Lower Bounds on the Bounded Coefficient Complexity of Bilinear Maps

Peter Bürgisser and Martin Lotz
Universität Paderborn

We prove lower bounds of order $n \log n$ for both the problem of multiplying polynomials of degree n, and of dividing polynomials with remainder, in the model of bounded coefficient arithmetic circuits over the complex numbers. These lower bounds are optimal up to order of magnitude. The proof uses a recent idea of R. Raz [Proc. 34th STOC 2002] proposed for matrix multiplication. It reduces the linear problem of multiplying a random circulant matrix with a vector to the bilinear problem of cyclic convolution. We treat the arising linear problem by extending J. Morgenstern's bound [J. ACM 20, pp. 305-306, 1973] in a unitarily invariant way. This establishes a new lower bound on the bounded coefficient complexity of linear forms in terms of the singular values of the corresponding matrix. In addition, we extend these lower bounds for linear and bilinear maps to a model of circuits that allows a restricted number of unbounded scalar multiplications.

Categories and Subject Descriptors: F.1.1 [**Computation by Abstract Devices**]: Models of Computation; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*Computations on polynomials*; I.1.2 [**Symbolic and Algebraic Manipulation**]: Algorithms

General Terms: Algorithms, Theory

Additional Key Words and Phrases: algebraic complexity, bilinear circuits, cyclic convolution, Fast Fourier Transform, lower bounds, polynomial multiplication, singular values, rigidity

1. INTRODUCTION

Finding lower bounds on the complexity of polynomial functions over the complex numbers is one of the fundamental problems of algebraic complexity theory. It becomes more tractable if we restrict the model of computation to arithmetic circuits, where the multiplication with scalars is restricted to constants of bounded absolute value. This model was introduced in a seminal work by [Morgenstern 1973; 1975], where it was proved that the complexity of multiplying a vector with some given square matrix A is bounded from below by the logarithm of the absolute value of the determinant of A. As a consequence, Morgenstern derived the lower bound $\frac{1}{2} n \log n$ for computing the Discrete Fourier Transform.

[Valiant 1976; 1977] analyzed the problem to prove nonlinear lower bounds on the complexity of the Discrete Fourier Transform and related linear problems in the unrestricted model of arithmetic circuits. However, despite many attempts, this problem is still open today.

Author's address: Faculty of Computer Science, Electrical Engineering, and Mathematics, University of Paderborn, 33095 Paderborn, Germany. Email: {pbuerg,lotzm}@math.uni-paderborn.de.
This work was supported by the Forschungspreis 2002 der Universität Paderborn and by the Paderborn Institute for Scientific Computation (PaSCo).

To motivate the bounded coefficient model (b.c. for short), we note that many algorithms for arithmetic problems, like the Fast Fourier Transform and the fast algorithms based on it, use only small constants. [Chazelle 1998] advocated the b.c. model as a natural model of computation by arguing that the finite representation of numbers is essentially equivalent to bounded coefficients.

[Chazelle 1998] refined Morgenstern's bound by proving a lower bound on the b.c. linear complexity of a matrix A in terms of the singular values of A. His applications are nonlinear lower bounds for range searching problems. Several papers [Nisan and Wigderson 1995; Lokam 1995; Pudlák 1998] studied b.c. arithmetic circuits. The concept of matrix rigidity, originally introduced in [Valiant 1977], hereby plays a vital role. A geometric variant of this concept (Euclidean metric instead of Hamming metric) is closely related to the singular value decomposition of a matrix and turns out to be an important tool, as worked out in [Lokam 1995]. [Raz 2002] recently proved a nonlinear lower bound on the complexity of matrix multiplication in the b.c. model. To our knowledge, this paper and [Nisan and Wigderson 1995] are the only ones which deal with the complexity of bilinear maps in the b.c. model of computation. However, the proof of the $\Omega(n \log n)$ lower bound in [Nisan and Wigderson 1995](Cor. 3) is incorrect, as it assumes that the derivative inequality [Baur and Strassen 1983] carries over to the b.c. model. The counterexample $2^n \sum_{1 \leq i \leq n} X_i Y_i$ pointed out by [Pudlák 2003] shows that this is not true.

The main result of this paper (Theorem 4.1) is a nonlinear lower bound of order $n \log n$ to compute the cyclic convolution of two given vectors in the b.c. model. This bound is optimal up to a constant factor. The proof is based on ideas in [Raz 2002] to establish a lower bound on the complexity of a bilinear map $(x, y) \mapsto \varphi(x, y)$ in terms of the complexity of the linear maps $y \mapsto \varphi(a, y)$ obtained by fixing the first input to a (Lemma 2.4). However, the linear circuit for the computation of $y \mapsto \varphi(a, y)$ resulting from a hypothetical b.c. circuit for φ has to be transformed into a *small* one with bounded coefficients. This can be achieved with a geometric rigidity argument by choosing a vector a at random according to the standard normal distribution in a suitable linear subspace of $\mathbb{C}^m$ (Lemma 4.2).

[Raz 2002] proceeded by analyzing the complexity of the resulting linear map with a geometric rigidity bound and the Hoffman-Wielandt inequality. There one has to study the multiplication with a random matrix. In our situation, however, we have to estimate the complexity of the multiplication with a random circulant matrix. We treat this by extending Morgenstern's bound in a new way. We define the r-*mean square volume* of a complex matrix A, which turns out to be the square root of the r-th elementary symmetric function in the squares of the singular values of A. An important property of this quantity is that it is invariant under multiplication with unitary matrices from the left or the right. We prove that the logarithm of the r-mean square volume provides a lower bound on the b.c. complexity of the matrix A (Proposition (3.1)). This implies that the logarithm of the product of the largest r singular values is a lower bound on the b.c. complexity.

Recently, [Raz 2003] pointed out to us a technically simpler proof of our main result, avoiding a study of correlations. His proof is based on the rigidity bound combined with a lower bound for the sum of squares of the smallest r singular values of a random circulant matrix.

We also study an extension of the bounded coefficient model of computation by allowing a limited number of *help gates* corresponding to scalar multiplications with unbounded constants. We can show that our proof technique is robust in the sense that it still allows to prove $n \log n$ lower bounds if the number of help gates is restricted to $(1-\epsilon)n$ for fixed $\epsilon > 0$. This is achieved by an extension of the mean square volume bound (Proposition 6.1), which is related to the spectral lemma in [Chazelle 1998]. The proof is based on some matrix perturbation arguments.

From the lower bound for the cyclic convolution we obtain nonlinear lower bounds for polynomial multiplication, inversion of power series, and polynomial division with remainder by noting that the well-known reductions between these problems [Bürgisser et al. 1997] preserve the b.c. property. These lower bounds are again optimal up to order of magnitude.

1.1 Organization of the paper

In Section 2, we introduce the model of computation and discuss known facts about singular values and matrix rigidity. We also introduce some notation and present auxiliary results related to (complex) Gaussian random vectors. In Section 3 we first recall previously known lower bounds for b.c. linear circuits. Then we introduce the mean square volume of a matrix and prove an extension of Morgenstern's bound in terms of this quantity. Section 4 contains the statement and proof of our main theorem, the lower bound on cyclic convolution. In Section 5, we derive lower bounds for polynomial multiplication, inversion of power series and division with remainder. Finally, in Section 6 we show that our results can be extended to the case, where a limited number of unbounded scalar multiplications (help gates) is allowed.

2. PRELIMINARIES

We start this section by giving a short introduction to the model of computation.

2.1 The model of computation

We will base our arguments on the model of algebraic straight-line programs over $\mathbb{C}$, which are often called arithmetic circuits in the literature. For details on this model we refer to chapter 4 of [Bürgisser et al. 1997]. By a result in [Strassen 1973b], we may exclude divisions without loss of generality.

