

Lower Bounds on the Bounded Coefficient Complexity of Bilinear Maps^{*}

Peter Bürgisser

Martin Lotz

Department of Mathematics and Computer Science

University of Paderborn

33095 Paderborn, Germany

{pbuerg,lotzm}@math.uni-paderborn.de

Abstract

We prove lower bounds of order $n \log n$ for both the problem to multiply polynomials of degree n , and to divide 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 to multiply 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.

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 J. Morgenstern [9, 10], where he 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, he derived the lower bound $\frac{1}{2}n \log n$ for computing the Discrete Fourier Transform. L. Valiant [17, 18] 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.

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. B. Chazelle [3] also motivated the b.c. model as a natural model of computation. His argument is that the finite representation of numbers is essentially equivalent to bounded coefficients.

Several papers [11, 8, 12] provided size-depth trade-offs for b.c. arithmetic circuits. The concept of matrix rigidity, originally introduced by Valiant [18], 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 by Lokam [8].

Chazelle [3] 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.

R. Raz [13] recently proved a nonlinear lower bound on the complexity of matrix multiplication in the b.c. model. To our knowledge, this paper and the paper by Nisan and Wigderson [11] are the only ones which deal with the complexity of bilinear maps in the b.c. model of computation.

The main result of this paper 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 uses the idea of Raz [13] 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 2.1). 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 4.1).

In the case of matrix multiplication, Raz [13] proceeded by again applying a geometric rigidity bound to the result-

^{*}To appear in Proc. 43rd FOCS 2002, Nov. 16-19, Vancouver, ©IEEE

ing linear problem via the Hoffman-Wielandt inequality. This approach does not seem to yield good enough bounds in our situation, where we have to estimate the complexity of structured random matrices; in the case of the convolution these are circulant matrices.

Instead, we treat the arising linear problem 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 essentially provides a lower bound on the b.c. complexity of the matrix A (Theorem 3.1). This result contains a restricted version of Chazelle's Spectral Lemma [3] as a special case.

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 (see, e.g., [2]) 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 models of computation and recall some previously known results on lower bounds for b.c. linear circuits. We also recall some facts on complex random variables. In Section 3 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. However, the proof of a key technical lemma is postponed to Section 5. In Section 6, known reductions are applied to derive lower bounds on polynomial multiplication, inversion of power series and division with remainder.

1.2 Acknowledgments

The authors would like to thank Joachim von zur Gathen for bringing the paper [13] to their attention. The second author thanks Tom Schmitz at ETH Zürich for some useful probabilistic advice.

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 [2]. By a result of V. Strassen [16], we may exclude divisions without loss of generality.

Definition 2.1. A *straight-line program* Γ expecting inputs of length n is a sequence $(\Gamma_1, \dots, \Gamma_r)$ of instructions $\Gamma_k = (\omega_k; i_k, j_k)$, $\omega_k \in \{*, +, -\}$ or $\Gamma_k = (\omega_k; i_k)$, $\omega_k \in \mathbb{C}$ with integers i_k, j_k satisfying $-n < i_k, j_k < k$. A sequence of polynomials b_{-n+1}, \dots, b_r is called the *result sequence* of Γ on input variables a_1, \dots, a_n , if for $-n < k \leq 0$, $b_k = a_{n+k}$, and for $1 \leq k \leq r$, $b_k = b_{i_k} \omega_k b_{j_k}$ if $\Gamma_k = (\omega_k; i_k, j_k)$ and $b_k = \omega_k b_{i_k}$ if $\Gamma_k = (\omega_k; i_k)$, $\omega_k \in \mathbb{C}$. Γ is said to *compute* a set of polynomials F on input a_1, \dots, 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 $\Gamma = (\Gamma_1, \dots, \Gamma_r)$ expecting inputs X_1, \dots, X_n is called a *linear circuit*, if $\omega_k \in \{+, -\}$ for every instruction $\Gamma_k = (\omega_k; i_k, j_k)$, or $\omega_k \in \mathbb{C}$ if the instruction is of the form $(\omega_k; i_k)$. A circuit expecting inputs $X_1, \dots, X_m, Y_1, \dots, Y_n$ 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 $(*; i, j)$, with $\Gamma_i \in \Gamma^{(1)}$ and $\Gamma_j \in \Gamma^{(2)}$,
4. $\Gamma^{(4)}$ is a linear circuit with the previously computed results as inputs.

