n} \cdot \frac{n}{2m} = 2m$. Thus, the sets of variables out of $a_1, \ldots, a_{(2m-1)k+2m}$ that participate in rows with index in $\{\ell \; : \; \ell \equiv r \mod s\}$ are pairwise disjoint, and the equations we got on these rows are linearly independent as equations over the variables $a_1, \ldots, a_{(2m-1)k+2m}$. Since we got at least $(2m^2 - |T|)/s$ independent linear equations on completely random bits, all equations hold simultaneously with probability at most $2^{(-2m^2+|T|)/s}$. The fact that $\{X^i_\ell\}_{\ell=1}^m$ and $\{Y^j_\ell\}_{\ell=1}^m$ passed the check in Step 5 means that $|T| \leq 24m^5/n^2$, and using $s = 2m^2/n$, we get a probability bound of $2^{(-2m^2+24m^5/n^2)\cdot \frac{n}{4m^2}}$, which completes the proof. □

Theorem 5.12 (Almost all random Hankel matrices have high AN2-complexity). *A random Hankel matrix A is rejected by Test 2 with parameter m (which implies it has direct complexity at least m) with probability at least*

$$1 - \left(\frac{n}{2m}\right)^2 \cdot \binom{2m^2}{\leq 12m^3/n}^3 \cdot 2^{-n/2+6m^3/n} \; .$$

[11] In fact, we cannot expect this set of equations to be linearly independent simply because there are too many equations (i.e., more equations than variables).

In particular, for $m = \frac{n^{2/3}}{10(\log n)^{1/3}}$*, this probability is at least* $1 - 2^{-n/4}$*, for large enough* n*.*

PROOF. We use a union bound over all the $\binom{n}{2m}^2 \cdot \binom{2m^2}{\leq 12m^3/n}^3$ possible ways to pick $i, j, \{X^i_\ell\}_{\ell=1}^m, \{Y^j_\ell\}_{\ell=1}^m$ and L^j_{row}, and employ Lemma 5.11 to bound each possibility. □

Explicit 3-Linear Functions with $\mathtt{C}_2 = \tilde{\Omega}(n^{2/3})$. The following is a corollary of Theorem 5.12.

COROLLARY 5.13. *Let* $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{2n} \to \{0,1\}$ *be the trilinear function defined by* $F(x,y,z) = \sum_{i=1}^n \sum_{j=1}^n z_{i+j} x_i y_j$*. Then,* $\mathtt{C}_2(F) = \Omega(n^{2/3}/\log^{1/3} n)$*.*

We omit the proof, since it is identical to that of Corollary 3.12.

6. Digest and open problems

6.1. Digest. It is well known that random matrices have high rank; specifically, a random n-by-n matrix has rank at least $n/2$ with probability at least $1 - \exp(-\Omega(n^2))$. Our basic strategy is to obtain similar bounds for pseudorandom matrices, where the bound is "similar" in the sense that it is exponentially (in the randomness complexity) close to 1. We restrict ourselves to randomness complexity $O(n)$, since we aim at replacing the random bits by auxiliary inputs, making the construction explicit (to ensure multilinearity, we also use the fact that the sampled matrix is multilinear in the randomness). This means that we can afford a union bound over $\exp(n)$-many events.

The first instantiation of our strategy appears in the proof of Theorem 3.6, where we handle random Toeplitz matrices. Firstly, we consider a partition of the random n-by-n matrix into m-by-m matrices such that each submatrix depends on $\Theta(n)$ random bits. Next, using Lemma 3.1 and a union bound over all m-by-m matrices that are s'-sparse, we prove that (with high probability) all submatrices have rigidity $s' = \Omega(n/\log n)$ for rank $m/2$. It follows that a random Toeplitz matrix has rigidity $(n/m)^2 \cdot s' = \widetilde{\Omega}(n^3/m^2)$

for rank $m/2$, which yields new results for any $m \in [n^{0.51}, n^{0.99}]$. A slightly better result is obtained for structured rigidity, since in this case we may consider slightly less sparse matrices.

