Towards Arithmetical Lower Bounds with Unbounded Coefficients

Maurice J. Jansen
University at Buffalo

Kenneth W.Regan*
University at Buffalo

Abstract

Raz [Raz03] and Bürgisser and Lotz [BL04] proved tight $\Omega(n \log n)$ lower bounds on the size of bounded-coefficient (bc) arithmetical circuits computing certain functions f, namely matrix multiplication and cyclic convolution, respectively. The latter holds in the presence of $(1-\epsilon)n$ extra "help gates" with unbounded coefficients. For such f we consider the collections of functions $f(Dx)$ or $f(Dx, Ey)$ where D and E belong to certain matrix groups G; these are called the *orbits* of f under action by G. The point is that D and/or E are allowed to have unbounded coefficients, and no charge is made for the complexity of D and/or E themselves. We prove a lower bound of $\Omega(n(\log n)^{1/3})$ on the size of bc arithmetical circuits computing cyclic convolution when the circuits get such free help from the action of the group $SLD_n(\mathbf{C})$ of diagonal complex matrices of determinant 1. Since adding just one extra "help gate" to this action is equivalent to the *un*bounded coefficient case, the results build a bridge to that case.

1 Introduction

One of the central mysteries in arithmetic circuit complexity over infinite fields F is the computational power conferred by the ability to use "for free" constants of arbitrary magnitude and/or precision from F. These constants are a major technical obstacle in relating arithmetic complexity to Boolean circuit complexity theory. It is commonly observed (e.g. by [Mor73, Cha98, Mul99]) that classic important algorithms employ only simple constants. A major exception is *polynomial interpolation*, but even here it seems that over fields containing the rationals, small constants with enough bits of precision are equally as useful as large ones.

To probe the significance of field constants, several researchers have obtained size lower bounds on arithmetical circuits with a uniform bound on the absolute value of constants. Morgenstern [Mor73] proved that bounded-coefficient circuits (henceforth, bc-circuits) need size $\Omega(n \log n)$ to compute the linear transformation for the Fast Fourier Transform. Chazelle [Cha98] obtained similar bounds for geometric range-searching problems, while Lokam [Lok01] obtained related size-depth tradeoffs for bc-circuits computing linear transformations with certain degrees of *rigidity*. More recently Raz [Raz02] broke through by obtaining $\Omega(n \log n)$ lower bounds for a natural *bi*-linear function, namely multiplication of two $\sqrt{n} \times \sqrt{n}$ matrices. Bürgisser and Lotz [BL04] extended Raz's ideas to obtain tight $\Omega(n \log n)$ bounds on

*Part of this work by both authors supported by NSF Grant CCR-9821040. Corresponding author: Dept. of CSE at UB, 201 Bell Hall, Buffalo, NY 14260-2000; (716) 645-3180 x114, fax 645-3464; regan@cse.buffalo.edu

bc-circuits for *cyclic convolution*, and thence for polynomial multiplication and related bi-linear functions. They show them even when the bc-restriction is relaxed by allowing the circuits to have $(1-\epsilon)n$-many multiplication gates in which one of the two fan-in wires is a scalar of unbounded magnitude. These results allow such "help gates" anywhere in the circuits, but we will focus on help gates whose non-scalar wire is one of the input variables.

The natural question is, can one obtain similar lower bounds without the bc-restriction at all? A mark of difficulty is that the best-known lower bounds on general arithmetical circuit size of *any explicit family of* $n^{O(1)}$*-degree polynomials* $f(x_0,\ldots,x_{n-1})$ are $\Omega(n\log n)$. These employ the "Derivative Lemma" of Baur and Strassen [BS82], which converts circuits of size s computing f into n-output circuits of size $O(s)$ computing $(\partial f) : F^n \to F^n = (\partial f/\partial x_1,\ldots,\partial f/\partial x_n)$. They apply to functions like $f = x_0^n + \ldots + x_{n-1}^n$ such that the *geometric degree* of the ideals I_f in $F[x_0,\ldots,x_{n-1},y_0,\ldots,y_{n-1}]$ generated by $(y_0 - \partial f/\partial x_0,\ldots,y_{n-1} - \partial f/\partial x_{n-1})$ grow as $\exp(n\log n)$. For bilinear forms, however, the first derivatives are linear, and so the associated geometric degree is just 1. Thus extending the results of [Raz02, BL04] to unbounded-coefficient cases seems a hard task, hence one worthy of attack by building bridges of partial progress.

Our main bridging concept allows certain linear transformations of unit determinant at the inputs free of charge. For a bilinear function $f(x,y)$ of $2n$ variables, we consider the *orbit* of f under the natural "double action" $Gf = \{\,\lambda x,y.f(Ex,Dy) : D,E\in G\,\}$ of some group G of $n\times n$ matrices. This is a special case of a "single action" by a group of $2n\times 2n$ (block) matrices. Such actions on multilinear maps f like the determinant and permanent polynomials form the basis of Mulmuley and Sohoni's recently-proposed attack on super-polynomial (arithmetical or Boolean) circuit lower bounds [MS02]. If G is allowed to be the group $GL_n(\mathbf{C})$ of all nonsingular $n\times n$ complex matrices—or even its subgroup $GLD_n(\mathbf{Q})$ of diagonal matrices with nonzero rational entries—then the orbit under G is equivalent to having n "help gates," and hence to the unbounded-coefficient model. Our approach becomes interesting when G is the special linear group $SL_n(\mathbf{C})$ of complex matrices with determinant 1, or its subgroup $SLD_n(\mathbf{C})$ of diagonal matrices. Note that these orbits allow "help gates" to $n-o(n)$ of the inputs, so long as the remaining inputs are given small multipliers to compensate. Also note that this model does not constrain the linear circuit complexity of D and E themselves, which may be as high as quadratic.

By a simple extension of the Bürgisser-Lotz method—and as a non-redundant way of including their details for later reference—we first extend their $\Omega(n\log n)$ lower bounds to cases where E and D have bounded *condition number*. This subsumes the case where E and D belong to the group $U_n(\mathbf{C})$ of complex unitary matrices, since those have condition number 1. Matrices of bounded condition number do not form a group, however. The point is that matrices of unit determinant—even diagonal ones—can be arbitrarily ill-conditioned.

We attack this via an extension of their method that makes major use of recent advances in Fourier sampling by Candès, Romberg, and Tao [CRT04]. These advances still have a "$\log N$ bottleneck" [CT04] that knocks the lower bounds we obtain from $\Omega(n\log n)$ down to $\Omega(n(\log n)^{1/3})$.

Main Theorem. Bounded-coefficient arithmetical circuits for any element of the GLD_n-orbit of cyclic convolution require size $\Omega(n(\log n)^{1/3})$.

Our tweaking of the Fourier techniques from [CRT04] produces a lower bound of $\Omega(n^2/l_n)$, where l_n is the maximum of $n/\log n$ and the number of entries in a unit-determinant diagonal matrix D whose absolute value is less than 1. We balance this with an extension of the "help-gates" theorem of [BL04] to the case of $n - n/(\log n)^{1/3}$ help gates, to obtain the stated bound. The main ingredient—and the frontier issue in [CRT04]—involves lower bounds on the expected magnitude of the determinant of large minors of the Discrete Fourier Transform matrix, under certain partly-adversarial distributions.

2 Definitions and Background

Throughout the paper, we let F_n abbreviate $\left(\frac{DFT_n}{\sqrt{n}}\right)$, where $(DFT_n)_{ij} = \omega^{ij}$ for $\omega = e^{2\pi i/n}$. Arithmetical circuits computing a function $f : F^n \to F^m$ over some field F have binary addition ($+$) and binary multiplication ($*$) gates, n-many input gates labeled $x_0, \ldots, x_{n-1}$, and m-many output gates. The wires into a $+$ or $*$ gate may contain multiplicative constants from F. For instance, a binary $+$ gate with constants c, d on the two incoming wires computes the operation $cy + dz$ given arguments y and z. With $c = 1$, $d = -1$ the gate computes subtraction. The size of the circuit is the number of binary gates, or equivalently in case of bounded fan-in, the number of wires.

