Orbits and Arithmetical Circuit Lower Bounds

Maurice J. Jansen[1] and Kenneth W.Regan[⋆1]

University at Buffalo

Abstract. The orbit of a regular function f over a field F under action by a matrix group G is the collection of functions $f(Ex)$ for $E \in G$. We show that some lower bounds of Bürgisser and Lotz [BL03] and Shpilka and Wigderson [SW99] in restricted arithmetical circuit/formula models extend to orbits, where E does not count against the complexity bounds and is not subject to the (same) restriction(s). Our "orbit model" and a second "linear-combination model" aim to bridge the gap between the bounded-coefficient linear/bilinear circuit model of [Mor73,NW95,Raz03,BL03] and the arbitrary-coefficient case. We extend size-depth tradeoff methods of Lokam [Lok01] to the latter model. Variants of the Baur-Strassen "Derivative Lemma" are developed, including one that can be iterated for sums of higher partial derivatives.

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 [BL03] extended Raz's ideas to obtain tight $\Omega(n \log n)$ bounds on bc-circuits for *cyclic convolution*, and thence for polynomial multiplication and related bi-linear functions. These lower bounds hold even when the bc-restriction is lifted for $O(n^{1-\epsilon})$-many "help gates." The natural question is, can one obtain similar lower bounds without the bc-restriction at all?

A flag of difficulty is that the best-known lower bounds on general arithmetical circuit size of *any explicit family of $n^{O(1)}$-degree polynomials* $f(x_0, \dots, x_{n-1})$ 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 $(\partial f) : F^n \to F^n = (\partial f/\partial x_1, \dots, \partial f/\partial x_n)$. They apply to functions like $f = x_0^n + \dots + x_{n-1}^n$ such that the *geometric degree* of the ideals I_f in $F[x_0, \dots, x_{n-1}, y_0, \dots, y_{n-1}]$ generated by $(y_0 - \partial f/\partial x_0, \dots, y_{n-1} - \partial f/\partial x_{n-1})$ 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,BL03] 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

[⋆] 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

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]. Note that this model not only works past the above-mentioned $O(n^{1-\epsilon})$ limit on "help" gates with unbounded constants, it also does not constrain the linear circuit complexity of D and E themselves, which may be as high as quadratic. We note first that taking G to be all of $SL_n(\mathbf{C})$, the group of complex matrices of determinant 1, is close to the arbitrary-coefficients case from the standpoint of lower bounds. This means, however, that partial progress should further restrict either the matrices D, E or some other aspect of the circuits. We extend the lower bounds in [BL03] when D, E (also) have bounded *condition number*, and we extend the size-depth tradeoff method of Lokam [Lok01] to handle some other cases of orbits.

Orbits have no effect on the multiplicative-complexity (ℓ_*) measure for sum-of-products-of-sums ($\Sigma\Pi\Sigma$) formulas (always with arbitrary coefficients) used by Shpilka and Wigderson [SW99]. We show that they do matter for additive-complexity (ℓ_+) and overall formula size ($\ell = \ell_+ + \ell_*$), by proving $\Omega(n^{4/3})$ $\Sigma\Pi\Sigma$-formula size lower bounds on a tri-linear function whose orbit under triple-action by unitary matrices includes a polynomial with linear-size $\Sigma\Pi\Sigma$ formulas. The function is the inner product of cyclic convolution with a third vector of variables. For several functions in [SW99] we combine a closed-form derivative lemma for $\Sigma\Pi\Sigma$ formulas with their methods to obtain lower bounds on ℓ_+ and ℓ that are sharper for low degree than those they gave for ℓ_*, and that do carry over to entire $GL_n(\mathbf{C})$ orbits.

Finally, in place of bc-linear or bc-bilinear circuits computing m-ary functions $f = (f_1, \ldots, f_m)$, we consider bc circuits with $n + m$ inputs and one output that compute $L_f = \sum_{j=1}^{m} a_j f_j(x_0, \ldots, x_{n-1})$. As reported in [BL03], Pudlák has noted the counterexample $f = 2^n \sum_{i=0}^{n-1} x_i y_i$ to the erroneous assertion in [NW95] that the Derivative Lemma preserves asymptotic bc-complexity. Nevertheless, we give a new form of the Derivative Lemma that preserves asymptotic bc-complexity in passing from f to $L_{\partial f}$. This form can be iterated, yielding for example circuits of size $O(s(f))$ with inputs $x_0, \ldots, x_{n-1}, a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ that compute

$$\sum_{i,j=0}^{n-1} a_i b_j \frac{\partial^2 f}{\partial x_i \partial x_j},$$

and likewise bc-circuits of size $O(s_{bc}(f))$. Going to L_f rules out the way spectral methods are applied in [Mor73,Raz03,BL03], because all diagonal matrices D with entries of absolute value at most 2^n have linear-size circuits for L_D. Nevertheless, we show that the spectra/rigidity technique used by Lokam [Lok01] can be extended to prove size-depth tradeoffs on L_f for certain f that are vectors of linear transformations. A short concluding section assesses the prospects of extending the two bridges to prove strongly super-linear lower bounds on general arithmetic circuits.

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 for 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 $f = (f_1, \ldots, f_n) = T_3(\ell_1 r_1, \ldots, \ell_k r_k)$.