In other words, $\Gamma^{(1)}$ and $\Gamma^{(2)}$ compute linear functions f_1, \dots, f_k in the X_i and g_1, \dots, g_k 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. [2, 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 $\varphi: \mathbb{C}^m \times \mathbb{C}^n \rightarrow \mathbb{C}^p$ 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 $\varphi^A: \mathbb{C}^n \rightarrow \mathbb{C}^m$ (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. (There is no serious danger of confusion arising from this, since these complexity notions differ at most by a constant factor.)

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

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 [13].

Lemma 2.1. *Let $\varphi: \mathbb{C}^m \times \mathbb{C}^n \rightarrow \mathbb{C}^p$ be a bilinear map and Γ be a b.c. bilinear circuit computing φ . If f_1, \dots, f_k are the linear maps computed by the circuit on the first set of inputs, then for all $a \in \mathbb{C}^m$:*

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

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 with $\gamma/2$ by 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 $\varphi(a, -): \mathbb{C}^n \rightarrow \mathbb{C}^p$. Since there are p outputs, the size increases by at most $p \log \gamma$. \square

2.2 Singular values and geometric rigidity

Following Raz [13] we define a geometric notion of rigidity of a matrix as follows. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with column vectors $a_i \in \mathbb{C}^m$ and $r \in \mathbb{N}$. The *(geometric) rigidity* of a matrix $A \in \mathbb{C}^{m \times n}$ is defined as

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

where the minimum is taken over all complex subspaces $V \subseteq \mathbb{C}^m$ with dimension r and dist denotes the usual euclidean distance, i.e., $\text{dist}(a, V) = \min_{v \in V} \|a - v\|$.

This notion is closely related to the singular values of the matrix A . For the following definition see [5].

Definition 2.4. Let $A \in \mathbb{C}^{m \times n}$ and $\lambda_1 \geq \dots \geq \lambda_m \geq 0$ be the eigenvalues of the hermitian matrix AA^* . The *singular values* of A are defined as $\sigma_k = \sqrt{\lambda_k}$ for $1 \leq k \leq p = \min\{m, n\}$.

In [5, Thm. 2.5.3] the following characterization of the singular values of A is given:

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

where the matrix norm $\|\cdot\|_2$ denotes the 2-norm.

We show now that the r -geometric rigidity is equal to the $(r+1)$ -th singular value up to a small constant factor.

Lemma 2.2. *For $A \in \mathbb{C}^{m \times n}$ and $r \in \mathbb{N}$ we have*

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

Proof. Assume according to (1) that $\sigma_{r+1}(A) = \|A - B\|_2$ with a matrix B of rank at most r . Let V be the subspace spanned by the column vectors b_i of B . Then we have

$$\|A - B\|_2 \geq \max_i \|a_i - b_i\| \geq \max_i \text{dist}(a_i, V) \geq \text{rig}_r(A).$$

This shows the right-hand inequality.

Assume now $\text{rig}_r(A) = \max_i \text{dist}(a_i, V)$ for a subspace V of dimension r . Choose $b_i \in V$ such that $\text{dist}(a_i, V) = \|a_i - b_i\|$ and consider the matrix B of rank at most r with column vectors b_i . Then

$$\max_i \text{dist}(a_i, V) = \max_i \|a_i - b_i\| \geq \frac{1}{\sqrt{n}} \|A - B\|_2,$$

where the rightmost estimate follows from

$$\max_i \|a_i - b_i\|^2 \geq \frac{1}{n} \sum_i \|a_i - b_i\|^2 \geq \frac{1}{n} \|A - B\|_2^2.$$

\square

To explain the naming of the geometric rigidity, we remark that if we replace in (1) the 2-norm by the following matrix norm (a_i are the columns of A)

$$\|A\| := \max_{1 \leq i \leq n} \|a_i\|,$$

then we just obtain $\text{rig}_r(A)$. On the other hand, if we replace in (1) the 2-norm by the the Hamming metric, then we get the usual matrix rigidity as introduced by Valiant [18]. For similar reasonings we refer to Lokam [8].