Linear circuits have no $*$-gates. *Bilinear* circuits as described in [Raz02, BL02] have a single layer of k-many $*$-gates, and clusters T_1, T_2, T_3 of addition gates. The gates in T_1 compute k-many linear forms $\ell_1, \ldots, \ell_k$ in the variables $x_0, \ldots, x_{n-1}$ (so T_1 is a linear transformation from $\mathbf{C}^n$ to $\mathbf{C}^k$), while those in T_2 compute linear forms $r_1, \ldots, r_k$ in $y_0, \ldots, y_{n-1}$. The multiplication gates compute the respective products $\ell_1 r_1, \ldots, \ell_k r_k$, while the gates in T_3 compute a linear transformation from $\mathbf{C}^k$ to $\mathbf{C}^n$ that gives the final output $f = (f_1, \ldots, f_n) = T_3(\ell_1 r_1, \ldots, \ell_k r_k)$. Except where indicated in case (c) above, it does not matter whether "ℓ_i" and "ℓ_j" for different i, j are computed by the same node in T_1, and likewise for T_2.

We follow [Raz02] in defining *bounded coefficient* (bc) circuits to have all constants on wires be of magnitude at most 1. Bürgisser and Lotz use a bound of 2, but theirs can be converted to Raz's circuits by inserting extra $+$ gates, and the only effect on our results would be a halving of the depth tradeoff obtained under restriction (c). As attested in [Raz02], every (bc) circuit computing bilinear forms $(f_1, \ldots, f_n)$ can be converted into a *bilinear* (bc) circuit computing $(f_1, \ldots, f_n)$ with a small constant-factor increase in size and depth. Thus we may restrict attention to bilinear circuits for the functions we consider. The complexity of a linear transformation A for bounded-coefficient (linear) circuits is denoted by $s^{bc}_{lin}(A)$.

The *cyclic convolution* $x \circ y$ of two n-vectors x, y as above is the n-vector $(z_0, \ldots, z_{n-1})$ with components

$$z_k = \sum_{i+j \equiv k \bmod n} x_i y_j$$

for $0 \leq k < n$. For example, for $n = 5$, we get

$$x \circ y = \begin{pmatrix} x_0y_0 + x_4y_1 + x_3y_2 + x_2y_3 + x_1y_4 \\ x_1y_0 + x_0y_1 + x_4y_2 + x_3y_3 + x_2y_4 \\ x_2y_0 + x_1y_1 + x_0y_2 + x_4y_3 + x_3y_4 \\ x_3y_0 + x_2y_1 + x_1y_2 + x_0y_3 + x_4y_4 \\ x_4y_0 + x_3y_1 + x_2y_2 + x_1y_3 + x_0y_4 \end{pmatrix}$$

When fixing $x = a = (a_0, \ldots, a_{n-1})^T$, the induced map on y is computed by the circulant matrix

$Circ(a) =$

$$\begin{pmatrix} a_0 & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \cdots & a_3 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n-2} & a_{n-3} & \cdots & a_0 & a_{n-1} \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{pmatrix}.$$

We also find it convenient to consider the "half convolution" defined by $HCirc(x)y$, where $HCirc(a)$ is the lower-triangular matrix

$$\begin{pmatrix} a_0 & 0 & \cdots & 0 & 0 \\ a_1 & a_0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ a_{n-2} & a_{n-3} & \cdots & a_0 & 0 \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{pmatrix}.$$

Then $x \circ y$ can be obtained by adding $HCirc(x)y$ to the inverted vector $HCirc(x_{n-1}, x_{n-2}, \ldots, x_1)(y_1, y_2, \ldots, y_{n-1})$, which can be done by bilinear (bc) circuits with linearly many extra $+$ gates. Thus lower bounds on $x \circ y$ extend immediately to $HCirc(x)y$. The convenience is that $HCirc(x)y$ is definable by recursion from $HCirc(x_1, \ldots, x_{n-2})(y_1, \ldots, y_{n-2})$, needing only linearly-many extra binary $*$ gates applied to x_0, y_0 and elements of $x_0, \ldots, x_{n-1}$ and $y_0, \ldots, y_{n-1}$ and preserving the bilinear format. Namely, zero out the first column and main diagonal of $HCirc(a)$, observe that the piece in between is the lower triangle of $HCirc(a_1, \ldots, a_{n-2})$ multiplying the interior $n-2$ elements of y, and restore the summands in the first column and main diagonal involving x_0 and y_0. We use this fact in the proof of Theorem 2.1 below.

Now we define our orbit model.

Definition 2.1. Let D and E be $n \times n$ non-singular complex matrices. An *orbit circuit* is the composition $\Gamma(Dx, Ey)$, where Γ is a bounded-constants bilinear circuit. The size of the circuit is taken to be the size of Γ.

To emphasize, the entries of the matrices E and D above are not restricted to be of norm at most one. An orbit circuit thus has the potential help of $2n^2$-many unbounded constants, although flowing through only $2n$-many input gates.

First, *any* bilinear circuit C can be converted to an orbit circuit Γ of the same size with diagonal matrices E and D. If g is a $+$ gate with m outgoing wires with constants $c_1, \ldots, c_m$ and constants d, e on its incoming wires, then we may take c to be the maximum of $|c_1|, \ldots, |c_m|$, replace each c_i by c_i/c (which has norm at most 1), and make cd, ce the new constants on the incoming wires. If g is a $*$ gate, we need only propagate cd, e upward. Iterating this from the outputs up pushes all unbounded constants up to the wires from the inputs. Repeating this one more time pushes the unbounded constants onto the inputs themselves as nonnegative reals, and they can be the entries of E and D. None of the final constants will be zero unless the corresponding input was already zeroed out. Thus the orbit model with $G = GL_n(\mathbf{C})$, namely the group of all invertible complex matrices, is no less general than the unbounded-coefficients case (possibly more so, if D and E have high circuit complexity by themselves).

Things become more interesting with $G = SL_n(\mathbf{C})$. If (the function computed by) C ignores inputs x_0 and y_0, then we can create diagonal matrices D, E of determinant 1 by taking the first entry to be $1/K^{n-1}$ and the remaining entries to be K, where K is the maximum real constant obtained in the pushing-up process. The tiny entry in D and E gets thrown away while the large ones feed the bc-circuit Γ left over from the process. If we insist on attention to functions f that depend on all of their inputs, then lower bound techniques that tolerate two unbounded "help gates" (not needing the $(1-\epsilon)n$ allowance in [BL02]) still imply lower bounds in the general case, with x_0 and y_0 becoming the help gates. If we disallow this but "relax" orbit circuits Γ by allowing access also to the un-transformed inputs x_0 and y_0, we can still prove rigorously that $SL_n(\mathbf{C})$-orbit bc-circuit lower bounds imply unbounded-coefficient lower bounds, for half-convolution and functions with a similar recursion:

Theorem 2.1 *Bilinear circuits C of size s computing $HCirc(x)y$ can be converted into "relaxed" SL_n-orbit circuits Γ of size $s + O(n)$ computing $HCirc(x)y$.*

Proof. Convert C to Γ_0 by pushing up constants as before, along with the above diagonal $D, E \in SL_n(\mathbf{R})$. Now reduce Γ_0 by zeroing the constants out of x_0 and y_0, splicing out gates their wires connect to. The resulting circuit computes $HCirc(x_1, \ldots, x_{n-2})(y_1, \ldots, y_{n-2})$. Finally use the free access to the untransformed inputs x_0 and y_0 to re-create $HCirc(x)y$ as above, adding $2n$-many $*$ gates and $2n-1$ $+$ gates at the outputs. On products $x_0 y_i$ with $i > 0$, the constant K on y_i from D is counter-acted by a constant $1/K$ on the wire from x_0, and similarly for products $x_i y_0$. This yields the desired "relaxed" orbit bc-circuit Γ. $\square$