More radical savings appear in Section 5, where the structured rigidity is exploited much further. Here we are considering a number of bad events that exceeds $\exp(O(n))$, whereas our randomness complexity is still $O(n)$. This is done by "covering" these bad events by a "net" of $\exp(n)$ bad super-events and taking a union bound on the latter (partially explicitly and partially implicitly). Firstly, since we start with restricted notions of structured rigidity (which suffice for our application), we can upper bound the number of linear dependencies (in the $2m$-by-$2m$ submatrix) by $\exp(n)$ (rather than by $\exp(m^2)$). Secondly, relying on structured rigidity, we cover all relevant s''-sparse $2m$-by-$2m$ matrices by $\exp(n)$ such matrices, where $s'' = m^2/\operatorname{poly}\log(n)$ and T covers S if the non-zero entries of S are a subset of the non-zero entries of T. Finally, rather than considering the probability that some m^2 linear equations involving the elements of $R + S$ hold, where R is a pseudorandom matrix and S is a fixed matrix, we consider all matrices covered by some matrix T simultaneously. We do so by considering the probability that some $m^2 - s''$ related linear equations involving only the elements of R hold, where the latter equations are obtained by eliminating the variables corresponding to non-zero entries of T from the former equations. We stress that the final step accounts for more than $2^{s''}$ sparse matrices, whereas the amount of randomness is $O(n) \ll s''$.

6.2. Randomness-rigidity tradeoff. We would like to remark that it is straight-forward to derive other rigidity results if we allow different amounts of randomness. Naturally, the results gets better as the randomness increases. We quantify the number of random bits by ℓ, and towards applying Lemma 3.1 choose $m = 2r$ and $k = \ell/2m$, while assuming $\ell \geq 2m$ and $m \geq k$. Applying Lemma 3.1, we get that $\mathbf{Pr}[\operatorname{rank}(A + S) \leq r] \leq 2^{-\Omega(\ell)}$ where S is any fixed matrix and the probability is over the random bits of A (recall that A is a random Hankel-like matrix with k new bits in each row). The number of random bits used is $mk + m - k \leq \ell$ by our assumptions.

Next, we consider the $n \times n$ matrix consisting of $(n/m)^2$ copies of A. Any sparse $n \times n$ matrix with sparsity s will have an $m \times m$ submatrix S whose sparsity is at most $s' = sm^2/n^2$ and on this submatrix the rank of $A + S$ will be at most r with probability at most $\binom{m^2}{\leq s'} \cdot 2^{-\Omega(\ell)}$. Hence we can handle sparsity $s \leq c_0 \cdot \frac{n^2}{m^2 \log n} \cdot \ell$ for some universal constant c_0. Overall, we get an $n \times n$ matrix using ℓ bits of randomness, which has rigidity $\Omega(\frac{n^2\ell}{r^2 \log n})$ for rank $r \in [\sqrt{\ell}, \ell/4]$ (where the lower bound follows from the assumption $k \leq m = 2r$, and the upper bound from the assumption $2r = m \leq \ell/2$).

6.3. Open problems. Our work brings up a lot of natural open problems; some of which are stated next. We state all problems in the affirmative, although we actually do not know whether or not they can be resolved in that direction.

Random Toeplitz matrices. While Theorem 1.6 provides an almost tight lower bound on the AN2-complexity of the corresponding bilinear functions, their AN-complexity remains undetermined: Theorem 1.4 asserts a $\Omega(n^{0.6})$ lower bound, whereas Goldreich & Wigderson (2013, Thm. 3.1) states an $O(n^{2/3})$ upper bound.

Open Question 6.1 (Tight AN-complexity lower bound for random Toeplitz matrices). *Prove that, with high probability, bilinear functions that correspond to random Toeplitz matrices have AN-complexity* $\Omega(n^{2/3})$.