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 size-depth tradeoffs. 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, and likewise for linear circuits. Thus we may restrict attention to (bi-)linear circuits for the functions we consider. The complexity of a linear transformation A for bounded-coefficient 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

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

for $0 \leq k < n$. When fixing $x = a = (a_0, \ldots, a_{n-1})^T$, the induced map on y is computed by the circulant matrix *Circ*(a) at left. We also find it convenient to consider the "half convolution" defined by *HCirc*$(x)y$, where *HCirc*(a) is the lower-triangular matrix at right:

$$Circ(a) = \begin{pmatrix} a_0 & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \cdots & a_3 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n-2} & a_{n-3} & \cdots & a_0 & a_{n-1} \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{pmatrix}, \qquad HCirc(a) = \begin{pmatrix} a_0 & 0 & \cdots & 0 & 0 \\ a_1 & a_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \\ a_{n-2} & a_{n-3} & \cdots & a_0 & 0 \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{pmatrix}.$$

Then 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, gives $x \circ y$. 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 1 below. Now we define our orbit model in the bilinear case.

Definition 1. *Let E and D be $n \times n$ non-singular complex matrices. An* orbit circuit *is the composition $\Gamma(Ex, Dy)$, 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 a bc-bilinear orbit circuit Γ of the same size for some 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 functions f that depend on all of their inputs, then techniques that tolerate two unbounded "help gates" (not needing the $n^{1-\epsilon}$ 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 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. □

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. We refer to 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 1 ([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 $\|f\|^2$. And $T := \frac{|S|^2}{2\|f\|^2}$ 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, the real and imaginary parts of each component z_i are normal distributed random variables with mean 0.

Lemma 2 ([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* $\Pr[|z_1|^2 \cdots |z_r|^2 \geq \delta^r \det(\Sigma)] > 1/2$*, 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. 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 ([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*

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

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 ([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 3 ([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 3. *For $A \in \mathbf{C}^{m\times n}$, and $1 \le r \le m$, 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 3 ([BL02]). *Let $f_1, \ldots, f_k$ be linear forms and $1 \le r < n$. 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*

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

3 Well-Conditioned Orbit Circuits

In this section, we will consider orbit circuits $\Gamma(Ex, Dy)$ for which matrices E and D are *well conditioned* in the following traditional sense, with reference to [GvL96].

Definition 4. *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 equals the product $||E||_2 \cdot ||E^{-1}||_2$. We fix some absolute constant κ_1, and stipulate that a* well-conditioned *matrix E has $\kappa(E) \le \kappa_1$.*

Unitary matrices have condition number 1. That the results of [BL02,BL03] 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:

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

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 $b_i \ge \kappa(B)^{-1}$. Hence $\mathrm{msv}_r^2(AB) =$

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

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]. The proof is in Appendix 1.

Lemma 4. *Let $1 \le r < n$, 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*

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

where c is some absolute constant.

Combining this with Lemma 3 in the same manner as in [BL02] yields the main theorem of this section.

Theorem 4. *Any orbit circuit $\Gamma(Ex, Dy)$, where E and D have determinant 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. For completeness the remaining details, mostly un-altered from [BL02], are in Appendix 1. □

We are able to drop the well-conditioning requirement in the above theorem for circuits Γ with exactly n multiplication gates, but have not been able to push that even to $n + 1$.

4 Orbits of $\Sigma\Pi\Sigma$-Formulas

We consider orbit circuits of the form $C(Ex)$, where $E \in GL_n(\mathbf{C})$ and C is a $\Sigma\Pi\Sigma$-formula—namely, a formula with one layer of unbounded-fanin $+$ gates, one layer of unbounded-fanin $*$ gates, and output $+$ gate(s) of unbounded fanin. In this entire section, constants on wires are unrestricted. Shpilka and Wigderson [SW99] proved super-linear lower bounds on the multiplicative complexity ℓ_*, which counts the wires fanning in to the layer of $*$ gates, on $\Sigma\Pi\Sigma$ formulas for certain natural families of polynomials that we also analyze below. We obtain sharper and more general bounds, but on additive complexity and total formula size instead. Let $l_3^o(f)$ denote the smallest number of wires for a $\Sigma\Pi\Sigma$-formula C for which there exists invertible matrix E such that $C(Ex) = f$. Regular $\Sigma\Pi\Sigma$-formula size, that is fixing E to be the identity map in the above, is denoted by $l_3(f)$. We refer to [SW99] for definitions and basic results used in the following.

To separate the orbit-$\Sigma\Pi\Sigma$-formula model from the original, for ℓ and ℓ_+, consider the tri-linear function $g = z^T \text{Circ}(x) y$. In Theorem 10, we show an $\Omega(n^{4/3})$ lower bound on $\ell_+(g)$. To find a polynomial h in the orbit of g that has $O(n)$ $\Sigma\Pi\Sigma$-formula size, apply DFT_n^{-1} to x, F_n to y and F_n^{-1} to z. Since $\text{Circ}(x) = F_n \text{diag}(\lambda) F_n^{-1}$ for $\lambda = DFT_n x$, we get $z^T F_n^{-1} \text{Circ}(DFT_n^{-1} x) F_n y = z^T \text{diag}(x) y^T$. We can divide out constants to make the three matrices unitary. To show which lower bounds *do* extend to orbits, we first extend a lemma of [SW99].