2.3 Complex normal random variables

A random vector $X = (X_1, \dots, X_n)$ in \mathbb{R}^n is called (multivariate) standard normal distributed 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 normal distributed.

Throughout this paper, we will be working with random vectors Z assuming complex 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) normal distributed.

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 normal distributed in U* iff for some orthonormal basis b_1, \dots, b_r of U we have $Z = \sum_j \zeta_j b_j$, where the random vector (ζ_j) of the components is standard normal distributed 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 complex normal distributed vector with a unitary matrix is again standard normal distributed, since a unitary transformation $\mathbb{C}^r \rightarrow \mathbb{C}^r$ induces an orthogonal transformation $\mathbb{R}^{2r} \rightarrow \mathbb{R}^{2r}$.

The easy proof of the following lemma is left to the reader.

Lemma 2.3. *Let (Z_1, \dots, Z_n) be standard normal distributed in \mathbb{C}^n . Consider a complex linear combination $S = f_1 Z_1 + \dots + f_n Z_n$ with $f = (f_1, \dots, f_n) \in \mathbb{C}^n$. Then the real and imaginary parts of S are independent and normal distributed, each with mean 0 and variance $\|f\|^2$. Moreover, $T := |S|^2/2\|f\|^2$ 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 U are both equal to 1.*

2.4 Two useful inequalities

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

uating the corresponding integrals yields

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

Proposition 2.1. *Let Z be a complex random variable, which is centered and normal distributed. Then*

$$0 \leq \log \mathbb{E}[|Z|^2] - \mathbb{E}[\log |Z|^2] \leq \gamma, \quad \text{Var}(\log |Z|^2) \leq \theta.$$

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 $\Delta := \log \mathbb{E}[|Z|^2] - \mathbb{E}[\log |Z|^2]$ is non-negative, since \log is concave (Jensen's inequality). By linearity of the mean, Δ as well as $\text{Var}(\log |Z|^2)$ are invariant under multiplication of Z with scalars. We may therefore w.l.o.g. assume that $1 = \lambda_1 \geq \lambda_2$. From this we see that

$$\begin{aligned} \log \mathbb{E}[|Z|^2] &= \log \mathbb{E}[X^2 + \lambda_2^2 Y^2] \leq \log \mathbb{E}[X^2 + Y^2] = 1 \\ \mathbb{E}[\log |Z|^2] &= \mathbb{E}[\log(X^2 + \lambda_2^2 Y^2)] \geq \mathbb{E}[\log X^2] = 1 - \gamma, \end{aligned}$$

which implies the first claim. The estimates

$$\begin{aligned} \text{Var}(\log |Z|^2) &\leq \mathbb{E}[\log^2 |Z|^2] \leq \mathbb{E}[\log^2(X^2 + Y^2)] = \theta. \end{aligned}$$

prove the second claim. \square

3 An extension of Morgenstern's bound

Morgenstern's bound [9] states that if $A \in \mathbb{C}^{n \times n}$, then $\mathcal{C}(A) \geq \log |\det(A)|$, see also [2, Chapter 13] for details. An immediate generalization of Morgenstern's bound is as follows. We define the *r -volume* $\text{vol}_r(A)$ of a matrix $A \in \mathbb{C}^{m \times n}$ as the maximum among the absolute values of all $r \times r$ subdeterminants of A . Then we have

$$\mathcal{C}(A) \geq \log \text{vol}_r(A) \tag{2}$$

for any matrix $A \in \mathbb{C}^{m \times n}$ and $1 \leq r \leq \min\{m, n\}$.

It will be important to use a variant of the r -volume that is invariant under unitary transformations. Instead of taking the maximum, we will use the sum of the squares of the absolute values of all $r \times r$ minors of the matrix under consideration.