The above would be resolved by proving the following rigidity bound[12]

Open Question 6.2 (Rigidity of random Toeplitz matrices). *Prove that, with high probability, random Toeplitz matrices have rigidity* $\Omega(n^2)$ *for rank* $\Omega(n^{2/3})$.

A similar challenge holds with respect to matrices sampled from an 2^{-n}-biased distribution. In fact, it may be easier to settle the following:

[12] Indeed, a bound on structured rigidity (or even on the restricted notions of structured rigidity considered in Section 4) would suffice.

Open Question 6.3. (Rigidity of small-biased distribution of matrices). *Prove that, with high probability, a matrix sampled from an 2^{-n}-biased sample space has rigidity $\Omega(n^2)$ for rank $\Omega(n^{2/3})$.*

Recall that the proof of Theorem 1.7 establishes an almost tight lower bound on the AN-complexity of the corresponding bilinear functions, but this is done via a much more restricted notion of rigidity. We also mention that it is easy to prove that these matrices have rigidity $\widetilde{\Omega}(n^3/r^2)$ for rank $r \in [\sqrt{n}, n/32]$ (by degenerating the proof of Theorem 3.6).[13]

Explicit matrices and bilinear functions. Our lower bounds refer to distributions over n-by-n matrices that are generated using $O(n)$ random bits, and we obtain explicit multilinear functions by using these random bits as auxiliary variables (hence these functions are trilinear or 4-linear, depending on the way the distribution is generated). The begging challenges are to get rid of the randomness.

Open Question 6.4 (AN-complexity lower bound for explicit bilinear function). *For any $\alpha \in (0.5, 2/3]$, present an explicit bilinear function that has AN-complexity $\Omega(n^\alpha)$.*

Needless to say, the larger the α, the better. Even an AN2-complexity lower bound would be welcome. Open Question 6.4 would be resolved by proving the following rigidity bound

Open Question 6.5 (Rigidity of some explicit matrices). *For any $\alpha \in (0.5, 2/3]$, present an explicit matrix that has rigidity $\Omega(n^{3\alpha})$ for rank $\Omega(n^\alpha)$.*

[13] Specifically, Lemma 3.1 can be replaced by a simpler proof that refers to an m-by-m submatrix whose entries are taken from an 2^{-n}-biased distribution over $\{0,1\}^{m^2}$. In this case, it is easy to bound the probability that the submatrix has rank at most $r = m/2 \leq m-(n/m)$ by $\binom{m}{r} \cdot (2^r)^{n/m} \cdot (2^{-(n/m)\cdot m} + 2^{-n})$, where the bound holds by considering n/m rows that depend on at most r other rows (which cover the row basis). Using $r = m/2$, we get a probability bound of $2^m \cdot 2^{n/2} \cdot 2^{-n+1}$. For the rest of the proof, one may select an arbitrary partition of the n-by-n matrix to m-by-m submatrices, since any such partition will yield submatrices as above.

As noted in Section 1.1, this rigidity challenge refers to a range of parameters that differs from the standard one.

Higher AN-complexity lower bounds. Theorem 1.7 provides an AN-complexity lower bound of $\Omega(n^{2/3}/(\log n)^{1/3})$ for some explicit 4-linear function. This is not necessarily tight, since by Goldreich & Wigderson (2013, Thm. 3.1) any t-linear function has AN2-complexity $O((tn)^{t/(t+1)})$. More importantly, we wish to surpass the aforementioned lower bound.

Open Question 6.6 (AN-complexity lower bounds for explicit multilinear functions). *For any $\alpha \in (2/3, 1)$, present an explicit $O(1)$-linear function that has AN-complexity $\Omega(n^\alpha)$.*

By the strategy outlined in Section 5.2, it suffices to meet this challenge with a random tensor of constant dimension sampled using $O(n)$ random bits, provided that its entries may be expressed as $O(1)$-linear functions in the random bits. Here too, even an AN2-complexity lower bound would be welcome.