Without any relaxation, allowing such matrices D and E certainly defeats the proof methods of Raz and Bürgisser-Lotz. These rely on bounding the volume-expansion factor on all r-dimensional subspaces of $\mathbf{C}^n$, for some value $r = \Theta(n)$. Matrices of this form can expand volume in many of these subspaces by the unbounded factor K (or rather by K^r), and it seems not to matter that the first co-ordinate is crushed by $1/K^{n-1}$. We adapt these methods for cases where we can avoid or contain this problem. As in [Raz02, BL02], the backbone of the lower bound technique will be to simplify the bilinear circuit into a linear circuit using the probabilistic method. The idea is to nonzero probability of fixed values a for $(x_0, \ldots, x_{n-1})$ such that the "x side" of the bilinear bc-circuit Γ keeps the values $\ell_1(a), \ldots, \ell_k(a)$ reasonably small while leaving the complexity of the induced linear map $A(y_0, \ldots, y_{n-1})$ high. Substituting those values at the $*$ gates and building them up additively from bounded constants leaves a bc-linear circuit C computing A of the same order of size as Γ, hence Γ must obey the size lower bounds known for C. We now review the probabilistic machinery from [BL02].

2.1 Standard Gaussian vectors

When we need to distinguish row and column vectors x, we write the latter as $(x_0, \ldots, x_{n-1})^T$. The *conjugate transpose* of a vector x and matrix A are written x^* and A^*, respectively. A random vector $x \in \mathbf{C}$ is called *standard Gaussian* if the real and imaginary parts of all components x_i comprise $2n$ independent standard normally distributed random variables. An important fact is that if F is any unitary transformation, then Fx is again standard Gaussian distributed.

For an r-dimensional linear subspace U, we say that a random vector a is standard Gaussian distributed in U if we can write $a = \beta_1 v_1 + \ldots + \beta_r v_r$, where β is standard Gaussian in $\mathbf{C}^r$ and

$\{v_i\}_i$ is an orthonormal basis. This representation is independent of the choice of orthonormal basis.

We will use the following two lemmas from [BL02]. A random variable t is *exponentially distributed with parameter 1* if it has density function $p(t) = e^{-t}$ for $t \geq 0$, and $p(t) = 0$ otherwise.

Lemma 2.2 ([BL02]) *Let $(x_1, \ldots, x_n)^T$ be standard Gaussian in $\mathbf{C}^n$. Let $f = (f_1, \ldots, f_n)^T \in \mathbf{C}^n$. Then $S := f_1 x_1 + \ldots + f_n x_n$ is normally distributed with mean 0 and variance $\|f\|^2$. Furthermore, $T := \frac{|S|^2}{2\|f\|^2}$ is exponentially distributed with parameter 1. Hence T has mean and variance both equal to 1.*

As in [BL02], when we say a vector $z \in \mathbf{C}^r$ is normal distributed with mean 0, we mean that the real and imaginary parts of each component z_i are normal distributed random variables with mean 0.

Lemma 2.3 ([BL02]) *Let $z = (z_1, \ldots, z_r)^T$ be a normal distributed random vector in $\mathbf{C}^r$ with mean 0. Define the complex covariance matrix of z to be the outer product $\Sigma = zz^*$. Then we have*

$$\Pr[|z_1|^2 \cdots |z_r|^2 \geq \delta^r \det(\Sigma)] > \frac{1}{2},$$

for some absolute constant $\delta > 0$.

2.2 Mean Square Volume & Matrix Rigidity

It is well known that the volume of the parallelepiped subtended by the rows of a matrix $A \in \mathbf{C}^{n \times n}$ is given by $|\det(A)|$. Morgenstern [Mor73] proved that $\log|\det(A)|$ is an asymptotic lower bound on the size of a linear arithmetical circuit with bounded coefficients computing the linear transformation given by A. For further lower bounds it is useful to define variations of volume for r-subsets of the n coordinates. The two versions [BL02, BL04] of the work by Bürgisser and Lotz refer to two different "r-volume" notions, and it suits our purposes to include both, giving them different names. Given an $m \times n$ matrix A and sets $I \subseteq \{1, \ldots, m\}$ of row indices and $J \subseteq \{1, \ldots, n\}$ of column indices, define $A_{I,J}$ to be the matrix of elements with row index in I and column index in J, and A_I to be $A_{I,\{1,\ldots,n\}}$.

Definition 2.2 ([Raz02, BL02]). Given $A \in \mathbf{C}^{m \times n}$, and r such that $1 \leq r \leq \min m, n$, define

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

$$\mathrm{vol}'_r(A) = \max_{|I|,|J|=r} (|\det(A_{I,J})|). \tag{2}$$

The centerpiece definition in [BL02, BL04], however, involves taking the Euclidean norm rather than the max-norm.

Definition 2.3 ([BL02]). Given $A \in \mathbf{C}^{m \times n}$, and r such that $1 \leq r \leq \min m, n$, define the

r*-mean square volume* $\mathrm{msv}_r(A)$ of A by

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

These definitions are related by:

Lemma 2.4 ([BL02, BL04], respectively) *For A and r as above,*

$$\mathrm{vol}'_r(A) \leq \mathrm{msv}_r(A) \leq \binom{m}{r}^{1/2}\binom{n}{r}^{1/2}\mathrm{vol}'_r(A), \tag{3}$$
$$\mathrm{vol}_r(A) \leq \mathrm{msv}_r(A) \leq \binom{m}{r}^{1/2}\mathrm{vol}_r(A). \tag{4}$$

An important fact is that mean square r-volume is invariant under unitary transformations. That is, for A as above and all unitary matrices $U \in \mathbf{C}^{m\times m}$ and $V \in \mathbf{C}^{n\times n}$,

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

As we have remarked above, msv_r is not preserved under transformations in $SL_n(\mathbf{R})$ (unless $r = n$).

The following theorem states the use of the mean square volume measure for proving lower bounds.

Theorem 2.5 ([BL02]) *For $A \in \mathbf{C}^{m\times n}$, and $1 \leq r \leq \min(m, n)$, we have that a linear bounded-constant circuit computing A has size at least* $\log \mathrm{msv}_r(A) - \frac{1}{2}\log\binom{m}{r}\binom{n}{r}$.

Next we introduce Raz's notion of geometric rigidity.

Definition 2.4 ([Raz02]). Let $A \in \mathbf{C}^{n\times n}$ be an matrix with with row vectors a_i, The r*-rigidity* of A is defined to be

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

where V ranges over all linear subspaces of $\mathbf{C}^m$, and $dist(a, V) = \min_{v\in V} \|a - v\|$.

Lemmas 3.1 and 3.2 in [Raz02] give the following

Theorem 2.6 *For $A \in \mathbf{C}^{m\times n}$, and $1 \leq r \leq m$, we have that a linear bounded-constant circuit computing A has size at least $r \log rig_r(A) - n$.*

The term n in the above stems from the fact that in [Raz02] norms are bounded by 1 instead of 2. We will use the following lemma from [BL02]. Here for $f, a \in \mathbf{C}^n$, we think of f as a linear form via $f(a) = f^*a$.

Lemma 2.7 ([BL02]) *Let $f_1, \ldots, f_k$ be linear forms and $1 \leq r < n$. Then there exists a linear subspace U of $\mathbf{C}^n$ of dimension r such that for $a \in U$ standard Gaussian, we have that*

$$\Pr[\max_i |f_i(a)| \leq 2\sqrt{\ln 4k}\, rig_{n-r}(f_1^T, \ldots, f_k^T)] \geq \frac{1}{2}.$$

3 Well-Conditioned Orbit Circuits

In this section, we will consider orbit circuits $\Gamma(Ex, Dy)$ for which the matrices E and D are *well conditioned* in the following traditional sense.