Definition 2.1. A *straight-line program* Γ expecting inputs of length n is a sequence $(\Gamma_1, \ldots, \Gamma_r)$ of instructions $\Gamma_s = (\omega_s; i_s, j_s)$, $\omega_s \in \{\times, +, -\}$ or $\Gamma_s = (\omega_s; i_s)$, $\omega_s \in \mathbb{C}$, with integers i_s, j_s satisfying $-n < i_s, j_s < s$. A sequence of polynomials $b_{-n+1}, \ldots, b_r$ is called the *result sequence* of Γ on input variables $a_1, \ldots, a_n$, if for $-n < s \le 0$, $b_s = a_{n+s}$, and for $1 \le s \le r$, $b_s = b_{i_s} \omega_s b_{j_s}$ if $\Gamma_s = (\omega_s; i_s, j_s)$ and $b_s = \omega_s b_{i_s}$ if $\Gamma_s = (\omega_s; i_s)$. Γ is said to *compute* a set of polynomials F on input $a_1, \ldots, a_n$, if the elements in F are among those of the result sequence of Γ on that input. The *size* $\mathcal{S}(\Gamma)$ of Γ is the number r of its instructions.

In the sequel we will refer to such straight-line programs briefly as circuits. A circuit in which the scalar multiplication is restricted to scalars of absolute value at most 2 will be called a *bounded coefficient circuit* (b.c. circuit for short). Of course, the bound of 2 could be replaced by any other fixed bound. Any circuit can be transformed into a b.c. circuit by replacing a multiplication with a scalar λ with at most $\log|\lambda|$ additions and a multiplication with a scalar of absolute value at most 2. Unless otherwise stated, log will always refer to logarithms to the base 2.

We now introduce restricted notions of circuits, designed for computing linear and bilinear maps.

Definition 2.2. A circuit $\Gamma = (\Gamma_1, \ldots, \Gamma_r)$ expecting inputs $X_1, \ldots, X_n$ is called a *linear circuit*, if $\omega_s \in \{+, -\}$ for every instruction $\Gamma_s = (\omega_s; i_s, j_s)$, or $\omega_s \in \mathbb{C}$ if the instruction is of the form $(\omega_s; i_s)$. A circuit on inputs $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ is called a *bilinear circuit*, if its sequence of instructions can be partitioned as $\Gamma = (\Gamma^{(1)}, \Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)})$, where

(1) $\Gamma^{(1)}$ is a linear circuit with the X_i as inputs,

(2) $\Gamma^{(2)}$ is a linear circuit with the Y_j as inputs,

(3) each instruction from $\Gamma^{(3)}$ has the form $(\times; i, j)$, with $\Gamma_i \in \Gamma^{(1)}$ and $\Gamma_j \in \Gamma^{(2)}$,

(4) $\Gamma^{(4)}$ is a linear circuit with the previously computed results of $\Gamma^{(3)}$ as inputs.

In other words, $\Gamma^{(1)}$ and $\Gamma^{(2)}$ compute linear functions $f_1, \ldots, f_k$ in the X_i and $g_1, \ldots, g_\ell$ in the Y_j. $\Gamma^{(3)}$ then multiplies the f_i with the g_j and $\Gamma^{(4)}$ computes linear combinations of the products $f_i g_j$.

It is clear that linear circuits compute linear maps and that bilinear circuits compute bilinear maps. On the other hand, it can be shown that any linear (bilinear) map can be computed by a linear (bilinear) circuit such that the size increases at most by a constant factor (cf. [Bürgisser et al. 1997, Theorem 13.1, Proposition 14.1]). This remains true when considering bounded coefficient circuits, as can easily be checked. From now on, we will only be concerned with bounded coefficient circuits.

Definition 2.3. By the *b.c. complexity* $\mathcal{C}(\varphi)$ of a bilinear map $\varphi\colon \mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}^p$ we understand the size of a smallest b.c. bilinear circuit computing φ. By the *b.c. complexity* $\mathcal{C}(\varphi^A)$ of a linear map $\varphi^A\colon \mathbb{C}^n \to \mathbb{C}^m$ (or the corresponding matrix $A \in \mathbb{C}^{m\times n}$), we understand the size of a smallest b.c. linear circuit computing φ^A.

By abuse of notation, we also write $\mathcal{C}(F)$ for the smallest size of a b.c. circuit computing a set F of polynomials from the variables.

Let $\varphi\colon \mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}^p$ be a bilinear map described by $\varphi_k(X, Y) = \sum_{i,j} a_{ijk} X_i Y_j$. Assuming $|a_{ijk}| \le 2$, it is clear that $\mathcal{C}(\varphi) \le 3mnp$. Therefore, if $f_1, \ldots, f_k$ are the linear maps computed on the first set of inputs by an optimal b.c. bilinear circuit for φ, we have $k \le \mathcal{S}(\Gamma) \le 3mnp$.

The complexity of a bilinear map φ can be related to the complexity of the associated linear map $\varphi(a, -)$, where $a \in \mathbb{C}^m$. We have taken the idea behind the following lemma from [Raz 2002].

Lemma 2.4. *Let $\varphi\colon \mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}^p$ be a bilinear map and Γ be a b.c. bilinear circuit computing φ. If $f_1, \ldots, f_k$ are the linear maps computed by the circuit on the first set of inputs, then for all $a \in \mathbb{C}^m$:*

$$\mathcal{C}(\varphi(a, -)) \le \mathcal{S}(\Gamma) + p \log\left(\max_j |f_j(a)|\right).$$

Proof. Let $a \in \mathbb{C}^m$ be chosen and set $\gamma = \max_j |f_j(a)|$. Transform the circuit Γ into a linear circuit Γ' by the following steps:

(1) replace the first argument x of the input by a,
(2) replace each multiplication by $f_i(a)$ with a multiplication by $2\gamma^{-1} f_i(a)$,
(3) multiply each output by $\gamma/2$, simulating this with at most $\log(\gamma/2)$ additions and one multiplication with a scalar of absolute value at most 2.

This is a b.c. linear circuit computing the map $\varphi(a, -) \colon \mathbb{C}^n \to \mathbb{C}^p$. Since there are p outputs, the size increases by at most $p \log \gamma$. □

2.2 Singular values and matrix rigidity

The *Singular Value Decomposition* (SVD) is one of the most important matrix decompositions in numerical analysis. Lately, it has also come to play a prominent role in proving lower bounds for linear circuits [Chazelle 1998; Lokam 1995; Raz 2002]. In this section, we present some basic facts about singular values and show how they relate to notions of matrix rigidity. For a more detailed account on the SVD, we refer to [Golub and Van Loan 1996]. We also find [Courant and Hilbert 1931, Chapt. 1, Sect. 4] a useful reference.

The singular values of $A \in \mathbb{C}^{m\times n}$, $\sigma_1 \geq \ldots \geq \sigma_{\min\{m,n\}}$, can be defined as the square roots of the eigenvalues of the Hermitian matrix AA^*. Alternatively, they can be characterized as follows:

$$\sigma_{r+1} = \min\{\|A - B\|_2 \mid B \in \mathbb{C}^{m\times n}, \mathrm{rk}(B) \leq r\},$$

where $\|\cdot\|_2$ denotes the matrix 2-norm, that is, $\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2$. An important consequence is the Courant-Fischer min-max theorem stating

$$\sigma_{r+1} \; = \min_{\mathrm{codim} V = r} \; \max_{x \in V - \{0\}} \frac{\|Ax\|_2}{\|x\|_2}.$$

This description implies the following useful fact from matrix perturbation theory:

$$\sigma_{r+h}(A) \leq \sigma_r(A + E) \tag{1}$$

if the matrix E has rank at most h.

More generally, for any metric d on $\mathbb{C}^{m\times n}$ (or $\mathbb{R}^{m\times n}$) and $1 \leq r \leq \min\{m, n\}$, we can define the *r-rigidity* of a matrix A to be the distance of A to the set of all matrices of rank at most r with respect to this metric:

$$\mathrm{rig}_{\mathrm{d},r}(A) = \min\{\mathrm{d}(A, B) \mid B \in \mathbb{C}^{m\times n}, \mathrm{rk}(B) \leq r\}.$$

Using the Hamming metric, we obtain the usual matrix rigidity as introduced in [Valiant 1977]. On the other hand, using the metric induced by the 1, 2-norm $\|A\|_{1,2} := \max_{\|x\|_1=1} \|Ax\|_2$, we obtain the following geometric notion of rigidity, as introduced in [Raz 2002]:

$$\mathrm{rig}_r(A) \; = \min_{\dim V = r} \; \max_{1\leq i\leq n} \mathrm{dist}(a_i, V).$$

Here, the a_i are the column vectors of $A \in \mathbb{C}^{m\times n}$ and dist denotes the usual Euclidean distance, i.e., $\mathrm{dist}(a_i, V) := \min_{b\in V} \|a_i - b\|_2$.