Definition 3.1. Let $A \in \mathbb{C}^{m \times n}$ be a matrix. Then the *mean square volume* $\text{msv}_r(A)$ of A is defined as

$$\text{msv}_r(A) = \left(\sum_{I, J} |\det A_{I, J}|^2 \right)^{1/2},$$

where I and J run over all subsets of $\{1, \dots, m\}$ and $\{1, \dots, n\}$ of cardinality r , respectively, and $A_{I, J}$ is the $r \times r$ submatrix consisting of the rows indexed by I and columns indexed by J .

We remark that for $A \in \mathbb{C}^{m \times n}$ and $\lambda \in \mathbb{C}$

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

where A^* denotes the complex transpose of A . Moreover, we have $\text{msv}_n(A) = |\det A|$ for $A \in \mathbb{C}^{n \times n}$.

The next lemma presents further properties of this new invariant.

Lemma 3.1. *Let $A \in \mathbb{C}^{m \times n}$ and let $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ be unitary matrices. Then, for $1 \leq r \leq \min\{m, n\}$,*

$$\text{vol}_r(A) \leq \text{msv}_r(A) \leq \sqrt{\binom{m}{r} \binom{n}{r}} \text{vol}_r(A),$$

and (unitary invariance)

$$\text{msv}_r(A) = \text{msv}_r(UAV).$$

Proof. The first claim is obvious. For the second claim we use an argument based on the Gramian determinant. Let $A = (a_1, \dots, a_m)^\top$ be a matrix in $\mathbb{C}^{m \times n}$ with rows $a_j \in \mathbb{C}^n$. For a subset $I \subseteq \{1, \dots, m\}$ with $|I| = r$ let A_I be the submatrix of A consisting of the rows indexed by I . Define the Gramian matrix $G_I = A_I A_I^* = (\langle a_j, a_k \rangle)_{j,k \in I} \in \mathbb{C}^{r \times r}$. By the Binet-Cauchy formula (see [1, Chapter 4]), we have that

$$\det G_I = \det A_I A_I^* = \sum_J |\det A_{I,J}|^2, \quad (3)$$

where J runs over the subsets of $\{1, \dots, n\}$ of cardinality r . Since the scalar product defining G_I is invariant under unitary transformations, we get from (3) for a unitary $V \in \mathbb{C}^{n \times n}$ that

$$\sum_J |\det (AV)_{I,J}|^2 = \det G_I = \sum_J |\det A_{I,J}|^2,$$

hence $\text{msv}_r(AV) = \text{msv}_r(A)$. Using the fact that the mean square volume is invariant under taking the complex transpose, we conclude that also $\text{msv}_r(A) = \text{msv}_r(UA)$. \square

Remark 3.1. The r -volume can be seen as the maximum-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., [7] 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 then follows from the fact that Λ^r is equivariant with respect to unitary transformations and that the Frobenius norm is invariant under such.

Lemma 3.1, combined with the bound (2), immediately yields the following fundamental theorem.

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

$$\mathcal{C}(A) \geq \log \text{msv}_r(A) - \frac{1}{2} \log \binom{m}{r} \binom{n}{r}.$$

Note that this theorem implies Morgenstern's bound $\mathcal{C}(A) \geq \log |\det A|$ for $A \in \mathbb{C}^{n \times n}$ in the case $r = n$.

The fact that the mean square volume of A is invariant under unitary transformations allows us to express it in terms of the singular values of the matrix A .

Proposition 3.1. *Let $A \in \mathbb{C}^{m \times n}$ with singular values $\sigma_1, \dots, \sigma_p$, $p = \min\{m, n\}$. Then we have for $1 \leq r \leq p$ that*

$$\text{msv}_r^2(A) = \text{msv}_r^2(\text{diag}(\sigma_1, \dots, \sigma_p)) = \sum_I \prod_{k \in I} \sigma_k^2,$$

where I runs over all subsets of $\{1, \dots, p\}$ with r elements. Hence, the square of the r -mean square volume of a matrix is the r -th elementary symmetric polynomial in the squares of the singular values.