Definition 3.1. The *condition number* $\kappa(E)$ of a non-singular matrix E is defined to be the ratio $\frac{\sigma_1(E)}{\sigma_n(E)}$ of it largest and smallest singular value. This is the same as the product $||E||_2 \cdot ||E^{-1}||_2$ (see [GvL96]). We will fix some absolute constant κ_1, and stipulate that a *well-conditioned* matrix E has $\kappa(E) \leq \kappa_1$.

Unitary matrices have condition number 1. That the results of [BL02, BL04] carry over to orbits under unitary matrices follows immediately on the "x side" because the image of a standard-Gaussian vector under unitary transformation is standard Gaussian, and on the "y side" because unitary transformations preserve msv_r. For bounded condition number the "y side" needs only the following easy proposition:

Proposition 3.1 *For any two $n \times n$ nonsingular matrices A and B where B has determinant equal 1, for any $1 \leq r \leq n$, $\mathrm{msv}_r^2(AB) \geq \kappa(B)^{-2r}\mathrm{msv}_r^2(A)$.*

Proof. Let $B = UDV$ be the singular value decomposition of B. Then $\mathrm{msv}_r^2(AB) = \mathrm{msv}_r^2(AUDV) = \mathrm{msv}_r^2(AUD)$. So the general case reduces to the case where B is diagonal with real entries. So assume $B = \mathrm{diag}(b_1, \ldots, b_n)$. Observe that each $b_i \geq \kappa(B)^{-1}$. Hence

$$\begin{aligned}
\mathrm{msv}_r^2(AB) &= \Sigma_{I,J} |\det(AB)_{I,J}|^2 \\
&= \Sigma_{I,J} \prod_{j \in J} |b_j|^2 |\det A_{I,J}|^2 \\
&\geq \kappa(B)^{-2r} \Sigma_{I,J} |\det A_{I,J}|^2 \\
&= \kappa(B)^{-2r} \mathrm{msv}_r^2(A).
\end{aligned}$$

□

However, the "x side" needs more care that the deviation from standard Gaussian distribution incurred in going from x to Ex does not disturb the statistical machinery by too much. The crux of the matter lies in the following generalization of a lemma in [BL02].

Lemma 3.2 *Let $1 \leq r < n$, and Let E and D be an $n \times n$ complex matrices with determinant 1 that are well-conditioned. Let U be a linear subspace of dimension r, and let a be standard Gaussian in U. Then*

$$\Pr[s_{lin}^{bc}(Circ(Ea)D) \geq \frac{1}{2} r \log n - cn] > \frac{1}{2},$$

where c is some absolute constant.

Proof. We can write

$$Circ(Ea) = F_n \mathrm{diag}(\lambda_0, \ldots, \lambda_{n-1}) F_n^{-1},$$

where

$$(\lambda_0, \ldots, \lambda_{n-1})^T = DFT_n Ea.$$

See for example [GvL96]. Let $\alpha = \frac{\lambda}{\sqrt{n}}$. By invariance of mean-square-volume under unitary transformation, we get that

$$\begin{aligned} & \mathrm{msv}_r^2(\mathit{Circ}(Ea)) \\ = \; & \mathrm{msv}_r^2(\mathrm{diag}(\lambda_0, \ldots, \lambda_{n-1})) \\ = \; & \sum_J \prod_{j \in J} |\lambda_j|^2 \\ = \; & n^r \sum_J \prod_{j \in J} |\alpha_j|^2, \end{aligned}$$

where J ranges over all subsets of $\{1, \ldots, n\}$ of size r. By definition of standard Gaussian, we can write $\alpha = V\beta$, where V is an $n \times r$ matrix with orthonormal column vectors $v_1, \ldots, v_r$ and β standard Gaussian in $\mathbf{C}^r$. Let $W = F_n E V$. Then $\alpha = F_n E a = F_n E V \beta = W\beta$.

For a subset J of $\{1, \ldots, n\}$ of size r, let W_J be the sub-matrix of W consisting of rows indexed by J, and let $\alpha_J = (\alpha_j)_{j \in J}^T$. Observe that $\alpha_J = W_J \beta$. The covariance matrix of α_J is given by

$$\begin{aligned} \Sigma &= E[\alpha_J \alpha_J^*] \\ &= E[W_J \beta \beta^* W_J^*] \\ &= W_J E[\beta\beta^*] W_J^* \\ &= W_J W_J^*. \end{aligned}$$

The last line follows because β is standard Gaussian distributed. We get that $\det(\Sigma) = |\det(W_J)|^2$. The *Binet-Cauchy theorem* as cited in [BL02, BL04] yields

$$\sum_J |\det W_J|^2 = \det(W^* W) = \det(V^* E^* E V).$$

We *claim* now that $\det(V^* E^* E V) \geq \kappa^{-r}$, where $\kappa > 0$ is a global constant. To prove the claim, observe that for any $\beta \in \mathbf{C}^r$, if $V\beta$ is an eigenvector of E^*E with eigenvalue γ, then β is an eigenvector of V^*E^*EV with eigenvalue γ. Namely, $V^*E^*EV\beta = V^*\gamma V\beta = \gamma\beta$. For E^*E we have positive real eigenvalues $\gamma_1 \geq \ldots \geq \gamma_n > 0$. Associate a basis $v_1, \ldots, v_n$ of eigenvectors to these. Since the rank of V is r, there must be indices $i_1, \ldots, i_r$, so that $v_{i_1}, \ldots, v_{i_r}$ are in the range of V. Hence V^*E^*EV has eigenvalues $\gamma_{i_1}, \ldots, \gamma_{i_r}$. Since E is well-conditioned $\gamma_1 \leq \kappa\gamma_i$, for each i, for some global constant κ. Since $\det(E) = 1$, $\gamma_1 \geq 1$. So $\gamma_i \geq \kappa^{-1}$. Hence $\det(V^*E^*EV) = \gamma_{i_1} \cdots \gamma_{i_r} \geq \kappa^{-r}$, proving the claim.

Hence we conclude that there exists a set J such that

$$|\det(W_J)|^2 \geq \kappa^{-r} \binom{n}{r}^{-1}.$$

Applying Lemma 2.3 to the vector α_J, we get that with probability greater than $\frac{1}{2}$ that

$$\prod_{i \in J} |\alpha_i|^2 \geq \delta^r \kappa^{-r} \det(\Sigma) \geq \delta^r \binom{n}{r}^{-1},$$

where δ is an absolute constant. Hence

$$\mathrm{msv}_r^2(\mathit{Circ}(Ea)) \geq n^r \delta^r \kappa^{-r} \binom{n}{r}^{-1} \geq n^r \delta^r \kappa^{-r} 2^{-n}.$$

Hence by proposition 3.1,

$$\mathrm{msv}_r^2(Circ(Ea)D) \geq n^r \delta^r \kappa^{-3r} 2^{-n}.$$

Hence applying theorem 2.5 we get:

$$\begin{aligned} s^{bc}_{lin}(Circ(Ea)D) &\geq \log \mathrm{msv}_r(Circ(Ea)D) - n \\ &\geq \frac{r}{2} \log n - cn, \end{aligned}$$

where c is an absolute constant. $\square$

Combining the above lemma with Lemma 2.7 in the same manner as in [BL02] yields the main theorem of this section.

Theorem 3.3 *Any orbit circuit $\Gamma(Ex, Dy)$, where E and D have determinant equal to 1 and are well-conditioned, computing cyclic convolution $x \circ y$ must have $\Omega(n \log n)$ gates.*

Proof. The main idea is that the two lemmas show the existence of a value a to fix for x, so that simultaneously the values of the linear forms $\ell_1(a), \ldots, \ell_k(a)$ are manageably small and the bc-complexity of the resulting linear map in y is high. The values $\ell_1(a), \ldots, \ell_k(a)$ are small enough that the linear circuit obtained from the original bilinear bc-circuit Γ by plugging them in and deleting the "x side" can be converted into a linear bc-circuit adding not too many gates, leading to the conclusion that Γ itself must have been large.