Acknowledgements

O. G. work was partially supported by the Minerva Foundation with funds from the Federal German Ministry for Education and Research. During the work on this project, A. T. was student at Weizmann Institute of Science, Rehovot, Israel. His research was supported by an Adams Fellowship of the Israel Academy of Sciences and Humanities, by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

An extended abstract of this paper has appeared as Goldreich & Tal (2016).

A. Appendices

A.1. Generalization to larger fields. As stated in Section 3, the choice of field $\mathbb{F}_2$ was not crucial in the proofs of Lemma 3.1, Theorem 3.6 and Theorem 3.9. One can syntactically replace the field size 2 by any prime power q, keeping the proofs intact. Furthermore, in Theorem 3.6, we slightly benefit from taking a larger

field. We state the generalized theorems and point to the improvements over Theorem 3.6.

LEMMA A.1 (Main Lemma for $\mathbb{F}_q$). *Let $m, k \in \mathbb{N}$, $16 \leq k \leq m$. Let $A \in \mathbb{F}_q^{m \times m}$ be the random matrix*

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_m \\ a_{k+1} & a_{k+2} & a_{k+3} & \dots & a_{k+m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{(m-1)k+1} & a_{(m-1)k+2} & a_{(m-1)k+3} & \dots & a_{(m-1)k+m} \end{pmatrix}$$

where $a_1, \dots, a_{(m-1)k+m}$ are uniform scalars from $\mathbb{F}_q$, and let $S \in \mathbb{F}_q^{m \times m}$ be some fixed matrix. Then, $\mathbf{Pr}_A[\operatorname{rank}(S + A) \leq m/2] \leq q^{-km/16}$.

The proof of Lemma A.1 is identical to the original proof of Lemma 3.1, replacing 2 with q.

THEOREM A.2 (random Hankel matrices over $\mathbb{F}_q$ are rigid). *Let $A \in \mathbb{F}_q^{n \times n}$ be a random Hankel matrix $A_{i,j} = a_{i+j}$ where $a_2, \dots, a_{2n}$ are uniform independent scalars from $\mathbb{F}_q$. Then, for every $\sqrt{n} \leq r \leq n/32$, with probability $1 - o(1)$, the matrix A has rigidity $\frac{n^3}{80r^2 \log_q(qn^2)}$ for rank r. In particular, if $q \geq n^{\Omega(1)}$, then we get rigidity $\Omega(\frac{n^3}{r^2})$ for rank r.*

PROOF. We highlight the difference from the proof of Theorem 3.6. When considering the probability that A is simple we can write

(Union Bound)
$$\mathbf{Pr}\left[\exists i,j \ : \ A^{i,j} \text{ is simple}\right] \leq \sum_{i,j} \mathbf{Pr}[A^{i,j} \text{ is simple}]$$

(Union Bound)
$$\leq \sum_{i,j} \sum_{\substack{S \in \mathbb{F}_q^{m \times m}: \\ \mathtt{wt}(S) \leq s'}} \mathbf{Pr}[\operatorname{rank}(A^{i,j} + S) \leq \tfrac{m}{2}]$$

(Lemma A.1)
$$\leq \left(\frac{n}{m}\right)^2 \cdot \binom{m^2}{s'} \cdot q^{s'} \cdot q^{-mk/16}$$
$$< n^2 \cdot (m^2 q)^{s'} \cdot q^{-n/16}$$

Using $s' \leq \frac{n}{20\log_q(qm^2)}$, which follows from $s \leq \frac{n^3}{80r^2\log_q(qn^2)}$, we get $\mathbf{Pr}[\exists i,j \ : \ A^{i,j} \text{ is simple}] = o(1)$, which completes the proof. □

THEOREM A.3 (random Hankel matrices over $\mathbb{F}_q$ are structured rigid). *Let $A \in \mathbb{F}_q^{n\times n}$ be a random Hankel matrix. Then, for every $\sqrt{n} \leq r \leq n/32$ and $s \leq n^3/1000r^2$, with probability $1 - o(1)$, the matrix A has structured rigidity s for rank r.*

The proof of Theorem A.3 is identical to the original proof of Theorem 3.9, replacing 2 with q.