Lemma 5. *Let $g \in \mathbf{C}[y_1, \ldots, y_n]$ and let $E \in GL_n(\mathbf{C})$. Suppose $f = g(Ex)$. If it holds that for every affine subspace A of codimension κ, $dim(\partial_d(f)_{|A}) > D$, then also for every affine subspace B of codimension κ, $dim(\partial_d(g)_{|B}) > D$.*

Proof. Suppose there exists an affine subspace B of codimension κ such that $dim[\partial_d(g)_{|B})] \leq D$. Let $S = \partial_d(g)$, $S(Ex) = \{s(Ex) : s \in S\}$ and $T = \partial_d(f)$. Observe that $T \subseteq \text{span}(S(Ex))$. Suppose restriction to B is represented by the substitution $(Bx + b)$. $E^{-1}B$ is also affine of codimension κ,

$$dim[\partial_d(f)_{|E^{-1}B}] = dim[\{p(E^{-1}Bx + E^{-1}b) : p \in T\}]$$

Since $\{p(E^{-1}Bx + E^{-1}b) : p \in T\}$ is contained in the span of $S(Bx + b)$, we obtain a contradiction. □

Theorem 5. *Let $f \in \mathbf{C}[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of codimension κ, $dim(\partial_{d+1}(f)_{|A}) > D$. Then*

$$l_3^o(f) \geq \min(\frac{\kappa^2}{d+2}, \frac{D}{\binom{\kappa+d}{d}}).$$

Proof. Suppose $f = C(Ex)$, where C is a $\Sigma\Pi\Sigma$ formula with $l_3^o(f)$ many wires and E is some invertible matrix. Write Let $g = C(y)$. By lemma 5, for any affine A of codimension κ,

$$\sum_{i=1}^{n} dim[\partial_d(\frac{\partial g}{\partial y_i})_{|A}] \geq dim[\partial_{d+1}(g)_{|A}] > D. \tag{1}$$

Let $M_1, \ldots, M_s$ be the multiplication gates of C. We have that $g = \sum_{i=1}^{s} M_i$, where for $1 \leq i \leq s$, $M_i = \Pi_{j=1}^{d_i} l_{i,j}$ with $deg(l_{i,j}) = 1$ and $d_i = indeg(M_i)$. Write $l_{i,j} = c_{i,j,1}y_1 + c_{i,j,2}y_2 + \ldots + c_{i,j,n}y_n + c_{i,j,0}$. Computing the partial derivative of g w.r.t. variable y_k we get:

$$\frac{\partial g}{\partial y_k} = \sum_{i=1}^{s} \sum_{j=1}^{d_i} c_{i,j,k} \frac{M_i}{l_{i,j}} \ . \tag{2}$$

This embodies a closed-form of the Baur-Strassen Derivative Lemma for $\Sigma\Pi\Sigma$ formulas, and confers extra power for ℓ_+ over the methods in [SW99]. Let $S = \{i | dim(M_i^h) \geq \kappa\}$. If $|S| \geq \frac{\kappa}{d+2}$, then $l_3^o(f) \geq \frac{\kappa^2}{d+2}$. Suppose $|S| < \frac{\kappa}{d+2}$. If $S = \emptyset$, then let A be an arbitrary affine subspace of codimension κ. Otherwise, we have $d + 2 < \kappa$. It is possible to pick $d + 2$ input linear forms $l_{j,1}, \ldots, l_{j,d+2}$ of each multiplication gate M_j with $j \in S$, such that $\{l_{j,1}^h, \ldots, l_{j,d+2}^h | j \in S\}$ is a set of at most κ independent homogeneous linear forms. Define $A = \{y | l_{i,j}(y) = 0, i \in S, j \in [d+2]\}$. We have $codim(A) \leq \kappa$. Wlog. assume $codim(A) = \kappa$. For each $i \in S$, d+2 linear forms of M_i vanish on A. This implies that

$$dim(\partial_d(\frac{M_i}{l_{i,j}})_{|A}) = 0, \quad \text{while for } i \notin S, \quad dim(\partial_d(\frac{M_i}{l_{i,j}})_{|A}) < \binom{\kappa+d}{d}$$

by Proposition 2.3 in [SW99]. Let $D_k = dim(\partial_d(\frac{\partial g}{\partial y_k})_{|A})$. By equation (1), $\sum_{k=1}^{n} D_k > D$. By Proposition 2.2 of [SW99] and equation (2),

$$D_k \leq \sum_{\substack{i,j \\ c_{i,j,k} \neq 0}} dim(\partial_d(\frac{M_i}{l_{i,j}})_{|A}).$$

Hence there must be at least $\frac{D_k}{\binom{\kappa+d}{d}}$ terms on the RHS, i.e. there are at least that many wires from y_k to gates in the next layer. Hence in total the number of wires to fanning out from the inputs of C is at least $\sum_{i=1}^{n} \frac{D_i}{\binom{\kappa+d}{d}} > \frac{D}{\binom{\kappa+d}{d}}$. □

In case we just want to prove lower bounds on $\Sigma\Pi\Sigma$-formula size, we observe that the above proof actually yields the following:

Theorem 6. *Let $f \in \mathbf{C}[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of codimension κ, $\sum_{i=1}^{n} dim(\partial_d(\frac{\partial f}{\partial x_i})_{|A}) > D$. Then*

$$l_3(f) \geq \min(\frac{\kappa^2}{d+2}, \frac{D}{\binom{\kappa+d}{d}}).\square$$

Typically, in applications the expression $\sum_{i=1}^{n} dim(\partial_d(\frac{\partial f}{\partial x_i})_{|A})$ is bounded from below by considering $dim(\partial_{d+1}(f)_{|A})$, and hence we get no stronger lower bound without orbits. However, we'll see one example where we do need to call upon Theorem 6 to prove a nonlinear $\Sigma\Pi\Sigma$-formula lower-bound, while there exists no such lower bound in the orbit model.