Notions of rigidity can be related to one another the same way the underlying norms can. In particular, we have the following relationship between the geometric rigidity and the singular values:

$$\frac{1}{\sqrt{n}}\,\sigma_{r+1}(A) \leq \mathrm{rig}_r(A) \leq \sigma_{r+1}(A).$$

The proofs of these inequalities are based on well known inequalities for matrix norms. To be precise, note that if B is a matrix of rank at most r with columns b_i, we have

$$\|A-B\|_{1,2}^2 = \max_i \|a_i - b_i\|_2^2 \geq \frac{1}{n}\sum_{i=1}^{n}\|a_i-b_i\|_2^2 \geq \frac{1}{n}\|A-B\|_2^2 \geq \frac{1}{n}\sigma_{r+1}^2,$$

which shows the left inequality. The other inequality follows from the fact that $\|A\|_{1,2} \leq \|A\|_2$, which is a consequence of $\|x\|_2 \leq \|x\|_1$ for $x \in \mathbb{C}^n$.

2.3 Complex Gaussian vectors

A random vector $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$ is called *standard Gaussian* iff its components X_i are i.i.d. standard normal distributed. It is clear that an orthogonal transformation of such a random vector is again standard Gaussian.

Throughout this paper, we will be working with random vectors Z assuming values in $\mathbb{C}^n$. However, by identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$, we can think of Z as a $2n$-dimensional real random vector. In particular, it makes sense to say that such Z is (standard) Gaussian in $\mathbb{C}^n$.

Let U be an r-dimensional linear subspace of $\mathbb{C}^n$. We say that a random vector Z with values in U is *standard Gaussian in U* iff for some orthonormal basis $b_1, \ldots, b_r$ of U we have $Z = \sum_j \zeta_j b_j$, where the random vector (ζ_j) of the components is standard Gaussian in $\mathbb{C}^r$. It is easy to see that this description does not depend on the choice of the orthonormal basis. In fact, the transformation of a standard Gaussian vector with a unitary matrix is again standard Gaussian, since a unitary transformation $\mathbb{C}^r \to \mathbb{C}^r$ induces an orthogonal transformation $\mathbb{R}^{2r} \to \mathbb{R}^{2r}$.

The following lemma is a direct consequence of some facts about the normal distribution.

LEMMA 2.5. *Let $(Z_1, \ldots, Z_n)$ be standard Gaussian in $\mathbb{C}^n$. Consider a complex linear combination $S = f_1 Z_1 + \ldots + f_n Z_n$ with $f = (f_1, \ldots, f_n) \in \mathbb{C}^n$. Then the real and imaginary parts of S are independent and normal distributed, each with mean 0 and variance $\|f\|_2^2$. Moreover, $T := |S|^2/2\|f\|_2^2$ is exponentially distributed with parameter 1. That is, the density function is e^{-t} for $t \geq 0$ and the mean and the variance of T are both equal to 1.*

Proof. If $X_1, \ldots, X_n$ are standard Gaussians in $\mathbb{R}$ and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, then $\sum_{j=1}^n a_j X_j$ is again Gaussian, with mean 0 and variance $\|a\|_2^2$. Note also that if $Z = X + iY$ with independent standard Gaussians X, Y in $\mathbb{R}$, and $f \in \mathbb{C}$, then the real and imaginary parts of fZ are again independent, with mean 0 and variance $|f|^2$. This follows from the fact that complex multiplication corresponds to a rotation and scaling. From these observations we obtain the first statement of the lemma. In particular, $\|f\|_2^{-1} S = X + iY$ with independent standard Gaussians

X, Y in $\mathbb{R}$. It is well known that in this case, $\frac{1}{2}(X^2+Y^2)$ is exponentially distributed with parameter 1, see [Feller 1971, II.2-3] for details □

2.4 Two useful inequalities

Let X, Y be i.i.d. standard normal random variables and set $\gamma := 1 - \mathrm{E}[\log X^2]$ and $\theta := \mathrm{E}[\log^2(X^2+Y^2)]$. Evaluating the corresponding integrals yields

$$\gamma = -\frac{1}{\sqrt{\pi}} \int_0^\infty t^{-1/2} e^{-t} \log t \, dt \approx 2.83$$
$$\theta = \frac{1}{2} \int_0^\infty e^{-t/2} \log^2 t \, dt \approx 3.45.$$

LEMMA 2.6. *Let Z be a centered Gaussian variable with complex values. Then*

$$0 \le \log \mathrm{E}[|Z|^2] - \mathrm{E}[\log |Z|^2] \le \gamma, \ \mathrm{Var}(\log |Z|^2) \le \theta.$$

Proof. By a principal axis transformation, we may assume that $Z = \lambda_1 X + i\lambda_2 Y$ with independent standard normal X, Y. The difference $\Delta := \log \mathrm{E}[|Z|^2] - \mathrm{E}[\log |Z|^2]$ is nonnegative, since log is concave (Jensen's inequality). By linearity of the mean, Δ as well as $\mathrm{Var}(\log |Z|^2)$ are invariant under multiplication of Z with scalars. We may therefore w.l.o.g. assume that $1 = \lambda_1 \ge \lambda_2$. From this we see that

$$\log \mathrm{E}[|Z|^2] = \log \mathrm{E}[X^2 + \lambda_2^2 Y^2] \le \log \mathrm{E}[X^2 + Y^2] = 1$$
$$\mathrm{E}[\log |Z|^2] = \mathrm{E}[\log (X^2 + \lambda_2^2 Y^2)] \ge \mathrm{E}[\log X^2] = 1 - \gamma,$$

which implies the first claim. The estimates

$$\mathrm{Var}(\log |Z|^2) \le \mathrm{E}[\log^2 |Z|^2] \le \mathrm{E}[\log^2(X^2+Y^2)] = \theta.$$

prove the second claim. □

3. THE MEAN SQUARE VOLUME BOUND

Morgenstern's bound [Morgenstern 1973] states that $\mathcal{C}(A) \ge \log |\det(A)|$ for a square matrix A, see also [Bürgisser et al. 1997, Chapter 13] for details. We are going to study several generalizations of this bound.

Let $A \in \mathbb{C}^{m \times n}$ be a matrix. For an r-subset $I \subseteq [m] := \{1, \dots, m\}$ let A_I denote the submatrix of A consisting of the rows indexed by I. The Gramian determinant $\det A_I A_I^*$ can be interpreted as the square of the volume of the parallelepiped spanned by the rows of A_I (A^* denotes the complex transpose of A).

[Raz 2002] defined the *r-volume* of A by

$$\mathrm{vol}_r(A) := \max_{|I|=r} (\det A_I A_I^*)^{1/2}$$

and observed that the proof of Morgenstern's bound extends to the following *r-volume bound*:

$$\mathcal{C}(A) \ge \log \mathrm{vol}_r(A). \tag{2}$$

Moreover, [Raz 2002] related this quantity to the geometric rigidity as follows:

$$\mathrm{vol}_r(A) \ge (\mathrm{rig}_r(A))^r,$$

which implies the *rigidity bound*,

$$\mathcal{C}(A) \geq r \log \mathrm{rig}_r(A). \tag{3}$$

For our purposes it will be convenient to work with a variant of the r-volume that is completely invariant under unitary transformations. Instead of taking the maximum of the volumes $(\det A_I A_I^*)^{1/2}$, we will use the sum of the squares. We define the r-*mean square volume* $\mathrm{msv}_r(A)$ of $A \in \mathbb{C}^{m\times n}$ by

$$\mathrm{msv}_r(A) := \left(\sum_{|I|=r} \det A_I A_I^*\right)^{1/2} = \left(\sum_{|I|=|J|=r} |\det A_{I,J}|^2\right)^{1/2}.$$

Hereby, $A_{I,J}$ denotes the $r \times r$ submatrix consisting of the rows indexed by I and columns indexed by J. The second equality is a consequence of the Binet-Cauchy formula $\det A_I A_I^* = \sum_{|J|=r} |\det A_{I,J}|^2$, see [Bellman 1997, Chapter 4]. The choice of the L_2-norm instead of the maximum norm results in the following inequality

$$\mathrm{vol}_r(A) \leq \mathrm{msv}_r(A) \leq \sqrt{\binom{m}{r}}\, \mathrm{vol}_r(A). \tag{4}$$

The mean square volume has the following nice properties:

$$\mathrm{msv}_r(A) = \mathrm{msv}_r(A^*),\ \mathrm{msv}_r(\lambda A) = |\lambda|^r \,\mathrm{msv}_r(A),\ \mathrm{msv}_r(A) = \mathrm{msv}_r(UAV),$$

where $\lambda \in \mathbb{C}$ and U and V are unitary matrices of the correct format. The first two properties are straightforward to verify. As for unitary invariance, let $A = (a_1, \ldots, a_m)^\top$. Then for $I \subseteq [m]$ with $|I| = r$ we have $A_I A_I^* = (\langle a_j, a_k\rangle)_{j,k\in I}$, where $\langle\cdot,\cdot\rangle$ denotes the complex scalar product. From the unitary invariance of this scalar product, we get $\mathrm{msv}_r(AV) = \mathrm{msv}_r(A)$ for a unitary matrix V of the correct format. Unitary invariance on the right follows now from the invariance under complex conjugation.