A.2. The structure of matrices associated with general bilinear circuits. The following proof is essentially given in Goldreich & Wigderson (2013, Thm. 4.4), although the result is not spelled out. We give it here for completeness.

PROPOSITION A.4 (Proposition 4.5, restated). *If $\mathtt{C}(F) = m$, then F can be expressed as*

$$\text{(Eq. (4.6))} \qquad \sum_{i=1}^{m} L_i(x)L'_i(y) + \sum_{i=1}^{m} M'_i(x)M_i(y) + \sum_{\ell=1}^{m} Q_\ell(x,y)$$

where $L_1,\ldots,L_m$ and $M_1,\ldots,M_m$ are m-sparse linear functions, $L'_1,\ldots,L'_m$ and $M'_1,\ldots,M_m$ are general linear functions, and each Q_ℓ is a bilinear function of at most m variables from x and at most m variables from y. The matrix associated with F has the form

$$\text{(Eq. (4.7))} \qquad A = L_{\text{col}}B + CL_{\text{row}} + \sum_{\ell=1}^{m} S_\ell$$

where L_{col} is an $n\times m$ matrix with m-sparse columns, B is a general $m\times n$ matrix, C is a general $n\times m$ matrix, L_{row} is an $m\times n$ matrix with m-sparse rows, and each S_ℓ is an $n\times n$ matrix whose ones reside in an $m\times m$ rectangle.

PROOF. We are given a bilinear circuit computing F of AN-complexity m, and we wish to show that we can write F as in Eq. (4.6). The deduction of Eq. (4.7) from Eq. (4.6) is obvious.

We perform simplification rules on the circuit. Let $\mathcal{F}_1, \ldots, \mathcal{F}_m$ denote the gates of the circuits, where $\mathcal{F}_m$ is the output gate. For any $i < m$, if $\mathcal{F}_j$ appears as a free term in $\mathcal{F}_i$ (i.e., the ith gate adds $\mathcal{F}_j$ to its output), then we can "delay" this addition by feeding $\mathcal{F}_j$ to the gates that are fed by $\mathcal{F}_i$, and performing the addition there. If $\mathcal{F}_j$ was only added to $\mathcal{F}_i$ we omit $\mathcal{F}_j$ from $\mathcal{F}_i$ (i.e., omit the edge feeding $\mathcal{F}_j$ into the ith gate).[14] We repeat this simplification rule until one cannot perform it anymore. Since the circuit is a directed acyclic graph, the process will stop. Call a gate an x-gate (resp. y-gate) if there are directed paths only from the x variables (resp. y variables) to this gate, and call it an xy-gate if there are paths both from x and from y. Recall that by multilinearity (Definition 2.3) an x-gate (resp. y-gate) cannot be multiplied by another x-gate (resp. y-gate), and an xy-gate cannot be multiplied at all. We observe that after the transformation the circuit is of height at most 3 (where the height of a gate is the longest path from an input variable to it). This is shown as follows.

- Next to the inputs (i.e., at height 1) we have either x-gates (resp. y-gates) computing linear functions in x (resp. y) on m variables, or xy-gates computing a bilinear function on m variables from both x and y.
- For $i < m$, if $\mathcal{F}_i$ is at height 2, then it must be an xy-gate, since by the simplification rules it must multiply two of its inputs, and these inputs must come from different blocks of variables.
- By the same reasoning, if $\mathcal{F}_i$ is of height 3, then it must be the output gate (i.e., $i = m$), since otherwise it must be fed by an xy-gate of height 2, which cannot be added nor multiplied.

We shall express the functions computed by gates at heights 1, 2, and finally by the output gate.