4.1 Lower Bounds

The degree-d symmetric polynomial in n variables is defined to be $S_n^d = \sum_{|I|=d} \prod_{i \in I} x_i$.

Theorem 7. *For $1 \le d \le \log n$, $l_3^o(S_n^{2d}) = \Omega(\frac{n^{\frac{2d}{d+1}}}{d+1})$.*

Proof. By lemma 4.14 in [SW99], for any affine subspace A of codimension κ and $d \ge 0$,

$$dim(\partial_{d+1}(S_n^{2d+2})_{|A}) \ge \binom{n-\kappa}{d+1}$$

Applying Theorem 5 we get that

$$l_3(S_n^{2d+2}) \ge \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d+1}}{\binom{\kappa+d}{d}}) = \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{n-\kappa-d-1}{d+1}) \ge \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{n-2\kappa}{d+1})), \tag{3}$$

subject to the condition $(d+1) < \kappa$. Set $\kappa = \frac{1}{9}n^{\frac{d+1}{d+2}}$. Then we have that

$$\frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{n-2\kappa}{d+1} \ge (\frac{n-\kappa}{\kappa+d})^d\frac{n-2\kappa}{d+1} \ge (\frac{8/9n}{2/9n^{\frac{d+1}{d+2}}})^d\frac{n-2\kappa}{d+1} = 4^d n^{\frac{d}{d+2}}\frac{n-2\kappa}{d+1} \ge \frac{n^{\frac{2d+2}{d+2}}}{d+1}.$$

Hence (3) is at least $\min(\frac{n^{\frac{2d+2}{d+2}}}{81(d+2)}, \frac{n^{\frac{2d+2}{d+2}}}{d+1}) = \Omega(\frac{n^{\frac{2d+2}{d+2}}}{d+2})$. □

Comparing this with [SW99] we see that we get a slightly stronger lower bound on the size of an orbital $\Sigma\Pi\Sigma$-formula if we count the wires in the first layer rather than the wires in the multiplication layer. The number of wires in the multiplication layer of a formula for S_n^{2d} was bounded there to be $\Omega(\frac{n^{2d/(d+2)}}{d})$ for $d \le \log n$. This result includes the extension to the orbit model, since the composition with an invertible linear map does not alter the number of wires in the multiplication layer. Our results show that the extension to the orbit model is also possible, even when counting wires at the input layer of the circuit.

Definition 5. *Define product-of-inner-products by $PIP_n^2 = (\sum_{j=1}^{n} a_j b_j)(\sum_{i=1}^{n} c_i d_i)$.*

Theorem 8. *$\ell_3^o(PIP_n^2) = \Omega(n^{4/3})$.*

Proof. Set $d = 1, \kappa = n^{2/3}$. Observe that $\frac{\partial PIP_n^2}{\partial a_i c_j} = b_i d_j$. Let A be any affine subspace of codimension κ with basis B. At least $n - \kappa$ variables in $\{b_1, \ldots, b_n\}$ are not in B. Symmetrically, at least $n-\kappa$ variables in $\{d_1, \ldots, d_n\}$ are not in B. So for at least $(n-\kappa)^2$ indices (i,j), $\frac{\partial PIP_n^2}{\partial a_i c_j}_{|A} = \frac{\partial PIP_n^2}{\partial a_i c_j}$. These are independent terms, hence $dim(\partial_2(PIP_n^2)_{|A}) \ge (n-\kappa)^2$. Applying theorem 5 we get that $\ell_3^o(PIP_n^2) \ge \min(\frac{n^{4/3}}{3}, \frac{(n-n^{2/3})^2}{n^{2/3}+1}) = \Omega(n^{4/3})$. □

We can generalize this to taking d inner products:

Definition 6. *Define the product-of-d-inner-products polynomial by* $PIP^d_n = \prod_{i=1}^{d}(\sum_{j=1}^{n} a^i_j b^i_j)$*, for variables* a^i_j, b^i_j *with* $i, j \in \{1, \ldots, n\}$*.*

Theorem 9. *For constant* $d \geq 2$*,* $\ell^o_3(PIP^d_n) = \Omega(n^{\frac{2d}{d+1}})$*.*

Compare with $\ell^*_3(PIP^d_n) = \Omega(n^{\frac{2d}{d+2}})$ in [SW99], which for the special case $d = 2$ even becomes trivial. As far as we know the above theorem is the first non-linear lower bound on the (orbital) $\Sigma\Pi\Sigma$-formula size of PIP^2_n.