The remaining details are mostly unaltered from [BL04]: Let $\Gamma(Ex, Dy)$ be an orbit circuit computing $x \circ y$. Fix $r = \frac{1}{2}n$. Canceling the matrices E and D, we get that $\Gamma(x, y)$ computes $Circ(E^{-1}x)D^{-1}y$. Let $f_1, \ldots, f_k$ be the linear forms computed by the circuit in $\Gamma(x, y)$. in the variables $x_1, \ldots, x_n$. To be precise, if a gate computes $c_1x_1 + \ldots + c_nx_n$, then it corresponding linear form as a vector is $(c_1, \ldots, c_n)^T$. Let $R = rig_{n-r}(f_1^T, \ldots, f_k^T)$. Observe that E^{-1} and D^{-1} have determinant 1 and are well-conditioned as well. By lemmas 3.2 and 2.7, there exists an $a \in \mathbf{C}^n$ such that:

1. $s^{bc}_{lin}(Circ(E^{-1}a)D^{-1}) \geq \frac{1}{2} r \log n - cn$, for absolute constant c, and
2. $\max_i |f_i(a)| \leq 2\sqrt{\ln 4k} R$.

Let $\alpha = \max_i |f_i(a)|$. Then $\Gamma(a, y)$ computes the linear mapping $Circ(E^{-1}a)D^{-1}$. As in [BL02], we can make this circuit into a bounded-constant linear circuit by:

1. replacing each multiplication with $f_i(a)$ with a multiplication by $2\alpha^{-1} f_i(a)$, and
2. multiplying each output with $\frac{\alpha}{2}$ using at most $\log(\frac{\alpha}{2})$ additions and one scalar multiplication of absolute value at most 2.

Letting $S(\Gamma)$ denote the size of Γ, we thus obtain a bounded-constant linear circuit that has at most $S(\Gamma) + n \log \alpha \leq S(\Gamma) + n \log(2\sqrt{\ln 4k} R)$ gates computing $Circ(E^{-1}a)D^{-1}$. We can assume $k \leq n^2$, and by the rigidity bound of theorem 2.6:

$$S(\Gamma) \geq s^{bc}_{lin}(f_1^T, \ldots, f_k^T) \geq (n - r) \log R - n. \tag{5}$$

So we obtain the inequality

$$S(\Gamma) + n \log(2\sqrt{4n^2} R) \geq n \log n - cn,$$

which together with (1) yields $S(\Gamma) = \Omega(n \log n)$. $\square$

4 Help Gates

To provide the balancer for the bounds in the next section, we note the following extension of Theorem 6.4 in [BL04].

Theorem 4.1 *For cyclic convolution of length-n vectors, the bc-complexity with $n-n/(\log n)^{1/3}$ help gates at one of the two inputs x, y is $\Omega(n(\log n)^{1/3})$.*

Proof Sketch. Following [BL04], let us describe the help gates as being on the "y-side"; by symmetry, it doesn't matter which. Thus with $l = n(\log n)^{-1/3}$, we have the case $h_1 = 0$, $h_2 = h = n-l$ of their proof of Theorem 6.4, Omitting floor notation, set $r = (h+n)/2 = n-l/2$. Given a bilinear circuit Γ that is bc except for the help gates, let $(f_1, \ldots, f_k)$ denote the linear functionals on the "x side" as before, and set $R = rig_{l/2}(f_1, \ldots, f_k)$. The conditions $h_2 = h < r < n - h_1$ needed to apply Lemma 6.5 and other referenced lemmas in [BL04] apply, and this yields

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

Substituting for r and h_1, h_2 gives

$$(1 + \frac{2n}{l})s(\Gamma) + \frac{nk}{l} + O(n \log\log n) \geq \frac{1}{4} l \log n.$$

If $k = \Omega(\frac{l^2 \log n}{n})$, then automatically the circuit Γ has $\Omega(\frac{l^2 \log n}{n})$-many gates. If $k = o(\frac{l^2 \log n}{n})$, then the nk/l term drops out asymptotically, and we get

$$s(\Gamma) = \Omega(\frac{l^2 \log n}{n}).$$

With $l = n(\log n)^{-1/3}$, we have $s(\Gamma) = \Omega(n(\log n)^{1/3}$. $\square$

5 Main Theorem on Diagonal Orbits

In this section we prove orbit lower bounds in the special case the free maps are diagonal with respect to the standard basis and of determinant equal 1.

Theorem 5.1 *Suppose $\{D_n\}_{n>0}$ is a family of sets, where each D_n is a set of n complex numbers $d_1^n, \ldots, d_n^n$ such that $\prod_{i=1}^n |d_i^n| = 1$, and suppose $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuits such that for all n,*

$$\Gamma_n(x_1 \cdot d_1^n, \ldots, x_n \cdot d_n^n, y) = x^T \mathit{Circ}(y)$$

Define $l_n = |D_n \cap (0,1)|$, and suppose $l_n = o(n)$. Then there exists $\epsilon > 0$ so that for infinitely many n,

$$s(\Gamma_n) \geq \min(\epsilon n \log n, \epsilon \frac{n^2}{l_n}).$$

Note that if $l_n \neq o(n)$, then there exists a constant $\epsilon_0 > 0$ so that for infinitely many n, $l_n \geq \epsilon_0 n$. In this case we think of the d_i^n of magnitude larger than 1 as help gates as in [BL02]. There are at most $(1-\epsilon_0)n$-many such help gates. Theorem 6.4 of [BL02] then yields that $s(\Gamma_n) \geq \epsilon n \log n$ for some fixed constant $\epsilon > 0$, and here, for infinitely many n. This observation is absorbed by our Theorem 4.1 above, for cases $l_n \geq n/(\log n)^{1/3}$, but in case of future improvements it helps to know that for $l_n \geq \epsilon_0 n$ we can get $\Omega(n \log n)$ lower bounds this way. It is a convenience to have $l_n = o(n)$ in what follows.

Proof. Without loss of generality we may suppose that each D_n consists of distinct positive real numbers. For each n, let $i_{1n}, \ldots, i_{nn}$ be such that

$$d^n_{i_{1n}} < d^n_{i_{2n}} < \ldots < d^n_{i_{nn}}.$$

In case

$$\log \prod_{j=l_n+1}^{n} d^n_{i_{jn}} = o(n \log n),$$

then we can replace the constants which are bigger than 1 by bounded constant repeated additions. Which takes at most $\sum_{j=l_n+1}^n \log d^n_{i_{jn}} = o(n \log n)$ additional gates. Hence we would obtain a family of regular bounded-coefficient bilinear circuits of size $s(\Gamma_n) + o(n \log n)$ computing $x^T \mathrm{Circ}(y)$, but such a family must have size $\Omega(n \log n)$ by [BL02]. Hence we would conclude $s(\Gamma_n) = \Omega(n \log n)$. In this case we can see that both conlusions of the remaining two cases of the theorem are satisfied.

So assume that there is a $\delta > 0$ such that for infinitely many n, $\prod_{j=l_n+1}^n d^n_{i_{jn}} > 2^{\delta n \log n}$. This implies that for infinitely many n,

$$\prod_{j=1}^{l_n} d^n_{i_{jn}} < 2^{-\delta n \log n}.$$

We first prove a lemma about minors of circulant matrices.

Lemma 5.2 *There exists an integer $c > 0$ such that for all large enough n, for any $1 \leq l \leq \frac{n}{c \log n}$ and any l distinct rows $I \subset \{0, \ldots, n-1\}$, for any U of dimension $n - l$, we have that for b standard Gaussian in U,*

$$\Pr[\textit{Circ}(b) \textit{ has an } l \times l \textit{ minor } M_0 \textit{ with rows in } I \textit{ with } |\det M_0| \geq c_0^l 2^{-(2n+l)}] > \frac{1}{2},$$

where $c_0 > 0$ is a constant depending on c.

Proof. Let $r = n - l$. Wlog. asume r is even. We have that $\mathrm{Circ}(b) = F_n \mathrm{diag}(\lambda) F_n^*$, where $\lambda = DFT_n b$. First we observe the following easy proposition about sequences of numbers.