Height-1 Gates: Denote by $L_1, \ldots, L_k$ (resp., $M_1, \ldots, M_{k'}$) the linear functions computed by x-gates (resp. y-gates) at height 1. Denote by $Q_i(x, y)$ the bilinear function computed by

[14] The condition accounts for the case that $\mathcal{F}_i$ does not compute a homogeneous polynomial. In that case it is possible that the value of $\mathcal{F}_j$ appears in the output of $\mathcal{F}_i$ both as a free term and as a factor in a product with some other gate or variables (e.g., $\mathcal{F}_2(x, y) = x_7 \cdot \mathcal{F}_1(y) + x_2 y_3 + \mathcal{F}_1(y)$).

an xy-gate $\mathcal{F}_i$ at height 1. Each of L_i, M_i and Q_i computes a function that depends on at most m variables from x and at most m variables from y.

Height-2 Gates: Each xy-gate $\mathcal{F}_i$ (which is not the output gate) at height 2 is fed with x-gates, y-gates and variables (we may assume that it is not fed by xy-gates by the simplification rule). It computes a sum of products of terms in x and terms in y; we break these products according to whether or not each term is a gate or an input variable. For any linear function $L_j(x)$, we gather all linear functions in y and all variables in y that are multiplied by L_j in the computation of gate $\mathcal{F}_i$; this gives a general (not necessarily sparse) linear function in y, denoted $L_{i,j}(y)$. For any linear function $M_j(y)$, we gather all variables from x that are multiplied by M_j to get a linear function in x, denoted by $M_{i,j}(x)$. Letting $Q_i(x,y)$ be the bilinear function with all variable by variable products, we get

$$\mathcal{F}_i(x,y) = \sum_{j=1}^{k} L_j(x)L_{i,j}(y) + \sum_{j=1}^{k'} M_{i,j}(x)M_j(y) \; + Q_i(x,y) \;, \tag{A.5}$$

where Q_i depends on at most m variables from x and at most m variables from y.

Output Gate: If the output gate is at height 2, then we are done. Otherwise, we may assume that the output gate is only fed by xy-gates, since if it is fed by some other gates (or variables) then their contribution (multiplied by other gates or variables) can be computed by an auxiliary xy-gate (of height 2) that feeds the output gate. Hence, the output gate computes the sum of at most m gates, each of which computes a bilinear function as in Eq. (A.5), where this also covers xy-gates of height 1 (since they also compute a function of this very form). Overall, we get that

$$\begin{aligned} F(x,y) &= \sum_{i=1}^{m} \left(\sum_{j=1}^{k} L_j(x)L_{i,j}(y) + \sum_{j=1}^{k'} M_{i,j}(x)M_j(y) + Q_i(x,y) \right) \\ &= \sum_{j=1}^{k} L_j(x) \cdot \left(\sum_{i=1}^{m} L_{i,j}(y) \right) + \sum_{j=1}^{k'} \left(\sum_{i=1}^{m} M_{i,j}(x) \right) \cdot M_j(y) \\ &\quad + \sum_{i=1}^{m} Q_i(x,y) \end{aligned}$$

Letting $L'_j(y) = \sum_{i=1}^{m} L_{i,j}(y)$ and $M'_j(x) = \sum_{i=1}^{m} M_{i,j}(x)$, we get

$$F(x,y) = \sum_{j=1}^{k} L_j(x)L'_j(y) + \sum_{j=1}^{k'} M'_j(x)M_j(y) + \sum_{i=1}^{m} Q_i(x,y)$$

where the L_j's are m-sparse linear functions in x, the L'_j's are general linear functions in y, the M'_j's are general linear functions in x, the M_j's are m-sparse linear functions in y, and each Q_i is a bilinear function that depends on at most m variables in x and y. □

A.3. Characterization of AN-complexity for bilinear forms. In this section, we prove Proposition 4.8 and 4.4, which show that Proposition 4.5 and 4.1 capture $\mathtt{C}(F)$ and $\mathtt{C}_2(F)$, respectively.