Proof. It is well known (see [5]) that there are unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that $U^* A V = \text{diag}(\sigma_1, \dots, \sigma_p)$. The claim then follows from the unitary invariance of the mean square volume. \square

Proposition 3.1 implies that $\text{msv}_r(A) \geq \sigma_r^r$. Combining this with Theorem 3.1 and Lemma 2.2, we conclude that

$$\mathcal{C}(A) \geq r \log \sigma_r - n \geq r \log \text{rig}_r(A) - n. \quad (4)$$

We note that this bound implies the rigidity bound in Raz [13, Cor. 3.4] up to the additive term n .

The mean square volume can be applied to obtain a variant of Chazelle's Spectral Lemma [3]. Using entropy considerations, Chazelle obtained a lower bound for the b.c. complexity of linear maps $\mathbb{C}^n \rightarrow \mathbb{C}^n$ in terms of its singular values, even if up to $n/2$ help gates are allowed (nodes in the circuit that are allowed to compute *any* function of the previous intermediate results). We now show that the mean square volume bound allows us to deal with a weaker variant of help gates. This result is not needed for understanding the rest of the paper.

In what follows, we will call an *unbounded* gate in a linear circuit an instruction corresponding to a scalar multiplication with a constant of absolute value greater than two.

Proposition 3.2. *Let $\varphi^A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ be a linear map and Γ be a b.c. circuit for φ^A containing $k < n$ unbounded gates. Then for any $k < r \leq n$,*

$$\mathcal{S}(\Gamma) \geq (r - k) \log \sigma_r - (n - k),$$

where σ_r is the r -th largest singular value of A .

Proof. Let g_i , $i \in I$, be the linear forms computed at the unbounded gates of Γ . We transform the circuit Γ into a b.c. circuit Γ' by replacing each unbounded gate with a multiplication by zero. This new circuit is obviously a b.c. circuit of size $\mathcal{S}(\Gamma') = \mathcal{S}(\Gamma) - k$, computing a linear map φ^B corresponding to a matrix B . Let V be the orthogonal complement of $\text{span}\{g_i \mid i \in I\}$. Clearly, $\text{codim}V \leq k$. The linear maps φ^A and φ^B coincide on V and we denote their common restriction to V by φ .

Let $\sigma_1^A \geq \dots \geq \sigma_n^A$ and $\sigma_1^B \geq \dots \geq \sigma_n^B$ be the singular values of φ^A and φ^B , respectively, and $\sigma_1 \geq \dots \geq \sigma_{n-k}$ be the first $n-k$ singular values of φ . The Courant-Fischer min-max Theorem (see [4, Chapt. 1, Sect. 4]) gives the following description of the l -th singular value of a matrix A in terms of its associated quadratic form x^*A^*Ax :

$$\sigma_l = \min_W \max_{x \in W, \|x\|=1} x^*A^*Ax,$$

where the minimum is taken over all subspaces W of codimension $l-1$. This description easily implies that the l -th singular value of the linear map φ^B restricted to a hyperplane lies in the interval $[\sigma_{\ell+1}^B, \sigma_\ell^B]$. Using this observation repeatedly, we obtain that (see [4, Chapt. 1, Sect. 4.2])

$$\sigma_{r-k}^B \geq \sigma_{r-k} \geq \sigma_r^A \text{ for } k < r \leq n. \quad (5)$$

Using Theorem 3.1, Proposition 3.1, and (5) we thus get

$$\begin{aligned} \mathcal{S}(\Gamma) - k &= \mathcal{S}(\Gamma') \geq \log \text{msv}_{r-k} B - n \\ &= \frac{1}{2} \log \left(\prod_{i=1}^{r-k} (\sigma_i^B)^2 \right) - n \\ &\geq (r-k) \log \sigma_{r-k}^B - n \\ &\geq (r-k) \log \sigma_r^A - n, \end{aligned}$$

and the proof is finished. \square

We note that this bound does not give useful results if the number of unbounded gates is close to n .