Theorem 10. $\ell_3(z^T Circ(x)y) = \Omega(n^{\frac{4}{3}})$*.*

Proof. Let $f = z^T\text{Circ}(x)y$. Apply theorem 6 for $d = 1$. Since $\partial_1(\frac{\partial f}{\partial z_i})$ contains all variables $x_1, \ldots, x_n$, we conclude $dim[\partial_1(\frac{\partial f}{\partial z_i})_{|A}]$ is at least $n - \kappa$ for any affine A of codimension κ. Hence $\ell_3(f) \geq \min(\kappa^2/3, \frac{n(n-\kappa)}{\kappa+1})$. Taking $\kappa = n^{2/3}$ yields $\ell_3(f) = \Omega(n^{4/3})$. □

Note that $z^T\text{Circ}(x)y$ can be computed in $O(n \log n)$ size using a bounded constant $\Sigma\Pi\Sigma\Pi\Sigma$ circuit, and also note that theorem 3.1 and 3.2 of [SW99] are rendered useless for this polynomial, because the dimension of the set of first partials and also the dimension of the set of second partials is just $O(n)$. Indeed, we cannot prove a non-linear lower bound on $\ell^o_3(z^T\text{Circ}(x)y)$ because this is $O(n)$ as shown at the top of this section! The proof of our last lower bound is in the Appendix:

Definition 7. *For* $d \geq 1$*, define the linear-sum of the product of* d $n \times n$ *matrices* $X^1, \ldots, X^d$ *to be the polynomial* $LMM_d = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(X^1 \cdot X^2 \ldots \cdot X^d)_{ij}$

Theorem 11. *For constant* $d \geq 1$*,* $\ell^o_3(LMM_{2d+1}) = \Omega(n^{4-\frac{4}{d+2}})$*.*

5 Derivative Lemmas and Linear Combinations

In this section inputs are not considered gates, fan-in is bounded by two, and size is measured by counting gates. Given a function $f : F^n \to F^m$, $f = (f_1, \ldots, f_m)$, define $L_f : F^{n+m} \to F$ by

$$L_f(x_0, \ldots, x_{n-1}, a_1, \ldots, a_m) = \sum_{j=1}^{m} a_j f_j(x_0, \ldots, x_{n-1}).$$

Define the *linear combination complexity* of f, denoted by subscripting "lc" to a complexity measure, to be the complexity of L_f in that measure. Thus $s^{bc}_{lc}(f)$ denotes $s^{bc}(L_f)$. For unbounded-coefficient circuits, applying the standard Baur-Strassen Derivative Lemma [BS82] to L_f shows that $s_{lc}(f)$ and $s(f)$ are asymptotically equivalent. Pudlák's example $g = 2^n \sum_{j=0}^{n-1} x_j y_j$ shows that s^{bc} cannot be preserved, since $\partial g = (2^n x_0, \ldots, 2^n y_{n-1})$ requires $\Omega(n^2)$ size for bc-circuits.

Note, however, that $L_{\partial g}$ has $O(n)$-size bc circuits that form $a_0 x_0 + \cdots + a_{2n-1} y_{n-1}$ and then add this to itself iteratively n times. Thus $s^{bc}_{lc}(\partial g) = O(n)$, which separates the lc-bc model from the bc-model. Moreover, ∂g has $O(n)$ size in the "inhomogeneous" bc-circuit model, which allows the circuits a separate constant-1 input and allows them to build up higher constants by iteratively adding this constant to itself. Inspection of standard proofs of the Derivative Lemma shows that this example is basically the worst possible thing that can happen:

Proposition 2. *For any nontrivial function f, $s^{bc}_{lc}(\partial f) = O(s^{bc}(f))$, and $s^{bc}_{lc}(f)$ equals the minimum size of bc circuits with constant-1 input that are allowed to multiply their output gates by built-up constants.* □

The equivalences apply to linear/bilinear circuits modulo charging for the multiplications by the linear-combining variables. Now we observe that in contrast to s^{bc} complexity, the Derivative Lemma can be adapted to preserve s^{bc}_{lc} complexity, in a form that also allows iteration to higher partial derivatives (fixing $b_0 = 0$ computes $L_{\partial f}$ exactly). The proof is in the Appendix.

Theorem 12. *Given a bc-circuit of size s computing $f = (f_1, \ldots, f_m)$ in variables $x_0, \ldots, x_{n-1}$, we can construct a bc-circuit of size at most $5s$ with extra inputs $b_0, b_1, \ldots, b_n$ and output gates for $j = 1, \ldots, m$ each of which computes*

$$b_0 f_j + \sum_{i=1}^{n} b_i \frac{\partial f_j}{\partial x_i}.$$

6 Size-Depth Tradeoffs for Linear Combination Circuits

In this section we consider bilinear lc-circuits of the following structure. There are three sets of input vectors namely x, y and special interpolation inputs z. There are two top-level mappings computing separately for input vectors x and y. Both these mapping are computed by depth-d circuits, where the length of each path from variable to linear form is exactly d. Multiplication gates are allowed, but are restricted to have exactly one of it's inputs taken to be a z variable. Say the outputs of these circuits are $l_1(x), ..., l_k(x)$ and $r_1(y), \ldots, r_k(y)$. (These are actually bilinear, but we drop the z in the notation, because we want to think of these as linear forms, when z is fixed). Then there are k multiplication gates computing $m_i = \ell_i(x) r_i(y)$ for $1 \leq i \leq k$. Finally there is a bottom layer with a single output. This layer is restricted to be formula consisting of addition gates only. Constants on the wires are assumed to have norm at most one.