Note also that $\mathrm{msv}_n(A) = |\det A|$ for $A \in \mathbb{C}^{n\times n}$. The unitary invariance allows to express the mean square volume of A in terms of the singular values $\sigma_1 \geq \ldots \geq \sigma_p$ of A, $p := \min\{m, n\}$.

It is well known [Golub and Van Loan 1996] that there are unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that $U^*AV = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. Hence we obtain

$$\mathrm{msv}_r^2(A) = \mathrm{msv}_r^2(\mathrm{diag}(\sigma_1, \ldots, \sigma_p)) = \sum_{|I|=r} \prod_{i\in I} \sigma_i^2 \ \geq\ \sigma_1^2\sigma_2^2 \cdots \sigma_r^2, \tag{5}$$

where I runs over all r-subsets of $[p]$. Hence, the square of the r-mean square volume of a matrix is the r-th elementary symmetric polynomial in the squares of its singular values.

Combining the r-volume bound (2) with (4) we obtain the following *mean square volume bound.*

PROPOSITION 3.1. *For a matrix $A \in \mathbb{C}^{m\times n}$ and $r \in \mathbb{N}$ with $1 \leq r \leq \min\{m, n\}$ we have*

$$\mathcal{C}(A) \geq \log \mathrm{msv}_r(A) - \frac{m}{2}. \tag{6}$$

Remark 3.2. The r-volume can be seen as the $1,2$-norm of the map $\Lambda^r A$ induced by A between the exterior algebras $\Lambda^r\mathbb{C}^n$ and $\Lambda^r\mathbb{C}^m$ (see e.g., [Lang 1984] for background on multilinear algebra). Similarly, the mean square volume can be interpreted as the Frobenius norm of $\Lambda^r A$. The unitary invariance of the mean square volume also follows from the fact that Λ^r is equivariant with respect to unitary transformations and that the Frobenius norm is invariant under such.

4. A LOWER BOUND ON CYCLIC CONVOLUTION

In this section we use the mean square volume bound (6) to prove a lower bound on the bilinear map of the cyclic convolution.

Let $f = \sum_{i=0}^{n-1} a_i x^i$ and $g = \sum_{i=0}^{n-1} b_i x^i$ be polynomials in $\mathbb{C}[X]$. The cyclic convolution of f and g is the polynomial $h = \sum_{i=0}^{n-1} c_i x^i$, which is given by the product of f and g in the quotient ring $\mathbb{C}[X]/(X^n - 1)$. Explicitly:

$$c_k = \sum_{i+j\equiv k \bmod n} a_i b_j, \quad 0 \le k < n.$$

Cyclic convolution is a bilinear map on the coefficients. For a fixed polynomial with coefficient vector $a = (a_0, \dots, a_{n-1})$, this map turns into a linear transformation with the circulant matrix

$$\mathrm{Circ}(a) = \begin{pmatrix} a_0 & a_1 & \dots & a_{n-1} \\ a_{n-1} & a_0 & \dots & a_{n-2} \\ \dots & \dots & \dots & \dots \\ a_1 & a_2 & \dots & a_0 \end{pmatrix}.$$

Let $\mathrm{DFT}_n = (\omega^{jk})_{0\le j,k<n}$ be the matrix of the Discrete Fourier Transform, with $\omega = e^{2\pi i/n}$. It is well known [Golub and Van Loan 1996, Sect. 4.7.7] that

$$\mathrm{Circ}(a) = \big(\frac{1}{\sqrt{n}}\mathrm{DFT}_n\big)^{-1} \mathrm{diag}(\lambda_0, \dots, \lambda_{n-1}) \frac{1}{\sqrt{n}}\mathrm{DFT}_n,$$

where the eigenvalues λ_k of $\mathrm{Circ}(a)$ are given by

$$(\lambda_0, \dots, \lambda_{n-1})^\top = \mathrm{DFT}_n (a_0, \dots, a_{n-1})^\top. \tag{7}$$

Hence the singular values of $\mathrm{Circ}(a)$ are $|\lambda_0|, \dots, |\lambda_{n-1}|$ (in some order). Note that $n^{-1/2}\mathrm{DFT}_n$ is unitary.

We recall that the Fast Fourier Transform provides a b.c. bilinear circuit of size $O(n \log n)$ that computes the n-dimensional cyclic convolution. The main result of the paper is the optimality of this algorithm in the b.c. model.

THEOREM 4.1. *The bounded coefficient complexity of the n-dimensional cyclic convolution* conv_n *satisfies* $\mathcal{C}(\mathrm{conv}_n) \ge \frac{1}{12} n \log n - O(n \log\log n)$.

In fact, the proof of the theorem shows that we can replace the constant factor $1/12$ by the slightly larger value 0.086. We state the theorem with $1/12$ for simplicity of exposition.

4.1 Bounding the absolute values of linear forms

To prepare for the proof, we need some lemmas. The idea behind the following lemma is already present in [Raz 2002]. We will identify linear forms on $\mathbb{C}^n$ with vectors in $\mathbb{C}^n$.

Lemma 4.2. *Let $f_1, \dots, f_k \in \mathbb{C}^n$ be linear forms and let $1 \le r < n$. Then there exists a complex subspace $U \subseteq \mathbb{C}^n$ of dimension r such that for a standard Gaussian vector a in U, we have*

$$\mathrm{P}\left[\max_i |f_i(a)| \le 2\sqrt{\ln(4k)}\ \mathrm{rig}_{n-r}(f_1, \dots, f_k)\right] \ge \frac{1}{2}.$$

Proof. Set $R = \mathrm{rig}_{n-r}(f_1, \dots, f_k)$. Then there exists a linear subspace $V \subseteq \mathbb{C}^n$ of dimension $n-r$ such that $\mathrm{dist}(f_i, V) \le R$ for all $1 \le i \le k$. Let f_i' be the projection of f_i along V onto the orthogonal complement $U := V^\perp$ of V. By our choice of the subspace V we have $\|f_i'\| \le R$.

Let $(b_1, \dots, b_n)$ be standard Gaussian in $\mathbb{C}^n$ and a be the orthogonal projection of b onto U along V. Then a is standard Gaussian in U. Moreover, we have $f_i'(b) = f_i(a)$. By Lemma 2.5, the random variable $T = |f_i'(b)|^2/(2\|f_i'\|^2)$ is exponentially distributed with parameter 1.

The assertion now follows from standard large deviations arguments. For any real λ, we have

$$\mathrm{P}[T \ge \lambda] = \mathrm{E}[1_{T \ge \lambda}] \le \mathrm{E}[e^{(T-\lambda)/2}] = e^{-\lambda/2}\mathrm{E}[e^{T/2}].$$

On the other hand,

$$\mathrm{E}[e^{T/2}] = \sum_{k=0}^{\infty} \frac{1}{2^k k!}\mathrm{E}[T^k] = \sum_{k=0}^{\infty} \frac{1}{2^k} = 2,$$

since $\mathrm{E}[T^k] = \int_0^\infty x^k e^{-x} dx = k!$. It follows that

$$\mathrm{P}[T \ge \lambda] = \mathrm{P}\left[|f_i'(b)|^2 \ge 2\lambda\|f_i'\|^2\right] \le 2e^{-\lambda/2}.$$

Since $\|f_i'\| \le R$, we have for a fixed i that

$$\mathrm{P}\left[|f_i(a)| \ge \sqrt{2\lambda}\, R\right] \le 2e^{-\lambda/2}.$$

By the union bound we obtain

$$\mathrm{P}\left[\max_i |f_i(a)| \ge \sqrt{2\lambda}\, R\right] \le 2ke^{-\lambda/2}.$$

Setting $\lambda = 2\ln(4k)$ completes the proof. □

4.2 Proof of the main result

In the next lemma, we state a lower bound on the b.c. linear complexity of a circulant $\mathrm{Circ}(a)$ with standard Gaussian parameter vector a in a subspace of $\mathbb{C}^n$.