Proposition A.6 (Proposition 4.4, restated). *Any bilinear form F that can be written as in Eq. (4.2), has $\mathtt{C}_2(F) = O(m)$.*

Proof. We start by recalling Eq. (4.2). Let F be a bilinear form that can be written as

$$F(x,y) = \sum_{(i,j)\in P} L_i(x)L'_j(y) + \sum_{\ell=1}^{m} Q_\ell(x,y)$$

where $P \subseteq [m] \times [m]$, the L_i-s and the L'_j-s are m-sparse linear functions and the Q_ℓ-s are bilinear functions that depend on at most m variables from x and at most m variables from y. We shall build a depth-2 bilinear circuit computing F of AN-complexity $O(m)$.

The construction is straightforward: We compute each Q_ℓ by a single gate which is fed by the $2m$ input variables on which it depends. Similarly, we compute each L_i (L'_j, resp.) by a single linear gate fed by at most m input variables from x (y, resp.). Lastly, we feed the output gate with the aforementioned $3m$ gates (computing L_i, L'_j and Q_ℓ for $i,j,\ell \in [m]$); that is, the output gate computes $\sum_{(i,j)\in P} L_i(x)L'_j(y) + \sum_{\ell=1}^{m} Q_\ell(x,y)$ by simply multiplying the L_i-s with the L'_j-s according to the subset P and adding all the bilinear functions $Q_1, \ldots, Q_m$. Hence, we have built a bilinear circuit

computing F of depth-2 with $3m+1$ gates and fan-in at most $3m$, and $\mathtt{C}_2(F) \leq 3m+1$ follows. □

PROPOSITION A.7 (Proposition 4.8, restated). *Any bilinear form F that can be written as in Eq. (4.6), has $\mathtt{C}(F) = O(m+\sqrt{n})$.*

PROOF. We start by recalling Eq. (4.6). Let F be a bilinear form that can be written as

$$F(x,y) = \sum_{i=1}^{m} L_i(x)L'_i(y) + \sum_{i=1}^{m} M'_i(x)M_i(y) + \sum_{\ell=1}^{m} Q_\ell(x,y)$$

where $L_1,\ldots,L_m$ and $M_1,\ldots,M_m$ are m-sparse linear functions, $L'_1,\ldots,L'_m$ and $M'_1,\ldots,M_m$ are general linear functions, and each Q_ℓ is a bilinear function of at most m variables from x and at most m variables from y. We shall show how to compute $F(x,y)$ by a bilinear circuit of AN-complexity $O(m+\sqrt{n})$.

As in the proof for the depth-2 case, we compute each Q_ℓ by a single gate which is fed by $2m$ input variables on which it depends. We compute each L_i (M_i, resp.) by a single linear gate fed by at most m input variables from x (y, resp.). To compute $\sum_{i=1}^{m} L_i(x)L'_i(y)$ we partition the variables in y to $\sqrt{n}$ buckets of size $\sqrt{n}$ each. For $k \in [\sqrt{n}]$ we denote the kth bucket by B_k; for a concrete choice let $B_k = \{(k-1)\sqrt{n}+1,\ldots,k\sqrt{n}\}$. We write $L'_i(y) = \sum_{j=1}^{n} a_{i,j}y_j$ where $a_{i,j} \in \mathbb{F}_2$. For $k \in \sqrt{n}$, we denote by $L'_{i,k}(y) \triangleq \sum_{j\in B_k} a_{i,j}y_j$. Clearly

$$\sum_{i=1}^{m} L_i(x)L'_i(y) = \sum_{i=1}^{m}\sum_{k=1}^{\sqrt{n}} L_i(x)L'_{i,k}(y) = \sum_{k=1}^{\sqrt{n}}\sum_{i=1}^{m} L_i(x)L'_{i,k}(y) \ .$$

For a given $k \in [\sqrt{n}]$ we denote the inner sum in the RHS of the last equation by $F_k(x,y) \triangleq \sum_{i=1}^{m} L_i(x)L'_{i,k}(y)$. For $k \in [\sqrt{n}]$, we compute F_k by a single gate which is fed by the linear functions $L_1(x),\ldots,L_m(x)$, which have been computed already, and the variables $\{y_j : j \in B_k\}$. The gate multiplies each $L_i(x)$ with the variables in y according to $L'_{i,k}$, and then sums up all the products yielding $F_k(x,y)$. We put a gate summing all F_k-s together to get

$\sum_{k=1}^{\sqrt{n}} F_k(x,y) = \sum_{i=1}^{m} L_i(x)L'_i(y)$. In a completely symmetric fashion we compute using $\sqrt{n}+1$ additional gates $\sum_{i=1}^{m} M'_i(x)M_i(y)$.