4 A lower bound on cyclic convolution

In this section we use Theorem 3.1 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 $\mathbb{C}[X]/(X^n - 1)$. More explicitly:

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

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

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

Let $\text{DFT}_n = (\omega^{jk})_{0 \leq j, k < n}$ be the matrix of the Discrete Fourier Transform, with $\omega = e^{2\pi i/n}$. It is well known (see, e.g., [5, Sect. 4.7.7]) that

$$C(a) = \left(\frac{1}{\sqrt{n}} \text{DFT}_n \right)^{-1} \text{diag}(\lambda_0, \dots, \lambda_{n-1}) \frac{1}{\sqrt{n}} \text{DFT}_n,$$

where the eigenvalues λ_k of $C(a)$ are given by

$$(\lambda_0, \dots, \lambda_{n-1})^\top = \text{DFT}_n(a_0, \dots, a_{n-1})^\top. \quad (6)$$

Hence the singular values of $C(a)$ are $|\lambda_0|, \dots, |\lambda_{n-1}|$. Note that $n^{-1/2} \text{DFT}_n$ 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.

We now state the main result of the paper.

Theorem 4.1. *Let $\varphi_n: \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ be the bilinear map of the n -dimensional cyclic convolution. Then we have for $n \rightarrow \infty$*

$$\mathcal{C}(\varphi_n) \geq \frac{1}{12} n \log n - O(n \log \log n).$$

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.

To prepare for the proof, we need some lemmas. The idea behind the following lemma is already present in [13]. We will identify linear forms on \mathbb{C}^n with vectors in \mathbb{C}^n .

Lemma 4.1. *Let $f_1, \dots, f_k \in \mathbb{C}^n$ be linear forms and let $1 \leq r < n$. Then there exists a complex subspace $U \subseteq \mathbb{C}^n$ of dimension r such that for a standard normal distributed complex random vector a with values in U , we have*

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

Proof. Set $R = \text{rig}_{n-r}(f_1, \dots, f_k)$. Then there exists a linear subspace $V \subseteq \mathbb{C}^n$ of dimension $n-r$ such that $\text{dist}(f_i, V) \leq R$ for all $1 \leq i \leq k$. 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 $\|f'_i\| \leq R$.

Let (b_1, \dots, b_n) be standard normal distributed in \mathbb{C}^n , and a be the orthogonal projection of b onto U along V . Then a is standard normal distributed with values in U .

Moreover, we have $f'_i(b) = f_i(a)$. By Lemma 2.3, the random variable $T = |f'_i(b)|^2 / (2\|f'_i\|^2)$ is exponentially distributed with parameter 1. For any real λ , we get:

$$\mathbb{P}[T \geq \lambda] = \mathbb{E}[1_{T \geq \lambda}] \leq \mathbb{E}[e^{-(T-\lambda)/2}] = e^{-\lambda/2} \mathbb{E}[e^{T/2}].$$

On the other hand,

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

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

$$\mathbb{P}[T \geq \lambda] = \mathbb{P}[|f'_i(b)|^2 \geq 2\lambda\|f'_i\|^2] \leq 2e^{-\lambda/2}.$$

Since $\|f'_i\| \leq R$, we have for a fixed i that

$$\mathbb{P}[|f_i(a)| \geq \sqrt{2\lambda} R] \leq 2e^{-\lambda/2}.$$

By the union bound we obtain

$$\mathbb{P}\left[\max_i |f_i(a)| \geq \sqrt{2\lambda} R\right] \leq 2ke^{-\lambda/2}.$$

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

In the next lemma, we state a lower bound on the b.c. linear complexity of the circulant.

Lemma 4.2. *Let $U \subseteq \mathbb{C}^n$ be a subspace of dimension r . For a standard normal distributed complex random vector a with values in U , we have*

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

where $c = \frac{1}{2}(3 + \gamma + \sqrt{2\theta}) \approx 4.23$, and γ, θ are the constants introduced in Section 2.2.3.

The proof of this lemma is given in the next section. We proceed with the proof of the main theorem.