Lemma 4.3. *Let $U \subseteq \mathbb{C}^n$ be a subspace of dimension r. For a standard Gaussian vector a in U, we have*

$$\mathrm{P}\left[\mathcal{C}(\mathrm{Circ}(a)) \ge \frac{1}{2} r \log n - cn\right] > \frac{1}{2},$$

where $c = \frac{1}{2}(2+\gamma+\sqrt{2\theta}) \approx 3.73$, and γ, θ are the constants introduced in Section 2.4.

We postpone the proof of this lemma and proceed with the proof of the main theorem.

Proof. (of Theorem 4.1) Let Γ be a b.c. bilinear circuit for conv_n, which computes the linear forms $f_1, \ldots, f_k$ on the first set of inputs. Fix $1 \le r < n$, to be specified later, and set $R = \mathrm{rig}_{n-r}(f_1, \ldots, f_k)$. By Lemma 4.2 and Lemma 4.3 there exists an $a \in \mathbb{C}^n$, such that the following conditions hold:

(1) $\max_{1 \le i \le k} |f_i(a)| \le 2\sqrt{\ln(4k)}\, R$,
(2) $\mathcal{C}(\mathrm{Circ}(a)) \ge \frac{1}{2} r \log n - cn$.

By Lemma 2.4 and the fact that $k \le 3n^3$, we get

$$\mathcal{S}(\Gamma) + n \log(2\sqrt{\ln(12n^3)}\, R) \ge \mathcal{C}(\mathrm{Circ}(a)). \tag{8}$$

On the other hand, the rigidity bound (3) implies the following upper bound on R in terms of $\mathcal{S}(\Gamma)$:

$$\mathcal{S}(\Gamma) \ge \mathcal{C}(f_1, \ldots, f_k) \ge (n - r) \log R.$$

By combining this with (8) and using the second condition above, we obtain

$$\Big(1 + \frac{n}{n-r}\Big)\mathcal{S}(\Gamma) \ge \frac{r}{2} \log n - O(n \log\log n).$$

Setting $\epsilon = r/n$ yields

$$\mathcal{S}(\Gamma) \ge \frac{\epsilon(1-\epsilon)}{2(2-\epsilon)} n \log n - O(n \log\log n).$$

A simple calculation shows that the coefficient of the $n \log n$ term attains the maximum 0.086 for $\epsilon \approx 0.58$. Choosing $\epsilon = 1/2$ for simplicity of exposition finishes the proof. □

Before going into the proof of Lemma 4.3, we provide a lemma on bounding the deviations of products of correlated normal random variables.

LEMMA 4.4. *Let $Z = (Z_1, \ldots, Z_r)$ be a centered Gaussian vector in $\mathbb{C}^r$. Define the complex covariance matrix of Z by $\Sigma_r := (\mathrm{E}(Z_j \overline{Z}_k))_{j,k}$ and put $\delta := 2^{-(\gamma + \sqrt{2\theta})} \approx 0.02$. Then we have $\mathrm{E}(|Z_1|^2 \cdots |Z_r|^2) \ge \det \Sigma_r$ and*

$$\mathrm{P}\left[|Z_1|^2 \cdots |Z_r|^2 \ge \delta^r \det \Sigma_r\right] > \frac{1}{2}.$$

Proof. For proving the bound on the expectation decompose $Z_r = \xi + \eta$ into a component ξ in the span of $Z_1, \ldots, Z_{r-1}$ plus a component η orthogonal to this span in the Hilbert space of quadratic integrable random variables with respect to the inner product defined by the joint probability density of Z. Therefore, $|Z_r|^2 = |\xi|^2 + \xi\overline{\eta} + \overline{\xi}\eta + |\eta|^2$, hence by independence

$$\begin{aligned}\mathrm{E}(|Z_1|^2 \cdots |Z_{r-1}|^2 |Z_r|^2) &= \mathrm{E}(|Z_1|^2 \cdots |Z_{r-1}|^2 |\xi|^2) + \mathrm{E}(|Z_1|^2 \cdots |Z_{r-1}|^2)\, \mathrm{E}(|\eta|^2) \\ &\ge \mathrm{E}(|Z_1|^2 \cdots |Z_{r-1}|^2)\, \mathrm{E}(|\eta|^2).\end{aligned}$$

Let $\xi = \sum_{i<r} \lambda_i Z_i$. Then the complex covariance matrix Σ_r' of $(Z_1, \ldots, Z_{r-1}, \eta)$ arises from Σ_r by subtracting the $\overline{\lambda}_i$-th multiple of the i-th column from the r-th

column, and by subtracting the λ_j-th multiple of the j-th row from the r-th row, for all $i, j < r$. Therefore, using $\mathrm{E}(Z_i \overline{\eta}) = 0$, we obtain

$$\det \Sigma_r = \det \Sigma_r' = \det \Sigma_{r-1} \, \mathrm{E}(|\eta|^2).$$

The desired bound on the expectation $\mathrm{E}(|Z_1|^2 \cdots |Z_r|^2) \geq \det \Sigma_r$ thus follows by induction on r. Noting that $\mathrm{E}(|Z_r|^2) \geq \mathrm{E}(|\eta|^2)$, we also conclude from the above equation that

$$\mathrm{E}(|Z_1|^2) \cdots \mathrm{E}(|Z_r|^2) \geq \det \Sigma_r. \tag{9}$$

In order to prove the probability estimate for the random product $|Z_1|^2 \cdots |Z_r|^2$, we first transform the product into a sum by taking logarithms. For every $\epsilon > 0$ Chebychev's inequality yields the bound

$$\mathrm{P}\Big[\frac{1}{r}\Big|\sum_{j=1}^{r}(\log|Z_j|^2 - \mathrm{E}[\log|Z_j|^2])\Big| \geq \epsilon\Big] \leq \frac{\mathrm{Var}(\sum_{j=1}^{r} \log|Z_j|^2)}{\epsilon^2 r^2}. \tag{10}$$

For the variance we have by Lemma 2.6

$$\begin{aligned}\mathrm{Var}(\sum_{j=1}^{r} \log|Z_j|^2) &= \sum_{j,k} \mathrm{Cov}(\log|Z_j|^2, \log|Z_k|^2) \\ &\leq \sum_{j,k} \sqrt{\mathrm{Var}(\log|Z_j|^2)\mathrm{Var}(\log|Z_k|^2)} \leq r^2\theta.\end{aligned}$$

Setting $\epsilon^2 = 2\theta$ in this equation and after exponentiating in (10) we obtain

$$\mathrm{P}\left[|Z_1|^2 \cdots |Z_r|^2 \leq 2^{-\epsilon r + \sum_{j=1}^{r} \mathrm{E}[\log|Z_j|^2]}\right] \leq \frac{1}{2}. \tag{11}$$

By combining the bound (9) with Lemma 2.6 we get

$$\log\det\Sigma_r \leq \sum_{i=1}^{r} \log \mathrm{E}[|Z_i|^2] \leq \gamma r + \sum_{i=1}^{r} \mathrm{E}[\log|Z_i|^2].$$

Hence we conclude from (11) that

$$\mathrm{P}\left[|Z_1|^2 \cdots |Z_r|^2 \leq 2^{-(\epsilon+\gamma)r} \det \Sigma_r\right] \leq \frac{1}{2},$$

from which the lemma follows. □

Proof. (of Lemma 4.3) By equation (7) we have $\lambda = \mathrm{DFT}_n a$ and the singular values of the circulant $\mathrm{Circ}(a)$ are given by the absolute values of the components of λ. Setting

$$\alpha = n^{-1/2}\lambda = n^{-1/2}\mathrm{DFT}_n a,$$

we obtain for the r-mean square volume by (5)

$$\mathrm{msv}_r^2(\mathrm{Circ}(a)) = n^r \sum_{|I|=r} \prod_{i \in I} |\alpha_i|^2. \tag{12}$$