We identify a bilinear form $p(x, y)$ on $n + n$ variables in a natural way with the $n \times n$ matrix of coefficients $(p)_{ij} =$ coefficient of monomial $x_i y_j$. Linear forms $\ell_i(x)$ and $r_i(y)$ are identified with row vectors. So multiplication gate m_i computes $\ell_i^T r_i$, when z is fixed. The function computed by the circuit is required to be of the form $\sum_{k=1}^{m} z_k (x^T A_k y)$, for certain $n \times n$ matrices A_k. In this situation, we say the circuit is an lc-circuit for computing $A_1, \ldots, A_m$.

Definition 8 ([Lok01]). *Given $1 \leq r \leq n$ and an $n \times m$ matrix A, define its L_2-r-rigidity to be $\Delta_r^2 = min\{||A - B||_F^2 : B$ is an $n \times m$ matrix of rank at most $r\}$, where $||A - B||_F$ denotes the Frobenius norm.*

Theorem 13. *Let C be a depth-d lc-circuit of structure as defined above that computes $A_1, \ldots, A_n$. Then for $1 \leq r \leq n$, the number of wires of C is at least*

$$r \left(\min_{|I|=n-r} \sum_{i \in I} \Delta_r^2(A_i) \right)^{\frac{1}{2d+1}} n^{\frac{-2}{2d+1}}.$$

The proof is in the Appendix. The above theorem yields lower bounds whenever the bilinear forms that are computed have associated matrices of high L_2-r-rigidity. For example:

Corollary 1. *Let $n = 2^i$ and let $A_1, \ldots A_n$ be a set of n Hadamard matrices. Then any depth d bilinear lc-circuit, of the structure defined above, that computes $A_1, \ldots, A_n$ has size $\Omega(4^{-2d-1} n^{1+\frac{1}{2d+1}})$.*

Proof. It is well-known that for a Hadamard matrix H, $\Delta_r^2(H) \geq n(n-r)$. Applying the above theorem with $r = n/2$ yields the corollary. □

Finally, let us note that because the Frobenius norm is invariant under unitary transformations, the results in this section extend to an orbit model of the form $C(U_1 x, U_2 y, z)$, where U_1 and U_2 are unitary matrices and C is a depth d lc-circuit restricted as before.

7 Conclusions

We have introduced two "bridging models" and shown that some lower bounds for stricter models extend to them. We have separated them from the stricter models, and have noted that the modifications of spectral methods made recently by Raz [Raz02] and Bürgisser and Lotz [BL03] may be powerless in our models. The natural first question to ask is, can they be separated from the arbitrary-coefficients case, or are they equivalent to it? For the latter, "linear combination" model, equivalence may follow if one can prove a good bound $g(s,d)$ such that for every linear circuit C of size s and depth d computing the mapping of an $n \times n$ matrix A in the standard basis, there is a circuit of size s and depth d (or $O(s), O(d)$) computing A with constants bounded in magnitude by $g(s,d) \max_{i,j} |A_{ij}|$. However, even then it is not clear how one can scale such constants down to bounded and restore the original values of the circuit by a single multiplication with a constant of magnitude 2^s at the output gates. Certainly we have opened new combinatorial ground for further progress. Our iterable Derivative Lemma may also be useful for work involving higher derivatives.

References

[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.

[BL03] P. Bürgisser and M. Lotz. Lower bounds on the bounded coefficient complexity of bilinear maps. *J. Assn. Comp. Mach.*, 2003. to appear; also at arXiv.org/cs/0301016.

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

[Bür00] Peter Bürgisser. Cook's versus Valiant's hypothesis. *Theor. Comp. Sci.*, 235:71–88, 2000.

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

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

[Koi96] Pascal Koiran. Hilbert's Nullstellensatz is in the polynomial hierarchy. *Journal of Complexity*, 12(4):273–286, December 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.

[NW95] Noam Nisan and Avi Wigderson. On the complexity of bilinear forms. In *Proc. 27th Annual ACM Symposium on the Theory of Computing*, pages 723–732, 1995.

[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.

[SW99] A. Shpilka and A. Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. Technical Report 23, ECCC, 1999.

A Proofs from Section 3

Proof of Lemma 4.

We can write

$$\mathit{Circ}(Ea) = F_n \mathrm{diag}(\lambda_0, \ldots, \lambda_{n-1}) F_n^{-1} \;\text{ where }\; (\lambda_0, \ldots, \lambda_{n-1})^T = \mathit{DFT}_n Ea.$$

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

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

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 $W = F_n EV$. Then $\alpha = F_n Ea = F_n EV\beta = W\beta$.

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

$$\Sigma = E[\alpha_J \alpha_J^*] = E[W_J \beta \beta^* W_J^*] = W_J E[\beta\beta^*] W_J^* = W_J W_J^*.$$

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,BL03] yields

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

We *claim* now that $\det(V^* E^* EV) \geq \kappa^{-r}$, 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. Thus we conclude that there exists a set J such that

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

Applying Lemma 2 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, and with applying Proposition 1,

$$\mathrm{msv}_r^2(\mathit{Circ}(Ea)) \geq n^r \delta^r \kappa^{-r} \binom{n}{r}^{-1} \geq n^r \delta^r \kappa^{-r} 2^{-n} \geq n^r \delta^r \kappa^{-3r} 2^{-n}.$$

Hence applying theorem 2 we get, for some absolute constant c,

$$s_{lin}^{bc}(\mathit{Circ}(Ea)D) \geq \log \mathrm{msv}_r(\mathit{Circ}(Ea)D) - n \geq \frac{r}{2} \log n - cn. \square$$

Proof of Theorem 4.