Proposition 5.3 *For any n, for any even r with $\frac{n}{2} < r \leq n$, and for any n numbers $1 \geq \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n \geq 0$ such that $\sum_{i=1}^n \gamma_i = r$, we have that $\gamma_{\frac{r}{2}} \geq \frac{2r}{3n}$.*

Proof. For purpose of contradiction assume the opposite. Then

$$\sum_{i=1}^n \gamma_i \quad < \quad \frac{r}{2} - 1 + (n - \frac{r}{2} + 1)\frac{2r}{3n}$$

$$\begin{aligned}
&\le \frac{r}{2} - 1 + (n - \frac{r}{2} + 1)\frac{2}{3} \\
&< \frac{r}{2} - 1 + (\frac{n}{2} + 1)\frac{2}{3} \\
&\le \frac{r}{2} - 1 + (\frac{r}{2} + 1)\frac{2}{3} \\
&< r.
\end{aligned}$$

□

Lemma 5.4 *There exists a set* $S \subset \{0, 1 \ldots, n-1\}$ *of* $\frac{r}{2}$ *indices such that for any* $0 < \epsilon < 1$,

$$\Pr[\textit{for each } s \in S,\ |\lambda_s|^2 \ge \tfrac{2}{3}\epsilon] \ge 1 - \epsilon.$$

Proof. We can represent λ as $\lambda = \sqrt{n}B\beta$, where B is an $n \times r$ matrix with orthonormal columns and β standard Gaussian in $\mathbf{C}^r$. We have $||B||_F^2 = r$. Let γ_i be the ith largest $||s||_2^2$ of a row s in B. Each $\gamma_i \in [0, 1]$ since B can be extended to a unitary matrix and $\sum_{i=1}^n \gamma_i = r$. Since $r = n - l \ge n - \frac{n}{c \log n} > \frac{n}{2}$, for large enough n, we concude using above proposition that $\gamma_{r/2} \ge \frac{2r}{3n} > \frac{1}{3}$. Hence there exists a set S of $\frac{r}{2}$ rows of $\sqrt{n}B$ with $||s||_2^2 \ge \frac{n}{3}$ for each $s \in S$. Write $\lambda_s = s\beta$. By Lemma 2.7, $\frac{|\lambda_s|^2}{2||s||_2^2}$ has density function e^{-t}. Hence for any a, $\Pr[|\lambda_s|^2 \ge 2||s||_2^2 a] = e^{-a}$. So for $s \in S$, $\Pr[|\lambda_s|^2 \ge \frac{2}{3}na] \ge e^{-a}$. Set $a = \frac{\epsilon}{n}$. Using the union bound we see that

$$\Pr[\exists s \in S, |\lambda_s|^2 < \frac{2}{3}\epsilon] \le \frac{r}{2}(1 - e^{-\epsilon/n}) \le \epsilon.$$

The last inequality follows using Taylor series. Namely, $r(1 - e^{-\epsilon/n}) = r(1 - \sum_{k=0}^{\infty} \frac{(-\epsilon/n)^k}{k!}) \le n \sum_{k=1}^{\infty} -1^{(k+1)} \frac{\epsilon^k}{k! n^k} = \sum_{k=1}^{\infty} -1^{(k+1)} \frac{\epsilon^k}{k! n^{(k-1)}} \le \epsilon$. □

Let S be the set whose existence is stated in the above lemma. Fix a constant $0 < \tau < \frac{1}{2}$. Define a set of independent random variables as follows. For each $s \in S$,

$$I_s = \begin{cases} 0 & \text{with probability } 1 - \tau \\ 1 & \text{with probability } \tau. \end{cases}$$

We define the random set

$$\Omega = \{k : I_k = 1\}.$$

Each $k \in \Omega$ corresponds to frequency $\frac{2\pi n}{k}$. We will abuse notation by letting $\omega \in \Omega$ denote both the index and its corresponding frequency. It will be clear from the context which is meant.

Observe $E[|\Omega|] = \tau|S|$ and $Var[|\Omega|] = |S|(\tau - \tau^2)$. Hence using Chebyshev, for any $\delta > 0$,

$$\Pr[||\Omega| - \tau|S|| \ge \delta n] \le \frac{|S|(\tau - \tau^2)}{\delta^2 n^2} \le (\tau - \tau^2)\delta^{-2} n^{-1}.$$

Fix $\delta \in (0, \frac{\tau}{2})$. Then if n is large enough,

$$\Pr[|\Omega| \ge (\frac{\tau}{2} - \delta)n] > \frac{7}{8}.$$

For any set $T \subset \{0, 1, \ldots, n-1\}$, let $l_2(T)$ be the coordinate subspace of $\mathbf{C}^n$ of vectors with support contained in T. Let $\iota : l_2(I) \to \mathbf{C}^n$ denote the canonical embedding of $l_2(I)$ into $\mathbf{C}^n$. Its conjugate $\iota^* : \mathbf{C}^n \to l_2(I)$ is the projection map defined by setting coordinates to zero outside I. Define the linear transformation $H : l_2(I) \to \mathbf{C}^n$, by

$$Hf(t) = -\sum_{\omega\in\Omega}\sum_{t'\in I : t'\neq t} e^{i\omega(t-t')} f(t'),$$

for $f \in l_2(I)$. Observe that

$$\iota^*\iota - \frac{1}{|\Omega|}\iota^* H = \frac{n}{|\Omega|} F^n_{I,\Omega} F^{n*}_{\Omega,I}, \tag{6}$$

where $F^n_{I,\Omega}$ is the map $l_2(\Omega) \to l_2(I)$ defined by the minor of F^n with rows I and columns Ω.

Even though we have modified the distribution of Ω compared to [CRT04], we can still prove an analogue of their Theorem 3.1:

Theorem 5.5 *Let $H_0 = \iota^* H$ and let $c_\tau = e\log((1-\tau)/\tau)$. For integer k, let*

$$a_k = (2k-1)^{2k} c_\tau^{-(2k-1)} n |I|^{2k}$$

and

$$b_k = \frac{(2k)!}{k! 2^k} \left(\frac{\tau}{1-\tau}\right)^k n^k |I|^{k+1}.$$

Then

$$E[Tr(H_0^{2k})] \leq k\phi^{2k} \max(a_k, b_k), \text{ where}$$

$\phi = \frac{1+\sqrt{5}}{2}$ *is the golden ratio.*

Proof. We sketch how to adapt the proof of Theorem 3.1 in [CRT04]. Adopt the convention that $t_{2n+1} = t_1$. Define $\Omega_e = \Omega \cup \{k : I_k = 1, k \in \{0, ..., n-1\}/S\}$. This is the same distribution utilized as in [CRT04]. We have

$$\begin{aligned} E[Tr(H_0^{2k})] &= \sum_{t_1,\ldots,t_{2k} : t_j \neq t_{j+1}} E[\sum_{\omega_1,\ldots,\omega_{2k}\in\Omega} e^{i\sum_{j=1}^{2k}\omega_j(t_j - t_{j+1})}] &(7)\\ &= \sum_{t_1,\ldots,t_{2k} : t_j \neq t_{j+1}} \sum_{\omega_1,\ldots,\omega_{2k}\in S} e^{i\sum_{j=1}^{2k}\omega_j(t_j - t_{j+1})} E[\prod_{j=1}^{2k} I_{\omega_j\in\Omega}] &(8)\\ &= \sum_{t_1,\ldots,t_{2k} : t_j \neq t_{j+1}} \sum_{\omega_1,\ldots,\omega_{2k}\in S} e^{i\sum_{j=1}^{2k}\omega_j(t_j - t_{j+1})} E[\prod_{j=1}^{2k} I_{\omega_j\in\Omega_e}] &(9)\end{aligned}$$