Now let a be a standard Gaussian vector in the subspace U of dimension r. Let W be the image of U under the unitary transformation $n^{-1/2}\mathrm{DFT}_n$. As a unitary transformation of a, α is standard Gaussian in the subspace W (cf. Section 2.3). This means that there is an orthonormal basis $b_1, \dots, b_r$ of W such that

$$\alpha = \beta_1 b_1 + \cdots + \beta_r b_r,$$

where (β_i) is standard Gaussian in $\mathbb{C}^r$. Let $B \in \mathbb{C}^{n\times r}$ denote the matrix with the columns $b_1, \dots, b_r$ and let B_I be the submatrix of B consisting of the rows indexed by I, for $I \subseteq [n]$ with $|I| = r$. Setting $\alpha_I = (\alpha_i)_{i\in I}$ we have $\alpha_I = B_I\beta$. The complex covariance matrix of α_I is given by $\Sigma := E[\alpha_I\alpha_I^*] = B_I B_I^*$, hence

$$\det\Sigma = |\det B_I|^2.$$

We remark that $|\det B_I|^2$ can be interpreted as the volume contraction ratio of the projection $\mathbb{C}^n \to \mathbb{C}^I, \alpha \mapsto \alpha_I$ restricted to W. For later purposes we also note that $\mathrm{E}(|\alpha_i|^2) = \sum_j |B_{ij}|^2 \le 1$.

By the Binet-Cauchy formula and the orthogonality of the basis (b_i) we get

$$\sum_{|I|=r} |\det B_I|^2 = \det\left(\langle b_i, b_j\rangle\right)_{1\le i,j\le r} = 1.$$

Therefore, we can choose an index set I such that

$$|\det B_I|^2 \ge \binom{n}{r}^{-1} \ge 2^{-n}.$$

By applying Lemma 4.4 to the random vector α_I and using (12), we get that with probability at least 1/2,

$$\mathrm{msv}_r^2(\mathrm{Circ}(a)) \ge n^r\delta^r \det\Sigma \ge n^r\delta^r 2^{-n}, \tag{13}$$

where $\delta = 2^{-(\gamma+\sqrt{2\theta})}$. The mean square volume bound (6) implies that

$$\mathcal{C}(\mathrm{Circ}(a)) \ge \log \mathrm{msv}_r(\mathrm{Circ}(a)) - \frac{n}{2} \ge \frac{1}{2} r\log n - \frac{1}{2}(2 + \log\delta^{-1})n,$$

with probability at least 1/2. This proves the lemma. □

5. MULTIPLICATION AND DIVISION OF POLYNOMIALS

By reducing the cyclic convolution to several other important computational problems, we are going to derive lower bounds of order $n\log n$ for these problems. These bounds are optimal up to a constant factor.

5.1 Polynomial multiplication

Let $f = \sum_{i=0}^{n-1} a_i x^i$, $g = \sum_{i=0}^{n-1} b_i x^i$ be polynomials in $\mathbb{C}[X]$ and $fg = \sum_{i=0}^{2n-2} c_i x^i$. Clearly, we can obtain the coefficients of the cyclic convolution of f and g by adding c_k to c_{k+n} for $0 \le k < n$. This observation and Theorem 4.1 immediately imply the following corollary.

COROLLARY 5.1. *The bounded coefficient complexity of the multiplication of polynomials of degree less than n is at least $\frac{1}{12} n\log n - O(n\log\log n)$.*

5.2 Division with remainder

We will first derive a lower bound on the inversion of power series mod X^{n+1} and then use this to get a lower bound for the division of polynomials.

Let $\mathbb{C}[[X]]$ denote the ring of formal power series in the variable X. We will study the problem to compute the first n coefficients $b_1, , \dots, b_n$ of the inverse in $\mathbb{C}[[X]]$

$$f^{-1} = 1 + \sum_{k=1}^{\infty} b_k X^k$$

of the polynomial $f = 1 - \sum_{i=1}^{n} a_i X^i$ given by the coefficients a_i. We remark that the b_k are polynomials in the a_i, which are recursively given by

$$b_0 := 1, \quad b_k = \sum_{i=0}^{k-1} a_{k-i} b_i.$$

Note that the problem to invert power series is not bilinear. [Sieveking 1972] and [Kung 1974] designed a b.c. circuit of size $O(n \log n)$ solving this problem.

We now prove a corresponding lower bound on the b.c. complexity of this problem by reducing polynomial multiplication to the problem to invert power series.

Theorem 5.2. *The map assigning to $a_1, \dots, a_n$ the first n coefficients $b_1, \dots, b_n$ of the inverse of $f = 1 - \sum_{i=1}^{n} a_i X^i$ in the ring of formal power series has bounded coefficient complexity greater than $\frac{1}{324} n \log n - O(n \log\log n)$.*

Proof. Put $g = \sum_{i=1}^{n} a_i X^i$. The equation

$$1 + \sum_{k=1}^{\infty} b_k X^k = \frac{1}{1-g} = \sum_{k=0}^{\infty} g^k.$$

shows that g^2 is the homogeneous quadratic part of $\sum_{k=1}^{\infty} b_k X^k$ in the variables a_i.

Let Γ be an optimal b.c. circuit computing $b_1, \dots, b_n$. According to the proof in [Bürgisser et al. 1997, Theorem 7.1], there is a b.c. circuit of size at most $9\,\mathcal{S}(\Gamma)$ computing the homogeneous quadratic parts of the $b_1, \dots, b_n$ with respect to the variables a_i. This leads to a b.c. circuit of size at most $9\,\mathcal{S}(\Gamma)$ computing the coefficients of the squared polynomial g^2.

Now let $m := \lfloor n/3 \rfloor$, and assume that $g = g_1 + X^{2m} g_2$ with g_1, g_2 of degree smaller than m. Then

$$g^2 = g_1^2 + 2 g_1 g_2 X^{2m} + g_2^2 X^{4m},$$

By the assumption on the degrees we have no "carries" and we can therefore find the coefficients of the product polynomial $g_1 g_2$ among the middle terms of g^2. Thus we obtain a b.c. circuit for the multiplication of polynomials of degree $m - 1$. The theorem now follows from Corollary 5.1. □

We now show how to reduce the inversion of power series to the problem of dividing polynomials with remainder. The reduction in the proof of the following corollary is from [Strassen 1973a], see also [Bürgisser et al. 1997, Section 2.5].

Corollary 5.3. *Let f, g be polynomials with $n = \deg f \geq m = \deg g$ and g be monic. Let q be the quotient and r be the remainder of f divided by g, so that*

$f = qg + r$ and $\deg r < \deg g$. The map assigning to the coefficients of f and g the coefficients of the quotient q and the remainder r has bounded coefficient complexity at least $\frac{1}{324} n \log n - O(n \log\log n)$.

Proof. Dividing $f = X^{2n}$ by $g = \sum_{i=0}^{n} a_i X^{n-i}$, where $a_0 = 1$, we obtain:

$$X^{2n} = \Big(\sum_{i=0}^{n} q_i X^i\Big)\Big(\sum_{i=0}^{n} a_i X^{n-i}\Big) + \sum_{i=0}^{n-1} r_i X^i.$$

By substituting X with $1/X$ in the above equation and multiplying with X^{2n}, we get

$$1 = \Big(\sum_{i=0}^{n} q_i X^{n-i}\Big)\Big(\sum_{i=0}^{n} a_i X^i\Big) + \sum_{i=0}^{n-1} r_i X^{2n-i}.$$

Since the remainder is now a multiple of X^{n+1}, we get

$$\Big(\sum_{i=0}^{n} a_i X^i\Big)^{-1} \equiv \Big(\sum_{i=0}^{n} q_i X^{n-i}\Big) \bmod X^{n+1}.$$

From this we see that the coefficients of the quotient are precisely the coefficients of the inverse mod X^{n+1} of $\sum_{i=0}^{n} a_i X^i$ in the ring of formal power series, and the proof is finished. □

6. UNBOUNDED SCALAR MULTIPLICATIONS

We extend our model of computation by allowing some instructions corresponding to scalar multiplications with constants of absolute value greater than two, briefly called *help gates* in the sequel. If there are at most h help gates allowed, we denote the corresponding bounded coefficient complexity by the symbol $\mathcal{C}_h$.