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 4 and 3, there exists an $a \in \mathbf{C}^n$ such that:

1. $s_{lin}^{bc}(Circ(E^{-1}a)D^{-1}) \geq \frac{1}{2}r \log n - cn$, 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 3:

$$S(\Gamma) \geq s_{lin}^{bc}(f_1^T, \ldots, f_k^T) \geq (n - r) \log R - n. \tag{4}$$

So we obtain the inequality

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

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

Proof of Theorem 11.

Rewrite $LMM_{2d+1} = \sum_{i_0,\ldots,i_{2d+1} \in \{1,\ldots,n\}} a_{i_0,i_{2d+1}} x^1_{i_0,i_1} x^2_{i1,i2} \ldots x^{2d+1}_{i_{2d},i_{2d+1}}$. Consider fixed indices $i_0, \ldots, i_{2d+1}$. Taking $(d+1)$-order partials w.r. to variables $x^1_{i_0,i_1}, x^3_{i_2,i_3}, \ldots, x^{2d+1}_{i_{2d},i_{2d+1}}$ of LMM_{2d+1} yields the monomial $a_{i_0,i_{2d+1}} x^2_{i_1,i_2} x^4_{i_3,i_4} \ldots x^{2d}_{i_{2d-1},i_{2d}}$. Consider an arbitrary affine subspace A of codimension κ. Since in each matrix there are at least $n^2 - \kappa$ unassigned variables when doing the substitution corresponding to restriction to A, we conclude that there are at least $(n^2 - \kappa)^{d+1}$ choices for the indices, which produce a partial derivative that is not altered by restricting to A. Since each choice yields a different partial we conclude $dim[\partial_{d+1}(LMM_{2d+1})_{|A}] \geq (n^2 - \kappa)^{d+1})$. Taking $\kappa = n^{\frac{2d+2}{d+2}}$ in theorem 5 yields the theorem. □

B Proof of Theorem 12

We use induction on the number of gates r other than the outputs. The base case is when $r = 0$. In this case each f_j is a gate taking both inputs directly from the input variables, $s = m$ and the theorem follows readily. Suppose $r > 0$. Let h be a gate taking both inputs from the variables. Let Γ' be the circuit obtained from Γ by replacing h with a new variable x_{n+1}. That is, whenever there is a wire from h to a gate, have the same wire (with identical constant) to that gate from x_{n+1}, and finally remove h. Say the new circuit computes $(f_1', \ldots, f_m')$. By induction, we obtain a bounded coefficient circuit Γ'' with inputs $x_1, \ldots, x_{n+1}$ and $b_0, \ldots, b_{n+1}$ computing

$$b_0 f_j' + \sum_{i=1}^{n+1} b_i \frac{\partial f_j'}{\partial x_i},$$

for all $j = 1 \ldots m$ of size at most $5(s-1)$. Note that for each i, $f'_i[x_{n+1} \leftarrow h] = f_i$. The chain rule gives us the follow equality for any $j = 1 \ldots m$ and $k = 1 \ldots n$,

$$\frac{\partial f_j}{\partial x_k} = \frac{\partial f'_j}{\partial x_k}[x_{n+1} \leftarrow h] + \frac{\partial f'_j}{\partial x_{n+1}}[x_{n+1} \leftarrow h] \cdot \frac{\partial h}{\partial x_k}$$

Let Γ''' be the circuit obtained from Γ'' by replacing the input variable x_{n+1} with the gate h. That is, whenever there is a wire from x_{n+1} to a gate have exactly the same wire (with identical constant) from h to that gate, and finally remove x_{n+1}. We see that Γ''' has a gate g_j computing

$$g_j = b_0 f'_j[x_{n+1} \leftarrow h] + \sum_{i=1}^{n+1} b_i \frac{\partial f'_j}{\partial x_i}[x_{n+1} \leftarrow h],$$

for $j = 1 \ldots m$. Hence we obtain the required circuit by substituting $b_{n+1} \leftarrow \sum_{i=1}^{n} b_i \frac{\partial h}{\partial x_i}$, since

$$\begin{aligned} g_j[b_{n+1} \leftarrow \sum_{i=1}^{n} b_i \frac{\partial h}{\partial x_i}] &= b_0 f'_j[x_{n+1} \leftarrow h] + \sum_{i=1}^{n} b_i \frac{\partial f'_j}{\partial x_i}[x_{n+1} \leftarrow h] + \sum_{i=1}^{n} b_i \frac{\partial h}{\partial x_i} \cdot \frac{\partial f'_j}{\partial x_{n+1}}[x_{n+1} \leftarrow h] \\ &= b_0 f_j + \sum_{i=1}^{n} b_i \frac{\partial f_j}{x_i}. \end{aligned}$$

for any $j = 1 \ldots m$. The substitution for b_{n+1} can be done by adding at most 3 gates. That is, in case $h = \alpha x_i + \beta x_j$, we substitute $\alpha b_i + \beta b_j$, which takes one gate. In case $h = \alpha x_i \cdot \beta x_j$, we substitute $\alpha\beta b_i x_j + \alpha\beta b_j x_i$, which takes 3 gates. In both cases constants on the wires are 1 or constants from Γ. We conclude that Γ''' has size at most $5(s-1) + 4 \leq 5s$ and that it is a bc-circuit. □