Compare the last line with:

$$\sum_{t_1,\ldots,t_{2k} : t_j \neq t_{j+1}} \sum_{\omega_1,\ldots,\omega_{2k}\in\{0,\ldots,n-1\}} e^{i\sum_{j=1}^{2k}\omega_j(t_j - t_{j+1})} E[\prod_{j=1}^{2k} I_{\omega_j\in\Omega_e}] \tag{10}$$

In [CRT04] it is proved that (10) is bounded by $k\phi^{2k}\max(a_k, b_k)$. Hence we must show that (9) is always less than (10). Define K_0 to be the $|I| \times |I|$ matrix defined by $K_0(t, t') = \sum \omega \in \Omega_e/\Omega e^{i\omega(t-t')}$, if $t \neq t'$, and 0 otherwise. Define $H_0' = H_0 + K_0$. We have that (10) is equal to $E[tr(H_0'^{2k})]$. The remainder follows upon showing that $E[tr(H_0^{2k})] \leq E[tr(H_0'^{2k}]]$. For $k = 1$, we have $E[tr(H_0'^2)] = E[tr((H_0 + K_0)^2)] = E[tr(H_0^2)] + 2E[tr(H_0 K_0)] + E[tr(K_0^2)]$. We are thus done with this step if we show that $E[tr(H_0 K_0)]$ vanishes. We have:

$$\begin{aligned} tr(H_0 K_0) &= \sum_{t,t' \in T} H^0_{t,t'} K^0_{t't} \\ &= \sum_{t \neq t' \in T} \sum_{\omega \in \Omega} e^{i\omega(t-t')} \sum_{\omega \in \Omega_e/\Omega} e^{i\omega(t'-t)} \\ &= \sum_{t \neq t' \in T} \sum_{0 \leq \omega_1, \omega_2 < n} e^{i\omega(t-t')(\omega_1 - \omega_2)} I_{\omega_1 \in \Omega} I_{\omega_2 \in \Omega_e/\Omega}. \end{aligned}$$

Note the two indicators appearing in each term are independent and have expectation τ. Hence when taking expectation we obtain:

$$\begin{aligned} \tau^2 \sum_{t \neq t' \in T} \sum_{0 \leq \omega_1, \omega_2 < n} e^{i\omega(t-t')(\omega_1 - \omega_2)} &= \tau^2 \sum_{t \neq t' \in T} \sum_{0 \leq \omega_1 < n} e^{i\omega(t-t')\omega_1} \sum_{0 \leq \omega_2 < n} e^{i\omega(t'-t)\omega_2} \\ &= 0. \end{aligned}$$

[1] □

Continuing with the proof of Lemma 5.2, let $k = \lceil \log n \rceil$ and apply Theorem 5.5. Using the Markov inequality,

$$\Pr[Tr(H_0^{2k}) \geq 8k\phi^{2k} \max(a_k, b_k)] \leq \frac{1}{8}$$

Hence

$$\Pr[Tr(H_0^{2k}) < 8k\phi^{2k} \max(a_k, b_k) \text{ and } |\Omega| \geq (\frac{\tau}{2} - \delta)n] > \frac{3}{4}.$$

Since H_0 is self-adjoint we have

$$||H_0||_2^{2k} = ||H_0^k||_2^2 \leq ||H_0^k||_F^2 = Tr(H_0^{2k}).$$

Hence with probability bigger than 3/4 we have that

$$||H_0||_2 < 8^{\frac{1}{2k}} k^{\frac{1}{2k}} \phi \max(a_k, b_k)^{\frac{1}{2k}}$$

together with $|\Omega| \geq (\frac{\tau}{2} - \delta)n$. Since $\iota^*\iota$ is the identity on $l_2(I)$. We only need to to verify that

$$||H_0||_2 \leq \epsilon_0 n, \tag{11}$$

for some constant $0 < \epsilon_0 < (\frac{\tau}{2} - \delta)$ to be guaranteed that with probability bigger than 3/4,

$$\iota^*\iota - \frac{1}{|\Omega|} H_0$$

has determinant at least

$$(1 - \frac{\epsilon_0}{\tau/2 - \delta})^l.$$

[1]Note added late: We have found a gap in our obtaining this inequality for all k, although it appears to be an exercise in estimates.

using Weyl's perturbation theorem. See for example, [Bha97]. We have

$$\begin{aligned}8^{\frac{1}{2k}}k^{\frac{1}{2k}}\phi a_k^{\frac{1}{2k}} &= 8^{\frac{1}{2k}}k^{\frac{1}{2k}}\phi(2k-1)c_\tau^{\frac{-(2k-1)}{2k}}n^{\frac{1}{2k}}|I| \\ &= O(1)\cdot(2k-1)|I| \\ &\leq O(1)\cdot(2\lceil\log n\rceil-1)\frac{n}{c\log n}.\end{aligned}$$

and using $\frac{(2k)!}{k!2^k} \leq 2^{k+1}e^{-k}k^k$ we have

$$\begin{aligned}8^{\frac{1}{2k}}k^{\frac{1}{2k}}\phi b_k^{\frac{1}{2k}} &\leq 8^{\frac{1}{2k}}k^{\frac{1}{2k}}\phi 2^{\frac{k+1}{2k}}e^{-1/2}k^{\frac{1}{2}}(\frac{\tau}{1-\tau})^{\frac{1}{2}}n^{\frac{1}{2}}|I|^{\frac{k+1}{2k}} \\ &= O(1)\cdot k^{\frac{1}{2}}n^{\frac{1}{2}}|I|^{\frac{1}{2}+\frac{1}{2k}} \\ &= O(1)\cdot\frac{n}{\sqrt{c}}.\end{aligned}$$

So we conclude that if we choose c large enough, that inequality (11) is satisfied.

Lemma 5.6 *Let A be the minor of F_n consisting of rows I and columns Ω. Then*

$$\Pr[|det(AA^*) \geq c_0^l)] > \frac{3}{4},$$

where $c_0 > 0$ is a constant depending on c.

Proof. By equation 6,

$$AA^* = \frac{|\Omega|}{n}(\iota^*\iota - \frac{1}{|\Omega|}\iota^* H)$$

So

$$\det(AA^*) \geq (\frac{|\Omega|}{n})^l(1-\frac{\epsilon_0}{\tau/2-\delta})^l \geq (\frac{\tau}{2}-\delta)^l(1-\frac{\epsilon_0}{\tau/2-\delta})^l$$

with probability greater than $\frac{3}{4}$. $\square$

Corollary 5.7 *With probability bigger than 3/4 there exists an $l \times l$ minor M of F_n with rows in I and columns in Ω such that*

$$|det(M)|^2 \geq c_0^l 2^{-n},$$

where $c_0 > 0$ is a constant depending on c.

Proof. Using Binet-Cauchy,

$$\sum_{|J|=l}|\det(A_{I,J})|^2 = \det(AA^*) \geq c_0^l.$$

Since there are at most 2^n terms in this sum the corollary follows. $\square$

Set $\epsilon = 1/4$. Let J_M be the columns of M. By Lemma 5.4, with probability at least 3/4, all $|\lambda_s|^2$ with $s \in J_M$ are at least 2ϵ. Hence with probability bigger than 1/2 we have this happens together with the event described in Corollary 5.7. Using Binet-Cauchy once more, we get

$$\begin{aligned}\det(\mathrm{Circ}(b)\mathrm{Circ}(b)^*)_{I,I} &= \det(F_n \mathrm{diag}(|\lambda_i|^2)F_n^*)_{I,I}) \\ &= \sum_{|J|=l} \prod_{j\in J} |\lambda_j|^2 |\det(F^n_{I,J})|^2 \\ &\geq \prod_{j\in J_M} |\lambda_j|^2 |\det(M)|^2 \\ &\geq (2\epsilon)^l |\det(M)|^2.\end{aligned}$$

Using Binet-Cauchy again, we see that

$$\begin{aligned}\sum_{|J|=l} |\det(\mathrm{Circ}(b)_{I,J})|^2 &= \det(\mathrm{Circ}(b)\mathrm{Circ}(b)^*)_{I,I} \\ &\geq (2\epsilon)^l |\det(M)|^2.\end{aligned}$$

Hence there exists an $l \times l$ minor M_0 of $\mathrm{Circ}(b)$ with $|\det(M_0)|^2 \geq (2\epsilon)^l |\det(M)|^2 2^{-n} \geq (2\epsilon)^l c_0^l 2^{-2n} = c_0^l 2^{-(2n+l)}$. This finishes the proof of Lemma 5.2. □

Popping back into the proof of Theorem 5.1, take c to be the constant provided by Lemma 5.2. Let $m_n = \min(l_n, \lfloor \frac{n}{c \log n} \rfloor)$, and let $e_n = \frac{l_n}{m_n}$. We have for infinitely many n that

$$\prod_{j=1}^{m_n} \frac{1}{d^n_{i_j n}} > 2^{\frac{\delta}{e_n} n \log n}. \tag{12}$$

Let us consider some large enough n for which (12) holds, and let us drop the sub and superscipts n on our variables.