To wrap up, we feed the output gate by the m gates computing $Q_1,\ldots,Q_m$, in addition to the two gates computing $\sum_{i=1}^{m} L_i(x)L'_i(y)$ and $\sum_{i=1}^{m} M'_i(x)M_i(y)$ respectively. The output gate adds up all its entries to output the required function. Overall, we constructed a bilinear circuit computing F with $O(m+\sqrt{n})$ gates and fan-in $O(m+\sqrt{n})$, which shows that $\mathtt{C}(F) = O(m+\sqrt{n})$. □

References

N. Alon, O. Goldreich, J. Håstad & R. Peralta (1992). Simple Construction of Almost k-wise Independent Random Variables. *Random Structures and Algorithms* **3**(3), 289–304.

A. E. Andreev (1987). On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.* **42**, 63–66. In Russian.

P. Bürgisser & M. Lotz (2004). Lower bounds on the bounded coefficient complexity of bilinear maps. *J. ACM* **51**(3), 464–482.

J. Friedman (1993). A note on matrix rigidity. *Combinatorica* **13**(2), 235–239.

O. Goldreich (2008). *Computational Complexity: A Conceptual Perspective*. Cambridge University Press.

O. Goldreich & A. Tal (2016). Matrix rigidity of random Toeplitz matrices. In *STOC*, 91–104.

O. Goldreich & A. Wigderson (2013). On the Size of Depth-Three Boolean Circuits for Computing Multilinear Functions. *Electronic Colloquium on Computational Complexity (ECCC)* **20**, 43.

J. Håstad (1989). Almost Optimal Lower Bounds for Small Depth Circuits. In *RANDOMNESS AND COMPUTATION*, 6–20. JAI Press.

S. Kopparty, M. Kumar & M. E. Saks (2014). Efficient Indexing of Necklaces and Irreducible Polynomials over Finite Fields. In *ICALP*, 726–737.

R. Lidl & H. Niederreiter (1997). *Finite Fields*, volume 20 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2nd edition.

S. V. Lokam (2009). Complexity Lower Bounds using Linear Algebra. *Foundations and Trends in Theoretical Computer Science* **4**(1–2), 1–155.

E. Mossel, A. Shpilka & L. Trevisan (2006). On epsilon-biased generators in NC^0. *Random Structures and Algorithms* **29**(1), 56–81.

J. Naor & M. Naor (1993). Small-Bias Probability Spaces: Efficient Constructions and Applications. *SIAM J. on Computing* **22**(4), 838–856.

R. Paturi, P. Pudlák, M. E. Saks & F. Zane (2005). An improved exponential-time algorithm for k-SAT. *J. ACM* **52**(3), 337–364.

R. Paturi, P. Pudlák & F. Zane (1999). Satisfiability Coding Lemma. *Chicago J. Theor. Comput. Sci.* **1999**.

R. Raz (2003). On the complexity of matrix product. *SIAM J. on Computing* **32**(5), 1356–1369.

M. A. Shokrollahi, M. A. Spielman & V. Stemann (1997). A Remark on Matrix Rigidity. *Inf. Process. Lett.* **64**(6), 283–285.

L. G. Valiant (1977). Graph-theoretic arguments in low-level complexity. In *Lecture notes in Computer Science*, volume 53, 162–176. Springer.

Manuscript received June 13, 2016

Oded Goldreich
Weizmann Institute of Science
Rehovot, Israel
oded.goldreich@weizmann.ac.il

Avishay Tal
Institute for Advanced Study
Princeton, NJ, USA
avishay.tal@gmail.com