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

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

By Lemma 2.1 and the fact that $k \leq 3n^3$, we get

$$\mathcal{S}(\Gamma) + n \log(2\sqrt{\ln(12n^3)} R) \geq \mathcal{C}(C(a)). \quad (7)$$

On the other hand, by Equation (4), we get the following upper bound on R in terms of $\mathcal{S}(\Gamma)$:

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

By combining this with (7) and using the second condition above, we obtain

$$\left(1 + \frac{n}{n-r}\right) \mathcal{S}(\Gamma) + \frac{n^2}{n-r} \geq \frac{r}{2} \log n - O(n \log \log n).$$

Setting $\epsilon = r/n$ yields

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

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. \square

5 Proof of Lemma 4.2

Before going into the proof, we provide a lemma on bounding the deviations of products of (possibly dependent) random variables. For this we need the following auxiliary result.

Lemma 5.1. *Let $A = (a_{kj})_{1 \leq k, j \leq r} \in \mathbb{C}^{r \times r}$ be a positive definite hermitian matrix. Then $\det A \leq \prod_{j=1}^r a_{jj}$.*

Proof. For a proof, see for example [1, Chapter 8]. \square

Lemma 5.2. *Let $Z = (Z_1, \dots, Z_r)$ be a normal distributed random vector in \mathbb{C}^r with mean 0. Define the complex covariance matrix of Z by $\Sigma := (E[Z_j \bar{Z}_k])_{j,k}$. Then we have*

$$\mathbb{P}[|Z_1|^2 \cdots |Z_r|^2 \geq \delta^r \det \Sigma] > \frac{1}{2},$$

where $\delta = 2^{-(\gamma + \sqrt{2\theta})} \approx 0.02$.

Proof. We first transform the product into a sum by taking logarithms. For every $\epsilon > 0$ the Chebychev inequality yields the bound

$$\mathbb{P}\left[\frac{1}{r} \left| \sum_{j=1}^r (\log |Z_j|^2 - E[\log |Z_j|^2]) \right| \geq \epsilon\right] \leq \frac{\text{Var}(\sum_{j=1}^r \log |Z_j|^2)}{\epsilon^2 r^2}. \quad (8)$$

For the variance we have by Proposition 2.1

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

Setting $\epsilon^2 = 2\theta$ in this equation and after exponentiating in (8) we obtain

$$\mathbb{P}\left[|Z_1|^2 \cdots |Z_r|^2 \leq 2^{-\epsilon r + \sum_{j=1}^r \mathbb{E}[\log |Z_j|^2]}\right] \leq \frac{1}{2}. \quad (9)$$

By combining Lemma 5.1 with Proposition 2.1 we get

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

Hence we conclude from (9) that

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

from which the lemma follows. \square

Proof. (of Lemma 4.2) By equation (6) we have $\lambda = \text{DFT}_n a$ and the singular values of the circulant $C(a)$ are given by the absolute values of the components of λ . Setting

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

we obtain for the r -mean square volume by Proposition 3.1

$$\text{msv}_r^2(C(a)) = n^r \sum_I \prod_{k \in I} |\alpha_k|^2, \quad (10)$$

where I runs over all subsets of $\{1, \dots, n\}$ with r elements.

Let now a be a standard normal distributed random vector with values in the subspace U of dimension r . Let W be the image of U under the unitary transformation $n^{-1/2} \text{DFT}_n$. As a unitary transformation of a , α is standard normal distributed with values 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 normal distributed 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 $I \subseteq \{1, \dots, n\}$, $|I| = r$.

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 $\mathbb{C}^n \rightarrow \mathbb{C}^I$, $\alpha \mapsto \alpha_I$ restricted to W .