We are going to perform the following substitution on the circuit. Set $x_{i_j} = 0$ for all $j > m$ and substitute $x_{i_j} = z_j / d_{i_j}$ otherwise. This yields a bounded coefficient bilinear circuit of no size no bigger than $s(\Gamma)$, and it computes

$$(z_1, \ldots, z_m)\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_m}^{-1})M,$$

where M is the $m \times n$ minor of $\mathrm{Circ}(y)$ corresponding to rows $I := \{i_1, \ldots, i_m\}$.

Now set $r = n - m$. Let $f_1, \ldots, f_k$ be the linear forms in y of Γ. Lemma 2.7 provides us with a linear subspace U of dimension $n - m$ such that for standard Gaussian $b \in_R U$, we have that

$$\Pr[\max_i |f_i(b)| \leq 2\sqrt{\ln 4k} \cdot rig_m(f_1^T, \ldots, f_k^T)] \geq \frac{1}{2}.$$

Applying lemma 5.2, we conclude there must exist an input b for y such that we have both of the following conditions:

1. $\mathrm{Circ}(b)$ has an $m \times m$ minor M_0 with rows in I with $|det M_0| \geq {c_0}^m 2^{-(2n+m)}$, and
2. $\max_i |f_i(b)| \leq 2\sqrt{\ln 4k} \cdot rig_m(f_1^T, \ldots, f_k^T)$.

Hence we obtain from Γ_n a bounded coefficient *linear* circuit computing the $\mathbf{C}^m \to \mathbf{C}^m$ map

$$(z_1, \ldots, z_m)\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_m}^{-1})M_0,$$

by removing the outputs not corresponding to M_0 and replacing multiplications with $f_i(b)$ by $f_i(b)/\mu$ and correcting this by adding at most $m \log \mu$ repeated additions at the output gates, where $\mu = \max_i |f_i(b)|$.

By Theorem 2.6,

$$\log rig_m(f_1^T, \ldots, f_k^T) \leq \frac{s(f_1, \ldots, f_k) + n}{m}.$$

Hence the number of gates we added is at most

$$\begin{aligned} m \log \mu &\leq m \log \sqrt{\ln 4k} + m \log rig_m(f_1^T, \ldots, f_k^T) \\ &\leq m \log \sqrt{\ln 4n^2} + m \frac{s(\Gamma) + n}{m} \\ &= s(\Gamma_n) + O(n). \end{aligned}$$

However, by Morgenstern's bound any bounded coefficient circuit computing $\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_m}^{-1}) M_0$ requires a number of gates at least

$$\begin{aligned} \log |\det\ \mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_m}^{-1}) M_0| &\geq \log 2^{\frac{\delta}{e_n} n \log n} c_0^m 2^{-(2n+m)} \\ &= \frac{\delta}{e_n} n \log n - O(n) \\ &= \frac{\min(l_n, \lfloor \frac{n}{c \log n} \rfloor)}{l_n} \delta n \log n - O(n) \\ &\geq \min(\delta n \log n, \frac{\delta}{2c} \cdot \frac{n^2}{l_n}) - O(n). \end{aligned}$$

□

Proof of Main Theorem. In case $l_n \leq n/(\log n)^{1/3}$, Theorem 5.1 provides a lower bound of $\Omega(n^2/l_n) = \Omega(n(\log n)^{1/3})$. In case $l_n \geq n/(\log n)^{1/3}$, then we are in a case of $n - l_n \leq n - n/(\log n)^{1/3}$ help gates, and the conclusion follows from Theorem 4.1. □

The balancing here in general is between n^2/l_n and $\frac{l_n^2}{n} \log n$, and setting them asymptotically equal gives $n^3 = l_n^3 \log n$, hence the choice of breakpoint for l_n.

6 Conclusions

We have proved super-linear lower bounds for a circuit model midway between the bounded-coefficients and the unbounded-coefficients case. It appears these bounds should also carry over to the $SL_n(\mathbf{C})$ action, in view of the SVD $A = UDV$ into unitary and diagonal matrices, but this is not immediate as the $SL_n(\mathbf{C})$ action involved is not simply a product of the unitary and GLD actions in our theorems.

The obvious desire is to improve the $(\log n)^{1/3}$ factor to $\log n$ itself. By our present techniques, this appears to require major improvements to estimates by Candès, Romberg, and Tao [CRT04], and also showing that these improvements survive our tweaking of the distributions. Candès, Romberg, and Tao do conjecture the needed improvements, and cite computer runs supporting this in [CRT04], but remarks in [CT04] and [Tao04] indicate that such a proof is still far off.

Acknowledgments We thank Peter Bürgisser and Ran Raz for comments on early stages of this work. We thank Ryan O'Donnell for pointing us toward the work by Terence Tao while the second author was at a Schloss Dagstuhl workshop in October. Last and most, we thank Terence Tao himself for help understanding the analytical issues involved.

References

[Bha97] R. Bhatia. *Matrix Analysis.* Springer Verlag, 1997.

[BL02] P. Bürgisser and M. Lotz. Lower bounds on the bounded coefficient complexity of bilinear maps. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 659–668, 2002.

[BL04] P. Bürgisser and M. Lotz. Lower bounds on the bounded coefficient complexity of bilinear maps. *J. Assn. Comp. Mach.*, 51:464–482, 2004.

[BS82] W. Baur and V. Strassen. The complexity of partial derivatives. *Theor. Comp. Sci.*, 22:317–330, 1982.

[Cha98] B. Chazelle. A spectral approach to lower bounds, with application to geometric searching. *SIAM J. Comput.*, 27:545–556, 1998.

[CRT04] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Technical report, California Institute of Technology, 2004. arXiv.math.CA.

[CT04] E.J. Candès and T. Tao. Near optimal signal recovery from random projections and universal encoding strategies. Technical report, California Institute of Technology, 2004. arXiv.math.CA/0410542 v1.

[GvL96] G.H. Golub and C. van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, 1996.

[Lok01] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. *J. Comp. Sys. Sci.*, 63, 2001.

[Mor73] J. Morgenstern. Note on a lower bound of the linear complexity of the fast Fourier transform. *J. Assn. Comp. Mach.*, 20:305–306, 1973.

[MS02] K. Mulmuley and M. Sohoni. Geometric complexity theory I: An approach to the P vs. NP and related problems. *SIAM J. Comput.*, 31(2):496–526, 2002.

[Mul99] K. Mulmuley. Lower bounds in a parallel model without bit operations. *SIAM J. Comput.*, 28:1460–1509, 1999.

[Raz02] R. Raz. On the complexity of matrix product. In *Proc. 34th Annual ACM Symposium on the Theory of Computing*, pages 144–151, 2002. Also ECCC TR 12, 2002.

[Raz03] R. Raz. On the complexity of matrix product. *SIAM Journal of Computing*, 32(5):1356–1369, 2003.

[Tao04] T. Tao, 11/2004. Personal communication.