We are going to show that our proof technique is robust in the sense that it still allows to prove $n \log n$ lower bounds if the number of help gates is restricted to $(1-\epsilon)n$ for fixed $\epsilon > 0$.

6.1 Extension of the mean square volume bound

As a first step we extend the mean square volume bound (5) and (6) for dealing with help gates.

Proposition 6.1. *Assume $A \in \mathbb{C}^{m \times n}$ has the singular values $\sigma_1 \geq \ldots \geq \sigma_p$, where $p := \min\{m, n\}$. For all integers s, h with $1 \leq s \leq p - h$ we have*

$$\mathcal{C}_h(A) \geq \sum_{i=h+1}^{h+s} \log \sigma_i - \frac{m}{2} + h \geq s \log \sigma_{h+s} - \frac{m}{2} + h.$$

Proof. Let Γ be a b.c. circuit with at most h help gates, which computes the linear map corresponding to A. Without loss of generality, we may assume that Γ has exactly h help gates. Let g_i, $i \in I$, be the linear forms computed at the help gates of Γ. We transform the circuit Γ into a b.c. circuit Γ' by replacing each help gate with a multiplication by zero. This new circuit is obviously a b.c. circuit of size

$\mathcal{S}(\Gamma') = \mathcal{S}(\Gamma) - h$, computing a linear map corresponding to a matrix $B \in \mathbb{C}^{m\times n}$. The linear maps corresponding to A and B coincide on the orthogonal complement of $\mathrm{span}\{g_i \mid i \in I\}$ in $\mathbb{C}^m$, therefore $B = A + E$ for a matrix E of rank at most h. From the perturbation inequality (1) we obtain that

$$\sigma_i(B) \geq \sigma_{i+h}(A) \quad \text{for } i \leq p - h.$$

By (5) this implies for $s \leq p - h$ that

$$\mathrm{msv}_s^2(B) \geq \sum_{0<i_1<\cdots<i_s\leq p-h} \sigma_{i_1}^2(B)\cdots\sigma_{i_s}^2(B) \geq \sum_{h<i_1<\cdots<i_s\leq p} \sigma_{i_1}^2(A)\cdots\sigma_{i_s}^2(A).$$

On the other hand, by the mean square volume bound (6) we have

$$\mathcal{S}(\Gamma) - h = \mathcal{S}(\Gamma') \geq \log \mathrm{msv}_s(B) - \frac{m}{2}.$$

Combining the last two estimates completes the proof. □

Remark 6.2. 1. Proposition 6.1 implies that $\mathcal{C}_{(1-\epsilon)n}(\mathrm{DFT}_n) \geq \epsilon(\frac{1}{2}n\log n - n)$ for the Discrete Fourier Transform DFT_n, provided $0 < \epsilon \leq 1$.

2. Note that the number h of help gates may be replaced by the dimension of the subspace spanned by the linear functions computed at the help gates.

3. Proposition 6.1 can be seen as a variant of the spectral lemma in [Chazelle 1998]. Using entropy considerations, Chazelle obtained the slightly worse lower bound $\Omega((r-2h)\log\sigma_r)$ for the b.c. complexity of a matrix $A \in \mathbb{R}^{n\times n}$ with at most h help gates. While this allows to handle at most $n/2$ help gates, Chazelle's result is stronger in the sense that it involves a more general notion of help gates, which are allowed to compute *any* function of the previous intermediate results.

6.2 Extremal values of Gaussian random vectors

In this section we derive the following auxiliary result about the distribution of the maximal absolute value of the components of a Gaussian random vector.

LEMMA 6.3. *1. A centered Gaussian random vector $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$ with $\max_i \mathrm{E}(X_i^2) \leq 1$ satisfies for any $\epsilon > 0$*

$$\lim_{n\to\infty} \mathrm{P}\left[\max_i |X_i| > \sqrt{2\ln n} + \epsilon\right] = 0.$$

2. A centered Gaussian random vector $(Z_1, \ldots, Z_n)$ in $\mathbb{C}^n$ with $\max_i \mathrm{E}(|Z_i|^2) \leq 1$ satisfies for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[\max_i |Z_i| > 2\sqrt{\ln(2n)} + \epsilon\right] = 0.$$

Proof. 1. Since X is centered we have for any $u \in \mathbb{R}$

$$\mathrm{P}\left[\max_i |X_i| \geq u\right] \leq \mathrm{P}\left[\max_i X_i \geq u\right] + \mathrm{P}\left[\max_i(-X_i) \geq u\right] \leq 2\mathrm{P}\left[\max_i X_i \geq u\right].$$

For proving the first assertion it is therefore sufficient to show that for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[\max_i X_i > \sqrt{2\ln n} + \epsilon\right] = 0. \tag{14}$$

For this we may assume that the components of X are uncorrelated. In fact, Slepian's inequality (see [Ledoux and Talagrand 1991]) implies that for centered Gaussian vectors $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ we have

$$\mathrm{P}\left[\max_i X_i \le u\right] \le \mathrm{P}\left[\max_i Y_i \le u\right]$$

provided $\mathrm{E}(X_i^2) = \mathrm{E}(Y_i^2)$ and $\mathrm{E}(X_iX_j) \le \mathrm{E}(Y_iY_j)$ for all i, j.

We may also assume that all the X_i have variance 1 since the distribution function

$$F_\sigma(u) := \frac{1}{\sigma\sqrt{2\pi}} \int_\infty^u \exp(-\frac{t^2}{2\sigma^2})dt.$$

of a centered normal random variable with variance $\sigma^2 \le 1$ satisfies $F_1(u) \le F_\sigma(u)$ for all $u \ge 0$. Hence, if X is a Gaussian vector with uncorrelated components X_i of variance $\sigma_i^2 \le 1$, we have

$$F_1(u)^n \le \prod_{i=1}^n F_{\sigma_i}(u) = \mathrm{P}\left[\max_i X_i \le u\right].$$

In the case where $X_1, \dots, X_n$ are independent and standard normal distributed we have according to [Cramér 1946] that

$$\mathrm{E}(\max_i X_i) = \sqrt{2\ln n} + o(1), \quad \mathrm{Var}(\max_i X_i) = \frac{\pi^2}{12}\frac{1}{\ln n}(1 + o(1)), \quad n \to \infty$$

and Claim (14) follows from Chebychev's inequality.

2. The second assertion follows from the first one applied to the Gaussian vector W with values in $\mathbb{R}^{2n}$ given by the real and imaginary parts of the Z_i (in some order). Note that $\max_{1\le i\le n} |Z_i| \le \sqrt{2}\max_{1\le j\le 2n} |W_j|$. □

6.3 Cyclic convolution and help gates

Our goal is to prove the following extension of Theorem 4.1.

THEOREM 6.4. *The bounded coefficient complexity with at most* $(1-\epsilon)n$ *help gates of the* n*-dimensional cyclic convolution* conv_n *is at least* $\Omega(n\log n)$ *for fixed* $0 < \epsilon \le 1$.

The proof follows the same line of argumentation as in Section 4. We first state and prove an extension of Lemma 4.3.

LEMMA 6.5. *Let* $U \subseteq \mathbb{C}^n$ *be a subspace of dimension* r *and* $h \in \mathbb{N}$ *with* $h < r$. *For a standard Gaussian vector* a *in* U, *we have*

$$\mathrm{P}\left[\mathcal{C}_h(\mathrm{Circ}(a)) \ge \frac{1}{2}(r-h)\log n - n(c + \log\log n)\right] > \frac{1}{2},$$

for some constant $c > 0$.