With $m = 1$ this gives $s^{bc}_{lc}(f, \partial(f)) = O(s^{bc}(f))$. We can also obtain the "transposed" theorem:

Theorem 14. *Given a bounded coefficient circuit Γ computing $f_1, \ldots, f_m$ at (non-input) gates of fanout zero in variables $x_1, \ldots, x_n$ of size s, we can construct a bounded-coefficient circuit of size at most $5s$ with extra inputs $b_1, \ldots, b_m$ computing $\sum_{i=1}^{m} b_i f_i$ and $\sum_{i=1}^{m} b_i \frac{\partial f_i}{\partial x_j}$, for all $j = 1 \ldots n$, whenever these are not identically zero.*

C Proof of Theorem 13

Let C be given as indicated. Fix $1 \leq r \leq n$. Let S equal the number of wires of C. We call a gate g special if the number of wires leaving g is at least S/r. Note there can be at most r special gates. No gate in the layer below the multiplication gates can be special. There are at least $n - r$ variables z_i that are not special. We now will consider what happens to corresponding outputs A_i as we remove a special gate g. That is, temporarily fix $z_i = 1$ and $z_k = 0$ for $k \neq i$ and then remove g.

case 1: g is an input variable x_j.
In this case we remove the wires fanning out from x_j. That means $\ell_i^{new} = \ell_i$ with jth entry set to zero. Hence $m_i^{new} = m_i$ with row j zeroed out. Since each output A_i s simply a linear combination of the matrices m_i, we get $A_i^{new} = A_i$ with jth row zeroed out, i.e. A_i gets modified by subtracting a rank-≤ 1 matrix.

case 2: g is an input variable y_j.
Similarly as above we can conclude each output gets modified by subtracting a rank-≤ 1 matrix.

case 3: g is a multiplication gate $m_i = \ell_i^T r_i$.
Each output gets modified by subtracting a scalar multiple of m_i. Observe that $rank(m_i) \leq 1$, hence each output gets again modified by subtraction of a rank-≤ 1 matrix.

case 4: g is an addition gate linear in x.
Suppose gate g computes the linear form l. Then $\ell_i^{new} = \ell_i - \gamma_i l$, for certain scalars γ_i. Hence $m_i^{new} = (\ell_i^{new})^T r_i = \ell_i^T r_i - \gamma_i l^T r_i$. Since $A_i = \Sigma_{j=1}^k \alpha_j m_j$, we get that

$$\begin{aligned} A_i^{new} &= \Sigma_{j=1}^k \alpha_j m_j^{new} \\ &= \Sigma_{j=1}^k \alpha_j m_j - \gamma_j l^T r_j \\ &= A_i - l^T \Sigma_{j=1}^k \alpha_j \gamma_j r_j. \end{aligned}$$

Observe that $l^T \Sigma_{j=1}^k \alpha_j \gamma_j r_j$ has rank at most 1. Hence again we have that each output is modified by a rank at most 1 matrix.

case 5: g is an addition gate linear in y.
Similarly as case 4, we can show that each output get modified by subtracting a rank at most 1 matrix.

case 6: g is a multiplication gate with input z_j.
If $j \neq i$, then A_i is unaltered. Otherwise, this case reduces to case 4 or 5.

Let C' be the circuit obtained by consecutively removing all special gates. From the above we conclude C' computes $A_{i1} - B_{i1}$, ..., $A_{i_{n-r}} - B_{i_{n-r}}$, where each B_{i_t} has rank at most r. Wlog we assume $i_1 = 1, i_2 = 2$, etc. The fanout of each gate in C' is at most S/r. We are now going to estimate the following quantity, which is the sum of norms of all entries of the computed matrices:

$$\Phi = \sum_{m=1}^{n-r} \sum_{i=1}^{n} \sum_{j=1}^{n} |(A_m - B_m)_{ij}|^2.$$

For a given triple (x_i, y_j) there are at most $(S/r)^{2d}$ paths starting in x_i and y_j that come together in the same multiplication gate. Then from that gate there are at most S/r choices. Then there is a fixed path to the output. On these paths there can be at most one multiplication gate with z-input. Hence there are at most $(S/r)^{2d+1}$ path starting in x_i, y_j and going to an output via the same multiplication gate. Φ can be computed by summing over all such paths the product of the constants on the wires. Since each constant has norm at most 1, we conclude each such path contributes at most 1 to Φ. Hence

$$\Phi \leq n^2 (S/r)^{2d+1}.$$

Thus

$$S \geq r \Phi^{\frac{1}{2d+1}} n^{\frac{-2}{2d+1}}.$$

Observe that

$$\Phi = \sum_{m=1}^{n-r} ||A_m - B_m||_F^2 \geq \min_{|I|=n-r} \sum_{i \in I} \Delta_r^2(A_i),$$

from which the theorem readily follows. □