By the Binet-Cauchy formula (compare (3)) and the orthogonality of the basis (b_i) we get

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

Therefore, we can choose an index set I such that

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

By applying Lemma 5.2 to the random vector α_I and using (10), we get that with probability at least $1/2$,

$$\text{msv}_r^2(C(a)) \geq n^r \delta^r \det \Sigma \geq n^r \delta^r \binom{n}{r}^{-1},$$

where $\delta = 2^{-(\gamma + \sqrt{2\theta})}$. Therefore, using $\binom{n}{r} \leq 2^n$, we get from Theorem 3.1 that

$$\begin{aligned} \mathcal{C}(C(a)) &\geq \log \text{msv}_r(C(a)) - n \\ &\geq \frac{1}{2} r \log n - \frac{1}{2} (3 + \log \delta^{-1}) n, \end{aligned}$$

with probability at least $1/2$. This proves the lemma. \square

6 Applications

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. However, we did not attempt to optimize these factors.

6.1 Polynomial multiplication

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 $\mathbb{C}[X]$ 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 $0 \leq k < n$. This observation and Theorem 4.1 immediately imply the following corollary.

Corollary 6.1. *Let $\varphi_n: \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}^{2n-1}$ be the bilinear map of multiplying polynomials of degree less than n . Then we have for $n \rightarrow \infty$*

$$\mathcal{C}(\varphi) \geq \frac{1}{12} n \log n - O(n \log \log n).$$

6.2 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. M. Sieveking [14] and H.T. Kung [6] 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 6.1. *Let φ_n be 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. Then we have for $n \rightarrow \infty$*

$$\mathcal{C}(\varphi_n) \geq \frac{1}{324} n \log n - O(n \log \log n).$$

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 proof of [2, 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

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

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 6.1. \square

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 due to V. Strassen [15], see also [2, Section 2.5].

Corollary 6.2. *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 $\deg r < \deg g$. If $\varphi_{n,m}$ denotes the function that maps the*

coefficients of f and g to the coefficients of q and r , then we have

$$\mathcal{C}(\varphi_{2n,n}) \geq \frac{1}{324} n \log n - O(n \log \log n).$$

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

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

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

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

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

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

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. \square

References

- [1] R. Bellman. *Introduction to matrix analysis*. SIAM, Philadelphia, PA, 1997.
- [2] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic complexity theory*. Springer-Verlag, Berlin, 1997.
- [3] B. Chazelle. A spectral approach to lower bounds with applications to geometric searching. *SIAM Journal on Computing*, 27(2):545–556, 1998.
- [4] R. Courant and D. Hilbert. *Methoden der mathematischen Physik. I*. Springer-Verlag, Berlin, 1931. Zweite Auflage.
- [5] G. H. Golub and C. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, 1996.
- [6] H. T. Kung. On computing reciprocals of power series. *Numer. Math.*, 22:341–348, 1974.
- [7] S. Lang. *Algebra*. Addison-Wesley, second edition, 1984.
- [8] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *Proc. 36th FOCS*, pages 6–15, 1995.
- [9] J. Morgenstern. Note on the lower bound of the linear complexity of the fast fourier transform. *Journal of the ACM*, 20:305–306, 1973.
- [10] J. Morgenstern. The linear complexity of computation. *J. ACM*, 22:184–194, 1975.
- [11] N. Nisan and A. Wigderson. On the complexity of bilinear forms. *Proc. of the 27th ACM Symposium on the Theory of Computing*, pages 723–732, 1995.

- [12] P. Pudlák. A note on the use of determinant for proving lower bounds on the size of linear circuits. *ECCC Report*, 42, 1998.
- [13] R. Raz. On the complexity of matrix product. In *Proc. 34th STOC*, 2002. Also available as ECCC Report 12, 2002.
- [14] M. Sieveking. An algorithm for division of power series. *Computing (Arch. Elektron. Rechnen)*, 10:153–156, 1972.
- [15] V. Strassen. Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. *Num. Math.*, 20:238–251, 1973.
- [16] V. Strassen. Vermeidung von Divisionen. *Crelles J. Reine Angew. Math.*, 264:184–202, 1973.
- [17] L. Valiant. Graph theoretic properties in computational complexity. *J. Comput. Syst. Sci.*, 13:278–285, 1976.
- [18] L. Valiant. Graph theoretic arguments in low-level complexity. Number 53 in *Lecture Notes in Computer Science*, pages 162–176. Springer, 1977.