Proof. As in the proof of Lemma 4.3 we assume that the random vector $\alpha = n^{-1/2}\mathrm{DFT}_n a$ is standard Gaussian with values in some r-dimensional subspace W. Recall that $\sqrt{n}\,|\alpha_i|$ are the singular values of $\mathrm{Circ}(a)$. We denote by $|\alpha^{(1)}| \ge$

$\ldots \geq |\alpha^{(n)}|$ the components of α with decreasing absolute values. In particular, $|\alpha^{(1)}| = \max_i |\alpha^{(i)}|$. Proposition 6.1 implies that

$$\begin{aligned}\mathcal{C}_h(\mathrm{Circ}(a)) \;&\geq\; \sum_{i=h+1}^{r} \log(\sqrt{n}\,|\alpha^{(i)}|) - \frac{n}{2} + h \\ &= \frac{1}{2}(r-h)\log n + \log\left(\prod_{i=h+1}^{r} |\alpha^{(i)}|\right) - \frac{n}{2} + h.\end{aligned}$$

In the proof of Lemma 4.3 (13) we showed that $\mathrm{msv}_r^2(\mathrm{Circ}(a)) \geq n^r \delta^r 2^{-n}$ with probability at least 1/2. In the same way, one can show that with probability at least 3/4 we have $\mathrm{msv}_r^2(\mathrm{Circ}(a)) \geq n^r c_1^n$ for some fixed constant $c_1 > 0$. From the estimate

$$\sum_{|I|=r} \prod_{i\in I} |\alpha_i|^2 \;\leq\; 2^n \prod_{i=1}^{r} |\alpha^{(i)}|^2$$

we thus obtain that $\prod_{i=1}^{r} |\alpha^{(i)}|^2 \geq (c_1/2)^n$ with probability at least 3/4.

By applying Lemma 6.3 to the centered Gaussian random variable α we obtain that with probability at least 3/4

$$\max_i |\alpha^{(i)}|^2 = |\alpha^{(1)}|^2 \leq c_2 \log n$$

for some fixed constant $c_2 > 0$. (Recall that $\mathrm{E}(|\alpha^{(i)}|^2) \leq 1$.)

Altogether, we obtain that with probability at least 1/2 we have

$$\prod_{i=h+1}^{r} |\alpha^{(i)}|^2 \;\geq\; \frac{\prod_{i=1}^{r} |\alpha^{(i)}|^2}{|\alpha^{(1)}|^{2h}} \;\geq\; \left(\frac{c_1}{2c_2\log n}\right)^n.$$

This completes the proof of the lemma. □

Proof. (of Theorem 6.4) Let Γ be a b.c. bilinear circuit computing conv_n using at most $h \leq (1-\epsilon)n$ help gates, $0 < \epsilon \leq 1$. Referring to the partition of instructions in Definition 2.2, we assume that $\Gamma^{(1)}$ uses h_1 help gates, and that $\Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)}$ use a total of h_2 help gates. Thus $h_1 + h_2 = h$. Let $f_1, \ldots, f_k$ denote the linear forms computed by $\Gamma^{(1)}$.

Assume $h_2 < r < n - h_1$ and set $R = \mathrm{rig}_{n-r}(f_1, \ldots, f_k)$. By Lemma 4.2 and Lemma 6.5 there exists an $a \in \mathbb{C}^n$, such that the following conditions hold:

(1) $\max_{1\leq i\leq k} \log|f_i(a)| \leq \log(2\sqrt{\ln(4k)}\,R) \leq \log R + O(\log\log n)$,
(2) $\mathcal{C}_{h_2}(\mathrm{Circ}(a)) \geq \frac{1}{2}(r-h_2)\log n - O(n\log\log n)$.

On the other hand, by Proposition 6.1 and using $\sigma_{n-r}(f_1, \ldots, f_k) \geq R$, we get

$$\mathcal{S}(\Gamma) \geq \mathcal{C}_{h_1}(f_1, \ldots, f_k) \geq (n - r - h_1)\log R - \frac{k}{2}.$$

The proof of Lemma 2.4 shows that

$$\mathcal{S}(\Gamma) + n \max_{1\leq i\leq k} \log|f_i(a)| \;\geq\; \mathcal{C}_{h_2}(\mathrm{Circ}(a)).$$

By combining all this we obtain

$$\Big(1+\frac{n}{n-r-h_1}\Big)\mathcal{S}(\Gamma)+\frac{nk}{2(n-r-h_1)}+O(n\log\log n)\geq\frac{1}{2}(r-h_2)\log n.$$

We set now $r:=\lfloor(h_2+n-h_1)/2\rfloor$. Then $r+h_1\leq(1-\frac{\epsilon}{2})n$ and $r-h_2\geq\frac{\epsilon}{2}\,n-1$. By plugging this into the above inequality we obtain

$$\frac{\epsilon+2}{\epsilon}\,\mathcal{S}(\Gamma)+\frac{k}{\epsilon}+O(n\log\log n)\geq\frac{\epsilon}{4}\,n\log n.$$

Let $\kappa:=\frac{\epsilon^2}{8}$. If $k\leq\kappa n\log n+n$, then $\mathcal{S}(\Gamma)\geq\frac{\epsilon^2}{8(\epsilon+2)}\,n\log n-O(n\log\log n)$. On the other hand, if $k>\kappa n\log n+n$, then trivially

$$\mathcal{S}(\Gamma)\geq\mathcal{C}_{h_1}(f_1,\ldots,f_k)\geq k-n\geq\kappa n\log n.$$

This completes the proof of the theorem. □

ACKNOWLEDGMENTS

We are indebted to Joachim von zur Gathen for bringing the paper [Raz 2002] to our attention. We are grateful to Satyanarayana Lokam for suggesting to extend our lower bounds involving help gates from the linear to the bilinear case. We thank Tom Schmitz and Mario Wschebor for useful discussions about probability. Special thanks go to Ran Raz for enlightening comments and suggestions. Thanks also go to the referees for useful comments.

REFERENCES

Baur, W. and Strassen, V. 1983. The complexity of partial derivatives. *Theoret. Comp. Sci. 22*, 317–330.

Bellman, R. 1997. *Introduction to matrix analysis.* SIAM, Philadelphia, PA.

Bürgisser, P., Clausen, M., and Shokrollahi, M. 1997. *Algebraic Complexity Theory.* Grundlehren der mathematischen Wissenschaften, vol. 315. Springer Verlag.

Chazelle, B. 1998. A spectral approach to lower bounds with applications to geometric searching. *SIAM Journal on Computing 27(2)*, 545–556.

Courant, R. and Hilbert, D. 1931. *Methoden der mathematischen Physik. I.* Springer-Verlag, Berlin. Zweite Auflage.

Cramér, H. 1946. *Mathematical Methods of Statistics.* Princeton Mathematical Series, vol. 9. Princeton University Press.

Feller, W. 1971. *An introduction to probability theory and its applications.* Vol. 2. John Wiley & Sons.

Golub, G. H. and Van Loan, C. 1996. *Matrix Computations.* The John Hopkins University Press, Baltimore.

Kung, H. 1974. On computing reciprocals of power series. *Num. Math. 22*, 341–348.

Lang, S. 1984. *Algebra*, Second ed. Addison-Wesley.

Ledoux, M. and Talagrand, M. 1991. *Probability in Banach Spaces.* Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, vol. 23. Springer Verlag.

Lokam, S. 1995. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *Proc. 36th FOCS.* 6–15.

Morgenstern, J. 1973. Note on a lower bound of the linear complexity of the fast Fourier transform. *J. ACM 20*, 305–306.

Morgenstern, J. 1975. The linear complexity of computation. *J. ACM 22*, 184–194.

Nisan, N. and Wigderson, A. 1995. On the complexity of bilinear forms. In *Proc. of the 27th ACM Symposium on the Theory of Computing*. 723–732.

Pudlák, P. 1998. A note on the use of determinant for proving lower bounds on the size of linear circuits. *ECCC Report 42*.

Pudlák, P. 2003. Private communication.

Raz, R. 2002. On the complexity of matrix product. In *Proc. 34th STOC*. 144–151. Also available as ECCC Report 12, 2002.

Raz, R. 2003. Private communication.

Sieveking, M. 1972. An algorithm for division of power series. *Computing 10*, 153–156.

Strassen, V. 1973a. Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. *Num. Math. 20*, 238–251.

Strassen, V. 1973b. Vermeidung von Divisionen. *Crelles J. Reine Angew. Math. 264*, 184–202.

Valiant, L. 1976. Graph theoretic properties in computational complexity. *J. Comp. Syst. Sci. 13*, 278–285.

Valiant, L. 1977. Graph theoretic arguments in low-level complexity. Number 53 in LNCS. Springer Verlag, 162–176.